



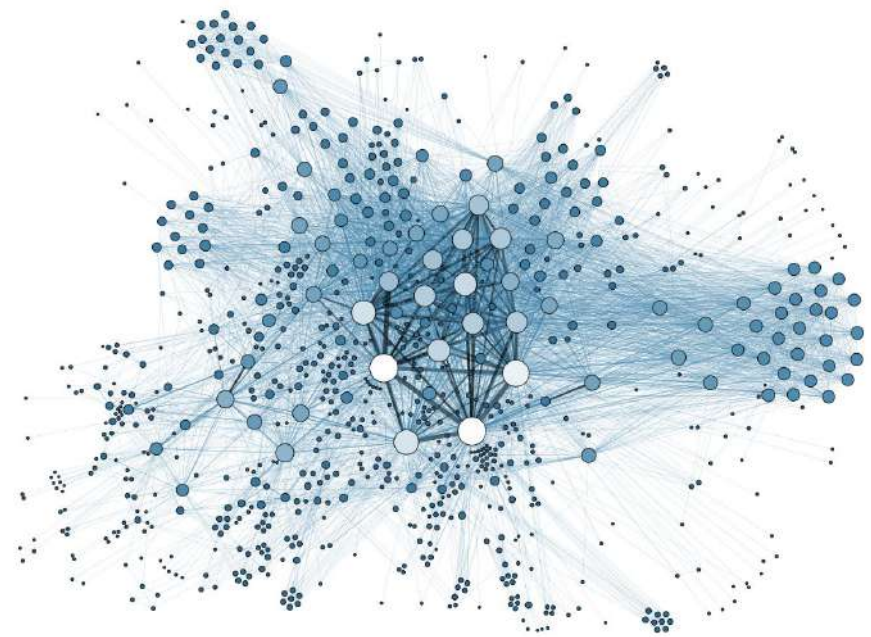
Curso de
Análisis
de Negocios para
Ciencia de Datos

Sílvia Ariza Sentís

¿Qué es ciencia
de datos y big data?
¿Cómo afectan
a mi negocio?

Aprender cómo surge
la ciencia de datos, en
qué consiste y dónde
está presente.

Big data



Gran volumen de información.

Las empresas **almacenan** datos diariamente.

Costos, eficiencia, entender a los clientes.

Solución matemática a un problema de negocio.

¿Qué tipo de
información podemos
analizar?

Descubrir qué tipos de información existen, qué industrias los usan y qué tipo de acciones podemos tomar a partir de ellos.

Tipología de datos

Personas

Transacciones

Navegación web

Machine 2 Machine

Biométricos



¿Cómo crear empresas
data-driven?

