

Big Data en la industria de las telecomunicaciones

Casos de uso prácticos apalancados por analítica avanzada

Mg. Alejandro M. Salevsky
Lic. Diego Tassara
Ing. Claudio Lupi

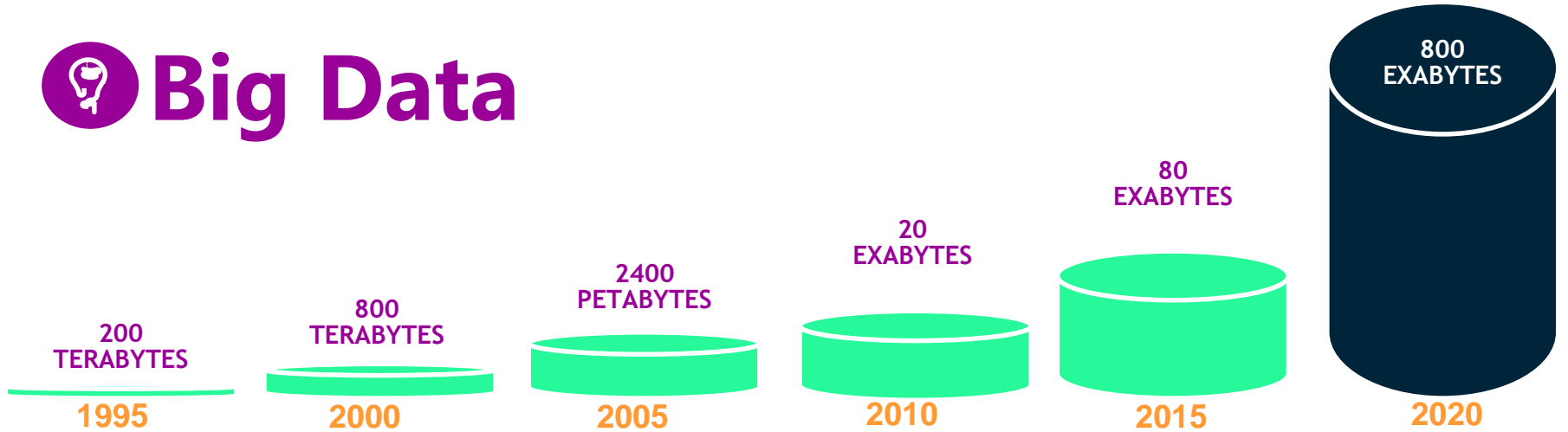


Telefónica

1. Para qué **Big Data** en una Telco?
2. Caso de uso 1: **Smart Cities**
3. Caso de uso 2: **Score** avanzado de riesgo

Volumen mensual de Información

Big Data



42K Llamados /
sms

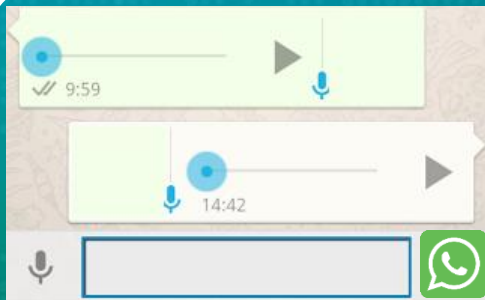
830K sesiones de
datos móvil



4,2K Palabras por
minuto

1.2 MM peticiones a
servidores





Telefónica BUSINESS SOLUTIONS



41 minutes ago

So over the moon to announce the birth of our baby boy Finley born this morning.... It's amazing to be mummy & daddy! Love him so much x

