

# PARTE 01

MÉTODOS Y  
APLICACIONES DE DATOS

# PARTE 01

MÉTODOS Y  
APLICACIONES DE DATOS

# PARTE 02

MARCO DEL  
PROYECTO DE DATOS

# PARTE 02

MARCO DEL  
PROYECTO DE DATOS

# ANÁLISIS DE DATOS Y SERVICIOS FINANCIEROS DIGITALES

# M A N U A L

---

# AGRADECIMIENTOS

IFC y The Partnership for Financial Inclusion de Mastercard Foundation desean reconocer el generoso apoyo de las instituciones que participaron en los estudios de caso para este manual. Airtel Uganda, Commercial Bank of Africa, FINCA República Democrática del Congo, First Access, Juntos, Lenddo, MicroCred, M-Kopa, Safaricom, Tiixa, Tigo Ghana, y Zoon. Este manual no hubiera sido posible sin la participación de estas instituciones.

IFC y Mastercard Foundation desean dar un agradecimiento especial a los autores Dean Caire, Leonardo Camiciotti, Soren Heitmann, Susie Lonie, Christian Racca, Minakshi Ramji, y Qiuyan Xu, así como a los revisores y colaboradores: Sinja Buri, Tiphaine Crenn, Ruth Dueck-Mbeba, Nicolais Guevara, Joseck Mudiri, Riadh Naouar, Laura Pippinato, Max Roussinov, Anca Bogdana Rusu, Matthew Saal y Aksinya Sorokina. Por último, los autores desean extender un agradecimiento especial a Anna Koblanck y Lesley Denyes por su extenso apoyo editorial.

Número ISBN: 978-0-620-76146-8

Primera Edición 2017

# MANUAL

# ANÁLISIS DE DATOS Y SERVICIOS FINANCIEROS DIGITALES

---



# PRÓLOGO

Este es el tercer manual sobre Servicios Financieros Digitales producido y publicado por The Partnership for Financial Inclusion, una iniciativa conjunta de IFC y Mastercard Foundation para expandir las microfinanzas y hacer avanzar los Servicios Financieros Digitales en África Subsahariana. El primer manual de la serie, *Manual de Canales de Distribución y Tecnología Alternativos*, ofrece una guía completa acerca de los componentes de la tecnología financiera digital, con especial atención en los elementos básicos de hardware y software, para un despliegue exitoso. El segundo *Manual Servicios Financieros Digitales y Gestión de Riesgo*, es una guía sobre los riesgos asociados al dinero electrónico y corresponsales bancarios, además ofrece un marco para la gestión de dichos riesgos. El presente manual tiene el objetivo de proporcionar apoyo y orientación útil sobre cómo aplicar la analítica de datos para expandir y mejorar la calidad de los servicios financieros.

Este manual está diseñado para cualquier tipo de proveedor de servicios financieros que ofrezca o tenga la intención de ofrecer Servicios Financieros Digitales. Los proveedores de Servicios Financieros Digitales incluyen todo tipo de instituciones, tales como instituciones de microfinanzas, bancos, operadores de redes móviles, empresas de fintech y proveedores de servicios de pago. Los canales, productos y procesos habilitados por la tecnología generan datos muy valiosos sobre las interacciones con los clientes; al mismo tiempo, se pueden habilitar vínculos con fuentes de datos externos que están cada

vez más disponibles. El manual ofrece un resumen de los conceptos básicos e identifica las tendencias de uso en el mercado, y también ilustra una gama de aplicaciones prácticas y casos de proveedores de Servicios Financieros Digitales que están traduciendo sus propios datos o datos externos en información para sus negocios. También ofrece un marco guía de proyectos de datos para proveedores de SFD que deseen aprovechar los insights de datos con el fin de satisfacer mejor las necesidades de los clientes, y mejorar las operaciones, los servicios y los productos. El manual se entiende como una cartilla sobre datos y análisis de datos, y no presupone conocimiento previo de ninguno. Sin embargo, se espera que el lector entienda los Servicios Financieros Digitales, y esté familiarizado con los productos, la función de los corresponsales, aspectos de la gestión operativa y el papel de la tecnología. El manual está organizado de la siguiente manera:

**Introducción:** Presenta el manual y establece, a grandes rasgos, la plataforma y las definiciones para Servicios Financieros Digitales y analítica de datos.



## Parte 1: Métodos y Aplicaciones de Datos

**Capítulo 1.1:** Habla de la ciencia de datos en el contexto de Servicios Financieros Digitales y proporciona una visión general de los tipos de datos, fuentes y metodologías, y herramientas utilizadas para derivar información a partir de datos.

**Capítulo 1.2:** Describe cómo aplicar el análisis de datos a Servicios Financieros Digitales. El capítulo resume las técnicas utilizadas para obtener información sobre el mercado a partir de datos y describe el papel que pueden desempeñar los datos en la mejora de la gestión operativa de los Servicios Financieros Digitales. El capítulo incluye ejemplos de la vida real y estudios de caso de las lecciones aprendidas por los expertos en el campo. Finaliza con una descripción de cómo los expertos pueden utilizar datos para desarrollar algoritmos basados en los modelos de scoring crediticio para la inclusión financiera.

## Parte 2: Marco del Proyecto de Datos

**Capítulo 2.1:** Ofrece un marco para la implementación de proyectos de datos y una guía paso a paso para resolver problemas prácticos de negocios mediante la aplicación de este marco, para obtener valor de fuentes de datos existentes y potenciales.

**Capítulo 2.2:** Proporciona un directorio de fuentes de datos y recursos tecnológicos, así como una lista de métricas de desempeño para evaluar proyectos de datos. También incluye un glosario que ofrece descripciones de los términos utilizados en esta guía y en la práctica de la industria.

**Conclusiones:** Incluye las lecciones aprendidas de los proyectos de datos a la fecha, basadas en la experiencia de IFC en África subsahariana con el programa The Partnership for Financial Inclusion de Mastercard Foundation.

# CONTENIDO

<b>PRÓLOGO</b>	<b>4</b>
<b>ACRÓNIMOS</b>	<b>7</b>
<b>RESUMEN EJECUTIVO</b>	<b>10</b>
<b>INTRODUCCIÓN</b>	<b>14</b>
<b>PARTE 1: MÉTODOS Y APLICACIONES DE DATOS</b>	<b>16</b>
Capítulo 1.1: Datos, Analíticas y Métodos.....	16
Definición de Datos	16
Fuentes de Datos	19
Protección de Datos y Protección del Cliente	23
Ciencia de los Datos: Introducción	26
Métodos	29
Herramientas	32
Capítulo 1.2: Aplicaciones de Datos Para Proveedores de Servicios Financieros Digitales .....	34
1.2.1 Analítica y Aplicaciones: Conocimiento de Mercado	36
1.2.2 Analítica y Aplicaciones: Gestión de Operaciones y Desempeño	54
1.2.3 Analítica y Aplicaciones: Scoring crediticio	79

<b>PARTE 2: MARCO DEL PROYECTO DE DATOS</b>	<b>100</b>
Capítulo 2.1: Gestión de un Proyecto de Datos.....	100
El Data Ring	100
Estructuras y Diseño	102
OBJETIVO (S)	104
Cuadrante 1: HERRAMIENTAS	107
Cuadrante 2: HABILIDADES	112
Cuadrante 3: PROCESO	117
Cuadrante 4: VALOR	124
APLICACIÓN: Uso del Data Ring	126
Capítulo 2.2 Recursos.....	136
2.2.1 Resumen de las Clasificaciones de Casos de Uso Analíticos	136
2.2.2 Directorio de Fuentes de Datos	137
2.2.3 Métricas para la Evaluación de Modelos de Datos	141
2.2.4 Data Ring y Data Ring Canvas	141
<b>CONCLUSIONES Y LECCIONES APRENDIDAS</b>	<b>145</b>
<b>GLOSARIO</b>	<b>149</b>
<b>BIOGRAFÍAS DE LOS AUTORES</b>	<b>156</b>

---

# ACRÓNIMOS

<b>IA</b>	Inteligencia Artificial
<b>ALA</b>	Anti Lavado de Activos
<b>API</b>	Interfaz de Programación de Aplicaciones
<b>ARPU</b>	Promedio de Ingresos por Usuario
<b>ATM</b>	Cajero Automático
<b>BI</b>	Inteligencia de Negocios
<b>CBA</b>	Banco Comercial de África
<b>CBS</b>	Sistema de Core Bancario
<b>CD</b>	Canales de Distribución Alternativos
<b>CDO</b>	Chief Data Officer [Director de Datos]
<b>CDR</b>	Registros Detallados de Llamadas
<b>CFT</b>	Contra la Financiación del Terrorismo
<b>CGAP</b>	Consultative Group to Assist the Poor
<b>COT</b>	Comisión Sobre Transacción
<b>CRISP-DM</b>	Cross Industry Standard Process para la Minería de Datos
<b>CRM</b>	Customer Relationship Management
<b>CSV</b>	Valores Separados por Comas
<b>BD</b>	Base de Datos
<b>ETL</b>	Extraer - Transformar - Cargar
<b>IF</b>	Institución Financiera
<b>FTC</b>	Comisión Federal de Comercio

<b>MLG</b>	Modelo Lineal Generalizado
<b>GPS</b>	Sistema de Posicionamiento Global
<b>GSM</b>	Sistema Global para Comunicaciones Móviles
<b>GSMA</b>	Asociación del Sistema Global para las Comunicaciones Móviles
<b>TIC</b>	Tecnologías de Información y Comunicación
<b>ID</b>	Documento de Identificación
<b>IFC</b>	Corporación Financiera Internacional
<b>IMF</b>	Instituciones Microfinancieras
<b>PI</b>	Propiedad Intelectual
<b>TI</b>	Tecnología Informática
<b>JSON</b>	Notación de Objetos JavaScript
<b>KCB</b>	Banco Comercial de Kenya
<b>KPI</b>	Indicador Clave de Desempeño
<b>KRI</b>	Indicador Clave de Riesgo
<b>KYC</b>	Conozca a su Cliente
<b>MEL</b>	Monitoreo, Evaluación y Aprendizaje
<b>MiPyMe</b>	Micro, Pequeñas y Medianas Empresas
<b>MNO</b>	Operador de Telefonía Móvil
<b>NDA</b>	Acuerdo de Confidencialidad
<b>NPL</b>	Préstamos en Mora
<b>NU</b>	Naciones Unidas
<b>OLA</b>	Acuerdo de Nivel Operacional
<b>OPS</b>	Operaciones del Proyecto
<b>OTC</b>	Servicios Financieros en Mostrador
<b>P2P</b>	Persona a Persona

<b>CeR</b>	Cartera en Riesgo
<b>PBAX</b>	Private Branch Automatic Exchange
<b>PIN</b>	Número de Identificación Personal
<b>PMV</b>	Producto Mínimo Viable
<b>PLN</b>	Procesamiento del Lenguaje Natural
<b>POS</b>	Punto de Venta
<b>PYME</b>	Pequeña y Mediana Empresa
<b>QA</b>	Garantía de Calidad
<b>RCT</b>	Prueba Aleatoria Controlada
<b>RDC</b>	República Democrática del Congo
<b>RFP</b>	Solicitud de Propuesta
<b>SFD</b>	Servicios Financieros Digitales
<b>SIG</b>	Sistema de Información Gerencial
<b>SIM</b>	Módulo de Identidad del Suscriptor
<b>SLA</b>	Acuerdo de Nivel de Servicio
<b>SMS</b>	Servicio de Mensajes Cortos
<b>SNA</b>	Análisis de Redes Sociales
<b>SOC</b>	Sistema de Originación de Créditos
<b>SQL</b>	Lenguaje de Consultas Estructuradas
<b>SVM</b>	Máquina de Vector de Soporte
<b>SVN</b>	Red de Vector de Soporte
<b>TCP</b>	Protocolo de Control de Transmisión
<b>TPS</b>	Transacciones por Segundo
<b>UE</b>	Unión Europea
<b>USSD</b>	Servicio Suplementario de Datos no Estructurados

---

# Resumen Ejecutivo



“Deje que el conjunto de datos cambie su forma de pensar.”

– Hans Rosling

IFC apoya a las instituciones que buscan desarrollar Servicios Financieros Digitales para aumentar la inclusión financiera, y está dedicada a múltiples iniciativas en una gama de mercados a través de su portafolio de inversiones y proyectos de asesoría. Desde 2017, a través de su trabajo con The MasterCard Foundation y otros aliados, IFC trabaja con proveedores de SFD en África subsahariana en la expansión de la inclusión financiera a través de productos y servicios digitales. Las interacciones con clientes, así como la industria en general en la región y más allá, han identificado la necesidad de una guía sobre cómo utilizar el campo emergente de la ciencia de datos para desatar el valor de la data que sale de estas implementaciones. A pesar que el análisis de datos ofrece a los proveedores de SFD la oportunidad de conocer a sus clientes a un nivel granular y utilizar estos conocimientos para ofrecer servicios de mayor calidad, muchos expertos aún no han implementado un enfoque sistemático orientado a los datos en sus operaciones y organizaciones. Hay algunos ejemplos que han recibido mucha atención por su éxito en ciertos mercados, como la incorporación de datos alternativos para evaluar el riesgo de crédito de nuevos tipos de clientes. Sin embargo, lo prometedor de los datos va más allá de una o dos aplicaciones en casos específicos. Las barreras comunes en la aplicación de la información o ideas proporcionadas por los datos en los SFD incluyen la falta de conocimiento, la falta de destreza y sentirse incómodos con un enfoque desconocido. Este manual busca proporcionar una perspectiva de la oportunidad de los datos para impulsar la inclusión financiera, junto con las medidas que los expertos pueden tomar para comenzar a adoptar un enfoque basado en datos en sus negocios y diseñar proyectos basados en datos para resolver problemas prácticos de negocio.

En la última década, los Servicios Financieros Digitales han transformado la oferta al cliente y el modelo de negocio del sector financiero, especialmente en los países en desarrollo. Un gran número de personas de bajos ingresos, micro empresarios, pequeñas empresas y poblaciones rurales que antes no tenían acceso a servicios financieros formales, ahora están digitalmente bancarizados por medio de una serie de proveedores de servicios financieros antiguos y nuevos, incluyendo–proveedores no tradicionales tales como operadores telefónicos (MNOs en inglés) y fintechs emergentes. Esto ha mostrado un impacto en la calidad de vida, como se ilustra en Kenia, donde un estudio realizado por investigadores del Instituto de Tecnología de Massachusetts (MIT) ha demostrado que la introducción de servicios financieros tecnológicos puede ayudar a reducir la pobreza.<sup>1</sup>

---

<sup>1</sup> Suri y Jack, 'The Long Run Poverty and Gender Impacts of Mobile Money', *Science* Vol. 354, Issue 6317 (2015): 1288–1292.

El estudio estima que desde 2008, el acceso a servicios de dinero electrónico que permiten a los usuarios almacenar e intercambiar dinero, aumentó los niveles de consumo per cápita diario de 194.000 personas, o aproximadamente el dos por ciento de los hogares kenianos, sacándolos de esta manera de la extrema pobreza. El impacto fue más destacado entre los hogares encabezados por mujeres, a menudo considerados particularmente marginados económicamente. Este es un buen argumento para una inclusión financiera más amplia y profunda en el África Subsahariana y otras economías emergentes. Los datos y el análisis de datos pueden ayudar a lograr esto.

Se estima que aproximadamente 2.5 quintillones de bytes de datos se producen cada día.<sup>2</sup> Para hacerse una idea de esta cantidad, este volumen de datos llenaría aproximadamente 10 millones de discos Blu-ray. La mayoría de estos datos son jóvenes - el 90 por ciento de la data existente en el mundo se creó en los últimos dos años.<sup>3</sup> La reciente revolución digital de datos se extiende tanto al mundo en desarrollo como al mundo desarrollado. En 2016, había 7.8 billones de suscripciones

de teléfonos móviles en el mundo, de los cuales el 74 por ciento estaban en países en desarrollo.<sup>4</sup> Se espera que el futuro sea aún más rico en datos. A medida que los costos de los teléfonos inteligentes caen, el acceso a Internet móvil aumentará del 44 por ciento en 2015 al 60 por ciento en 2020. En África subsahariana, se prevé que el uso de teléfonos inteligentes aumentará del 25% en 2015 al 50% de todas las conexiones en 2020.<sup>5</sup> Los objetos cotidianos también están cada vez más habilitados para enviar y recibir datos, conectándose y comunicándose directamente entre sí y a través de la interfaz de usuario en aplicaciones de teléfonos inteligentes, esto es conocido como el Internet de las Cosas.<sup>6</sup> Si bien este es principalmente un fenómeno de los países desarrollados, también hay ejemplos del mundo en desarrollo. En África oriental, por ejemplo, hay dispositivos solares que producen información sobre el uso de la unidad y los reembolsos de Servicios Financieros Digitales realizados por el propietario. Los datos se utilizan para realizar evaluaciones de crédito instantáneas que, en última instancia, pueden impulsar nuevos negocios. Para los proveedores de Servicios Financieros Digitales, los datos se

pueden extraer de una gama cada vez más amplia de fuentes: datos transaccionales, registros de llamadas móviles, grabaciones de call center, vinculaciones de clientes y corresponsales, patrones de compra de tiempo aire, información de buró de crédito, publicaciones en redes sociales y datos geoespaciales.

Estas fuentes emergentes de datos tienen la capacidad de impactar positivamente en la inclusión financiera. La analítica puede mejorar los procesos de negocio de las instituciones que sirven a hogares de bajos ingresos al permitirles identificar y atraer a nuevos clientes de manera más eficiente. Así, los datos pueden ayudar a las instituciones financieras a vincular personas nuevas y previamente excluidas. También profundiza la inclusión financiera ya que los clientes existentes aumentan su uso de productos financieros. Al mismo tiempo, los encargados de la formulación de políticas y otras partes interesadas del sector público pueden ahora obtener una visión detallada de la inclusión financiera examinando el acceso, el uso y otras tendencias. Esta evidencia puede desempeñar un papel en el desarrollo de políticas y estrategias futuras para mejorar la inclusión financiera.

<sup>2</sup> 'The 4 Vs of Big Data', IBM Big Data Hub, consultado el 3 de abril de 2017, <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

<sup>3</sup> 'The 4 Vs of Big Data', IBM Big Data Hub, consultado el 3 de abril de 2017, <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>

<sup>4</sup> 'The Mobile Economy 2017', GSMA Intelligence

<sup>5</sup> 'Global Mobile Trends', GSMA Intelligence

<sup>6</sup> *Internet of Things*. In Wikipedia. La enciclopedia libre, consultada el 3 de abril de 2017, [https://en.wikipedia.org/w/index.php?title=Internet\\_of\\_things&oldid=773435744](https://en.wikipedia.org/w/index.php?title=Internet_of_things&oldid=773435744)

La mayor disponibilidad de datos presenta desafíos y oportunidades. El mayor desafío es cómo aprovechar su utilidad mientras se garantiza la privacidad de las personas. Una gran proporción de nuevos datos disponibles se producen de forma pasiva como resultado de nuestras interacciones con servicios digitales, tales como teléfonos móviles, búsquedas en Internet, compras en línea y transacciones almacenadas electrónicamente. Las características de los individuos se pueden inferir a partir de los algoritmos complejos que hacen uso de estos datos, hecho posible debido a los avances en capacidad analítica. Así, la privacidad se ve más comprometida por el hecho de que los generadores primarios de datos no son conscientes de los datos que están generando y las formas en que pueden ser utilizados. Como tal, las empresas y los actores del sector público deben poner en práctica salvaguardas adecuadas para proteger la privacidad. Debe haber políticas y marcos legales claros tanto a nivel nacional como internacional que protejan a los productores de datos de los ataques de los hackers y las exigencias de los gobiernos, al tiempo que estimulan la innovación en el uso de datos para mejorar los productos y servicios. Igualmente, a nivel institucional deben establecerse políticas claras que regulen la aceptación y el rechazo de los usuarios para el uso de datos, extracción de datos, reutilización de datos por terceros, la transferencia y la difusión.

El uso de datos es relevante a lo largo del ciclo de vida de un cliente para de obtener un entendimiento más profundo de sus necesidades y preferencias. A grandes rasgos existen tres aplicaciones, para datos en Servicios Financieros Digitales: desarrollar conocimiento de mercado, mejorar la gestión operativa y el scoring crediticio. Este manual hace un extenso uso de estudios de caso para demostrar el uso de analítica de datos para los profesionales. En particular, el universo de datos está en constante expansión y las capacidades analíticas también están mejorando con el aumento de la capacidad tecnológica. Como tal, el potencial para el uso de datos se extiende mucho más allá de las aplicaciones descritas en este manual.

El desarrollo de un conocimiento de mercado basados en datos es clave para desarrollar un negocio orientado al cliente. Comprender los mercados y los clientes a un nivel granular permitirá a los expertos mejorar los servicios al cliente y resolver sus necesidades más importantes, liberando así el valor económico. Un negocio orientado al cliente entiende sus necesidades y deseos, asegurando que los procesos internos y externos, las iniciativas de marketing y la estrategia de producto, son el resultado de la ciencia de datos que promueve la fidelidad del cliente. Desde la perspectiva de las operaciones, los datos desempeñan un papel importante

en la automatización de los procesos y la toma de decisiones, permitiendo que las instituciones se vuelvan escalables rápida y eficientemente. Aquí los datos también juegan un papel importante en el monitoreo del desempeño y proporcionan información sobre cómo se puede mejorar. Por último, el uso extendido de Internet y teléfonos móviles son fuentes de nuevos datos, lo que permite a los proveedores de Servicios Financieros Digitales hacer una evaluación de riesgo más precisa de personas previamente excluidas que no tienen historial financiero formal para apoyar sus solicitudes de créditos.

El manual describe los pasos que los expertos pueden tomar para comprender los elementos esenciales necesarios para diseñar un proyecto de datos y aplicarlo en sus propias instituciones. Se introducen dos herramientas para guiar a los directores de proyectos por estos pasos: El Data Ring y como complemento el Data Ring Canvas. *El Data Ring* es una lista de verificación visual, cuya forma circular centra el 'corazón' de cualquier proyecto de datos como una meta estratégica de negocio. Se discute el proceso de fijación de metas, seguido de una descripción de las categorías de recursos básicos y estructuras de diseño necesarias para implementar el proyecto. Estos elementos incluyen recursos duros, tales como la data misma, junto con herramientas de software, hardware de procesamiento y

almacenamiento; así como recursos suaves, incluyendo habilidades, experiencia en el tema y recursos humanos necesarios para la ejecución. Esta sección también describe cómo se aplican estos recursos durante la ejecución del proyecto para ajustar los resultados y entregar el valor de acuerdo con una estrategia de implementación definida.

La herramienta complementaria incorpora estos elementos de diseño estructural en un Canvas, un espacio donde los directores de proyectos pueden articular y diseñar los recursos y definiciones clave de una manera organizada e interconectada. Las herramientas ayudan a definir las relaciones interconectadas entre las estructuras de diseño del proyecto, para ver visualmente cómo las piezas se unen, para identificar dónde pueden existir las brechas o donde los requisitos de recursos necesitan un ajuste. El enfoque de Canvas también sirve como una herramienta de comunicación, proporcionando un esquema de diseño de alto nivel en una hoja de papel que puede ser actualizada y discutida durante la implementación del proyecto.

Finalmente, se proporcionan tablas de recursos. El directorio de datos enumera fuentes prominentes de datos disponibles para los expertos en Servicios Financieros Digitales y una breve descripción de su

posible aplicación en un proyecto de datos. La base de datos tecnológica enlista las herramientas esenciales en la industria de la ciencia de datos, y los productos comerciales importantes para la gestión, análisis, visualización e informes de datos. También hay una lista de métricas para evaluar modelos de datos que serían discutidos comúnmente por consultores externos o proveedores de analítica. Se pueden descargar copias de las herramientas de Data Ring para referencia o uso.

El manual hace uso extenso de estudios de caso para ilustrar las experiencias de un conjunto diverso de proveedores de SFD en la implementación de proyectos de datos dentro de sus organizaciones. Si bien estos expertos principalmente operan en África y están ofreciendo Servicios Financieros Digitales a sus clientes en forma de dinero electrónico o banca de corresponsales, esto no quiere decir que el conocimiento proporcionado por los datos no pueda ser utilizado por cualquier tipo de Proveedor de Servicios Financieros utilizando diferentes modelos de negocio. Un común denominador visto en todos estos casos es que las instituciones pueden desarrollar sistemáticamente sus capacidades de datos a partir de pequeños pasos. Convertirse en

una organización basada en los datos, con actividades competitivas impulsadas por los datos, es un viaje que requiere visión y compromiso a largo plazo. Puede requerir cambios en la cultura organizacional y mejoras en las capacidades internas existentes. Es importante destacar que las instituciones deben asegurar que los procesos a través de los cuales se recopilan, almacenan y analizan los datos respetan la privacidad individual.

Este manual tiene como objetivo brindar orientación y apoyo útil a los proveedores de SFD para ampliar la inclusión financiera y mejorar el desempeño institucional. La ciencia de los datos ofrece una oportunidad única para que los proveedores de Servicios Financieros Digitales conozcan a sus clientes, corresponsales y comerciantes, así como mejorar sus procesos internos de operación y crédito, utilizando estos conocimientos para ofrecer servicios de mayor calidad. La ciencia de los datos requiere que las empresas adopten nuevas habilidades y maneras de pensar, que pueden no ser familiares para ellos. Sin embargo, estas habilidades son aceptables y permitirán a los expertos en Servicios Financieros Digitales optimizar tanto el desempeño institucional como la inclusión financiera.

---

# Introducción

En los mercados emergentes, los individuos que antes no tenían acceso a servicios bancarios acceden cada vez más a servicios financieros formales a través de canales digitales. Se está aprovechando la capacidad de computación ubicua, la conectividad generalizada, el almacenamiento masivo de datos y las tecnologías analíticas avanzadas, para ofrecer productos y servicios financieros adaptados de manera más eficiente y más directa a una gama más amplia de clientes; en forma colectiva, estos productos y servicios se denominan Servicios Financieros Digitales. Los proveedores de SFD, es decir, las instituciones que apalancan SFD para proporcionar servicios financieros, abarcan un conjunto diverso de instituciones incluyendo Proveedores de Servicios Financieros tradicionales, tales como bancos e instituciones microfinancieras (IMF), así como Proveedores de Servicios Financieros emergentes como operadores telefónicos, fintechs y proveedores de servicios de pago.

La data o datos son términos generales que se utilizan para describir cualquier información, hecho o estadística que se ha recopilado para cualquier tipo de análisis o para fines de referencia. Los datos existen en muchas formas, como números, imágenes, texto, audio y video. Tener acceso a datos es un activo competitivo. Sin embargo, no tiene sentido sin la capacidad de interpretarla y utilizarla para mejorar la orientación al cliente, impulsar las perspectivas del mercado y extraer valor económico. Las herramientas analíticas son las que ayudan a superar la brecha entre los datos y las ideas o insights en inglés. La ciencia de los datos es el término dado al análisis de los datos, que es un proceso creativo y exploratorio que toma prestadas destrezas de muchas disciplinas incluyendo negocios, estadística e informática. Se ha definido como un campo extenso y multidimensional que utiliza matemáticas, estadística y otras técnicas avanzadas para encontrar patrones y conocimientos significativos en los datos registrados.<sup>7</sup> Las herramientas tradicionales de Inteligencia de Negocios (BI) han sido de naturaleza descriptiva, mientras que la analítica avanzada puede utilizar los datos existentes para predecir el comportamiento futuro del cliente.

La naturaleza interdisciplinaria de la ciencia de los datos implica que cualquier proyecto de datos necesita ser realizado a través de un equipo que pueda apoyarse en múltiples habilidades. Requiere apoyo de la parte técnica. Sin embargo, también requiere la participación del equipo de negocios. Como ilustra el siguiente gráfico, la conversión de datos en un valor para las empresas y la inclusión financiera, es un todo

---

<sup>7</sup> 'Analytics: What is it and why it matters?', SAS, consultado el 3 de abril de 2017, [https://www.sas.com/en\\_zh/insights/analytics/what-is-analytics.html](https://www.sas.com/en_zh/insights/analytics/what-is-analytics.html)



**Figura 1:** La Cadena de Valor de Datos

un recorrido. Entender las fuentes de datos y las herramientas analíticas es sólo una parte del proceso. Este proceso es incompleto sin contextualizar los datos firmemente dentro de las realidades de negocios del proveedor de SFD. El proveedor adicionalmente debe integrar los resultados del análisis en sus procesos de toma de decisiones.

Para los proveedores de SFD, el análisis de datos presenta una oportunidad única. Los proveedores de Servicios Financieros Digitales son particularmente activos en mercados emergentes y atienden cada vez más a clientes que pueden carecer de historias financieras formales tales como registros crediticios. Atender a estos nuevos mercados puede ser especialmente desafiante. Descubrir las preferencias y niveles de conocimiento de los nuevos tipos de clientes puede tomar más tiempo y esfuerzo. A medida que el uso de la tecnología digital y los

teléfonos inteligentes se extiende en los mercados emergentes, los proveedores de Servicios Financieros Digitales están particularmente bien posicionados para aprovechar los datos y la analítica con el fin de expandir su base de clientes y brindar un servicio de mayor calidad. La analítica de datos puede utilizarse para un objetivo específico, como el score crediticio, pero también se puede emplear de manera más general para aumentar la eficiencia operativa. Cualquiera que sea el objetivo, un proveedor de SFD basado en los datos tiene la capacidad de actuar sobre la base de la evidencia, en lugar de la observación anecdótica o reaccionando a lo que los competidores están haciendo en el mercado.

Al mismo tiempo, es importante plantear la cuestión de la protección de los consumidores y la privacidad, ya que los productores primarios de datos a menudo no saben que los datos se está recopilando,

analizando y utilizando con fines específicos. La privacidad de datos inadecuada puede resultar en robo de identidad y prácticas irresponsables de préstamos. En el contexto del crédito digital, se requieren políticas para asegurar que la gente entienda las implicaciones de los datos que están compartiendo con los proveedores de SFD y para asegurar que tengan acceso a los mismos datos que el proveedor puede consultar. Para poder desarrollar políticas, las partes interesadas, tales como proveedores, encargados de la formulación de políticas, reguladores y otros, tendrán que reunirse para discutir las implicaciones de las inquietudes sobre la privacidad, las posibles soluciones y el camino a seguir. Para los que están en el sector de la inclusión financiera, los proveedores pueden educar proactivamente a los clientes sobre cómo se recopila la información y cómo usarla, y se comprometen a sólo recoger los datos que sean necesarios sin compartir esta información con terceros.



# PARTE 1

## *Métodos y Aplicaciones de Datos*

### *Capítulo 1.1: Datos, Analíticas y Métodos*

La creciente complejidad y variedad de datos que se están produciendo ha llevado al desarrollo de nuevas herramientas analíticas y métodos para explotar estos datos con el fin de obtener información importante -insights-. La intersección de datos y su conjunto de herramientas analíticas recae ampliamente en el campo emergente de la ciencia de los datos. Para los Proveedores de Servicios Financieros Digitales que buscan aplicar enfoques basados en datos a sus operaciones, esta sección brinda los antecedentes para identificar los recursos e interpretar las oportunidades operacionales desde el punto de vista de los datos, el método científico y el conjunto de herramientas analíticas.

#### **Definición de Datos**

Los datos son muestras de la realidad, registrados como mediciones y almacenados como valores. La forma en que se clasifican los datos, su formato, estructura y fuente determinan qué tipos de herramientas pueden utilizarse para analizarlos. Los datos pueden ser cuantitativos o cualitativos. Los datos cuantitativos generalmente son bits de información que pueden medirse objetivamente, por ejemplo, los registros transaccionales. Los datos cualitativos son bits de información sobre las cualidades y generalmente son más subjetivos. Las fuentes comunes de datos cualitativos son las entrevistas, observaciones u opiniones, y estos tipos de datos se utilizan a menudo para juzgar el sentimiento o el comportamiento del cliente. Los datos también se clasifican según su formato. En el sentido más básico, esto describe la naturaleza de los datos; número, imagen, texto, voz o biométricos, por ejemplo. La digitalización de los datos es el proceso que consiste en tomar estos bits de realidad observada o medida, y representarlos con números que las computadoras entiendan. El formato de datos digitalizados describe la forma en la que se codifica digitalmente una cierta medición. Hay muchas maneras de codificar la información, pero cualquier parte de la información digitalizada convierte las cosas en números que pueden proporcionar un análisis, sirviendo así, como una fuente potencial de conocimiento con valor operacional. La clasificación del formato es crítica porque ese formato describe cómo convertir la información digital de vuelta en una representación de la realidad y cómo usar las herramientas correctas de la ciencia de los datos para obtener resultados analíticos.

Los datos deben estar almacenados, con el fin de estar disponibles para el análisis. Pueden almacenarse de una forma estructurada o no estructurada. Los *datos estructurados* tiene un conjunto de atributos y relaciones que se definen durante el proceso de diseño de la base de datos; estos datos encajan en una organización predeterminada, también conocida como esquema. En una base de datos estructurada, todos los elementos de la base de datos tendrán el mismo número de atributos en una secuencia específica. Los datos transaccionales generalmente están estructurados; tiene las mismas características y se guardan de la misma manera. Los datos estructurados son más fáciles de consultar y analizar. Los *datos no estructurados* no se organizan de acuerdo con esquemas predeterminados. Son flexibles para crecer de manera y forma, donde los atributos fiables pueden o no existir. Esto los hace más difíciles de analizar, pero esto es una ventaja a medida que más datos se generan rápidamente a partir de nuevas fuentes como las redes sociales, correos electrónicos, aplicaciones móviles y dispositivos personales. Los datos no estructurados tiene la ventaja de poder ser guardados en el estado en el que se encuentra, sin necesidad de comprobar si cumple con regla organizacional alguna. Esto hace que su almacenamiento sea rápido y flexible. También hay datos que se consideran datos semi-estructurados. Considere un tweet de Twitter, por ejemplo, que está limitado a 140 caracteres. Esta es una estructura organizacional predeterminada, y el servicio está programado para verificar que cada tweet satisfaga este requisito. Sin embargo, el contenido de lo que está escrito en un tweet no está ni predefinido ni es obligatorio;

esta combinación prácticamente infinita de palabras y letras ejemplifica los datos no estructurados. En su conjunto, los tweets son, por tanto, datos semi-estructurados.

Los datos también se clasifican según su fuente. Los Proveedores de Servicios Financieros tienden a clasificar las fuentes de datos como tradicionales o no tradicionales, donde las fuentes de *datos tradicionales* se refieren a fuentes internas de datos tales como transacciones básicas de sistemas de gestión de cuentas, encuestas de clientes, formularios de registro o información demográfica. Las fuentes de datos tradicionales también incluyen fuentes externas como los burós de crédito. Generalmente, se trata de datos estructurados. Los *datos no tradicionales*, o datos alternativos, pueden ser estructurados, semi-estructurados o no estructurados, y no siempre están relacionados con el uso de los servicios financieros. Ejemplos de este tipo de datos incluyen, datos de uso de voz y Servicio de Mensajes Cortos (SMS) de los MNO, imágenes satelitales, datos geoespaciales, datos de redes sociales, correos electrónicos u otros datos proxy. Estos tipos de fuentes de datos son cada vez más utilizados por los Proveedores de Servicios Financieros para ampliar o profundizar la comprensión del cliente, o se utilizan en combinación con datos tradicionales para obtener información operacional. Por ejemplo, una IMF que desee asociarse con una cooperativa lechera para conceder créditos a productores lecheros podría usar sus niveles de producción lechera como una variable sustitutiva (proxy) del salario con el fin de evaluar la capacidad de brindar crédito a los agricultores que carecen de historial de crédito formal.<sup>8</sup>

---

<sup>8</sup> Transcripción de la sesión 'Deploying Data to Understand Clients Better' The MasterCard Foundation Symposium on Financial Inclusion 2016, consultado Abril 3 de 2017 <http://mastercardfdnsymposium.org/resources/>

¿Qué es  
Big Data?



*Big data* es típicamente el término general utilizado para describir la vasta escala y la naturaleza sin precedentes de datos que se está produciendo. El Big Data tiene cinco características. Los primeros especialistas en Big Data identificaron las primeras 3 características que se enumeran a continuación y que aún son llamadas "las tres Vs". Desde entonces, las características de Big Data han crecido a una lista más larga de cinco:

- 1. Volumen:** La increíble cantidad de datos que se están produciendo es alucinante. La madurez de estos datos se ha vuelto cada vez más joven, es decir, que la cantidad de datos que tiene menos de un minuto de edad ha ido aumentando constantemente. Se prevé que la cantidad de datos en el mundo aumentará 44 veces entre 2009 y 2020.
- 2. Velocidad:** Una gran parte de los datos disponibles se producen y se vuelven disponibles en tiempo real. Cada minuto se envían 204 millones de correos electrónicos. Como consecuencia, estos datos se procesan y almacenan a velocidades muy altas.
- 3. Variedad:** La era digital ha diversificado los tipos de datos disponibles. Hoy en día, el 80 por ciento de los datos que se generan no son estructurados, en forma de imágenes, documentos y videos.
- 4. Veracidad:** La veracidad se refiere a la credibilidad de los datos. Los administradores de negocios necesitan saber que los datos que usan en el proceso de toma de decisiones son representativos de las necesidades y deseos de sus clientes. Por lo tanto, es importante asegurar un proceso de limpieza de datos riguroso y continuo.
- 5. Complejidad:** La combinación de los cuatro atributos anteriores requiere procesos analíticos complejos y avanzados. Han surgido una variedad de procesos analíticos para hacer frente a estos grandes conjuntos de datos.

## Fuentes de Datos

Esta sección se centra en las principales fuentes de información que los proveedores de SFD pueden considerar para posibles *insights* operacionales o de mercado. Es importante destacar que una fuente de datos no debe considerarse en forma aislada; la combinación de múltiples fuentes de datos a menudo conducirá a una comprensión cada vez más matizada de las realidades que codifican los datos. El capítulo 2.2, sobre la recopilación y el almacenamiento de datos de SFD, brinda una visión general de las fuentes de datos tradicionales y alternativas más comunes disponibles para los proveedores.

### Fuentes Tradicionales de Datos

Como se mencionó anteriormente, los Proveedores de Servicios Financieros han obtenido datos de vinculación de clientes, datos transaccionales y estudios de mercado primarios. Gran parte de los datos relevantes para el crédito se han almacenado en forma de documentos (copias físicas o electrónicas de documentos), y sólo los datos básicos de registros de clientes y actividad bancaria se han almacenado en bases de datos centralizadas. Un reto para los Proveedores de Servicios Financieros hoy en día es asegurar que estos tipos de datos tradicional también se almacenen en un formato digital que facilite el análisis de datos. Esto puede requerir un cambio en la forma en que se recopilan los datos, o la introducción de tecnología que convierta los datos a un formato digital. Aunque hay tecnología nueva disponible para digitalizar datos tradicionales, la digitalización puede

ser una tarea demasiado grande para los datos existentes.

### Datos de Cliente y Corresponsales

Los expertos recopilan una gran cantidad de información sobre sus clientes durante los procesos de vinculación y solicitud de créditos por razones tanto comerciales como de cumplimiento con regulaciones. Asimismo, también recopilan información sobre sus corresponsales como parte del proceso de solicitud y durante las visitas de control. Para ambas categorías, esto puede incluir variables como género, localización e ingresos. Alguna parte de esta data es verificada por medio de documentos oficiales, mientras que otra parte es comentada y capturada durante las entrevistas. En el caso de los prestatarios, gran parte de esta información del cliente se captura digitalmente en un Sistema de Originación de Créditos (SOC) o un módulo de originación en el Sistema de Core Bancario (CBS). Es sorprendentemente común que dicha información se guarde sólo en papel o en archivos escaneados.

### Terceros

Los burós de crédito y los registros son excelentes fuentes de datos objetivos y verificables. Brindan una verificación de la credibilidad de la información reportada por los solicitantes de créditos y a menudo pueden revelar información que el solicitante puede no divulgar de buena gana. La mayoría de los informes de burós de crédito y registros públicos pueden ya ser consultados en línea pudiendo visualizar los datos relevantes digitalmente. Sin embargo, un reto es que

no todos los mercados emergentes tienen una infraestructura de centrales de riesgo totalmente funcional.

### Estudios de Mercado Primarios

La investigación de mercado se utiliza generalmente para comprender mejor a los clientes y segmentos de mercado, seguir las tendencias del mercado, desarrollar productos y buscar retroalimentación de los clientes. Puede ser cualitativa o cuantitativa, y puede ser útil entender cómo y por qué los clientes usan los productos. El comprador secreto es un método de investigación de mercados común para probar si los corresponsales brindan un buen servicio al cliente, mientras que algunos proveedores de SFD buscan comentarios directos de los clientes a través de encuestas de satisfacción generando una puntuación [Net Promoter Score] que mide el grado de disposición de los clientes de recomendar un producto o servicio.

### Datos de Call Center

Los datos de call centers son una buena fuente para entender qué problemas enfrentan los clientes y cómo se sienten sobre los productos y el servicio al cliente de un proveedor. Los datos del call center se puede analizar categorizando los tipos de llamada y los tiempos de resolución, y utilizando analíticas de voz para examinar los registros de audio. Los datos de los call center son particularmente útiles para entender los problemas que los clientes, corresponsales o comerciantes están teniendo con los productos o la nueva tecnología que acaba de ser lanzada.

## 1.1\_ANALÍTICA DE DATOS Y MÉTODOS



**Número**



**Imagen**



**Texto**



**Voz**



**Biométrico**

*Figura 2: Formatos de Datos*

### **Bases de datos Transaccionales**

Los datos transaccionales ofrecen información sobre los niveles de actividad y las tendencias de uso del producto. Las comparaciones simples de transacciones por valor frente a volumen, pueden ofrecer resultados muy diferentes sobre el comportamiento del consumidor. Para las Instituciones Financieras, tales como bancos o microfinancieras, los datos sobre el uso de cuentas bancarias (depósitos, débitos y créditos) y otros servicios (tarjetas, préstamos, pagos y seguros) por parte de los clientes se capturan normalmente en el CBS. El uso de cuentas bancarias y servicios deja rastros objetivos de datos que pueden ser analizados para ver patrones que señalan diferentes niveles de capacidad financiera y sofisticación. Los diferentes patrones de uso también pueden indicar diferentes niveles de riesgo. Para procesar solicitudes de créditos, las Instituciones Financieras pueden requerir documentación de otras instituciones, como burós de crédito, sin embargo, esto tiende a ser en papel y es difícil de digitalizar.

### **Fuentes Alternativas de Datos**

A medida que más de nuestra comunicación y negocios se hacen a través de teléfonos móviles, tabletas y computadoras, hay más fuentes de datos digitalizados que pueden proporcionar información sobre la capacidad financiera y el carácter de los clientes. Estas fuentes nos pueden contar cómo pasan su tiempo y gastan su dinero las personas, y dónde y con quién se lo gastan.

### **Registros Detallados de Llamadas (CDR) de los MNOs**

Desde sus operaciones centrales, los operadores móviles tienen acceso a los CDR y las coordenadas de Torres Retransmisoras. Los operadores telefónicos analizan los CDR para desarrollar campañas de marketing y promociones focalizadas y para ajustar los precios, por ejemplo. Como mínimo, un CDR incluye 1) llamadas de voz, tiempo de llamada, uso de servicios de datos y datos de SMS en el emisor, el receptor, el tiempo y la duración, y 2) tiempo aire, información de recarga de datos, incluyendo tiempo, ubicación y denominación. Además, esta información

puede combinarse con las señales de torres de celdas para generar ubicaciones de actividad del cliente. Los operadores telefónicos que ofrecen servicios de dinero electrónico tienen acceso tanto a los datos CDR como a la base de datos transaccionales de SFD y, cuando se combinan para el análisis, es más probable que esta información ayude a predecir la actividad y el uso del cliente, que los datos demográficos simples. En algunos mercados, los MNOs y los Proveedores de Servicios Financieros se alían entre sí para beneficiarse de los datos combinados. Por ejemplo, las recargas de tiempo aire pueden ser un buen indicador de los ingresos discrecionales. Los clientes que gastan totalmente su tiempo de llamadas y, rutinariamente y con frecuencia, hacen pequeñas recargas es probable que tengan menos ingresos discrecionales que los que recargan con menos frecuencia, pero en cantidades mayores.

### **Datos de Transacciones Asistidos por Corresponsales**

Comprender cuáles ubicaciones y corresponsales son los más activos puede proporcionar información para ayudar a mejorar el desempeño de la red de corresponsales. Para muchos proveedores de SFD, los corresponsales son la cara principal frente al cliente y hacer seguimiento del patrón de uso y actividad del corresponsal puede revelar información sobre las preferencias del cliente y el desempeño del corresponsal. Dicha información puede ser registrada directamente desde teléfonos móviles, terminales de punto de venta (POS) u computadores como puntos de transacción. Alternativamente, podría estar indirectamente asociado, como los formularios de registro de corresponsales, que se tienen que integrar en el flujo de datos transaccionales para que se lleve a cabo un análisis.

### **Datos Geoespaciales**

Los datos geoespaciales hacen referencia a datos que contienen información de localización, como las coordenadas del sistema de posicionamiento global (GPS), direcciones, ciudades y otros identificadores geográficos o de proximidad. En los últimos años, los datos geoespaciales de carácter muy granular

han permitido a los proveedores de Servicios Financieros Digitales examinar y hacer referencias cruzadas entre los factores del lado de la demanda tales como el nivel de inclusión financiera, la ubicación del cliente, los niveles de pobreza y el uso con móviles de voz y datos, actividad del corresponsal, características rurales o urbanas, presencia de infraestructura y similares. Esto puede ofrecer información que puede ser útil para la adquisición de clientes y estrategias de marketing, expansión de sucursales o corresponsales, y análisis de competidores o del mercado en general. Los datos geoespaciales pueden ofrecer una visión más granular que los indicadores socioeconómicos típicos, que generalmente sólo están disponibles en formato agregado.

### **Perfiles en Redes Sociales**

Cada vez más, los mercados de clientes potenciales y existentes se están desarrollando en línea y mantienen una presencia en redes sociales como Facebook, Twitter y LinkedIn. Los datos de conducta en línea puede ofrecer información sobre la retroalimentación, las actitudes, estilos de vida, metas y la forma en que los servicios financieros pueden desempeñar un papel en la vida del cliente. Los datos de las redes sociales incluyen datos sobre el grado de conexión social, tráfico iniciado

y el comportamiento en línea, incluyendo el tiempo, ubicación, frecuencia y secuencia de un sitio web o una serie de sitios web. Las redes sociales también pueden ser indicativas del estatus socioeconómico de un individuo. Por ejemplo, las personas con un perfil de LinkedIn que tienen muchas conexiones pueden, en promedio, ser de menor riesgo que las que no las tienen. Eso no se debe a que registrarse en una cuenta de LinkedIn indique la capacidad de atender una deuda per se, sino porque LinkedIn se dirige a profesionales y, en promedio, los profesionales ganan salarios más altos que los obreros. Los perfiles públicos de las redes sociales también pueden ser útiles para verificar los datos de contacto y la información personal básica del cliente. Sin embargo, las redes sociales como fuente de datos tienen sus limitaciones. Los Proveedores de Servicios Financieros generalmente sólo pueden acceder a las cuentas de redes sociales de los clientes que lo aceptan, y puede ser difícil conseguir suficientes clientes que acepten esto para construir una base de datos lo suficientemente grande como para un análisis significativo. Algunos clientes también pueden no ser activos en las redes sociales, debido a su propia elección o sus circunstancias. Los datos de perfil, incluso cuando están disponibles, también puede estar sesgados.

## 1.1\_ANALÍTICA DE DATOS Y MÉTODOS

### Fuentes de Datos Operacionales

Hay muchos procesos de negocio necesarios para ejecutar operaciones de Servicios Financieros Digitales, en los que cada departamento trabaja para completar tareas y cumplir objetivos de desempeño, al tiempo que dependen de datos procedentes de múltiples fuentes. Las posibles fuentes de datos externas e internas se ilustran en la figura siguiente y se enumeran con mayor detalle en el Capítulo 2.2. Cada departamento genera y consume datos a través de este ecosistema. Algunas de las fuentes de datos más importantes son:

#### Sistema Core de Datos

El sistema core proporciona la mayor parte de los datos. El motor transaccional es responsable de gestionar el flujo de trabajo de transacciones e interacciones, enviando tantos datos granulares y metadatos como sea posible a las bases de datos pertinentes. Esto incluye el movimiento de fondos más tarifas y comisiones, así como cualquier modelo de negocio en torno al reparto de comisiones y las normas tributarias. También debe proporcionar rastros de flujos de trabajo completamente auditables de actividades no financieras tales como cambios en el número de identificación personal (PIN), consultas de saldos, mini-estados y descargas de datos, así como funciones internas como transferencias de fondos entre cuentas.

### Informes del Sistema de Inteligencia de Negocio (BI)

Cuando los productos de SFD son nuevos y hay un volumen de datos relativamente bajo, es común que las empresas creen informes personalizados a partir de datos sin procesar utilizando herramientas sencillas como Excel. A medida que el negocio y los datos crecen, y el análisis requerido se vuelve más complejo, esto pronto se vuelve incontrolable. La mayoría de los grandes sistemas de Servicios Financieros Digitales pondrán en marcha un almacén de datos que utiliza sistemas de BI para aprovechar múltiples fuentes de datos, que proporcionan algunos informes básicos, así como la capacidad de personalizar.

### Archivos de Registro Técnico

Una rica fuente de datos se puede encontrar en los archivos de registro técnico. Los proveedores de Servicios Financieros Digitales más avanzados utilizan proactivamente los paneles de control para asegurar continuamente la salud del sistema y proporcionar detección temprana de fallos. También es común tener monitores de desempeño y alertas integradas en el sistema de monitoreo que pueden proporcionar información valiosa. Los proveedores que visualizan estos datos sólo cuando se requiere un análisis forense específico, pierden información disponible y útil.

### Datos Internos Periféricos

#### Datos de Private Branch Automatic Exchange (PBAX)

El PBAX controla las llamadas entrantes a un call center y puede proporcionar datos sobre el volumen de llamadas entrantes, el número de llamadas perdidas antes de que se contesten y el tiempo que se emplea en las llamadas. Estos datos son vitales para la planificación eficiente de los patrones de cambio y tamaño, así como la medición y la mejora general del desempeño del equipo.

#### Sistemas de Tickets

El sistema de tickets hace seguimiento al proceso de resolución de conflictos de negocio y proporciona una gran cantidad de información, desde los tipos de conflictos que se producen, hasta tiempos de resolución de conflictos.

## Protección de Datos y Protección del Cliente

Las nuevas metodologías analíticas y de recopilación de datos plantean varias cuestiones relacionadas con los derechos de privacidad del cliente y la protección del consumidor. En primer lugar, como se ha señalado anteriormente, muchos de los datos producidos y recopilados se recojen de forma pasiva, es decir, sin el conocimiento del productor de los datos. A veces, estos datos pueden ser compartidos con terceros sin el conocimiento del productor de los datos. Esto puede tener implicaciones negativas en la capacidad del individuo para obtener préstamos o seguros. El problema se agrava cuando el individuo no es consciente de esta información negativa o no tiene el derecho de disputar la información negativa. En la actualidad, no existen políticas estándar de participación para compartir datos. Algunos proveedores de SFD con aplicaciones que están instaladas en los teléfonos móviles de sus clientes pueden tener la capacidad de barrer la información de uso de Internet del cliente y otros datos, incluyendo mensajes SMS, contactos y datos de ubicación, entre otros.

Dada la diversidad de proveedores de Servicios Financieros Digitales, no todos están bajo el mismo régimen de vigilancia, lo que lleva a diferentes políticas de privacidad de datos para cada uno. Algunas de las violaciones a los derechos individuales a la privacidad podrían tener impactos

reputacionales negativos. En Kenia, muchos proveedores de crédito digital han surgido para satisfacer la demanda de crédito, pero operan fuera del ámbito regulador del Banco Central.<sup>9</sup> Uno de estos proveedores incluyó en sus términos y condiciones que el proveedor era libre de publicar los nombres de los morosos en su web y en sus muros de las redes sociales. En casos como este, los clientes pueden no ser conscientes que están de acuerdo en suspender sus derechos de privacidad hasta que sea demasiado tarde. Esto puede ser particularmente cierto en contextos de países en desarrollo en los que la alfabetización y la conciencia de los problemas son bajas.

Cabe destacar, que incluso en los países donde el consentimiento del usuario es frecuente, los consumidores pueden no entender los permisos que están concediendo. A modo de ejemplo, los usuarios en mercados sofisticados pueden no ser conscientes de todas las aplicaciones en sus teléfonos inteligentes que hacen uso de datos de ubicación. Los estudios muestran que el 80 por ciento de los usuarios de móviles tienen inquietudes sobre compartir su información personal mientras usan Internet móvil o las aplicaciones.<sup>10</sup> Sin embargo, el 82 por ciento de los usuarios están de acuerdo con los avisos de privacidad sin leerlos porque tienden a ser demasiado largos o usan terminología que no les es familiar. Debido a las inquietudes de seguridad y la

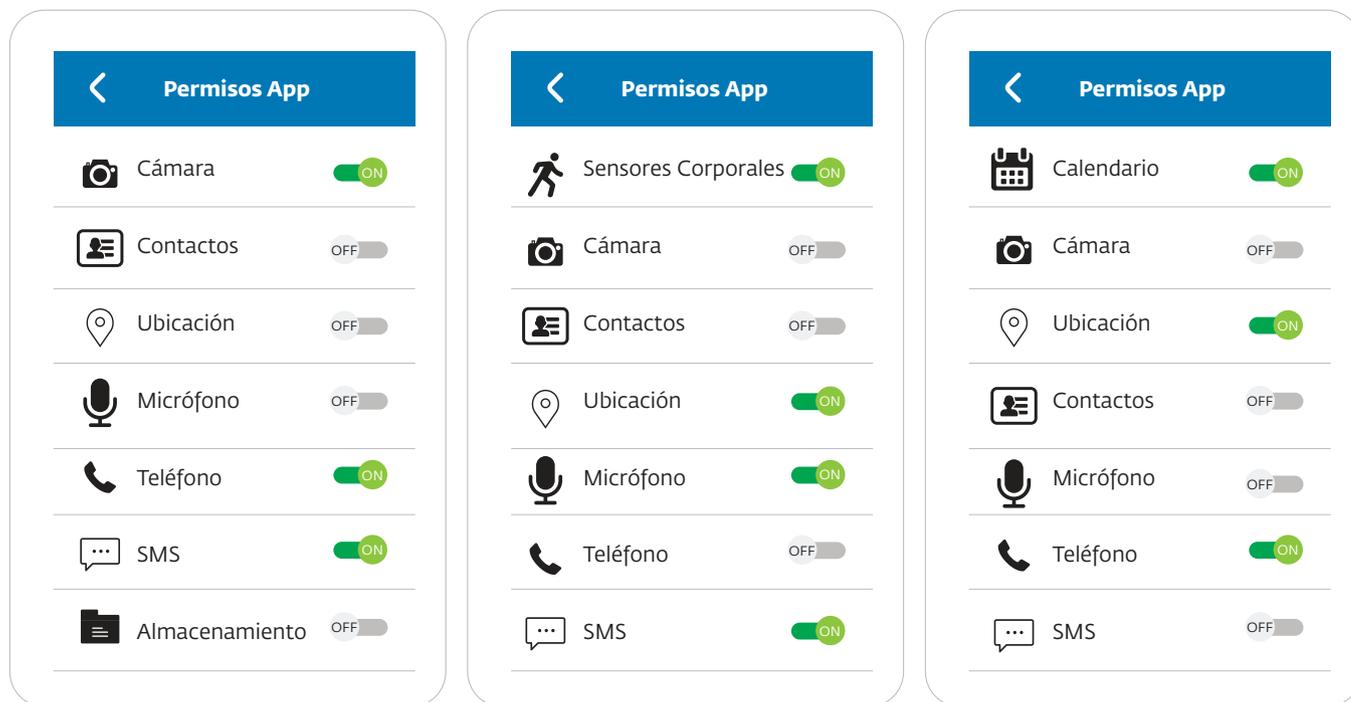


*Figura 3: Ejemplo de solicitud para guardar y acceder a la data del historial de ubicaciones de usuarios a través de la aplicación Google Maps*

<sup>9</sup> Ombija y Chege, 'Time to Take Data Privacy Concerns Seriously in Digital Lending', *Blog de Consultative Group Against Poverty*, 24 de octubre de 2016, consultado el 3 de abril de 2017, <https://www.cgap.org/blog/time-take-data-privacy-concerns-seriously-digital-lending>

<sup>10</sup> 'Mobile Privacy: Consumer research insights and considerations for policymakers', GSMA

## 1.1\_ANALÍTICA DE DATOS Y MÉTODOS



*Figura 4: Ejemplos de configuración para permisos de aplicaciones en teléfonos inteligentes*

disposición declarada de los clientes a dejar de usar las aplicaciones que encuentran demasiado intrusivas o que carecen de seguridad, la mayoría de las aplicaciones hoy en día ofrecen formas sencillas de aceptación y rechazo.

Las leyes de privacidad, donde existen, varían ampliamente por jurisdicción y más aún en cuanto al grado de obligatoriedad. En el contexto de los mercados

desarrollados, en la Unión Europea (UE) el derecho a la privacidad y la protección de datos está fuertemente regulado y se aplica activamente,<sup>11</sup> mientras que en los Estados Unidos no existe una ley federal integral de protección de datos. En el año 2016, la UE emitió una regulación sobre protección de datos que obliga a que todos los generadores de datos puedan recibir de vuelta la información que proporcionan a las empresas,

enviar la información a otras empresas y permitir a las empresas intercambiar información cuando sea técnicamente posible.<sup>12</sup> Este tipo de regulación aumenta el empoderamiento del consumidor al tiempo que mejora la competencia, ya que los consumidores pueden ahora moverse entre los proveedores con su historial de transacciones intacto. En los Estados Unidos, la Comisión Federal de Comercio (FTC) es el órgano regulador de

<sup>11</sup> Regulación que gobierna la protección de datos en la UE incluye la Directiva de protección de datos de la UE 95/46 CE y la Directiva de la UE sobre Privacidad y Comunicaciones electrónicas 2002/58 CE (modificada por la Directiva 2009/136)

<sup>12</sup> Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo (2016), consultado el 3 de abril de 2017, <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>

la privacidad de los datos. Sin embargo, el Código de principios justos de información de la FTC es sólo un conjunto de recomendaciones para mantener prácticas de captura de datos orientadas al consumidor y respetuosas de la privacidad - no es aplicable por ley. En ausencia de cualquier regla federal de privacidad, en los Estados Unidos han desarrollado estatutos y regulaciones federales y estatales para abordar la privacidad de la información personal y la seguridad de los datos, tanto en un sentido general como en una base sectorial a la cual todos los negocios relevantes deben adherirse.

En lo que se refiere a África subsahariana, Ghana, Sudáfrica y Uganda parecen destacarse como las mejores prácticas regionales. Lo que distingue a estos tres países es el hecho de que la regulación está guiada por un principio de orientación al cliente y, como tal, la regulación se centra en:

- Empoderar al consumidor para tomar decisiones pertinentes sobre el uso de sus datos personales, especialmente en relación con la toma de decisiones automatizada
- Establecer mecanismos claros a través de los cuales el consumidor puede solicitar compensación
- Dar al cliente el 'derecho a ser olvidado'

Los flujos transfronterizos de datos constituyen un asunto delicado, especialmente porque pueden afectar asuntos de seguridad nacional. La regulación en países como Angola, Sudáfrica y Tanzania estipula específicamente que los datos sólo puede ser transferidos a países donde la ley provee normas de protección igual o más estricta para los datos personales en cuestión. Zambia va aún más lejos al prohibir cualquier transferencia off-shore de datos que no esté anonimizada.<sup>13</sup> En el otro extremo del espectro, el proyecto de ley sobre protección de datos de Kenia de 2016 ha sido duramente criticado por expertos por no incluir ninguna disposición para la jurisdicción extraterritorial.<sup>14</sup>

Sin embargo, la privacidad de los datos de los clientes es un nuevo campo de políticas, y países como Mozambique y Zimbabue siguen confiando en la Constitución para interpretar los derechos de privacidad como resultado de no tener leyes regulatorias específicas. En este contexto, los mercados emergentes buscan frecuentemente mercados y reguladores más establecidos para obtener pistas sobre cómo abordar los temas en cuestión.