**Entender qué componentes
son necesarios para crear
una cultura data-driven
en una empresa.**

Empresas data-driven

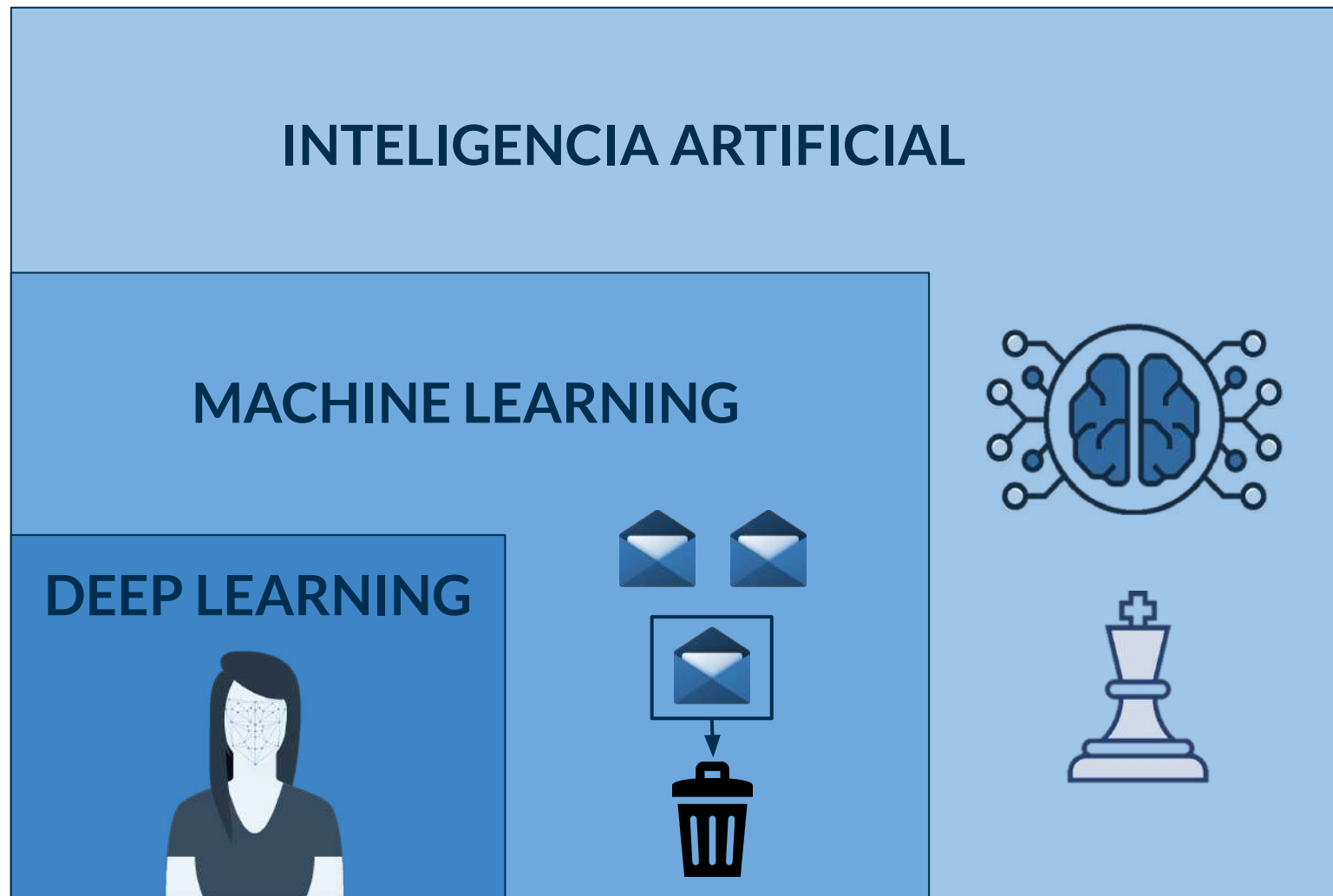


1. Crear una **cultura** de datos.
2. **Recolectar** información.
3. Medir **todo**.
4. Datos **relevantes** y **precisos**.
5. Testear y crear **hipótesis**.
6. Desde los **insights** de datos a las **acciones**.
7. Cumplir las **regulaciones** de datos.
8. **Automatizar**.

¿Qué es machine
learning?

Aprender cómo surge
el machine learning,
en qué consiste y dónde
está presente.