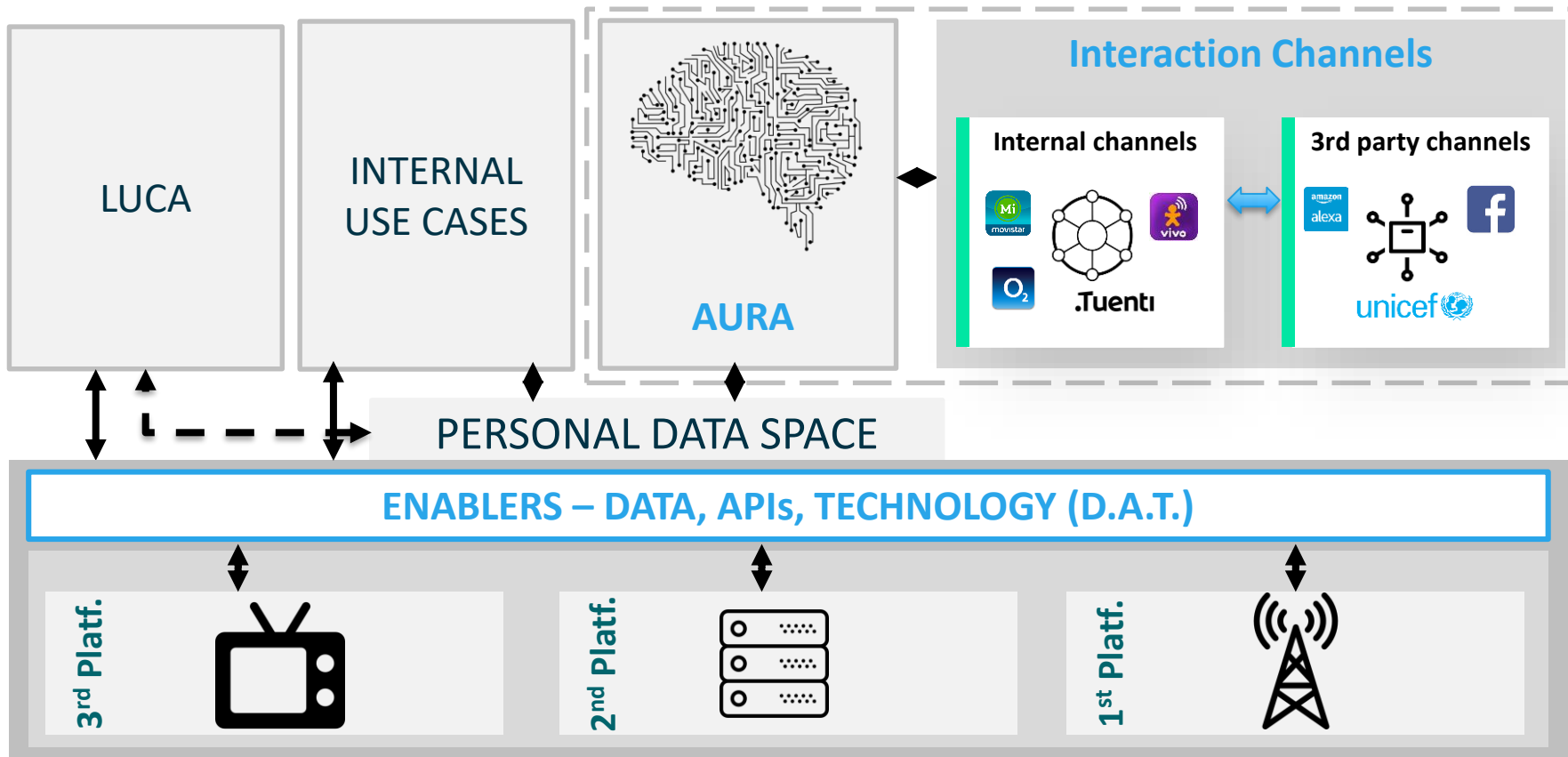
Like · Comment · Share

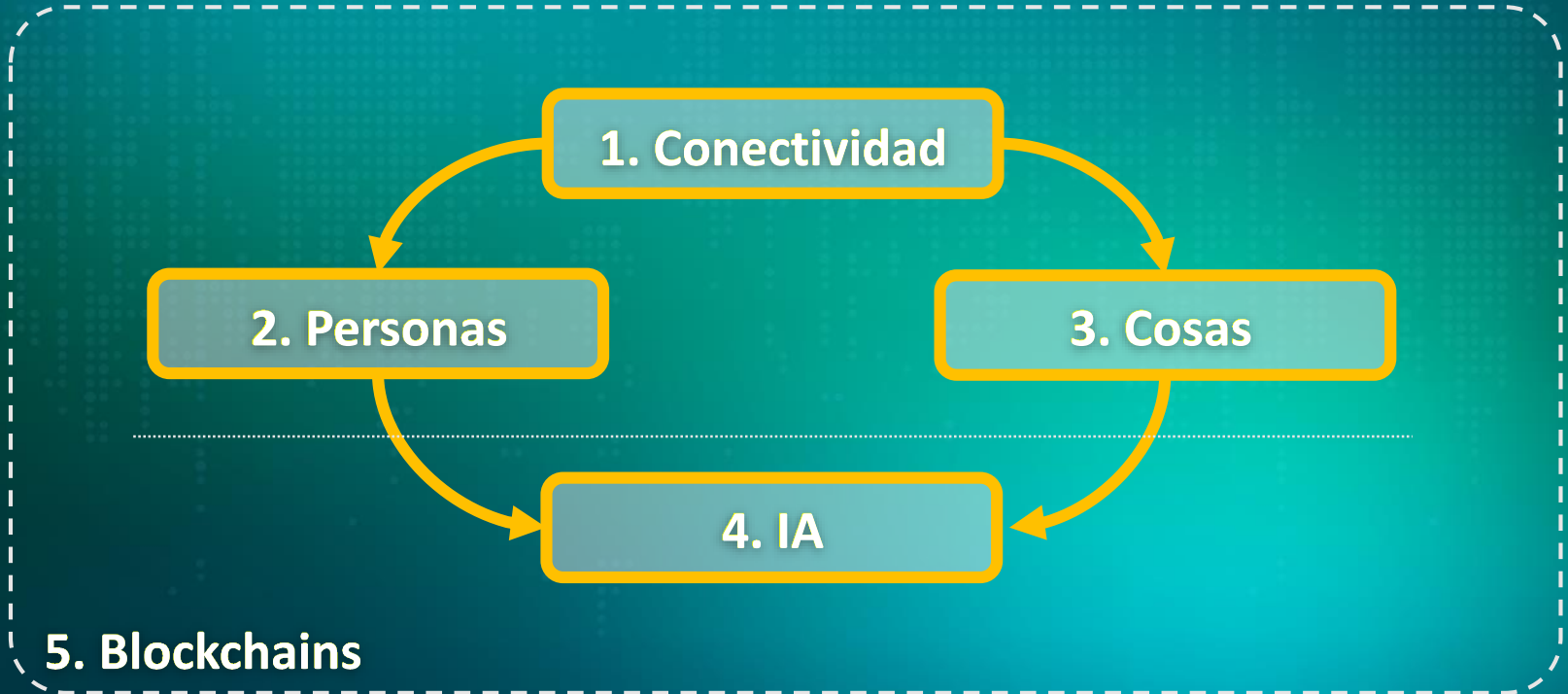
147





Una mirada conceptual de arquitectura de Big Data





1.