Dado este contexto, pero consciente de las diferencias entre el uso de la tecnología en mercados emergentes y desarrollados, las Naciones Unidas (ONU) han ofrecido alguna orientación general en términos del desarrollo de políticas. La ONU hace

hincapié en la necesidad de acelerar el desarrollo y la adopción de normas legales, técnicas, geoespaciales y estadísticas con respecto a:

- La apertura y el intercambio de metadatos
- Protección de datos personales como derecho humano<sup>15</sup>

Así las cosas, por el momento, no existe una política uniforme que rija las cuestiones de protección de datos. El primer paso para entender las implicaciones de la privacidad es asegurar una discusión sectorial que involucre a los proveedores de SFD, reguladores, encargados de la formulación de políticas, otros actores del sector público, inversionistas e Instituciones Financieras de desarrollo para diseñar soluciones y estándares. Al mismo tiempo, en el sector de la inclusión financiera, los proveedores de Servicios Financieros Digitales deben reconocer que, si bien los datos representan una oportunidad para mejorar la rentabilidad, también subraya la obligación de agregar valor. Esto puede lograrse utilizando los datos para mejorar el acceso a los servicios financieros. Los proveedores de SFD pueden intentar educar a la gente acerca de cómo se usará su información personal mientras sólo recopile información que sea necesaria.

<sup>13</sup> 'Global Data Privacy Directory', Norton Rose Fulbright

<sup>14</sup> Francis Monyango, 'Consumer Privacy and data protection in E-commerce in Kenya', *Nairobi Business Monthly*, Abril 1 de 2016, consultado Abril 3 de 2017, <http://www.nairobibusinessmonthly.com/politics/consumer-privacy-and-data-protection-in-e-commerce-in-kenya/>

<sup>15</sup> 'A World That Counts: Mobilizing the Data Revolution for Sustainable Development', Grupo Asesor de Expertos Independientes sobre la Revolución de la data para el Desarrollo Sostenible del Secretario General de la ONU

## 1.1\_ANALÍTICA DE DATOS Y MÉTODOS

### Ciencia de los Datos: Introducción

La ciencia de los datos es el uso interdisciplinario de métodos científicos, procesos y sistemas para extraer información y conocimientos de diversas formas de datos con el fin de resolver problemas específicos. Combina las ciencias numéricas tales como la estadística y las matemáticas aplicadas, con la informática

y la experiencia empresarial y sectorial. Es una disciplina exploratoria y creativa, que busca encontrar soluciones innovadoras a problemas complejos a través de un enfoque analítico. La ciencia de los datos se refiere al método científico de análisis: los científicos de datos se dedican a la resolución de conflictos estableciendo una hipótesis comprobable y probando y refinando asiduamente esa hipótesis para obtener resultados fiables y validados.

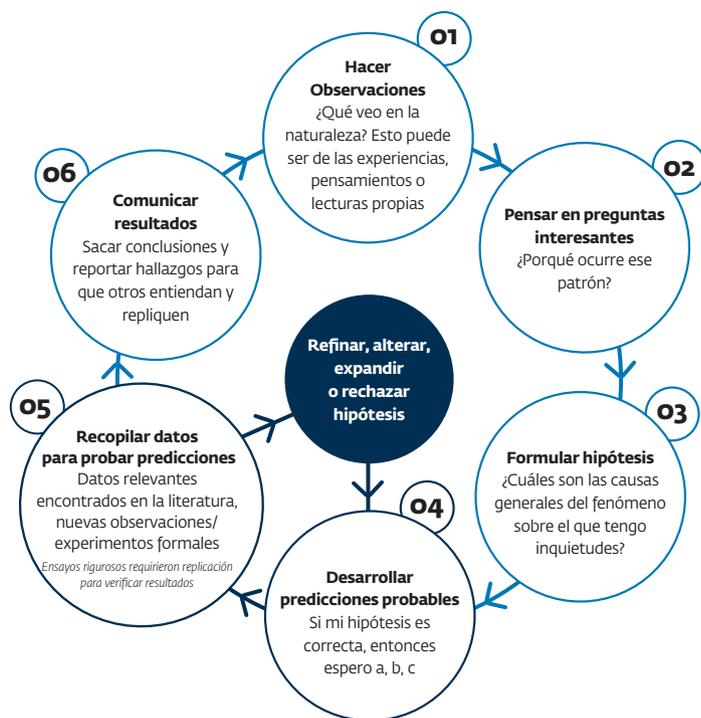
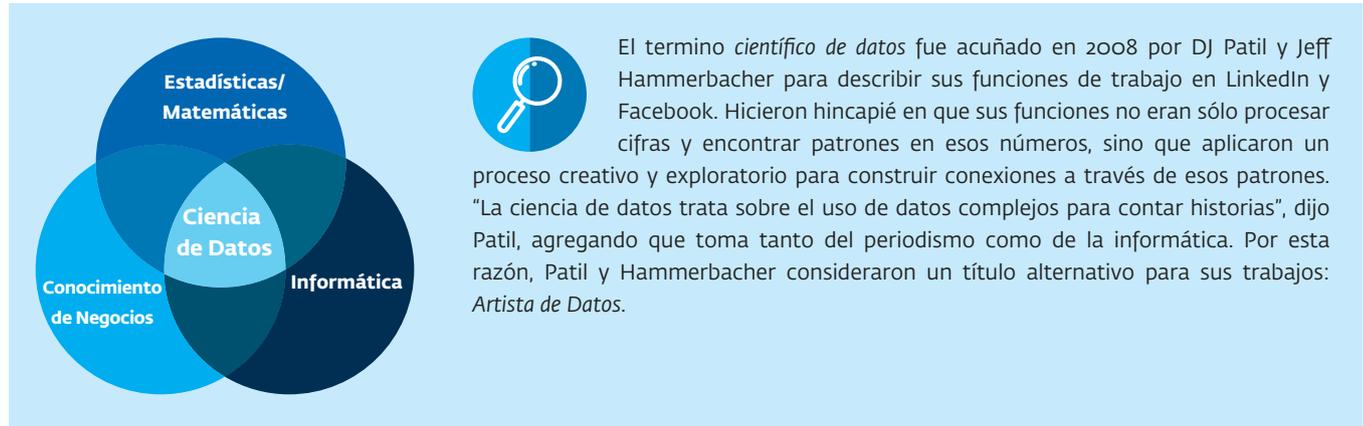


Figura 5: El método científico, el proceso analítico que se utiliza de manera similar para la 'ciencia de los datos'

## Ciencia de Datos



El término *científico de datos* fue acuñado en 2008 por DJ Patil y Jeff Hammerbacher para describir sus funciones de trabajo en LinkedIn y Facebook. Hicieron hincapié en que sus funciones no eran sólo procesar cifras y encontrar patrones en esos números, sino que aplicaron un proceso creativo y exploratorio para construir conexiones a través de esos patrones. “La ciencia de datos trata sobre el uso de datos complejos para contar historias”, dijo Patil, agregando que toma tanto del periodismo como de la informática. Por esta razón, Patil y Hammerbacher consideraron un título alternativo para sus trabajos: *Artista de Datos*.

**Figura 6:** Ciencia de Datos, la intersección de varias disciplinas

Para entregar BI, todos los análisis relacionados con datos deben comenzar por definir metas de negocio e identificar las preguntas o hipótesis de negocio correctas. El método científico brinda una orientación útil (véase la Figura 5). Es importante destacar que no es un proceso lineal. En su lugar, siempre hay un ciclo de aprendizaje y retroalimentación para asegurar una mejora incremental. Esto es clave para obtener resultados que permitan la toma de decisiones fiables y basadas en evidencia. El capítulo 2.1 de este manual proporciona un proceso paso a paso para implementar proyectos de datos para proveedores de SFD, utilizando la metodología del Data Ring.

La ciencia de los datos facilita el uso de nuevos métodos y tecnologías para BI, y análisis útiles se pueden derivar de

datos grandes y pequeños, tradicionales y alternativos. Las computadoras más rápidas y algoritmos complejos aumentan las posibilidades analíticas, pero ninguno de los dos reemplaza ni desplaza las herramientas validadas en el tiempo y los enfoques para obtener resultados basados en datos con el objetivo de resolver problemas de negocio. Más bien, es importante entender las fortalezas que ofrecen diferentes herramientas y aumentarlas adecuadamente para obtener los resultados deseados de una manera oportuna y rentable.

La Figura 7 proporciona una descripción de alto nivel de los métodos analíticos de BI, clasificados por su uso operacional y su relativa sofisticación. Muchas categorías y sus técnicas e implementaciones asociadas se superponen, pero es útil dividir las

cuatro casos de uso principales: *descriptivo, diagnóstico, predictivo y prescriptivo*. Las metodologías menos complejas suelen ser descriptivas por naturaleza, proporcionando descripciones históricas del desempeño institucional, cifras agregadas y estadísticas resumidas. También tienen menos probabilidad de ofrecer una ventaja competitiva, pero son de todos modos críticas para el monitoreo del funcionamiento operacional y el cumplimiento regulatorio. En el extremo opuesto, los análisis más innovadores y complejos son prescriptivos, optimizados para la toma de decisiones y ofrecen información sobre las expectativas futuras. Esta progresión también ayuda a clasificar los entregables y la estrategia de implementación de un proyecto de datos, que se analiza más adelante en el Capítulo 2.1.

## 1.1\_ANALÍTICA DE DATOS Y MÉTODOS

### Marco Analítico de la Ciencia de los Datos para la Inteligencia de Negocios

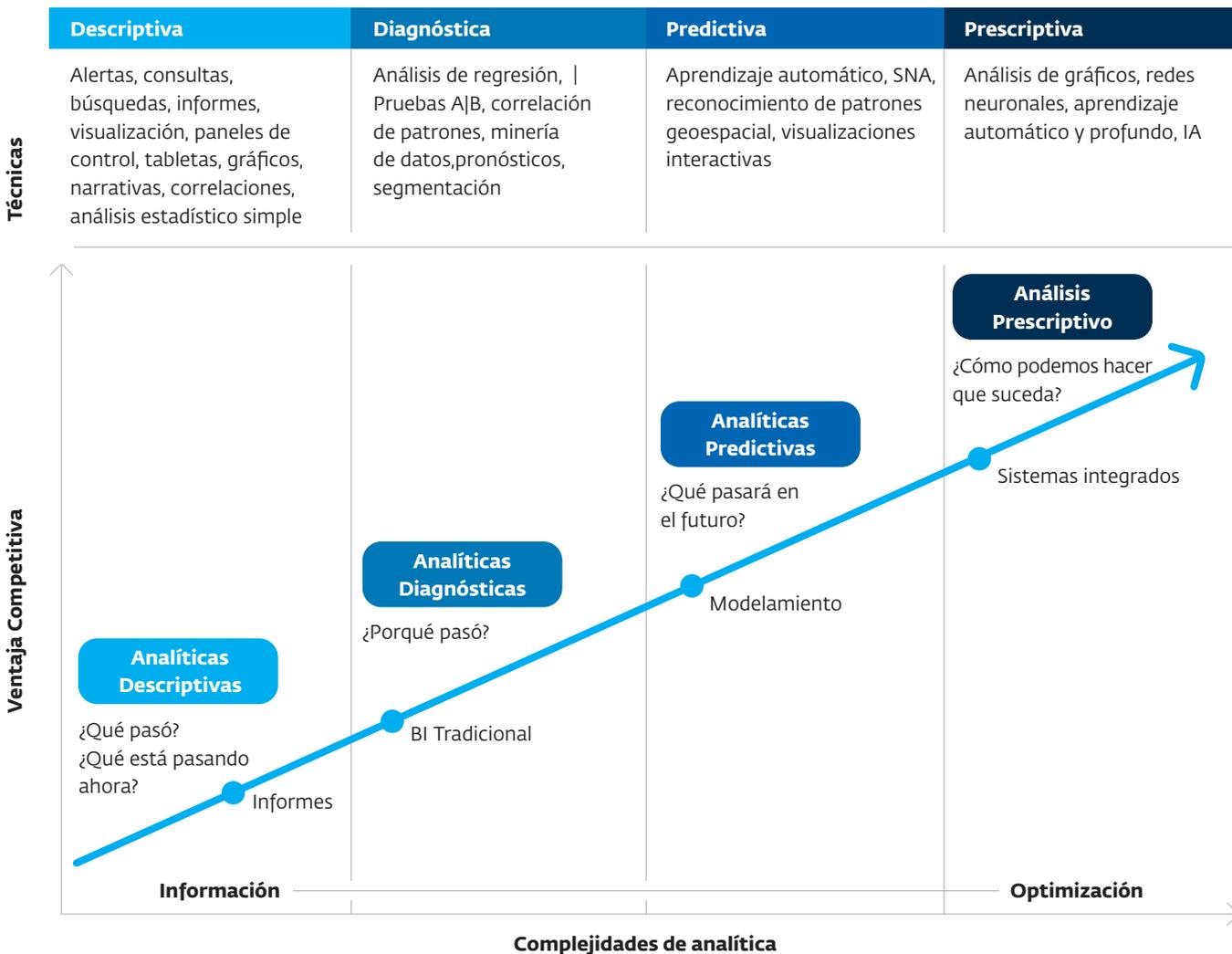


Figura 7: Las cuatro categorías de Análisis de Negocio

## Métodos

Los casos de uso analítico descritos en la Figura 7 ayudan a determinar el método, el tiempo, el costo y la complejidad de los proyectos de datos. Los siguientes métodos generalmente se incluyen dentro de las herramientas del científico de datos, y ayudan a combinar los métodos generales con propósitos analíticos. Estos métodos son especialmente relevantes para las discusiones con consultores externos o proveedores de soluciones para ayudar a enmarcar lo que están entregando o para evaluar una propuesta.

### Analítica Descriptiva

La analítica descriptiva ofrece informes agregados de alto nivel de registros históricos y responde a preguntas sobre qué ocurrió. Los Indicadores Clave de Desempeño (KPI) también están dentro de esta categoría.

- **Estadísticas Descriptivas:** También conocidas como estadísticas de resumen, las estadísticas descriptivas incluyen promedios, sumas, conteos y agregaciones. Las estadísticas de correlación que muestran las relaciones entre las variables también ayudan a describir los datos.
- **Tabulación:** El proceso de organizar los datos en un formato de tabla se conoce como tabulación. La tabulación cruzada resume los datos de una o más fuentes en un formato conciso para el análisis o la presentación de informes,

agregando valores frecuentemente. Es un método de segmentación, que permite que los agregados se tabulen por sexo o ubicación, por ejemplo, u otros segmentos de interés. Excel utiliza el término tabla dinámica para describir este tipo de análisis.

### Analítica de Diagnóstico

La búsqueda de los factores clave o la comprensión de los patrones de datos cambiantes es el análisis de diagnóstico. Se trata de preguntar por qué sucedió algo; por ejemplo, preguntando por qué los patrones de transacción cambiaron para determinar si no sólo hay correlación, sino causalidad. La analítica de diagnóstico generalmente requiere métodos más sofisticados y diseños de investigación, como se describe a continuación.

- **Pruebas A|B:** Se trata de un método estadístico en el que dos o más variantes de un experimento se muestran a los usuarios al azar para determinar cuál funciona mejor para un objetivo de conversión dado. Las pruebas A | B permiten a los negocios probar dos escenarios diferentes y comparar los resultados. Es un método muy útil para identificar las mejores estrategias promocionales o de marketing entre las opciones probadas.
- **Regresión:** La regresión estadística es uno de los tipos más básicos de modelado, y es muy potente. Permite el análisis multi-variable para estimar

las relaciones entre una variable dependiente, usualmente una métrica de interés comercial, y un conjunto de variables independientes con las que se correlaciona. Identificar las variables estadísticamente significativas<sup>16</sup> puede orientar la estrategia, ayudar a enfocar los objetivos y estimar los resultados.

- **Segmentación:** La segmentación es un método de clasificación de grupos en subgrupos basados en criterios definidos, comportamiento o características. La segmentación puede ayudar a identificar categorías de demografía de clientes o categorías de uso del producto, con umbrales cuantificados y estadísticamente significativos. Esto se utiliza a menudo junto con el análisis de regresión o técnicas de modelación más sofisticadas para predecir a qué segmento podría pertenecer un prospecto de cliente no identificado.
- **Análisis geoespacial:** Este método agrupa los datos de acuerdo con su ubicación en un mapa, o en relación con el lugar y la proximidad. Esto también puede ayudar a identificar segmentos de clientes y de comportamiento, tales como desde dónde y hacia dónde envían dinero, o qué sucursal tienden a visitar. Combinado con técnicas más avanzadas, también puede permitir que los servicios basados en la localización conecten de manera proactiva a los clientes que están cerca de la gente o de lugares de interés.

<sup>16</sup> Estadísticamente significante es la posibilidad de que una relación entre dos o más variables sea causada por algo distinto al azar

## 1.1 ANALÍTICA DE DATOS Y MÉTODOS

### Análisis Predictivo

Las predicciones permiten una toma de decisiones orientada hacia el futuro y estrategias basadas en datos. Desde el punto de vista de la ciencia de los datos, esto es posiblemente la categoría más importante de los métodos, ya que los algoritmos complejos y la potencia computacional se utilizan a menudo para impulsar modelos. Desde una perspectiva de negocio, los modelos predictivos pueden proporcionar eficiencias operacionales identificando segmentos de clientes de alta propensión y expandiendo el alcance a menores costos a través de campañas de marketing focalizadas. También pueden ayudar a mejorar el soporte al cliente previendo proactivamente las necesidades de servicio.

- **Aprendizaje Automático (Machine Learning):** Este es un campo de estudio que construye algoritmos para aprender y hacer predicciones sobre los datos. En particular, este método permite a un proceso analítico identificar patrones en los datos sin una instrucción explícita del analista y permite que los métodos de modelado identifiquen variables de interés y factores incluso para patrones no intuitivos. Es una técnica más que un método en sí mismo. Los enfoques basados en el aprendizaje automático se clasifican en términos de 'aprendizaje supervisado' o 'aprendizaje no supervisado', dependiendo de si hay evidencia empírica para entrenar el algoritmo de aprendizaje, donde las metodologías supervisadas tienen evidencia empírica.

- **Modelamiento:** Existen dos métodos de modelización primaria: regresión y clasificación. Ambos se pueden utilizar para hacer predicciones. Los modelos de regresión ayudan a determinar un cambio en una variable de salida con variables de entrada dadas; por ejemplo, ¿cómo aumentan las calificaciones crediticias con los niveles de educación? Los modelos de clasificación ponen los datos en grupos o a veces en grupos múltiples, respondiendo a preguntas como si un cliente es activo o inactivo, o a qué rango de ingresos pertenece. Existen numerosos tipos de técnicas de modelización para cualquiera, con detalles técnicos matizados. Los enfoques de modelado tienden a generar mucha atención, pero es importante señalar que el método de modelado probablemente no es una especificación de diseño de análisis importante. Normalmente, se intentan muchos tipos de modelos y luego se selecciona el mejor en métricas de desempeño predefinidas; en otras ocasiones se combinan, creando un enfoque de conjunto. Un consultor debe describir por qué se selecciona un enfoque recomendado, y no simplemente declarar, por ejemplo, que la solución se basa en un método específico, como el conocido método de 'bosque aleatorio'. Decidir qué método usar para el modelado debe considerar la importancia de ser capaces de interpretar por qué se han dado unos resultados en comparación con la exactitud de la predicción. Los modelos de regresión tienden a ser muy transparentes y

fácilmente interpretables, por ejemplo; mientras que el método de bosque aleatorio está en el otro extremo del espectro, proporcionando buenas predicciones, pero una comprensión insuficiente de lo que las genera.

### Análisis Prescriptivo

Los métodos en esta categoría tienden a ser categorizados por la predicción o la clasificación de los aspectos de comportamiento en las relaciones complejas, e incluye un conjunto avanzado de métodos, que se describen a continuación. La inteligencia artificial (IA) y los modelos de aprendizaje profundo pertenecen a este grupo. Sin embargo, esta clasificación se enmarca mejor dentro de la infraestructura que se estima necesaria para usar los resultados de un análisis, asegurando que ofrece valor operacional. Por ejemplo, esto podría tomar la forma de un conjunto de herramientas de los paneles de control, necesarias para ejecutar una visualización interactiva en un sitio web, o la infraestructura de tecnología de la información (TI) para automatizar un modelo de scoring crediticio. Lo que define un *producto basado en datos* es la integración de un algoritmo o proceso basados en datos en un sistema operativo mayor, o como filtro en un proceso automatizado que depende de él para brindar un servicio.

# Lecciones de la Industria: Google tiene Gripe

## Modelado Predictivo y Ajuste de Modelos: Riesgo de Fiabilidad de los Modelos Sin Supervisión

Los investigadores del buscador Google se preguntaban si podría haber una correlación entre las personas que buscan palabras como “tosser”, “estornudar” o “congestión nasal” - síntomas de gripe o resfriado - y la prevalencia actual de la gripe. En los Estados Unidos, la propagación de la gripe tiene datos rezagados; la gente se enferma y visita al médico, entonces el médico reporta las estadísticas, y por lo tanto los datos capturan lo que ya ha sucedido. ¿Podrían los modelos impulsados por palabras de búsqueda proporcionar datos en tiempo real a medida que se propagaba realmente la influenza? Este enfoque para reducir los retrasos en los datos se conoce como **nowcasting**. Para problemas como la gripe estacional, los beneficios para la salud

pública son obvios. El modelo fue un éxito y fue lanzado públicamente como Google Flu Trends [Tendencias de Gripe de Google]. El impresionante modelamiento de big data de Google fue destacado en la revista científica *Nature* en 2008. Sin embargo, seis años más tarde, el fracaso del mismo modelo se describió de manera prominente en la revista *Science*. ¿Qué pasó entre 2008 y 2014?

El número de usuarios de Internet creció sustancialmente durante estos seis años y los patrones de búsqueda de 2008 no permanecieron constantes. La cuestión central fue que Google Flu Trends se desarrolló utilizando técnicas de aprendizaje automático **no supervisado**: 45 frases de búsqueda impulsaron el modelo,

identificadas como correlaciones estadísticamente poderosas en 2008. Pero muchos de estos términos de búsqueda eran en realidad predictores de las estaciones, y las estaciones a su vez se correlacionaron con la gripe. Cuando los patrones de gripe cambiaron antes o después de lo que había sido en 2008, esos términos de búsqueda ya no se correlacionaban, tan fuertemente, con la gripe. Combinado con el cambio demográfico de los usuarios, el modelo se volvió poco fiable. Google Flu Trends se dejó en piloto automático, utilizando métodos de aprendizaje **no supervisado** y las correlaciones estadísticas se debilitaron con el tiempo incapaz de mantenerse al día con los patrones de cambio.



Cuando se utilizan métodos similares para las decisiones de negocios o para asuntos de salud pública, es importante tener en cuenta que la pérdida de fiabilidad con el tiempo puede presentar riesgos significativos.

### Método de Bosque Aleatorio



El *método de bosque aleatorio* ha generado mucha expectativa en la ciencia de los datos porque tiende a impulsar modelos altamente precisos. Es una forma de *modelo* de clasificación que utiliza una estructura de decisión tipo árbol o diagrama de flujo combinada con enfoques de selección aleatorios para identificar una ruta óptima entre el resultado deseado y un conjunto de variables de entrada 'bosque'. Es importante entender que algunos métodos de la ciencia de los datos se entienden fácilmente en un contexto de negocios, mientras que otros no. El método del bosque aleatorio puede, por ejemplo, generar modelos altamente precisos, pero su complejidad produce una 'caja negra' que hace que sea muy difícil de interpretar. Esto podría ser potencialmente problemático para un modelo de score crediticio; podría identificar a las personas más solventes dados los datos de entrada, pero no puede ayudar a describir lo que hace que estas personas sean solventes o qué determina la recomendación de crédito.

- **Minería de Texto (Procesamiento del Lenguaje Natural):** La minería de texto es el proceso de obtener información de alta calidad a partir del texto. El texto puede ayudar a identificar las opiniones y los sentimientos de los clientes sobre productos utilizando mensajes de redes sociales, Twitter o mensajes del CRM. El procesamiento del Lenguaje Natural (PLN) combina la lingüística computacional y los métodos de Inteligencia Artificial para ayudar a las computadoras a comprender la información de texto para su procesamiento y análisis.
- **Análisis de Redes Sociales (SNA)** Este es el proceso de análisis cuantitativo y cualitativo de una red social. Para propósitos de negocios, el SNA puede emplearse para evitar la rotación, detectar fraude y abuso, o para inferir atributos, como la capacidad crediticia basada en grupos de pares.
- **Procesamiento de Imágenes:** Este enfoque utiliza algoritmos informáticos para realizar análisis con el propósito de clasificación, extracción de características, análisis de señales o reconocimiento de patrones. Los negocios pueden utilizar esto para reconocer a las personas en las imágenes para ayudar con la detección de fraude o para detectar características geográficas relevantes para situar correspondientes mediante imágenes satelitales.

### Herramientas

La ciencia de los datos y sus métodos se desarrollan con lenguajes de programación de computadoras, o los algoritmos se ejecutan en plataformas computacionales. Los datos que alimenta estos algoritmos se extraen de las bases de datos. El conjunto de herramientas del científico de datos también incluye conocimientos profundos sobre la computación técnica y las habilidades blandas necesarias para desarrollar e implementar algoritmos de datos. Las especificaciones técnicas de estas herramientas están fuera del alcance de la Analítica de datos de SFD. Sin embargo, algunas tecnologías prominentes se destacan para anotar algunas herramientas que los científicos de datos tenderán a utilizar. Los productos de datos exitosos requieren una combinación de métodos, herramientas y habilidades, como se comentará más adelante en el Capítulo 2.1: Gestión de un Proyecto de Datos.

### Herramientas Duras

- **Bases de Datos:** La estructura de los datos guiará la solución de base de datos apropiada. Los datos estructurados suelen ser gestionados por *bases de datos relacionales* con esquemas fijos que pueden soportar la fiabilidad integral de los datos, lo que puede ayudar a los analistas a identificar anomalías en el valor de los datos o evitar que guarden datos erróneos desde el principio.

Las bases de datos relacionales organizan conjuntos de datos en tablas que están *relacionadas* entre sí por una clave, es decir, un atributo de metadatos compartido entre las tablas. Las soluciones de centros de datos empresariales y el almacenamiento de datos transaccionales suelen utilizar bases de datos relacionales. Los productos conocidos incluyen: Oracle, SQL Server y MySQL. Los datos no estructurados suele ser gestionados por bases de datos *no relacionales* que carecen de esquemas rígidos, comúnmente denominados bases de datos NoSQL. Ofrecen ventajas en escala y distribución, y se depende de estas a menudo para Big Data y las aplicaciones interactivas en línea. A medida que los conjuntos de datos grandes se hacen más grandes, el espacio en disco duro se vuelve limitado y el tiempo computacional de búsqueda tarda más. La ventaja de las bases de datos NoSQL es que están diseñadas para ser *escalables horizontalmente*, lo que significa que otra computadora, o dos, o cien, pueden ser añadidas de manera transparente para aumentar el espacio de almacenamiento y la potencia computacional de búsqueda. Si bien las soluciones relacionales también pueden escalarse y distribuirse, a menudo son más complejas de manejar y afinar cuando se guardan datos en múltiples computadoras. Los productos más conocidos incluyen: Hadoop, MongoDB y BigTable.

- **Frameworks:** Estos son conjuntos de paquetes de software que combinan una solución de almacenamiento de datos con un interfaz de programación de aplicaciones (API) que integran herramientas de gestión o analíticas en la base de datos. En otras palabras, se trata de soluciones de fuente única para administrar y analizar datos. Los productos prominentes incluyen Spark y Hive. Hadoop, mencionado anteriormente, es un producto entre una base de datos NoSQL y un framework. Se utiliza para administrar y escalar datos distribuidos utilizando un método de búsqueda conocido como MapReduce, un método desarrollado por Google para almacenar y consultar datos a través de sus vastas redes de datos.
- **Computación en la Nube:** Los proveedores externos ofrecen soluciones de hosting que proporcionan acceso a la potencia computacional, almacenamiento de datos y frameworks. Esta es una excelente solución para las empresas que quieren participar en analítica de datos más sofisticada, especialmente Big Data, pero no tienen la capacidad de invertir en servidores de computadoras y contratar a técnicos para administrarlos. Los productos conocidos incluyen: Amazon Web Services (AWS), Cloudera, Microsoft Azure y IBM SmartCloud.

## Herramientas Blandas

- **Lenguajes:** 'R' y Python son dos lenguajes de programación que se han vuelto esenciales para la ciencia de los datos. Ambos ofrecen los beneficios de elaboración rápida de prototipos y el análisis exploratorio que pueden poner proyectos de datos en marcha rápidamente. Ambos también incluyen bibliotecas complementarias construidas para la ciencia de datos, permitiendo técnicas sofisticadas de aprendizaje automático o técnicas de modelamiento con relativa simplicidad de programación. Los frameworks y bases de datos también tienen sus propios conjuntos de lenguajes de programación. SQL es necesario para los sistemas de bases de datos relacionales, mientras que otras soluciones pueden requerir Java, Scala, Python o Pig para Hadoop.
- **Diseño y Visualización:** Los lenguajes básicos de la ciencia de datos básica suelen incluir bibliotecas de visualización para ayudar a explorar los patrones de datos y visualizar los resultados finales. Como muchos proyectos de datos producen paneles de control interactivos o herramientas de monitoreo basadas en datos, hay una serie de proveedores externos que ofrecen soluciones completas. Algunos proveedores externos de productos incluyen: IBM, Microsoft, Tableau, Qlik, Salesforce, DataWatch, Platfora, Pyramid y BIME, entre otros, algunos de los cuales se ejemplifican en los estudios de casos operativos del Capítulo 1.2.



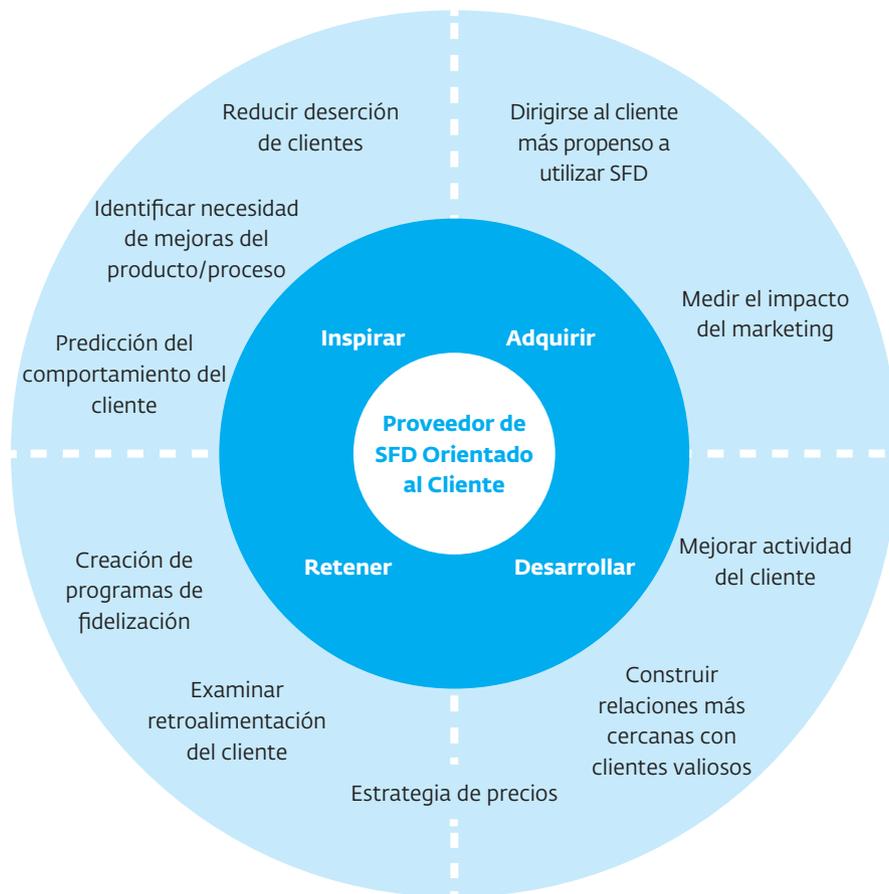
# PARTE 1

## *Capítulo 1.2: Aplicaciones de Datos Para Proveedores de Servicios Financieros Digitales*

Este capítulo cubre las tres áreas principales en las que el análisis de datos permite a las empresas enfocarse en el cliente, construyendo así una mejor propuesta de valor y generando valor de negocio para el proveedor de Servicios Financieros Digitales. En primer lugar, examina el papel que el análisis de los datos puede desempeñar para mejorar el conocimiento de los clientes por parte del proveedor de SFD. En segundo lugar, ilustra cómo los datos pueden desempeñar un papel más importante en las operaciones cotidianas de un proveedor de Servicios Financieros Digitales típico. Finalmente, analiza el uso de datos alternativos en las evaluaciones y decisiones de crédito. Estas secciones presentarán una serie de casos de uso para demostrar el potencial de la ciencia de los datos para los proveedores de SFD, pero no son exhaustivos. Las posibilidades de negocios que ofrece la ciencia de los datos está limitada sólo por la disponibilidad de datos, métodos y habilidades necesarias para utilizar los datos. Seguidamente, presentamos una serie de ejemplos para alentar a los proveedores de Servicios Financieros Digitales a comenzar a pensar formas en que los datos puede ayudar a sus operaciones existentes a alcanzar el siguiente nivel de desempeño e impacto.

La Figura 8 ilustra cómo el análisis de datos puede desempeñar un papel de apoyo en la toma de decisiones de un negocio de Servicios Financieros Digitales, a lo largo del ciclo de vida del cliente y en las tareas operativas correspondientes. Como tal, los datos juegan un papel clave en ayudar a los proveedores de SFD a enfocarse más en el cliente. No hace falta decir que todas las organizaciones dependen de la fidelidad del cliente. La orientación al cliente consiste en establecer una relación positiva con los clientes en cada etapa de la interacción, con el objetivo de impulsar la fidelidad, beneficios y los negocios de los clientes. Esencialmente, los servicios orientados al cliente proporcionan productos basados en las necesidades, preferencias y aspiraciones de su segmento, incorporando esta comprensión en los procesos operativos y en la cultura.

### Ciclo de Vida del Cliente



*Figura 8: Oportunidades existentes para aplicaciones de datos a lo largo del Ciclo de Vida del Cliente*

## 1.2\_APLICACIONES DE DATOS

Ser receptivo a los clientes es clave cuando se desea enfocar en el mismo. Es útil entender por qué los clientes se van y cuándo es más probable que se vayan, para que se puedan tomar medidas apropiadas. Algunos clientes se irán inevitablemente y se convertirán en antiguos clientes. El uso de analítica de datos para entender cómo estos clientes se han comportado a lo largo del ciclo de vida puede ayudar a los proveedores a desarrollar indicadores que alerten al negocio cuando sea probable que los clientes se vayan. También puede ofrecer nociones sobre cuáles de estos clientes pueden ser recuperados por el proveedor de servicios y cómo hacerlo.

Los proveedores de Servicios Financieros Digitales suelen atender a personas que anteriormente no tenían acceso a bancos u otros servicios financieros, así como a otros clientes mal atendidos. Esto plantea desafíos especiales para los proveedores ya que primero tienen que establecer confianza y fe en un nuevo sistema para sus clientes. Tales clientes pueden tener ingresos irregulares, ser más susceptibles a los shocks económicos y pueden tener un comportamiento de gastos diferente. Por último, la necesidad de protección

del consumidor para este segmento es mayor porque podrían tener menos acceso a información, menores niveles de alfabetización y un mayor riesgo de fraude en comparación con otros segmentos. Los proveedores de SFD tendrán que entender las necesidades particulares de estos clientes y luego diseñar procesos operativos que reflejen esta comprensión. Así, entender a los clientes y la entrega de valor al cliente es crucial para los proveedores de SFD, y los datos pueden ayudar a ser más orientados al cliente.

### 1.2.1 Analítica y Aplicaciones: Conocimiento de Mercado

Esta sección muestra cómo usar los datos para desarrollar un entendimiento más preciso y matizado de los clientes y los mercados, lo que a su vez puede ayudar a un proveedor de servicios a desarrollar productos y servicios que estén alineados con las necesidades del cliente. Como se describió en el capítulo anterior, los proveedores de Servicios Financieros Digitales tienen acceso a datos valiosos de clientes en una variedad de formas.

Estos datos pueden ser manipulados y analizados para ofrecer un conocimiento granular del mercado. Este tipo de análisis suele implicar un conjunto diverso de métodos y datos cuantitativos y cualitativos. Esta sección comienza con un estudio de caso para ilustrar cómo los pequeños pasos, para incorporar un enfoque basado en datos, pueden aportar mayor precisión al entendimiento de las preferencias del cliente. A continuación, sigue con una discusión de cómo se pueden utilizar los datos para entender la gestión de clientes con un producto de Servicios Financieros Digitales con el fin de mejorar la actividad del cliente y reducir la rotación de los mismos. Luego, explica cómo utilizar la segmentación para identificar grupos específicos dentro de la base de clientes y cómo utilizar este conocimiento para mejorar los esfuerzos de focalización. Esto es seguido por una discusión de cómo los proveedores de Servicios Financieros Digitales pueden aprovechar las nuevas tecnologías para predecir el comportamiento financiero y mejorar la adquisición de clientes. Por último, esta sección examina formas de interpretar la retroalimentación de los clientes para mejorar los productos y servicios existentes.

# CASO 1

## Zoona - Pruebas de estrategias de marketing para un impacto óptimo

### Como Desarrollar Hipótesis para Mensajes de Marketing Exitosos y Probarlos

Zoona es un Proveedor de Servicios de Pagos con operaciones en Zambia, Malawi y Mozambique, donde busca convertirse en el principal proveedor de transferencias de dinero y cuentas de ahorro simples para las masas. El marketing es a menudo una actividad que consume mucho tiempo y requiere muchos recursos, y puede ser difícil medir el impacto. Zoona manejó algunos de estos desafíos utilizando un enfoque orientado al cliente para probar tres estrategias de marketing diferentes para un producto nuevo de depósito llamado Sunga. Primero, realizó un piloto de tres meses del producto Sunga en una zona, luego extendió el piloto a otros tres municipios para probar tres estrategias de marketing diferentes, todo con el fin de identificar el enfoque más impactante

para el lanzamiento a nivel nacional. La primera estrategia se llamó 'Gratificación instantánea', y otorgó a todos los clientes que abrieron una cuenta un brazalete gratis, así como una alta probabilidad de recibir un pequeño reembolso de efectivo cada vez que hicieran un depósito. En la segunda estrategia, llamada 'Lotería', los clientes tenían una baja probabilidad de ganar un premio grande, con sólo cuatro ganadores seleccionados durante dos meses. El tercer enfoque incluía embajadores de apertura de cuentas que acudían a áreas de alta actividad, como los mercados, para invitar a las personas a abrir cuentas.

Las estadísticas del primer mes de este piloto ampliado se presentan a continuación. Los números se han indexado con respecto al municipio

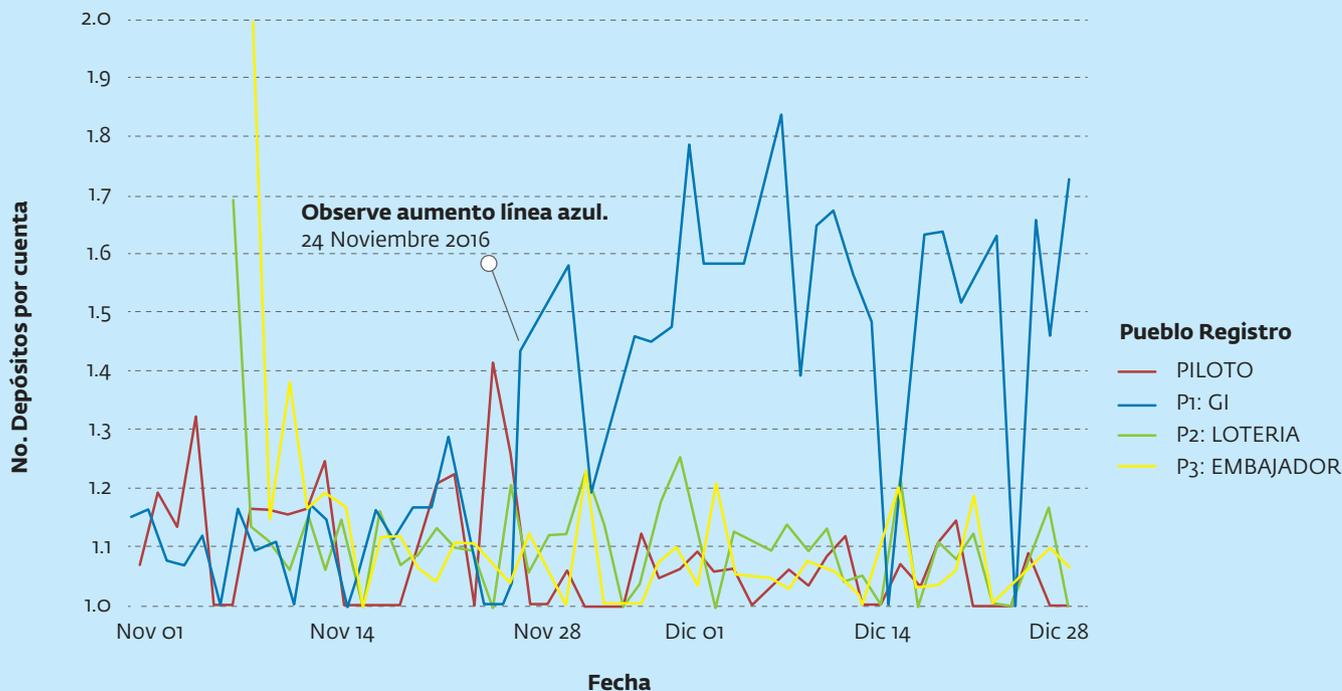
piloto inicial, por lo que 1.3 indica que los resultados son un 30 por ciento mejores que el piloto de referencia. El análisis muestra que la metodología de lotería fue la menos popular, mientras que el mayor número de cuentas abiertas fue acreditado a la estrategia del embajador. Estas cuentas también tenían altos valores de depósito. Zoona también analizó las tasas de actividad de los clientes, medidas como el número de depósitos por cuenta. El enfoque de gratificación instantánea fue el claro ganador. En la Figura 9, el 24 de noviembre es la fecha en que los depositantes comenzaron a ganar pequeños reembolsos de efectivo cada vez que depositaban en sus cuentas: la línea azul muestra que los depósitos aumentan significativamente.

#### Comparación de Estrategias de Marketing, Tabla de Resultados

INDEXADO (Primeros 30 días)	#Registros	Valor Depositado
Piloto	1.0	1.0
P1: Gratificación Instantánea	1.4	1.9
P2: Lotería	1.1	1.8
P3: Embajador	3.0	3.8

Tabla 1: Comparación de Resultados, estrategia 'embajador' aumenta aperturas de cuenta 300% sobre línea base

## 1.2\_APLICACIONES DE DATOS



**Figura 9:** Resultados de las pruebas de campaña de incentivos para clientes

El resultado del análisis fue apoyado por llamadas de seguimiento a los clientes. La retroalimentación reveló que la gratificación instantánea también impulsó el marketing boca-a-boca, ya que el 88 por ciento

de los del grupo de gratificación instantánea le contó a un familiar o amigo acerca del producto. Como resultado, la estrategia de marketing a nivel nacional ahora combina las estrategias 'Embajador'

y 'Gratificación instantánea' - la primera para estimular las aperturas de cuentas y la segunda para estimular los niveles de actividad de los clientes.



Este estudio de caso ilustra que un enfoque riguroso para probar las estrategias de marketing no tiene que conllevar metodologías complicadas. Por el contrario, un enfoque sistemático y la planificación, utilizando la iteración rápida de las técnicas medidas por las tasas de respuesta de los clientes, pueden crear resultados medibles. También se destaca el beneficio de combinar metodologías para llegar al comportamiento deseado del cliente.

## Caso de Uso: Entendiendo la Interacción con el Producto para Ofertas de SFD

Entender cómo un cliente usa o no usa un producto o servicio es importante para hacer mejoras en el área apropiada de operaciones con el fin de extender el alcance y aumentar la adopción. Los datos transaccionales y los datos de perfil del cliente proporcionan información valiosa sobre cómo los clientes se vinculan con un producto con el paso del tiempo. Esta retroalimentación puede utilizarse

para desarrollar mensajes efectivos para el producto o para desarrollar acciones para gestionar la interacción del cliente con el producto. Los altos niveles de registro combinados con bajos niveles de actividad suelen implicar que el costo de adquirir y mantener activos a los clientes es innecesariamente alto. Los datos transaccionales, así como los datos geoespaciales, puede ofrecer a los proveedores información sobre los niveles de actividad tanto de los clientes como de los corresponsales. Esta información pueden

ayudar al proveedor de servicios a efectuar cambios en todo el negocio para alinearse con el comportamiento y las necesidades del cliente. Por ejemplo, este tipo de análisis puede ayudar a informar estrategias de marketing, estrategias de reclutamiento de corresponsales o la adopción de procesos de mejores prácticas para corresponsales. La Figura 10 proporciona una ilustración sencilla de cómo se pueden interpretar los datos transaccionales. El proceso analítico de datos también se explora con mayor detalle en el Capítulo 2.1.



Figura 10: El Proceso de Análisis e Interpretación de Datos

## 1.2\_APLICACIONES DE DATOS

### Mejorando la Actividad de Clientes

Un análisis transaccional simple como se ha visto anteriormente puede, por ejemplo, revelar que los clientes muy activos están asociados con corresponsales específicos. Para poder actuar sobre esta información, será necesario averiguar por qué este es el caso. ¿Podría ser debido a las mejores prácticas adoptadas por los corresponsales, debido a la ubicación geográfica, o debido a alguna otra variable? Como ejemplo, se podrían realizar entrevistas para comprender mejor las técnicas de los corresponsales, y los datos geoespaciales podrían utilizarse para comprender mejor el impacto de la ubicación en la actividad de los corresponsales y los clientes. Grupos de actividad muy alta o muy baja indican a menudo la necesidad de investigación y discusiones de grupos focales más profundas, para entender las razones detrás de estas.

### Reduciendo la Rotación de Clientes

Mirar de cerca los datos transaccionales puede proporcionar pistas sobre por qué los clientes están abandonando el servicio y cómo conservarlos. La frecuencia con la que los clientes interactúan con un servicio puede indicar si se acaba de adquirir, si son clientes activos del servicio o si es necesario volver a ganárselos de vuelta al servicio. Diferentes mensajes y canales son relevantes para los clientes en cada una de estas etapas. En general, mantener

a los clientes existentes es mucho menos costoso que adquirir nuevos. Un gran número de clientes que nunca han hecho transacciones indican un enfoque inadecuado en la etapa de vinculación. Un número elevado de clientes inactivos puede indicar otras limitaciones en la oferta de servicios, que pueden mejorarse mediante pequeñas mejoras de productos o procesos.

### Caso de Uso: Segmentación

Los segmentos pueden ser delineados por marcadores demográficos, marcadores de comportamiento tales como patrones de uso de SFD, datos geográficos u otros datos externos de MNOs, tales como uso y compra de tiempo aire y datos. Comprender los segmentos es necesario para descubrir las necesidades y deseos de grupos específicos, así como para diseñar estrategias de ventas y marketing bien orientadas. La información de la segmentación, destinada a ampliar las perspectivas de generación de ingresos en cada segmento único, es fundamental dentro la hoja de ruta estratégica de una institución. La segmentación de clientes es un aspecto crucial para convertirse en una organización orientada al cliente, que los atienda bien, tome decisiones inteligentes de inversión y mantenga un negocio saludable.

En principio, muchos proveedores de Servicios Financieros Digitales reconocen la importancia de la segmentación. Sin embargo, en la práctica, la mayoría de

los proveedores de SFD prestan servicios al mercado masivo en contextos de países en desarrollo como un solo segmento o utilizan segmentación demográfica básica para entender a los clientes. La razón de la limitada incorporación de la segmentación en la generación de información de clientes tiene dos vertientes. En primer lugar, los proveedores de SFD asediados en mercados altamente competitivos pueden verse alentados por el éxito de ciertos productos y pueden sentirse obligados a adoptar un enfoque centrado en el producto, en lugar de orientado al cliente. Así, los proveedores de SFD pueden dejar de pensar en los diferentes usos posibles para sus ofertas en función de las necesidades y preocupaciones del cliente. Por el contrario, pueden optar por resaltar casos de uso muy particulares y mensajes para un producto. Por ejemplo, mientras que el producto de transferencia de dinero electrónico de M-Pesa fue muy exitoso en Kenia, los MNOs en otros mercados no han tenido el mismo éxito, enfatizando la necesidad de mirar el mercado y el comportamiento de los clientes y sus necesidades, mercado por mercado antes de desplegar productos. En segundo lugar, hay una falta de conciencia sobre cómo segmentar eficazmente la base de clientes y cómo usar este análisis de segmentación. La segmentación no necesita ser complicada o costosa. Expertos en el área deben definir claramente los objetivos de negocio, los cuales pueden guiar el ejercicio de segmentación.

## Segmentación de Clientes



Figura 11: Ejemplos de Segmentos de Clientes de SFD según actividad del producto

El siguiente marco presentado por el Consultative Group to Assist the Poor (CGAP) ilustra cómo los diferentes tipos de segmentación pueden ser empleados por un profesional dependiendo de sus necesidades:<sup>17</sup>

Tipo de Segmentación	Ejemplo	Necesidades de Datos	Ventajas	Desventajas
<b>Demográfica</b>	<ul style="list-style-type: none"> <li>Rural vs. Urbano</li> <li>Hombre vs. Mujer</li> <li>Adulto vs. Joven</li> </ul>	<ul style="list-style-type: none"> <li>Información de Vinculación y Conozca a Su Cliente (KYC)</li> </ul>	<ul style="list-style-type: none"> <li>Sencillo</li> <li>Los datos son fáciles de encontrar</li> </ul>	<ul style="list-style-type: none"> <li>Falta de uniformidad dentro de los grupos</li> <li>Menos esclarecedor</li> </ul>
<b>Comportamiento</b>	<ul style="list-style-type: none"> <li>Nunca realizó transacciones vs. usuarios inactivos vs. activos</li> <li>Ahorradores vs. Retiradores</li> </ul>	<ul style="list-style-type: none"> <li>Base de Datos transaccional</li> </ul>	<ul style="list-style-type: none"> <li>Los datos son fáciles de encontrar</li> <li>Fácil de atribuir valor al cliente</li> </ul>	<ul style="list-style-type: none"> <li>Falta de conclusiones sobre la vida del cliente, necesidades, aspiraciones</li> <li>Menos útil para los mensajes de marketing</li> </ul>
<b>Demográfica y por Comportamiento</b>	<ul style="list-style-type: none"> <li>Estudiantes</li> <li>Trabajadores migrantes que envían dinero a casa</li> </ul>	<ul style="list-style-type: none"> <li>Información de Vinculación y KYC</li> <li>Base de Datos transaccional</li> <li>Investigación primaria de mercado</li> </ul>	<ul style="list-style-type: none"> <li>Asigna valor a un cliente y proporciona información sobre su vida y necesidades</li> <li>Mensajes de marketing más fáciles de desarrollar</li> </ul>	<ul style="list-style-type: none"> <li>Los datos son relativamente más difíciles de encontrar</li> <li>Puede tener segmentos superpuestos</li> </ul>
<b>Psicográfica</b>	<ul style="list-style-type: none"> <li>Mujeres que quieren un sitio seguro para ahorrar</li> <li>Los clientes que creen que el acceso al dinero electrónico implica mayor estatus</li> <li>Consciente de su Presupuesto</li> </ul>	<ul style="list-style-type: none"> <li>Datos transaccionales históricos profundos y ricos</li> <li>Investigación primaria</li> </ul>	<ul style="list-style-type: none"> <li>Respuesta fuerte a las aspiraciones del cliente</li> <li>Propuesta de valor fuerte</li> <li>Mensajes de marketing más fáciles de desarrollar</li> </ul>	<ul style="list-style-type: none"> <li>Datos difíciles de encontrar</li> <li>Puede tener segmentos superpuestos</li> <li>Podría ser un segmento muy dinámico, es decir, quiere, podría cambiar</li> </ul>

Tabla 2: Marco de Segmentación del Cliente CGAP

<sup>17</sup> CGAP (2016). Customer Segmentation Toolkit

## CASO 2

### *Tigo Cash Ghana Aumenta el Uso de la Billetera Electrónica Activa*

#### **Modelos de segmentación de clientes mejoran la vinculación y activación de clientes**

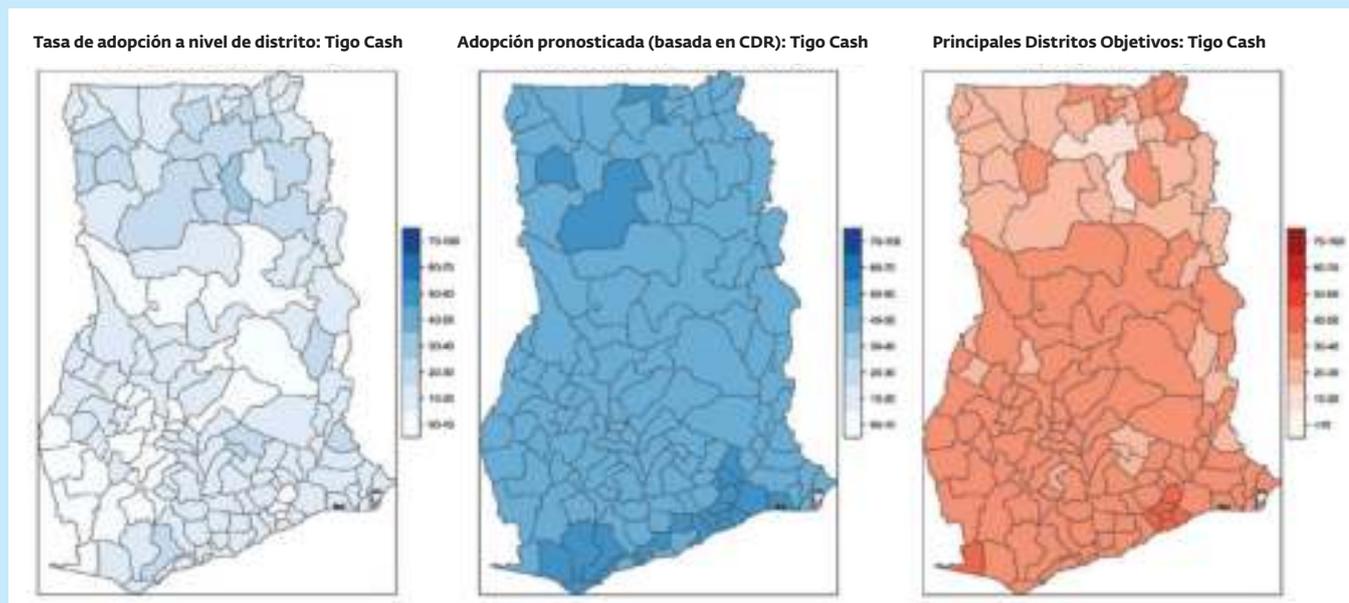
Tigo Cash, lanzado en Ghana en abril de 2011, es el segundo mayor proveedor de dinero electrónico en términos de usuarios registrados. A pesar de las altas tasas de inscripción, conseguir que los clientes realicen varias transacciones por medio de dinero electrónico sigue siendo un reto y un enfoque clave. Las tasas de vinculación de clientes

y el mantenimiento de las tasas de actividad siguieron siendo un objetivo clave después del lanzamiento del servicio. Una base de clientes con transacciones activas no es sólo un desafío en Ghana; el GSMA estima que las tasas de actividad globales son tan bajas como del 30 por ciento.

En 2014, Tigo Cash Ghana se asoció con IFC para un análisis predictivo

con el fin de identificar usuarios móviles de voz y datos que tuvieran alta probabilidad de convertirse en usuarios activos de dinero electrónico. Para ello, un equipo de científicos de datos analizó durante seis meses casi dos terabytes de CDR y datos transaccionales.

Los resultados del análisis sugieren que existen diferencias entre los



**Figura 12:** Distritos actuales, previstos y principales en el uso de dinero móvil

clientes a través de un gran número de métricas de uso de teléfonos móviles, la estructura de la red social y la movilidad individual y de grupo. Hay fuertes diferencias entre suscriptores de voz y sólo datos, suscriptores inactivos de dinero electrónico y suscriptores activos de dinero electrónico. Se puede observar una fuerte correlación entre los altos usuarios de los servicios tradicionales de telecomunicaciones y la probabilidad de que esos usuarios también se conviertan en usuarios regulares de dinero electrónico.

Con la ayuda de algoritmos de aprendizaje automático, el equipo de investigación identificó perfiles que coinciden, entre clientes de voz y sólo datos que aún no son suscriptores de dinero electrónico, pero que es probable que se conviertan en usuarios activos. El equipo también geo-mapeó los datos (ver Figura 12) para su posterior análisis. Además, el análisis de los CDRs y los datos transaccionales se complementaron con encuestas para no sólo comprender lo que sucedió, sino también el por qué.

### **Determinantes en la Adopción de Dinero Electrónico**

La necesidad de más educación al cliente y la adaptación del producto es algo que salió claramente a través de las encuestas individuales. Sólo

una pequeña proporción de usuarios de dinero electrónico informó que la no disponibilidad de corresponsales les impidió utilizar servicios de dinero electrónico. Los bajos niveles de uso estaban más estrechamente vinculados a la falta de conocimiento de las personas acerca de la propuesta de valor del dinero electrónico o la percepción de que no tenían suficiente dinero para usar los servicios.

### **Nuevos Clientes**

El modelado predictivo resultó en 70.000 nuevos usuarios activos de dinero electrónico debido al uso del modelo único. Los resultados trazaron el grupo de posibles adoptantes de dinero electrónico e identificaron lugares en los que las actividades de marketing no convencionales estaban logrando el mayor impacto. Tener una idea exacta del potencial de marketing en diferentes áreas evita la sobreoferta de personal de ventas y aumenta la eficiencia del marketing. El enfoque basado en datos proporcionó una forma más inteligente y más informada para orientar a los suscriptores telefónicos existentes a adoptar el dinero electrónico.

### **Mejores Tasas de Actividad**

El uso de SMS y el alto volumen de voz y uso de datos móviles son

factores clave que se utilizaron para identificar a los potenciales usuarios activos de dinero electrónico. Lo que comenzó como un análisis de las CDRs históricas, aportó un valor de prueba de concepto y llevó a desarrollar un enfoque basado en datos que permitió a Tigo Cash superar la marca de 65 por ciento de actividad entre sus clientes de dinero electrónico. La base de clientes activos creció de 200.000 antes del ejercicio, a más de 1 millón de clientes activos en 90 días.

### **Cambio de Mentalidad Institucional**

El resultado de la colaboración se convirtió en la base de todo el trabajo de adquisición de clientes de Tigo Cash Ghana. Por encima de todo, la analítica de datos mostró el valor de conocer a los clientes. Tigo Cash Ghana planea aumentar su capacidad interna de ciencia de datos, así como mejorar aún más su entendimiento del cliente con investigación primaria adicional. El objetivo ahora ha pasado de vincular nuevos clientes que se espera que sean activos, a pensar a futuro sobre maneras de mantener niveles de actividad alta de una manera sostenible.



Un enfoque institucional hacia la adquisición y retención de clientes puede cambiarse y mejorarse de forma fundamental simplemente haciendo uso de los datos existentes para tomar decisiones operativas más informadas.

## 1.2\_APLICACIONES DE DATOS

### Programas de Marketing Focalizados

Focalizarse en los grupos de mercado adecuados, con las campañas de publicidad y marketing adecuadas, puede aumentar mucho la eficacia de una campaña en términos de captación y uso. Usando una combinación de fuentes de datos, los proveedores de Servicios Financieros Digitales pueden segmentar datos transaccionales por parámetros demográficos para identificar grupos estratégicos dentro de su base de clientes. Los programas de marketing se pueden personalizar para enfocarse en estos grupos, a menudo con mayor eficiencia y eficacia que los enfoques estándar. Se ha visto que los proveedores de SFD combinan el conocimiento de segmentos con los datos sobre la rentabilidad a fin de enfocar los esfuerzos de marketing en segmentos con probabilidad de optimizar los beneficios. Del mismo modo, otros proveedores de SFD han utilizado los ciclos de vida del cliente para hacer las ofertas de productos correctas para los clientes adecuados. El principal desafío aquí es encontrar qué preocupa a los grupos de clientes con el fin de diseñar una campaña de marketing adecuada. Mientras que el universo de datos disponibles para los proveedores de SFD está creciendo cada día, en ausencia de análisis para arrojar luces sobre esto, una vez que se identifican los grupos de clientes, los proveedores de SFD pueden utilizar la investigación

primaria para identificar qué le preocupa a los segmentos. Todos los datos de los clientes pueden utilizarse para desarrollar programas de marketing focalizados. Sin embargo, es probable que los resultados sean más precisos si el análisis se realiza en los miembros de segmentos determinados de clientes.