Contexto



Machine learning

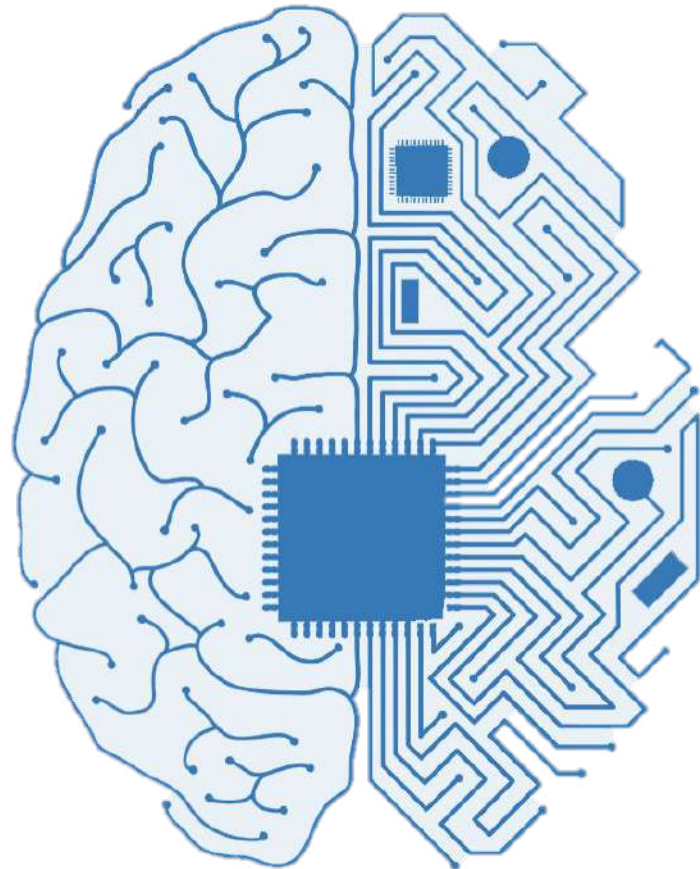
Detección de fraudes

Búsqueda web

Anuncios a tiempo real

Análisis de textos

Next best action



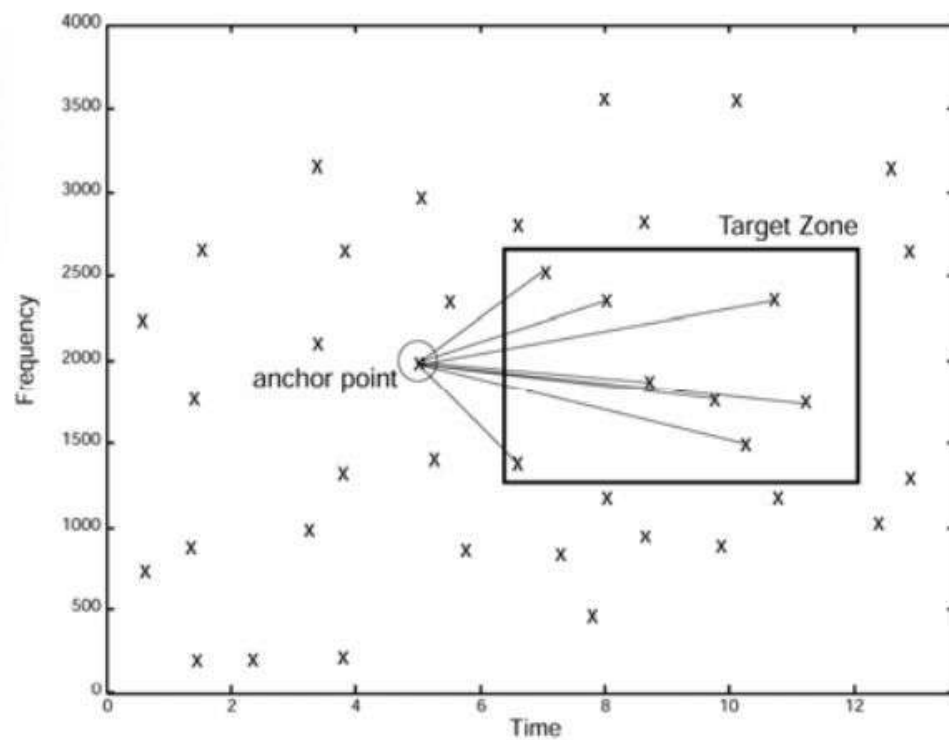
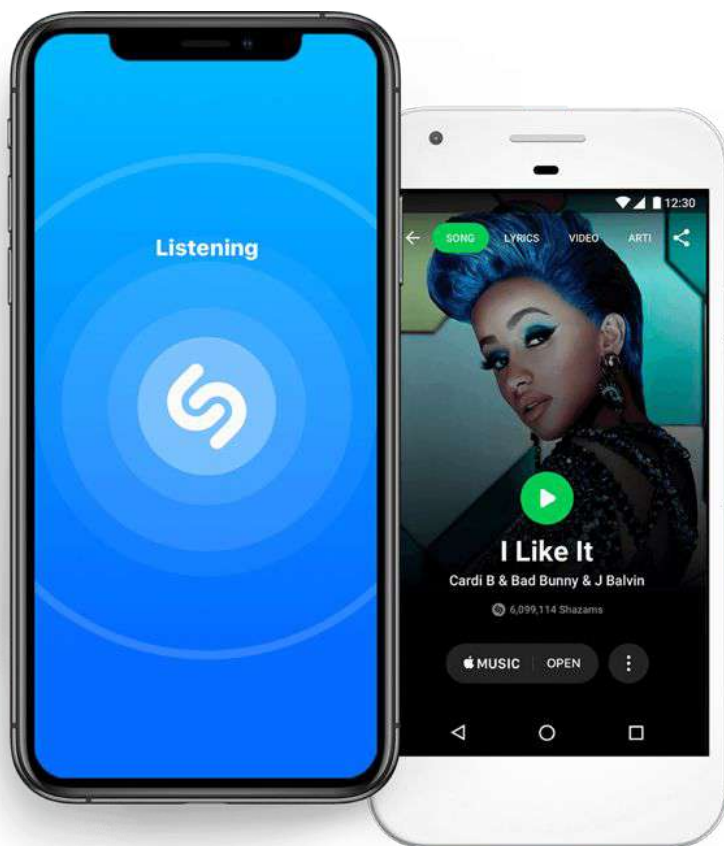
¿Qué es
deep learning?