Conectividad

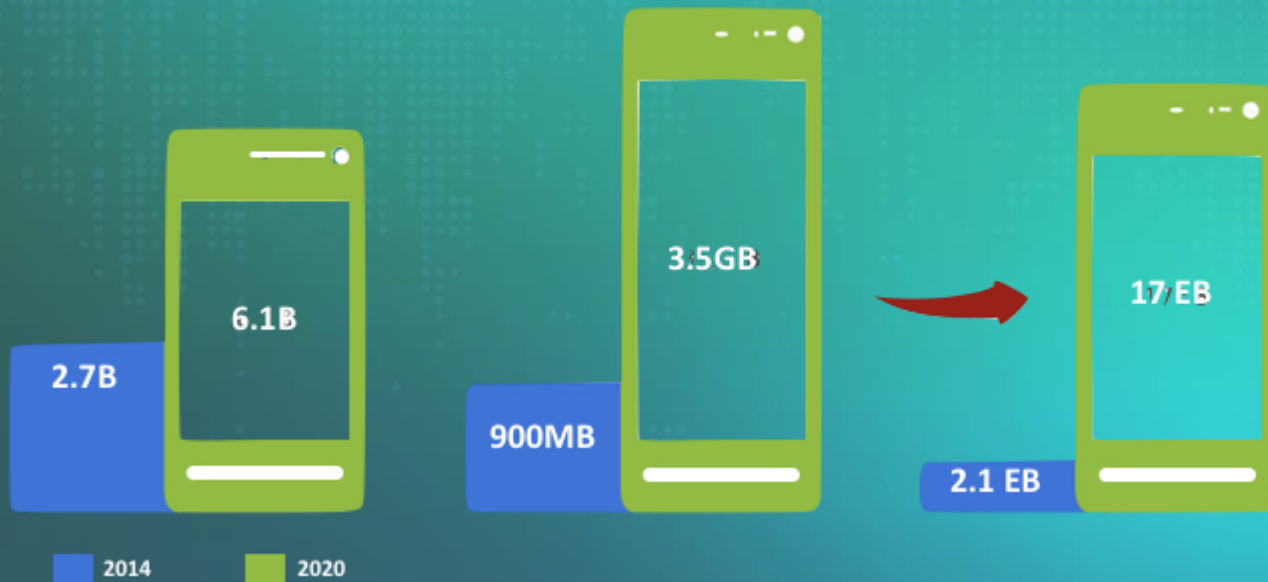
52% de la población mundial con cobertura de internet en **2020**