### Campañas de Fidelización y Promoción

Puede haber segmentos de clientes que realicen un número muy alto de transacciones en el canal SFD. Estos segmentos pueden desear recompensas de fidelidad para transacciones específicas tales como pagos a ciertos tipos de comerciantes. Alternativamente, el proveedor de SFD puede ser capaz de empujar a otros segmentos hacia ciertos tipos de transacciones ofreciendo campañas promocionales. Las transacciones específicas en la base de datos y los perfiles de clientes ayudarían a identificar qué grupos se beneficiarían de dichas campañas.

### Relaciones de Alto Valor con Clientes

La segmentación de los clientes basada en rentabilidad es una aplicación común del proceso de segmentación. Además, se pueden evaluar los grupos propensos a convertirse en importantes en el futuro. Los proveedores de SFD pueden usar esta información para aumentar su cuota de mercado de este grupo y para reducir la asignación de recursos a grupos menos rentables.

Los datos necesarios para este tipo de análisis son datos demográficos de los clientes, datos transaccionales y datos sobre la rentabilidad del cliente.

Esto es igualmente aplicable a la identificación de corresponsales de alto desempeño basándose en la segmentación. Trabajando con FINCA en la República Democrática del Congo (RDC), IFC analizó los datos transaccionales de corresponsales y los formularios de vinculación en la República Democrática del Congo para mostrar que ser mujer y estar involucrada en un negocio orientado a servicios está altamente correlacionado con ser un corresponsal de alto desempeño.<sup>18</sup>

### Mejoras en Productos o Procesos

La clasificación de los clientes en segmentos también permite a los proveedores de SFD prestar mayor atención a las necesidades específicas de una cohorte representativa. En un grupo más grande, estas necesidades pueden perderse de vista - pero prestar atención a los segmentos más pequeños permite a los proveedores de SFD afilar su enfoque y explorar las necesidades o deseos desatendidos o ignorados. Por ejemplo, dentro de un grupo de personas que no utilizan un servicio, podría haber algunos que son *clientes inactivos*, o aquellos que transaron un par de veces, pero luego dejaron de usar el servicio. Hablar con estos usuarios podría revelar la necesidad de realizar pequeños cambios en el producto o proceso. De forma

<sup>18</sup> Harten and Rusu Bogdana, 'Women Make the Best DFS Agents'. IFC Field Note 5, The Partnership for Financial Inclusion

alternativa, los clientes de un segmento pueden utilizar la gama completa de productos ofrecidos por un proveedor de SFD, mientras que otro segmento puede utilizar sólo uno o dos de estos productos. En tales casos, la segmentación proporciona información para una investigación de mercado focalizada y para el desarrollo de productos con el objetivo de desbloquear la demanda de los clientes.

### **Oportunidad de Mercado y Productos Prioritarios**

Una vez completado el ejercicio de segmentación, los proveedores de Servicios Financieros Digitales pueden evaluar hasta qué punto su oferta de productos satisface las necesidades y deseos de cada segmento. Pueden estimar qué segmentos representan la mayor oportunidad a lo largo del tiempo y qué tan competitiva es su oferta dentro de estos segmentos cruciales de crecimiento. De esta forma, un análisis basado en la segmentación puede jugar un papel importante en la hoja de ruta estratégica de un proveedor de SFD.

La segmentación demográfica tradicional - que puede basarse en la edad, en los ingresos o en la geografía - es útil, pero la experiencia demuestra que la segmentación

demográfica es menos predictiva de la relación futura de una institución con un cliente que la segmentación basada en características de comportamiento. La agrupación de clientes basada en datos demográficos tiende a dar el mismo tratamiento a todos los clientes de un grupo, independientemente de su nivel de actividad en el canal. La demografía también puede ser de naturaleza estática, donde - particularmente en el mundo del acceso financiero habilitado por tecnología - el comportamiento del cliente es dinámico y en constante cambio.

El acceso a las bases de datos transaccionales puede transformar la segmentación tradicional en una poderosa herramienta para generar información de los clientes. Con la mayor disponibilidad de datos, nuevas herramientas de análisis de datos y múltiples canales disponibles para los clientes, los proveedores de Servicios Financieros Digitales ahora tienen la opción de usar información de comportamiento individual. Esta información predice mejor las necesidades y usos financieros de las personas. Además, refleja las necesidades y actividades cambiantes del cliente. Sin embargo, los datos de comportamiento pueden no tener mucha información sobre

las necesidades y aspiraciones del cliente, lo que dificulta el desarrollo de mensajes perceptivos en torno a estos segmentos.

La realización de un ejercicio de segmentación en la base de datos de clientes requiere recursos dedicados y un plan detallado. En particular, las estrategias de segmentación que hacen uso de múltiples fuentes de datos son más exitosas a la hora de describir de manera útil y precisa a los grupos de clientes. Así, el proceso para desarrollar la segmentación del cliente debe incorporar este enfoque. La analítica de datos juega un papel importante en este proceso, ya que permite a los proveedores de SFD segmentar exactamente por las variables que juegan un papel en dirigir el uso y la adopción. Este informe sólo analiza el papel de la analítica de datos para facilitar este proceso, pero es importante señalar que esos segmentos pueden crearse a través de múltiples tipos de investigación y análisis.

# CASO 3

## *Airtel Money - Aumento de Actividad con Modelos Predictivos de Segmentación de Clientes*

### **Modelo de Segmentación con Aprendizaje Automático Ofrece Valor Operacional y Visión Estratégica**

*Airtel Money, la oferta Servicios Financieros Digitales de Airtel Uganda, se lanzó en 2012. La adopción inicial fue baja, con sólo una fracción de sus 7.5 millones de suscriptores GSM vinculándose al servicio. Los niveles de actividad también fueron bajos, con alrededor del 12,5 por ciento de usuarios activos. IFC y Airtel Uganda colaboraron en un estudio de investigación para utilizar la analítica de Big Data y el modelado predictivo para identificar a los clientes GSM existentes con probabilidad de convertirse en usuarios activos de Airtel Money.*

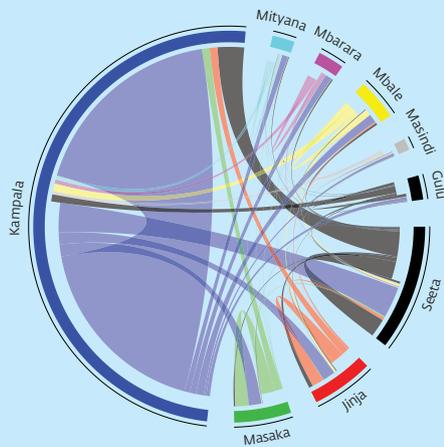
*El proyecto analizó seis meses de transacciones de CDR y Airtel Money. El análisis buscó segmentar usuarios de dinero electrónico muy activos, activos y no activos. El estudio identificó tres categorías diferenciadoras: Los niveles de actividad GSM, el gasto móvil mensual y la conectividad del usuario. Usando métodos de aprendizaje*

*automático, un modelo predictivo fue capaz de identificar a los posibles usuarios activos con una precisión del 85%. Esto produjo 250.000 clientes de Airtel Money de ‘alta probabilidad’, nuevos y activos, de la base de suscriptores GSM para que Airtel les llegara con marketing focalizado. El análisis de redes geoespaciales y de clientes ayudó a identificar nuevas áreas de interés estratégico, mapeadas en función del nuevo potencial de adopción.*

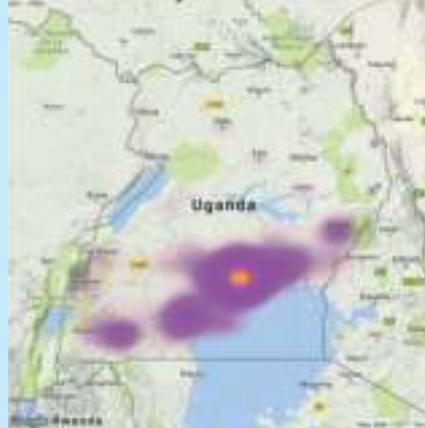
*El modelo de aprendizaje automático identificó algunas variables con alta fiabilidad estadística, pero que no tenían sentido comercial, tales como la ‘entropía de duración de voz’. Como resultado, un análisis complementario dio métricas de reglas de negocio, o indicadores que tenían una buena correlación con la actividad potencial y también tenían relaciones altas con KPIs del negocio. Cada métrica tenía un punto de corte numérico para*

*orientar a los clientes por encima o por debajo de un punto de corte determinado. Si bien no es tan preciso como otros modelos, proporcionó un sólido ‘corte rápido’ que se podía utilizar contra los KPI para evaluar rápidamente las expectativas.*

*Finalmente, el estudio analizó los corredores del movimiento de dinero electrónico dentro de la región. Encontró que el 60 por ciento de todas las transferencias ocurren dentro de un radio de 19 kilómetros en Kampala y sus alrededores. El entendimiento de esta necesidad de remesas de corta distancia también informó los esfuerzos de marketing de Airtel Money para transferencias P2P. Por otra parte, este análisis de redes de transacciones P2P identificó otras ciudades y zonas rurales con corredores de actividad que podrían dar lugar a relaciones estratégicas más allá de Kampala para que Airtel se centre en el crecimiento.*



Transacciones P2P enviadas por número de origen



Ubicación de los clientes del CDR



**Figura 13:** Análisis de red (izquierda) de los flujos P2P entre ciudades y robustez del canal. También se muestra la densidad geoespacial de las transacciones P2P de Airtel Money (centro), en comparación con la distribución de uso GSM (derecha). Datos a 2014.



El análisis avanzado de analítica de datos puede proporcionar información sobre segmentos de clientes activos y altamente activos que pueden impulsar modelos de propensión para identificar clientes potenciales con alta precisión. El análisis de redes y geoespacial puede proporcionar información para priorizar la planificación estratégica del crecimiento.

## 1.2\_APLICACIONES DE DATOS

### Caso de Uso: Pronóstico del Comportamiento del Cliente

El modelado predictivo es una herramienta de toma de decisiones que utiliza datos pasados de clientes para determinar la probabilidad de resultados futuros. Los proveedores de Servicios Financieros Digitales evalúan la información multidimensional de los clientes para identificar las características del cliente que están correlacionadas con los resultados deseados. Como parte del modelado, a cada cliente se le asigna un score o clasificación que calcula la probabilidad del cliente de tomar una determinada acción.

Para una institución orientada al cliente, el modelo predictivo puede informar cómo entiende y responde a las necesidades del cliente. Sin embargo, quedan algunos impedimentos que evitan su uso más extendido. Hay una percepción - que ahora está cambiando gradualmente entre los proveedores de SFD - que los proveedores ya conocen su base de clientes lo suficientemente bien como para entender qué productos y campañas de marketing funcionan. Alternativamente, algunos proveedores de Servicios Financieros Digitales miran lo que ha funcionado en otros lugares y tratan de replicar productos y servicios similares en sus propios mercados. Muchos proveedores tampoco saben exactamente cómo y dónde comenzar el proceso.

El análisis predictivo puede ayudar a los expertos en el área a alcanzar los siguientes objetivos:

- Adquisición de nuevos clientes
- Desarrollo de una oferta de producto óptima
- Identificar los objetivos del cliente y predecir el comportamiento del cliente
- Prevención de la rotación
- Estimar el impacto del marketing

### Nueva adquisición e Identificación de Objetivos

Como lo demuestran las investigaciones y la experiencia de los especialistas, los expertos en el área han registrado con éxito un gran número de nuevos clientes para sus SFD. Sin embargo, la transformación de estos clientes registrados en clientes activos sigue siendo una tarea difícil que sólo unos pocos proveedores de SFD han podido dominar. En promedio, alrededor de un tercio de los clientes registrados han realizado una sola transacción en los últimos 90 días.<sup>19</sup> Una de las razones identificadas para estos bajos niveles de actividad es la focalización inadecuada en la etapa de vinculación. La mayoría de las ofertas de Servicios Financieros Digitales apuntan al vasto mercado masivo. Como tal, son capaces de vincular un gran número de clientes, pero han tenido un éxito limitado convirtiendo a estos clientes en una base de clientes activa y generadora de ganancias.

El análisis predictivo podría ayudar a identificar a los clientes en la etapa de vinculación que son mucho más propensos a convertirse en usuarios activos en el futuro a través de una técnica estadística conocida como modelado de respuesta. El modelado de respuesta utiliza el conocimiento existente de una base de clientes potenciales para proporcionar una score de propensión a cada cliente potencial. Cuanto mayor sea el score, más probable es que el cliente se convierta en un usuario activo. Los MNO que son proveedores de Servicios Financieros Digitales han utilizado este tipo de modelado para predecir qué miembros de su base de clientes de voz y datos pueden convertirse en usuarios activos de sus SFD. El modelo se basa en la hipótesis de que los clientes que probablemente gasten más en voz y datos también son propensos a adoptar Servicios Financieros Digitales. Usando datos CDR, el modelo es capaz de predecir con un alto grado de exactitud la probabilidad de que un cliente se convierta en un usuario activo de SFD.

### Desarrollo de Ofertas Óptimas de Producto

Existen modelos predictivos que pueden usarse para descubrir qué paquetes de productos pueden ser utilizados conjuntamente por los clientes. Así, el modelo identificará segmentos que tienden a utilizar sólo un único producto como las transferencias P2P y otros que hacen uso de múltiples productos,

<sup>19</sup> 'State of the Industry Report on Mobile Money', Decade Edition 2006 - 2016, GSMA

como servicios de depósito, compra de tiempo aire y transferencias P2P. Sin embargo, el segundo grupo nunca puede usar el servicio para microcréditos. Esta es la información que el proveedor de Servicios Financieros Digitales puede utilizar para fines de marketing y desarrollo de productos.

### **Predicción del Comportamiento del Cliente**

Este análisis también puede usarse para entender el valor potencial futuro de cada cliente. Esto incluye el valor del cliente, la fidelidad del cliente, la compra esperada y el comportamiento de uso, y la respuesta esperada a campañas y programas. Del mismo modo, los proveedores de Servicios Financieros Digitales pueden aumentar sus oportunidades de venta mejorada y venta cruzada al predecir el uso futuro a través de la canasta actual de productos y patrones de uso. Determinar qué paquetes de productos funcionan juntos a través de analítica de datos transaccionales también presenta una oportunidad para la venta cruzada. Por ejemplo, un Proveedor de Servicios de Pago puede encontrar que los usuarios están usando la billetera electrónica como una cuenta de almacenamiento, una indicación de que estos clientes pueden ser atendidos más eficazmente a través de una cuenta de ahorros.

Esta información puede utilizarse a través de varias funciones operativas: diseño de campañas y marketing, proyecciones financieras, asignación de la inversión del

cliente y desarrollo futuro de productos. Este tipo de predicción también se puede utilizar a nivel individual del cliente o a nivel agregado para un segmento como un todo.

En particular, un análisis predictivo completo del valor del cliente requiere un alto nivel de clientes activos en las áreas de producto y canal. Esto puede no ser realista para muchos proveedores de SFD. Sin embargo, a medida que las organizaciones crecen, ser capaces de pronosticar los patrones y tendencias de los clientes futuros no sólo será posible, sino imprescindible para hacer crecer un negocio saludable. Así, estar al tanto de esta funcionalidad puede ayudar a los proveedores de SFD a incorporarla en su proceso de toma de decisiones cuando sea relevante.

### **Prevención de Rotación**

*La rotación del cliente* ocurre cuando un cliente deja el servicio de un proveedor de SFD. El costo de la rotación incluye tanto la pérdida de ingresos futuros del cliente, así como los costos de marketing y adquisición relacionados con la sustitución del cliente perdido. Adicionalmente, en el momento de la rotación, los ingresos obtenidos del cliente pueden no haber cubierto la adquisición de ese cliente. Así, la analítica en torno a la rotación de clientes tiene dos objetivos: predecir qué clientes van a rotar y entender qué pasos de marketing pueden convertir a un cliente con un alto riesgo de abandonar, en un cliente retenido.

### **Estimar el Impacto del Marketing**

El marketing de Servicios Financieros Digitales tiende a ser intensivo en recursos debido a su relativa novedad en muchos mercados. Esto se ve favorecido por la comprensión de que un producto requiere creación de conciencia antes de lograr la aceptación del cliente. Sin una herramienta para medir el éxito, los gerentes se ven obligados a confiar en su intuición y en datos generales de ventas para evaluar el valor de sus esfuerzos de marketing. Dado que los clientes están ahora interactuando con los proveedores de SFD en múltiples canales, digitales y de otro tipo, también es difícil aislar los efectos de campañas específicas, ya que los clientes están expuestos a múltiples mensajes en un momento dado.

El modelado predictivo permite medir el impacto del marketing en el comportamiento del cliente. Dependiendo de los datos disponibles, el análisis puede permitir a los proveedores de SFD estimar el alza en las ventas que se pueden atribuir al marketing. El modelado predictivo identificará cómo las medidas específicas de marketing pueden afectar el comportamiento del cliente entre segmentos. Puede demostrar, por ejemplo, que una determinada acción de marketing o publicidad en un determinado canal puede tener una respuesta mucho mayor de ciertos segmentos en comparación con la respuesta promedio de la población.

## 1.2\_APLICACIONES DE DATOS

### Mensajes de Marketing Personalizados

Las secciones anteriores ya han discutido cómo el marketing objetivo puede utilizar un entendimiento más profundo de los segmentos de clientes. El *Marketing personalizado* es marketing focalizado en un nivel extremadamente individualizado, donde los deseos y las necesidades de un cliente individual se anticipan usando su comportamiento pasado y otra información reportada. Muchos clientes potenciales tienen experiencia limitada con los servicios financieros y muchas veces desconfían de la capacidad de ser relevantes para sus vidas. La mensajería personalizada permite a los proveedores de SFD 'hablar' a sus clientes como si los conocieran, lo que permite a los proveedores ganar la confianza de los clientes. Además, los clientes pueden tener una relación altamente personalizada con su proveedor. En los mercados competitivos, mensajes personalizados ayudarían a construir una afinidad por un servicio sobre otro. Los clientes son mucho más propensos a responder a la mensajería que responde a sus intereses, en lugar de mensajería impersonal que se refiere a una

propuesta de valor general no específica para Servicios Financieros Digitales. Por último, el mensaje de marketing correcto hará que el cliente tome medidas basándose en los mensajes que reciben, presumiblemente porque hablan de los puntos críticos para el cliente.

Algunos mensajes personalizados pueden fallar en sus objetivos específicos, ya que los mensajes no solicitados pueden ser ignorados fácilmente o, lo que es peor, pueden causar asociaciones negativas con el proveedor de SFD. Así, los mensajes personalizados deben ser cuidadosamente diseñados y focalizados a fin de garantizar que están llegando a los clientes que requieren la información.

¿Cómo pueden los proveedores de SFD personalizar mensajes de marketing?

#### 1. Recopilar datos e identificar clientes:

En primer lugar, los proveedores de SFD necesitan recopilar datos sobre sus clientes. Las fuentes de estos datos incluyen transacciones, datos demográficos, preferencias e inputs en redes sociales de los clientes.

#### 2. Entender a los Clientes:

Posteriormente, los proveedores de SFD necesitan examinar estos datos y considerar la segmentación en grupos basados en características comunes.

#### 3. Desarrollar Mensajes e Interactuar con los Clientes:

Los proveedores de SFD deben luego desarrollar mensajes e identificar los canales adecuados para comunicar mensajes a su base de clientes. El siguiente paso es involucrarse con la base de clientes a través de mensajería.

#### 4. Prueba de la Eficacia de la Mensajería:

El impacto del mensaje se puede medir usando las pruebas A/B. La personalización debe ir acompañada de pruebas para que sea posible evaluar su impacto.

#### 5. Refinar el Mensaje:

La retroalimentación de los clientes y la medición del impacto deben alimentar el refinamiento del mensaje.

## CASO 4

# *Juntos Envía Mensajes Escalables y Personalizados para Involucrar a los Clientes*

### **Fuentes de Datos: Los datos Cualitativos y Cuantitativos Mejoran la Segmentación y Alcance**

*Juntos, una compañía de tecnología de Silicon Valley, se asoció con proveedores de Servicios Financieros Digitales para construir relaciones de confianza con los usuarios finales; mejorando las tasas generales de actividad de los clientes. A nivel global, muchos proveedores de SFD experimentan inactividad alta y bajo uso del servicio. Esto desalienta a los proveedores, cuyas inversiones pueden sufrir rendimientos financieros insuficientes y cuyos clientes pueden tener acceso a servicios que no están utilizando lo suficiente. Juntos ofrece una solución a este problema mediante el uso de mensajes personalizados de involucramiento de clientes basados en estrategias de segmentación según datos, que dan resultados cuantificados.*

*Los buenos datos respaldan este enfoque. En primer lugar, Juntos realiza una investigación etnográfica para comprender mejor a los clientes del mercado. El involucramiento está siempre basado en datos cuantitativos proporcionados por el aliado de SFD, por la investigación cualitativa conductual hecha en el país y por los aprendizajes extraídos de la experiencia a nivel global. Habiendo desarrollado un entendimiento inicial del usuario final, Juntos lleva a cabo una serie de pruebas controladas aleatorizadas (RCTs) antes del lanzamiento completo del producto. Estos experimentos controlados están diseñados para probar el contenido, la sincronización de los mensajes o los patrones de distribución, y para identificar el enfoque más efectivo para la gestión de clientes.*

*Para comenzar, los mensajes se le envían a los usuarios, y los usuarios pueden responder a esos mensajes. Esto desarrolla la relación de confianza requerida. Más importante aún, esas respuestas son recibidas por un “chatbot” automatizado de Juntos que analiza los resultados de acuerdo a tres KPIs:*

- **Porcentaje de Participación:** *¿Qué porcentaje de usuarios respondió al chatbot? ¿Con qué frecuencia respondieron?*
- **Contenido de Respuestas:** *¿Qué dicen las respuestas? ¿Qué información compartieron o solicitaron?*
- **Comportamiento Transaccional:** *¿Cambió el comportamiento transaccional después de recibir mensajes durante una semana? ¿Un mes? ¿Dos meses?*

## 1.2\_APLICACIONES DE DATOS

Estos experimentos permiten entender qué clientes inactivos se activaron debido al alcance de los mensajes de Juntos, y comprender qué mensajes habilitaron una actividad más alta y más consistente. Por ejemplo, se envía un mensaje de control a un grupo de usuarios seleccionado al azar: “¿Usted puede usar su cuenta para enviar dinero a casa!” Otros podrían extraer datos del servicio para incluir el nombre del cliente: “Hola Juan, ¿sabías que puedes usar tu cuenta para enviar dinero a casa?” Tal vez otros datos se incorporarán en el mensaje: “Usted usó su cuenta por última vez hace 20 días, ¿Adónde le gustaría enviar dinero hoy?” Estos son sólo ejemplos, pero muestran cómo un mensaje

genérico se compara con un mensaje personalizado con un indicador de tiempo. Los datos etnográficos base de Juntos mejoran la comprensión cualitativa de los clientes, ayudando a construir una hipótesis respecto de cuál de los mensajes tiene más probabilidad de tener eco y, para seguidamente, someter estos mensajes a pruebas estadísticas.

La primera pregunta es si los mensajes de prueba producen resultados estadísticamente mejores en comparación con el mensaje de control genérico. Cuando la respuesta es “sí”, es importante profundizar un poco más y preguntarse sobre el encuestado y encuestar entre segmentos tales como rural o urbano; masculino o femenino;

rango de ingresos; y patrones de uso, fusionando esta información con datos etnográficos sobre el sentimiento del consumidor.

Al probar una amplia variedad de mensajes, Juntos puede segmentar grupos de usuarios de acuerdo a los mensajes que muestran una mejora estadística con el uso a lo largo del tiempo. Esto significa que se pueden elaborar mensajes muy cautivadores creados para todos, desde mujeres rurales, a hombres jóvenes, hasta personas de altos ingresos en entornos urbanos. El enfoque de Juntos se adapta a cada contexto y se ajusta continuamente para acomodar ágilmente a los clientes que cambian sus interacciones con el tiempo.



La recolección del sentimiento cualitativo del cliente y los datos de mercado mejoran el entendimiento del comportamiento del cliente, lo que ayuda a los proveedores a elaborar mensajes que a la gente le gusta ver. Las pruebas estadísticas de hipótesis identifican qué mensajes tienen más eco con grupos específicos, permitiendo mensajes personalizados para públicos objetivo.

## Caso de Uso: Entendiendo la retroalimentación del Cliente y Analítica de Texto

Los proveedores de Servicios Financieros Digitales también pueden extraer conclusiones útiles sobre las preferencias y actitudes de los clientes a través de nuevas técnicas basadas en algoritmos llamadas minería de texto o analítica de texto. Hoy en día, muchas compañías pueden acceder a información sobre los gustos de los clientes a través de las redes sociales, correos electrónicos, sitios web y de las transcripciones de las conversaciones en los call centers. Cabe destacar que estos métodos se han aplicado en contextos de países desarrollados en Europa y América del Norte. Sin embargo, los proveedores de SFD en los mercados emergentes también pueden querer analizar estos datos para ayudar a crecer los negocios. La analítica de texto también puede realizarse manualmente. Con los avances tecnológicos, es probable que estos métodos se vuelvan más baratos y más adaptables a los contextos y lenguajes de los países en desarrollo.

La aplicación más común para la analítica de texto es a través de dos métodos:

**1. Métodos de Resumen de Texto:** Estos métodos proporcionan un resumen de toda la información clave de un texto. Este resumen se puede crear utilizando sólo el texto original (enfoque extractivo) o utilizando texto que no está presente en el texto (enfoque abstractivo).

**2. Análisis de Sentimientos:** El análisis de sentimientos o 'minería de opinión' es una herramienta basada en algoritmos utilizada para evaluar el lenguaje, tanto hablado como escrito, para determinar si la expresión de opinión es positiva, negativa o neutra y en qué medida. A través de este análisis, los proveedores de SFD entienden cómo los clientes se sienten acerca de sus productos, cómo se relacionan con la marca y cómo estas actitudes están cambiando con el tiempo. Los picos o valles en el análisis de sentimiento son de especial interés.

Actualmente, las evaluaciones de la analítica de texto pueden aplicarse en tres áreas:

### Mejora de Productos y Servicios

Los proveedores de SFD podrían hacer mejoras rápidas a los productos y servicios si pudieran escuchar directamente de los clientes. Las redes sociales, correos electrónicos y otros mecanismos de retroalimentación directa son una excelente manera de escuchar de inmediato y directamente a los clientes. La investigación de mercado puede ser una fuente limitada de comentarios de los clientes a este respecto.

### Marketing Boca a Boca

El marketing boca a boca sigue siendo la forma más fiable de publicidad para muchos clientes. Para los productos y los proveedores de SFD que tienen bases grandes de clientes existentes, no es difícil motivar a los clientes satisfechos para impulsar el marketing boca a boca. Sin embargo, para nuevos productos,

como Servicios Financieros Digitales, los proveedores deben encontrar un método para catalizar los niveles de educación entre las bases de clientes potenciales, especialmente entre los clientes que construyen el entusiasmo y el impulso para el producto dentro de la base de clientes objetivo. Normalmente, los clientes están más motivados para hablar acerca de uno o dos casos de uso específicos; rara vez se extenderá un mensaje genérico sobre la marca. Las fuentes de redes sociales y otra información en la web se pueden utilizar para identificar influenciadores por su nivel de conexión, nivel y naturaleza de la interacción y alcance potencial. Este tipo de análisis depende de datos de redes sociales no estructurados, datos de sitios evaluadores de contenido y datos de blogs.

### Impacto de Marketing y Monitoreo de Retroalimentación

La minería de opinión permite a los proveedores de SFD entender el proceso del pensamiento de un gran número de clientes. A través del análisis de sentimiento, es posible rastrear lo que dicen los clientes sobre nuevos productos, publicidad, servicios, marcas y otros aspectos del marketing. Este análisis también puede usarse para entender cómo el mercado percibe los productos y servicios de la competencia. Estos datos de las redes sociales, blogs, sitios web de revisión y otros sitios web en el ámbito social también son no estructurados.

### 1.2.2 Analítica y Aplicaciones: Gestión de Operaciones y Desempeño

El equipo de operaciones es responsable de operar la 'sala de máquinas' que es clave para el negocio de los Servicios Financieros Digitales porque realiza una variedad de tareas, incluyendo: recogida de datos, almacenamiento de datos y asegurar su conectividad fluida entre varios sistemas y aplicaciones para todo el entorno TI del proveedor de SFD; monitorear constantemente la calidad de los datos; inducción y gestión del desempeño de los corresponsales; asegurar que la tecnología esté funcionando según lo diseñado; prestar apoyo al cliente; suministrar la información y las herramientas necesarias para el equipo comercial, incluyendo la medición del desempeño, la supervisión del riesgo y la presentación de informes reglamentarios; resolver los problemas; monitoreo eficiente de los indicadores, excepciones y anomalías; gestión del riesgo; y asegurar que el negocio cumpla con las obligaciones regulatorias. Esto no se puede hacer eficientemente sin acceso a datos precisos, presentados en una forma que sea relevante, fácilmente digerible y oportuna.

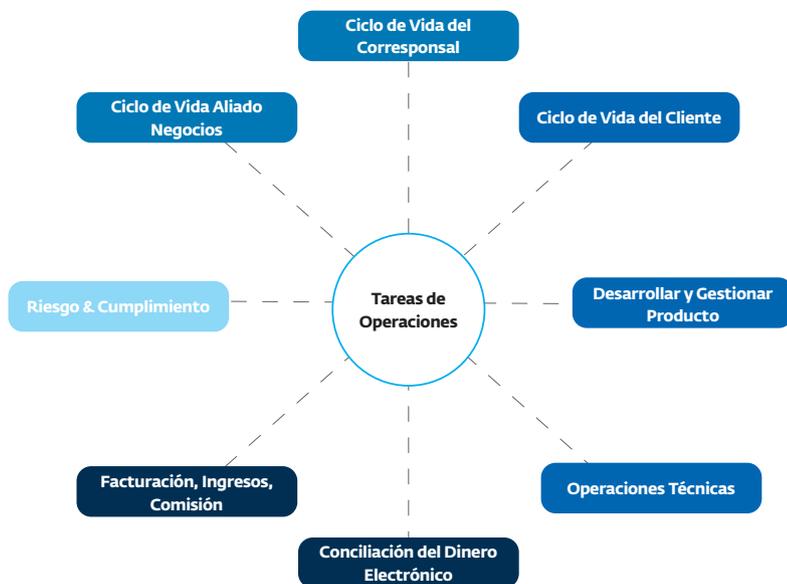


Figura 14: Tareas de Operaciones

Este equipo tiene un papel importante en la estructura organizacional, siendo independiente de otras funciones básicas y también estando integrado en las principales actividades empresariales. La naturaleza de las responsabilidades del equipo requiere habilidades técnicas, así como el conocimiento del negocio. Esta combinación permite interpretaciones significativas de los datos que eventualmente pueden ayudar en los procesos de toma de decisiones de las principales partes interesadas del negocio.

En esta sección se describe el papel que pueden desempeñar los datos para optimizar las operaciones diarias de un proveedor de SFD típico. Comienza por describir cómo los datos pueden convertirse en información útil, dando ejemplos de la vida real de analítica de datos en acción. Esto incluye algunos consejos sobre las mejores prácticas en el uso de datos en Servicios Financieros Digitales. A medida que el uso del panel de control de datos se vuelve cada vez más común, proporciona información sobre la creación y el contenido de los paneles de control.

#### Caso de Uso: Visualización del Desempeño con Paneles de Control

A menudo se dice que una imagen vale más que mil palabras. Así, encontrar una forma gráfica de representar datos es una manera poderosa de comunicar información y tendencias rápidamente, lo cual es crítico para el monitoreo constante del desempeño del negocio y crítico para identificar los riesgos antes que se materialicen. Los paneles de control bien estructurados, adaptados a varios grupos de usuarios, deben reflejar la demanda de las unidades de negocio y ayudarles a tomar decisiones más informadas.

Convertir los datos en gráficos y otras formas de visualización facilitan la comunicación de la información revelada y también ayuda a detectar tendencias y anomalías en los datos. Muchas personas en la organización no tienen el tiempo o los recursos para analizar los datos ellos mismos; simplemente quieren las respuestas a las preguntas que les ayudarán a hacer su trabajo más eficazmente.

Un panel de control da una foto instantánea de los KPIs relevantes para un departamento o para el negocio en general. Si raramente es necesario tomar medidas basadas en los datos reportados, las métricas de los paneles de control probablemente son incorrectas. Con el fin de diseñar paneles de control robustos, es importante incorporar retroalimentación de los usuarios finales, con el fin de satisfacer sus necesidades específicas. Sin esta retroalimentación, los paneles de control podrían quedar obsoletos y todos los esfuerzos para desarrollarlos se desperdiciarían. Así, el desarrollo de los paneles de control es un proyecto compartido entre los equipos de operaciones y de negocios, que podría pasar por varias iteraciones para recorrer el ciclo de retroalimentación de las diversas partes interesadas.

Algunos paneles de control deben ser en tiempo real. Por ejemplo, un equipo de operaciones técnicas tiene que actuar respecto de las alertas elevadas en tiempo real: los gerentes de atención al cliente evalúan en forma activa los volúmenes de llamadas para asignar trabajo de equipo y gestionar incidentes, los equipos de gestión de riesgo son informados constantemente sobre los impagos, y los equipos de ventas pueden tomar acciones tempranas

respecto a las cuentas de baja actividad para reactivar al cliente y no dejar que la cuenta se vuelva inactiva. Algunos de estos paneles de control permitirían a los usuarios finales manipular los datos para visualizar varios cortes y segmentos de datos. A menudo, estos tipos de paneles de control se presentan en tiempo real en una pantalla grande en el espacio de trabajo del equipo para que todos lo puedan ver. Para el personal de campo, donde el acceso a Internet puede ser de calidad variable, los paneles de control en línea se pueden descargar y almacenar en caché localmente para su uso en el campo.

Otros paneles de control administrativos proporcionan conclusiones a través del análisis de datos del día, la semana, el mes o el año anterior y, por lo tanto, pueden entregarse de múltiples formas, incluidos informes, presentaciones o por un portal en línea. En consecuencia, cada departamento y equipo de proyecto necesita paneles de control personalizados para las metas e iniciativas del departamento. Por lo general, como mínimo, las soluciones de SFD deben tener múltiples paneles de control de operaciones que abarcan las siguientes áreas, cada una de las cuales proporciona acceso basado en roles para audiencias específicas:

- **Riesgo:** Fuga de ingresos; Préstamos en Mora (NPLs); conclusiones sobre Anti Lavado de Activos (ALA); suficiencia de capital; detección de fraude.
- **Finanzas:** Percepciones de ganancias y pérdidas; supervisión de dinero electrónico.
- **Marketing:** Información de los clientes y tendencias para diversas ofertas.

- **Ventas:** Desempeño del corresponsal; desempeño del comerciante y del facturador; desempeño del equipo de ventas.
- **Operaciones:** Gestión de liquidez de corresponsales.
- **Atención al Cliente:** Estadísticas e información del call center.
- **Operaciones Técnicas:** Información sobre operaciones técnicas.

Las herramientas de gestión de datos comercialmente disponibles han avanzado enormemente en los últimos años. Es probable que los paneles de control estándar estén disponibles como parte del paquete de productos de los proveedores de tecnología. Con el fin de obtener las conclusiones más profundas necesarias y hacerlo de una manera reproducible, hay dos enfoques estándar:

1. **Volver al Proveedor:** A menudo hay presupuesto disponible para que los proveedores hagan cambios en los paneles de control, pero las solicitudes de varios departamentos y los clientes de varios proveedores que compiten por la atención pueden conducir a problemas de capacidad y retrasos.
2. **Utilizar Excel para Manipular Informes sin Procesar Descargados de 'Cubos de Datos' del sistema:** Cuando se hace una pregunta al equipo de soporte de decisiones de negocios, éste creará un del panel de control personalizado y entregará un informe o una presentación de PowerPoint con un intento de respuesta. Esta es otra forma ad hoc de creación de paneles de control.

## 1.2\_APLICACIONES DE DATOS

La última generación de herramientas de gestión de datos permite la libertad de investigar áreas de interés sin necesidad de experiencia en manipulación de datos. Sin embargo, las bases de datos subyacentes deben diseñarse y optimizarse para desplegar y utilizar con éxito este tipo de herramientas. Cualquiera que sea el proceso de gestión de datos o el sistema que se esté utilizando, éstos son los puntos a tener en cuenta al crear un panel:

**1. Piense en Contestar “¿Y qué?”:** Los resultados deben ser accionables, no sólo 'bueno saberlos'. Muchos paneles de control sólo muestran el estado actual del negocio y no dan contexto de resultados anteriores o tendencias basadas en el tiempo.

**2. Decidir qué Pregunta se está Respondiendo Antes de Empezar:** A menudo, los informes son el vertedero para todos los datos disponibles, sean útiles o no. Este tipo de informes no contiene las métricas motivacionales y las medidas que aumentan el desempeño.

**3. Diseñar el Informe para Contar una Historia:** Una vez que los datos correctos son medidos y recopilados, el informe debe contener información llamativa para llevar al lector a los puntos más importantes. Que sea visual, interesante y útil.

### Informes de Operaciones Estándar

Con el fin de mejorar sus negocios, los proveedores de Servicios Financieros Digitales están tratando de encontrar la respuesta a preguntas tales como:

- ¿Cuál fue el volumen y el valor de la transacción?
- ¿Cuántos clientes y corresponsales estaban activos?
- ¿Qué ingresos hicimos?
- ¿Cómo se compara esto con el mes pasado y con el presupuesto?
- ¿Hay indicadores de riesgo fuera de rangos aceptables?
- ¿Existen transacciones inusuales recurrentes, picos de actividad o cualquier anomalía que señale actividad inusual?

El punto de partida es centrarse en los KPIs, o métricas con objetivos cuantificables que la estrategia operacional está tratando de conseguir y con relación a qué tipo de desempeño se está evaluando. Los KPI generales del negocio deben relacionarse directamente con los objetivos estratégicos de la organización y, como resultado, determinar los KPIs específicos de cada departamento. Los datos más útiles son aquellos que pueden convertirse en la información necesaria para tomar decisiones. Antes

de crear un informe, se debe identificar exactamente lo que quiere saber y confirmar que se tomarán medidas como resultado de obtener los datos.

Los KPIs departamentales bien estructurados proporcionan a los equipos de operaciones una información a partir de la cuál pueden medir el desempeño frente a los objetivos. Ayudan a los equipos a comprender lo que está sucediendo sobre el terreno y dónde existe el potencial de mejora.

Los informes estándar de KPI sobre los principales motores del negocio suelen estar segmentados por área operativa. Los KPI principales de cada área operacional respectiva se encuentran en la Tabla 3 a continuación.

Departamento	Temas de Prioridad para los KPIs
<b>Finanzas y Tesorería</b>	Ingresos, ingresos y gastos por intereses, comisiones y honorarios, cantidad retenida en depósito, volumen y valor de transacción, volumen de clientes y corresponsales (activos), costos indirectos y emisión de dinero electrónico para no bancarios, reconciliación de estados de cuenta bancarios.
<b>Ciclo de Vida de los Aliados Comerciales</b> (comerciantes, facturadores, conmutadores, corresponsales, bancos aliados, otros PSP)	Vinculación, niveles de actividad, resolución de conflictos, administración de desempeño, reconciliación y liquidación.
<b>Gestión del Ciclo de Vida del Cliente</b>	Administración de KYC, niveles de actividad, comportamiento transaccional, resolución de conflictos (servicio al cliente) y gestión de cuentas.
<b>Operaciones Técnicas</b>	Monitoreo del desempeño del producto, monitoreo de los niveles de servicio de los socios, gestión del cambio, integración de socios, resolución de fallos, administración de incidentes y gestión de acceso de usuarios.
<b>Riesgo Crediticio</b>	Estructura de riesgo del portafolio, préstamos en mora, bajas y pérdidas por riesgo, provisión de créditos.
<b>Riesgo Operacional y Cumplimiento</b>	Administración de riesgo operacional, monitoreo y seguimiento de actividades sospechosas, cumplimiento regulatorio, debida diligencia e investigaciones ad hoc.
<b>Ciclo de Vida de la Red de Corresponsales (específicamente de SFD)</b>	Vinculación, niveles de actividad, administración de conciliación del fondo de dinero electrónico y efectivo, resolución de conflictos, administración de desempeño, conciliación y liquidación, y auditoría.
<b>Otro</b>	Dependiendo de la naturaleza de los Servicios Financieros Digitales, pueden requerirse otros informes, por ejemplo, las organizaciones que ofrecen crédito realizarán calificaciones crediticias, recuperación de cartera y tareas relacionadas.

*Tabla 3: Principales KPIs por Área Operacional*

Dependiendo de la estrategia de negocio y los objetivos departamentales, una selección de los datos anteriores se presentan como los KPI del negocio y sus departamentos. Éstos pueden, idealmente, presentarse como paneles de control, o como un conjunto de informes. Es importante que cada departamento segregue sus datos en KPIs y datos de

apoyo, ya que siempre existe la tentación de incluir datos periféricos, que no son estrictamente necesarios para entender la salud de su departamento, dentro de los informes de administración. Esto puede distraer o dar lugar a una priorización inadecuada. Los datos de apoyo son vitales para ayudar a entender los factores que impulsan los KPIs y determinar cómo

pueden ser mejorados, pero generalmente no tienen que ser reportados a un público más amplio a menos que haya un punto específico que se deba hacer. Un buen ejemplo de esto es el enfoque que se ilustra a continuación con el uso de los paneles de control de MicroCred.

# CASO 5

## *MicroCred Utiliza Paneles de Control para Mejores Sistemas de Gestión*

### **Visualizaciones de Datos y Paneles de Control para el Desempeño Diario y el Monitoreo del Fraude**

*MicroCred es una red de microfinanzas centrada en la inclusión financiera en África y Asia. En Senegal, opera un negocio de microfinanzas en crecimiento que ofrece servicios financieros a personas que carecen de acceso a bancos u otros servicios financieros. El alcance se ha extendido a todo el país mediante la creación de una red de más de 500 corresponsales de Servicios Financieros Digitales. Los dispositivos POS del corresponsal pueden realizar transacciones de venta de mostrador (Servicios Financieros en Mostrador) para pagos de facturas y remesas, y también facilitar el depósito y los retiros a cuentas de MicroCred. La confirmación de la transacción es proporcionada mediante un recibo vía SMS. A finales de 2016, casi un tercio de los clientes registraron su cuenta para usar el canal de corresponsales y más de un cuarto usaban activamente los puntos de venta de corresponsales para realizar transacciones. Esto generó importantes datos operativos y de desempeño del canal.*



**Figura 15:** Ejemplo de Datos del Panel de Control de MicroCred

MicroCred fue uno de los primeros en adoptar los sistemas de gestión de datos de próxima generación, adquiriendo e implementando BIME, una herramienta de visualización para ayudar a optimizar las operaciones. Este sistema permitió a MicroCred desarrollar paneles de control interactivos, diseñados para responder a preguntas operativas específicas.

MicroCred utiliza con mayor frecuencia dos paneles de control:

#### **Paneles de Control de Operaciones Diarias**

Esto da una perspectiva diaria de los portafolios de ahorros y préstamos, destacando cualquier problema. Presenta datos durante un período de tres meses, pero puede ajustarse de acuerdo con las necesidades del usuario. Este panel de control utiliza alertas automatizadas para advertir al equipo de operaciones de posibles problemas. Los informes, personalizados para los equipos operativos, incluyen medidas tales como:

- Seguimiento de KPIs, incluyendo volúmenes de transacciones, comisiones y cargos

- Actividad del corresponsal, con alertas para mostrar corresponsales de bajo desempeño que no están haciendo transacciones
- Actividad sospechosa y posibles alertas de fraude, como actividad inusual del corresponsal o del cliente
- Monitoreo del proceso de inscripción a SFD, con énfasis en vinculaciones no exitosas
- Distribución geográfica de las transacciones

#### **Panel de Control Estratégico Mensual**

Esto da una visión más estratégica a más largo plazo y es utilizada principalmente por el equipo administrativo para visualizar medidas de negocio críticas más complejas. Fue desarrollado para considerar el comportamiento durante el ciclo de vida del cliente, incluyendo cómo el uso del servicio evoluciona a medida que los clientes se familiarizan con la tecnología y los servicios ofrecidos. También es posible realizar fácilmente análisis ad hoc para dar seguimiento a cualquier pregunta planteada por los datos presentados en el panel de control. Se enfoca en:

- Uso de sucursales de MicroCred vs. corresponsales
- Adopción y uso por parte del cliente de SFD
- Despliegue del canal de Servicios Financieros Digitales
- Evolución de los KPIs fundamentales frente a los objetivos a largo plazo

Con herramientas de visualización como BIME, es fácil crear gráficos para ilustrar los datos operativos, lo que facilita la detección de tendencias y anomalías y la comunicación eficaz. La implementación del sistema de gestión de datos también presentó algunos desafíos, tanto técnicos como culturales. MicroCred recomienda que se adopte un enfoque paso a paso, comenzando con algunos paneles de control básicos y construyendo con el tiempo paneles de control más sofisticados.



Las herramientas de visualización y los paneles de control interactivos se pueden integrar en los sistemas de gestión de datos y proporcionar informes dinámicos adaptados que sirven para el monitoreo de operaciones, gestión y desempeño estratégico.

## 1.2\_APLICACIONES DE DATOS

### Datos Utilizados en Paneles de Control

Hay dos niveles principales de registro de datos necesarios para desarrollar los paneles de control: transacción y de cliente. Atienden a diferentes objetivos, pero ambos son importantes.

### Datos de Transacción

Los datos de transacción se caracterizan por su alta frecuencia y heterogeneidad. Sin embargo, los proveedores de Servicios Financieros Digitales deben tratar de estandarizar la tipología de transacciones con el fin de rastrear la rentabilidad del producto, monitorear y analizar el comportamiento de los clientes (y corresponsales) y elevar las alertas tempranas de bajo desempeño o baja actividad de cuentas. Los tipos de transacciones deben ser claramente diferenciados y deben ser fácilmente identificables en la base de datos, incluso cuando las transacciones parecen técnicamente similares. Por ejemplo, una causa común de confusión ocurre cuando hay múltiples maneras de obtener fondos en una cuenta de cliente, como P2P entrante, pagos masivos o crédito, pero todos los datos se combinan y simplemente se reportan como “depósitos”. Estos tres tipos de transacciones deben tratarse por separado debido a su impacto muy diferente en los ingresos - uno es un costo directo, el otro es una fuente de

ingresos y el siguiente es potencialmente neutral en términos de costos - y debido a sus implicaciones para la estrategia de marketing.

### Datos de los Clientes

Tener un identificador de cliente único es crucial, especialmente cuando el panel de control está obteniendo datos de múltiples aplicaciones. A través de la integración de datos, los proveedores pueden controlar la integridad de los datos y así garantizar la calidad del registro de datos, lo cual es necesario para el seguimiento de la concentración del portafolio, el cálculo de la penetración del producto, la venta cruzada y la cobertura del personal de ventas y el análisis de otras métricas importantes. Generalmente hay dos grupos grandes de datos que tienen que ser grabados a nivel de cliente: demográfico y financiero. Las listas completas de métricas de datos se pueden encontrar en el Capítulo 1.2. La combinación de datos a nivel de transacciones y de clientes puede proporcionar información útil sobre el comportamiento de ciertos segmentos de clientes y puede conducir a una gestión óptima del desempeño.

### Caso de Uso: Gestión del Desempeño del Corresponsal

La gestión de corresponsales es probablemente el reto más importante para proporcionar Servicios Financieros

Digitales exitosos, ya que requiere una intervención práctica regular de un equipo de ventas de campo, así como un soporte de operaciones de back-office. Puede ser problemático difundir información, porque el equipo y los corresponsales están geográficamente dispersos con distintos niveles de conectividad y suelen estar equipados con tecnología básica. Sin embargo, sus necesidades de datos son muchas. Los administradores de relaciones, los agregadores y los corresponsales con múltiples puntos de venta en múltiples localidades necesitan información sobre el desempeño y la administración de la conciliación. Los trabajadores de la fuerza de ventas de campo que rara vez regresan a la oficina necesitan acceder a la información remotamente. El corresponsal necesita información sobre su propio desempeño en términos de transacción y conteo de clientes, volumen de negocios, eficiencia de ventas (conversión) y rentabilidad. Potencialmente, será útil la información sobre los servicios de reposición de efectivo disponibles, particularmente en mercados donde los corresponsales pueden proporcionar servicios de conciliación de dinero electrónico y servicios de administración de efectivo entre sí. En los mercados con aliados independientes de gestión de efectivo, los corresponsales también deben disponer de datos sobre los niveles del fondo de dinero electrónico y efectivo.

La gestión de desempeño del corresponsal necesita datos granulares, vinculados directamente a los equipos responsables de la gestión de los puntos de venta. Los datos de desempeño del corresponsal tienen que segmentarse fácilmente de la misma manera que el equipo de ventas está estructurado; cada sección y cada individuo pueden ver su propio desempeño. Esta es la base para establecer objetivos de desempeño que pueden ser evaluados y recompensados con precisión. En el siguiente ejemplo, tanto los equipos como las personas responsables de cada nivel de la jerarquía de corresponsales, desde el director de ventas hasta los representantes de ventas del distrito, necesitan datos precisos y oportunos relacionados directamente con sus responsabilidades. La información más útil que se puede dar al equipo de ventas se refiere a los corresponsales por los que son responsables.

### **Cobertura de Brechas de Corresponsal**

No hay respuestas definitivas para el número óptimo de corresponsales necesarios para que cada cliente tenga un acceso razonablemente fácil a un corresponsal y para que cada corresponsal tenga suficientes clientes para generar un ingreso aceptable. La investigación indica

que entre 200 y 600 clientes activos por cada corresponsal activo son óptimos para los proveedores de Servicios Financieros Digitales, dependiendo de las condiciones del mercado. Una tarea clave de ventas es monitorear los datos del corresponsal y del cliente, controlando el crecimiento y la ubicación de los puntos de venta de los corresponsales para asegurar que estén en línea con la actividad del cliente.

### **Cómo Identificar a los Corresponsales más Fuertes**

Los corresponsales de calidad deben ser recompensados por sus esfuerzos. Los incentivos, incluidas las actividades de marketing y extra cupo, o bonos relacionados con el desempeño, pueden basarse en estos datos. Tener objetivos personalizados de corresponsales basados en las condiciones del mercado local, y tener una manera de mostrar claramente al corresponsal cómo se desempeña comparado con sus propios objetivos y sus pares, puede ser muy poderoso. Los objetivos incluyen la liquidez y la actividad del cliente. Una característica clave de un buen corresponsal es que rara vez se les acaba el dinero electrónico o efectivo en caja. Los objetivos del agregador de corresponsales debe basarse en la actividad de gestión de liquidez, para cuyo apoyo se le contrató, así como en el desempeño de su equipo de corresponsales.

### **Cómo Identificar los Corresponsales más Débiles**

En la mayoría de los mercados, alrededor del 80 por ciento de los corresponsales están activos. Esto significa que los clientes que deseen realizar transacciones con el otro 20 por ciento de corresponsales probablemente no podrán hacerlo porque no hay fondos suficientes o el corresponsal está ausente. Los corresponsales de bajo desempeño tienen que ser llevados a un nivel aceptable, o si esto resulta imposible, ser retirados del servicio. Debido a que la falta de liquidez del dinero electrónico tiene una fuerte correlación con el incumplimiento, una métrica clave utilizada a menudo para el análisis del desempeño del corresponsal es el número de días en que no disponen de stock por mes (es decir, niveles de conciliación por debajo de un valor de umbral).

Este tipo de analítica de datos de corresponsales es muy eficaz, pero bastante detallado y a menudo se realiza manualmente, lo que puede ser lento e intensivo en mano de obra. Proporcionar al equipo de ventas herramientas de gestión de datos automatizadas que pueden utilizar en el campo, así como métricas de desempeño personalizadas, puede ser muy eficaz. El estudio de caso de Zoono a continuación demuestra bien estos puntos.

# CASO 6

## Zoona Zambia - Optimización de la Gestión de Desempeño del Corresponsal

### Cultura de Datos: Un Enfoque Integrado Basado en Datos para Productos, Servicios y Generación de Informes

Zoona es el proveedor líder de Servicios Financieros Digitales en Zambia, que ofrece transacciones de servicios financieros en mostrador a través de una red de corresponsales Zoona dedicados. Los servicios de corresponsales incluyen: registro de clientes, envío y recepción de pagos de remesas, suministro de depósitos o retiros en efectivo para cuentas y desembolso de pagos masivos de terceros, tales como salarios y pagos de G2P. Zoona tiene una cultura empresarial basada en datos y cuenta con un equipo centralizado de analistas de datos para perfeccionar constantemente la sofisticación y la eficacia de sus servicios y operaciones.

#### Ubicación del Corresponsal

Zoona ha desarrollado un simulador interno para determinar la ubicación óptima de los quioscos de corresponsales. El enfoque utiliza *simulaciones*<sup>20</sup> Monte Carlo para probar millones de escenarios posibles de localización

de corresponsales para identificar qué configuraciones maximizan el crecimiento del negocio. Factores como el número de clientes atendidos por día por corresponsales existentes y tiempo de espera se utilizan para determinar la demanda local y el potencial de crecimiento hasta que se alcance la saturación. Para asegurar la fiabilidad, los escenarios modelados son referenciados con los inputs del equipo de ventas de campo, que tiene conocimiento local de la zona y los puntos de venta bajo la mayor presión. En ubicaciones clave, el equipo también usa Google Maps y camina físicamente por las calles, observando el nivel de actividad y dónde pueden estar los puntos calientes potenciales. Por ejemplo, miles de personas pueden llegar a una estación de autobuses, luego se dispersan en varias direcciones; Zoona traza las rutas más populares, creando corredores donde es probable encontrar clientes potenciales. Zoona también mapea

la ubicación de los competidores en estas rutas.

#### Ciclo de vida del Corresponsal

Un corresponsal relativamente nuevo en una carretera principal puede no ser tan productivo como un corresponsal maduro en un mercado activo, debido a la ubicación y que el corresponsal maduro haya desarrollado una base de clientes leales. Sin embargo, un servicio robusto de SFD necesita corresponsales en ambas ubicaciones - y los objetivos establecidos para cada corresponsal deben ser realistas y alcanzables. Zoona analiza los datos de los corresponsales para proyectar las expectativas de desempeño futuro de los segmentos de corresponsales, como urbano y rural, produciendo curvas de “desempeño en el tiempo” para cada corresponsal, hasta el nivel del suburbio. Estos son compatibles con los KPIs de gestión de corresponsales.

<sup>20</sup> Las simulaciones Monte Carlo toman muestras de una distribución de probabilidad para cada variable para producir miles de resultados posibles. Los resultados se analizan para obtener probabilidades de los diferentes resultados que se producen.

## Gestión de Liquidez

Los corresponsales requieren una fuente de liquidez cómoda para atender las transacciones, por lo que se incluye en los escenarios de ubicación la proximidad a los bancos cercanos o cajeros automáticos (ATM). La dificultad para reponer el fondo también puede deberse a una sobreconcentración de corresponsales, que colectivamente cargan las fuentes del fondo cercanas y socavan el valor de la red de corresponsales locales. Las simulaciones de Zoona miran ambos escenarios como parte de la optimización. Además, entendiendo que la conciliación del corresponsal es un factor clave en el desempeño, Zoona está pilotando una solución innovadora para recoger los saldos de efectivo y fondos de dinero electrónico de un corresponsal para ayudar a los corresponsales a administrar su conciliación de manera más eficaz. Esto proporciona a los corresponsales acceso a herramientas de administración de desempeño, que se desarrollan utilizando la herramienta de visualización de gestión de datos Qlikview. Este proporciona a Zoona datos que los corresponsales podrían no querer informar.



Las analíticas pueden soportar muchos aspectos de las operaciones y desarrollo de producto: optimización de la ubicación de corresponsales, administración de desempeño y herramientas que crean incentivos para el informe voluntario de datos. Una cultura empresarial basada en datos impulsa la integración.

## 1.2\_APLICACIONES DE DATOS

### Administración de Back Office del Corresponsal

El equipo de back office del corresponsal es responsable de todas las tareas necesarias para configurar nuevos corresponsales y administrar sus interacciones con SFD en curso. A menudo, esto también incluye obtener los datos necesarios por el equipo de ventas (arriba). Para ser eficaces, necesitan una gran cantidad de datos, incluidos los informes estándar y el acceso a los datos para ejecutar informes ad hoc centrados en consultas específicas. Además de proporcionar los datos del equipo de ventas, también deben medir el tiempo que demoran sus procesos de negocios, para asegurar que su equipo tenga capacidad de cumplir contra los niveles de servicio internos. Esto se logra midiendo los problemas planteados por tipo y volumen, y midiendo el tiempo de resolución de conflictos, a menudo a través de un sistema de tickets.

### Back Office para Aliados de Negocio

Para efectos de la administración del back office, se pueden combinar varios tipos de aliados de negocios no corresponsales. Estos incluyen facturadores y otros PSP, comerciantes, organizaciones que usan el SFD para propósitos de administración de negocios, incluyendo pagos de nómina y otros pagos masivos, y otras Instituciones Financieras, incluyendo bancos y proveedores de SFD. El equipo de back office de administración de aliados de negocio es responsable de tareas similares a la administración de corresponsales,

pero con requerimientos regulatorios diferentes (y no necesita administración de conciliación). En consecuencia, las métricas claves que necesitan son similares a las de los corresponsales, pero con algunos procesos de negocio y objetivos diferentes.

### Optimización de la Eficiencia del Corresponsal

Los datos pueden ser utilizados con mayor eficiencia por los equipos de gestión de corresponsales cuando tienen acceso móvil y en línea a estos datos. Algunas de estas tareas incluyen:

- Planificación de la carga de trabajo
- Registro de entradas y salidas de los puntos de venta del corresponsal en las visitas de campo
- Actualizar o verificar la ubicación y otra información demográfica del punto de venta
- Mostrar estadísticas de desempeño personalizadas al corresponsal directamente al llegar
- Mostrar comisión percibida tanto hasta la fecha como en el mes
- Mostrar los ingresos percibidos de los clientes que el corresponsal está atendiendo
- Permitirles agregar fotos a la base de datos
- Rellenar directamente las medidas básicas de la encuesta de corresponsales de Garantía de Calidad (QA)
- Notificar que la información de KYC está en tránsito

- Establecer nuevos objetivos de desempeño e incentivos
- Enviar solicitudes de servicio del corresponsal y consultas directamente al equipo de operaciones
- Capturar potenciales interesados para nuevas locaciones de puntos de venta de corresponsales

El acceso a este tipo de datos puede resultar en corresponsales más motivados y exitosos, así como mejorar el desempeño general del negocio de SFD. Se pueden abordar preguntas importantes como: ¿Cuánto fondo de dinero electrónico necesitan los corresponsales? Con el fin de administrar efectivo y fondos digitales, es útil comprender los momentos más ocupados del día, la semana y el mes, y proporcionar orientación sobre los requerimientos de fondos esperados. También es útil tener indicadores en el sistema de tal manera que, si el fondo de un corresponsal cae por debajo de un nivel mínimo, la persona responsable de la gestión del fondo del corresponsal reciba una alerta automatizada. En operaciones más sofisticadas, los algoritmos se pueden utilizar para predecir proactivamente cuántos fondos necesitará cada corresponsal cada día y para aconsejarles acerca del balance óptimo inicial antes de comenzar a comercializar o después del cierre de transacciones. Esto también se puede hacer para la cantidad de efectivo que el corresponsal deba tener a mano para atender retiros.