**Aprender en qué consiste
el deep learning y dónde
está presente.**

Deep learning



Deep learning



Deep learning



Flujo de trabajo en ciencia de datos: fases, roles y oportunidades laborales

**Aprender las diferentes fases
de los datos y los diferentes
empleos que existen en cada
una de ellas.**

Roles en datos

INGENIERO/A DE DATOS

Base de Datos

ETLs / APIs

SQL y NoSQL

ANALISTA BUSINESS INTELLIGENCE

Extracción y dashboards

Automatización

SQL y Excel

DATA SCIENTIST

Machine Learning

Modelos estadísticos

R y Python

DATA TRANSLATOR

Data Scientist - Decision Makers

Destiladores de data

Expertos necesidad de negocio



Herramientas para cada etapa del análisis de datos

**Entender cuál es la función
de los programas más
usados en cada etapa del
estudio de la información.**

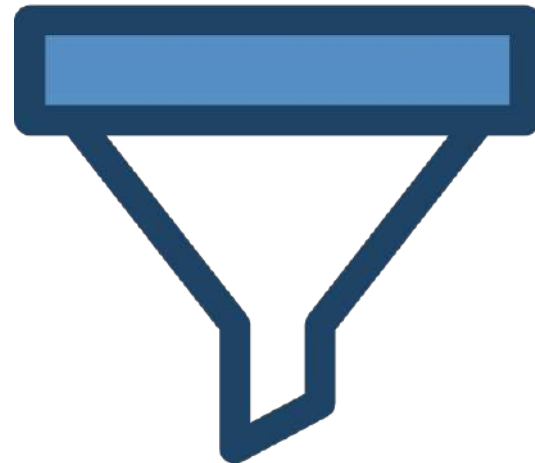
Extracción de información con SQL



Extracción de información

Síntesis de la base de datos

Cuadros de control de la operación



Análisis y visualización con R y Python

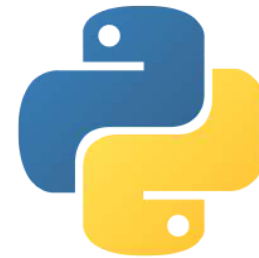


Análisis de datos con
enfoque estadístico

Más complejo al inicio

Análisis descriptivo
y exploratorio

Packages: ggplot2, dplyr



Análisis de datos con
enfoque en ingeniería

Parecido a otros lenguajes

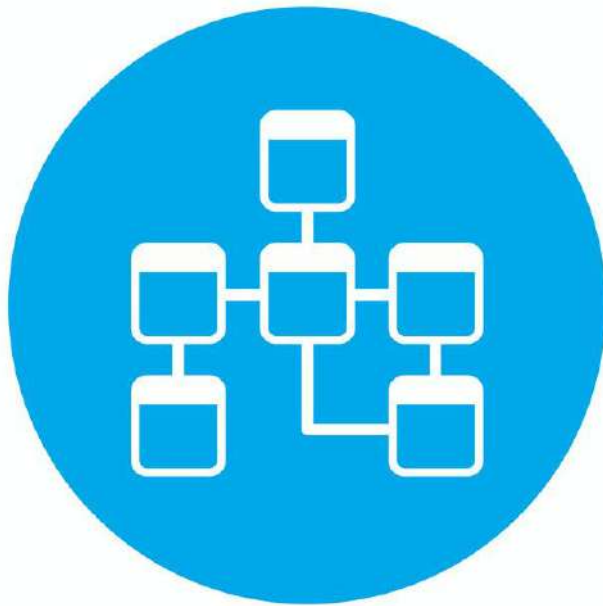
Análisis descriptivo
y exploratorio

Packages: Pandas, Numpy

Bases de datos y funciones principales de SQL

**Entender la sintaxis
y principales funciones
de SQL.**

Bases de datos



Ventas

orden_id	Fecha	Valor	cliente_id
1	2020/08/01	1,400MXN	2
2	2020/06/14	1,500MXN	4
3	2020/02/24	5,400MXN	4
4	2020/04/18	280MXN	6
5	2020/04/16	8,800MXN	3

Instrucciones de comandos SQL

COMANDOS **SELECT**

Selección de los campos (columnas)
para hacer el análisis o para
sintetizar la tabla de origen.

Instrucciones de comandos SQL

CLÁUSULAS

FROM Tabla donde se almacena la información.

WHERE Especificar las condiciones.

GROUP BY Campos (columnas) de agrupación.

ORDER BY Campos (columnas) de ordenación.

Instrucciones de comandos SQL

OPERADORES LÓGICOS

AND Une varias condiciones que tienen que ser cumplidas para obtener resultados.

OR Evalúa dos o más condiciones y obtienes resultados si una de ellas se cumple.

NOT Excluye un valor de la información a obtener.

Instrucciones de comandos SQL

FUNCIONES AGREGADAS

AVG Promedio (average) de un campo (columna).

COUNT Recuento de valores de una columna.

DISTINCT Encontrar valores únicos.