2. Personas

Total de Smartphones en uso

Datos por Smartphone

Total de datos



2. Personas



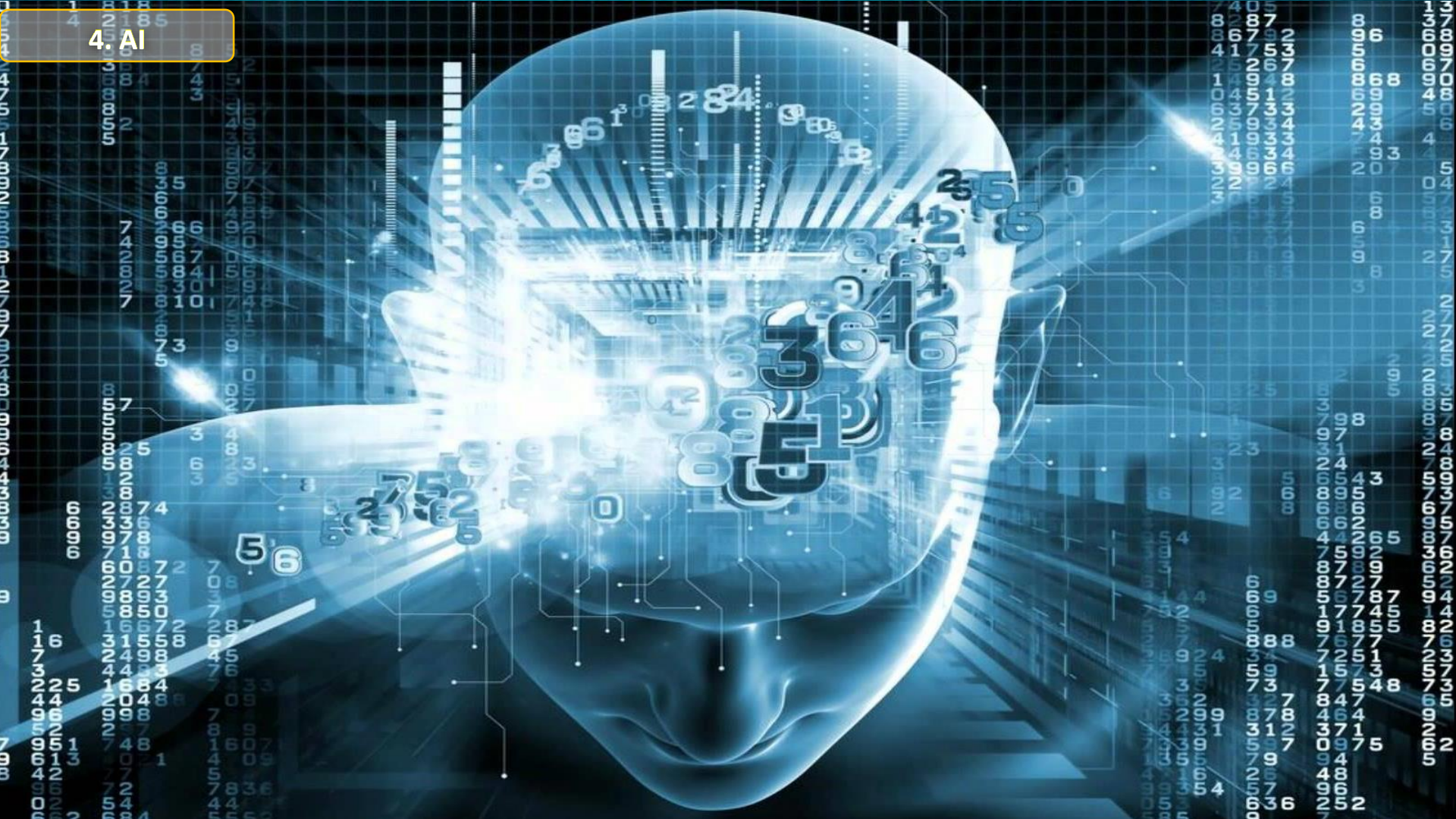
+





25B de dispositivos conectados en el **2020**

4. AI



El avance significativo en el poder de procesamiento de datos ha dado lugar en los últimos años a una nueva generación de estadística llamada "Machine Learning"...

4. AI

Evolución de algoritmos estadísticos

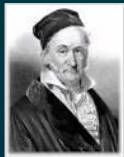
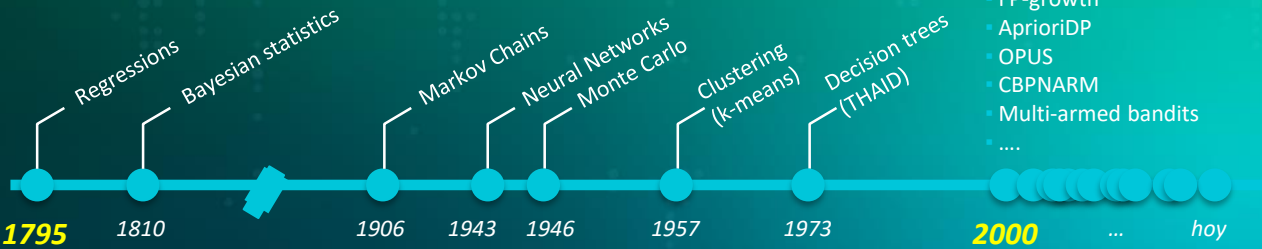
Classical stats (pre Big Data era)

Big Data native stats (Machine Learning)

- Support Vector Machines
- Gradient Boosting (GBM)
- Boltzmann Machines
- Learning Vector Quantization (LVQ)
- Random Forest
- Adaboost
- FP-growth
- AprioriDP
- OPUS
- CBPNARM
- Multi-armed bandits
-

Velocidad de Procesamiento + Explosión de datos

Netflix
Google
Amazon
Waze
...



Gauss



Von Neumann



Morgan



Breiman

Diversas áreas e industrias están y serán impactadas por la IA. Por ejemplo:

4. AI

Automotriz



1 GB / s de procesamiento de datos

Salud



55% de crecimiento de IoT en 2018



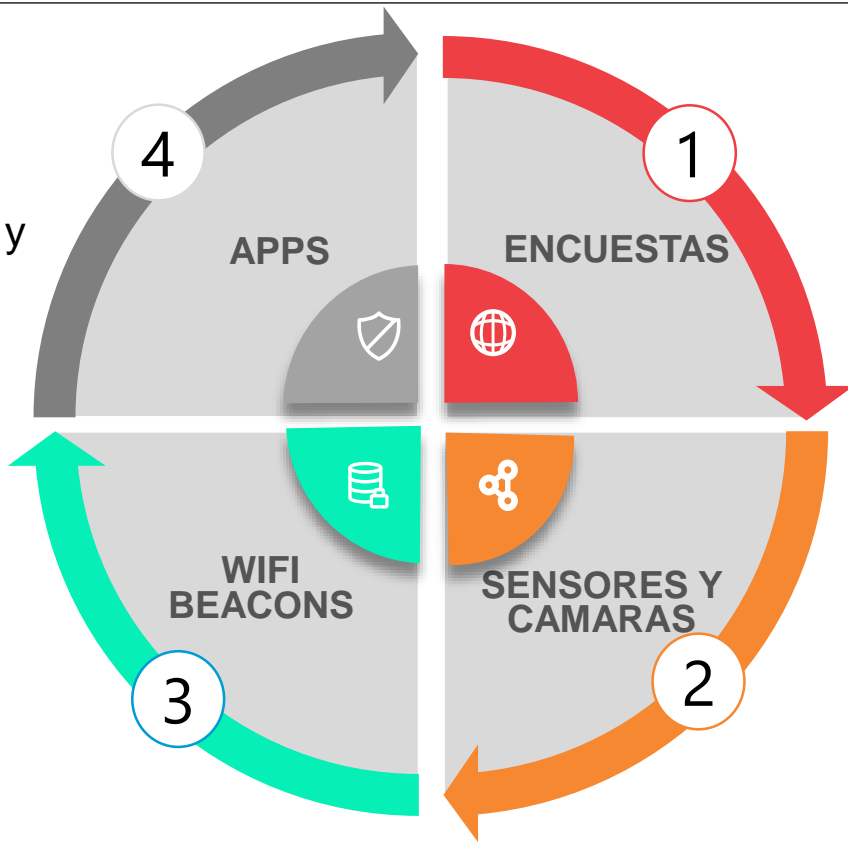
Cómo transformar datos
en **insights** para una
ciudad inteligente?



LOS INSIGHTS TRADICIONALES SON INSUFICIENTES PARA LAS NUEVAS DEMANDAS

Depende de si el usuario tiene la aplicación descargada y los servicios de localización activados. Muestra sesgada

Depende de que el usuario tenga el WiFi encendido. Muestra sesgada.

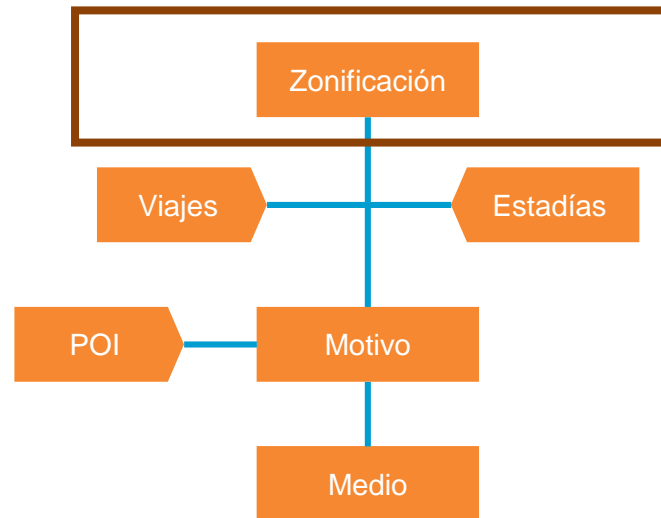
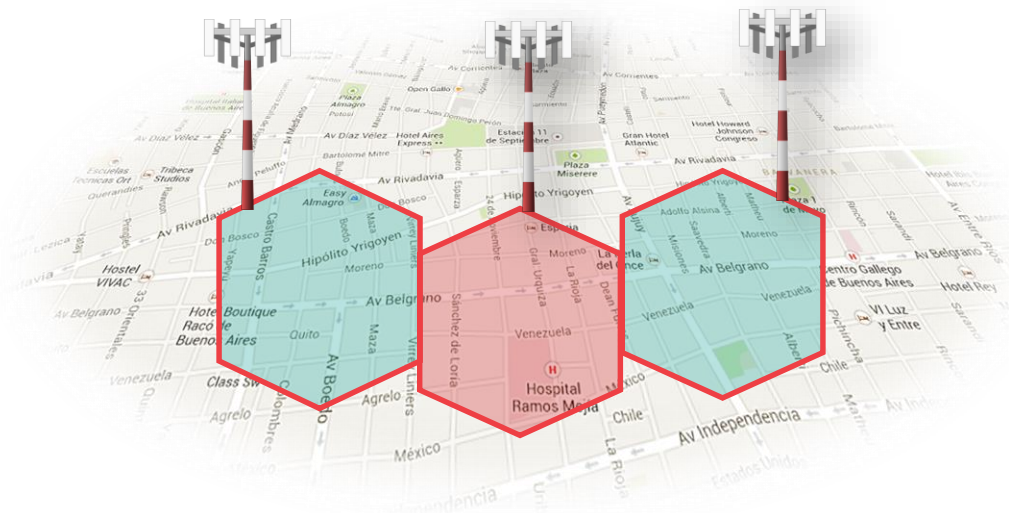


Muestras pequeñas, comportamiento declarado, poco ágil y caro.

Infraestructura cara para instalar, falta información de perfilado y procedencia.