# CASO 7

## FINCA DRC – Cómo es Un Corresponsal Exitoso y Cómo Poner los Resultados en Marcha

### Recopilación de Datos: Optimización del Proceso para Tener Mejor Información y Una Implementación Exitosa

Con una tasa de penetración bancaria apenas por debajo del 11 por ciento, la RDC tiene una de las tasas más bajas de acceso financiero en África. En 2011, la institución microfinanciera FINCA DRC introdujo su red de corresponsales, empleando propietarios de pequeños negocios para ofrecer servicios bancarios. La red de corresponsales creció rápidamente, y en el momento en que comenzó la recopilación de datos de corresponsales en 2014, recibía más del 60 por ciento de las transacciones totales de FINCA DRC. Para el año 2017, las transacciones de corresponsales habían crecido hasta el 76 por ciento del total de transacciones. Sin embargo, el crecimiento se concentró principalmente en la capital del país, Kinshasa y en uno de los centros comerciales del país, Katanga. FINCA DRC buscó expandir la red a áreas rurales y así construyeron un modelo predictivo para identificar los criterios que definen a un corresponsal exitoso. Los resultados se incorporaron a las encuestas de reclutamiento de corresponsales, ayudando a FINCA DRC a seleccionar buenos corresponsales para las áreas de

expansión. Más aún, la disponibilidad de una red de corresponsales exitosa que los clientes pueden utilizar para pagar convenientemente sus préstamos ayuda a FINCA DRC a reducir su riesgo de cartera.

El modelo predictivo definió “corresponsales exitosos” tanto en términos de números de transacciones como en volúmenes más altos. Los datos para el modelo lineal generalizado (MLG) provienen de tres fuentes principales:

- **Formularios de Solicitud de Corresponsales:** Estos proporcionan información sobre los datos del negocio y sociodemográficos del propietario.
- **Formularios de Monitoreo de Corresponsales:** Los oficiales de FINCA DRC monitorean regularmente a los corresponsales, recopilando información sobre el efectivo y fondo electrónico del corresponsal, la condición del almacén, los datos de sentimiento sobre la interacción del cliente con el corresponsal y el branding del producto FINCA DRC. Esto se compila en una calificación de monitoreo.

- **Datos Transaccionales de Corresponsales:** Estos datos incluyen información sobre el volumen y número de transacciones de depósito, retiros y transferencia realizadas por corresponsales individuales.

La disponibilidad de datos y la calidad de los mismo fueron los principales desafíos en el desarrollo del modelo de desempeño de los corresponsales. Los datos digitalizados son necesarios para las fuentes que normalmente sólo se recopilan en papel, como la aplicación de los corresponsales y formularios de monitoreo. Los datos que faltan deben minimizarse, tanto para que los conjuntos de datos sean más robustos como para permitir la fusión de conjuntos de datos mediante campos de metadatos coincidentes. Esto requiere la estandarización de datos recopilados por diferentes personas, que pueden estar utilizando diferentes métodos de recopilación. La falta de datos consistentes puede conducir a una reducción significativa de la muestra, socavando la precisión y el desempeño de predicción del modelo.

## 1.2\_APLICACIONES DE DATOS

Los corresponsales exitosos en la República Democrática del Congo se identifican por los siguientes criterios estadísticamente significativos: ubicación geográfica, sector del negocio principal de un corresponsal, género del corresponsal y si reinvierten los beneficios. Los corresponsales de propiedad de mujeres son, por ejemplo, un 16 por ciento más rentables con su negocio de corresponsales que los negocios de sus contrapartes masculinas; el valor de su inventario de negocios es un 42 por ciento más alto. También se encontró que devuelven más dinero al inventario del negocio, en lugar de guardarlo en una cuenta bancaria que da poco interés. Esto dio como resultado un promedio total de transacciones un 5 por ciento más alto mensualmente.

Estos resultados se implementaron para mejorar y agilizar el proceso de selección de corresponsales, lo que en última instancia ayudó a expandir la red hacia las zonas rurales mediante la incorporación de factores en las encuestas de corresponsales y la estrategia de despliegue. Para 2016, la red de corresponsales había crecido hasta albergar el 70 por ciento del total de transacciones. El modelo identificó la ubicación

como un criterio clave, revelando otra oportunidad de investigación. Como un estudio de seguimiento, FINCA DRC e IFC utilizarán una metodología RCT para identificar la ubicación óptima del corresponsal.



Comparar los datos de los perfiles de los corresponsales con las métricas del corresponsal puede resaltar las características clave que conducen a un mejor desempeño del corresponsal. La integración de estos aprendizajes con los procesos de gestión y direccionamiento de corresponsales garantiza el aprovechamiento completo de los datos para la gestión de desempeño.

## Caso de Uso: Administración del Back Office

### Automatización de Procesos

A pesar que los proveedores de Servicios Financieros Digitales están poniendo mucho esfuerzo en el desarrollo de automatización de front-end (móviles, banca en línea), algunos todavía luchan para desarrollar funciones de back-end altamente automatizadas. Tareas automatizadas que pueden ayudar a las operaciones de back-office - tales como la suscripción y originación de créditos, el procesamiento de transacciones y la conciliación automatizada - tienen un valor tremendo. Los proveedores están ahora avanzando hacia la automatización robótica de los procesos simples y repetidos, que pueden ser llevados a cabo de manera mucho más barata y precisa por las máquinas que por los seres humanos. De acuerdo con AT Kearney, la Automatización Robótica de Procesos (RPA) hace operaciones en promedio 20 veces más rápido que los seres humanos e incluye beneficios del 25 al 50 por ciento de ahorro de costos para los que los adoptan.<sup>21</sup> Varias áreas de automatización generalmente se pueden agrupar en la automatización del registro de datos y el procesamiento de datos.

El enfoque principal del registro de datos radica en la digitalización de flujos de trabajo hechos en papel. Observamos que muchos proveedores todavía usan formularios de solicitud hechos en papel

para recopilar información de apertura de cuenta. Múltiples errores que se producen a lo largo del proceso de la digitación manual obligan a que estos formularios atraviesen múltiples ciclos de reelaboración. Eventualmente, después de pasar por un proceso de verificación de varios pasos, la información clave se registra en el sistema manualmente por el front o middle-office, creando una carga adicional sobre el personal y causando una repartición de tiempo ineficiente. Estos formularios luego tienen que guardarse en almacén físico y mantenerse durante cierto período de tiempo. Racionalizar y simplificar el proceso de recopilación de datos a través de la interfaz de front-end y mediante un sistema de controles de datos incorporados aumenta la eficiencia y reduce los costos laborales. Por supuesto, con el fin de registrar los datos de una manera robusta, la arquitectura de TI debe ser lo suficientemente fuerte como para clasificar correctamente, corroborar y almacenar datos.

El procesamiento de datos puede ser automatizado en casi todas las etapas de la relación con el cliente. Establecer pasos estándar de verificación que puedan acelerar la apertura de cuentas y cambios de cuentas, y las decisiones crediticias para ciertos segmentos se pueden activar por medio de modelos de scoring probados, bien estructurados. Además, los mapas de calor para acción pueden automatizar los desembolsos, y los formularios automatizados de solicitud

y retroalimentación pueden digitalizar los cierres de cuentas. Las analíticas avanzadas, que se describen en el capítulo anterior y las cuales pueden incluir la generación de prospectos para campañas de ventas o gestión multicanal, pueden utilizarse para descubrir oportunidades y riesgo no explotados dentro de la cartera. Una vez identificadas, las notificaciones automatizadas pueden enviarse directamente al personal de front-office o directamente a los clientes. Por ejemplo, para la prevención de la rotación, los clientes que se acercan al estado inactivo pueden recibir mensajes de texto o correos electrónicos de reactivación. Los prestatarios pueden recibir notificaciones sobre los próximos pagos o productos de mejor precio disponibles para la refinanciación. Algunas funciones que requieren intervenciones humanas, tales como el análisis financiero y de negocios y la gestión de relaciones personales, complementarán y se beneficiarán del proceso automatizado.

### Supervisión del Riesgo y Cumplimiento Regulatorio

Después de la crisis financiera de 2008, los reguladores nacionales han estado ajustando continuamente la regulación de la industria financiera para proteger tanto a los clientes como a la industria en general. El aumento de los requisitos de capital, liquidez y transparencia imponen una pesada carga a la industria financiera regulada, al tiempo que crean una ventaja competitiva para los actores no regulados,

<sup>21</sup> 'Robotic Process Automation: Fast, Accurate, Efficient', A.T. Kearney, consultado Abril 3 de 2017, <https://www.atkearney.com/financial-institutions/ideas-insights/robotic-process-automation>

## 1.2\_APLICACIONES DE DATOS

como los proveedores de tecnología financiera. En consecuencia, los bancos tienen que presupuestar mayores costos de cumplimiento para cumplir con los requisitos reglamentarios. Los informes regulatorios requieren reunir datos de diversos sistemas, entre ellos: contabilidad financiera, sistema de contabilidad, tesorería, monitoreo de la calidad de los activos y bases de datos de cobro, entre otros. Las pruebas de estrés periódicas requieren una sólida infraestructura de TI con una alta capacidad para almacenar y procesar grandes cantidades de datos. Además, el cumplimiento de los requisitos de KYC requiere datos reales para la toma de decisiones oportunas y seguras. Los datos necesarios para medir y monitorear los riesgos de mercado, crédito, AML, y liquidez están idealmente alojados en un repositorio unificado para permitir que un proveedor de SFD tenga una visión completa del riesgo en toda su cartera.

Este repositorio unificado también permite al proveedor de SFD ejecutar análisis de escenarios y pruebas de estrés para cumplir con los requisitos reglamentarios. El cumplimiento regulatorio incurre en costos directos a través del mayor costo del capital, así como de los costos indirectos, tales como el establecimiento de procesos de informes, la repartición del tiempo del personal y, en algunos casos, la inversión en tecnología nueva.

### Prevención del Fraude

Con las tendencias globales avanzando hacia la computación en la nube, la gobernanza y la protección de datos se vuelven cada vez más importantes. Los proveedores de SFD tienen que prestar mayor atención al comportamiento transaccional de los clientes. También deben efectuar cumplimiento con KYC para detectar posibles actividades fraudulentas - tales como el lavado de dinero y la identidad

falsa - mientras se evitan o reducen los riesgos operativos y financieros. Las nuevas intervenciones y regulaciones sobre ciberseguridad requerirán que los proveedores de SFD desarrollen y mantengan herramientas que apunten a proteger contra amenazas externas y posibles actividades delictivas. Mantener y agregar los datos apropiados necesarios para construir modelos de prevención del fraude y del riesgo operacional puede reducir la exposición del proveedor de SFD. La transmisión y procesamiento de datos en tiempo real les permite detectar fraudes con mayor rapidez y precisión, reduciendo así los posibles riesgos de pérdidas. Por ejemplo, si las tarjetas de crédito o de débito de un cliente se utilizan desde una ubicación geográfica inusual o con una frecuencia inusual, los proveedores de SFD pueden alertar al cliente y potencialmente bloquear el procesamiento de estas transacciones sospechosas.

## Seguimiento de Datos para Detección de Fraude



En el contexto de los proveedores de Servicios Financieros Digitales que ofrecen servicios P2P, los proveedores pueden utilizar una variedad de herramientas para determinar si las transacciones se depositan de forma fraudulenta en la cuenta de otra persona con el fin de evitar cargos. En lugar de usar su propia cuenta y pagar cargos, los clientes pueden iniciar una transferencia P2P pagada desde la cuenta de otra persona. La velocidad de transacción puede dar una indicación básica; si el dinero se deposita en una cuenta y luego se retira de nuevo en un período muy corto de tiempo, hay una buena probabilidad de que se trate de un depósito directo. La ubicación de la transacción da una indicación aún mejor porque si la ubicación de los corresponsales que realizan el depósito y el retiro se encuentra a cierta distancia, es improbable, o incluso imposible, que el cliente pudiera haber viajado entre esos puntos en el intervalo entre las transacciones. Debe ser posible generar alertas para este tipo de comportamiento, y los corresponsales que hacen cantidades inusualmente altas de depósitos directos se les puede dar seguimiento. Esto no capturará las transacciones entre clientes que viven cerca, así que muchos proveedores de SFD también realizan investigación de comprador secreto para entender mejor los niveles de depósitos directos.

## Caso de Uso: Gestión de Interacción con el Usuario

La gestión de los clientes a lo largo del ciclo de vida, fomentar un mayor uso, y el manejo de nuevos comportamientos se encuentra dentro del alcance del equipo de marketing. Sin embargo, también hay un aspecto operacional de la gestión de clientes que es sobre todo motivo de preocupación para los equipos técnicos, de servicio al cliente, y riesgo. Estos equipos son responsables de asegurar que la interacción del usuario, según fue diseñada, esté detectando y arreglando cualquier problema. También son responsables de gestionar la interacción del usuario para clientes empresariales y usuarios internos.

A este respecto, es importante definir el uso y el comportamiento esperado 'normal' del sistema, de manera que se puedan hacer pronósticos tanto para la planificación técnica como para la comercial. Las medidas suelen establecerse de arriba hacia abajo, como metas mensuales de negocios y objetivos estratégicos. Dicho esto, algunas métricas de resultados deben recogerse de 'abajo hacia arriba', tales como las mediciones del uso promedio de un servicio. Como se comentó anteriormente, el uso de promedios puede ser engañoso, y el comportamiento puede tener que dividirse en sectores, y luego juntarse en una 'vista promedio' de actividad contra la cual se puedan elaborar planes. Por ejemplo, el equipo técnico necesita saber tanto el número esperado de transacciones por día como también los períodos de máximo uso probable para poder garantizar que el sistema pueda hacer frente a estos puntos de máximo uso.

Definir patrones de 'comportamiento normal' es fundamental para la gestión de riesgo. Los patrones de actividad que se desvían de las normas acordadas, en particular los datos transaccionales y de uso de servicios, deben generar alertas. Estos patrones deben ser revisados para determinar si el comportamiento atípico era legítimo, o era un posible caso de fraude. Además del comportamiento de los clientes y los corresponsales, también es aconsejable perfilar la 'actividad normal' de las interacciones de los empleados en el sistema. Por ejemplo, ¿hay un empleado mirando significativamente más registros de clientes que un empleado 'normal' en el mismo rol, o está entrando al sistema fuera de sus patrones normales de turno? Esta actividad anormal podría apuntar a una actividad fraudulenta potencial.

### Mejoras en la Eficiencia del Servicio al Cliente

Los equipos de servicio al cliente en los call centers son los empleados más cercanos al cliente de SFD en el día a día. Debido a esto, pueden proporcionar una alerta temprana de cualquier problema importante que pueda surgir. A menudo, serán los primeros en enterarse de un fallo del sistema o comportamiento fraudulento de un corresponsal, por lo que se requiere un proceso para alertar al equipo apropiado de cualquier problema potencial basado en la información (comprobada la precisión) recibida de los clientes. Es probable que estos equipos también se enteren de problemas menores que afectan al servicio impidiendo que los clientes realicen transacciones de manera óptima,

tales como la falta de corresponsales, límites restrictivos de transacciones y tiempos de espera de transacción cortos. Es importante, entonces, recopilar datos estadísticos sobre las llamadas recibidas, incluidas las quejas y sugerencias. El aprovechamiento de este tipo de datos se ejemplifica en el Caso 8 a continuación.

El monitoreo del número de llamadas a medida que el servicio crece ayuda a determinar cuántos asesores de call center son necesarios. Para algunos servicios de alta demanda, sólo una proporción de las llamadas presentadas realmente llegan a una línea de atención al cliente. En este caso, los intentos de llamadas frente a las llamadas presentadas son una cifra importante, ya que esto indica un problema importante o niveles de personal inadecuados. Los problemas de call center más frecuentes son PINES olvidados, teléfonos o tarjetas perdidos, transacciones enviadas a los destinatarios equivocados y códigos de comprobantes perdidos. El número de llamadas que se pueden realizar depende de la velocidad del sistema de back-office y de la rapidez con la que pueda responder al resolver el problema. Como los costos del call center son generalmente altos, los datos que proporcionan deben ser utilizados para acelerar el proceso de resolución de conflictos e incrementar el número de llamadas que cada representante puede tomar. Estos datos también se pueden utilizar para mejorar la experiencia del usuario y de esta manera lograr que el cliente cometa menos errores.

# CASO 8

## *Safaricom M-Pesa - Uso de KPIs para Mejorar el Servicio al Cliente y los Productos*

### **Uso de Análítica de datos para Identificar los Cuellos de Botella Operacionales y Priorizar las Soluciones**

*M-Pesa en Kenia fue el pionero de los Servicios Financieros Digitales a gran escala, con 20.7 millones de clientes, una base activa de treinta días de 16,6 millones,<sup>22</sup> e ingresos reportados en 2016 de \$4.5 billones.<sup>23</sup> Cuando Safaricom lanzó el servicio en 2007, no había plantillas ni mejores prácticas; todo fue diseñado desde cero. La mejora continua de la operación fue esencial a medida que el servicio escalaba.*

*La adopción del servicio fue inesperadamente alta desde el inicio, con más de 2 millones de clientes en su primer año, superando las previsiones en un 500 por ciento. Esta creciente demanda obligó a una rápida escalada y las operaciones necesarias para anticipar proactivamente los problemas de escalada en la tecnología y los procesos de negocios, ya que una mala experiencia del*

*cliente podría erosionar rápidamente la confianza. Las métricas basadas en datos ayudaron al equipo a planificar y guiar las operaciones en forma apropiada.*

*Como la adopción de servicios fue inesperadamente alta desde el principio, el número de llamadas al call center fue en consecuencia mucho mayor de lo previsto, dando lugar a un gran volumen de llamadas sin respuesta. Este problema estableció un KPI que el equipo de atención al cliente necesitaba para resolver a niveles aceptables.*

*El problema se abordó primero mediante la contratación de personal adicional, pero la contratación por sí sola no podía seguir el ritmo del aumento del número de clientes. Para identificar cuellos de botella y priorizar las soluciones, el equipo analizó sus datos. Se examinaron*

*los datos de llamadas de PABX y los registros de resolución de problemas y se encontró lo siguiente:*

- **Duración de Llamada:** *La duración media de la llamada fue de 4.5 minutos, alrededor del doble del tiempo presupuestado para cada llamada.*
- **Temas Clave para Resolución Rápida:** *Los dos tipos de llamadas clave a abordar para la optimización eran los clientes olvidando el PIN y los clientes enviando dinero al número de teléfono equivocado; esto cubrió del 85 al 90 por ciento de las llamadas largas que llegaban al call center.*

*El análisis logró dos cosas. En primer lugar, los cuellos de botella se identificaron con éxito, resultando en información clave para operaciones. En segundo lugar, otras cuestiones operacionales*

<sup>22</sup> Richard Mureithi, 'Safaricom announces results for the financial year 2016'. *Hapa Kenya*, 12 de mayo de 2017, consultado el 3 de abril de 2017, <http://www.hapakkenya.com/2016/05/12/safaricom-announces-results-for-the-financial-year-2016/>

<sup>23</sup> Chris Donkin, 'M-Pesa continues to dominate Kenyan market'. *Mobile World Live*, Enero 25 de 2017, consultado el 3 de Abril de 2017, <https://www.mobileworldlive.com/money/news-money/m-pesa-continues-to-dominate-kenyan-market/>

*fueron descubiertas, principalmente, el grado en el que los clientes erróneamente enviaron dinero y se olvidaron de sus pines. De esta manera, gestionar frente al KPI de Llamadas No Contestadas generó beneficios operativos más amplios.*

*Utilizando los resultados analíticos, operaciones implementó una estrategia de resolución. Primero, al entender los tipos de problemas largos frente a los cortos, los problemas difíciles se podían identificar y pasar rápidamente a un equipo de back-office. Esto redujo los tiempos de espera y los cuellos de botella para los clientes, permitiendo procesar más clientes por día. En segundo lugar, los equipos de operaciones y desarrollo de productos trabajaron para reducir los tiempos en todos los tipos de llamadas. Esto se logró mejorando la infraestructura técnica y el interfaz de usuario, mitigando los problemas que causaron llamadas largas. La combinación de iniciativas redujo el KPI de Duración de Llamada y el KPI de número de Llamadas No Contestadas, cambiando ambos a niveles aceptables a pesar de que los números de clientes siguieron creciendo más allá de los niveles previstos.*



*La gestión por KPIs es un elemento crítico de las operaciones. El análisis detallado de los datos detrás de los KPI puede ayudar a identificar cuellos de botella operacionales e incluso revelar otros factores operacionales que impulsan las métricas más allá de los umbrales. Entender los datos que impulsa un KPI puede hacerlo más útil.*

## 1.2\_APLICACIONES DE DATOS

### Caso de Uso: Datos de Operaciones Técnicas

Por su propia naturaleza, un SFD tiene que estar disponible las 24 horas del día, siete días a la semana, y normalmente está diseñado para procesar grandes volúmenes de interacciones del sistema, tanto financieras como no financieras. Por esta razón, el servicio debe ser monitoreado proactivamente tomando medidas preventivas para asegurar la disponibilidad continua del servicio. Los datos de los diagnósticos de servicio se utilizan normalmente para realizar este análisis. Los paneles de control técnicos deben actualizarse en tiempo real para mostrar la integridad del sistema. Deben ser supervisados automáticamente y diseñados para alertar a las funciones y personas responsables si se detecta un problema potencial. El concepto de usar datos para 'entender lo normal' se utiliza para detectar de forma proactiva los fallos en varias capas del servicio, y se establecen soluciones de monitoreo automático para detectar cuándo se superan los valores de umbral. Por ejemplo, si un sistema de SFD normalmente procesa un número dado de transacciones por segundo (TPS) cada jueves por la noche, pero un jueves la cifra es mucho menor, esto indica que probablemente existe un problema que requiere acción.

Las tendencias se pueden usar para predecir los problemas de desempeño al mismo tiempo que identifican incidentes específicos; debido a esto, el equipo también debe considerar el desempeño en el tiempo. El análisis de tendencias es vital en la planificación de la capacidad, y los patrones de uso y crecimiento del sistema dan pistas importantes sobre

cuándo se necesitará capacidad adicional del sistema. Tanto si el sistema es externo o si es un desarrollo interno, es importante que el equipo técnico supervise los niveles de servicio y las tendencias de capacidad, planificando acciones correctivas. Los datos clave normalmente requeridos incluye disponibilidad del sistema, tiempo de inactividad esperado y no esperado, volumen de transacciones y capacidad máxima y sostenida.

### Transacciones e Interacciones



Una transacción es un movimiento financiero de dinero, generalmente el acto de debitar una cuenta y acreditar a otra. Para que esto suceda, el usuario tiene que interactuar con el sistema. Las interacciones mismas pueden ofrecer información, y se utilizan con frecuencia en el desarrollo de productos digitales de teléfonos inteligentes y servicios web para ayudar a comprender mejor al cliente.

Las interacciones de SFD, incluso utilizando teléfonos de gama baja, se pueden medir y pueden proporcionar datos útiles sobre la experiencia del cliente para un servicio. Por ejemplo, es posible medir interacciones como 'intentos abandonados de realizar una transacción financiera', y luego diagnosticar lo que impidió que los clientes completaran estas transacciones. Otro ejemplo es cuando servicio al cliente interactúan con el sistema en nombre de un cliente, por ejemplo, restableciendo un PIN olvidado. Estas interacciones rara vez se miden, pero también pueden proporcionar información útil para mejorar las operaciones del servicio.

Los Servicios Financieros Digitales exitosos tienen una buena comunicación entre los equipos comerciales y técnicos. El equipo comercial debe discutir proactivamente sus planes de marketing y pronósticos, así como cualquier actividad competitiva con el fin de preparar al equipo técnico para posibles cambios de volumen. Se necesitan reuniones regulares (al menos trimestralmente) para revisar los últimos pronósticos de volumen basados en los resultados del trimestre anterior y la actividad de marketing planificada. Esto permite al equipo técnico planificar debidamente. El equipo técnico debe, a su vez, asesorar a cualquier socio que pueda verse afectado por un cambio en el pronóstico. Esto es particularmente relevante para los socios MNO, ya que ha habido varios casos de requerimientos de volumen de SMS inmanejables durante promociones inusualmente exitosas. Del mismo modo, si se planean cambios técnicos o revisiones, el marketing debe ser consciente y debe evitar actividades que puedan poner una presión adicional sobre el sistema en ese momento.

### Lecciones Aprendidas de Operaciones y Administración del Desempeño

**Registrar el Beneficio Comercial de Ventas de Tiempo aire:** Los informes pueden ser engañosos cuando los clientes usan SFD para comprar tiempo aire. Dependiendo del negocio principal del proveedor de SFD, vender tiempo aire prepagos puede ser una fuente

de ingresos o un ahorro de costos. Para los no-MNO, cada venta de tiempo aire traerá una pequeña comisión, ya que están actuando como un distribuidor de tiempo aire. Estos ingresos se deben considerar parte de los ingresos de los SFD. Para los MNOs, en lugar de ingresos, esta transacción representa un ahorro de costos con un impacto significativo, ya que elimina las comisiones (típicamente) del 2% al 3% y los costos de distribución. Sin embargo, muchos MNO no atribuyen este ahorro de costos al negocio de SFD porque se ha contabilizado dentro de la línea de pre-pagados de tiempo aire. Aunque esto puede ser correcto en términos contables, para medir con precisión el valor de los SFD para el negocio, este ahorro de costos debe incluirse en las cuentas de administración interna de los SFD.

**Cuidado con los Promedios:** Por su naturaleza, las ofertas de SFD tienden a

atraer a personas de recursos limitados que carecen de acceso a los bancos y las personas de mayores ingresos (y las empresas) que interactúan con ellos. Esto lleva a volúmenes muy altos de transacciones de bajo valor junto con un pequeño número de transacciones de alto valor. La visualización de datos puede ser muy eficaz para identificar dónde es inapropiado el uso de promedios. Por ejemplo, la Figura 16 muestra una curva de frecuencia de distribución típica de los valores de transacción para un proveedor de SFD, siendo la mayoría de las transacciones (moda) de \$20. Sin embargo, el valor medio de transacción es de \$86, debido a que un número relativamente pequeño de transacciones de alto valor sesga el promedio. Estos promedios pueden conducir a una visión equivocada e inflada de la riqueza y la actividad financiera del cliente 'promedio'.

### Observar Tendencias a Largo Plazo y

**Resultados a Corto Plazo:** Las tendencias proporcionan información mucho más densas que un punto de datos aislado. Los cambios deben ser entendidos en el contexto del tiempo, ya que puede haber un efecto estacional, como un día festivo, que es responsable de un salto en la actividad. Este punto de máximo uso puede ser seguido por un descenso, luego un retorno al status quo, que es común alrededor de la Navidad. También puede haber un impacto estacional; por ejemplo, durante el período de cosecha, los agricultores con cultivos comerciales obtienen la mayor parte de sus ingresos anuales y son mucho más activos económicamente en comparación con otras épocas del año. Otras causas de cambios a corto plazo en el desempeño pueden ser la actividad competitiva, el clima extremo y la incertidumbre política.

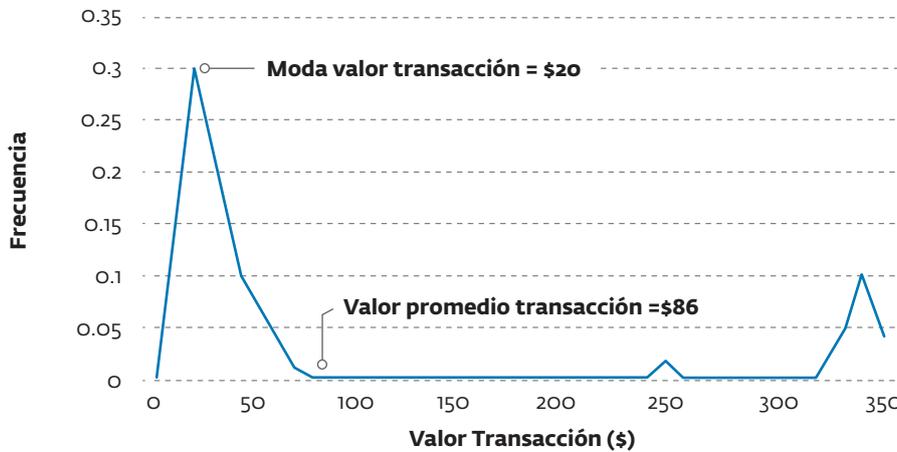


Figura 16: Gráfico de frecuencia del valor de transacción que demuestra que los promedios pueden conducir a conclusiones equivocadas

## 1.2\_APLICACIONES DE DATOS

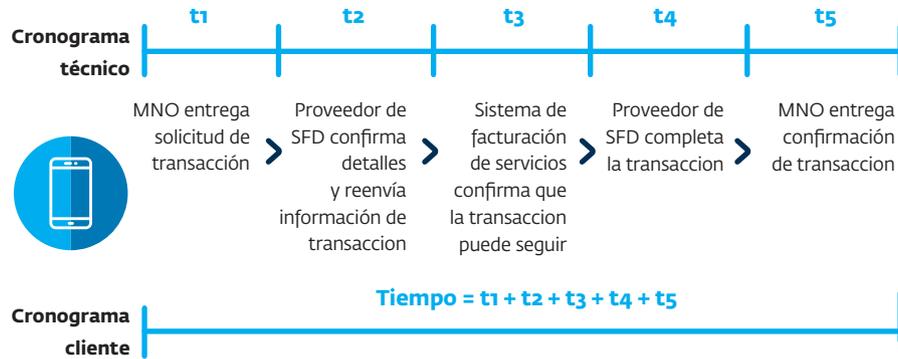
**Cuidado con Métricas de Vanidad:** Las métricas de vanidad pueden verse bien en el papel, pero pueden dar una visión falsa del desempeño del negocio. Son fácilmente manipulables y no necesariamente se correlacionan con los datos que realmente importan, como la vinculación, el costo de adquisición y, finalmente, los ingresos y beneficios. Un ejemplo típico de las métricas de vanidad de SFD es informar de los clientes registrados, en lugar de los activos. También, reportar corresponsales totales en lugar de corresponsales activos. Sólo al centrarse en los KPI reales y métricas críticas es posible comprender adecuadamente la salud de la compañía. Si una empresa se centra en las métricas de vanidad, puede obtener una falsa sensación de éxito.

**Los Datos de Nivel de Servicio deben ser Relevantes para los Objetivos de Negocio:** Cada equipo de operaciones recopila una gran cantidad de datos sobre cómo está funcionando su sistema. Sin embargo, en SFD complejos y multi-socios, pueden no considerar el desempeño del servicio de extremo a extremo y su efecto sobre la experiencia del usuario. Para un cliente, el indicador de desempeño que es relevante es el desempeño de la transacción de extremo a extremo; ¿se completó la transacción y cuánto tiempo tomó? Es sorprendente cómo pocos SFD miden este desempeño de transacción de extremo a extremo dado su papel fundamental en establecer

y mantener la confianza del cliente, estableciendo la aceptación de los SFD y manteniendo la reputación del negocio. La Figura 17 ilustra el problema de un cliente que utiliza su teléfono para pagar una factura. En este caso, hay tres 'propietarios del sistema' involucrados: un MNO que suministra conectividad, el proveedor de SFD que suministra la transacción y el facturador al que se paga.

Cada sistema devuelve sus propios datos de eficiencia, pero la experiencia del cliente puede ser muy diferente si hay demoras de entrega entre sistemas. Otro ejemplo común es cuando los MNOs proporcionan sesiones de Servicio

Suplementario de Datos no Estructurados (USSD) con un tiempo de espera demasiado corto o un fallo de abandono de USSD de manera que algunos clientes físicamente no puedan completar una transacción en el tiempo asignado. Debe ser sencillo en una relación proveedor-vendedor pedir datos que muestren información relevante, por ejemplo, deserciones de USSD o colas de transacciones. Sin embargo, a menudo es un problema crítico en la provisión de SFD que no hay Acuerdos de Nivel de Servicio (SLA) directos o integrales, lo que a veces puede hacer imposible entender la información en este detalle.



*Figura 17: Tiempo de Transacción: Medidas del Sistema versus la Experiencia del Cliente*

**Filtrado del “Diluvio de Datos”:** Cada interacción con un sistema de SFD puede generar un gran número de puntos de datos. Algunos de éstos serán financieros, y algunos registrarán qué interfaz se está utilizando, o incluso cuánto tiempo le toma al usuario navegar por la experiencia de usuario. La intensidad de la información recogida aumenta enormemente a medida que los sistemas hacen un uso creciente de interfaces de usuario más avanzadas, tales como los teléfonos inteligentes. Esto puede conducir a la sobrecarga de información y al ‘fallo de filtro’ - esencialmente, una inhabilidad de ver árboles por culpa del bosque. Esto, junto con las restricciones en torno a asegurar los recursos necesarios para administrar estos mecanismos de envío de datos, es la razón por la que tan poca de esta información está siendo utilizada por el negocio para la toma de decisiones. Clasificar y correlacionar la información externa con datos internos puede conducir a una pérdida de información clave.



# CASO 9

## M-Kopa Kenia - Modelos de Negocios Innovadores y Estrategias Basadas en Datos

### La Cultura de Negocios Impulsada por Datos Incorpora la Analítica entre Operaciones, Productos y Servicios

Fundada en Kenia en 2011, M-Kopa comenzó como un proveedor de sistemas de energía solar para el hogar, principalmente para la iluminación, a la vez que para recarga de pequeños dispositivos como teléfonos móviles y radios. El negocio combina la tecnología de máquina a máquina, utilizando tarjetas SIM integradas con una solución de micro-pagos lo que significa que la tecnología puede ser monitoreada y puesta a disposición sólo cuando se recibe el pago anticipado. Los clientes compran sistemas M-Kopa usando 'créditos' a través del servicio M-Pesa de dinero electrónico, luego pagan los sistemas usando M-Pesa hasta que el saldo se pague por completo y el producto sea de su propiedad. En los últimos años, el negocio se ha expandido a otras áreas, incluyendo la provisión de electrodomésticos y préstamos, utilizando unidades solares de propiedad del cliente como garantía de refinanciamiento. Estos productos se ofrecen a los clientes que han construido un score crediticio de 'capacidad de

pago', evaluada por la compra inicial del sistema y el pago posterior. M-Kopa también está disponible en Uganda, Tanzania y Ghana.

M-Kopa utiliza datos proactivamente en toda la dimensión del negocio para mejorar la eficiencia operativa. Sus bases de datos acumulan información sobre la demografía de los clientes, la dependencia del cliente en el dispositivo y el comportamiento de pagos. Cada unidad solar transmite automáticamente los datos de uso y la información de diagnóstico del sistema a M-Kopa, informándoles cuando, por ejemplo, las luces están encendidas. Todo esto se puede analizar para mejorar la calidad del servicio, la eficiencia operacional y la comprensión del comportamiento del cliente.

#### Gestión de la Capacidad Técnica

Un análisis del comportamiento del cliente en uso y pagos muestra que los usuarios prefieren comprar créditos por adelantado para asegurar una

energía fiable para los siguientes días. Sabiendo cuándo es probable que los clientes paguen (y con cuánto tiempo de antelación), M-Kopa puede pronosticar las expectativas y planificar en consecuencia, asegurando que sus clientes no se verán afectados por interrupciones anunciadas de M-Pesa que podrían impedir que estos pagos se ejecuten.

#### Servicio al Cliente

Los dispositivos M-Kopa comunican los datos de la batería cuando se registran y el análisis de los datos permiten al servicio de atención al cliente comprobar si las unidades están funcionando según lo previsto y permite un mantenimiento proactivo y preventivo que se puede realizar de forma remota:

- Si un cliente se queja de que no está recibiendo la cantidad esperada de energía, los paneles de control de la batería se utilizan para diagnosticar el problema. Por ejemplo, la batería no se está cargando completamente durante las horas diurnas.

- *A pesar de los buenos controles de calidad de manufactura, siempre hay variaciones en el desempeño de la batería cuando las unidades están en el campo, determinado por factores tales como patrones de uso o condiciones ambientales. M-Kopa ha creado algoritmos de mantenimiento predictivo para detectar el desempeño sub-óptimo de la batería, lo que le permite intervenir y organizar un reemplazo gratuito antes que se produzca un ‘fallo’ de la batería.*

### **Gestión de Equipos de Ventas**

*El equipo de ventas de campo vende productos y servicios de M-Kopa directamente a los clientes. Los*

*representantes de ventas utilizan una aplicación de teléfono inteligente para registrar todas sus actividades digitalmente, en tiempo real. Esto permite una comprensión detallada de su desempeño y una resolución rápida cuando hay problemas. Las medidas dinámicas de desempeño en línea y las tablas de clasificación pueden ser desglosadas por persona y están a disposición del equipo de gestión de ventas y los líderes de equipo para fomentar las mejoras de desempeño a través de la gamificación.<sup>24</sup> La aplicación también permite a los miembros del equipo realizar un seguimiento de su comisión y cualquier bono e incentivo adicional.*

### **Enfoque en Clientes Potenciales para Ventas Adicionales**

*El comportamiento de pagos del cliente puede proporcionar una gran cantidad de información sobre la salud financiera y la solvencia. Los datos de la batería muestran la dependencia del cliente con el dispositivo para iluminación, lo que añade un nivel más profundo de comprensión. Esta información se utiliza para identificar y activamente enfocarse en los clientes existentes para actualizaciones y servicios adicionales. M-Kopa también comparte esta información con burós de crédito para ayudar a proporcionar a los clientes una calificación crediticia.*



*Una cultura corporativa impulsada por los datos es necesaria para integrar la analítica y elaboración de informes en toda la empresa. Esto ayuda a aprovechar las fuentes de datos y analíticas en múltiples áreas para atraer a nuevos clientes, administrar equipos de ventas, ofrecer un mejor servicio al cliente y desarrollar nuevos productos.*

<sup>24</sup> Gamificación es la aplicación de elementos de diseño de juegos y principios de juego en contextos distintos al juego. Se pueden encontrar más ejemplos de Servicios Financieros Digitales en el sitio web de CGAP: <https://www.cgap.org/blog/series/gamification-and-financial-services-poor/>

## 1.2\_APLICACIONES DE DATOS

### **Interacciones del Sistema de Almacenamiento:**

Incluso hace unos años, cuando se lanzaban muchas ofertas de Servicios Financieros Digitales, la captura y el almacenamiento de datos eran relativamente costosos y engorrosos, por lo que no se conservaban los datos que no eran inmediatamente necesarios para administrar un negocio. La nueva tecnología permite un almacenamiento de datos barato y abundante. Aunque normalmente se ignoran, también hay nuevas herramientas para analizar los datos que están en archivos de registro en los servidores que hacen posible, con las herramientas adecuadas, correlacionar múltiples fuentes de datos para proporcionar información más densa sobre los servicios. Se recomienda muy especialmente que los proveedores de SFD recopilen y almacenen todos los bits de datos que puedan sobre cada interacción del sistema, incluso aquellas que se rechazaron. Si bien puede no parecer útil o relevante para las operaciones actuales, puede ser de valor a futuro para análisis de datos o investigación de fraude.

Los principios de no repudio requieren que estos cambios se registren como eventos adicionales, en lugar de intentar editar registros previamente finalizados. Por ejemplo, si una comisión debe ser recuperada de un corresponsal, esto debe ser registrado explícitamente como una actividad separada (pero vinculada), en lugar de pagar silenciosamente una cantidad menor, o simplemente ajustar el archivo de la comisión por pagar.

### **Combinación de Datos para Agregar**

**Contexto:** La combinación de datos de proveedores de SFD con datos de aliados puede tener muchos beneficios operacionales. Por ejemplo, cuando hay colaboración con un MNO, también hay información sobre dónde se encontraban físicamente el remitente y el destinatario, la tarjeta SIM utilizada, el tipo de teléfono utilizado, los registros potenciales de llamadas y los patrones de recarga del cliente. Como muchos mercados tienen un mandato de registro de tarjeta SIM estricto, la información de KYC del cliente también se puede utilizar para completar y cruzar los registros de referencia. Aunque algunos de estos parámetros no son de importancia primordial para las transacciones, estos datos son útiles para determinar las anomalías del sistema; por ejemplo, si un cliente normalmente realiza transacciones desde un teléfono determinado y ese teléfono ha cambiado, puede ser que la transacción sea fraudulenta. Se pueden obtener más pruebas cruzando la ubicación donde la transacción tuvo lugar con el registro de ubicación normal del cliente.

Puede haber desafíos al intentar correlacionar datos de diferentes fuentes, que requieren consideración durante el proceso de diseño de la base de datos. Por ejemplo, incluso cuando el MNO es parte de la misma organización que el proveedor de SFD, el intercambio de datos puede ser un problema porque los dos sistemas no han sido diseñados para proporcionar servicios de información entre sí. Intentar vincular retrospectivamente los datos de telecomunicaciones de una interacción del cliente con un sistema con la información transaccional

financiera de SFD no es sencillo. Esto suele ser porque no hay ninguna pieza común de los datos que vincule a los dos registros, e incluso los relojes que marcan la hora del evento en los dos sistemas es poco probable que estén perfectamente sincronizados. Debido a esto, muchos sistemas sólo realizan la actividad de combinación de datos por excepción, generalmente para investigaciones de fraude caso por caso. Sin embargo, el contexto adicional proporcionado por los datos combinados puede agregar capas de valor, particularmente en el caso del monitoreo proactivo del fraude. Hacer más fácil combinar los datos para que puedan ser utilizados en las actividades operacionales 'normales', especialmente para las operaciones de SFD más maduras.

**Intentos Fallidos:** Es común que los proveedores de Servicios Financieros Digitales retengan los datos asociados con las transacciones exitosas, cuando se completó la actividad solicitada. Sin embargo, las transacciones fallidas también pueden proporcionar información. Las razones por las que se rechazaron determinadas transacciones pueden apuntar a necesidades muy específicas, como la necesidad de proporcionar información y educación un fallo técnico o una deficiencia en el diseño del servicio que debe modificarse para proporcionar una experiencia de usuario más intuitiva.

Con el fin de realizar estas analíticas avanzadas, cada bit de información sobre cada interacción del sistema debe ser recogido y almacenado, incluso si su relevancia no es inmediatamente obvia.

**Única Fuente de Verdad:** Cuando hay múltiples sistemas, es común tener la misma data duplicada en múltiples lugares. Esto es a menudo porque la infraestructura actual hace que sea difícil combinar fuentes de datos de otra manera. Esta duplicación de datos puede conducir a problemas con respecto a la 'fuente de la verdad', en otras palabras, en cuál fuente de datos se puede confiar cuando hay información contradictoria. Todos los sistemas están ocasionalmente sujetos a errores, y cuando hay una disputa sobre los detalles de la transacción o un debate sobre si los fondos fueron transferidos, tiene que haber un acuerdo claro sobre los datos de los que fiarse. El manejo de estos detalles es

parte de cualquier proyecto que combina y compara fuentes de información; también es importante entender claramente si un registro es definitivo o aún puede actualizarse. Tratar incorrectamente un registro no final como final puede llevar a estragos en la analítica de datos, creando desconfianza en la integridad de la plataforma.

### 1.2.3 Analítica y Aplicaciones: Scoring crediticio

El scoring crediticio puede describirse ampliamente como el estudio del comportamiento pasado y características

del prestatario para predecir el comportamiento futuro de los prestatarios nuevos y existentes.<sup>25</sup> La aparición de Big Data y las fuentes y formatos de estos datos han presentado enfoques adicionales para el proceso de scoring crediticio. La incorporación de estas fuentes alternativas de datos impulsa modelos de calificación de crédito alternativos. Esta sección examina la forma cómo los datos impulsan el scoring crediticio y qué tipos de datos funcionan mejor para diversas necesidades. Las relaciones fundamentales de scoring crediticio se representan como una línea de tiempo en la siguiente figura.



Figura 18: Cronograma de Definición de Scoring Crediticio

<sup>25</sup> Schreiner, 'Credit scoring for microfinance: Can it work?', *Journal of Microfinance/ESR Review*, Vol. 2.2 (2009): 105-118

## 1.2\_APLICACIONES DE DATOS

A continuación, se presentan los puntos clave ilustrados en la Figura 18:

1. **Pasado:** Se estudian los datos (o, en su ausencia, la experiencia) para entender cuáles son características del prestatario que están más significativamente relacionadas con el riesgo de impago. Este estudio del pasado informa la elección de factores y ponderaciones en el panel de control.
2. **Presente:** El Scorecard (construido sobre los datos pasados característicos del prestatario) se utiliza para evaluar las mismas características en los nuevos solicitantes de préstamos. El resultado es un scoring numérico que se utiliza para situar al solicitante en un 'grupo de riesgo', o rango de puntuaciones con tasas de pago similares observadas.
3. **Futuro:** El modelo supone que los nuevos solicitantes con las mismas características que los prestatarios anteriores presentarán el mismo comportamiento de pagos que los prestatarios anteriores. Así, la tasa de morosidad observada en el pasado para un grupo de riesgo dado es la tasa de morosidad pronosticada para los nuevos prestatarios en ese mismo grupo de riesgo.

Se podría escribir todo un manual sobre el scoring crediticio, y de hecho varios textos completos y accesibles se han publicado sobre el tema durante la última década.<sup>26</sup> Además, CGAP publicó recientemente una introducción al scoring crediticio en el contexto de los Servicios Financieros Digitales.<sup>27</sup> Para efectos de este manual, el resto de esta sección de crédito se centra en:

1. Cómo se convierten los datos en puntuaciones de crédito
2. Cómo se utilizan los datos para hacer frente a los desafíos de la evaluación de crédito en los mercados en desarrollo

### Desarrollo del Scorecard

Los scorecards de crédito se desarrollan mirando una muestra de datos sobre préstamos pasados que han sido clasificados como 'buenos' o 'malos'. Una definición común de préstamos 'malos' (o 'deficientes') es 90 o más días consecutivos en mora,<sup>28</sup> pero para el desarrollo de un scorecard, un préstamo malo debe ser descrito como uno (visto en retrospectiva) que las Instituciones Financieras elegirían no hacer otra vez en el futuro. Para cada nuevo solicitante de crédito, el modelo de calificación calculará e informará qué porcentaje de

los prestatarios anteriores con la misma combinación de características del prestatario fueron 'malos'.

Es importante llevar a cabo análisis sobre los préstamos buenos y malos. Estudiar las relaciones de riesgo en los datos de crédito es tan simple como mirar el número de préstamos buenos y malos para diferentes características del prestatario. Cuantos más préstamos malos haya como proporción del total de préstamos para una característica dada de prestatario, más riesgo habrá.

La tabulación cruzada, o tabla de contingencia, es una herramienta analítica simple que se puede utilizar para construir y administrar scorecards de crédito. La Tabla 4 muestra el número de préstamos buenos y malos entre rangos de valores, en un ejemplo desde un campo de datos de un MNO, en este caso, el tiempo desde la vinculación a la red móvil. Supongamos que la expectativa es que los solicitantes con un historial más largo en la red móvil tendrán menos riesgo (por lo general, los historiales más largos, ya sea en el empleo, en los negocios, en la residencia o como cliente de un banco, están vinculados a un menor riesgo).

<sup>26</sup> Vea por ejemplo: Siddiqi, 'Credit risk scorecards: developing and implementing intelligent credit scoring', *John Wiley and Sons*, Vol. 3 (2012). Anderson, 'The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation', Oxford University Press, 2007

<sup>27</sup> 'An Introduction to Digital Credit: Resources to Plan a Deployment', *Consultative Group Against Poverty via Slide Share*, Junio 3 de 2016, consultado April 3 de 2017, <http://www.slideshare.net/CGAP/an-introduction-to-digital-credit-resources-to-plan-a-deployment>

<sup>28</sup> Para los Servicios Financieros Digitales y los operadores de microcrédito, la definición de préstamo 'malo' a menudo puede ser un período de morosidad mucho más corto, como 30 o 60 días de moras consecutivas. El diseño del producto (incluidas las sanciones y los cargos por mora) y la mano de obra involucrada en los procesos de recaudación influirán en el punto en el que un cliente es 'malo' o es mejor evitarlo.

Fila		<= 2 meses	> 2 meses y <= 1 año	> 1 año y <= 2 años	> 2 años y >= 3 años	> 3 años	Total
A	Buenos	115	161	205	116	203	800
B	Malos	48	48	50	24	30	200
C	Tasa Negativa	29.4%	23.0%	19.8%	17.3%	12.7%	20.0%
D	Total	163	210	255	140	233	1,000
E	% Créditos Totales	16.3%	21.0%	25.5%	14.0%	23.3%	

**Tabla 4: Tabulación Cruzada de Créditos**

La Tabla 4 se puede leer de la siguiente manera:

**Fila A:** Número de contratos malos en el grupo (columna)

**Fila B:** Número de contratos malos en el grupo (columna)

**Fila C:** Número de contratos malos (fila B) / Número de contratos totales (fila D)

**Fila D:** Número de contratos totales (fila A + fila B)

**Fila E:** Total de contratos en el grupo (columna) dividido por todos los contratos (1.000)

Para llevar a cabo el análisis, el siguiente paso es buscar patrones sensibles e intuitivos. Por ejemplo, la tasa negativa en la fila C de la Tabla 4 disminuye claramente a medida que aumenta el tiempo transcurrido desde el registro en la red. Esto coincide con la expectativa inicial. Una forma fácil de pensar en el riesgo de

cada grupo es considerar su tasa negativa relativa al 20 por ciento (promedio) de tasa de impago por tiempo desde el registro:

- Menos de 2 meses, la tasa negativa es de 29 por ciento, una vez y media el promedio.
- Entre 1 año y 2 años, la tasa negativa de 19,8 por ciento, o riesgo promedio.
- Más de 3 años, la tasa negativa es del 12,7 por ciento, un poco más de la mitad del riesgo promedio.

En el desarrollo tradicional de scorecards de crédito, los analistas buscan patrones sencillos -incluyendo tasas malas en ascenso o descenso constante - que encajan con el sentido de negocios (y el común). Los scorecards de crédito desarrollados de esta manera se traducen bien al uso operativo como herramientas de negocio que son transparentes y bien comprendidas por la

administración. Un enfoque alternativo al desarrollo de un scorecard es la minería de datos, o el uso de algoritmos de aprendizaje automatizado más complejos para cualquier relación en un conjunto de datos, ya sea entendido o no por un analista humano. Aunque un enfoque puramente de aprendizaje automático puede resultar en una predicción mejorada en algunas situaciones, también hay ventajas prácticas pero difíciles de medir cuando los negocios y la gestión de riesgo entienden completamente cómo se calculan las puntuaciones.

La tabulación cruzada o el análisis similar de predictores únicos es el componente básico de los modelos de score crediticio.<sup>29</sup> Crear tabulaciones cruzadas como las del ejemplo anterior es fácil con cualquier software estadístico comercial o con el software libre 'R' de código abierto.

<sup>29</sup> De hecho, se pueden calcular coeficientes de regresión logística directamente a partir de una tabulación cruzada para una única variable.

## 1.2\_APLICACIONES DE DATOS

### Caso de Uso: Desarrollo de Scorecard

Los puntos del scorecard son transformaciones de los patrones de tasa negativa observados en las tabulaciones cruzadas. Aunque existen muchos métodos matemáticos que se pueden utilizar para construir scorecards (véase el capítulo 1.2.3), los diferentes métodos dan resultados similares. Esto se debe a que el poder predictivo del modelo de scoring estadístico no proviene de la matemática, sino de la fuerza de los datos mismos. Dados los datos adecuados sobre las características relevantes del prestatario, los métodos simples producirán un buen modelo y los métodos complejos pueden producir un modelo ligeramente mejor. Si no hay datos buenos (o muy pocos datos), ningún método dará buenos resultados. La verdad es que el desarrollo del scorecard

no sólo favorece modelos simples, sino que también significa que un proveedor de Servicios Financieros Digitales impulsado por datos debe centrarse inicialmente en capturar, limpiar y almacenar más y mejores datos.

La Tabla 5 abajo es otra tabulación cruzada, esta vez para el factor 'edad'. Al igual que en la tabla anterior, las tasas negativas en la fila C muestran el riesgo (la 'tasa negativa'), que disminuye a medida que aumenta la edad.

### Diferencias de Tasa Negativa

Una manera muy simple de convertir las tasas negativas en el scorecard es calcular las diferencias en las tasas negativas. Como se muestra en la fila G, la tasa negativa para cada grupo es restada de

la tasa más alta para todos los grupos (aquí es 30.9 por ciento para '23 o menor'), que se multiplica por 100 (para obtener números enteros en lugar de decimales). Los resultados (mostrados en la fila F) podrían utilizarse como puntos en el scorecard. En un esquema de puntos de este tipo, el grupo más arriesgado siempre recibirá 0 puntos y el grupo de menor riesgo (es decir, el grupo con la tasa negativa más baja) recibirá el mayor número de puntos.

Para los scorecards desarrollados mediante regresión (véase el capítulo 1.1), la transformación de los coeficientes de regresión en puntos positivos implica unos pocos pasos adicionales. Los cálculos no se muestran aquí, pero los resultados de clasificación son muy similares, como se muestra en la fila H.

Fila		23 años o menos	24 a 30 años	31 a 47 años	48 años o más	Total
A	Buenos	46	238	374	142	800
B	Malos	20	74	82	23	200
C	Tasa Negativa	30.9%	23.8%	18.0%	14.0%	20.0%
D	Total Columna	66	312	456	166	1,000
E	Porcentaje de Préstamos Totales	6.6%	31.2%	45.6%	16.6%	
F	PUNTOS	0	7	13	17	
G	Cálculo [multiplicado por 100]	$(.309 - .309) = 0$	$(.309 - .238) = 7$	$(.309 - .18) = 13$	$(.309 - .14) = 17$	
H	PUNTOS LOGIT	0	10	21	29	

Tabla 5: Tabulación Cruzada por Edad

## Factores que obtienen el mayor número de puntos en los scorecards de crédito



Cuanto más grandes sean las diferencias en tasas negativas entre los grupos, más puntos recibe un factor de riesgo en un scorecard. Utilizando el método simple de 'diferencias de tasas negativas' (descrito anteriormente), podemos ver en la Tabla 6 a continuación, que 'score crediticio de buró' toma un máximo de 39 puntos, mientras que el 'estado civil' toma un máximo de sólo ocho puntos. Esto se debe a que hay diferencias mucho mayores en las tasas negativas más altas y más bajas para el historial de crédito que las que hay para el estado civil.

Puntajes de Buró de Crédito					
Grupo	< 590 puntos	590 - 670 puntos	671 - 720 puntos	> 720 puntos	Tasa Negativa de Muestra
Tasa Negativa	39%	23%	13%	0%	20%
PUNTOS	0	16	26	39	
Estado Civil					
Grupo	Divorciado	Soltero	Casado	Viudo	Tasa Negativa de Muestra
Tasa Negativa	25%	24%	19%	17%	20%
PUNTOS	0	1	6	8	

**Tabla 6: Ejemplos de Importancia del Factor del Scorecard**

Dado que la clasificación del riesgo entre algoritmos es a menudo muy similar, muchos expertos prefieren utilizar métodos más sencillos en la práctica. El importante autor de scoring crediticio, David Hand, ha señalado que "los métodos sencillos típicamente producen un desempeño casi tan bueno como los métodos más sofisticados, en la medida en que la diferencia en el desempeño puede verse inundada por otras fuentes de incertidumbre que generalmente no se consideran".<sup>30</sup> La práctica generalizada y establecida de utilizar la regresión logística para el scoring crediticio habla de la facilidad con la que tales modelos se presentan como scorecards. Estos scorecards son bien comprendidos por la gerencia y pueden ser usados para manejar proactivamente los riesgos y las recompensas de la actividad crediticia.

<sup>30</sup> David Hand, 'Classifier technology and the illusion of progress', *Statistical Science*, Vol. 21.1 (2006): 1-14

### Scorecards de Expertos



Cuando no hay datos históricos, pero el proveedor tiene una buena comprensión de las características del prestatario que afectan al riesgo en el segmento, un cuadro de scorecard puede funcionar razonablemente bien haciendo clasificación en función del riesgo a los prestatarios.

Un *scorecard de expertos* utiliza puntos para clasificar a los prestatarios por riesgo, tal como lo hace un scorecard estadístico. La principal diferencia (y una importante) es que, sin datos pasados, incluyendo datos sobre morosidad, una IF no tiene manera de saber con certeza si su comprensión (o expectativa) de las relaciones de riesgo es correcta.

Por ejemplo, si sabemos que la edad es un factor de riesgo relevante para los préstamos de consumo y hemos visto (en la práctica) que el riesgo generalmente disminuye con la edad, podríamos crear grupos de edad similares a los de la Tabla 5. En este escenario, asignamos puntos usando un esquema sencillo donde el grupo percibido como más riesgoso siempre obtiene cero puntos y el grupo de menor riesgo siempre obtiene 20 puntos. En este caso, una ponderación de scorecard de expertos de la variable 'edad' podría parecerse a la Tabla 7. Estos puntos no son tan diferentes de los puntos estadísticos de edad mostrados en las filas F y H de la Tabla 5.

	23 años o menos	24 a 30 años	31 a 47 años	48 años o más
PUNTOS	0	7	15	20

*Tabla 7: Puntos 'Expertos' para 'Edad'*

Siempre y cuando la clasificación del riesgo sea correcta para cada factor de riesgo individual en un scorecard de expertos, la puntuación de un scorecard de expertos hará una clasificación del riesgo de los prestatarios de manera similar a cómo los clasifica un scorecard estadístico.<sup>31</sup> Esto significa que los scorecards de expertos pueden ser una herramienta útil para lanzar un nuevo producto para el cuál no hay datos históricos. También son una buena manera para los proveedores de Servicios Financieros Digitales que tienen la intención de ser impulsados por datos, para obtener algunos beneficios de scoring - incluyendo mayor eficiencia y consistencia - al tiempo que construye una mejor base de datos.

<sup>31</sup> Por lo general, utilizando sólo el juicio de expertos, los proveedores especifican incorrectamente la relación de clasificación de riesgo de uno o más factores. Una vez que se recopila la data de desempeño (pago del préstamo), se puede utilizar para corregir cualquier relación mal especificada, lo que conducirá a una mejor clasificación de riesgo del modelo estadístico resultante.

## Elección de un Conjunto de Factores de Riesgo

Si bien los campos de datos específicos disponibles para el scoring crediticio varían mucho según el producto, el segmento y el proveedor, generalmente los datos del modelo de puntuación deben ser:

- Muy relevantes
- Fácil de recopilar de manera consistente
- Objetivos, no auto informados

Algunos tipos de datos tienden a ser buenos predictores del pago de créditos entre segmentos y mercados. La Tabla 8 presenta algunos de ellos junto con sus patrones de riesgo comúnmente observados.

El 'mejor' conjunto de predictores de variable única se combinan en un modelo de variables múltiples. Si bien esto puede hacerse algorítmicamente para maximizar la predicción, un enfoque atractivo para los proveedores de Servicios Financieros Digitales es elegir un conjunto de factores que juntos crean un perfil de riesgo integral para el prestatario,<sup>32</sup> siguiendo las líneas de las conocidas cinco Cs del crédito: *capacidad, capital, colateral, condiciones, y carácter*. Este modelo es fácil de entender para los banqueros y la administración de los bancos, y es consistente con los marcos de gestión de riesgo como los Acuerdos de Capital de Basilea.

A medida que cada predictor individualmente fuerte se añade a un modelo de múltiples factores, su clasificación de riesgo mejora. Sin embargo, después de un número relativamente pequeño de buenos predictores individuales (normalmente de 10 a 20), la mejora incremental para cada factor adicional disminuye bastante bruscamente. Incluso si selectivamente seleccionamos los factores que no parecen estar altamente correlacionados entre sí, en realidad, muchos de los factores estarán correlacionados hasta cierto punto, llevando a los retornos decrecientes de factores adicionales.