SUM Suma de valores de una columna.

MAX Valor más alto de una columna.

MIN Valor más bajo de una columna.

Instrucciones de operadores SQL

IN	DISTINCT	BETWEEN	ROUND
=	-	*	+
()	''	<>	ELSE
CASE WHEN	THEN	END	AS

Nueva columna a partir de otro campo (columna)

CASE WHEN Mexicano = 'Yes' **THEN** 'Mexicano' **ELSE** 'Extranjero' **END AS** Nacionalidad

Creación de queries en SQL

Queries para
extraer información
de una tabla.

Ejemplo de SQL

Objetivo

Saber CUÁNTAS bocinas hemos vendido por más de 600 MXN desde 2019

Tabla en Excel

VENTAS_2020

Día	Mes	Año	Producto	id	Valor
1	2	1998	Bocina	24	\$528
12	4	2004	Auriculares	31	\$240
14	8	2016	Auriculares	14	\$315
16	10	2019	Bocina	200	\$1,050
21	12	2020	Bocina	304	\$680

Comprensión

TABLA: VENTAS_2020

COLUMNAS

Día (1-31)

Mes (1-12)

Año (1990-2020)

Producto: bocinas
y auriculares

Valor (MXN)

Código SQL

```
SELECT COUNT (DISTINCT id)
FROM VENTAS_2020
WHERE Producto = 'Bocina'
AND Valor > 600
AND Año >= 2019
```

Resultado

2

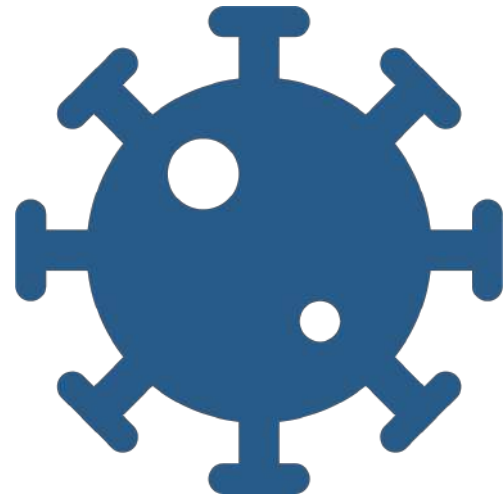
Conflictos y retos actuales sobre la ética y tratamiento de datos

**Distinguir información
sensible y crear un
criterio ético sobre
los usos de los datos.**

Ética en el procesamiento de imágenes



Ética en el procesamiento de imágenes



COVID 19 - HONG KONG

Aplica técnicas de
Storytelling para
convertir problemas
de datos en historias

**Aprender a desagregar
(extraer) un problema
en una estructura lógica.**

Estructura del problema

PROBLEMA

Algunos clientes contactan a soporte en exceso.

No los podemos identificar.

No podemos prevenir este comportamiento.

SOLUCIÓN

Script que identifique y clasifique a los Top Offenders.

Entender sus motivaciones - clasificarlos.

Definir acciones para prevenir esta tendencia.

ALCANCE

LATAM con distinción por ciudades.

Clientes.

Actualización mensual.

Cómo estructurar un caso de negocio

**Desglosar un problema
de negocio en una
hipótesis estructurada.**

Hipótesis / Storytelling

QUÉ

Algunos clientes contactan a soporte en exceso.

POR QUÉ

- a) Motivaciones económicas
- b) Preguntas
- c) Problemas tecnológicos
- d) Política de empresa

CÓMO

1. Análisis cuantitativo.
2. Análisis cualitativo.
3. Matriz cuantitativa - cualitativa.
4. Definir acciones de prevención.
5. Validación.



Análisis cuantitativo en un caso de negocio

**Identificar las variables
cuantitativas que nos
ayudarán a resolver
el ejercicio.**

Análisis cuantitativo

DESCARGAR INFORMACIÓN

Clientes con ≥ 1 queja
Datos por un mes
Macros por ciudad y mes

IDENTIFICAR

Patrones de comportamiento
Variables significativas

- a) Madurez (compras realizadas)
- b) Quejas mensuales (contactos)
- c) Compras mensuales
- d) Gasto mensual
- e) Créditos y dinero devuelto
- f) Margen operativo neto