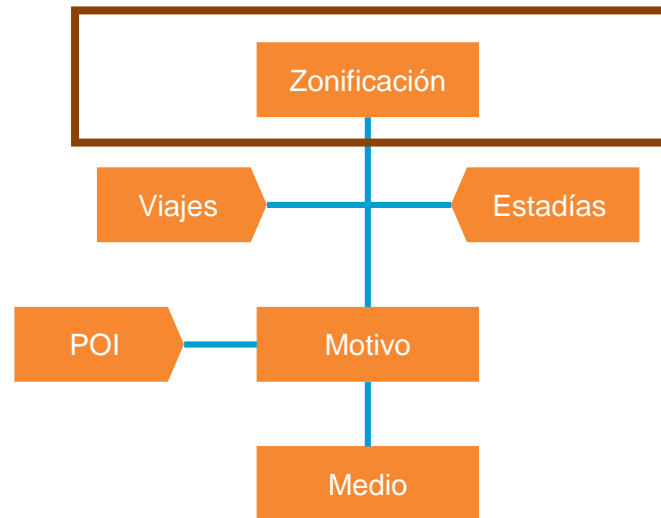
INPUTS GENERADOS

Zonificación



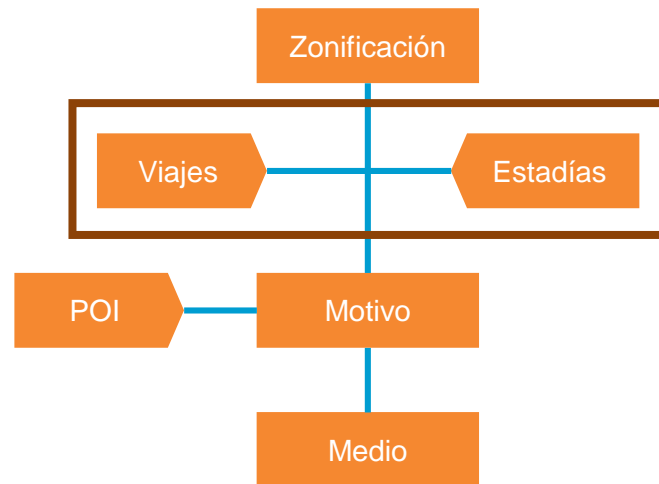
INPUTS GENERADOS

Viajes - Estadías



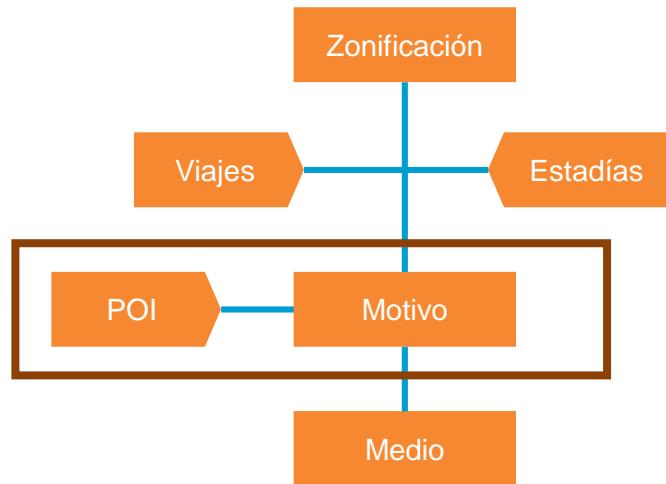
INPUTS GENERADOS

Motivo de viaje



INPUTS GENERADOS

Motivo de viaje



¿CÓMO SE PROCESA LA INFORMACIÓN QUE SE ALMACENA EN LA PLATAFORMA SMART STEPS?

PROCESO OBTENCIÓN DE INFORMACIÓN



EXTRACCIÓN

Se extrae la información de los sistemas de red sobre cada desplazamiento (usuario móvil) y su localización en cada celda



ANONIMIZACIÓN

Todos los datos personales se eliminan y se reemplazan por un identificador que impide dar marcha atrás.

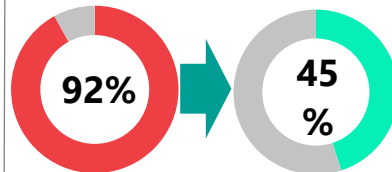


Se eliminan los datos personales



EXTRAPOLACIÓN

Una vez anonimizados, se extrapolan, de modo que se pueden establecer conclusiones sobre la población total (no solo clientes de Telefónica)



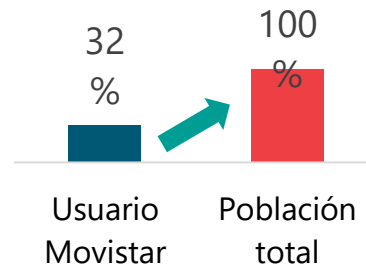
Usuarios Nacionales

Hombres



AGREGACIÓN

Una vez anonimizados y extrapolados se agregan de modo que se cumplan las restricciones impuestas por la Dirección Nacional de Protección de Datos de Argentina.