Tipo de Datos	Factor	Relación de Riesgo
Comportamiento	Compras	El riesgo disminuye a medida que aumenta el ingreso disponible
	Depósitos y movimiento de la cuenta	El riesgo disminuye a medida que aumentan depósitos y movimientos
	Historial de Crédito	El riesgo disminuye a medida que aumenta el historial crediticio positivo
	Pago de Cuentas	El riesgo disminuye en línea con la puntualidad de los pagos de facturas
Historial	Tiempo en residencia, trabajo, negocios	La estabilidad reduce el riesgo
	Tiempo como cliente	Los clientes con un historial más largo son de menor riesgo
Demográfico	Años	El riesgo disminuye con la edad y aumenta de nuevo alrededor de la edad de jubilación (principalmente debido a los riesgos de salud)
	Estado Civil	Las personas casadas frecuentemente están asentadas y estables, lo que reduce el riesgo
	Número de dependientes	El aumento del número de dependientes puede aumentar el riesgo (especialmente para las personas solteras), pero en algunas culturas en lugar de ello reduce el riesgo (mayor red de apoyo)
	Propietario de vivienda	Los propietarios son menos riesgosos que los inquilinos

Tabla 8: Datos Frecuentemente Eficaces para Scoring Crediticio

<sup>32</sup> Siddiqi, 'Credit risk scorecards: developing and implementing intelligent credit scoring', John Wiley y Sons, Vol. 3 (2012)

## 1.2\_APLICACIONES DE DATOS

Cuando una IF tiene suficientes datos, debe dar preferencia a los datos que:

- Son objetivos y se pueden observar directamente, en lugar de ser obtenidos del solicitante
- Evidencian relaciones con el riesgo de crédito que confirman el juicio experto o intuitivo
- Cuesta menos recolectarlos
- Pueden recopilarse de la mayoría, si no todos, los solicitantes
- No discrimina con base en factores que el prestatario no puede controlar (es decir, edad, sexo, raza) o que son potencialmente divisivos (es decir, religión, etnia, idioma)

### Caso de Uso: Nano-Créditos

Debido a que los bancos deben reportar los pagos de nano-créditos a burós de crédito y bancos centrales, el nano-crédito ha traído a millones de personas que antes carecían de acceso a los bancos, al sector financiero formal en todo el mundo, estableciendo un historial de crédito que es un paso para desbloquear el acceso a otro tipo de productos crediticios. Sin embargo, a algunas personas les preocupa que los nano-créditos creen un ciclo de deuda para individuos de bajos ingresos. Varios millones de personas con malas experiencias con nano-créditos podrían terminar en listas negras en los burós de crédito locales, lo cual avala aún más la necesidad de protección del consumidor.

En esta sección se examina cómo se utilizan los datos para superar algunos de los desafíos que han sido durante mucho tiempo barreras a la inclusión financiera. En particular, son los datos digitales generados por los teléfonos móviles, el dinero electrónico e Internet, los que están ayudando a poner a millones de personas que nunca han tenido cuentas bancarias o préstamos bancarios en el radar de las Instituciones Financieras formales.

Los estudios de caso que siguen investigan cómo se han utilizado los operadores telefónicos, las redes sociales y la datos bancarios tradicionales para lanzar nuevos productos, para ayudar a que más prestatarios se hagan elegibles para créditos formales y para evaluar pequeñas empresas, que son menos homogéneas que los consumidores individuales.

### Desafío de Crédito 1: Verificación de Ingresos y Gastos

Un importante desafío de actividad crediticia a nivel de consumidor en los mercados en desarrollo, es la obtención de datos fiables sobre el flujo de efectivo de nuevos clientes, tanto para las personas como para las empresas. El flujo de caja, o los ingresos que quedan después de gastos, es la principal fuente de pago del crédito y por lo tanto un foco de los modelos de créditos minoristas. Los niveles de ingresos también se utilizan para determinar la capacidad de endeudamiento de un individuo.

El crecimiento de la telefonía móvil y el uso del dinero electrónico - particularmente en África y Asia - ha creado registros digitales de terceros verificables sobre los patrones de pago reales, tales como recargas y pagos de dinero electrónico. Estos datos, en poder de los operadores telefónicos, proporcionan un cuadro de los flujos de efectivo de un usuario de SIM. Los terminales POS y las cajas de dinero electrónico también pueden pintar un cuadro algo más completo de los flujos de efectivo para los comerciantes.



Cuando se sabe cuánto dinero maneja una persona o compañía de forma diaria, semanal y mensual, se puede calcular mejor qué tamaño de créditos serían capaces de pagar.

Los siguientes dos casos analizan cómo los datos digitales han ayudado a abrir grandes mercados para nano-créditos de consumo.

# CASO 10

## M-Shwari Lanza un Mercado Para Nano Créditos

### Soluciones de Datos para Evaluar la Solvencia de los Prestatarios sin Historial de Crédito Formal

El Banco Comercial de África (CBA) y el operador móvil Safaricom comenzaron a reconocer el poder de los datos de telefonía móvil y dinero electrónico.

M-Shwari, el primer producto de ahorro y crédito digital de gran éxito, es bien conocido por los seguidores de 'fintech' y la inclusión financiera. Ha otorgado pequeños créditos llamados nano-créditos por medio de teléfonos móviles a millones de prestatarios, trayéndolos al sector financiero formal. Productos similares han sido lanzados desde entonces en otras partes de África, y la nueva competencia ha atestado el mercado en Kenia. La historia de M-Shwari es también un excelente estudio en el uso creativo de datos para traer un nuevo producto al mercado.

#### Modelando lo Desconocido

La tecnología de scoring crediticio analiza las características pasadas de los prestatarios y el comportamiento

de pagos para predecir los pagos futuros de créditos. ¿Qué pasa en el caso de que no haya un comportamiento de pagos pasado? Los operadores móviles tienen datos amplios sobre los teléfonos móviles de sus clientes y, en muchos casos, sobre el uso del dinero electrónico, pero es menos claro cómo se pueden usar esos datos para predecir la capacidad y la voluntad de pago de un crédito, sin datos sobre el pago de las obligaciones anteriores.

Por definición, no hay datos específicos pasados del producto, para un producto nuevo. Una forma para seguir utilizando el score crediticio con un producto nuevo es utilizando el juicio de expertos y el conocimiento del dominio para crear un 'scorecard de expertos', una herramienta que orienta las decisiones crediticias basadas en las clasificaciones de riesgo del prestatario. Ver cuadro en la página 84.

Otra forma de utilizar el scoring crediticio con un producto nuevo es estudiar un conjunto relevante de datos de clientes, como los datos de los operadores móviles, en relación con la información de pagos de préstamos, tales como:

- **Historial Crediticio General o Informe de Buró:** Esto sólo funciona para los clientes con un expediente en el buró.
- **Productos de Crédito Similares:** Otro producto de crédito lo suficientemente similar como para ser relevante para el nuevo producto se puede utilizar como indicador. Si bien el pago pasado de dicho producto puede ser o no representativo del pago futuro del nuevo producto, puede ser una aproximación aceptable, o 'variable sustitutiva', a efectos del modelado inicial.

## 1.2\_APLICACIONES DE DATOS

El primer scorecard de M-Shwari fue desarrollado usando datos de Safaricom y el historial de pagos de clientes que habían utilizado su producto Okoa Jahzi de crédito para comprar tiempo aire.<sup>33</sup> Los dos productos eran claramente diferentes, como se muestra en la Tabla 9 a continuación.

El producto M-Shwari ofrece a los prestatarios más dinero, flexibilidad de uso y tiempo de pago. La presunción era que los que habían utilizado exitosamente los créditos

muy pequeños de Okoa Jahzi serían mejores riesgos para el producto crediticio más grande.

El primer modelo de scoring crediticio de M-Shwari desarrollado con los datos de Okoa Jahazi,<sup>34</sup> junto con políticas conservadoras de cupos y procesos empresariales bien diseñados, permitieron el lanzamiento del producto, que rápidamente se convirtió en un éxito masivo.

El CBA esperaba que el scorecard basado en los datos de Okoa Jahazi

se volvieran a desarrollar lo antes posible utilizando el comportamiento de pagos del producto mismo de M-Shwari. Algunos comportamientos predictivos del uso del crédito para tiempo aire no se tradujeron directamente en el uso de M-Shwari y los cambios apropiados al modelo basados en los datos reales de uso de M-Shwari redujeron los préstamos en mora en un 2 por ciento. M-Shwari sigue actualizando su scorecard periódicamente con base en información nueva.

Producto	Okao Jahzi	M-Shwari
<b>Monto</b>	El menor gasto de tiempo aire entre los últimos 7 días o 100 chelines kenianos	100 a 10.000 chelines kenianos
<b>Propósito</b>	Utilizado sólo para tiempo aire	Utilizado para cualquier propósito
<b>Términos de pago</b>	72 horas	30 días

Tabla 9: Comparación de productos Okao Jahzi y M-Shwari



El lanzamiento y desarrollo exitoso de M-Shwari demuestran que hay formas de utilizar soluciones de scoring basadas en datos para segmentos completamente nuevos. También refuerza la verdad general acerca del scoring crediticio en el sentido de que un scorecard siempre es un trabajo en proceso. No importa lo bien que se desempeñe un scorecard con datos de desarrollo, siempre se debe monitorear y administrar usando informes estándar y se le debe afinar siempre que haya cambios significativos en los riesgos de mercado o en los tipos de clientes que solicitan el producto.

<sup>33</sup> Cook y McKay, 'How M-Shwari works: The story so far', Consultative Group to Assist the Poor y Financial Sector Deepening

<sup>34</sup> Mathias, 'What You Might Not Know', Abacus, 18 de Septiembre de 2012, consultado Abril 3 de 2017, <https://abacus.co.ke/okoa-jahazi-what-you-might-not-know/>

El producto de nano-créditos M-Shwari tuvo éxito gracias a la oportuna confluencia de:

- **Acceso a Datos del MNO:** CBA tuvo una ventaja de pionero debido a su fuerte alianza con Safaricom. Hoy, Safaricom vende sus datos telefónicos a todos los bancos en Kenia.
- **Un Producto Bien Diseñado:** Los productos pequeños, a corto plazo, se ajustan mejor al scoring crediticio, especialmente para los productos nuevos. La rápida retroalimentación sobre el desempeño de pagos de la población objetivo permite un desarrollo oportuno del modelo y controla el riesgo.
- **Buenos Sistemas y Personas:** El equipo directivo de M-Shwari es pequeño y flexible, reuniendo una combinación única de habilidades administrativas y técnicas, así como los sistemas para garantizar una implementación sin problemas.
- **Aprovechamiento de Recursos Externos:** FSD Kenia apoyó a CBA con conocimiento experto en modelado de riesgo, que era crucial para desarrollar el primer modelo de scoring y transferir destrezas al equipo de M-Shwari.

Aunque la historia de éxito de M-Shwari es inspiradora, hay muchos proveedores de Servicios Financieros Digitales que quieren entrar en el espacio de los nano-créditos, pero pueden encontrarlo difícil. Estos proveedores de SFD pueden no tener relaciones con MNO o pueden carecer de la capacidad interna para diseñar productos de ahorro y crédito digitales y modelos de scoring. El siguiente caso describe cómo los proveedores están facilitando la entrada de proveedores de SFD en el mercado masivo de nano-créditos.



# CASO 11

## Tiaxa, un Enfoque Completo de Nano-Crédito

### Desarrollo de Productos y Servicios de Datos a Través de Servicios de Suscripción Externalizados

Reconociendo que muchas Instituciones Financieras en los mercados en desarrollo carecen de los recursos para abordar el mercado de Servicios Financieros Digitales solo utilizando recursos internos, Tiaxa ofrece sus NanoCredits™ patentados dentro de una solución completa que incluye:

- Diseño de producto
- Adquisición de clientes (basada en modelos de scoring patentados)
- Gestión del riesgo Crediticio de la Cartera
- Despliegue de hardware y software
- Gestión de servicios las veinticuatro horas del día
- Línea de financiación para cartera (en algunos mercados africanos)

Tiaxa reúne Instituciones Financieras y MNOs y forma alianzas tripartitas en las que:

- MNOs suministran los datos que nutren sus modelos de decisión de crédito
- IF proporcionan las licencias de crédito necesarias (y la reglamentación formal del sector financiero) y la financiación
- Tiaxa proporciona la solución de producto de nano-crédito de extremo a extremo

Además de proporcionar los modelos de diseño y scoring de nano-créditos basados en datos de MNO, en la mayoría de los casos, Tiaxa asume y gestiona el riesgo de crédito de la cartera. El riesgo de pérdida se

gestiona mediante el débito directo de las cuentas MNO del prestatario para evitar la morosidad, que se acuerda con los prestatarios en los términos y condiciones del producto. Su modelo de negocio de alianzas a largo plazo funciona en términos que varían de la participación en los beneficios a los modelos de pago por transacción.

### Datos que Impulsan los Modelos de scoring de Tiaxa

Aunque los conjuntos de datos MNO varían según los países y los mercados, los conjuntos de datos que informan a los modelos patentados de Tiaxa típicamente incluyen una combinación de los siguientes tipos de datos:

Uso GSM	Nómina, pagos regulares	Transferencias de dinero	Información de KYC	Pagos de servicios públicos	Depósitos
<ul style="list-style-type: none"> <li>• Frecuencia de recarga, montos</li> <li>• Información de consumo de GSM</li> </ul>	<ul style="list-style-type: none"> <li>• Nómina, subsidios</li> <li>• Flujo de caja, necesidades de crédito</li> </ul>	<ul style="list-style-type: none"> <li>• Frecuencia y valor</li> <li>• ¿Recepción o envío?</li> </ul>	<ul style="list-style-type: none"> <li>• Nombre completo</li> <li>• Tipo de cuenta</li> <li>• Registrar fecha</li> <li>• Estado de KYC</li> <li>• Fecha de nacimiento región</li> </ul>	<ul style="list-style-type: none"> <li>• Indicador de flujo de caja</li> <li>• Sofisticación financiera</li> </ul>	<ul style="list-style-type: none"> <li>• Información de flujo de caja</li> </ul>

**Tabla 10:** Tipos de Datos que Informan a los Modelos Patentados de Tiaxa

*Tiaxa utiliza una gama de métodos de aprendizaje automático para reducir cientos de posibles predictores a un modelo óptimo. Se diseñan modelos a medida para cada consultoría.*

*Tiaxa cuenta ahora con más de 60 instalaciones, con 28 clientes, en 20 países, en 11 grupos MNO, con más de 1.5 billones de usuarios finales entre ellos. En la actualidad, la compañía*

*procesa más de 12 millones de nano-créditos por día en todo el mundo, principalmente en forma de créditos de tiempo aire.*



*A medida que evoluciona el panorama de la analítica de datos, se espera que los proveedores externos desarrollen soluciones completas que se conecten a fuentes de datos internas y entreguen valor a los productos existentes. Las empresas que no pueden invertir en analítica de datos a la medida o que prefieren un enfoque de ‘esperar y ver’ pueden aprovechar los servicios de suscripción en el futuro enviando datos a proveedores externos.*

## 1.2\_APLICACIONES DE DATOS

Para las Instituciones Financieras, la elección entre trabajar con proveedores o trabajar directamente con operadores telefónicos para llegar al segmento de nano-créditos sólo puede hacerse considerando las condiciones del mercado y los recursos disponibles. A continuación, se presentan algunos de los pros y contras de cada enfoque.

### Caso de Uso: Datos Alternativos

Las fuentes alternativas de datos se muestran prometedoras a la hora de verificar la identidad y hacer una evaluación básica del riesgo. Otra forma en que los proveedores de SFD recogen

datos de los nuevos solicitantes es pidiéndoles directamente que proporcionen información. Estas solicitudes pueden tomar la forma de:

- Formularios de Solicitud
- Encuestas
- 'Permisos' para Acceder a Datos del Dispositivo: Esto puede incluir permisos para acceder a contenido multimedia, registros de llamadas, contactos, comunicaciones personales, información de ubicación o perfiles de redes sociales en línea

Estas fuentes de datos en línea no tradicionales pueden y están siendo utilizadas para ofrecer servicios de verificación de identidad y score crediticio. La historia de la empresa de análisis de redes sociales Lenddo proporciona más información y algunos conocimientos sobre cómo los datos de redes sociales pueden agregar valor en el proceso de crédito.

Enfoque	Oportunidades	Desafíos
Trabajar con Datos de MNOs	<ul style="list-style-type: none"> <li>• Control total de los productos</li> <li>• Potencialmente más rentable</li> </ul>	<p>Necesidad de habilidades internas en:</p> <ul style="list-style-type: none"> <li>• Desarrollo de productos</li> <li>• Modelado de riesgo</li> </ul> <p>Necesita sistemas y software para administrar productos de SFD</p>
Trabajar con el Proveedor Externo	<ul style="list-style-type: none"> <li>• Proporciona conocimientos sobre productos, modelado y sistemas</li> <li>• Toma decisiones crediticias</li> <li>• Soluciones de software listas</li> </ul>	<ul style="list-style-type: none"> <li>• Dependencia en proveedor externo</li> <li>• Los detalles del modelo no se pueden compartir</li> <li>• Habilidades técnicas no transferidas</li> </ul>

*Tabla 11: Trabajar con MNOs o Proveedores Externos: Oportunidades y Desafíos*

## CASO 12

# Lenddo hace Minería de Datos de Redes Sociales para la Verificación de Identidad y Perfil de Riesgo

### Uso de Técnicas Analíticas Avanzadas y Fuentes Alternativas de Datos para Productos Nuevos

Los cofundadores de Lenddo, Jeffrey Stewart y Richard Eldridge, concibieron inicialmente la idea mientras trabajaban en la industria de tercerización de procesos de negocios en Filipinas en 2010. Ellos se sorprendieron por el número de sus empleados que habitualmente les pedían adelantos salariales y se preguntaban por qué estas personas jóvenes y brillantes con empleos estables no podían obtener créditos de las instituciones financieras formales.

El desafío particular en Filipinas era que el país no tenía burós de crédito ni números de identificación nacionales. Si la gente no usaba cuentas o

servicios bancarios - y menos del 10 por ciento lo hacían - ellos eran 'invisibles' para las IF formales y no podían obtener crédito. Al desarrollar su idea, los fundadores de Lenddo muy pronto reconocieron que sus empleados eran usuarios activos de la tecnología y estaban presentes en las redes sociales. Estas plataformas generan grandes cantidades de datos y esperaban que el análisis estadístico de estos pudiera ayudar a predecir la capacidad crediticia de un individuo.

Los solicitantes de créditos de Lenddo dan permiso para acceder a los datos almacenados en sus teléfonos móviles. Los datos en bruto del solicitante se visualizan,

extraen y califican, pero luego son destruidos (en lugar de almacenados) por Lenddo. Para un solicitante típico, su teléfono contiene miles de puntos de datos que hablan del comportamiento personal:

- Tres Grados de Conexiones Sociales
- Actividad (fotos y vídeos publicados)
- Membresías en Grupos
- Intereses y Comunicaciones (mensajes, correos electrónicos y tweets)

Más de 50 elementos en todos los perfiles de redes sociales proporcionan 12.000 puntos de datos por usuario promedio:

Entre Todas las Cinco Redes Sociales:	7900+ Comunicaciones de Mensajes Totales:
<ul style="list-style-type: none"><li>• 250+ conexiones de primer grado</li><li>• 800+ conexiones de segundo grado</li><li>• 2.700+ conexiones de tercer grado</li><li>• 372 fotos, 18 vídeos, 13 grupos, 27 intereses, 88 enlaces, 18 tweets</li></ul>	<ul style="list-style-type: none"><li>• 250+ conexiones de primer grado</li><li>• 5,200+ mensajes de Facebook, 1,100+ Facebook me gusta</li><li>• 400+ actualizaciones de estado de Facebook, 600+ comentarios de Facebook</li><li>• 250+ correos electrónicos</li></ul>

Tabla 12: Promedios de Puntos de Datos de Redes Sociales por Usuario Promedio

## 1.2\_APLICACIONES DE DATOS

### Uso de Datos

La confirmación de la identidad de un prestatario es un componente importante para conceder el crédito a los solicitantes sin antecedentes crediticios anteriores. La aplicación en formato de tableta de Lenddo pide a los solicitantes de crédito que completen un formulario digital corto pidiendo su nombre, fecha de nacimiento, número de contacto principal, dirección de correo electrónico principal, colegio y empleador. Luego, se pide a los solicitantes ingresar a Lenddo entrando y concediendo permisos a Facebook. Los modelos de Lenddo usan esta información para verificar la identidad del cliente en menos de 15 segundos. La verificación de identidad puede reducir significativamente el riesgo de fraude, que es mucho mayor para

los productos crediticios digitales, donde no hay contacto personal durante el proceso de suscripción. A continuación, se presenta un ejemplo del trabajo de Lenddo con el MNO más grande de Filipinas.

Lenddo trabajó con un gran MNO para aumentar la cuota de planes postpago que podría ofrecer a sus 40 millones de suscriptores de prepago (90 por ciento del total de suscriptores). La elegibilidad para planes postpago dependía de una verificación de identidad exitosa, y el proceso de verificación existente de la Telco requería que los clientes visitaran las tiendas y presentaran sus tarjetas de identificación, las cuales luego eran escaneadas y enviadas a una oficina central para su verificación. El tiempo promedio para completar el proceso de verificación fue de 11 días.

La plataforma SNA de Lenddo se utilizó para proporcionar verificación de identidad en tiempo real en segundos basándose en el nombre, a fecha de nacimiento, y el empleador. Esto mejoró la experiencia del cliente, redujo los posibles fraudes y errores causados por la intervención humana y redujo el costo total del proceso de verificación.

Además de sus modelos de verificación de identificación, Lenddo utiliza una serie de técnicas de aprendizaje automático para mapear las redes sociales y agrupar a los solicitantes en términos de patrones de comportamiento (uso). El resultado final es un LenddoScore™ que puede ser utilizado inmediatamente por las Instituciones Financieras para filtrar a los solicitantes o para alimentar y complementar los scorecards propios de una IF.



Estos algoritmos convierten un número inicialmente grande de puntos de datos sin procesar por cliente, en un número manejable de características y comportamientos del prestatario con relaciones conocidas respecto del pago de créditos.

## Caso de Uso: Scoring Crediticio para Pequeñas Empresas

Los ejemplos mencionados hasta ahora se han centrado en los productos digitales dirigidos a consumidores de mercado masivo y comerciantes. El flujo de datos de comportamiento creado en los canales digitales ha generado comprensiblemente la mayor expectativa sobre las oportunidades de la analítica de datos. Sin embargo, la mayoría de las Instituciones Financieras también tienen muchas oportunidades para hacer un mejor uso de los datos en el análisis de crédito y la gestión de riesgo de los productos tradicionales y fuera de línea que incluyen, pero no se limitan a:

- Créditos de Consumo
- Tarjetas de Crédito
- Préstamos y Arrendamientos de Micro, Pequeña y Mediana Empresa (MiPyMe)
- Pequeños Préstamos y Arrendamientos Agrícolas
- Cadena de Valor y Financiación de Cadena de Suministros.

Para estos productos, las IF han recogido tradicionalmente una gran cantidad

de datos, pero no necesariamente han digitalizado o sistematizado su captura, análisis y almacenamiento. En los mejores casos, el software LOS facilita la captura digital de datos tradicionales de una manera conducente a la analítica de datos, incluyendo el desarrollo de scorecards de crédito. A medida que la cadena de valor y los pagos en cadena de suministro se digitalizan, existe la oportunidad de aprovechar estos datos para proyectar flujos de efectivo y generar puntuaciones de crédito.

### Metodologías de Score Crediticio

Las IF tienen varias opciones para usar los datos que ya recogen para el modelado del riesgo de crédito. Tres de las soluciones más comunes son el desarrollo de scorecards de crédito propios, ya sea a través del conocimiento experto interno, mediante el trabajo con consultores externos, o mediante la tercerización del scoring crediticio a un proveedor externo.

### Desarrollar Scorecards de Crédito Propios

Los bancos de los principales mercados financieros (por ejemplo, Sudáfrica, América del Norte, Europa continental y Singapur) emplean grandes equipos que desarrollan

y mantienen modelos, incluidos modelos separados para el apoyo a las decisiones de aplicación, la gestión permanente de cartera (comportamiento) y provisiones. Como primer paso al desarrollo de modelos internos, las IF pueden optar por utilizar consultores externos para hacer los desarrollos iniciales y para crear capacidad con el personal interno para llevarlo adelante.

Muchos proveedores de Servicios Financieros Digitales tienen datos, analistas de datos y especialistas en TI internos capaces de administrar sus propios sistemas de scoring. Lo que estos equipos tienden a carecer es experiencia en el desarrollo de scorecards de crédito. Los buenos proyectos de analítica de datos requieren conocimientos expertos para tener éxito. La asistencia externa puede ayudar a la transferencia de conocimientos a construir la experiencia interna como parte del apoyo al proyecto. Cuando se trabaja con consultores externos, los proveedores de SFD deben asegurarse que las herramientas y habilidades necesarias se transfieran a los equipos internos para que los scorecards puedan ser administrados y monitoreados en el futuro.

### Una Mirada más Cercana a los Scorecards Propios



Un reciente proyecto de IFC con un banco en Asia ejemplifica cómo el proceso puede funcionar:

1. El banco compartió sus datos de cartera anteriores con el consultor.
2. El consultor preparó los datos para el análisis utilizando el software estadístico de código abierto 'R'.
3. El banco convocó a un grupo de trabajo de scoring crediticio para trabajar con el consultor. En un ambiente de taller, el consultor y el grupo de trabajo analizaron y seleccionaron los factores de riesgo para scorecards de créditos de consumo y de micro-empresas.
4. El banco vinculó a un nuevo analista para que asumiera la responsabilidad principal de los scorecards (y el analista también participó en los talleres 'R').
5. El grupo de trabajo de scoring crediticio y el consultor revisaron las fortalezas y debilidades de los modelos resultantes para alinear las estrategias de uso con los objetivos comerciales del banco y el apetito por el riesgo.
6. Con la orientación inicial del consultor, el banco y su proveedor de software local desarrollaron una plataforma de software para desplegar el scorecard.
7. El consultor proporcionó soporte remoto en el monitoreo y administración del scorecard.

Los pros y los contras de tales arreglos incluyen:

<b>Pros:</b>	<b>Contras:</b>
<ul style="list-style-type: none"><li>• El Banco aprende las habilidades necesarias para apropiarse de los modelos</li><li>• El Banco tiene control completo sobre sus scorecards</li><li>• Los scorecards son totalmente transparentes</li></ul>	<ul style="list-style-type: none"><li>• Requiere la vinculación activa de los gerentes senior y junior</li><li>• Requiere capacitación del personal o la integración de analítica de datos y especialistas en modelado de riesgo.</li><li>• Requiere software de despliegue adicional, tal como un LOS con funcionalidad de scoring</li><li>• El desarrollo interno trae consigo requisitos de mantenimiento a largo plazo</li></ul>

*Tabla 13: Pros y Contras de los Scorecards Propios*

### Subcontratar el Scoring Crediticio a un Proveedor Externo

La mayoría de los proveedores externos ofrecen el desarrollo de modelos personalizados utilizando los datos de buró (cuando están disponibles), los datos propios del banco, así como datos de terceros como los datos de CDR. Los proveedores normalmente también proporcionan software de despliegue del scorecard y mantienen los modelos para el FI. Trabajar con proveedores de score crediticio externaliza el conocimiento experto de scoring y las plataformas de software, a menudo aportando nuevos datos que de otro modo serían inalcanzables. También aporta experiencia internacional y credibilidad inmediata a la solución de scoring.

A continuación, se muestra un ejemplo del trabajo de First Access con un banco en África Oriental en el segmento de créditos para pequeñas empresas, un segmento para el cuál los datos de MNO solos no bastan para evaluar de forma exhaustiva el riesgo crediticio del solicitante.

## CASO 13

# First Access: Scoring crediticio con Proveedor Externo de Servicio Integral

### Tercerización de Conocimiento en Datos y Trabajo con Aliados Externos

Muchas Instituciones Financieras están interesadas en utilizar el scoring crediticio para aumentar la consistencia y eficiencia de la evaluación de crédito para pequeños créditos. Sin embargo, un número menor de IF en los mercados en desarrollo tiene la capacidad interna de desarrollar e implementar scorecards de manera eficiente sin ayuda alguna externa.

Como se mencionó anteriormente, el trabajo con proveedores externos de scoring crediticio subcontrata la experiencia de scoring y las plataformas de software, y a menudo aporta experiencia internacional y credibilidad inmediata a la solución de scoring.

First Access es uno de los muchos proveedores de scoring de crédito, pero uno de los relativamente pocos que se centra en los desafíos particulares que enfrentan los mercados emergentes. Fundada en julio de 2012, la

compañía inicialmente trabajó extensamente con Vodacom Tanzania, aprovechando sus datos MNO para desarrollar una herramienta de auto-decisión para proveedores de Servicios Financieros Digitales que atiende a clientes de bajos ingresos sin historial de crédito formal. Desde entonces, ha ampliado su presencia a la RDC, Malawi, Nigeria, Uganda y Zambia, trabajando más ampliamente en las soluciones de scoring para el segmento de micro y pequeñas empresas.

First Access trabajó con un banco en el este de África para desarrollar un scorecard para sus (micro) créditos a pequeñas empresas, centrado en préstamos de hasta \$3,000. El banco tomó un promedio de seis días para evaluar las solicitudes de préstamos, y además de los largos tiempos de espera, sus NPL habían estado aumentando. Al igual que muchos bancos en los mercados emergentes, no tenía herramientas para seleccionar o puntuar a los clientes, y por lo tanto

utilizó un sólo proceso para todos los solicitantes que entraran por la puerta.

First Access estudió los datos históricos de la cartera del banco para el segmento y construyó un algoritmo de scoring utilizando sólo la información disponible en el momento de cada solicitud de crédito, sin incluir datos adicionales normalmente recopilados en visitas dispendiosas a la sede de la empresa del solicitante, una característica típica de un proceso de suscripción de microcrédito. A pedido del banco, el modelo clasificó a los solicitantes en cinco segmentos de riesgo.

Una ‘prueba ciega’ de todos los microcréditos vencidos, desembolsados durante los seis meses anteriores, indicó que las puntuaciones clasificaron a los prestatarios por riesgo, como lo indican las tasas negativas en la Tabla 14 a continuación.

Segmento de Riesgo	A	B	C	D	E
CeR (Cartera en Riesgo)	1.00%	3.53%	9.97%	22.42%	26.78%

Tabla 14: Clasificaciones de Prestatarios de Microcréditos por Riesgo

## 1.2\_APLICACIONES DE DATOS

Utilizando el algoritmo de scoring, cada solicitante podría ser puntuado inmediatamente y asignado a uno de los segmentos de riesgo. El banco ajustó su proceso de evaluación de crédito para ofrecer aprobación el mismo día a sus clientes recurrentes en los segmentos A y B, que representaban el 22 por ciento de los solicitantes de préstamos. El tiempo de aprobación para este grupo de clientes se redujo de un promedio de seis días a un día, lo que mejoró la experiencia del cliente y la eficiencia y satisfacción del personal del banco.

Dado que los resultados del algoritmo en la práctica han validado

la prueba ciega original, el banco está ampliando el uso del algoritmo para llevar a cabo más aprobaciones y rechazos de créditos en el mismo día para clientes repetidos y nuevos. La aceleración de los grupos A y B han aumentado la eficiencia de la institución en la suscripción de microcréditos en un 18 por ciento, y ambos grupos han superado los resultados de las pruebas ciegas, con CeR combinado<sup>1</sup> del 1,26 por ciento en lugar del 3 por ciento esperado.

La plataforma de software First Access permite a las Instituciones Financieras configurar y gestionar

sus propios algoritmos de scoring personalizados y utilizar sus propios datos sobre su base de clientes y producto crediticios. Actualmente, First Access está desarrollando nuevas herramientas en su plataforma para dar a las IF más control y transparencia en la gestión de sus reglas de decisión, cálculos de scoring y umbrales de riesgo, con monitoreo continuo del desempeño del algoritmo. Dichos paneles de control de analíticas de desempeño pueden ayudar a las IF a gestionar mejor el riesgo en respuesta a los cambios en el mercado.

Pros:	Contras:
<ul style="list-style-type: none"><li>• Acceso a habilidades de modelamiento de primer nivel y experiencia internacional</li><li>• Proporcionar software de despliegue</li><li>• Potencialmente acortar el tiempo necesario para desarrollar e implementar el scorecard</li><li>• Gestionar y supervisar el scorecard y el software</li></ul>	<ul style="list-style-type: none"><li>• El banco no es dueño del modelo y por lo general no sabe el cálculo de scoring</li><li>• Los costos fijos del uso del modelo y desarrollo de modelos intermitentes</li></ul>

Tabla 15: Pros y Contras de la Tercerización del Scoring Crediticio a un Proveedor Externo



Un enfoque de Tercerización para el desarrollo de productos de datos proporciona soluciones rápidas y conocimientos especializados, pero también puede traer consigo un riesgo de mantenimiento a más largo plazo, problemas de propiedad intelectual (PI) y un requisito de que los diseños de proyectos se delimiten en detalle para asegurar resultados útiles.

## Accesibilidad y Privacidad

Existen dos retos fundamentales para el uso de nuevas formas de datos digitales: la accesibilidad y la privacidad. Para beneficiarse de nuevas fuentes de datos digitales, los Proveedores de Servicios Financieros deben tener acceso a estos datos en un formato que pueda ser analizado. Dos de las principales formas de acceder a dichos datos es comprando la data o colaborar con el proveedor externo. Algunos MNO, como Safaricom de Kenia, venden campos de datos agregados pre-procesados - tales como el gasto promedio mensual o el uso de llamadas - directamente a los Proveedores de Servicios Financieros. Algunos proveedores externos también procesan grandes conjuntos de datos sin procesar extraídos de los MNO, redes sociales y datos de

los dispositivos, y convierten estos en perfiles de clientes utilizables y vendibles. Las consideraciones de privacidad han limitado la disponibilidad de algunos datos, y no hay garantía de que, por ejemplo, los datos de redes sociales sigan siendo una fuente de datos consultable para los modelos de crédito en el futuro. Facebook ya ha tomado medidas para limitar la cantidad de datos que los servicios de terceros pueden extraer de los perfiles de usuario,<sup>35</sup> y los datos que hace accesibles a través de su API sólo se pueden utilizar legalmente para la verificación de identidad. En los Estados Unidos, la FTC, que monitorea las reglas sobre los datos de crédito y de consumo, ha indicado que las redes sociales corren el riesgo de ser reguladas como agencias de información al consumidor si sus datos son utilizados como criterios de crédito.<sup>36</sup>

---

<sup>35</sup> Seetharaman y Dwoskin, 'Facebook's Restrictions on User Data Cast a Long Shadow', *Wall Street Journal*, Septiembre 21 de 2015

<sup>36</sup> 'Facebook Settles FTC Charges That It Deceived Consumers By Failing To Keep Privacy Promises', *Sitio de Noticias de la Comisión Federal DEL Comercio*, Noviembre 29 de 2011, consultado Abril 3 de 2017, <https://www.ftc.gov/news-events/press-releases/2011/11/facebook-settles-ftc-charges-it-deceived-consumers-failing-keep/>

# PARTE 2

## Marco del Proyecto de Datos

### Capítulo 2.1: Gestión de un Proyecto de Datos



#### El Data Ring

Administrar cualquier proyecto es complejo y requiere los ingredientes adecuados; la intuición empresarial, la experiencia, las habilidades técnicas, el trabajo en equipo y la capacidad para manejar eventos imprevistos, determinarán el éxito. No hay receta para el éxito. Dicho esto, hay maneras de mitigar los riesgos y maximizar los resultados, aprovechando los marcos organizacionales para la planificación y aplicando buenas prácticas establecidas. Esto también es válido para un proyecto de datos. Esta sección presenta los componentes básicos necesarios para planificar un proyecto de datos bien gestionado utilizando un marco visual denominado el *Data Ring*.

Los componentes organizativos del marco se basan en las mejores prácticas de la industria, reconociendo los requisitos generales de recursos y los pasos del proceso que son comunes en la mayoría de los proyectos de datos. Tiene puntos en común con el Cross Industry Standard Process para la Minería de Datos (CRISP-DM), un enfoque de proceso de análisis de datos que cobró importancia después de su lanzamiento en 1996 y fue ampliamente utilizado a principios de los años 2000.<sup>37</sup> Su énfasis en la minería de datos y las herramientas computacionales que prevalecen hace dos décadas ha hecho que el uso del método disminuya considerablemente con el surgimiento del Big Data y técnicas contemporáneas de la ciencia de los datos. El sitio web original de CRISP-DM se desconectó hacia 2014, quedando sin un estándar específico de la industria para los proyectos de datos actuales.

El marco Data Ring aprovecha los conceptos de los métodos establecidos de la industria, con un enfoque modernizado para las tecnologías actuales y las necesidades de los equipos de ciencias de la información. Fue desarrollado por Christian Racca y Leonardo

<sup>37</sup> *Cross Industry Standard Process for Data Mining*. En Wikipedia, La Enciclopedia Libre, consultado el 3 de abril de 2017, [https://en.wikipedia.org/wiki/Cross\\_Industry\\_Standard\\_Process\\_for\\_Data\\_Mining/](https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining/)

Camiciotti<sup>38</sup> como una herramienta de planificación para ayudar a reconocer los elementos centrales del proyecto y reflexionar sobre los requisitos de recursos del proyecto de datos y sus relaciones de forma estructurada. En colaboración con los autores originales y Soren Heitmann, el Data Ring y la herramienta asociada, el Data Ring Canvas, fueron adaptados para este manual. La idea clave es proporcionar una herramienta que apoye a los directores de proyectos durante todo el proceso. A continuación, se muestra una lista de distintas maneras en que debe usarse la herramienta:

- **Lista de control:** Una lista de verificación o “lista de la compra”, mediante la cual se analiza la presencia (y los vacíos relacionados) de los ingredientes necesarios para emprender un proceso basado en datos.
- **Herramienta Descriptiva:** El Data Ring es un potente marco para explicar el proceso basado en datos (puede ser un informe interno, una presentación pública o una publicación científica).
- **Espejo de Retroalimentación Continua:** A partir de la definición de los objetivos y terminando en los resultados, cada ciclo de iteración proporciona retroalimentación para refinar el proceso y reevaluar el diseño.
- **Herramienta Focal:** Para mantener el *foco* del proyecto sobre los objetivos, mientras se hace monitoreo de metas claras.

El enfoque de Data Ring se diseña alrededor de la mitigación del riesgo y de la mejora continua; está diseñado para evitar inicios defectuosos, para asegurar un enfoque orientado a objetivos y evitar los peores escenarios. Puede usarse como una guía continua para definir y refinar metas. Esto ayuda a mantener la fase de ejecución bajo control y ofrece resultados de la mejor manera posible. El proceso de razonamiento es circular, pidiendo a los gerentes que vuelvan a examinar las preguntas centrales de planificación con cada iteración, refinando, ajustando y ejecutando. Cuando surgen problemas, la idea es instar a los gerentes a dar una vuelta completa, considerando cada cuadrante del anillo como una fuente de solución potencial.

El diagrama del Data Ring es bastante complejo, ya que describe el conjunto principal de consideraciones necesarias para planificar un proyecto completo. Los gerentes de proyecto pueden considerar la impresión del diagrama como una referencia visual singular para diseñar un proyecto de datos. En las siguientes secciones, cada una de estas estructuras detalladas se desglosará paso a paso y se discutirán. La sección concluye con un caso de uso paso a paso para ejemplificar cómo el Data Ring puede ser utilizado adicionalmente como una herramienta de planificación.

---

<sup>38</sup> Camiciotti y Christian ‘Creare valore con i BIG DATA’. *Data Ring adaptado para Servicios Financieros Digitales Handbook*, Edizioni LSWR (2015): <http://dataring.eu/>

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

### Estructuras y Diseño

#### Cinco Bloques Estructurales

El Data Ring ilustra el objetivo en el centro, rodeado por cuatro cuadrantes. Cinco Bloques Estructurales: *Objetivo, Herramientas, Habilidades, Proceso, y Valor*. Los cuatro cuadrantes se subdividen en 10 componentes: *Datos, Infraestructura, Informática, Ciencia de Datos, Negocios, Planificación, Ejecución, Interpretación, Ajuste, e Implementación*. Un plan de proyecto debe encaminarse a encapsular estos componentes y a entender en forma profunda sus relaciones interconectadas. El enfoque organizacional del Data Ring ayuda a los gerentes de proyecto a definir recursos y articular estas relaciones; cada componente está provisto de un conjunto de preguntas guía, que están alineadas visualmente en forma perpendicular al componente. Estas preguntas guía sirven como una lista gráfica de planificación de recursos.

#### Objetivo: Bloque Central

Establecer objetivos claros es la base de cada proyecto. Para una solución basada en datos a un problema, sin objetivos cuantitativos y medibles, todo el proceso de análisis de datos está en alto riesgo de fracaso. Esto se traduce en un bajo conocimiento del valor agregado y puede dar lugar a interpretaciones engañosas.

#### Herramientas y Habilidades

Los bloques superiores del Data Ring se centran en evaluar los recursos 'duros' y 'blandos' requeridos para implementar un proyecto de datos:

- **Recursos Duros:** Incluyendo los datos mismos, herramientas de software, procesamiento y hardware de almacenamiento.
- **Recursos Blandos:** Incluye habilidades, conocimiento del sector y recursos humanos para la ejecución.

#### Proceso y Valor

Los bloques inferiores del Data Ring se centran en la implementación y lanzamiento, aunque éstos consisten en tres actividades concretas:

1. Planificación de la ejecución del proyecto
2. Generación y manejo de los datos - la fase de ejecución
3. Interpretación y ajuste de los resultados para implementar la meta del proyecto y extraer valor

#### Diseño Circular

Un elemento central del Data Ring es su diseño circular. Esto enfatiza la idea de mejora continua y optimización iterativa. Estos conceptos son especialmente críticos para los proyectos de datos, formando elementos establecidos de diseño y planificación de proyectos de buenas prácticas. Esto se debe a que el resultado de cualquier proyecto de datos es, simplemente, más datos. Tómese un modelo

de score crediticio, por ejemplo. Los datos numéricos se introducen: edad, ingresos e historial de morosidad, por ejemplo. Los resultados son score crediticio, o más datos numéricos. El proceso consiste en la entrada y salida de datos.

De hecho, este principio de "entran datos- salen datos" es continuamente aplicable en todo el proyecto de datos. Puede aplicarse a todas las exploraciones analíticas intermedias y pruebas de hipótesis, más allá de simples descripciones de las condiciones de inicio y terminación. El proceso circular del Data Ring ilustra de manera similar un enfoque iterativo que pretende refinar, a través de ciclos, la comprensión de los fenómenos desde el punto de vista del análisis de datos. Esto permite una descripción de las causas (entran datos) y efectos (salen datos), y la identificación de los comportamientos y patrones emergentes no obvios. Los cinco bloques organizacionales principales del Data Ring están diseñados para planificar y lograr el equilibrio entre especificidad y flexibilidad a lo largo del ciclo de vida del proyecto de datos.

En términos prácticos, la planeación de proyectos debe considerar el bloque de cada anillo en secuencia, iterando hacia el plan general. El enfoque circular busca establecer las medidas necesarias para lograr un proceso mínimo viable. Es decir, cuando se pueden introducir datos en el sistema, se analizan y se obtienen resultados satisfactorios - y luego se repiten

## El Data Ring

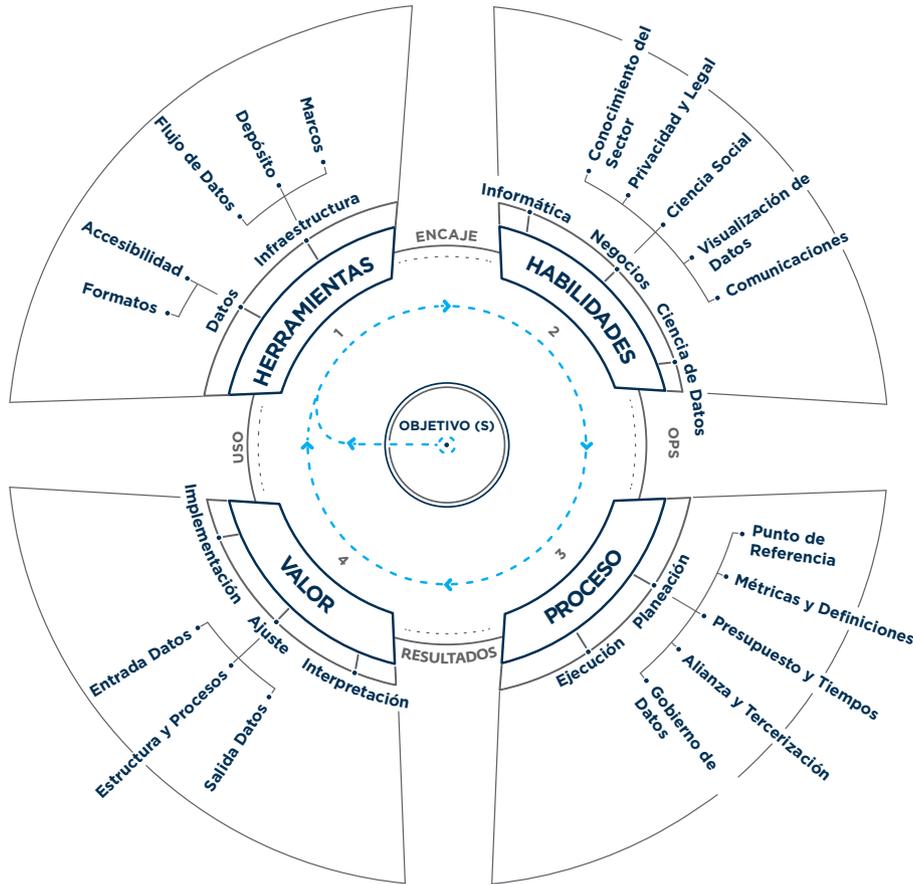


Figura 19: Data Ring, una Herramienta Visual de Planeación para Proyectos de Datos

sin romper el sistema; por ejemplo, con un conjunto de datos actualizado unos meses más tarde que incluye nuevos clientes. Una vez establecido, el proyecto puede iterar al siguiente nivel para entregar un producto mínimo viable (PMV). Este es el producto de datos más básico.

Un *producto de datos* es un modelo, algoritmo o procedimiento que toma datos y alimenta de forma fiable los resultados en el entorno a través de un proceso automatizado. En otras palabras, sus resultados de salida se integran en un contexto operativo más amplio sin cálculo manual. Esto es lo que establece un producto de datos aparte de un análisis singular. Un producto de datos puede ser sencillo - como una visualización interactiva del panel de control - pero también hay productos de datos altamente complejos, donde el score crediticio alimenta procesos semiautomáticos de toma de decisiones de crédito, influyendo en la generación de nuevos clientes con datos que retroalimentan el modelo de score crediticio para guiar las nuevas decisiones crediticias. El hecho que los productos de datos sean consumidores de sus propios resultados afirma su principio circular. El inventario de datos crece con cada iteración. Esto también enfatiza el enfoque organizacional del Data Ring con la meta situada en el centro, guiando los datos a analizar y si ha llegado o no el momento de detener la iteración y evaluar el objetivo alcanzado.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS



Empieza Pequeño. Para los nuevos proyectos de datos, un PMV es la meta recomendada. Este es un objetivo básico y modesto, creado para probar si un concepto de producto basado en datos tiene valor. Una vez logrado, los gerentes de proyecto pueden considerar los mismos conceptos de Data Ring para escalar el PMV a un prototipo.

### OBJETIVO (S)

El establecimiento de objetivos es el primer paso de la planeación del proyecto. El proyecto necesita saber hacia donde va con el fin de saber cuando ha llegado. En cierta medida, un enfoque basado en el destino para el análisis de datos, especialmente cuando se trata de estructuras, procesos y organizaciones complejas, podría conducir a descubrimientos inesperados y trayectorias no planeadas. El descubrimiento es de hecho un factor importante para los proyectos de datos, permitiendo la exploración y permitiendo al equipo de ciencia de los datos 'jugar' con los datos. Dicho esto, debe hacerse de forma estructurada, a través de

pruebas de hipótesis exploratorias, mediante la emulación del método científico (véase el capítulo 1.1, El Método Científico).

Cumplir el objetivo marca la finalización del proyecto. Con un enfoque iterativo, es especialmente importante saber cuál es el aspecto del proyecto terminado para evitar quedarse atascado en el ciclo de refinamiento. Establecer métricas y definiciones satisfactorias ayuda a guiar el camino del proyecto y advertirá de los riesgos si el proyecto empieza a desviarse. Al igual que con la gestión operativa, el proyecto debe monitorear y evaluar sus KPIs a lo largo del proceso iterativo, asegurando que estos puntos de referencia continúen sirviendo al proyecto de la mejor manera posible.

### Establecimiento de Objetivos

El objetivo es proponer una solución orientada a datos para un problema estratégico, con el fin de producir valor. Las necesidades operativas del proyecto se reflejan en los bloques estructurales y en las preguntas guía del Data Ring. Esto se traduce en claras necesidades de recursos, habilidades humanas y procesos concretos, todos ellos orientados por los planteamientos del problema que el proyecto busca resolver. Es probable que la definición de objetivos y el planteamiento del problema se definan en relación el uno con el otro; considerar si el objetivo previsto proporcionará la solución buscada; reflexionar sobre los matices del problema estratégico; refinar uno o ambos en consecuencia. Ayuda a descomponer

problemas más grandes en asuntos más específicos, con el fin de tener un objetivo claro para un problema claro.

### Planteamiento del Problema Estratégico

La idea de "vender el problema antes que la solución" ayuda a impulsar este enfoque y ayuda a comunicar a las partes interesadas cuál es el problema y quién lo tiene. Una vez que se discute el problema, explicar la solución se vuelve simple. A continuación, se presentan dos ejemplos de problemas estratégicos de Servicios Financieros Digitales:

- **Problema:** Los clientes existentes tienen bajas tasas de actividad de dinero electrónico.
- **Problema:** Los clientes potenciales están excluidos del acceso a productos de microcrédito.

### Definición del Objetivo

En el contexto de un proyecto de datos, el objetivo es entregar un proceso basado en datos y un producto con alguna especificación. Esto establece la senda del proyecto. También es importante saber si la senda es buena; en otras palabras, si el producto se basa en una hipótesis razonable sobre por qué funciona y por qué los resultados son fiables. La definición de objetivos tiene dos partes: la especificación del producto y su hipótesis estratégica. Aquí hay dos soluciones propuestas a los problemas planteados anteriormente:

- **Solución Propuesta:** Un modelo mínimo viable de predicción de segmentación de clientes para identificar usuarios activos de alta propensión a aumentar las tasas de actividad.
- **Solución Propuesta:** Un algoritmo de scoring crediticio del cliente a nivel de producción para la emisión automatizada de microcréditos.

### Especificación del Proceso y del Producto

Como se ha detallado anteriormente, los dos productos de datos ejemplificados son un modelo de predicción de segmentación de cliente y un algoritmo de scoring crediticio del cliente. Estos se especifican por su escala, lo que ayuda a describir cuán “grande” es el proyecto, o cómo se integra en sistemas más amplios.

La escala puede ser considerada a lo largo de la siguiente progresión:

- **Proceso:** datos introducidos que proporcionan resultados de forma fiable a través de un proceso automatizado
- **PMV:** un concepto de producto y proceso cuyos resultados evidencian un valor esencial
- **Prototipo:** un concepto de producto con implementación básica, usabilidad y fiabilidad
- **Producto:** un concepto probado con implementación fiable y una propuesta de valor demostrada
- **Producción:** un producto sistemáticamente implementado y a disposición de usuarios o clientes

Enmarcar el objetivo en términos de escala ayuda a definir los requerimientos de recursos y cómo los componentes del proyecto deben encajar. Una prueba de concepto de PMV podría ser entregada en un solo computador portátil en unas pocas semanas. En comparación, la escala a nivel de producción puede requerir servidores de datos especiales, expertos para mantenerlos y supervisión legal para garantizar la seguridad de los datos. Sin embargo, la producción de un PMV requiere recursos duros y blandos (es decir, infraestructura y personas), organizados de acuerdo con un proceso mínimo viable. Esto significa definir roles organizacionales claros, relaciones de gestión y de presentación de informes. Así es como una solución impulsada por datos a un problema estratégico se pone en marcha, cómo se identifican y resuelven los desafíos técnicos, y cómo se asegura que el producto concreto aporta valor estratégico.

### Hipótesis

Lo que estos productos de datos *hacen* es impulsado por una hipótesis subyacente, que sólo está implícita en estos dos ejemplos. La identificación de usuarios activos de alta propensión tiene una hipótesis operativa; existe una correlación entre las variables que definen estos segmentos de clientes y las tasas de actividad. Por ejemplo, los clientes con alto nivel de tiempo aire tienen mayores tasas de actividad. Esta es una hipótesis estadísticamente comprobable y, en última instancia, es la responsabilidad del equipo de ciencia de los datos demostrarlo. Si la

correlación es fuerte y fiable, esta hipótesis impulsada por objetivos de credibilidad y fiabilidad al producto de datos. Una hipótesis similar podría ser construida para un modelo de scoring crediticio con el fin de probar la hipótesis, por ejemplo: los clientes con pequeñas redes sociales tienen tasas de incumplimiento de créditos más altas. El ajuste de la hipótesis no se limita en modo alguno a los proyectos de datos basados en algoritmos. Un tablero de visualización también tiene una hipótesis, con respecto a las relaciones entre los datos que pretenden ser consultados. Tal hipótesis puede no ser estadísticamente probada por algoritmos, pero la fiabilidad de la visualización se basa en estas relaciones que son consistentes y válidas en el tiempo. Debido a esto, la visualización continuará contando una historia significativa o guiando la toma de decisiones útiles.

El principio de la “investigación reproducible” ha cobrado importancia entre los científicos de datos. *La investigación reproducible* describe enfoques transparentes y repetibles para el análisis y cómo se obtienen los resultados en el primer paso de escala del ‘proceso’. En principio, se trata de permitir una validación independiente de los resultados, que puede ser pertinente a efectos normativos o de auditoría. Esta es la razón por la cual el primer paso en la iteración cuando se utiliza el Data Ring es articular un proceso mínimo viable; esto establece el proyecto para lograr resultados fiables sobre los que se basa el valor esencial del producto. Este proceso apoya igualmente los

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

productos de datos para ver inmediatamente si y cuándo las hipótesis se vuelven poco fiables, lo que puede provocar el reajuste de modelos para garantizar la fiabilidad de manera permanente.

### Objetivos de Riesgos y Mitigación

Establecer las metas del proyecto en términos de hipótesis formuladas, probadas y refinadas ayuda a mitigar los riesgos comunes en los proyectos de datos. Los riesgos de una fijación inadecuada de objetivos son:

#### Riesgo: No establecer Objetivos

El principal riesgo es la ausencia de una motivación estratégica de proyectos y objetivos o no objetivos. En otras palabras, este riesgo encapsula las motivaciones para hacer algo significativo con los datos debido al atractivo, para poder usar palabras de moda, porque los competidores lo están haciendo, o simplemente porque es científicamente o tecnológicamente idóneo - sin embargo, las motivaciones carecen de una contrapartida basada en el valor. Este enfoque podría conducir a resultados inutilizables o presupuestos dilapidados, mientras que presenta una oportunidad perdida de aprovechar el análisis para obtener resultados orientados a objetivos que son relevantes para la organización. Para aquellos especialmente motivados en hacer algo, no es raro traer recursos externos que simplemente tienen la tarea de descubrir algo interesante. Esto puede correr el riesgo de tener resultados que no sólo son inutilizables, sino erróneos, ya que la exploración abierta puede permitir análisis sesgados o resultados forzados en la transmisión.

**Mitigación:** Saber lo que el proyecto pretende lograr. Si el equipo quiere hacer algo, pero no está seguro de por dónde empezar, debe contratar a especialistas en operaciones de datos para que revisen los datos y ayuden a arrojar luz sobre qué tipo de información relevante podrían proporcionar al negocio. El objetivo del proyecto se demuestra generalmente por la posibilidad de medir los resultados, pero es importante señalar que las pruebas de hipótesis a menudo resultan falsas. Esto es algo bueno. O iterar y tener éxito, o aceptar que la idea no funciona y volver a arrancar de cero. Esto es superior a un resultado bueno o interesante basado en datos incorrectos.

#### Riesgo: Falta de Foco

Igualmente, relacionado con los riesgos de proyectos que no establecen objetivos, están los proyectos cuyas metas son demasiado generales, están mal definidas o son demasiado flexibles y cambiantes. La meta establece la dirección y esboza lo que se logrará. La falta de claridad puede llevar a que los equipos se distraigan o analicen las preguntas complementarias, proporcionando así resultados auxiliares. Teniendo esto en cuenta, debe existir cierta flexibilidad para el refinamiento iterativo de las metas y permitir explorar y capitalizar el descubrimiento fortuito. La falta de foco también puede ser el resultado de un desajuste problema-solución. Esto sucede cuando el problema estratégico subyacente no se puede definir con precisión, o cuando la solución objetivo propuesta tiene una

inconsistencia lógica, como un negocio débil o una relación estratégica con el problema que se pretende resolver.

**Mitigación:** Establezca metas claras y precisas con relevancia comercial incorporadas en cada uno de los componentes de la hipótesis problema-producto-hipótesis. Asegúrese que pueden ser refinados a través de un enfoque iterativo y revisar estos a medida que el proyecto avanza. Además, asegúrese que existe una relevancia continua de los objetivos a medida que la estrategia empresarial evoluciona de manera independiente. Planee la exploración y flexibilidad dentro de la ejecución del proyecto. Establecer límites exploratorios es clave, ya que aseguran que los proyectos no se salgan de su curso, al mismo tiempo que dan oportunidad al descubrimiento. Esto también es apoyado por las unidades de medida específicas y los objetivos asociados, o KPIs, tanto para los objetivos intermedios como para el logro general de metas.

#### Riesgo: No Basado en Datos

El renombrado economista Roland Coase declaró: "si torturas los datos lo suficiente, confesarán". El riesgo está obligando a los datos a revelar lo que uno espera en un intento de validar el conocimiento, el comportamiento o la organización deseados. Pasar a un enfoque basado en datos significa estar dispuesto a observar las pruebas que resultan del análisis de datos. En otras palabras, analizar proyectos,

procesos o procedimientos a través de datos podría conducir a resultados que no estén alineados con las creencias, pensamientos o estrategias actuales, forzando a una organización a realizar un cambio profundo.

**Mitigación:** Emular el método científico para fijar los objetivos del proyecto en función del tiempo, apoyados por hipótesis rigurosamente contrastadas. Asegurar que la estrategia de ejecución utilice el concepto de investigación reproducible para permitir la repetición y la validación independiente de los resultados. Además, asegurarse que los promotores del proyecto entiendan que no se garantiza la búsqueda de patrones valiosos.

### Riesgo: No Pragmático

Las metas deben ser realistas con respecto a los recursos del proyecto y las expectativas del promotor, por ejemplo, competencia apropiada, infraestructura o presupuesto.

**Mitigación:** Asegúrese que la escala del producto se considera como parte de la definición de objetivos. Esto ayuda a limitar el proyecto y empujar a los directores de proyecto para que coincidan con los recursos y los requisitos. Además, asegúrese que un especialista en tecnología de la información y la comunicación (TIC) realice una evaluación técnica de TI del diseño del proyecto para asegurar el pragmatismo entre el objetivo del proyecto y las herramientas técnicas que se obtienen para entregarla.

## Cuadrante 1: HERRAMIENTAS

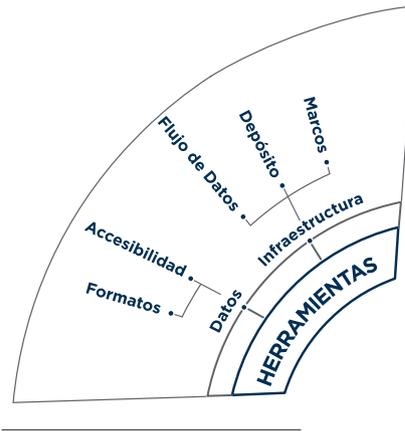


Figura 20: Data Ring Cuadrante 1: HERRAMIENTAS

El mundo y sus fenómenos dinámicos pueden ser observados y fragmentados en datos. En otras palabras, los datos son muestras de la realidad, registradas como mediciones y almacenados como valores. Además, los sistemas complejos desmienten el conocimiento adicional, que está integrado en el comportamiento colectivo de los diferentes componentes del sistema. Los componentes individuales no revelan nada, pero los patrones emergen de la observación del sistema entero.

La revolución de los datos ha proporcionado un aumento exponencial en el volumen, la velocidad y la variedad de datos digitales. Esta mayor disponibilidad de datos digitales permite mayor granularidad y precisión en la comprensión de procesos, actividades y sus vínculos interrelacionadas. Para obtener conocimiento y valor de su análisis,

los datos deben ser almacenados, descritos de manera apropiada y accesibles. Esto requiere que se establezca una infraestructura técnica adecuada para gestionar los datos, su accesibilidad y cálculo. Esto también permite el acceso a análisis de todo el sistema y a los patrones atractivos que pueden generar valor. El primer cuadrante del Data Ring pide a los directores de proyecto que consideren sus datos y la infraestructura técnica necesaria para analizarla a través de dos componentes: datos e infraestructura.

### Herramientas: Datos

Los datos son el insumo fundamental (y el resultado) de un proyecto de datos. Las preguntas guía del Data Ring están agrupadas en dos principios: accesibilidad y formato. Estos son elementos críticos que afectan profundamente las necesidades de recursos y decisiones de procesos.