DEFINIR

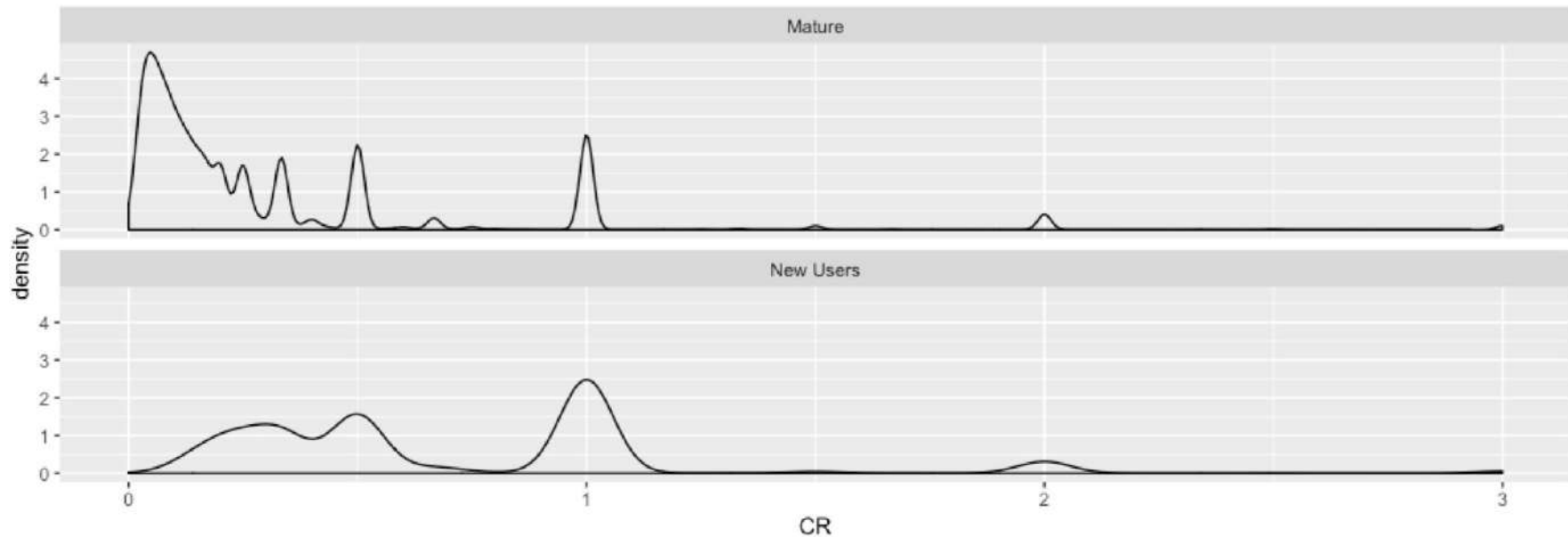
Segmentación según rentabilidad
Threshold (límite) Top Offender
Threshold para cada categoría

Clientes regulares (9 compras o menos).
Clientes bronce (10-19 compras).
Clientes plata (20-39 compras).
Clientes dorados (40 compras o más).

Análisis cuantitativo: mapeo

	Regular 280K	Bronce 56K	Plata 17K	Oro 6K	Total Activos 6M
Promedio compras	3	12	28	52	3
Promedio quejas	1.3	1.5	1.7	2	0.1
Ratio de contacto	43%	12%	6%	4%	3%

Hipótesis



	Regular 280K	Bronce 56K	Plata 17K	Oro 6K
Promedio CR (ratio contacto)	43%	5%	4%	3%
Promedio ajustado CR	12%*			

*Eliminar los clientes que hicieron menos de diez compras en total y 1&1 quejas vs. 1&2 compras (100% CR)

Análisis cuantitativo: aplicación

TOP OFFENDERS

	Regular	Bronce	Plata	Oro
Porcentaje límite	20%	5%	1%	1%
Volumen quejas	35%	17%	7%	7%
# Clientes	40K	3K	200	50

20% de las quejas las hacen estos usuarios

Análisis cualitativo en un caso de negocio

**Identificar las variables
cualitativas que nos
ayudarán a resolver
el ejercicio.**

Análisis cualitativo

**Clusterizar:
causas de contacto**

- a) Motivaciones económicas
- b) Preguntas
- c) Problemas tecnológicos
- d) Política de empresa