LA SOLUCIÓN LUCA TRANSPORTE PROVEE INFORMACIÓN AGREGADA DE ALTO VALOR

La información recogida y procesada en **LUCA TRANSPORTE** genera una **Matriz Origen - Destino** con los siguientes atributos:

DÍA Y FRANJA HORARIA

Discriminando por día de la semana y hora del día

MODO

Por carretera, tren, barco o avión para desplazamientos de más de 50Km

FRECUENCIA

del desplazamiento diferenciando entre esporádicos y habituales

MOTIVO

Movilidad ocupacional, movilidad personal, vuelta a casa ocupacional y vuelta a casa personal

VOLUMEN DE DESPLAZAMIENTOS

Trabajamos con distintos niveles geográficos: barrio, distrito, municipio, provincia, comunidad autónoma o agregaciones de secciones censales

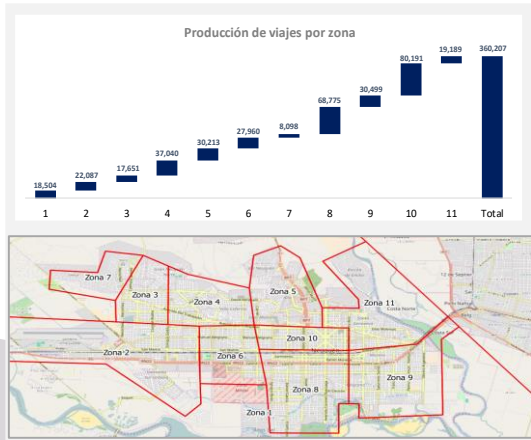
PERFIL DEL USUARIO

Género, edad, lugar de residencia nivel socioeconómico

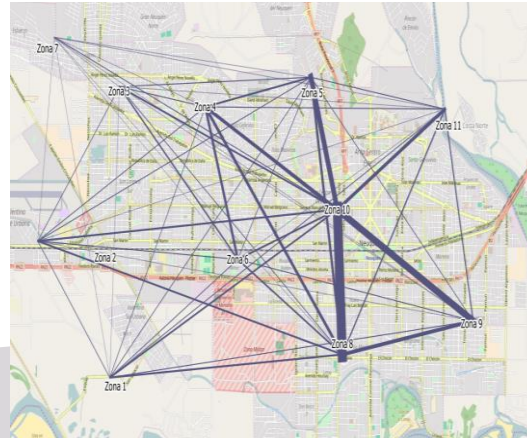


CASO DE ÉXITO: LUCA TRANSPORTE- MUNICIPALIDAD DE NEUQUÉN

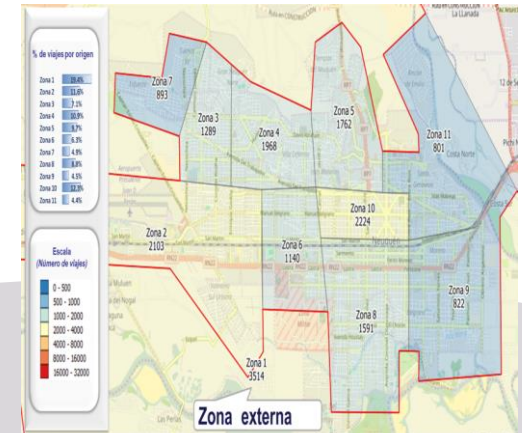
VIZUALIZACIÓN EN MAPA DE LA POBLACIÓN EXISTENTE



LINEAS DE VIAJES (Casa – Trabajo)



MAPAS DE CALOR DE LAS ZONAS





Telefónica

Cómo **crear** un score de
riesgo con **Big Data**?

Agenda



- **Modelo actual:** Universo scoreable, metodología del cálculo, variables. Análisis de sensibilidad.



- **iScore:** Generalidades del proyecto, visión del cálculo, estado general del proyecto.

1.- Modelo actual: marco conceptual

- ✓ **Tipo de modelo:** Modelo basado en metodología de credit risk scorecard.
- ✓ **Período de análisis:** Últimos 6 meses.
- ✓ **Universo scoreable:** Clientes móviles individuos pospagos con 3 o más facturas emitidas en el periodo de análisis.
- ✓ **Frecuencia:** 1 vez al mes.
- ✓ **Target 1:** Clientes morosos.
- ✓ **Score modelo:** El modelo arroja valores en el rango de 100 a 950.
- ✓ **Cálculo:**

$$\text{Score} = 104 + f(x_1, x_2, \dots, x_N) = 104 + f_1(x_1) + f_2(x_2) + \dots + f_N(x_N)$$

↑ Piso base

↑ Variables

— Cada variable aporta al score de forma independiente de las demás.

1.- Modelo actual: variables principales

- ✓ Débito automático
- ✓ Antigüedad
- ✓ Facturación promedio
- ✓ Zona geográfica de facturación.
- ✓ Velocidad de crecimiento ascendente de facturación
- ✓ Promedio de diferencia de días entre fecha de pago y de vencimiento de c/factura
- ✓ Suspensiones por falta de pago
- ✓ Porcentaje de saldo adeudado a 30 días del vencimiento.

1.- Modelo actual: análisis de sensibilidad – Cambio de terminales financiados contra factura

CATERS EFECTIVOS:

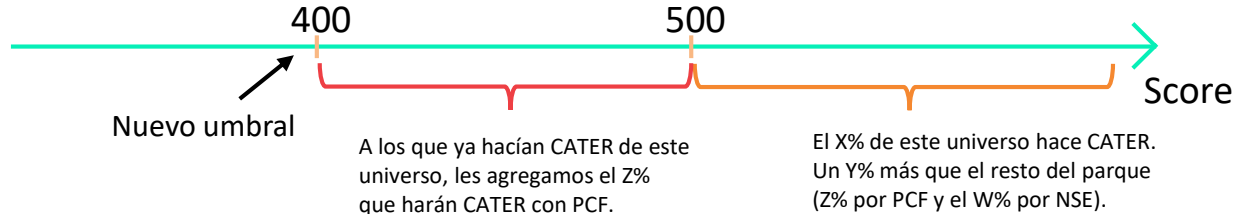
- ✓ Este concepto hace referencia a considerar los 2 escenarios que enfrentamos en las operaciones de CATER.
 - ❑ CATER NO MOROSO: operación de cater abonado. A nivel cantidad, suma en 1 unidad.
 - ❑ CATER MOROSO: operación de cater no abonado. A nivel cantidad, resta tantas veces como operaciones de CATER NO MOROSOS son necesarias para cubrir el costo en el que se incurrió al no cobrar la venta del terminal.
- Costo
- $C. \text{ CATER MOROSO} = C.D. \text{ DEL TERMINAL} + C.O. \text{ DEL BENEFICIO}$
 - Para ser comparado con un CATER NO MOROSO, lo dividimos por su margen bruto, y definimos así los #CATERS EFECTIVOS:

$$\#CATERS \text{ EFECTIVOS} = \#CATERS \text{ NO MOROSOS} - \frac{\#CATERS \text{ MOROSOS}}{MARGEN \text{ BRUTO}}$$

1.- Modelo actual: análisis de sensibilidad – Cambio de terminales financiados contra factura

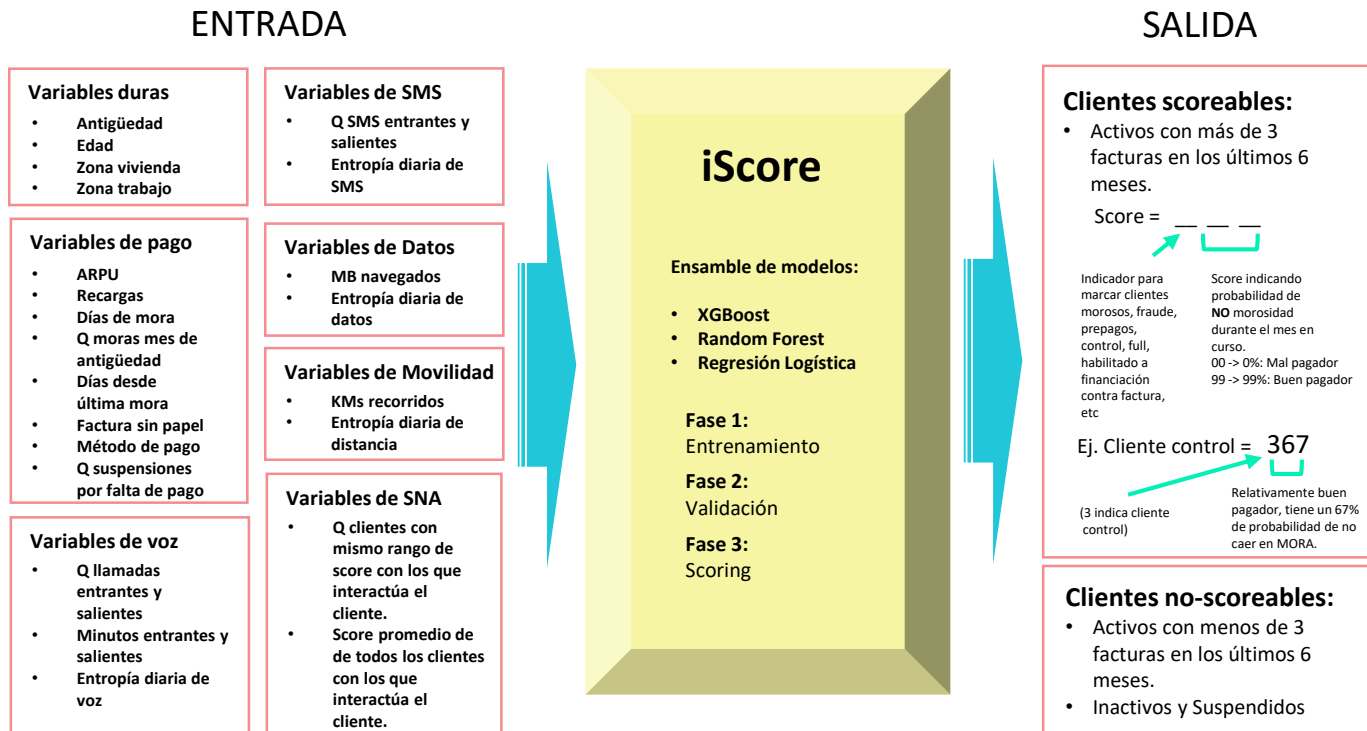
SIMULACIÓN DE CAMBIO DE UMBRAL CF (Actual Score 500)

- ✓ Haremos una simulación donde se hará un barrido del score umbral para la habilitación de cambio de terminal con pago contra factura (actual score 500).
- ✓ El score umbral lo haremos variar desde 100 a 750.
- ✓ Para cada caso, supondremos que los clientes por arriba del score umbral, harán cater CF en la misma proporción incremental que lo hacen los actuales clientes con $\text{Score} \geq 500$.



- ✓ Pero en esa misma proporción, entrarán en mora los clientes que hacen caters incrementales.

2.- iScore – El modelo adaptativo



Recalibración mensual automática