En primer lugar, es necesario saber cómo se describen los datos, sus propiedades y si representan números, texto, imágenes o sonido. Además, si están estructurados o no estructurados. Los datos también deben ser comprensible para los seres humanos y deben existir en un formato digitalizado y que pueda ser utilizado por una máquina. Estos parámetros básicos son relevantes para los datos de todos los tamaños y formas. Estos son factores críticos para determinar la mejor infraestructura técnica a utilizar para el proyecto. Consulte el Capítulo 1 para más información sobre los formatos de datos.

Recientemente, el concepto de Big Data ha cobrado importancia. Este es un concepto

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

útil, pero su importancia también ha creado conceptos erróneos. Particularmente, la simple disponibilidad de una cantidad apreciable o “grande” de datos puede aumentar el conocimiento o proporcionar mejores soluciones a un problema. A veces esto es cierto. Sin embargo, a veces no lo es. Aunque el big data puede proporcionar resultados, también es cierto que el small data puede lograr con éxito los objetivos del proyecto. Es importante que el director del proyecto asegure que hay datos correctos (y suficientes) disponibles para el trabajo y que las herramientas adecuadas estén en su lugar.

La definición de ‘big’ está cambiando constantemente, por lo que habitar en el término en sí rara vez beneficia a un proyecto. Lo que es más útil en el concepto de Big Data es entender que cuanto más grande sea un conjunto de datos, más tiempo tomará analizarlos. Con esto en mente, un conjunto de datos más grande también requiere capacidades de equipo técnico más específicas y la infraestructura técnica más compleja, sofisticada o cara para manejarlo. La ‘escala’ de los datos también pueden relacionarse con la escala de un objetivo; un PMV puede ser alcanzable con sólo una muestra de datos, mientras que la producción puede requerir datos transaccionales de alta velocidad continuos. Este es un elemento importante del proceso de diseño del proyecto; el hecho de tener terabytes de datos de transmisión puede no ser suficiente para cumplir los objetivos de un proyecto.

Las siguientes preguntas marco ayudan a identificar las fuentes de datos y su alcance en términos de requerimientos de recursos del proyecto. Si los sistemas internos de datos no capturan lo que se supone, esto obliga a la planificación de recursos del proyecto a cambiar, identificando nuevos recursos de datos necesarios:

- ¿Qué datos se producen o recogen a través de las actividades centrales?
- ¿Cómo se producen esos datos (por ejemplo, qué productos, servicios, puntos de contacto)?
- ¿Están los datos almacenados y organizados o pasan por el proceso?
- ¿Están listos los datos para su análisis en forma legible por una máquina?
- ¿Los datos están limpios, o hay irregularidades, valores faltantes o dañados, o errores?
- ¿Los datos disponibles son estadísticamente representativos, para permitir la prueba de hipótesis?
- ¿Cuál es la relación entre el tamaño de los datos y las necesidades de desempeño?

Estas preguntas ejemplifican el esfuerzo necesario en la fase inicial con el fin de adquirir con éxito, limpiar y preparar el conjunto de datos para su posterior análisis. Dependiendo de cuánto control esté disponible en todo el proceso basado en datos, esta fase de preparación será más larga o más corta, lo que significa costos de proyecto más altos o más bajos. La planificación inadecuada de los datos iniciales puede resultar en un aumento de

los costos las revisiones podrían significar la necesidad de seleccionar diferentes infraestructuras computacionales o diferentes capacidades de equipo.

### Accesibilidad de Datos

Se debe poder acceder a los datos para poder usarlos. Puede parecer trivial, pero este problema es complejo y debe considerarse al principio de cada proceso basado en datos para asegurar que los resultados estén a tiempo y dentro del presupuesto, o si los resultados son incluso posibles. La privacidad del cliente, la solicitud y la concesión de permisos de uso de datos y el establecimiento de quién tiene propiedad e interés legal una vez que se otorguen los permisos de acceso a datos son factores que dificultan la accesibilidad de los datos. La accesibilidad de los datos se puede juzgar según tres factores:

### Legal

Las regulaciones podrían impedir que un excelente y bien diseñado análisis basado en datos se realice en su totalidad. Esto interrumpiría el proceso en una fase intermedia, por lo que es vital ser consciente de las limitaciones legales desde el principio.

**La propiedad** de los datos debe ser establecida, identificando quién tiene permiso para analizarlos para obtener conclusiones. Si hay acuerdos de propiedad intelectual, deben cubrir tanto los trabajos existentes como las derivadas. Si el análisis es una colaboración de investigación, deben establecerse acuerdos de publicación,

incluida la claridad sobre lo que constituye información de propiedad exclusiva y lo que puede hacerse público.

**El uso ético** de la información también puede tener restricciones legales. Los datos sobre personas, grupos u organizaciones deben ser tratados con cuidado, anteponiendo la seguridad como primera consideración. Las regulaciones de protección de datos también pueden influir en cómo los datos pueden o no ser transferidos del propietario al analista, por ejemplo, si pueden ser enviados electrónicamente o por almacenamiento físico. Además, las regulaciones pueden esbozar procedimientos para los datos que salen de las fronteras nacionales, que se enrutan a través de terceros o se almacenan en servidores ubicados en países específicos.

### **Tecnológico**

Pueden existir barreras si los *formatos* de datos están desalineados con la tecnología seleccionada para el procesamiento y análisis de datos. Como un ejemplo simple, un algoritmo PLN no se puede aplicar significativamente a datos de imagen. De manera más práctica, las bases de datos generalmente se optimizan para tipos específicos de datos; y algunas tecnologías no están diseñadas para trabajar juntas, de forma similar a la construcción de un flujo de trabajo destinado a mezclar productos de Apple y Microsoft. Esto puede resultar en costos e ineficiencias, y puede crear problemas adicionales a resolver, intentando alineaciones forzadas.

**La información digital** es necesaria para analizarla a escala y velocidad de máquina. Puede haber algunas excepciones matizadas respecto de la regla y la IA está empujando estos límites.

**Se necesita compatibilidad** entre el formato de datos y la tecnología utilizada para gestionarlos. Incluso si se digitalizan los conjuntos de datos, pueden ser aislados e inaccesibles debido a elecciones tecnológicas incompatibles hechas por diferentes departamentos de la misma empresa, gobierno u organización. A veces pueden existir sistemas obsoletos, que también pueden prevenir interacciones con soluciones, lenguajes y protocolos modernos. La cantidad de esfuerzos para armonizar la infraestructura tecnológica podría ser una barrera no trivial desde una perspectiva de costo-tiempo.

### **Estratégico**

Los actores podrían tratar de preservar una ventaja competitiva mediante la intermediación del acceso a sus activos de datos. Esto generalmente puede tomar tres formas: al requerir hardware o software especial para leer formatos de datos de propiedad exclusiva; controlando cómo se pueden utilizar los datos; o mediante la exigencia de tarifas especiales de licencia. Considerando que los factores tecnológicos podrían ofrecer una solución alternativa, aunque sea a veces compleja o ineficiente, a menudo se establecen factores estratégicos para garantizar deliberadamente que el acceso sólo es

posible de acuerdo con la especificación del propietario de los datos, o tal vez se niegue acceso por completo.

### **Formatos de Datos**

Los datos digitales se puede representar de muchas *formas* y un *formato de datos* describe parámetros entendidos por humanos (es decir, texto, imagen, video, biométricos). A menudo, el formato es referido por el sufijo de tres o cuatro letras al final de un archivo de computador. Formato también puede referirse a estructuras de almacenamiento de datos y bases de datos de manera más general, por ejemplo: Oracle, MongoDB y JSON. (Consulte el Capítulo 1.1, Definición de Datos).

Existen numerosos formatos de datos, especialmente los de almacenamiento y procesamiento. El formato de datos está firmemente determinado por el contexto empresarial u organizacional y, en particular, por las personas responsables de la gestión de la creación, almacenamiento y procesamiento de datos. Para los gerentes de proyectos, *reconocer la fragmentación del formato y los problemas de incompatibilidad* son claves para establecer la alineación de datos requerida para proyectos bien diseñados. La comprensión de los valores registrados en un conjunto de datos, así como los metadatos más generales del conjunto de datos, ayuda a los directores de proyectos a planificar adecuadamente.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

El valor de un *punto de datos* se refiere al contenido intrínseco de un registro de datos. Este contenido se puede expresar en forma numérica, temporal o textual, denominada *tipo de dato*. Para el análisis de datos, el factor crucial es que estos valores subyacentes no se vean afectados por errores o sesgos sistemáticos debidos a fallos en la infraestructura o relacionados con el ser humano. En general, los gerentes de proyectos no consideran cómo se recogen los datos o si la instrumentación está bien ajustada. Es importante comprender cómo se realizan estas mediciones subyacentes y asegurar que exista una adecuada transferencia de conocimientos entre los propietarios de datos y los analistas de datos sobre cuestiones clave de medición. Como ejemplo práctico, si un sistema falló durante una actualización de TI, esta actualización se verá reflejada en una caída dramática en las transacciones. Los analistas deben ser conscientes de esta información para interpretar la anomalía correctamente. Las anomalías en los valores de los datos influyen en gran medida en el proceso de limpieza de datos y la planificación de proyectos relacionada.

*Metadatos* son 'datos sobre datos', lo cual incluye toda la información de antecedentes adicional que enriquece un conjunto de datos y lo hace más comprensible. Las columnas de títulos de encabezados de una hoja de Excel son metadatos (los títulos son datos de texto que describen los valores de las siguientes filas). Por ejemplo, imagine un conjunto de datos con las etiquetas, 'nombre del corresponsal' y 'volumen de transacción', procedido por una columna de números sin encabezado. ¿Están esos

números relacionados con los valores de las transacciones, tal vez los tiempos en que se realizaron las transacciones? Si el proyecto busca visualizar volúmenes en un mapa, la ubicación del corresponsal también se convierte en un requisito de datos; el proceso computacional debe ser capaz de pedir al conjunto de datos que proporcione todos los valores de ubicación. Si la categoría de ubicación no está compuesta por metadatos definidos, entonces el proceso no podrá encontrar coordenadas GPS que representen gráficamente. La solución podría ser sencilla, quizás, agregar un título de 'ubicación' a esta columna sin nombre. De esta manera, los equipos de proyecto pueden agregar información contextualizada a conjuntos de datos y proporcionar descripciones más detalladas de los datos (es decir, los metadatos) sobre los cuales el proceso analítico puede hacer preguntas y utilizar. En este sentido, los metadatos son sólo otro conjunto de datos. Los metadatos son especiales porque están inherentemente conectados al conjunto de datos subyacente, lo que permite que este proceso de preguntas y respuestas tenga lugar. Esto es sólo un ejemplo; los metadatos son algo más que encabezados de columna. Incluso en Excel, existen metadatos acerca de la hoja de cálculo sobre la cuál se trabaja, por ejemplo, tamaño del archivo, la fecha de creación y el autor son todos ejemplos de metadatos. Estos metadatos subyacentes permiten la búsqueda y clasificación de archivos, por ejemplo, el sistema operativo puede solicitar todos los archivos modificados en la última semana. Las respuestas se obtienen a través de los metadatos del archivo.

La comprensión de cómo los conjuntos de datos están conectados a través de metadatos es un elemento clave del diseño del proyecto y clave para identificar brechas y oportunidades de análisis. Los metadatos ayudan a identificar dónde pueden ser necesarios datos adicionales para cumplir con los objetivos del proyecto y cómo vincularse en nuevos conjuntos de datos cuando sea necesario. Los metadatos ayudan a identificar eficiencias donde ya existen conjuntos de datos suplementarios; licenciar datos de terceros puede cerrar algunas brechas y también se podrían crear metadatos derivados o sintéticos para ayudar a contextualizar los conjuntos de datos del proyecto. Para los gerentes de proyectos, es importante saber cuándo y dónde es probable que existan metadatos. Si no forman parte de conjuntos de datos iniciales, puede ser mejor pedir esta información a los propietarios de los datos, en lugar de contextualizarlos como parte del trabajo del proyecto.

### Herramientas: Infraestructura

Como se explicó anteriormente, los datos son el insumo (y el resultado) fundamental de un proyecto de datos. La infraestructura es el lugar a donde los datos físicamente entran y de donde salen. Los datos son información digital que debe ser adquirida, almacenada, procesada y calculada utilizando herramientas informáticas que se ejecutan en computadores virtuales o físicos.

La infraestructura tecnológica debe adecuarse a los objetivos que se planteen en cuanto al *volumen*, la *variedad* y la *velocidad* de los datos. Los recursos de la infraestructura permiten la usabilidad de

los datos y afectan fuertemente el 'poder' y la eficacia de los algoritmos científicos y modelos matemáticos aplicados. La infraestructura genérica basada en datos está constituida por estos bloques centrales:

### **Flujo Funcional de Datos**

El flujo funcional de datos es una cadena funcional de hardware o software en la que cada elemento recibe datos de entrada, se procesan y luego se reenvían al siguiente elemento. Es la forma cómo se cargan los datos en el proceso analítico; el flujo funcional de datos incluye el proceso de carga, las herramientas para procesar las cifras, cómo se descargan las cifras y cómo se introducen luego en un proceso operativo. Este flujo funcional, por ejemplo, ofrece la integración técnica de un producto de datos en sistemas corporativos más amplios. El flujo funcional debe ser planificado para asegurar un proceso fiable que tome datos sin procesar y proporcione resultados utilizables. El proyecto debe asegurar que se escriba un diagrama esquemático o de flujo para describir la implementación funcional del flujo de datos. La carga inicial en el flujo funcional generalmente marca el inicio operativo de un proyecto de datos, comenzando con el proceso de Extracción-Transformación-Carga (ETL) de datos. El ETL es un plan de procedimiento, establecido como parte de la gestión de datos del proyecto, que se analiza con mayor profundidad más adelante.

### **Depósito**

Una base de datos o sistema de archivos se llama *depósito*, o el elemento de infraestructura para almacenar datos. El depósito afecta la forma en que

se guardan y recuperan los datos y estos procesos de entrada y salida son fundamentales para diseñar un sistema que tenga buen desempeño. Se necesita tiempo para escribir datos en un disco y cuando llega una consulta, toma tiempo buscar la respuesta y enviarla al siguiente paso en el flujo funcional de datos. Las herramientas correctas de base de datos se guían a menudo por la naturaleza misma de los datos, su formato y su estructura. Además, la forma en que se utilizan los datos desempeña un papel en el depósito un sistema de archivo busca comprimir los datos en un volumen tan barato como sea posible, mientras que una base de datos transaccional asegura velocidad y fiabilidad para que los clientes no se queden esperando. Los marcos también guían la elección de la base de datos proporcionando herramientas integradas optimizadas para soluciones y diseños de depósitos específicos.

### **Marcos**

Un marco es un conjunto de soluciones diseñado para un grupo de problemas. Técnicamente, es un conjunto de bibliotecas predefinidas y herramientas comunes que permiten escribir código y programas de forma más rápida y sencilla. En el área de Big Data, estas incluyen plataformas que recogen herramientas, bibliotecas y características para simplificar los procesos de gestión y manipulación de datos (por ejemplo, Apache Spark, Apache Hadoop, Hortonworks, Cloudera. Véase el capítulo 2.2.3, Base de Datos Tecnológica). Cabe señalar que un proyecto puede integrar múltiples marcos. Se recomienda utilizar

un marco establecido, porque esto evita la necesidad de programar herramientas comunes desde cero, lo cual puede suponer un enorme ahorro de tiempo y de costos. Lo que se sacrifica es que el enfoque del proyecto, se debe adaptar a la manera en que el marco resuelve el conjunto de problemas que fue diseñado para abordar, lo cual puede o no ajustarse perfectamente a las necesidades precisas del proyecto. Seleccionar el marco equivocado corre el riesgo de incompatibilizar su enfoque de soluciones con los problemas del proyecto, introduciendo ineficiencias.

Los marcos se diseñan típicamente alrededor de especificaciones del hardware, y funcionan finalmente en las computadoras que procesan las cifras para el proyecto de datos. Aunque la potencia computacional bruta es igualmente un elemento crítico de la infraestructura del proyecto, lo mejor es primero planear el flujo funcional de datos, los requisitos del depósito y los marcos necesarios para satisfacer las necesidades del proyecto. Las especificaciones de computación adecuadas tienden a llegar después. El diseño y la gestión de la infraestructura no suele ser el papel de los gerentes de proyectos, pero ellos sí tienen que garantizar que las capacidades y los recursos estén disponibles para satisfacer las necesidades del proyecto. Esta es la razón por la cual una evaluación de TI se indica específicamente como parte de la gestión de riesgo y el establecimiento de objetivos pragmáticos. Depender de equipos de TI internos o asegurar la capacidad relevante en el equipo del proyecto de datos es fundamental para ayudar a evaluar los requisitos de

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

infraestructura y las necesidades técnicas, incluyendo escalabilidad, tolerancia a fallos, distribución o aislamiento del entorno. Estos términos técnicos son relevantes para la infraestructura computacional de la empresa a gran escala; los objetivos para un PMV se pueden lograr con mucho menos. Incluso los proyectos de datos pequeños probablemente involucrarán a la arquitectura empresarial en torno al flujo funcional de datos. Los datos que necesitan un proyecto se alimentarán casi con seguridad de los sistemas corporativos, y esto debe estar bien definido, planificado y coordinado con los equipos de TI.

### Cuadrante 2: HABILIDADES

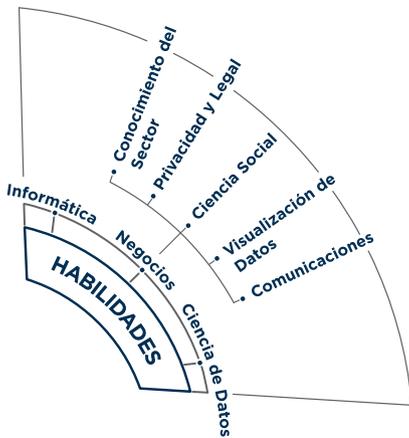


Figura 21: Data Ring Cuadrante 2: HABILIDADES

Los proyectos basados en datos necesitan científicos de datos. Dicho esto, “científico de datos” es un título relativamente vago y amplio, que todavía está siendo

definido. Mientras tanto, la industria y los medios de comunicación han generado exageraciones acerca de Big Data, el aprendizaje automático y una gran cantidad de tecnologías, al tiempo que crean una conciencia más amplia del enorme valor potencial de los datos. Esto ha generado presión para invertir en estos recursos con el fin de mantenerse al día con la competencia. Es fundamental que el gerente de proyecto de datos tenga en cuenta que se necesitan conjuntos muy específicos de habilidades y experiencia técnica para cumplir con los requisitos de un proyecto de datos. Como cosa igualmente crítica, deben ser conscientes que muchos de estos campos de experiencia se están formando dinámicamente al tiempo con el rápido cambio de la tecnología. El segundo cuadrante del Data Ring pide a los gerentes de proyecto que consideren los recursos humanos necesarios para lanzar el proyecto a través de tres componentes: la informática, negocios y la ciencia de datos.

### El Equipo

Armar la combinación correcta de habilidades es un reto para los gerentes de proyectos de datos debido a la evolución dinámica de la tecnología, los tamaños cada vez mayores de los conjuntos de datos y las habilidades necesarias para extraer valor de estos recursos.

**Un científico de datos suele ser un equipo de personas** gestionando datos. Más allá de una sola competencia, esto generalmente requiere un equipo

interdisciplinario de expertos técnicos que interactúan de cerca con todas las unidades - una sola persona o grupo - que gestionan los datos desde la adquisición hasta la visualización.

**Los equipos son dinámicos y colaborativos**, y es difícil seguir el ritmo de la innovación y el desarrollo de nuevas habilidades, conocimientos emergentes y creciente hiper-especialización. Las posibilidades de tercerización pueden lograr el dinamismo y las habilidades adecuadas para el objetivo perseguido. Alternativamente, la retención o la formación interna de científicos de datos generalistas pueden ayudar a asegurar la colaboración exitosa en un equipo multidisciplinario de especialistas en datos y operaciones de negocio.

**Se requiere una cultura abierta, científica y basada en datos.** Un enfoque científico adecuado y una cultura de datos deben existir dentro del equipo e, idealmente, dentro de toda la empresa. Debido a que el establecimiento de objetivos adecuados se basa en la emulación del método científico y las pruebas de hipótesis exploratorias, el equipo de científicos de datos debe estar impulsado por el sentido de la curiosidad y la exploración. El director del proyecto debe asegurarse que esa curiosidad sea dirigida y se mantenga apuntando a los objetivos.

Las siguientes preguntas marco ayudarán a los gerentes de proyectos a identificar recursos y necesidades:

- ¿Quién es responsable de administrar los datos en la empresa? ¿Cómo?
- ¿Existen colaboraciones en curso con instituciones de investigación u organizaciones calificadas para realizar las actividades de ciencia de datos?
- ¿Qué canales de reclutamiento existen en lo que respecta a los expertos en datos?
- ¿Cómo se fomenta la cultura de datos dentro de la empresa, y quién está involucrado?
- ¿Cómo se facilita la colaboración multidisciplinaria en la planificación y ejecución de proyectos?
- ¿Cómo se garantiza la validez científica en la elección de algoritmos y representaciones matemáticas de datos (modelado)? ¿Hay una persona calificada encargada de asegurar que los resultados sean verdaderos?
- ¿Quién asegura que están establecidas buenas prácticas y que los algoritmos están programados eficientemente?
- ¿Existe una colaboración abierta entre el equipo de datos y otras unidades de negocios?

Un equipo completo y altamente interdisciplinario es difícil de lograr, y es muy poco probable que la mayoría de las empresas tengan una amplia gama de habilidades relevantes a las que puedan recurrir cuando sea necesario. Entender estas brechas y ser el primer paso para ser consciente de todo el potencial y planificar las inversiones de tercerización, lo que se considera una parte de la planificación del proceso.

## Habilidades: Informática

Los datos son piezas digitales de información que se tienen que adquirir, almacenar, procesar y administrar por medio de herramientas informáticas, lenguajes de programación y de secuencias de comandos y bases de datos. Por tanto, las habilidades deben incluir conocimientos sobre:

### Computación en la Nube

Cuándo las fuentes de datos son grandes o enormes, las herramientas de programación normales y los recursos computacionales locales, tales como las computadoras personales, se vuelven rápidamente insuficientes. Las soluciones 'en la nube' son una respuesta práctica y efectiva a este problema, pero significan dominar los conocimientos esenciales sobre los sistemas de virtualización, paradigmas de escalamiento y la programación del marco. (Véase el capítulo 2.2.3, Base de Datos Tecnológica)

### Lenguajes de Scripting

Trabajar con infraestructura informática significa programar. Python o R son a menudo las mejores opciones para hacer prototipos rápidos y explorar los patrones de datos. Estas son opciones probables para una meta de PMV y para el desarrollo de proyectos en etapas tempranas. Ambos lenguajes de scripting están bien establecidos como herramientas indispensables de la ciencia de los datos, y el equipo idealmente debe 'hablar' ambos. (Véase el capítulo 2.2.3, Base de Datos Tecnológica)

Ciertas infraestructuras corporativas y requisitos de certificación pueden requerir diferentes opciones de programación, tales como Scala, Java o C++. Esto puede ser un problema para escalada de un objetivo;

más allá de la elaboración de un prototipo y la implementación en la producción, se requerirán invariablemente soluciones de programación de nivel empresarial, así como las habilidades para implementar. Esto también significa que puede ser necesaria la refactorización del código o la traducción entre lenguajes informáticos, así como interacciones cercanas entre el equipo de datos y los miembros del personal de TI y de ingeniería.

## Bases de datos y Almacenamiento de Datos

El Capítulo 1 habla de los datos estructurados versus los no estructurados. Un proyecto de datos puede basarse en ambos, que son manejados respectivamente por bases de datos relacionales y bases de datos no relacionales. El uso de estas herramientas requiere diferentes habilidades. Es probable que los datos procedente de bases de datos transaccionales empresariales procedan de bases de datos relacionales. Con mayor frecuencia, incluso los datos internos, tal como KYC o información biométrica, pueden ser almacenados por cualquiera de las soluciones, dependiendo del método de recolección. Sin embargo, es probable que un algoritmo de scoring crediticio que busque utilizar datos de redes sociales se base en datos no estructurados de fuentes de datos no relacionales.

## Control de Versiones y Colaboración

Las herramientas de control de versiones son esenciales para la evolución organizada del código, el mantenimiento y el trabajo en equipo y, por tanto, son esenciales para una buena planificación del proyecto.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

### Habilidades: Ciencia de los Datos

#### Herramientas Científicas

Los contextos diferentes requerirán una mezcla específica de acuerdo con las necesidades del proyecto, pero las siguientes son áreas académicas amplias de las que los proyectos de datos probablemente necesitarán sacar provecho:

- Bases Sólidas en Estadística: utilizadas para pruebas de hipótesis y validación de modelos.
- Ciencia de Redes: una disciplina que utiliza nodos y aristas para representar redes complejas matemáticamente; crítica para datos de redes sociales o mapeo de transacciones tipo P2P.
- Aprendizaje Automático: una disciplina que utiliza algoritmos para aprender de los comportamientos de datos sin una cosmología predefinida explícita; la mayoría de los proyectos que ofrecen un modelo o algoritmo.
- Las ciencias sociales, la PLN, la ciencia de la complejidad y el aprendizaje profundo son también habilidades deseables que podrían desempeñar un papel clave en áreas específicas de interés.

#### Curiosidad y Mente Científica

La actitud y las competencias conductuales son factores críticos para un equipo de ciencia de los datos exitoso. Personas que buscan explorar, hacer minería, agregar, integrar - y así identificar patrones y conexiones - conducirá a resultados superiores. En otras palabras, algunas 'habilidades de hacking' generales son

un valor agregado para el equipo de ciencia de datos; simplemente, el equipo debe poseer un enfoque mental para resolver problemas y un impulso interno para encontrar patrones a través de un análisis metódico.

Además, la validación científica es esencial para un proyecto de datos, y los científicos de datos deben tener una mente científica. Es decir, un enfoque metódico para hacer y responder preguntas, y un impulso para probar y validar resultados. Es importante destacar que los miembros del equipo deben encontrar motivación en los resultados y estar abiertos a cualquier interpretación que surja del análisis de los datos, incluso si los resultados pudieran contradecir las expectativas iniciales. De acuerdo con el método científico, este enfoque debe incorporarse en las competencias conductuales, por ejemplo: hacer observaciones; pensar en preguntas interesantes; formular hipótesis; y el desarrollo de predicciones comprobables.

#### Diseño y Visualización

Esto requiere habilidades multidisciplinares en términos de necesidades técnicas y comerciales. Desde el punto de vista técnico, 'DataViz' o la visualización de datos no debe considerarse exclusivamente como la parte final del proyecto destinado a embellecer los resultados. Es relevante a lo largo de la exploración y elaboración de prototipos, y está bien incorporado en etapas periódicas del proyecto, lo que lo convierte en una habilidad básica para que los científicos de datos identifiquen patrones.

### Habilidades: Negocios

El establecimiento de objetivos está esencialmente relacionado con la obtención de resultados relevantes para el negocio, y la comparación con las métricas apropiadas y los KPI. Saber cómo conectar estas métricas a la ejecución del proyecto es el objetivo mismo de hacer el proyecto. Esto requiere que el equipo del proyecto tenga conocimientos sólidos de negocio. Una perspectiva comercial clara también es esencial para la interpretación de los resultados y, en última instancia, para utilizar e implementar el proyecto para aportar valor. Con respecto a las habilidades, el mensaje clave es que un 'facilitador' necesita intermediar datos, especialistas técnicos, la administración y estrategia del negocio, para poder traducir los resultados obtenidos de los datos para personas no técnicas; el papel de este intermediario también articula las necesidades del negocio en términos de algoritmos y soluciones técnicas de vuelta al equipo. Hay especialidad en crecimiento llamada operaciones de datos que encapsula esta función.

#### Privacidad y Legal

Excepto en los casos en los que se liberan conjuntos de datos con una licencia abierta - explícitamente permitiendo su uso, remezcla y modificación - como en las iniciativas de datos abiertos, las cuestiones relacionadas con la privacidad, la propiedad de los datos y los derechos de uso para un fin específico no son insignificantes (Ver las barreras legales a los datos - en Accesibilidad de Datos en la página 108). Se debe consultar a los especialistas legales

corporativos para asegurarse que todas las inquietudes de los interesados sean debidamente atendidas. Dicho esto, los problemas de Big Data y de privacidad están entrando en nuevos terrenos, y la legislación que busca regular los datos está todavía en desarrollo. Muchas empresas hoy en día están construyendo sus negocios basados en datos, aprovechando las lagunas legales en las leyes locales. Esto puede presentar riesgo si las leyes cambian, al mismo tiempo que presentan oportunidades, trabajando para construir un ecosistema habilitador.

En términos de habilidades, cada uno de los miembros del equipo del proyecto debe tener algo de conocimiento legal básico. Esto permite identificar problemas potenciales y permite un diálogo constructivo con los expertos

legales a cargo. La conciencia respecto a los asuntos legales es particularmente relevante cuando se consiguen consultores externos y cuando se busca asegurar que los Acuerdos de Confidencialidad (NDAs, por sus siglas en inglés) son exhaustivos, cumplen con la normativa y pueden ser exigibles. Desde una perspectiva interna y externa, los datos también pueden ser una fuente de fraude. Los casos de fraude son cada vez más técnicamente sofisticados y orientados a los datos. A pesar que un equipo de ciencias de datos quiere tener habilidades de hacker como parte de un conjunto de habilidades balanceado, no quiere hackers reales. Es fundamental que todo el equipo esté bien versado en consideraciones legales y sea responsable tanto legal como moralmente de adherirse a estas.

# Lecciones de la Industria: Desanonimización de Datos

### Protección de Datos y Protección del Consumidor: La Anonimización de Datos de Usuarios es Necesaria y Difícil

En 2006, America Online (AOL), un proveedor de servicios de Internet, puso 20 millones de consultas de búsqueda a disposición pública para la investigación. Las personas fueron anonimizadas por medio de un número aleatorio. En un Artículo del *New York Times*, los periodistas Michael Barbaro y Tom Zeller describen cómo el número de cliente 4417749 fue identificado y posteriormente entrevistado para su artículo. Aunque que el usuario 4417749 era anónimo, sus búsquedas no lo eran. Ella era una ávida usuaria de Internet, buscando términos de búsqueda identificadores: “Dedos entumecidos”, “60 hombres solteros”, “perro que orina sobre todo”. Las búsquedas incluyeron nombres de personas y otra información específica, incluyendo paisajistas en Lilburn, Georgia, Estados Unidos de América. Ninguna búsqueda individual identifica, pero para un detective - o un periodista - es fácil identificar a las mujeres sesentonas con perros con mal comportamiento y bonitos jardines en Lilburn, Georgia. Thelma Arnold fue encontrada y

afirmó que los buscadores eran suyos. Fue una debacle de relaciones públicas para AOL.

Otra violación de datos llegó a los titulares en 2014 cuando Vijay Pandurangan, un ingeniero de software, desanonimizó 173 millones de registros de taxis publicados por la ciudad de Nueva York para una iniciativa de Datos Abiertos. Los datos se encriptaron usando una técnica que hace matemáticamente imposible hacer ingeniería inversa del valor encriptado. El conjunto de datos no tenía información de búsqueda identificadora como Arnold, pero los números de registro de taxis encriptados tenían una estructura conocida públicamente: número, letra, número, número (por ejemplo, 5H32). Pandurangan calculó que sólo había 23 millones de combinaciones, así que simplemente alimentó todas las opciones posibles de input en el algoritmo de encriptación hasta que produjo salidas coincidentes. Dado el poder informático de hoy, él fue capaz de desanonimizar a millones de taxistas en sólo dos horas.

Netflix, una compañía de cine y medios de comunicación en línea, patrocinó una competencia con crowdsourcing que desafiaba a los científicos de datos a mejorar en un 10 por ciento su algoritmo interno para predecir las puntuaciones de calificación de películas de los clientes. Uno de los equipos desanonimizó los hábitos de consumo de películas de los usuarios encriptados para la competencia. Haciendo referencia cruzada con la base de datos pública del Internet Movie Database (IMDB), que proporciona una plataforma de redes sociales para que los usuarios califiquen películas y escriban sus propias reseñas, los usuarios fueron identificados por los patrones de conjuntos de películas igualmente calificadas de los conjuntos de datos públicos de IMDB y encriptados de Netflix. Netflix llegó a acuerdos en las demandas presentadas por usuarios identificados y se enfrentó a investigaciones sobre protección de datos de consumidores presentadas por el gobierno de los Estados Unidos.

Anonimizar datos correctamente es muy difícil, habiendo muchas maneras de reconstruir la información. En estos ejemplos, la referencia cruzada de los recursos públicos (Netflix), la fuerza bruta y las potentes computadoras (taxis de Nueva York), y la investigación detectivesca a la antigua (AOL) condujeron a violaciones de la privacidad. Si se liberan datos para proyectos de datos abiertos, investigación u otros propósitos, se necesita tener mucho cuidado para evitar riesgo de desanonimización y graves consecuencias legales y de relaciones públicas.



## Ciencias Sociales y Datos

La intersección del conocimiento de datos y ciencias sociales es un nuevo ámbito de actividad académica y una competencia clave para los equipos de proyectos. La motivación del negocio para un proyecto de datos generalmente se reduce a los clientes, ya se trate de una mayor actividad, nuevos productos o nuevos datos demográficos. Para atraer a los clientes, uno necesita saber algo sobre ellos. Las habilidades de datos de ciencias sociales ayudan a interpretar los resultados a través de una lente que busca entender lo que los usuarios están o no están haciendo y por qué; así, los equipos son capaces de identificar patrones de datos útiles mejor y ajustar modelos alrededor de variables que representan las normas y actividades sociales del cliente.

## Conocimiento Sectorial

La experiencia en el tema, el conocimiento del mercado y el conocimiento sectorial describen la relación crítica entre los resultados del proyecto y el valor del negocio. A falta de conocimiento sectorial, se pueden analizar los datos equivocados, los modelos de alta precisión pueden probar la hipótesis equivocada o pueden seleccionarse variables estadísticamente significativas que no tienen relación con los KPIs del negocio. Con muchos modelos de aprendizaje automático entregando 'cajas negras' o marcos de infraestructura que usan enfoques automatizados, existen

riesgos significativos que un proyecto de datos pueda ofrecer resultados que parecen excelentes pero que son dirigidos, sin saberlo, sin verdadera Inteligencia de Negocios (BI). Entonces, el diálogo constante con los expertos del sector debe formar parte del diseño del proyecto.

## Comunicaciones

Los datos cuentan una historia. De hecho, las cifras exactas pueden contar algunas de las historias más poderosas de una manera concisa. Los vínculos entre las comunicaciones empresariales y los equipos de proyecto son un elemento importante para utilizar los resultados de los proyectos - tales como poder implementarlos de manera correcta, alineados con la estrategia de comunicación. También existe una fuerte relación de comunicación con la visualización y diseño de datos, especialmente para proyectos orientados al público. La visualización de datos es importante para comunicar resultados intermedios y finales. Asegurar las habilidades de diseño visual es tan importante como las habilidades técnicas para trazar gráficos, volver interactivos los resultados o servirlos al público a través de sitios web. Para muchos proyectos de datos, la visualización es un producto básico, tal es el caso de los paneles de control y de muchas de las metas del proyecto específicamente orientadas a dirigir las comunicaciones empresariales.

## Cuadrante 3: PROCESO

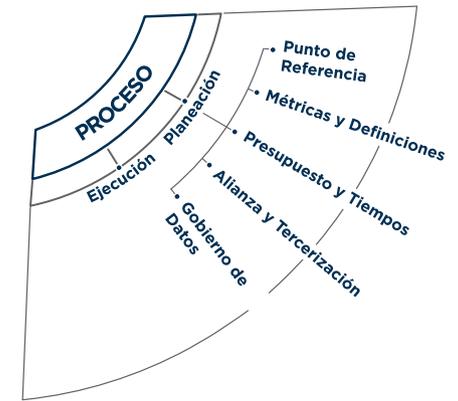


Figura 22: Data Ring Cuadrante 3: PROCESO

En las secciones anteriores se analizó la mitad superior del Data Ring, centrándose en los requisitos duros (infraestructura, datos y herramientas) y requisitos blandos (habilidades y competencias). Esta sección ahora cambia a la mitad inferior del Data Ring, que examina el proceso para diseñar y ejecutar un proyecto de datos.

Reconociendo que las corporaciones o instituciones tienen sus propios enfoques basados en una mezcla de historia organizacional, cultura corporativa, estándares de KPIs y regulaciones de gobierno de datos, lo que sigue son las que se consideran buenas prácticas generales para habilitar proyectos basados en datos y sus entregables.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

Los proyectos de datos deben definir sus entregables, los resultados de la Planificación y Ejecución del proyecto. Estos resultados median entre el Proceso y el subsiguiente bloque que apunta a convertirlos en Valor para el negocio. La siguiente lista especifica ocho elementos comunes a muchos proyectos de datos. Cuando sean aplicables, éstos deben estar en el cronograma de los entregables del proyecto, o especificarse dentro de los términos de referencia para la capacidad tercerizada.

### **Conjunto(s) de Datos:**

Los conjuntos de datos son todos los datos que se han recogido o analizado. Dependiendo del tamaño, el método de recolección y la naturaleza de los datos, el formato del conjunto de datos o conjuntos de datos puede variar. Todos estos deben documentarse, con información sobre dónde están ubicados - tales como en una red o una nube - y cómo acceder a ellos. La entrada de datos brutos se tiene que 'limpiar', proceso discutido en la sección de ejecución, abajo. Los conjuntos de datos limpios deben considerarse como entregables específicos, junto con los métodos con secuencias de comandos o los pasos metodológicos aplicados para limpiar los datos. Por último, los conjuntos de datos y métodos agregados también pueden considerarse como entregables específicos. Estos son necesarios para ayudarle a los promotores del proyecto a ver lo que se hizo con los

datos y, posiblemente, para detectar errores. Además, estos apoyan proyectos subsiguientes o análisis derivados que se basan en datos limpios y pre-agregados.

### **Cuestionarios y Herramientas de Recolección**

Los proyectos que requieren la recolección primaria de datos, tanto cuantitativos como cualitativos, pueden requerir el uso o desarrollo de herramientas de recolección de datos, tales como instrumentos de encuestas, cuestionarios, datos de registro de ubicaciones, informes fotográficos o discusiones o entrevistas en grupos focales. Estos instrumentos deben ser entregados, junto con los datos recogidos, incluyendo todos los lenguajes, traducciones y transcripciones. Éstos son necesarios para permitir encuestas posteriores o preguntas coherentes de serie cronológica, y también proporcionan los documentos necesarios de auditoría o verificación si surgen preguntas sobre los métodos de recopilación de datos en una etapa posterior.

### **Informe de Inventario de Datos**

Se trata de un informe con un resumen de los datos que se utilizaron para el análisis. Este informe incluye el tipo, tamaño y fecha de los archivos. Debe incluir discusiones sobre anomalías importantes o brechas en los datos, así como una evaluación de si las anomalías pueden estar sesgadas estadísticamente o presentar riesgo a la interpretación. Puede incluir gráficos que representan los puntos de datos principales para los segmentos centrales,

tales como las transacciones a lo largo del tiempo desglosadas por tipo de producto para mostrar tendencias, picos, bajas y brechas. Entregado al principio del proceso de ejecución, el informe de inventario de datos es una oportunidad para discutir los posibles riesgos del proyecto debidos a los datos subyacentes, así como las estrategias para la corrección de rumbo y la necesidad de refinamiento o re-adquisición de datos. Es especialmente útil medir el alcance de los requisitos de limpieza de datos y esforzarse por ajustar las anomalías de una manera estadísticamente imparcial.

### **Diccionario de Datos**

El diccionario de datos consolida la información de todas las fuentes de datos. Es una colección de la descripción de todos los elementos de datos, por ejemplo, tablas. Esta descripción normalmente incluye el nombre del campo de datos, su tipo, formato, tamaño, la definición del campo y, si es posible, un ejemplo de los datos. Los campos de datos que constituyen un conjunto deben enumerar todos los valores posibles. Por ejemplo, si un conjunto de datos de transacción tiene una columna denominada 'producto' que indica si una transacción fue una recarga, persona a persona, un débito, entonces el diccionario enumeraría todos los valores del producto y describiría que sus respectivos códigos observados en los datos, tales como TUP, P2P, y COT, respectivamente. Para los datos que no están en un conjunto discreto, como el dinero, normalmente se proporciona un valor de rango min-max, junto con su unidad

de medida, tal como el tipo de moneda. Las relaciones con otros conjuntos de datos también deben especificarse, cuando sea posible. Por ejemplo, el campo de datos del número de cuenta de un cliente puede estar presente en conjuntos de datos de transacciones de producto y también en conjuntos de datos KYC. Especificar esta conexión ayuda a comprender cómo se pueden combinar los datos o identificar dónde pueden ser necesarios requisitos de metadatos adicionales para facilitar dicha combinación. El diccionario de datos suele entregarse junto con el informe de inventario de datos, apoyando la discusión de diseño estratégico de un proyecto, la evaluación de riesgo o los requisitos de datos adicionales en sus primeras etapas.

### **Análisis Exploratorios y Libro de Registro**

Este es un conjunto de gráficos, cuadros o datos de tabla que resumen las principales características de una investigación específica o prueba de hipótesis. Todas las estadísticas descriptivas de los datos también se podrían incluir, por ejemplo, promedios, medianas o desviaciones estándar. La parte del análisis exploratorio de identificar tendencias y patrones descubiertos dentro de los datos es necesaria para refinar hipótesis analíticas, contextualizar metadatos o identificar características que se utilizan en un modelo. El análisis exploratorio se realiza como parte de la ejecución inicial del proyecto y, a menudo, continúa hasta completar el proyecto. Los resultados exploratorios

normalmente apoyan los entregables intermedios o las evaluaciones de hitos de proyectos. Estos resultados también se pueden resumir para ayudar a articular el estado y el progreso del proyecto poniendo de relieve las preguntas en la exploración actual, así como las preguntas que ya se han abordado. Un libro de registro de iniciativas exploratorias y conclusiones principales es útil a este respecto.

### **Gráficos de Validación de Modelos y Métricas de Rendimiento**

Para los proyectos de datos basados en modelos, se trata de una lista de gráficos con las métricas de rendimiento más relevantes del modelo predictivo. Véase el Capítulo 2.2.3: Métricas para la Evaluación de Modelos de Datos, que proporciona una lista de los indicadores top 10 utilizados en proyectos de modelado de datos. Estos gráficos y métricas se utilizarán para evaluar la eficacia y la fiabilidad del modelo. Los gráficos de validación pueden incluir los gráficos de ganancia y elevación, y las métricas de rendimiento dependerán del proyecto en particular. Estos pueden incluir, por ejemplo, la prueba de Kolmogorov-Smirnov (KS), la curva característica de funcionamiento del receptor (ROC) o el coeficiente Gini. Esta información es necesaria para evaluar los hitos de la finalización de los objetivos. La aprobación del modelo para el uso de producción o iteración de siguiente paso debe hacerse en términos de estas métricas.

### **Entregables Analíticos: Resultados, Algoritmos, Listas Blancas y Visualizaciones**

Estos son los resultados reales del proyecto. Un proyecto de segmentación de clientes puede incluir una lista blanca de clientes objetivo y sus puntuaciones de propensión asociadas, así como la información de geolocalización posible para asesorar a una campaña de marketing. Un algoritmo de score crediticio ofrece conjuntos de resultados para usuarios especificados en conjuntos de datos de control y tratamiento, y el código para el propio modelo, o una visualización que incluye scripts para representar KPIs y animarlos, así como webscripts u otros componentes para un interfaz de usuario. Cada proyecto tendrá su propio conjunto de entregables matizados. Estos deben definirse como parte del diseño del proceso del proyecto.

### **Informe Final de Análisis y Discusión del Costo-Beneficio de la Implementación**

Este es el informe final que presenta los resultados del análisis, responde a las preguntas y se refiere a los objetivos que se establecieron y acordaron al inicio del proyecto. Esto debe ser entregado junto con los resultados analíticos. Además de discutir la metodología, el proceso, los hallazgos y las soluciones a los desafíos clave, el informe final debe articular la propuesta de valor central de los resultados analíticos. Esto puede incluir: ganancias de eficiencia y ahorros de costos gracias al marketing mejorado basado en datos;

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

previsión de mayores oportunidades de otorgamiento de crédito; o los beneficios de productividad de los paneles de control. El informe final debe considerarse con respecto a la estrategia de implementación del proyecto, para reflejar el costo-beneficio de la propuesta de valor en los resultados analíticos y los requerimientos de recursos para implementarlos a la escala esperada por el proyecto.

### Proceso: Planificación

Las consideraciones siguientes son particularmente relevantes para la planificación de proyectos de datos y para ayudar a especificar el alcance de los entregables intermedios y finales.

### Punto de Referencia

Comprender quién más tenía un problema similar y cómo fue abordado y resuelto es crucial en la planificación de la fase de ejecución. La literatura científica es una fuente inmensa de información y los límites entre la investigación y la aplicación operativa a menudo se superponen en el campo de los datos. Desde la perspectiva de la gestión de proyectos, el benchmarking significa analizar los competidores de negocios y sus actividades en el campo de datos, asegurando que el proyecto esté alineado con las prácticas y operaciones internas de la compañía. En lenguaje sencillo, no reinvente la rueda.

### Métricas y KPIs

Las métricas son los parámetros que impulsan la ejecución del proyecto y

determinan si el proyecto tiene éxito. Por ejemplo: rechazar la hipótesis nula con un objetivo de confianza del 90 por ciento; logrando una tasa de exactitud del modelo de 85 por ciento; o tiempo de respuesta en una decisión de scoring crediticio por debajo de dos segundos. El ajuste de métricas previo evita los riesgos relacionados con la post-validación cuando, debido a umbrales ambiguos, los propietarios de proyectos ofrecen resultados 'bastante buenos'. Esto es a menudo en un esfuerzo por justificar la inversión, o peor aún, afirmando los resultados frente a lo que se creía, insistiendo en que deben funcionar. Véase el Capítulo 2.2.3: Métricas para la Evaluación de Modelos de Datos, que proporciona una lista de los indicadores top 10 utilizados en proyectos de modelado de datos. Las métricas relacionadas con la experiencia del usuario también son importantes, pero deben ser específicas para el contexto del proyecto. Por ejemplo, al evaluar cuánto tiempo es aceptable para que un usuario espere una decisión de scoring crediticio automatizada, entre más rápido mejor. Aun así, debe ser un KPI definido previamente para permitir que el equipo del proyecto entregue un producto bien ajustado.

### Presupuesto y Tiempo

El control de la planificación y la gestión deben tener en cuenta el estado abierto casi permanente de los proyectos de datos. Los objetivos y metas muestran un punto final, pero hasta que se alcanza, un proyecto de datos es a menudo una remodulación continua sobre la base de la

mejora del conocimiento y la definición de problemas. Algunos pueden creer que, si vuelven a ajustarlo de manera diferente, la próxima vez pueden llegar al 85 por ciento. Otros pueden pensar que podrían agregar nuevos datos de clientes para mejorar el modelo. Esta situación fluida no ayuda en la estimación de presupuestos, pero los parámetros de presupuesto deben ser utilizados por los gerentes de proyectos como una herramienta para sintonizar esfuerzos, compromiso y un espacio para probar diferentes hipótesis. Las inversiones iniciales deben comprender este proceso exploratorio e iterativo y sus riesgos. El concepto de escala de producto también ayuda a mitigar este riesgo; comience pequeño, haga iteración hacia arriba. Puede correrse el riesgo de ineficiencias para escalar y refactorizar, pero también mitiga los riesgos presupuestarios, tales como la compra de computadores nuevos para luego descubrir que la hipótesis no se cumple.

La planificación cronológica tiene consideraciones similares a la planificación presupuestaria. Una vez más, hay que buscar un equilibrio entre dar espacio a la exploración y la investigación manteniendo una alineación con las metas y métricas. Una técnica de gestión de proyectos de la industria del software conocida como el 'enfoque ágil' es útil para los proyectos de datos. Este enfoque analiza la progresión del proyecto a través de ciclos auto-sostenible donde la producción es algo medible y comprobable. Esto ayuda a enmarcar una exploración en un ciclo específico.

## Alianzas, Tercerización y Crowdsourcing

Este punto es particularmente importante desde la perspectiva de los recursos del proyecto. Hacer preguntas sobre el diseño de proyectos acerca de los requerimientos y su suficiencia, ayuda a identificar las brechas que los gerentes de proyectos deben llenar. En particular, esto no se limita a los recursos humanos. La computación en la nube es un hardware computacional tercerizado. Incluso los datos pueden ser de origen externo, ya sea mediante la concesión de licencias de los proveedores o mediante el establecimiento de alianzas que permiten el acceso. Crowdsourcing es una técnica emergente para solicitar equipos de datos completos con límites exploratorios muy amplios, generalmente con el objetivo de ofrecer creatividad pura y soluciones innovadoras a un problema fijo a cambio de un incentivo fijo. Como ejemplos, Kaggle es un destacado pionero de la ciencia de datos en forma de crowdsourcing; o el servicio 'Mechanical Turk' de Amazon para tareas o encuestas pequeñas hechas con crowdsourcing.

Un elemento importante a considerar es la PI. Los derechos deben especificarse en acuerdos contractuales. Esto incluye tanto la PI existente, así como la PI creada por medio del proyecto. Considere el proceso completo y la fase de ejecución a lo largo del flujo funcional de datos. La PI abarca más que los resultados finales que se pueden obtener; incluye scripts y códigos informáticos escritos para realizar

el análisis, e incluso conjuntos de datos intermedios, agregados y segmentaciones que se alimentan en otros procesos.

## Gobierno de Datos

Este punto se refiere a cómo y cuándo se utilizan los datos y quién tiene acceso a ellos. La planificación del gobierno de datos debe consultar políticas corporativas más amplias, requisitos legales y políticas de comunicaciones. El propósito del plan es permitir el acceso a los datos al equipo del proyecto y a las partes interesadas en la entrega, al tiempo que se ajusta a las necesidades de protección y seguridad de datos. El plan de gobierno de datos suele verse afectado por la escala del proyecto, donde los proyectos más grandes pueden tener mucho más riesgo que los proyectos más pequeños. Un desafío mayor es que el enfoque de la ciencia de los datos se beneficia del acceso a la mayor cantidad de datos disponibles para tender puentes entre conjuntos de datos y explorar patrones. Mientras tanto, más datos y más acceso también representan más riesgo. El gobierno de datos del proyecto también debe especificar el plan ETL. Esto también abarca el transporte, o la planificación para el movimiento físico o digital, que debe considerar el tránsito por políticas o entornos regulatorios, como por ejemplo, desde una empresa en África a un proveedor externo de analítica en Europa. El plan debe considerar los siguientes principios:

- **Encriptación:** La información sensible o identificadora debe ser encriptada, disimulada o anonimizada y mantenida durante todo el flujo funcional de datos.
- **Permisos:** El acceso a los conjuntos de datos debe definirse en forma granular por funciones de equipo o por punto de acceso (es decir, desde firewalls corporativos, versus desde redes externas).
- **Seguridad:** Los conjuntos de datos situados en el entorno del 'sandbox' del proyecto deben tener su propio dispositivo de seguridad o firewall y la capacidad de autenticar el acceso privilegiado.
- **Registro:** El acceso y el uso deben ser registrados y auditables, habilitados para el análisis y la generación de informes.
- **Regulación:** El plan debe garantizar que se cumplan los requisitos reglamentarios y que las NDA o los contratos legales deben estar en vigor para cubrir todas las partes interesadas del proyecto. Los derechos del cliente y la privacidad también deben ser considerados.

## Proceso: Ejecución

Exactamente como el Data Ring representa un proceso cíclico, la fase de Ejecución en muchos proyectos de datos tiende a reflejar una especie de ciclo dentro del circuito. Lo que generalmente se llama un 'análisis de datos' es en realidad más bien una colección de pasos progresivos e iterativos. Es un camino de exploración y validación de hipótesis hasta que un resultado logra las métricas objetivo definidas.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

La fase de ejecución se asemeja más a los marcos establecidos para el análisis de datos, tales como CRISP-DM u otras adaptaciones.<sup>39</sup> Los gerentes de proyectos que prefieren utilizar un marco de proceso analítico específico, o cuyos proyectos pueden ser mejor atendidos por un enfoque

dado, pueden incorporar fácilmente estos marcos en la especificación de diseño del proyecto de Data Ring durante la fase de ejecución. Los siguientes pasos se proporcionan de otro modo como un proceso general de buenas prácticas para la ejecución de analítica de datos.

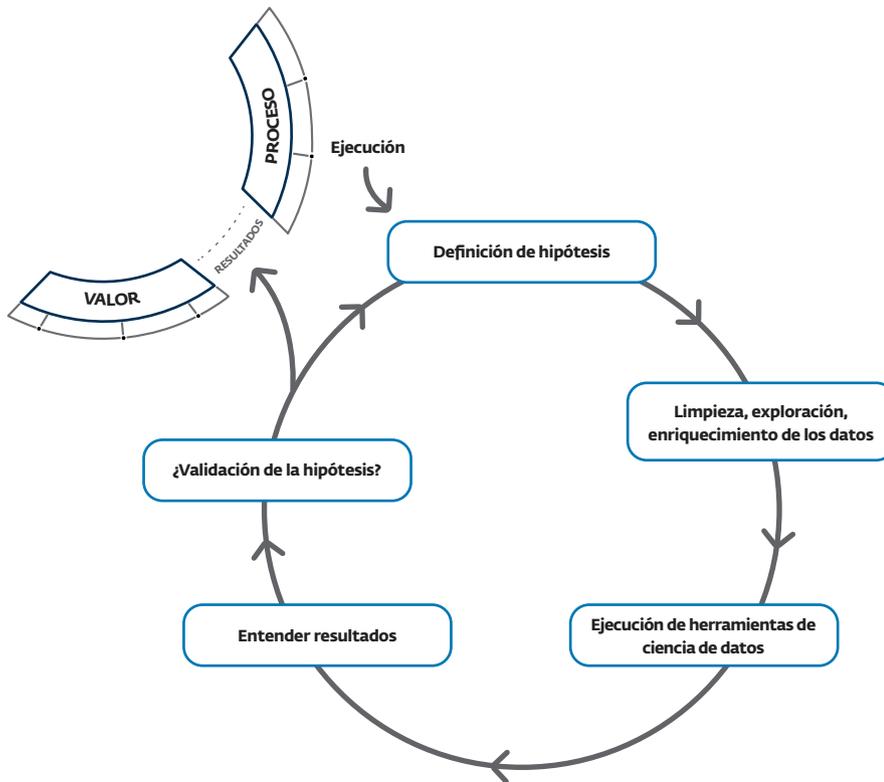


Figura 23: Proceso de ejecución del Data Ring

<sup>39</sup> Los métodos de procesos de analítica de datos relacionados incluyen, por ejemplo: 'Descubrimiento del Conocimiento en el Proceso de Bases de Datos' (Proceso KDD) de Usama Fayyad; 'Muestrear, Explorar, Modificar, Modelar, Evaluar' (SEMMA) por el Instituto SAS; 'Método de Soluciones de Analítica unificado para la Minería de Datos / Analítica Predictiva' (ASUM-DM) de IBM; 'Proceso del Equipo de Ciencia de Datos' (DSTP) de Microsoft

### Limpieza, Exploración y Enriquecimiento de los datos

Este paso es donde el equipo de ciencia de datos realmente comienza. La posibilidad que un conjunto de datos responda perfectamente a las necesidades del estudio es rara. Los datos tendrán que ser limpiados, lo que ha llegado a significar:

- a. Tratamiento:** Convertir los datos en un formato común, compatible con las herramientas de procesamiento.
- b. Entender:** Comprobar los metadatos y la documentación disponible para saber cuáles son los datos.
- c. Validar:** Identificar errores, campos vacíos y mediciones anormales.
- d. Unir:** Integrar descripciones numéricas (legibles por máquina) de eventos realizados manualmente por personas durante el proceso de recolección de datos con el fin de proporcionar una explicación clara de todos los eventos.
- e. Combinar:** Enriquecer los datos con otros datos, ya sea de la misma empresa, del dominio público, o de otro lugar.
- f. Análisis Exploratorio:** Utilizar técnicas de visualización de datos para explorar parcialmente datos y patrones.
- g. Iterar:** Iterar hasta que los errores son contabilizados y que exista un proceso para ir *de manera fiable* desde datos sin procesar hasta datos preparados para proyectos. Este es el proceso mínimo viable.

## Ejecución de Herramientas de Ciencia de Datos

Aquí es donde los científicos de datos aplican sus conocimientos. Se aplican el aprendizaje automático, la minería de datos, aprendizaje profundo, la PLN, la ciencia de redes, estadísticas o (usualmente) una mezcla de lo antes mencionado. Cuando se desarrollan proyectos de datos que incluyen modelos predictivos, es necesario tener una estrategia de validación del modelo antes de ejecutar el modelo. Esto permite probar estadísticamente la hipótesis del proyecto. De manera práctica, el conjunto de datos que impulsa el modelo debe segmentarse en un conjunto de 'control' y un conjunto de 'tratamiento' mediante selección aleatoria. Un reparto de 20 por ciento a 80 por ciento es un enfoque común y básico. El modelo se entrena en el conjunto de tratamiento. Luego, el modelo puede ejecutarse en el conjunto de control y los valores previstos del modelo pueden compararse con los valores conocidos del conjunto de control. Así se calculan las tasas de precisión y cómo se puede probar una hipótesis.

## Comprensión, Interpretación y Representación de Resultados

La interpretación de resultados se discutirá con más detalle en la siguiente sección en términos de aporte de valor al negocio. Pero desde la perspectiva del proceso, la comprensión de resultados se centra en asegurar una alineación entre los resultados obtenidos y el resultado esperado de la ejecución del proceso; y asegurando que sean computacionalmente válidos (es decir, controlando errores aritméticos o

errores de programación). El resultado de cualquier cálculo analítico o proceso, ya sea grande o pequeño, dará:

- Resultados Inutilizables (o incorrectos)
- Resultados Triviales o ya Conocidos
- Resultados Utilizables que Nutren Próximos Pasos
- Resultados Inesperados (a ser investigado con un nuevo flujo funcional, nuevos datos o un nuevo enfoque)

El diseño del proyecto debe reconocer estos posibles resultados y estar preparado para tratar cada caso. Salvo resultados inutilizables, es probable que todas las demás categorías de resultados ameriten una tarea de presentación o de informes para que sea comprensible para otros, incluidos los miembros internos del equipo, los administradores, los clientes y el público en general. Esto generalmente significa un resumen escrito, tabla, gráfico o una animación, que son medios para presentar y explicar los resultados. Los expertos en visualización de datos juegan un papel clave en este proceso, ya que no se trata simplemente de embellecer los resultados. La tarea difícil consiste en crear capas convincentes, interactivas y visuales para agregar brevemente a la narrativa más general del proyecto, lo que debería constituir una declaración del problema del proyecto en sí mismo.

La fase de ejecución también es la oportunidad para reevaluar los planes del proyecto, observando nuevamente que los proyectos de datos se entregan mejor utilizando un enfoque iterativo. La fase de ejecución de un proyecto es lo que pondrá

a prueba el proceso de diseño del proyecto y su enfoque, llevando a que se haga una revisión cuando surja lo inesperado. El marco del Data Ring también puede ayudar analizar problemas de ejecución para identificar soluciones; sus conceptos no se limitan a la planificación inicial. El Data Ring Canvas asociado (discutido en 2.1: Aplicación) está diseñado con esta intención, para proporcionar una plantilla que se puede actualizar continuamente y reflejar el estado del proyecto durante la ejecución.