**Clasificar:
causas de los TO
identificados**

**Profundizar:
motivos de contacto
geolocalización**

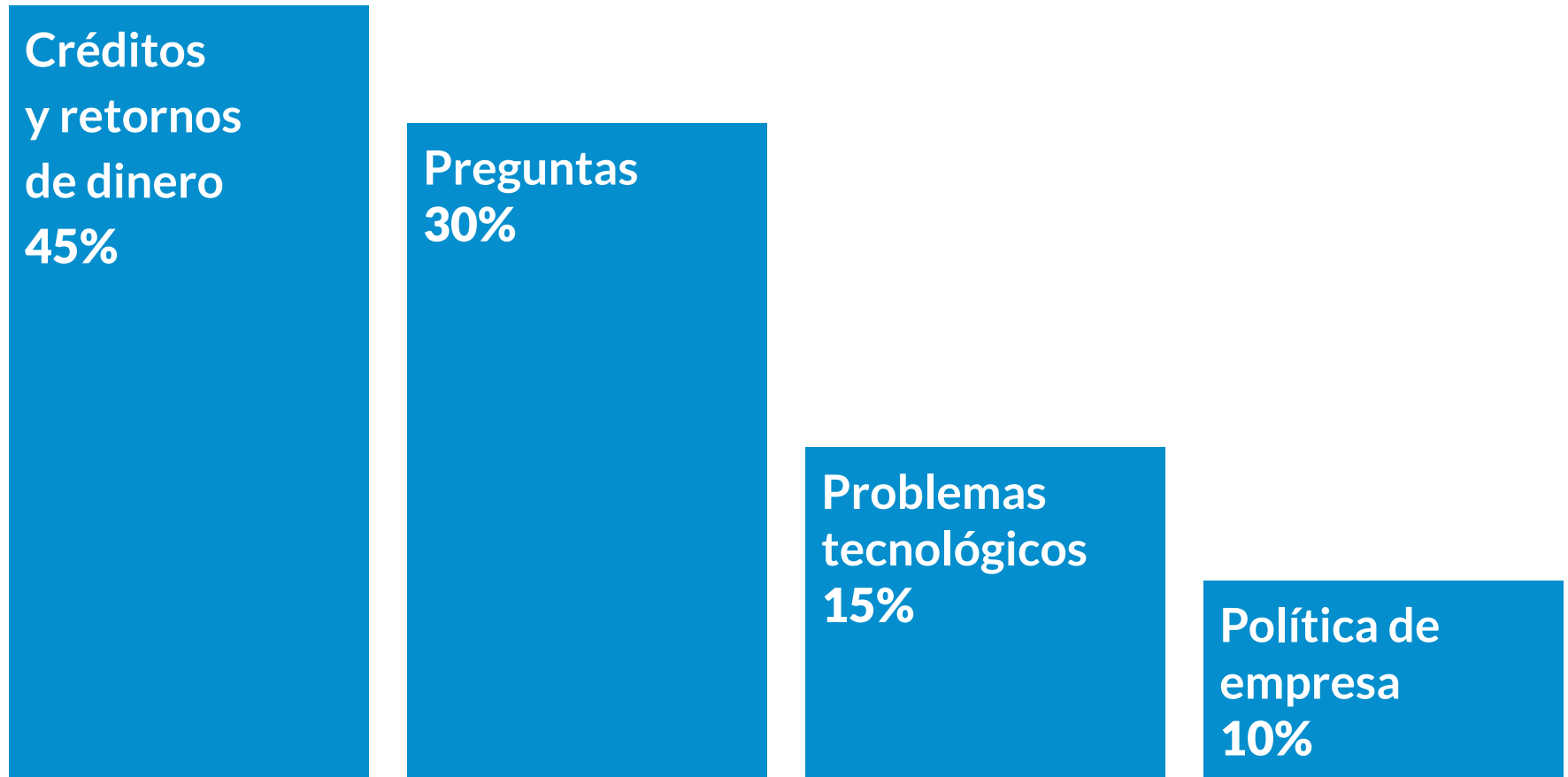
Análisis cualitativo: clusterización

Créditos
y retornos
de dinero
45%

Preguntas
30%

Problemas
tecnológicos
15%

Política de
empresa
10%



Fusión Cuanti-Cualitativa en un caso de negocio

**Resolver de manera
conjunta la información
cuantitativa y cualitativa
para sacar conclusiones.**

Matriz cuantitativa y cualitativa

	Créditos y retornos de dinero	Preguntas	Problemas tecnológicos	Política de empresa
Regular	35%	35%	20%	10%
Bronce	30%	25%	25%	20%
Plata	25%	20%	25%	30%
Oro	20%	25%	20%	35%

¿Qué es minería
de texto?

¿Cómo usarla para obtener
información adicional?