## Evaluación de las Métricas y Próximos Pasos

La eficacia del proyecto sólo se puede evaluar mediante una definición inicial cuantitativa y precisa de los objetivos y métricas del proyecto. Si los resultados no son satisfactorios, el proceso tiene que empezar de nuevo. Este paso de evaluación e iteración siempre es crítico, pero tiene consideraciones adicionales cuando se utilizan firmas externas. Los entregables pueden considerarse inadecuados a pesar de la calidad del trabajo. La rendición de cuentas de los resultados entregados debe acordarse de antemano, al igual que el grado de margen de maniobra para continuar iterando en busca de resultados satisfactorios. Exactamente cómo su parte en la hipótesis que fija el primer paso de este ciclo de ejecución, los gerentes de proyectos de datos de nuevo desempeñan un papel clave en mantener a los científicos centrados en los objetivos principales y potenciar futuras iteraciones.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

### Cuadrante 4: VALOR

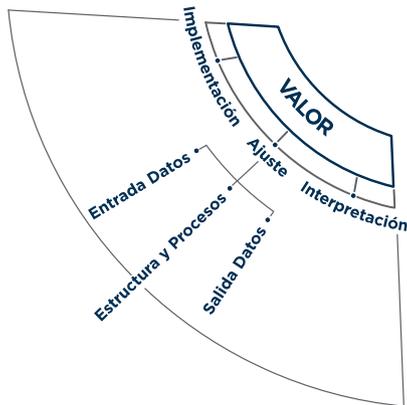


Figura 24: Data Ring Cuadrante 4: VALOR

Valor es la última parte del Data Ring o, por diseño, el punto de partida para futuras iteraciones para agregar o implementar componentes o ampliar el diseño. Esta etapa articula cómo los resultados de la ejecución del proceso se transforman en última instancia en 'información', y luego en 'conocimiento y valor' que se pueden implementar.

Este componente de creación de valor de los resultados suele ser una de las diferencias sustanciales entre un análisis de datos tradicional o un proyecto de BI y un proceso de analítica avanzada, en particular en el área de Big Data. Esto se debe a que los entregables del proyecto rara vez se definen en términos de informes escritos, por lo menos no exclusivamente. Los entregables de los

proyectos de datos suelen caracterizarse por paneles de control, modelos predictivos o palancas de toma de decisiones basadas en datos, herramientas de automatización e, idealmente, conclusiones de negocio poderosas. En otras palabras, un proyecto de datos rara vez termina con recomendaciones. En su lugar, entrega módulos que serán puestos a operar.

#### Valor: Interpretación

El primer paso después de una etapa de ejecución se centra en comprender la propuesta de valor inherente a los resultados y lo que puede ser necesario para refinar estos productos o sus procesos subyacentes para lograr el Objetivo. Un número podría significar nada o todo, dependiendo de la interpretación. Comprender los resultados no es una simple explicación de los fenómenos. En su lugar, significa poner los resultados en el contexto empresarial y abarcar la complejidad de las operaciones reales. Esto también requiere un enfoque transparente, colaborativo, discutiendo los resultados con todas las partes interesadas del proyecto para determinar lo que significan desde todos los ángulos. Teniendo en cuenta el papel de las operaciones de datos (ver, Habilidades de Negocios), no es raro que los científicos de datos puedan tener dificultades para explicar la relevancia operativa de los resultados a los gerentes. Si se hace un hallazgo importante, su valor debe ser comunicado con éxito a la gerencia, quienes pueden llevarlo a la acción.

#### Valor: Ajuste

Comprender los resultados es sólo la tarea inicial. El conocimiento derivado de datos debe convertirse en acciones concretas que se manifiestan mediante herramientas, modelos y algoritmos. Debido al enfoque iterativo, exploratorio de un proyecto de datos, la primera vez que se logra un resultado final exitoso, este invariablemente tendrá imperfecciones que se tienen que ajustar para volverse una herramienta de operación fluida. El ajuste se centra en tres áreas:

#### Datos de entrada

La elección y la calidad de los datos de entrada pueden determinar de manera decisiva la eficacia de los algoritmos utilizados para realizar el análisis. Considérese el aprendizaje automático, donde los algoritmos desarrollan una actitud de aprendizaje siguiendo una fase de entrenamiento que utiliza un subconjunto de datos. Así, al trabajar con datos, las operaciones aprenden progresivamente a recopilar mejores datos. Mejorar los datos sin procesar y minimizar las anomalías, los métodos de recopilación, los insumos manuales y errores de recopilación, derivará en resultados más ajustados a lo largo del tiempo.

#### Infraestructura, Habilidades y Proceso

Después de las primeras iteraciones de ejecución, habrá una mejor comprensión de la efectividad del equipo asignado

al proyecto, los procesos de gobierno de datos, así como las herramientas de software y hardware disponibles. Además, habrá una mayor comprensión de cómo toda la organización del proyecto trabaja en conjunto. Las ineficiencias se revelarán y, como se discutió anteriormente, todas las áreas del proyecto pueden servir como posibles fuentes de solución. Generalmente, el ajuste busca que todos los componentes funcionen mejor juntos. Esto se hace a través de: una mejor organización del equipo; comunicación más fuerte; aumento de las competencias del equipo; y tecnología, ya sea con métodos mejores, mayor poder computacional, o todo lo anterior.

### Datos de Salida

Finalmente, los datos de salida deben ser revisados. Es importante que los resultados de salida no estén sesgados o afectados por errores (humanos o de otro tipo), mala integración entre diferentes etapas del proceso o incluso errores comunes de programación. A menudo, esto significa revisar y fijar los datos de entrada. Aunque el proceso analítico es muy capaz de introducir sus propias anomalías. Esto es tanto un control de validación como una oportunidad de ajuste. En última instancia, la revisión los datos de salida apoya de manera general la organización y la fiabilidad del proceso, por ejemplo, asegurando que una visualización final exhiba los resultados correctos el 100 por cien del tiempo y bajo cualquier condición.

## Valor: Implementación

### Estrategia de Implementación

Para generar un impacto real, la estrategia de implementación debe ser diseñada desde el principio, como parte de la fijación de objetivos. Esta cuestión debe tenerse en cuenta a lo largo del proceso. Evite el riesgo de obtener datos brillantes que no puedan utilizarse en la práctica. Un aspecto clave de la estrategia de implementación es asegurar la adhesión de la gerencia. Presumiblemente, la asignación de recursos proporciona un cierto nivel de compromiso. Dicho esto, debido a que se aseguró a las partes interesadas que no garantizan los resultados en los procesos exploratorios, la estrategia de implementación tiene que asegurar un apoyo continuo y una comunicación fuerte respecto a los hallazgos intermedios.

Los tipos de analíticas, como se discutió en el Capítulo 1.1, también pueden ser relevantes para pensar cómo se usan los resultados:

- **Descriptivo:** Resumir o Agregar Información
- **Diagnóstico:** Identificación de subconjuntos de información basados en criterios específicos
- **Predictivo:** Usualmente construida sobre subconjuntos predictivos, combinados con palancas de decisión
- **Prescriptivo:** Totalmente integrado en sistemas automatizados; una pieza de operaciones

Estos elementos descriptivos pueden guiar la estrategia de implementación, formulando cómo es el caso de uso. Éste es también un componente importante para la generación de adhesión de la gerencia. Por ejemplo, si el caso de uso prevé la automatización completa, las preguntas de diseño del proyecto deben pedir que la infraestructura y los recursos sean suficientes para implementar un algoritmo totalmente automatizado. Si se necesita la inversión en un nuevo centro de datos para ejecutar el algoritmo y entregar las decisiones de crédito justo-a-tiempo, puede resultar difícil la adhesión para asegurar que los resultados del proyecto se utilicen, mientras que una estrategia de caso de uso basada en un piloto de pequeña escala implementado con los recursos existentes podría hacer un caso más fácil.

### Costo-Beneficio

La propuesta de valor anticipada debe articularse en el diseño inicial. Al comienzo, esto puede ser en términos generales, por ejemplo: una ganancia de eficiencia, una reducción de costos o retención de clientes. A medida que el proyecto se desarrolla y los resultados se obtienen y afinan, la propuesta de valor puede ser cuantificada. Una vez alcanzado el objetivo, esto ayudará a definir lo que realmente se ha obtenido y el valor que representa. El mismo proceso debe ser considerado para el uso de los resultados. Al principio, puede preverse alguna infraestructura general o requerimientos del sistema. Una vez que el proyecto esté maduro, el valor se debe comparar con el costo de implementación de la solución.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

### APLICACIÓN: Uso del Data Ring

#### Un Enfoque de Canvas

Como herramienta de planificación, el Data Ring adopta un enfoque de Canvas. Un 'Canvas' es una herramienta utilizada para hacer preguntas estructuradas y diseñar las respuestas de una manera organizada, todo en un solo lugar. Las respuestas son simples y descriptivas; incluso unas palabras bastarán. Desarrollar un Canvas sólido para impulsar la planeación de proyectos puede tomar semanas, ya que la interacción de las preguntas guía desafia la comprensión profunda de los problemas, las soluciones previstas y las herramientas para entregarlas. A continuación, se presenta una lista de las cuatro razones principales para adoptar un *enfoque de Canvas*:

1. Para obligar al dueño del proyecto a establecer una propuesta de valor de proyecto completamente transparente
2. Proporcionar autodiagnóstico y definir y respetar una estrategia interna de gobierno
3. Comunicar una representación completa del proceso 'en una página'
4. Planear de manera flexible con una herramienta que pueda redefinir los componentes a medida que el proyecto evoluciona

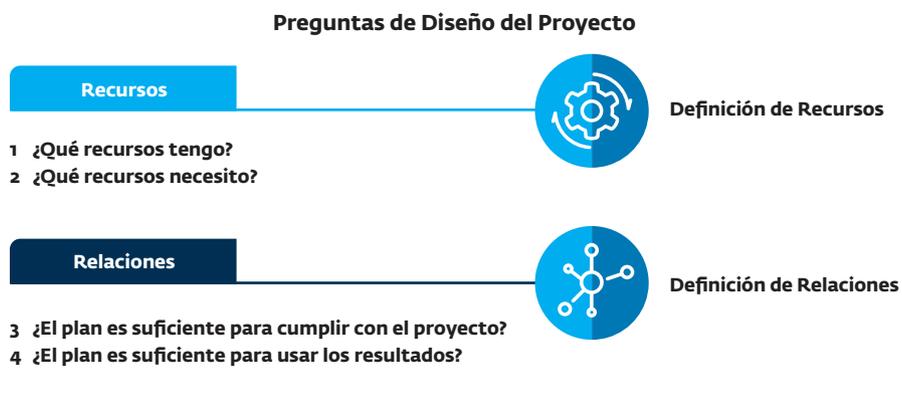
El concepto de Canvas fue presentado por Alex Osterwalder, quien desarrolló el Canvas de Modelo de Negocios [Business Model Canvas]. En los últimos años, se ha vuelto raro asistir a una competencia de startups, concursos de pitch, hackathon, o evento de lluvia de ideas de innovación

sin encontrarse con el Business Model Canvas, y observe a las personas que ponen notas Post-It de colores en tableros, comprometidos con la dura tarea de ofrecer una visión esquemática concisa y comprensiva de su modelo de negocios. La aplicación generalizada del marco entre los innovadores y emprendimientos de tecnología proporciona una base sólida para apoyar las necesidades de gestión de proyectos para proyectos de datos innovadores y orientados a la tecnología. Hay muchos recursos excelentes que proporcionan información adicional sobre el Business Model Canvas, pero no es un prerrequisito para entender o aplicar el Data Ring.

El Data Ring Canvas se inspira en este enfoque, aplicado a los requerimientos específicos de la gestión de proyectos de datos, al tiempo que enfatiza la necesidad de establecer objetivos claros y aplicar las herramientas y habilidades

adecuadas para la implementación exitosa del proyecto. Aquí, un resumen general paso a paso refina las cinco estructuras del Data Ring en términos de sus relaciones interconectadas. El punto es que cada uno de los bloques centrales del anillo representa un componente de un sistema dinámico interconectado. El enfoque iterativo y la aplicación de Canvas permite presentar estos en un diagrama singular para visualizar las piezas del plan integral, identificar las necesidades de recursos y las brechas, y construir un sistema armonioso.

Esto se hace mediante una planificación iterativa, donde primero se debe establecer un objetivo. Una vez se establece la meta, el enfoque va paso a paso alrededor del anillo para articular los recursos, las relaciones y el proceso necesarios para alcanzar la meta. Esto se realiza haciendo cuatro preguntas clave de diseño del proyecto en forma secuencial para cada uno de los bloques centrales. Las preguntas de diseño del proyecto son:



*Figura 25: Las Cuatro Preguntas de Diseño de Proyecto Requeridas por el Data Ring Canvas*

Antes de cerrar esta sección, es importante recordar el error más común cometido al utilizar este tipo de herramientas de negocios: no enfocarse demasiado en completar el Canvas. En pocas palabras, el Data Ring Canvas - al igual que el Business Model Canvas - es sólo un medio, no el objetivo mismo.

## Definición y Conexión de Recursos

### Definición de Recursos

Las dos primeras preguntas identifican las necesidades de recursos del proyecto. Éstos se identifican haciéndose en forma secuencial la primera pregunta guía: ¿Qué datos tengo? ... ¿Qué habilidades están disponibles para el proyecto? ... ¿Qué procesos internos ya están en marcha? ... Las preguntas guía para cada componente se deben considerar para detallar el proceso de planificación. Esto incluye preguntar, ¿Qué valor tengo? Respondiendo, quizás no en términos de resultados ya alcanzados, pero al principio, esto puede ser una pregunta útil y relevante. Puede haber métodos de ajuste que se puedan tomar de proyectos relacionados, o tal vez haya compromisos preexistentes de la gerencia para impulsar la implementación. Estos deben ser considerados entre los Recursos de Valor inicial que impulsan la planificación general.

Una vez que los recursos son delimitados por cada bloque, las preguntas se repiten:

- ¿Qué datos necesito?
- ¿Qué habilidades necesito?
- ¿Qué presupuesto, puntos de referencia, gobierno de datos o plan ETL necesito?

Esto es especialmente crítico para el valor, ya que explorar el valor requerido evidencia la motivación del proyecto. Además, el valor se relaciona con los recursos que son sacados a través de los propios resultados analíticos del proyecto. La planificación de las necesidades de los proyectos en términos de valor también ayuda a definir los entregables intermedios y finales del proyecto, incluyendo el desarrollo de informes o productos de conocimiento. Este enfoque secuencial e iterativo ayuda a identificar las brechas y los requerimientos de adquisición a medida que surgen en los diferentes pasos, construyendo el plan global de forma incremental.

### Conexión de Recursos

Con los recursos especificados para cada bloque estructural, un plan de proyecto debe tener como objetivo entender sus relaciones interconectadas. Las dos últimas preguntas de diseño del proyecto reflejan estas relaciones; es decir, dados los recursos previstos en una categoría de bloque, la necesidad de explorar si los recursos de las otras categorías están suficientemente ligados entre sí. De lo contrario, los requerimientos y vínculos pueden tener que ajustarse entre sí. Estos cuatro enlaces se especifican en la Figura 26: *encaje*, *operaciones del proyecto (ops)*, *resultados*, y *uso*. Cada enlace debe ser especificado para completar el Data Ring Canvas y articular un plan de proyecto integral. Estos se describen a continuación:

## Relaciones de Data Ring



Figura 26: Destacando Vínculos de Recursos en el Data Ring Canvas

### EMPALME: Herramientas y Habilidades

Todos los recursos duros y blandos del proyecto deben ser capaces de trabajar juntos, una relación descrita como *Encaje*. Puede parecer obvio, pero la experiencia práctica muestra que la fase de evaluación de los recursos suele subestimarse. Diferentes piezas de hardware y software tienen que hablar entre sí. La gente también debe hablar, no sólo para comunicarse entre sí dentro del equipo, sino también para utilizar la infraestructura técnica. El Canvas debe especificar la programación primaria [primary scripting] y lenguajes de bases de datos, así como los métodos de marco específicos necesarios para cumplir con el proyecto. En particular, estos lenguajes deben ser comunes entre equipos y herramientas.

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

Las herramientas y habilidades también deben calzar con el objetivo del proyecto. El riesgo principal relacionado con una evaluación incorrecta de los recursos es impulsar componentes avanzados de hardware, soluciones de software totalmente desarrolladas o habilidades humanas (por ejemplo, científicos de datos) al proyecto sin una adecuada integración con las infraestructuras existentes y expertos en el tema. La meta inicial recomendada para un proceso y producto mínimos viables ayuda a mitigar este riesgo mediante el establecimiento de metas en torno a recursos más pequeños; la idea es explorar ideas y probar conceptos de producto. Una vez demostrado, se puede aumentar gradualmente el proceso y el producto con los recursos duros y blandos necesarios para pasar al siguiente nivel.

### OPS: Habilidades y Proceso

Operaciones del proyecto, u *Ops*, es el proceso donde se abordan los cálculos reales y la exploración de datos necesarios para cumplir con el proyecto. Estas actividades son impulsadas por las preguntas analíticas específicas y los problemas operativos que el equipo del proyecto está tratando de resolver. Por ejemplo, un proyecto de scoring crediticio probablemente tendría un problema operativo específico para calcular variables que se correlacionan con las tasas de morosidad de los créditos. Del mismo modo, una visualización podría tener el problema técnico de cómo graficar una red de corresponsales en un mapa. *Ops* observa lo que las personas están haciendo. El bloque de Proceso articula cómo las personas toman medidas en términos de tiempo, presupuesto, requisitos de procedimiento o definición. El vínculo de las

operaciones del proyecto con Habilidades, en sentido de identificar soluciones viables a los problemas operativos, requiere conocimiento relevante sobre el tema. Las operaciones del proyecto (*ops*) en el Canvas deben especificar los problemas operativos centrales del proyecto que deben ser abordados; vinculados a las habilidades necesarias para abordarlos y el proceso para lograrlos.

### RESULTADOS: Proceso y Valor

Los *Resultados* computacionales de la ejecución del proceso se convertirán en valor. El canvas debe enumerar los resultados específicos que se esperan, ya sea un algoritmo, un modelo, una visualización de panel de control, o un informe analítico. El valor se logra a través del proceso de cómo los resultados son interpretados, afinados e implementados. Los métodos de validación de modelos se relacionan con los resultados de los datos del modelo seleccionado. La elección del modelo está vinculada por las definiciones y metas métricas establecidas en el Proceso y la interpretación del negocio, y las implementaciones de uso que crean valor. Los resultados numéricos y su interpretación tienen el riesgo de no poder entender correctamente los resultados obtenidos. También existe riesgo al convertir estos resultados en palancas de decisiones o de negocio que aportan valor. Para garantizar que los resultados sean interpretables según las necesidades del negocio, el canvas debe considerar sus principales entregables y puede incluir recursos adicionales que faciliten la interpretación del valor, como un informe analítico final. Es posible que también sea necesario especificar resultados de datos

adicionales o modelos suplementarios para asegurar una relación sólida entre los bloques de Proceso y de Valor.

### USO: Valor y Herramientas

La cuarta pregunta de diseño de proyecto mira más allá del cumplimiento, hacia lograr conseguir valor a partir del *Uso* del proyecto. El diseño del proyecto debe ser suficiente para utilizar los resultados del producto de datos. Una visualización del panel de control se ejecutará en un computador, por ejemplo, que está conectado a una intranet interna o a la red más general. Un servidor web lo pondrá en línea para que las personas puedan usarlo. Los datos que visualiza se almacenarán en alguna parte, adonde el panel de control debe conectarse y acceder a los datos. El personal de TI mantendrá estos servidores. Estos recursos pueden o no ser identificados en términos de lo que se necesita para cumplir con el proyecto en sí. La cuarta pregunta de diseño del proyecto ayuda a identificar las brechas de implementación que podrían surgir al completar el proyecto, asegurando que estas consideraciones se hagan como parte de la planificación de proyecto. El *Uso* enlaza el *Valor* que el proyecto entrega con las *Herramientas* necesarias para alimentar los datos de salida del proyecto en el sistema de implementación. Esto es especialmente importante para los proyectos que se basan en soluciones tercerizadas, en las que las necesidades de apoyo a la implementación deben ser delimitadas dentro de la contratación inicial. El *Uso* de canvas debe especificar cómo se conecta la estrategia de implementación con las herramientas de implementación.

# CASO 14

## Gestión del Proyecto de Big Data de Airtel Money

Este caso de administración de proyectos se basa en el caso de Airtel Money Uganda presentado en el Capítulo 1.2, Caso 3. Este proyecto fue diseñado y administrado por el equipo de investigación de Inclusión Financiera de IFC con sede en África. El caso, detallado a continuación, describe cada una de las preguntas de diseño del proyecto de Data Ring y considera los detalles de este proyecto. Un Data Ring Canvas terminado refleja este proceso, articulando los recursos clave del proyecto y las relaciones de diseño en una sola visualización. Mientras que este canvas es para un proyecto terminado, el proceso de usar el enfoque de canvas es dinámico; escribir y borrar componentes como desalineaciones obliga a nuevas consideraciones de diseño y requerimientos. Además, el uso de notas adhesivas es un buen enfoque, ya que permiten adiciones fáciles y nuevos elementos de diseño, mientras que también permiten el movimiento en el canvas hasta que se logre un plan satisfactorio.

### **El Establecimiento de Objetivos: Donde comienza el Data Ring**

Un objetivo es una solución para un problema estratégico, y el propósito del proyecto es entregar esa solución. En este ejemplo, el problema era la baja tasa de actividad de Airtel Money. IFC propuso una solución: un modelo para definir el perfil estadístico de un usuario activo y emparejar ese perfil con los no usuarios dentro de la base de suscriptores GSM existente. Una vez identificados, estos clientes podrían ser elegidos como usuarios de alta propensión para Airtel Money. Debido a que se desconocía si era posible que existiera un emparejamiento para el perfil, era importante establecer un ámbito modesto dirigido a una prueba de concepto:

- **El Objetivo:** Desarrollar un modelo mínimo viable de predicción de segmentación de clientes para identificar usuarios activos de alta propensión que permita aumentar las tasas de actividad.

- **La Hipótesis:** Existe una correlación entre la actividad de GSM y el comportamiento de la actividad de Airtel Money (es decir, se pueden crear y emparejar perfiles estadísticos).

### **Identificación de Recursos**

IFC no estaba en posesión de los datos de Airtel previamente, teniendo sólo un compromiso de la alianza con Airtel de proporcionar acceso a los datos de transacciones de CDR y Airtel Money. Aunque tanto IFC como Airtel tienen una **infraestructura sustancial** para sus operaciones, no estaban disponibles para el proyecto. El equipo de IFC encargó a un especialista en operaciones de datos que administrara el proyecto, trayendo **habilidades** relevantes como la **informática, ciencia de datos** y el **negocio de Servicios Financieros Digitales**. Especialistas en Servicios Financieros Digitales de IFC, especialistas en investigación de inclusión financiera y expertos regionales familiarizados con el mercado local y los comportamientos

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

de los clientes apoyaron el proyecto. Durante el proceso de **planificación**, el problema operacional era conocido previamente: baja actividad de Airtel Money. El equipo también tenía datos de referencia existentes de un proyecto de datos similar entregado para Tigo Ghana (véase el Capítulo 1.2, Caso 2: Tigo Cash Ghana, Segmentación), que ayudó a establecer la métrica de la **administración de proyecto**, como un 85 por ciento de precisión sobre el objetivo para el modelo previsto. Las definiciones del modelo también especificaron ‘actividad de 30 días’ como su variable dependiente. Finalmente, el **presupuesto** fue asignado a través del proyecto asesorado por IFC, financiado por la Fundación Bill y Melinda Gates; se estableció un cronograma de seis meses.

### Exploración de Recursos

A través de la alianza del proyecto IFC-Airtel, el equipo negoció acceso a seis meses de CDR histórico y los datos de Airtel Money, aproximadamente un terabyte, para ser extraídos de la bases de datos relacionales de Airtel y entregados en formato CSV. Esto requirió una **infraestructura técnica de Big Data** y las habilidades de ciencia de datos para analizarla. IFC emitió una Solicitud de Propuesta (RFP) para **tercerizar** estos elementos técnicos, para los cuales se seleccionó

Cignifi, Inc. Cignifi trajo: recursos de infraestructura adicionales, con sus grandes Hadoop-Hive clusters para Big Data; experiencia en el sector que trabaja con datos de CDR de MNO; habilidades en R y Python; estadísticas y aprendizaje automático; y recursos para la visualización de datos. El equipo de IFC-Airtel-Cignifi luego definió un **gobierno de datos** y el plan ETL que fue asesorado por los **requerimientos Legales** y de privacidad. Este plan envió al equipo de Cignifi a Kampala, Uganda para trabajar con el equipo de TI de Airtel para: entender sus bases de datos internas; definir los requerimientos de extracción de datos; encriptado y anonimizar datos sensibles; y luego transferir estos datos a un disco duro físico y seguro para ser cargado en los servidores de Cignifi. Las expectativas de valor de los proyectos se especificaron en el RFP para dar una lista con las puntuaciones de propensión del usuario, conocidas como ‘lista blanca’. También se especificaron analíticas adicionales, incluyendo un mapeo de redes sociales y análisis geoespacial.

### Suficiencia del Plan: Entrega

La revisión de suficiencia ayuda a asegurar la alineación entre todos los recursos, procesos y resultados

planeados. Es importante destacar que ayuda a identificar previamente los puntos que anticipan el refinamiento durante el proceso de implementación. También ayuda a reevaluar las áreas de proceso claves cuando se descubren problemas durante la ejecución analítica y estos requieren ajustes del plan.

El plan de **gobernanza de datos** esperaba refinamiento; la fase analítica y de ejecución del proyecto fue de 10 semanas, pero se planificó **relativo** a la fecha de inicio de la adquisición de datos, es decir, los **tiempos del proyecto** se verían afectados por la fecha real y cualquier problema de ETL. El **flujo funcional de datos** también tuvo suficiencia incierta; la planificación del flujo funcional y la asignación de recursos técnicos no fueron posibles hasta que los datos finales se pudieron examinar y se conoció su estructura. Este es un cuello de botella común. Anticipándose a estas incertidumbres, el **valor** agregado especificaba un entregable al principio: un ‘diccionario de datos’ que hablaba de todas las descripciones y relaciones de datos adquiridos, y que se usaría para refinar la suficiencia del proyecto una vez que se conocieran estos detalles. La fase de **ejecución** de cualquier proyecto de datos es donde las sorpresas ponen a prueba

los planes del proyecto. Como esto es de esperar, el proyecto también especificó un informe provisional de datos como una entrega temprana, que proporcionaría estadísticas descriptivas de alto nivel y resultados de análisis exploratorios iniciales, anomalías o brechas en los datos. El informe provisional de datos también incluiría cualquier cosa inesperada que pudiera requerir un ajuste estratégico.

### **Suficiencia del Plan: Implementación**

El objetivo del PMV del proyecto buscaba probar si el enfoque de modelado era relevante para Airtel y el mercado de Servicios Financieros Digitales de Uganda. En este sentido, el plan era suficiente. El proyecto proporcionó un (a) informe final, con hallazgos clave y análisis (b) con una lista blanca: un conjunto de datos de los millones de clientes GSM de Airtel - con un identificador encriptado -, cada uno con una puntuación de propensión asociada sobre la probabilidad del uso activo de Airtel Money.

El plan establecido no era suficiente en el sentido que los recursos fueron pre-asignados para usar la información de la lista blanca en las campañas de marketing, si el análisis resultaba exitoso.

La estrategia de entrega se acordó con la gerencia de Airtel: una reunión final permitiría la presentación y discusión del informe analítico, y el equipo de TI de Airtel tomaría la lista blanca para basar los próximos pasos en los resultados.

### **Ejecución del Proyecto: Ajustes de Planificación**

Las realidades en el terreno requieren el ajuste del plan del proyecto. Se descubrieron los siguientes desafíos durante la ejecución del proyecto y se requirió revisar el plan para asegurar que todas las áreas del proyecto estuvieran trabajando lo suficiente para alcanzar el logro de objetivos.

Una vez asegurado el conjunto de datos inicial, el proceso de flujo funcional de datos encontró irregularidades. El proceso de extracción insertó líneas vacías de alguna manera en el conjunto de datos sin procesar. Aunque los datos pudieron cargarse con éxito, se interpretaron incorrectamente; existían numerosas brechas de datos, aunque este no fuera el caso. Esto requirió cambios en el proceso ETL. El arreglo reveló un error más significativo. El conjunto de datos presentaba serias brechas en el primer mes, y este asunto requería revisar el plan de gobierno de datos,

ETL y diseño general del proyecto. El plan original del proyecto especificó datos de octubre de 2014 a marzo de 2015. La solución fue descartar completamente los datos de octubre y trabajar con Airtel para extraer datos de abril con el fin de mantener la serie de seis meses necesaria para asegurar un modelo estadísticamente fiable. También se descubrió que, según el plan, los datos mismos eran insuficientes. El análisis geoespacial y de red requería datos de ubicación de torres. Se descubrió que los conjuntos de datos de Airtel Money no registraban la ubicación de donde se realizaban las transacciones, sólo la fecha en que se realizaban. El equipo de Cignifi contextualizó estos metadatos creativamente emparejando marcas de tiempo en los datos de Airtel Money con marcas de tiempo de llamadas de voz para los usuarios coincidentes en la data GSM. El equipo usó una ventana de 30 minutos, que proporcionó una coordenada de ubicación que era fiable dentro de un tiempo-distancia de 30 minutos desde la ubicación de la transacción Airtel Money. En la discusión con el equipo de IFC, se acordó que esto era aceptable para que siguiera el análisis, aunque dependía de suponer que la mayoría de las personas, en promedio,

## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

no viajaban grandes distancias en un período de 30 minutos desde el momento en que hacían una transacción de Airtel Money y hacer una llamada telefónica.

La fase de **ajuste** requería una serie de cambios significativos. El resumen de estadísticas de los resultados de la primera ronda parecía inusual para los especialistas en Servicios Financieros Digitales; no coincidían con los patrones de conducta que los expertos en ciencias sociales conocían. Se descubrió que las **definiciones** originales del proyecto habían especificado de manera ambigua 'usuario activo' de manera tal que el equipo de análisis modeló una salida en términos de una transacción de Servicios Financieros Digitales dentro de los 30 días siguientes a la fecha de apertura de cuenta de Airtel Money, en lugar de una transacción dentro de cualquier período de 30 días sobre la totalidad del conjunto de datos. Esto requirió que el diseño

del modelo fuera rehecho. Esto fue en última instancia un beneficio, ya que el análisis inicial también reveló que las transacciones de depósitos y retiro no estaban proporcionando la robustez estadística deseada para lograr las **métricas** de precisión del proyecto. El equipo de IFC-Cignifi acordó rehacer los modelos utilizando los usuarios activos redefinidos y reorientar las transacciones P2P, ya que se consideró que proporcionaban la mayor precisión y, lo que es más importante, definir puntuaciones de propensión para el segmento de clientes generador de los ingresos más altos. Además, se agregó un modelo adicional para 'usuarios muy activos', o aquellos que hicieron transacciones al menos una vez cada 30 días durante un período consecutivo de tres meses. A pesar de ser un grupo pequeño, estos usuarios generaron casi el 70 por ciento de los ingresos totales de Airtel Money; el modelo adicional pretendía identificar a estos clientes de alto valor.

Por último, la interpretación de los **resultados** llevó a un entregable adicional del proyecto: reglas de negocios. Como se mencionó en el caso relacionado con Airtel, los algoritmos de aprendizaje automático del modelo establecieron una serie de variables significativas que eran difíciles de interpretar en un sentido comercial. El equipo de IFC consideró que el entregable a la dirección de Airtel se podía mejorar asegurando que el modelo y las puntuaciones asociadas de propensión de la lista blanca articularan el perfil estadístico de los usuarios activos en términos comerciales que se alinean con los KPIs relevantes para el negocio. Cignifi entregó tres métricas rápidas de segmentación con 'puntos de corte' para perfilar a los usuarios por: número de llamadas de voz por mes; ingresos totales de voz por mes; y duración mensual total de las llamadas de voz.

## Un Canvas Terminado: El diseño del proyecto Big Data de Airtel, utilizando el Data Ring Canvas

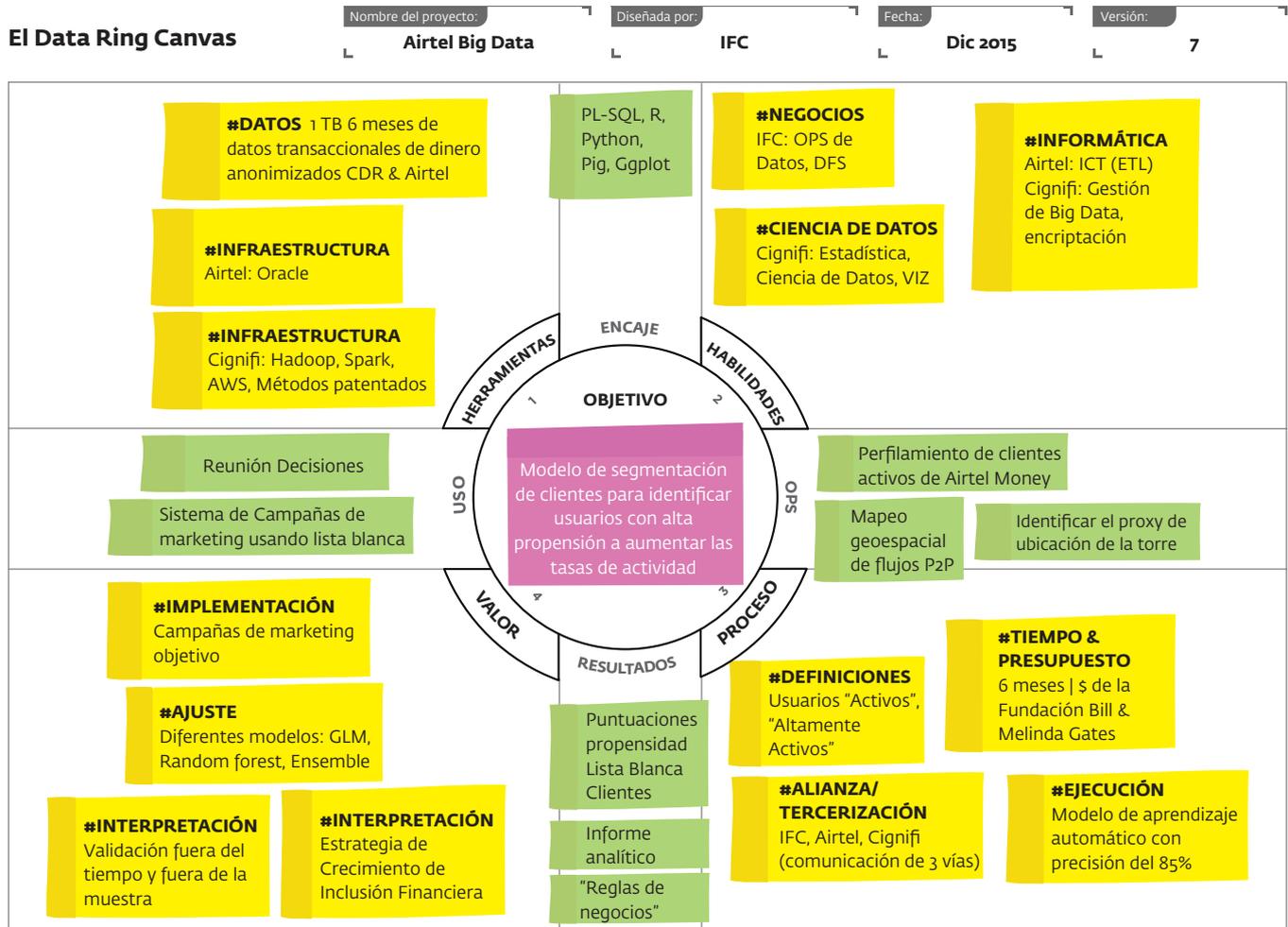


Figura 27: Un Data Ring Canvas terminado para el proyecto Airtel Big Data Fase 1



©2017 International Finance Corporation.

Data Analytics and Digital Financial Services Handbook (ISBN: 978-0-620-76146-8).

Este trabajo fue publicado bajo la licencia de Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License.

El Data Ring Canvas es un derivado del Data Ring de este Manual, adaptado por Heitmann, Camiciotti y Racca bajo licencia (CC BY-NC-SA 4.0).

Ver más en el siguiente enlace: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

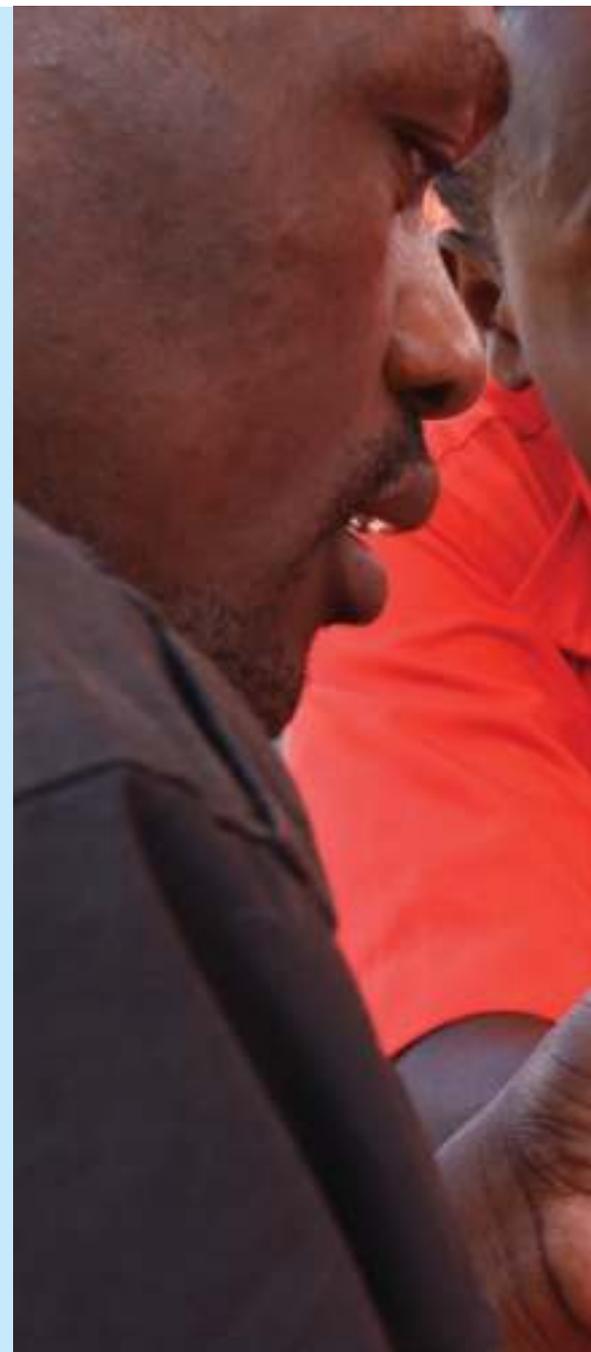
## 2.1\_GESTIÓN DE UN PROYECTO DE DATOS

### Entrega del Proyecto

*La lista blanca modelo identificó a aproximadamente 250.000 usuarios de mayor propensión para ser usuarios activos de dinero electrónico. En la lista blanca completa de varios millones de usuarios de GSM, el 30 por ciento de las puntuaciones de propensión pronosticaron la adopción de usuarios P2P 'altamente activos' para generar unos 1.45 billones de chelines ugandeses con las transacciones P2P; y 4.68 billones de chelines ugandeses por retiros de efectivo, o aproximadamente 1.7 millones de dólares en ingresos anuales adicionales.*

*Los resultados del proyecto fueron fuertes y convincentes. Sin embargo, la estrategia de implementación sólo se definió como punto de decisión. La fecha de entrega coincidió con*

*una campaña de marketing existente, poniendo los resultados de lista blanca en espera. Los suscriptores de Airtel Money crecieron significativamente durante los siguientes meses, lo que disminuyó el valor de la lista blanca, ya que muchos nuevos clientes fueron incorporados a través del marketing habitual. Durante este tiempo, los suscriptores GSM también crecieron, lo que proporcionó millones de usuarios potenciales nuevos de Airtel Money. IFC y Airtel acordaron un análisis de Fase II a finales de 2016. La meta del proyecto es similar, con un componente analítico adicional construido sobre la Fase I, diseñado para examinar los patrones de adopción y distribución de Airtel Money a través del tiempo y la geografía.*







# PARTE 2

## Capítulo 2.2 Recursos

### 2.2.1 Resumen de las Clasificaciones de Casos de Uso Analíticos

Resumen de las Clasificaciones de Casos de Uso Analíticos			
Clasificación	Pregunta Abordada	Técnicas	Implementación
<b>Descriptivo</b>	<ul style="list-style-type: none"> <li>• ¿Qué pasó?</li> <li>• ¿Qué está pasando ahora?</li> </ul>	Alertas, consultas, búsquedas, informes, visualización, paneles de control, tablas, gráficos, narrativas, correlaciones, así como análisis estadístico simple.	Informes
<b>Diagnóstico</b>	<ul style="list-style-type: none"> <li>• ¿Por qué pasó?</li> </ul>	Análisis de regresión, pruebas A/B, correlación de patrones, minería de datos, pronóstico, segmentación	BI Tradicional
<b>Predictivo</b>	<ul style="list-style-type: none"> <li>• ¿Qué ocurrirá en el futuro?</li> </ul>	Aprendizaje Automático, SNA, geoespacial, reconocimiento de patrones, visualizaciones interactivas	Modelamiento
<b>Preceptivo</b>	<ul style="list-style-type: none"> <li>• ¿Qué debe hacerse para que se produzca un determinado resultado?</li> </ul>	Las técnicas incluyen análisis de gráficos, redes neuronales, máquinas y aprendizaje profundo, IA	Soluciones Integradas, Decisiones Automatizadas

## 2.2.2 Directorio de Fuentes de Datos

Fuente: Sistema de Core Bancario y MNO		
Estructura: Típicamente, datos estructurados, usando bases de datos relacionales.		
Formato: Datos digitales, que pueden ser extraídos en varios formatos para reportes o análisis. Los datos heredados pueden incluir vinculaciones en papel o formularios de vinculación escaneados.		
Nombre	Datos	Ejemplos
Datos de Facturas de los Clientes	Duración del contrato; historial de pagos; tipos de compra	Marketing insights mejorados; potencial para crear score crediticio usando datos de facturas
Estado de Registro del Cliente	Estado de registro (por ejemplo, activo, inactivo, nunca usado)	Marketing insights; monitoreo del desempeño del negocio; cumplimiento de reglamentación
KYC del Cliente	Nombre, dirección, fecha de nacimiento, sexo, ingresos	Marketing insights; cumplimiento de reglamentación
Estado de la Cuenta	Tipo de cuenta, estado de actividad (activo, inactivo, envejecimiento de la actividad, inactivo con saldo)	Marketing insights; monitoreo del desempeño del negocio; cumplimiento de reglamentación
Actividad de Cuenta	Saldo de la cuenta, velocidad mensual, saldo promedio diario	Marketing insights; scoring crediticio; cumplimiento regulatorio
Datos de Transacciones Financieras (directos)	Volumen y valor de depósitos, retiros, pagos de facturas, transferencias u otras transacciones financieras	Monitoreo del desempeño del negocio y financiero; cumplimiento regulatorio; marketing insights; scoring crediticio
Datos de Transacciones Financieras (indirectos)	Transacciones fallidas; transacciones declinadas; canales utilizados; hora del día	Asuntos de desempeño del producto y problemas de diseño del producto; necesidades de capacitación y comunicaciones
Datos de Dinero Electrónico	Fondo de dinero electrónico, conciliaciones, transferencias de fondos entre corresponsales	Administración del desempeño del corresponsal; fraude y gestión de riesgo
Actividades No Financieras	Cambio de PIN; solicitud de saldo; solicitud de estado de cuenta	Marketing insights; mejoras en la eficiencia; desarrollo de productos
Originación de Créditos	Tipo de crédito, monto del crédito, colateral utilizado, duración, tasa de interés	Marketing insights; monitoreo de desempeño de la cartera; scoring crediticio; evaluación de créditos nuevos
Actividad Crediticia	Saldo de crédito, estado del crédito, fuente de la transacción de pago de crédito	Marketing insights; monitoreo de desempeño de la cartera; scoring crediticio; evaluación de créditos nuevos

## 2.2\_RECURSOS

### Fuente: Sistema de Dinero Electrónico

Estructura: Típicamente, datos estructurados, usando bases de datos relacionales.

Formato: Datos digitales, que pueden ser extraídos en varios formatos para reportes o análisis. Los datos heredados pueden incluir vinculaciones en papel o formularios de vinculación escaneados.

Nombre	Datos	Ejemplos
Del Cliente	Nombre, dirección, fecha de nacimiento, sexo, ingresos	Marketing insights; cumplimiento regulatorio
Estado de Registro	Estado de actividad (activo, inactivo, envejecimiento de la actividad, inactivo con saldo)	Marketing insights; monitoreo del desempeño del negocio; cumplimiento regulatorio
Actividad de Billetera Electrónica	Saldo de billetera electrónica, velocidad mensual, saldo promedio diario	Marketing insights; scoring crediticio; cumplimiento regulatorio
Datos de Transacción	Volumen y valor de depósitos, retiros, pagos de facturas, P2P, transferencia, recarga de tiempo aire u otras transacciones financieras	Monitoreo del desempeño empresarial y financiero; cumplimiento regulatorio; Marketing insights; scoring crediticio
Datos de Dinero Electrónico	Fondo de dinero electrónico, conciliaciones, transferencias de fondos entre corresponsales	Administración del desempeño del corresponsal; fraude y gestión de riesgo

### Fuente: Sistema de Gestión de Corresponsales

Estructura: Típicamente, datos estructurados, usando bases de datos relacionales.

Formato: Datos digitales, que pueden ser extraídos en varios formatos para reportes o análisis. Los datos heredados pueden incluir registros en papel, formularios de registro escaneados, o reportes de monitoreo de corresponsales o informes de desempeño.

Nombre	Datos	Ejemplos
Actividades del Corresponsal (directo)	Volumen y valor de transacciones de corresponsales; conciliación de transferencias; depósito y retiro de fondos; saldo del fondo; días sin fondos	Insights de marketing y ventas; scoring crediticio; gestión del desempeño de corresponsales
Actividades de Corresponsales (indirectas)	Cambio de PIN; solicitud de saldo; solicitud de estado de cuenta; crear asistente nuevo	Insights de marketing y ventas; gestión del desempeño de corresponsales
Actividades del Comerciante (directas)	Volumen y valor de transacciones del comerciante; número de clientes únicos	Insights de marketing y ventas; scoring crediticio; gestión del desempeño de comerciantes
Actividades del Comerciante (indirectas)	Cambio de PIN; solicitud de saldo; solicitud de estado de cuenta; crear asistente nuevo	Insights de marketing y ventas; gestión del desempeño de comerciantes
Datos Técnicos del Sistema	Número de TPS; listas de transacciones; tiempo de procesamiento	Planeación de capacidad; Supervisión del desempeño versus SLA; identificar problemas técnicos de desempeño
Informes de Visitas a Corresponsales y Comerciantes por Personal de Ventas	Presencia de materiales promocionales; conocimientos de asistentes; tamaño del fondo de caja; pueden incluir más frecuentemente datos semiestructurados o no estructurados, tales como informes de monitoreo en papel	Insights del cliente; gestión del desempeño de corresponsales

### Fuente: Sistema de Gestión de Relacionamento con el Cliente (CRM)

Estructura: A menudo incorporan datos tanto estructurados como semiestructurados que utilizan bases de datos relacionales o sistemas de almacenamiento basados en archivos, como grabaciones de voz o resúmenes de problemas etiquetados por categorías estructuradas.

Formato: Por lo general, datos digitales, aunque datos semiestructurados y no estructurados pueden no estar disponibles para la generación de informes (por ejemplo, para grabaciones de voz, si están disponibles).

Nombre	Datos	Ejemplos
Registros de Call Center	Registro de problemas, tipo de problemas, tiempo de resolución (puede incluir datos semiestructurados en informes)	Insights de los clientes; gestión de operaciones y del desempeño; mejoras al sistema
PBAX	Número de llamadas al Call Center; duración de las llamadas; tiempos de llamada en espera; llamadas perdidas	Gestión de Operaciones y del Desempeño
Datos de Retroalimentación de Atención al Cliente	Número de llamadas; estadísticas de tipo de llamada; estadísticas de resolución de problemas	Identificar: desempeño técnico y problemas de diseño del producto; necesidades de capacitación y comunicación; problemas de terceros (por ejemplo, corresponsal, facturador)
Datos de Retroalimentación del Corresponsal y del Comerciante	Número de llamadas de corresponsal o comerciante; estadísticas del tipo de llamada; estadísticas de resolución de problemas	Identificar: desempeño técnico y problemas de diseño del producto; necesidades de capacitación y comunicaciones de corresponsales; problemas del cliente
Interacciones del Canal de Comunicación	Volumen de visitas a sitios web, volúmenes del call center, consultas en redes sociales, solicitudes de chat en vivo	Insights de los clientes; administración de operaciones y del desempeño; mejoras al sistema
Datos de Comunicación Cualitativa	Tipo de consultas, satisfacción del cliente, análisis de redes sociales	Insights de los clientes

### Fuente: Registros de Clientes

Estructura: Incorporando a menudo datos estructurados, semi-estructurados y no estructurados, que van desde: Documentos de KYC que pueden incluir variedad de información personal dependiendo del tipo de documento; a encuestas de mercado o de clientes; hasta notas de grupos focales.

Formato: Se pueden usar una amplia variedad de formatos para almacenar datos de registro de clientes, incluyendo bases de datos relacionales, sistemas de almacenamiento de archivos o documentos en papel o escaneados.

Nombre	Datos	Ejemplos
Documentos de KYC	Identificación; certificado de salario; certificado de domicilio	Cumplimiento regulatorio; segmentación demográfica y geográfica
Formularios de Inscripción y de Solicitud	Abrir cuenta de Servicios Financieros Digitales; solicitud de crédito	Cumplimiento regulatorio; segmentación demográfica y geográfica
Investigación Cualitativa	Entrevistas con los clientes; grupos focales	Insights de marketing y productos
Investigación Cuantitativa	Estudios de conocimiento y uso; estudios de sensibilidad de precios; pruebas piloto	Insights de marketing y productos

## 2.2\_RECURSOS

### Registros de Corresponsales y Comerciantes

Estructura: Incorporando a menudo datos estructurados, semi-estructurados y no estructurados, que van desde: Documentos de KYC que pueden incluir variedad de información personal dependiendo del tipo de documento; hasta encuestas de marketing o de comerciantes; hasta notas de grupos focales.

Formato: Se puede usar una amplia variedad de formatos para almacenar datos de registros de corresponsales o comerciantes, incluyendo bases de datos relacionales, sistemas de almacenamiento de archivos o documentos en papel o escaneados.

Documentos de KYC	Certificados de constitución; declaraciones de impuestos; documentos KYC; extractos de cuentas bancarias	Cumplimiento regulatorio; segmentación demográfica y geográfica
Formularios de Registro	Registrarse como corresponsal o comerciante de Servicios Financieros Digitales	Cumplimiento regulatorio; segmentación demográfica y geográfica
Investigación Cualitativa	Entrevistas a los corresponsales; grupos focales	Insights de marketing, productos y ventas
Investigación Cuantitativa	Investigación de comprador secreto	Insights de marketing, productos y ventas

### Fuente: Alianzas con Terceros

Estructura: El tercero puede adoptar cualquier forma o estructura, dependiendo del contenido, fuente y proveedor que lo proporciona.

Formato: Los formatos pueden variar desde formatos .CSV comunes a APIs de acceso propietario y métodos de entrega.

Nombre	Datos	Ejemplos
Datos de Facturas Acerca de Clientes (empresas de servicios públicos)	Duración del contrato; historial de pagos; tipos de compra	Marketing insights mejorados; potencial para crear score de crédito usando datos de facturas
Datos del Pagador Acerca del Cliente (empleador, gobierno)	Historial de nómina; duración de pagos habituales	Marketing insights mejorados; scoring crediticio
Repositorios de Información del Cliente (por ejemplo, burós de crédito, listas de vigilancia, registros policiales)	Datos KYC; calificación crediticia; actividad fraudulenta previa	Scoring; investigaciones de fraude; gestión de riesgo
Datos geoespaciales (datos satelitales)	Demografía regional; densidad de población; topografía; infraestructuras como carreteras y electricidad; puntos de acceso financiero	Insights del mercado; gestión de corresponsales
Medios de Comunicación y Redes Sociales	Tipo y frecuencia de actividades de red; información personal; número de conexiones; tipo de conexiones	Insights del mercado; scoring crediticio

## 2.2.3 Métricas para la Evaluación de Modelos de Datos

Lista Top-10 de métricas de desempeño para evaluar modelos de datos	
Métrica	Definición
Curva Característica de Operación del Receptor (ROC)	La curva ROC se define como representación gráfica entre la tasa positiva real y la tasa de falsos positivos. Esta ilustra el desempeño del modelo a medida que su umbral de discriminación se varía. Cuanto mayor sea el área entre la curva ROC y la línea base, mejor será el modelo.
AUC	Área Debajo de la Curva (AUC) mide el área bajo la curva ROC. Proporciona una estimación de la probabilidad de que la población esté clasificada correctamente. Representa la capacidad del modelo para producir una buena clasificación de instancia relativa. Valor igual a uno es un modelo perfecto.
KS	La prueba estadística Kolmogorov-Smirnov (KS) mide la separación vertical máxima entre la distribución acumulada de buenos y malos. Representa la capacidad del modelo para separar la población de interés 'buena' de la población 'mala'.
Lift Chart	Mide la efectividad de un modelo predictivo calculado como la relación entre los valores predictivos positivos sobre el número de positivos en la muestra para cada umbral. Cuanto mayor sea el área entre la curva Lift y la línea base, mejor será el modelo.
Ganancias Acumuladas	Mide la efectividad de un modelo predictivo calculado como la relación entre los valores predictivos positivos sobre el número de positivos en la muestra para cada umbral. Cuanto mayor sea el área entre la curva de ganancias acumuladas y la línea base, mejor será el modelo.
Coefficiente GINI	El coeficiente Gini está relacionado con el AUC; $G(i) = 2AUC - 1$ . También proporciona una estimación de la probabilidad de que la población esté clasificada correctamente. Valor igual a uno es un modelo perfecto. Esta es la definición estadística de lo que impulsa el Índice Gini económico para la distribución del ingreso.
Exactitud	La exactitud es la capacidad del modelo para hacer una predicción correctamente. Se define como el número de predicciones correctas sobre todas las predicciones hechas. Esta medida sólo funciona cuando los datos están equilibrados (es decir, la misma distribución para lo bueno y lo malo).
Precisión	La precisión es la probabilidad de que una instancia seleccionada al azar sea positiva o buena. Se define como la proporción entre el total de instancias positivas verdaderas previstos y el total de instancias positivas previstos.
Sensibilidad	Sensibilidad es la probabilidad de que una instancia seleccionada al azar sea buena o positiva. Se define como la proporción entre el total de instancias positivas verdaderas y el total de instancias positivas.
Error Cuadrático Medio de Predicción (RMSEP)	El RMSEP es una medida de la diferencia entre los valores previstos por un modelo y los valores realmente observados. La métrica se utiliza en predicciones numéricas. Un buen modelo debe tener un RMSEP pequeño.

## 2.2.4 Data Ring y Data Ring Canvas

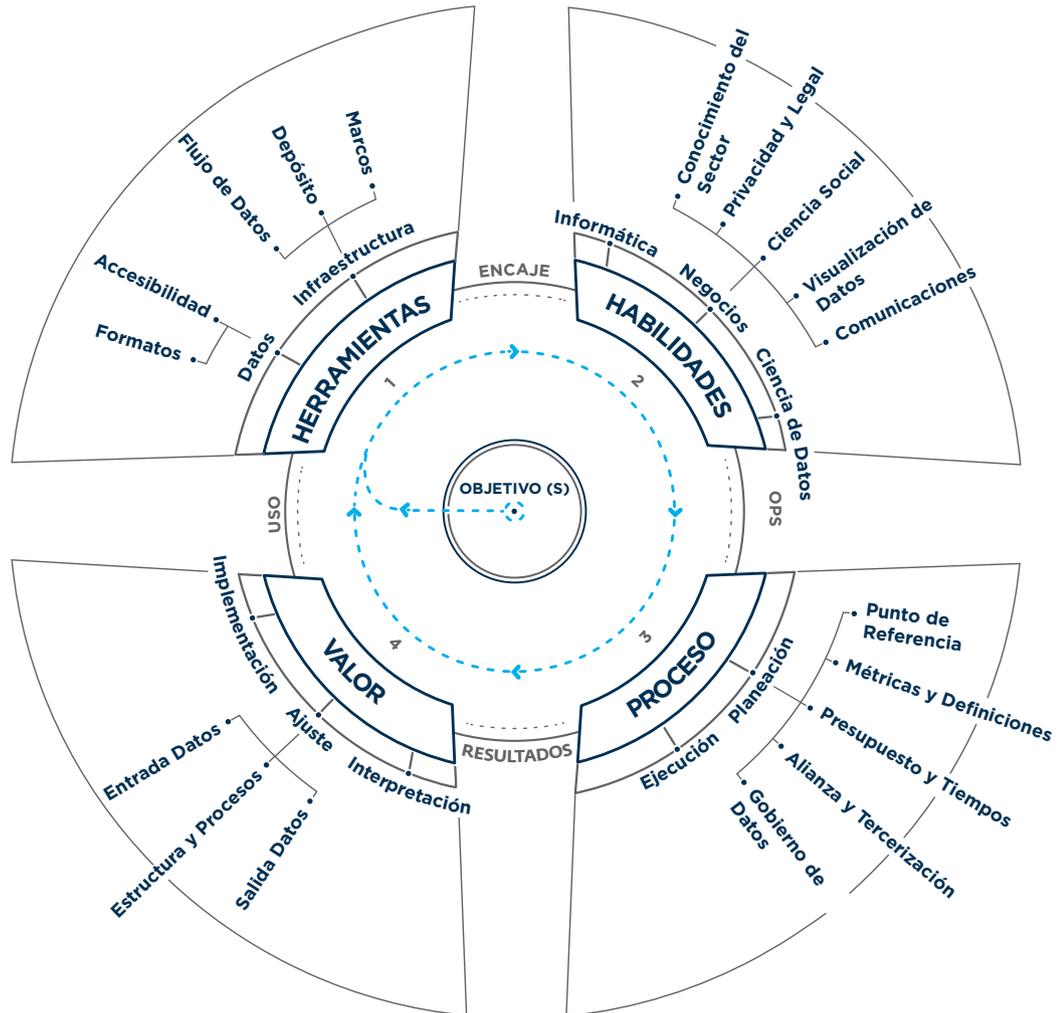
Las herramientas de Data Ring y Data Ring Canvas también están disponibles para descargar desde el sitio web de The Partnership for Financial Inclusion aquí: [www.ifc.org/financialeinclusiografica](http://www.ifc.org/financialeinclusiografica)

La siguiente página desprendible ofrece una copia del Data Ring y el Data Ring Canvas, para ser utilizada.

## 2.2\_RECURSOS



## El Data Ring



©2017 International Finance Corporation.

Data Analytics and Digital Financial Services Handbook (ISBN: 978-0-620-76146-8).

Este trabajo fue publicado bajo la licencia de Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License.

El Data Ring fue adaptado de Camiciotti y Racca, 'Creare Valore con i BIG DATA'. Edizioni LSWR (2015) bajo licencia (CC BY-NC-SA 4.0).

Ver más en el siguiente enlace: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

## 2.2\_RECURSOS

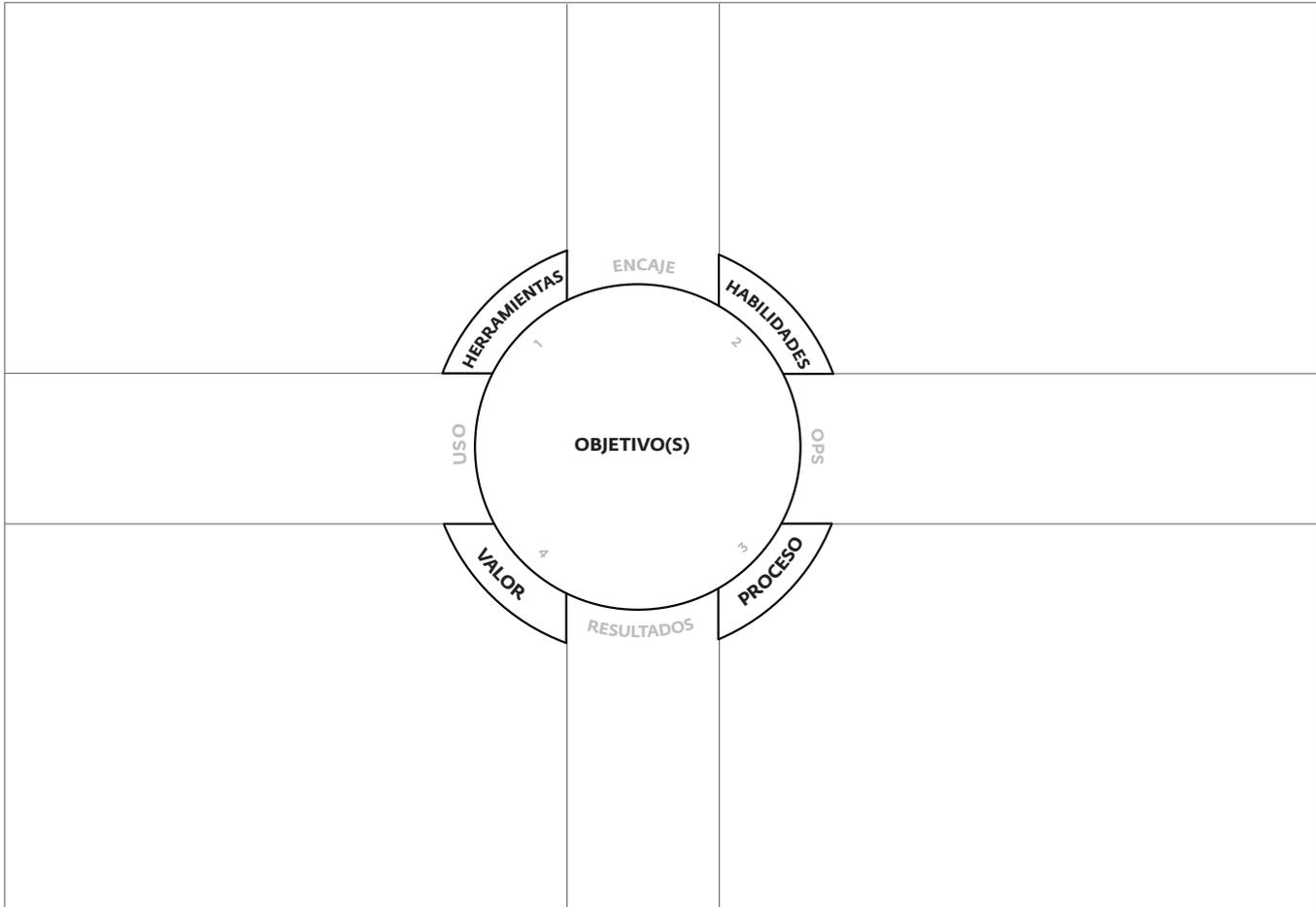
### El Data Ring

Nombre del proyecto:

Diseñada por:

Fecha:

Versión:



©2017 International Finance Corporation.

*Data Analytics and Digital Financial Services Handbook (ISBN: 978-0-620-76146-8).*

Este trabajo fue publicado bajo la licencia de Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License.

El Data Ring Canvas es un derivado del Data Ring de este Manual, adaptado por Heitmann, Camiciotti y Racca bajo licencia (CC BY-NC-SA 4.0).

Ver más en el siguiente enlace: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

---

# Conclusiones y Lecciones Aprendidas

El universo de datos se está expandiendo cada hora. La capacidad analítica computacional también se está volviendo más y más avanzada, y el costo del almacenamiento de datos está disminuyendo. El potencial analítico de datos descrito en este manual - y en estos casos - resalta cómo los proveedores de Servicios Financieros Digitales pueden aprovechar todo tipo de datos para construir nuevos servicios y lograr mayores eficiencias en sus operaciones actuales mediante la incorporación de enfoques basados en datos. Los profesionales en campo deben esforzarse por adoptar un enfoque basado en datos en sus negocios. Esto traerá mayor precisión a sus actividades y un enfoque basado en la evidencia para la toma de decisiones.

## Construyendo una Cultura Orientada a Datos

La cultura organizacional es crucial. Las organizaciones deben fomentar un entorno favorable a los datos en el que se fomente el poder de los datos, y donde las personas estén empoderadas y estimuladas a explorar para encontrar formas de mejorar los resultados. Como resultado, existe la necesidad de invertir en habilidades operativas de equipo, herramientas e ideas con el fin de hacer justicia a los datos. El liderazgo organizacional debe articular claramente la visión y los estándares fundamentales que formarán la base de su programa de gestión de datos. El liderazgo también debe formar un fuerte compromiso para desarrollar la capacidad de datos de la compañía, tanto en términos de visión

como de presupuesto. Además, es esencial que haya un departamento claramente definido o un individuo con influencia dentro de la organización que conduzca el proceso. Algunas organizaciones que están avanzadas en la curva de madurez han optado por crear una posición de alto nivel llamada Chief Data Officer o Director de Datos (CDO); esta persona trabaja en estrecha colaboración con la dirección para administrar todos los datos relacionados con la estrategia y la administración.

La organización debe considerar sus capacidades y experiencia actuales para articular claramente el futuro. Consideraciones importantes incluyen el tamaño de la organización, así como los recursos de TI existentes, tales como habilidades y experiencia. Además, pasar a un enfoque basado en datos implicará grandes cambios para la cultura organizacional, específicamente en torno a cómo se comparten los datos y cómo se toman las decisiones. La organización deberá estar preparada para brindar apoyo continuo durante el cambio y estar preparada para gestionar las expectativas del personal y de la gerencia. Los niveles actuales de madurez de la gestión de datos también son importantes. El proveedor de SFD puede querer examinar las fuentes de datos actuales, el marco de informes y el uso de los datos en la toma de decisiones para situarse en la curva de madurez. Entender dónde se encuentra en la escala de madurez de la gestión de datos ayudará

al proveedor a desarrollar una hoja de ruta que conduzca hacia el objetivo deseado.

Orientarse hacia los datos también incluye revisar las habilidades del personal existente y evaluar los niveles de confort de los miembros del equipo en tecnología e informática. El personal existente puede ser capacitado para manejar nuevas tecnologías. Están en una posición ideal para aplicar las nuevas tecnologías a los viejos problemas porque ya conocen la organización, su mercado y sus desafíos. Normalmente, el personal necesitará capacitación y entrenamiento en campo en gestión de datos. El proveedor de SFD tal vez desee identificar a los miembros del personal que tienen una aptitud y una actitud adecuadas para adoptar nuevas prácticas habilitadas por tecnología, y luego preparar un plan para el desarrollo intensivo de habilidades.

No importa en qué nivel esté una organización en su adopción de análisis basados en datos, hay margen para incorporar sistemáticamente datos en sus procesos y toma de decisiones. Los profesionales en campo pueden tomar pequeños pasos para comenzar a rigurosamente probar las necesidades y preferencias de sus clientes, para monitorear el desempeño internamente y entender el impacto en sus actividades de negocio. Más importante aún, las metas que establece una organización para el seguimiento del desempeño del negocio deben ser cuantificables y medibles.

## 2.2\_RECURSOS

### Toda los Datos son Buenos Datos

La analítica de datos ofrece a los proveedores de SFD la oportunidad de obtener una comprensión mucho más granular de sus clientes. Esta información se puede utilizar para diseñar mejores procesos y procedimientos que se alineen con las necesidades y preferencias del cliente. La analítica de datos trata sobre la comprensión de los clientes, con el objetivo de que el cliente obtenga mayor valor del producto.

En particular, la combinación de conocimientos de diferentes metodologías y fuentes de datos puede enriquecer el entendimiento. Como ejemplo, aunque los datos cuantitativos pueden proporcionar información sobre lo que está sucediendo, los datos cualitativos y la investigación aclararán por qué está sucediendo. De manera similar, varios proveedores de SFD han utilizado una combinación de modelado predictivo y análisis de geolocalización para identificar las áreas objetivo en las que deben concentrar sus esfuerzos de marketing.

Para el vasto mercado masivo que los proveedores de SFD atienden, en muchos casos no puede haber historial financiero formal o historial de datos de pagos para usar como base. En estas situaciones, los datos alternativos puede permitir a los proveedores de SFD verificar los flujos de efectivo a través de información de variables sustitutivas (proxy), tales como los datos del MNO. Aquí, los proveedores de SFD tienen la opción de trabajar directamente con un MNO o con un proveedor. La decisión depende de los mercados respectivos, así como de la preparación de la institución. Muchos proveedores pueden no tener el

conocimiento técnico para diseñar modelos de scoring basados en datos del MNO – en este caso, asociarse con un proveedor que proporciona este servicio es una buena opción.

### Uso de la Visualización de Datos

Una imagen vale más que mil palabras, o quizá mil números. El uso de visualizaciones para ilustrar gráficamente los resultados de los informes de gestión de datos estándar puede ayudar a la toma de decisiones y el monitoreo. Las representaciones gráficas permiten a la audiencia identificar las tendencias y los valores o las desviaciones rápidamente. Esto es cierto con respecto a los equipos internos de ciencia de datos que están explorando los datos, y también para comunicaciones más amplias, cuando las tendencias de los datos y los resultados pueden tener más impacto que las tablas visualizando relaciones o conclusiones basadas en datos.

Un cuadro o una gráfica es una visualización de datos, en su sentido más básico. Dicho esto, la 'visualización' como concepto y disciplina emergente es mucho más amplia, tanto con respecto a las herramientas disponibles como con los resultados posibles. Por ejemplo, una infografía puede ser una visualización de datos en muchos contextos, pero no es necesariamente una representación gráfica [plot]. En algunos casos, esta amplitud también puede incluir medios mixtos. Un pionero en esta área, por ejemplo, es Hans Rosling, cuyo trabajo combinando la visualización de datos con la narración interactiva de medios mixtos le valió un lugar en la lista de las 100 personas más influyentes de la revista Time.<sup>60</sup> Estos elementos de dinamismo e interactividad

han elevado el campo de visualización de datos muy por encima de cuadros y gráficos, aunque el campo también abarca estas herramientas más tradicionales.

La visualización de datos está relacionada con, pero separada, de los paneles de control. Un panel de control probablemente incluiría una o más visualizaciones discretas. Los paneles de control son puntos de referencia de acceso, que a menudo sirven como puntos de entrada a datos más detallados o herramientas de reportes. Aquí es donde se visualizan los KPIs para proporcionar información instantánea, típicamente para los administradores que necesitan una instantánea concisa del estado operacional. Los paneles de control simples se pueden implementar en Excel, por ejemplo. Por lo general, el concepto de panel de control se refiere a representaciones de datos más sofisticadas, incorporando las ideas de interactividad y dinamismo que abarca el concepto más amplio de visualización de datos. Además, los paneles de control más sofisticados probablemente incluirán datos en tiempo real y respuesta a las consultas de los usuarios. Mientras que la visualización de datos y los paneles de control de datos están inherentemente relacionados y a menudo se superponen, también es importante reconocer que son conceptualmente diferentes y juzgados por criterios diferentes. Hacer esto ayuda a certificar que las herramientas adecuadas se aplican para el trabajo correcto y garantiza que los proveedores y los productos se adquieran para los fines previstos.

### La Ciencia de los Datos es el Arte de los Datos

El capítulo 1 anotó la historia de la 'ciencia de datos' como término. Curiosamente,

<sup>60</sup> Hans Rosling. En Wikipedia, la Enciclopedia Libre, consultada el 3 de abril de 2017, [https://en.wikipedia.org/wiki/Hans\\_Rosling](https://en.wikipedia.org/wiki/Hans_Rosling).

aquellos que lo acuñaron vacilaron respecto de llamar a los expertos en el área de esta disciplina 'científicos de datos y 'artistas de datos'. Aunque que la ciencia de los datos ganó el título oficial, es importante reconocer que la creatividad, el diseño y la sensibilidad artística siguen siendo críticos para el campo. Después de la discusión anterior sobre la visualización de los datos, el proceso de convertir bits de datos en una información informativa, interactiva, estéticamente agradable y visualmente atractiva requiere tanto habilidades técnicas como ideas creativas. En referencia a Rosling, el proceso de hacer de la visualización de datos el personaje principal en lo que más correctamente se puede describir como una representación teatral evidencia aún más la interacción entre la ciencia de los datos y el arte de los datos. El papel de los científicos de datos, independientemente del título funcional, es recurrir a la habilidad técnica y la intuición creativa para explorar patrones, extraer valor de esas relaciones y comunicar su importancia.

Este dualismo de organización estructurada y patrones emergentes describe una de las complejidades generales de muchos proyectos de datos. Por otra parte, hay la necesidad de objetivos claros, arquitectura definida y experiencia precisa para asegurar que la entrega del proyecto sea puntual y dentro del presupuesto. Por otro lado, existe la necesidad muy importante de una flexibilidad abierta que permita descubrir patrones, explorar nuevas ideas, minería de datos para descubrir posibles anomalías, probar hipótesis y diseñar creativamente visualizaciones para contar la historia de los datos.

## Industria Global

El campo de la ciencia de los datos ha existido menos de una década, y el término mismo solo llega a prominencia en el 2008 (véase la Figura 6 en la Parte 1). Desde entonces, los teléfonos inteligentes se han vuelto ubicuos, el poder computacional ha crecido sustancialmente y los costos de almacenamiento se han desplomado. Las empresas de tecnología han introducido nuevos productos que se han asimilado rápidamente a la vida cotidiana, como Google Maps, el chat de video FaceTime de Apple y la IA de Amazon, Alexa. Los productos basados en datos se están apoderando rápidamente de todos los sectores, ya que grandes conjuntos de datos y herramientas de ciencias de datos ofrecen valor innovador en mercados establecidos. A mediados de los años 2000, se vio cómo la aparición de la analítica de los datos creció de manera destacada más allá de la industria de la tecnología, particularmente haciendo avances tempranos en el sector de bienes de consumo de rotación rápida (Fast Moving Consumer Goods abreviado en inglés como FMCG), como en supermercados y grandes almacenes. La industria a nivel global ha cambiado en pocos años, resumida por la observación ampliamente difundida por Tom Goodwin: Uber, la compañía de taxis más grande del mundo, no tiene vehículos. Facebook, el propietario de medios más popular del mundo, no crea contenido. Alibaba, el minorista más valioso, no tiene inventario. Y Airbnb, el mayor proveedor de alojamiento del mundo, no posee bienes raíces. Algo interesante está pasando. Las soluciones basadas en datos han permitido a los nuevos operadores irrumpir en los sectores establecidos, y las empresas de tecnología siguen impulsando el desarrollo.

Métodos alternativos de scoring crediticio son la búsqueda de nuevas fuentes de datos que permiten a los productos llegar a nuevos segmentos de clientes, a menudo a partir de la tecnología de redes sociales. Las estrategias de marketing están afinadas por rigurosas pruebas estadísticas, pruebas A/B, que fueron promulgadas por compañías como Amazon o Yahoo! para refinar sus diseños de sitios web. Adicionalmente, el análisis de segmentación geográfica de clientes, la asignación de flujos P2P y la identificación de la ubicación óptima del corresponsal, son ayudados por el análisis geoespacial y las herramientas que ofrecen Google Maps y la tecnología OpenStreetMap. A medida que la tecnología continúa evolucionando, los proveedores de SFD pueden esperar que surjan nuevas soluciones para ayudar a comprender mejor a los clientes, llegar a mercados más grandes y ofrecer productos y servicios adaptados a las necesidades de los clientes.

## Datos para la Inclusión Financiera

En el sector de inclusión financiera, los datos son importante porque la base de clientes objetivo a menudo carece de acceso a los bancos u otros servicios financieros o tiene una exposición limitada y no está familiarizada con los servicios financieros. Sus necesidades y patrones de gasto son diversos y diferentes. Los datos permiten a los proveedores de SFD crear productos y servicios que reflejen mejor las preferencias y aspiraciones del cliente. Los Servicios Financieros Digitales han cambiado el acceso y la asequibilidad de los servicios financieros en los mercados emergentes atendiendo a las necesidades de los clientes de bajos ingresos, aumentando así la inclusión financiera.

## 2.2\_RECURSOS

Los datos traen consigo la oportunidad de mejorar la inclusión financiera. Sin embargo, esto debe hacerse garantizando al tiempo que la protección del consumidor y la protección de datos no se vean comprometidos. Los datos se producen y recopilan en forma pasiva a través de dispositivos digitales como teléfonos celulares y computadores, entre otros. Muchas partes interesadas han expresado su preocupación por el hecho que los hogares de bajos ingresos, los productores primarios de estos datos en el contexto de la inclusión financiera, no sean conscientes que estos datos se están recopilando, analizando y monetizando. En ausencia de una política uniforme, existen normas diferentes aplicadas entre los tipos de proveedores y algunos casos en que se han violado los derechos de los consumidores. Con la proliferación de analítica de datos, es crítico que todas las partes interesadas - proveedores de SFD, reguladores, encargados de la formulación de políticas, instituciones financieras de desarrollo e inversionistas - discutan los problemas asociados con la protección de datos y la protección del consumidor para encontrar soluciones. Algunos profesionales en campo pueden sentirse presionados a adoptar nuevas tecnologías o metodologías para mantenerse al día con las tendencias predominantes o debido a las acciones tomadas por sus competidores. Cabe decir que tales esfuerzos podrían anularse si la organización no tiene la habilidad técnica para administrar el proyecto o no tiene la capacidad de actuar sobre la base de los resultados o la nueva información que los datos proveen. Así, los profesionales en campo deben identificar los problemas de negocio que están tratando de resolver, evaluar qué datos y capacidad analítica que poseen actualmente, y luego tomar decisiones sobre cómo implementar el proyecto de datos. El objetivo comercial debe estar en el centro de cualquier proyecto de gestión de datos.



# Glosario

Término	Explicación
<b>Acuerdo de Nivel de Servicio (SLA)</b>	Un SLA es el componente del contrato de servicio entre un proveedor de servicios y un cliente. SLAs proporciona aspectos específicos y medibles relacionados con las ofertas de servicios. Por ejemplo, los SLA se incluyen a menudo en los acuerdos firmados entre los proveedores de servicios de Internet y los clientes. SLA también se conoce como un Acuerdo de Nivel Operativo (OLA), cuando se utiliza en una organización sin una relación establecida o formal entre el proveedor y el cliente.
<b>Algoritmo</b>	En matemáticas y ciencias de la computación, un algoritmo es una secuencia autónoma de acciones a realizar. Los algoritmos realizan cálculos, procesamiento de datos o tareas de razonamiento automatizado.
<b>Almacenamiento de Datos</b>	Almacenamiento de datos es un término general para archivar datos en formas electromagnéticas o de otro tipo, para su uso por un computador o dispositivo. Los diferentes tipos de almacenamiento de datos juegan diferentes funciones en un entorno informático. Además de las formas de almacenamiento de hard data, ahora hay nuevas opciones para el almacenamiento remoto de datos, como la computación en la nube, que pueden revolucionar las formas en que los usuarios acceden a los datos.
<b>Análisis de Minería de Texto</b>	Minería de texto, también conocido como minería de datos de texto y aproximadamente equivalente a análisis de texto, es el proceso de derivar información de alta calidad a partir de texto. La información de alta calidad se obtiene típicamente a través de la creación de patrones y tendencias a través de medios tales como el aprendizaje de patrones estadísticos. La minería de textos generalmente involucra: estructurar el texto de entrada, (usualmente analizar, junto con la adición de algunas características lingüísticas derivadas y la eliminación de otras, y posterior inserción en una BD); derivar patrones dentro de los datos estructurados; y la evaluación e interpretación de los resultados.
<b>Análisis de Redes Sociales (SNA)</b>	El análisis de redes sociales, o SNA, es el proceso de investigación de estructuras sociales a través del uso de teorías de red y gráficas. Caracteriza estructuras en red en términos de nodos (actores individuales, personas o cosas dentro de la red) y los vínculos, aristas o enlaces (relaciones o interacciones) que los conectan.
<b>Análítica de Datos</b>	El análisis de datos se refiere a técnicas y procesos cualitativos y cuantitativos utilizados para generar información, mejorar la productividad y generar beneficios a nivel de negocio. Los datos se extraen y categorizan para identificar y analizar datos y patrones de comportamiento, y las técnicas de análisis de datos varían según los requisitos de la organización.
<b>Metodologías de Analítica Predictiva</b>	Los análisis predictivos proporcionan un análisis mucho más complejo de los datos existentes para hacer una proyección a futuro. Las técnicas incluyen análisis de regresión, estadísticas multivariantes, correlación de patrones, minería de datos, modelado predictivo y pronóstico.
<b>Metodologías de Analítica Prescriptiva</b>	El análisis prescriptivo va un paso más allá: proporciona información para alimentar decisiones óptimas para un conjunto de resultados futuros previstos. Las técnicas incluyen análisis de gráficos, redes neuronales, máquinas y aprendizaje profundo.
<b>Metodologías de Analítica Descriptiva</b>	Las metodologías analíticas menos complejas son de naturaleza descriptiva; proporcionan descripciones históricas del desempeño institucional, análisis en torno a las razones para este desempeño e información sobre el desempeño institucional actual. Las técnicas incluyen alertas, consultas, búsquedas, informes, visualización, tablas, paneles de control, gráficos, narrativas, correlaciones, así como análisis estadístico simple.
<b>Anti Lavado de Activos y Lucha Contra el Terrorismo (ALA/CFT)</b>	ALA/CFT son controles legales aplicados al sector financiero para ayudar a prevenir, detectar y reportar actividades de lavado de dinero. ALA/Los controles ALA/CFT incluyen montos máximos que pueden mantenerse en una cuenta o ser transferidos entre cuentas en cualquier transacción, o en un día cualquiera. También incluye la información financiera obligatoria de KYC para todas las transacciones de más de \$10,000, incluyendo la declaración de la fuente de los fondos, así como la razón de la transferencia.
<b>Aprendizaje Automático</b>	El aprendizaje automático es un tipo de inteligencia artificial (IA) que proporciona a los computadores la capacidad de aprender sin ser programados explícitamente. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden cambiar cuando se exponen a nuevos datos.

<b>Aprendizaje No Supervisado</b>	El aprendizaje no supervisado es un método que permite a las máquinas clasificar objetos tangibles e intangibles sin proporcionar ninguna información previa sobre los objetos. Las cosas que las máquinas necesitan clasificar son variadas, como los hábitos de compra de los clientes, patrones de comportamiento de las bacterias o ataques de hackers. La idea principal detrás del aprendizaje no supervisado es exponer a las máquinas a grandes volúmenes de datos variados y permitirles aprender e inferir de los datos. Sin embargo, las máquinas deben programarse primero para aprender de los datos.
<b>Aprendizaje Supervisado</b>	El aprendizaje supervisado es un método utilizado para permitir que las máquinas clasifiquen objetos, problemas o situaciones basados en datos relacionados introducidos en las máquinas. Las máquinas se alimentan de datos tales como características, patrones, dimensiones, color y altura de objetos, personas o situaciones de forma repetitiva hasta que las máquinas son capaces de realizar clasificaciones precisas. El aprendizaje supervisado es una tecnología o concepto popular que se aplica a escenarios de la vida real. El aprendizaje supervisado se utiliza para proporcionar recomendaciones de productos, segmentar clientes basados en sus datos, diagnosticar enfermedades basadas en síntomas anteriores y realizar muchas otras tareas.
<b>Arquitectura de Datos</b>	La arquitectura de datos es un conjunto de reglas, políticas, estándares y modelos que rigen y definen el tipo de datos recopilados; y cómo se usa, almacena, administra e integra dentro de una organización y sus sistemas de Base de Datos. Ofrece un enfoque formal para crear y administrar el flujo de datos, y cómo se procesan a través de los sistemas y aplicaciones de TI de una organización.
<b>Banca Electrónica</b>	El suministro de productos y servicios bancarios a través de canales de distribución electrónicos.
<b>Banca móvil</b>	El uso de un teléfono móvil para acceder a los servicios bancarios convencionales. Esto abarca tanto los servicios transaccionales como los no transaccionales, como la consulta de información financiera y la ejecución de transacciones financieras.
<b>Big Data</b>	Big Data son grandes conjuntos de datos, cuyo tamaño se mide por cinco características distintas: volumen, velocidad, veracidad, variedad, y complejidad.
<b>Billeteras electrónicas</b>	Una cuenta de dinero electrónico que pertenece a un cliente de Servicios Financieros Digitales y se accede a través de un teléfono móvil.
<b>Byte</b>	Es una unidad de información digital, considerada una unidad de tamaño de memoria. Consta de 8 bits y 1024 bytes es igual a 1 kilobyte.
<b>Call Center</b>	Una oficina centralizada utilizada para recibir o transmitir un gran volumen de solicitudes telefónicas. Además de manejar las quejas y consultas de los clientes, también se utiliza como un canal de distribución alternativo (ADC) para mejorar la cercanía con el cliente y atraer nuevos clientes a través de varias campañas promocionales.
<b>Canal</b>	El punto de acceso del cliente a un proveedor de servicios financieros, es decir, con quién o con qué interactúa el cliente para acceder a un servicio financiero o producto.
<b>Canales de Distribución Alternativos</b>	Canales que amplían el alcance de los servicios financieros más allá de la sucursal tradicional. Estos incluyen cajeros automáticos (ATMs), banca por internet, banca móvil, billeteras electrónicas, algunos servicios de tarjeta/dispositivo POS y servicios de concesión de créditos.
<b>Científico de Datos</b>	Un científico de datos es un individuo, una organización o un equipo que realiza análisis estadísticos, minería de datos y procesos de recuperación en una gran cantidad de datos para identificar tendencias, cifras y otra información relevante.
<b>Comerciante</b>	Una persona o negocio que proporciona bienes o servicios a un cliente a cambio de un pago.
<b>Complejidad</b>	Combinar los cuatro atributos de Big Data (volumen, velocidad, variedad y veracidad) requiere procesos analíticos avanzados. Hay una variedad de procesos analíticos que han surgido para hacer frente a estos grandes conjuntos de datos. Los procesos analíticos apuntan a tipos específicos de datos como texto, audio, web y redes sociales. Otra metodología que ha recibido gran atención tiene que ver con el aprendizaje automático, donde un algoritmo es creado y alimentado a un computador junto con datos históricos. Esto permite que el algoritmo haga predicciones de relaciones entre variables aparentemente sin conexión.
<b>Conozca a su Cliente (KYC)</b>	Reglas relacionadas con la ALA/CFT que obligan a los proveedores a llevar a cabo procedimientos para identificar a un cliente, y que evalúan el valor de la información para detectar, supervisar y reportar actividades sospechosas.

<b>Corresponsal</b>	Una persona o negocio contratado para procesar las transacciones de los usuarios. Las más importantes son depósitos y retiros (es decir, el cargue de valor en el sistema de dinero electrónico, y luego convertirlo de nuevo en un retiro). En muchos casos, los corresponsales registran nuevos clientes también. Los corresponsales suelen ganar comisiones por realizar estos servicios. También suelen ofrecer un servicio de atención al cliente de primera línea, como enseñar a los nuevos usuarios a realizar transacciones desde su teléfono. Por lo general, los corresponsales llevarán a cabo otros tipos de negocios, además de dinero electrónico. Los corresponsales en algunos casos estarán limitados por regulaciones, pero los comerciantes a pequeña escala, las IMFs, las cadenas de tiendas y las sucursales bancarias sirven como corresponsales en algunos mercados. Algunos participantes de la industria prefieren los términos 'comerciante' o 'minorista' para evitar ciertas connotaciones legales del término 'corresponsal', ya que se utiliza en otras industrias. (GSMA, 2014).
<b>Corresponsal Maestro</b>	Una persona o negocio que compra dinero electrónico de un proveedor de Servicios Financieros Digitales al por mayor y luego lo revende a corresponsales, quienes a su vez lo venden a los usuarios. (A diferencia de un super-corresponsal, los corresponsales maestros son responsables de gestionar los requisitos de liquidez en efectivo y de valor electrónico de un determinado grupo de corresponsales.)
<b>Cuenta Activa</b>	Una cuenta activa es aquella que ha sido usada para al menos una transacción en el periodo anterior, normalmente es reportada como "activa 30-días" o "activa 90-días". No incluye transacciones no financieras, tales como el cambio del código PIN.
<b>Data/Datos</b>	Data/Datos es un término general que se utiliza para describir cualquier información, hecho o estadística que se ha recopilado para cualquier tipo de análisis o para fines de referencia. Hay muchos tipos diferentes de datos de una variedad de fuentes diferentes. La data/datos generalmente se procesa, agrega, manipula o consolida para producir información que proporcione insights.
<b>Datos Cualitativos</b>	Datos que se aproximan o caracterizan, pero que no miden los atributos, características o propiedades de una cosa o fenómeno. Los datos cualitativos describen, mientras que los datos cuantitativos definen.
<b>Datos Cuantitativos</b>	Datos que pueden ser cuantificados y verificados, y son susceptibles a manipulación estadística. Los datos cualitativos describen, mientras que los datos cuantitativos definen.
<b>Data Cube</b>	En informática, datos multidimensionales, a menudo con el tiempo como una tercera dimensión de columnas y filas. En las operaciones comerciales, se trata de un término genérico que se refiere a sistemas corporativos que permiten a los usuarios especificar y descargar informes de datos sin procesar. Muchos incluyen campos de "arrastrar y soltar" para diseñar una solicitud de informe o agregaciones de datos sencillas.
<b>Lago de Datos</b>	Un Lago de Datos es un repositorio masivo, fácilmente accesible, centralizado, de grandes volúmenes de datos estructurados y sin estructurar.
<b>Datos No Estructurados</b>	Normalmente se refiere a información que no reside en una BD tradicional de columna-fila. Los archivos de datos no estructurados suelen incluir texto y contenido multimedia. Los ejemplos incluyen: mensajes de correo electrónico, documentos de procesamiento de textos, videos, fotos, archivos de audio, presentaciones, páginas web y muchos otros tipos de documentos empresariales.
<b>Data Scraping (Raspado de Datos)</b>	Es una técnica en la cual un programa de computadora extrae datos de salidas legibles por el ser humano, que provienen de otra fuente digital tal como un sitio Web, informes o pantallas de computadora.
<b>Datos Tradicionales</b>	Los datos tradicionales se refieren a datos internos estructurados de uso común (como transaccionales) y a datos externos (como información de burós de crédito) que se utilizan en el proceso de toma de decisiones. Puede incluir datos que se generan a partir de la interacción con clientes tales como encuestas, formularios de registro, salario e información demográfica.
<b>Data Warehouse (Almacén de Datos)</b>	Una colección de información corporativa y datos derivados de sistemas operativos y fuentes externas de datos. Un data warehouse está diseñado para soportar decisiones empresariales, al permitir la consolidación, análisis y generación de informes de datos, a diferentes niveles agregados.
<b>Datos Alternativos</b>	Datos no financieros de MNO, redes sociales y sus bases de datos transaccionales. El acceso a otros datos alternativos, como el historial de pagos y las facturas de servicios públicos, también puede permitir la creación de scores crediticios para los clientes que de otro modo no podrían ser atendidos.
<b>Datos Estructurados</b>	Datos Estructurados se refiere a cualquier dato que resida en un campo fijo dentro de un registro o archivo. Esto incluye datos contenidos en bases de datos relacionales.

<b>Datos Geoespaciales</b>	Información sobre un objeto físico que puede ser representado por valores numéricos en un sistema de coordenadas geográficas.
<b>Datos Periféricos</b>	Normalmente, las fuentes de datos periféricos más útiles son los datos de call centers, los datos de CRM (sistemas de tickets), la información de la base de conocimientos de preguntas frecuentes, los correos de aprobación, rastreadores de listas negras y listas blancas, rastreadores de Excel compartidos.
<b>Datos Semi-estructurados</b>	Los datos semi-estructurados son una forma de data estructurada que no se ajusta a la estructura formal de los modelos de datos asociados con BDs relacionales u otras formas de tablas de datos. Sin embargo, contienen etiquetas u otros marcadores para separar los elementos semánticos y hacer cumplir jerarquías de registros y campos dentro de la data.
<b>Desviación Estándar</b>	En la estadística, la desviación estándar es una medida que se utiliza para cuantificar la cantidad de variación o dispersión de un conjunto de valores de datos. Una desviación estándar baja indica que los puntos de datos tienden a estar cerca de la media (o promedio) del conjunto, mientras que una desviación estándar alta indica que los puntos de datos se extienden sobre un rango más amplio de valores.
<b>Dinero electrónico</b>	Es el valor almacenado en billeteras virtuales o tarjetas. Normalmente, el valor total de la moneda electrónica emitida es igualado por los fondos mantenidos en una o más cuentas bancarias. Normalmente se mantienen en un fideicomiso, de modo que incluso si el proveedor de billeteras electrónicas fallara, los usuarios podrían recuperar el valor total almacenado en sus cuentas.
<b>Distribución Estadística</b>	La distribución de una variable es una descripción del número relativo de veces, que cada posible resultado ocurrirá en una serie de pruebas.
<b>Fondo (Fondo de Corresponsal)</b>	El saldo del dinero electrónico, o efectivo físico, o dinero en una cuenta bancaria a la que un corresponsal puede acceder inmediatamente para satisfacer las demandas de los clientes de comprar (depositar) o vender (retirar) dinero electrónico.
<b>Estadísticas Paramétricas</b>	La estadística paramétrica es una rama de la estadística que asume que la data de la muestra viene de una población que sigue una distribución de la probabilidad basada en un sistema fijo de parámetros. La mayoría de los métodos estadísticos elementales bien conocidos son paramétricos.
<b>Exabyte (EB)</b>	El Exabyte (EB) es un múltiplo de la unidad byte para la información digital. En el Sistema Internacional de Unidades, el prefijo indica la multiplicación por la sexta potencia de 1000 (10 <sup>18</sup> ). Así, un EB es uno quintillion de bytes (escala corta). El símbolo para el Exabyte es EB.
<b>Gestión de Datos</b>	La gestión de datos es el desarrollo, ejecución y supervisión de planes, políticas, programas y prácticas que controlan, protegen, entregan y mejoran el valor de los datos y los activos de información.
<b>Hipótesis</b>	Una hipótesis es una predicción informada que puede ser probada.
<b>Historial de Crédito</b>	Un historial de crédito es un registro del pago de deudas de un prestatario; el pago responsable se interpreta como un historial de crédito favorable, mientras que la morosidad o incumplimientos son factores que crean un historial de crédito negativo. Un informe de crédito es un registro del historial de crédito del prestatario de un número de fuentes, que tradicionalmente incluyen bancos, compañías de tarjetas de crédito, agencias de cobranzas y gobiernos.
<b>Indicador Clave de Desempeño (KPI)</b>	Un KPI es un valor medible que demuestra la eficacia con la que una empresa está logrando los objetivos clave del negocio. Las organizaciones usan KPIs en múltiples niveles para evaluar su éxito en el logro de objetivos. Los KPIs de alto nivel pueden enfocarse en el desempeño general de la empresa, mientras que los KPIs de bajo nivel pueden enfocarse en procesos en departamentos como ventas, marketing o un call center.
<b>Indicador de Riesgo Clave (KRI)</b>	Un Indicador de Riesgo Clave es una medida usada para indicar el grado de riesgo de una actividad. Se diferencia de un Indicador de Desempeño Clave (KPI) en que este último se entiende como una medida de lo bien que se está haciendo algo, mientras que el primero indica lo perjudicial que puede ser algo si se produce y lo probable es que ocurra.
<b>Insight</b>	Insight es un término utilizado en Psicología proveniente del inglés que se puede traducir al español como "visión interna" o más genéricamente "percepción" o "entendimiento". Mediante un insight el sujeto "capta", "internaliza" o comprende, una "verdad" revelada. Puede ocurrir inesperadamente, luego de un trabajo profundo, simbólicamente, o mediante el empleo de diversas técnicas afines.
<b>Datos Abierta</b>	Los Datos Abiertos son los datos a los que cualquiera puede acceder, utilizar o compartir.
<b>Institución Financiera (IF)</b>	Un proveedor de servicios financieros, incluyendo cooperativas de ahorro y crédito, bancos, instituciones financieras no bancarias, instituciones microfinancieras y proveedores de servicios financieros móviles.
<b>Instituciones Microfinancieras (IMF)</b>	Una institución financiera especializada en servicios bancarios para grupos de bajos ingresos, pequeñas empresas o individuos.

<b>Inteligencia Artificial (IA)</b>	IA es un área informática que enfatiza la creación de máquinas inteligentes que funcionan y reaccionan como seres humanos.
<b>Investigación Primaria y Secundaria</b>	La investigación primaria es la información original recogida a través de su propio enfoque, a menudo un estudio o encuesta. La investigación secundaria utiliza los resultados existentes de los estudios realizados previamente y la recopilación de datos.
<b>Limpieza de Datos</b>	La limpieza de datos es el proceso de alterar los datos en un recurso de almacenamiento dado, para asegurarse que son precisos y correctos.
<b>Máquinas de Vector de Soporte (SVM)</b>	Una máquina vectorial de soporte, o SVM, es un algoritmo de aprendizaje automático que analiza los datos para su clasificación y análisis de regresión. SVM es un método de aprendizaje supervisado que examina los datos y los clasifica en una de dos categorías. Un SVM genera un mapa de datos ordenados con los márgenes entre los dos tan separados como sea posible. SVMs se utilizan en la categorización del texto, la clasificación de la imagen, el reconocimiento de la escritura, y en las ciencias. Una máquina de vectores de soporte también se conoce como una red de vectores de soporte (SVN).
<b>Metadatos</b>	Los metadatos describen otros datos. Proporciona información sobre el contenido de un determinado elemento. Por ejemplo, una imagen puede incluir metadatos que describe el tamaño de la imagen, la profundidad de color, la resolución de la imagen, cuando se creó la imagen y otros datos.
<b>Método Científico</b>	Solución de problemas mediante un enfoque paso a paso que consiste en (1) identificar y definir un problema, (2) acumular datos relevantes, (3) formular una hipótesis (4) realizar experimentos para probar la hipótesis, (5) interpretar los resultados objetivamente, y (6) repetir los pasos hasta encontrar una solución aceptable.
<b>Metodología no Paramétrica</b>	Un método comúnmente usado en estadísticas donde se usan pequeños tamaños de muestra para analizar datos nominales. Un método no paramétrico se utiliza cuando el investigador no sabe nada sobre los parámetros de la muestra elegida de la población.
<b>Métodos Monte Carlo</b>	Modelos que utilizan aproximaciones aleatorias para modelar sistemas complejos estableciendo un peso probabilístico a varios puntos de decisión en el modelo. Los resultados muestran un patrón de distribución estadística que puede usarse para predecir la probabilidad de ciertos resultados dados los inputs en el sistema que se está modelando. Estos modelos se usan típicamente para problemas de optimización o análisis de probabilidad.
<b>Minería de datos</b>	La minería de datos es el proceso computacional para descubrir patrones en grandes conjuntos de datos. Es un sub-campo interdisciplinario de la informática. El objetivo general del proceso de minería de datos es extraer información de un conjunto de datos y transformarla en una estructura comprensible para uso posterior.
<b>Modelado Predictivo</b>	El modelado predictivo es un proceso que utiliza la minería de datos y la probabilidad para pronosticar resultados. Cada modelo se compone de una serie de predictores, que son variables que probablemente influyan en los resultados futuros. Una vez que se han recopilado los datos para los predictores relevantes, se formula un modelo estadístico.
<b>Modelo de Calificación Psicométrica</b>	La psicometría se refiere a la medición de conocimientos, habilidades, actitudes y rasgos de personalidad. En los modelos de calificación psicométrica, los principios psicométricos se aplican al scoring crediticio mediante el uso de técnicas estadísticas avanzadas para pronosticar la probabilidad de incumplimiento de un solicitante.
<b>Operador de Red Móvil (MNO)</b>	Una empresa que tiene una licencia expedida por el gobierno para proporcionar servicios de telecomunicaciones a través de dispositivos móviles.
<b>Persona a Persona (P2P)</b>	Transferencia de fondos de persona a persona.
<b>Panel de Control</b>	Un Panel de Control de BI, es una herramienta de visualización de datos que muestra el estado actual de métricas y KPIs para una empresa. Los paneles de control consolidan y organizan números, métricas y, a veces, scorecards en una sola pantalla.
<b>Pequeñas y Medianas Empresas (PyMEs)</b>	Las pequeñas y medianas empresas, o PYMES, son empresas no filiales e independientes que emplean a menos de un determinado número de empleados. Este número varía según los países.
<b>Probabilidad</b>	La probabilidad es la probabilidad asumida de que el evento ocurrirá. La probabilidad se cuantifica como un número entre cero y uno (donde '0' indica imposibilidad y '1' indica certeza). Cuanto mayor sea la probabilidad de un evento, más seguro será que ocurrirá el evento.
<b>Procesamiento de Datos</b>	El procesamiento de datos es, en general, la recopilación y manipulación de elementos de datos para producir información significativa. En este sentido, puede considerarse un subconjunto del procesamiento de información, o el cambio (procesamiento) de información de cualquier manera detectable por un observador.
<b>Procesamiento de Imágenes</b>	El procesamiento de imágenes es un término de significado algo amplio, que se refiere al uso de herramientas analíticas para procesar o mejorar imágenes. Muchas definiciones de este término especifican operaciones matemáticas o algoritmos como herramientas para el procesamiento de una imagen.

<b>Procesamiento del Lenguaje Natural (PLN)</b>	El campo de estudio que se centra en las interacciones entre el lenguaje humano y las computadoras se llama Procesamiento del Lenguaje Natural [Natural Language Processing], o PLN en forma abreviada. Se sitúa en la intersección de la informática, la IA y la lingüística computacional. El PLN es un campo que cubre la comprensión y manipulación de un lenguaje humano.
<b>Interfaz de Programación de Aplicaciones (API)</b>	Un método para especificar un componente de software en términos de sus operaciones destacando un conjunto de funcionalidades que son independientes de su respectiva implementación. Las API se utilizan para la integración en tiempo real al Sistema de Core Bancario o Sistema de Información Gerencial (SIG), que especifican cómo dos sistemas diferentes pueden comunicarse entre sí a través del intercambio de 'mensajes'. Existen diferentes tipos de APIs, incluidas las basadas en la Web, la comunicación Protocolo de Control de Transmisión (TCP) e integración a una base de datos, o APIs escritas para sistemas específicos.
<b>Promedio</b>	Un promedio es la suma de una lista de números dividida por el número de números en la lista. En matemáticas y estadística, esto sería llamado la media aritmética.
<b>Promedio de Ingresos por Usuario (ARPU)</b>	El ARPU es una medida utilizada principalmente por operadores de redes móviles (MNOs), definida como el ingreso total dividido por el número de suscriptores.
<b>Protección de Datos</b>	La protección de datos, también llamada privacidad de la información, es el aspecto de TI que se ocupa de la capacidad que una organización o un individuo tiene para determinar qué datos en un sistema informático pueden ser compartidos con terceros.
<b>Protocolo de Transferencia de Archivos (FTP)</b>	El Protocolo de transferencia de archivos (FTP) es un protocolo cliente-servidor utilizado para transferir archivos o intercambiar archivos con un equipo host. FTP es el estándar de Internet para mover o transferir archivos de un computador a otro utilizando redes TCP o IP.
<b>Pruebas A/B</b>	La prueba A/B es un método para revisar dos versiones diferentes de un producto o servicio con el fin de evaluar la forma en la que un pequeño cambio en los atributos del producto puede impactar en el comportamiento del cliente. Este tipo de experimento permite a los proveedores de Servicios Financieros Digitales elegir múltiples variaciones de un producto o servicio, probar estadísticamente la adopción resultante en los clientes y comparar los resultados entre grupos objetivo.
<b>Prueba Aleatoria Controlada (RCT)</b>	Una prueba aleatoria controlada es un experimento científico en el que las personas que participan en la prueba se asignan aleatoriamente a diferentes contextos de intervención y luego se comparan entre sí. La aleatorización minimiza el sesgo de selección durante el diseño del experimento científico. Los grupos de comparación permiten a los investigadores determinar los efectos de la intervención en comparación con el grupo sin intervención (control), mientras que otras variables se mantienen constantes.
<b>Punto de Venta (POS)</b>	Dispositivo electrónico utilizado para procesar pagos con tarjeta en el punto en el que un cliente hace un pago al comerciante a cambio de bienes y servicios. El dispositivo POS es un dispositivo de hardware (fijo o móvil) que ejecuta software para facilitar la transacción. Dispositivos de PC originales, pero cada vez más incluyen teléfonos móviles, teléfonos inteligentes y tabletas.
<b>Reconocimiento de Patrones</b>	En TI, el reconocimiento de patrones es una rama del aprendizaje automático que enfatiza el reconocimiento de patrones de datos o regularidades de datos en un escenario determinado. Es una subdivisión del aprendizaje automático y no debe confundirse con un estudio de "actual machine learning". El reconocimiento de patrones puede ser 'supervisado', donde se pueden encontrar patrones previamente conocidos en un dato determinado, o 'sin supervisión', donde se descubren patrones enteramente nuevos.
<b>Registros Detallados de Llamadas (CDR)</b>	Este es el registro del MNO de una llamada de voz o un SMS, con detalles como origen, destino, duración, hora del día o cantidad cobrada por cada llamada o SMS.
<b>Regresión Lineal</b>	Técnica matemática para encontrar la recta que mejor se ajusta a los valores de una función lineal, trazada en un gráfico de dispersión como puntos de datos.
<b>Scoring crediticio</b>	Un análisis estadístico realizado por prestamistas e instituciones financieras para acceder a la capacidad crediticia de una persona. Los prestamistas utilizan el score crediticio, entre otras cosas, para llegar a una decisión sobre si conceder el crédito. Los scores crediticios de una persona es un número entre 300 y 850, siendo 850 la calificación crediticia más alta posible.
<b>Segmentación de Mercado</b>	El proceso de definición y subdivisión de un gran mercado homogéneo en segmentos claramente identificables con necesidades, deseos o características de demanda similares. Su objetivo es diseñar una mezcla comercial que coincida exactamente con las expectativas de los clientes en el segmento objetivo.
<b>Segmentación psicográfica</b>	La segmentación psicográfica implica dividir el mercado en segmentos basados en diferentes rasgos de personalidad, valores, actitudes, intereses y estilos de vida de los consumidores.

<b>Seguridad de Datos</b>	La seguridad de datos se refiere a medidas de protección de la privacidad digital que se aplican para impedir el acceso no autorizado a los computadores, bases de datos, sitios web y cualquier otro lugar donde se almacenan datos. La seguridad de datos también protege de la corrupción de datos. La seguridad de datos es un aspecto esencial de la TI para las organizaciones de todos los tamaños y tipos.
<b>Servicio de Dinero Electrónico, Servicios Financieros Móviles (SFM)</b>	Un servicio financiero digital que se proporciona mediante la expedición de cuentas virtuales contra una sola cuenta bancaria agrupada como billeteras electrónicas, a las que se accede mediante un teléfono móvil. La mayoría de los proveedores de dinero electrónico son MNO o Proveedor de Servicios de Pago (PSP).
<b>Servicio de Mensajes Cortos (SMS)</b>	Un canal de comunicación que 'almacena y reenvía' que implica el uso de la red de telecomunicaciones y el protocolo SMPP para enviar una cantidad limitada de texto de un teléfono a otro o de uno a muchos teléfonos.
<b>Servicio Suplementario Datos No Estructurados (USSD)</b>	Protocolo utilizado por los dispositivos móviles GSM para comunicarse con los computadores/red del proveedor de servicios. Este canal está soportado por todos los teléfonos GSM, permitiendo una sesión interactiva consistente en un intercambio bidireccional de mensajes basados en un menú de aplicación definido.
<b>Servicios Financieros Digitales (SFD)</b>	El uso de medios digitales para ofrecer servicios financieros. Los Servicios Financieros Digitales abarcan todas las ofertas de telefonía móvil, de tarjeta, de POS y de comercio electrónico, incluidos los servicios suministrados a los clientes a través de redes de corresponsales.
<b>Asociación del Sistema Global para las Comunicaciones Móviles (GSMA)</b>	La Asociación GSM (comúnmente conocida como 'la GSMA') es una organización gremial que representa los intereses de los operadores móviles en todo el mundo. Aproximadamente 800 operadores móviles son miembros de la GSMA y otros 300 en el ecosistema móvil más amplio son miembros asociados.
<b>Super-Corresponsal</b>	Un negocio, a veces un banco, que compra dinero electrónico de un proveedor de Servicios Financieros Digitales al por mayor y luego lo revende a los corresponsales, que a su vez lo venden a los usuarios.
<b>Tipo de teléfono móvil - Smartphone</b>	Un teléfono móvil que tiene la capacidad de procesamiento para realizar muchas de las funciones de un computador, que normalmente tiene una pantalla relativamente grande y un sistema operativo capaz de ejecutar un complejo conjunto de aplicaciones, con acceso a internet. Además del servicio de voz digital, los teléfonos inteligentes modernos ofrecen mensajes de texto, correo electrónico, navegación por Internet, cámara fotográfica y de vídeo y , un reproductor de MP3 y reproducción de vídeo con capacidades de transferencia de datos integradas, capacidades GPS.
<b>Tipo de Teléfono Móvil - Teléfono de Gama Baja</b>	Un teléfono móvil de gama baja que puede hacer y recibir llamadas, enviar mensajes de texto y acceder al canal USSD, pero tiene funcionalidades adicionales muy limitadas.
<b>Tipo de teléfono móvil - Teléfono de Gama Media</b>	Un teléfono de gama media es un tipo de teléfono móvil que tiene más funciones que un teléfono de gama baja, pero no es equivalente a un teléfono inteligente. Los teléfonos de gama media pueden ofrecer algunas de las funcionalidades avanzadas que se encuentran en un teléfono inteligente, como un reproductor de media portátil, cámara digital, organizador personal y acceso a internet, pero normalmente no admiten aplicaciones complementarias.
<b>Variedad</b>	La era digital ha diversificado los tipos de datos disponibles. Los datos tradicionales estructurados encajan en las BD existentes que están destinadas a información bien definida, que sigue un conjunto de reglas. Por ejemplo, una transacción bancaria tiene una marca de tiempo, cantidades y ubicación. Sin embargo, hoy en día, el 90 por ciento de los datos que se generan son no estructurados, es decir, viene en forma de tweets, imágenes, documentos, archivos de audio, historiales de compra de clientes y videos.
<b>Velocidad</b>	Una gran parte de los datos se produce y se sirve en tiempo real. Para 2018, se estima que se van a cargar y descargar en Internet 50.000 gigabytes de datos cada segundo. Cada 60 segundos, se envían 204 millones de correos electrónicos. Como consecuencia, estos datos tienen que ser almacenados, procesados y analizados a velocidades muy altas, a veces a un ritmo de decenas de miles de bytes cada segundo.
<b>Veracidad</b>	La veracidad se refiere a la fiabilidad de los datos. Los gerentes de negocios necesitan saber que los datos que usan en el proceso de toma de decisiones son representativos de las necesidades y deseos de sus segmentos de clientes. Así, las prácticas de gestión de datos en las empresas deben garantizar que el proceso de limpieza de datos es continua y rigurosa. Esto evitará la inclusión de datos engañosos e incorrectos en el análisis.
<b>Volumen</b>	La gran cantidad de datos que se está produciendo es alucinante. Se estima que aproximadamente 2,5 quintillones de bytes de datos se producen cada día. Para tener una idea de la cantidad, este volumen de datos llenaría aproximadamente 10 millones de discos Blu-ray. La madurez de esta data se ha vuelto cada vez menor, es decir, que la cantidad de datos que tiene menos de un minuto de edad ha ido aumentando constantemente. De hecho, el 90 por ciento de esta data se han producido en los últimos dos años. Se prevé que la cantidad de datos en el mundo aumentará 44 veces entre 2009 y 2020.

---

# Biografías de los Autores

## DEAN CAIRE

*Especialista en Scoring Crediticio, IFC*

Dean trabajó durante los últimos 15 años como consultor de scoring crediticio, 12 con la empresa DAI Europe y posteriormente como consultor independiente. Durante este tiempo, ha ayudado a clientes de 77 instituciones financieras en 45 países a desarrollar más de 100 modelos de scoring crediticio personalizados para los siguientes segmentos: crédito de consumo (incluyendo Servicios Financieros Digitales), arrendamientos estándar de activos, créditos para microempresas, créditos para pequeñas empresas (incluyendo Servicios Financieros Digitales para comerciantes), créditos agrícolas y leasing de equipos (incluyendo Servicios Financieros Digitales), microcréditos a grupos solidarios, y grandes créditos a compañías no cotizadas en bolsa. Dean se esfuerza por transferir el desarrollo de modelos y habilidades de gestión a las contrapartes de IF para que se puedan apropiar en forma plena de los modelos y administrarlos en el futuro.

## LEONARDO CAMICIOTTI

*Director Ejecutivo, Consorcio TOP-IX*

Reportando a la Junta Directiva, Leonardo es responsable de las actividades estratégicas, administrativas y operativas del Consorcio TOP-IX. El maneja el Programa de Desarrollo TOP-IX, que fomenta la creación de nuevas empresas mediante el apoyo infraestructural (es decir, ancho de banda de Internet, cloud computing y elaboración de prototipos de software) a startups y promueve proyectos de innovación en diferentes sectores, como big data y computación de alto rendimiento; fabricación abierta y tecnologías cívicas. Anteriormente, fue Científico Investigador, Oficial de Estrategia y Desarrollo de Negocios, y Dueño de Philips Corporate Research. Se graduó como Ingeniero Electrónico de la Universidad de Florencia y tiene un MBA de la Universidad de Turín.

## SOREN HEITMANN

*Oficial de Operaciones, IFC*

Soren dirige el programa de investigación aplicada y Monitoreo, Evaluación y Aprendizaje (MEL) de la alianza IFC-Mastercard Foundation. Trabaja en el vínculo de la investigación y la tecnología basadas en datos para ayudar a impulsar el aprendizaje y la innovación para los proyectos de IFC en África Subsahariana. Anteriormente, Soren lideró la medición de resultados para la VPU de riesgo de IFC y el equipo de Administración de Cartera de Evaluación y Monitoreo Regional, para Europa y Asia Central. Tiene experiencia en gestión de bases de datos, ingeniería de software y tecnología web, que ahora incorpora en su trabajo de soporte de operaciones de datos a clientes de IFC. Soren tiene un título en Antropología Cultural de la Universidad de Boston y una Maestría en Economía del Desarrollo de Johns Hopkins SAIS.

## SUSIE LONIE

*Especialista en Servicios Financieros Digitales, IFC*

Susie pasó tres años en Kenia creando y haciendo operar el servicio de pagos móviles de M-PESA, después de lo cual facilitó su lanzamiento en varios otros mercados incluyendo India, Suráfrica y Tanzania. En 2010, Susie fue co-ganadora del Economist Innovation Award for Social and Economic Innovation por su trabajo en M-PESA. Se convirtió en consultor independiente de Servicios Financieros Digitales en 2011 y trabaja con bancos, MNOs, y otros clientes en todos los aspectos del suministro de servicios financieros a los desbancarizados en mercados emergentes, incluyendo dinero electrónico, banca de correspondientes, transferencias internacionales de dinero, e interoperabilidad. Susie trabaja en estrategia Servicios Financieros Digitales, evaluación financiera, diseño de productos y requerimientos funcionales, operaciones, gestión de correspondientes, evaluaciones de riesgo, evaluaciones de investigación, y ventas y marketing. Sus títulos son en Ingeniería Química de Edinburgo y Manchester, Reino Unido.

## CHRISTIAN RACCA

*Ingeniero de Diseño, Consorcio TOP-IX*

Christian maneja el programa TOP-IX BIG DIVE que apunta a ofrecer cursos de formación para científicos de datos, iniciativas educativas orientadas a datos para empresas, organizaciones y proyectos de consultoría en el campo de la explotación de datos (Big Data). Después de graduarse en ingeniería de telecomunicaciones en el Politécnico di Torino, Christian se unió al Consorcio TOP-IX, trabajando en streaming de datos y cloud computing, y más tarde en nuevos emprendimientos web. Ha asesorado varios proyectos sobre el modelo de negocios, desarrollo de productos y arquitectura de infraestructura y ha cultivado relaciones con inversionistas, incubadoras, aceleradores y el ecosistema de Innovación en Italia y Europa.

## MINAKSHI RAMJI

*Oficial de Operaciones Asociado, IFC*

Minakshi dirige proyectos sobre Servicios Financieros Digitales e inclusión financiera dentro del Grupo de Instituciones Financieras de IFC en África Subsahariana. Antes de esto, fue consultora en MicroSave, una firma de consultoría de inclusión financiera con sede en la India, donde fue Analista Senior en su práctica de Servicios Financieros Digitales. También trabajó en el Centro de Microfinanzas del IFMR Trust en India, con foco en políticas relacionadas con problemáticas de acceso financiero en la India. Tiene una maestría en Desarrollo Económico de la London School of Economics y un BA en Matemáticas de Bryn Mawr College en los Estados Unidos.

## QIUYAN XU

*Científico de Datos Principal, Cignifi*

Qiuyan Xu es el Científico de Datos Principal de Cignifi Inc., dirigiendo el equipo de Big Data Analytics. Cignifi es una empresa de tecnología financiera de rápido crecimiento en Boston, Estados Unidos, que ha desarrollado la primera plataforma analítica comprobada para ofrecer puntuaciones de crédito y marketing para los consumidores que usan datos de comportamiento del teléfono móvil. La Doctora Xu tiene experiencia en análisis de Big Data, cloud computing, modelado estadístico, aprendizaje automático, optimización de operaciones y gestión de riesgo. Fue Directora de Analítica Avanzada en Liberty Mutual y Gerente de Gestión de Riesgo Empresarial de Travelers Insurance. La Doctora Xu tiene un doctorado en estadística de la Universidad de California, Davis y una certificación como Gerente de Riesgo Financieros de The Global Association of Risk Professionals.

## The Partnership for Financial Inclusion

The Partnership for Financial Inclusion es una iniciativa conjunta de \$ 37.4 millones de la IFC y Mastercard Foundation para expandir las microfinanzas y avanzar los servicios financieros móviles en el África subsahariana. La Asociación también recibe el apoyo de la Fundación Bill & Melinda Gates y del Banco de Desarrollo de Austria (OeEB, Oesterreichische Entwicklungsbank.AG). Trabaja con instituciones microfinancieras, bancos, operadores de redes móviles y proveedores de servicios de pago en todo el continente para probar modelos de negocios innovadores para la inclusión financiera. El programa tiene un componente fuerte de intercambio de información. Este manual es el segundo en una serie de manuales acerca de cómo hacer una implementación exitosa de servicios financieros digitales, es una de muchas publicaciones de la Alianza. Para mayor información y acceso a todos los informes, por favor visite [www.ifc.org/financialeinclusiionafrika](http://www.ifc.org/financialeinclusiionafrika)

## Acerca de IFC

La IFC, miembro del Grupo del Banco Mundial, es la institución global de desarrollo más grande y centrada exclusivamente en el sector privado. Trabajando con más de 2.000 negocios mundialmente, utilizamos nuestro capital, conocimiento experto, e influencia, para generar oportunidades donde más se necesitan. En el año fiscal 2015, nuestras inversiones a largo plazo en países en desarrollo subieron a casi \$18 billones, ayudando al sector privado a jugar un papel esencial en el esfuerzo global de acabar con la pobreza extrema e impulsar la prosperidad compartida. Para mayor información, visite [www.ifc.org](http://www.ifc.org)

## Acerca de Mastercard Foundation

Mastercard Foundation trabaja con organizaciones visionarias para proveer mayor acceso a la educación, capacitación técnica y servicios financieros para las personas que viven en pobreza, primordialmente en África. Como una de las fundaciones independientes más grandes, su trabajo es guiado por su misión de avanzar el conocimiento y promover la inclusión financiera para aliviar la pobreza. Con sede en Toronto, Canadá, su independencia fue establecida por MasterCard cuando se creó la Fundación en 2006. Para mayor información y para registrarse para el boletín de la Fundación, por favor visite [www.mastercardfdn.org](http://www.mastercardfdn.org).

Este manual es uno de los tres manuales sobre Servicios Financieros Digitales publicados por The Partnership for Financial Inclusion, una iniciativa conjunta de la IFC y Mastercard Foundation para fomentar la expansión de la inclusión financiera. Los otros dos manuales están disponibles a solicitud desde IFC o para descargar desde el sitio web de la Alianza: [www.ifc.org/financialeinclusiionafrica](http://www.ifc.org/financialeinclusiionafrica):



El **Manual de Canales de Distribución Alternativos y Tecnología** ofrece una guía práctica, paso a paso, para desarrollar canales de distribución alternativos que vinculan opciones de tecnología con el proceso general del negocio.



El **Manual de Servicios Financieros Digitales y Gestión de Riesgo** está diseñado para cualquier tipo de institución financiera que ofrece, o planea ofrecer, servicios financieros digitales, y presenta una perspectiva general de los riesgos relacionados y la mejor manera de aplicar un marco de gestión de riesgo para hacer frente a estos.

## CONTACT DETAILS

Anna Koblanck  
IFC, África Subsahariana  
[akoblanck@ifc.org](mailto:akoblanck@ifc.org)

[www.ifc.org/financialeinclusiionafrica](http://www.ifc.org/financialeinclusiionafrica)

2017