**Explora nueva
información a partir
del texto.**

Motivos de contacto

Regular

Tarifa de devolución
Tasa de envío
Cómo embalar para devolución

Bronce

Tarifa de devolución
Tasa de envío
Estado del producto

Plata

Tarifa de devolución
Facturas
Estado del producto

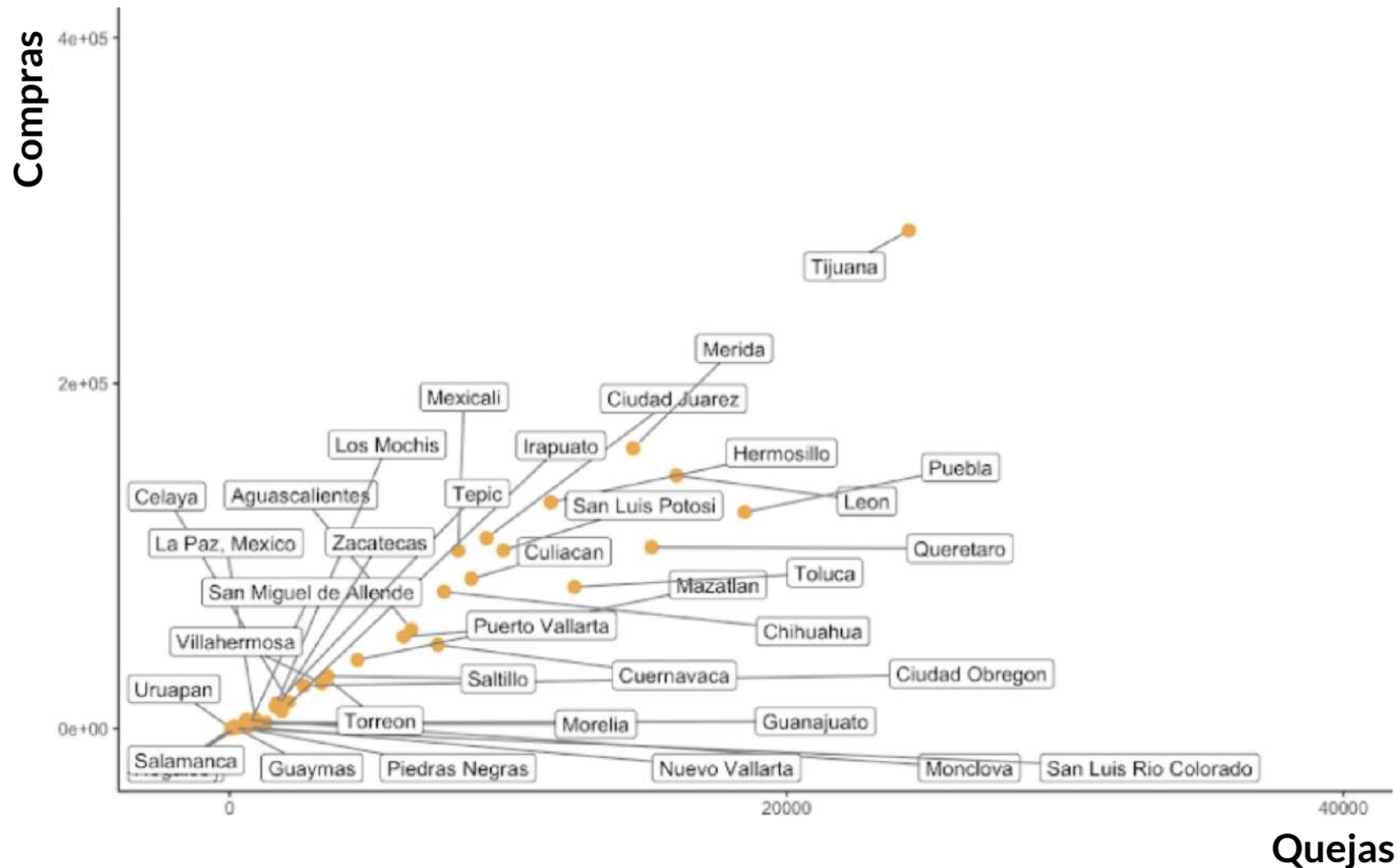
Oro

Facturas
Estado del producto
Log in

Variación de comportamientos a partir de la geolocalización

**Entender la relevancia
de hacer una distinción
geográfica para el estudio
y validarla en el caso de
negocio analizado.**

Geolocalización



**Toma decisiones
según los resultados
del análisis**

**Convertir la información
obtenida del análisis en
una estrategia o toma
de decisiones.**

Acciones derivadas del análisis

Algoritmos usados

- a) Minería de datos para clasificación de motivos de contacto.
- b) Correlaciones y patrones de comportamiento.
- c) Árboles de decisión y teoría de juegos para predecir y tomar decisiones.
- d) Validación con bayesianos y MCMC.

Acciones

- a) Taggear a los Top Offenders identificados mensualmente.
- b) Advertirlos.
- c) Llamar usuarios.
- d) Bloquear usuarios.
- e) Validación con A/B Tests.

Disminuyeron las quejas en un 30% a nivel de LatAm

Continúa aprendiendo
ciencia y análisis de
datos para ejecutar
estrategias efectivas