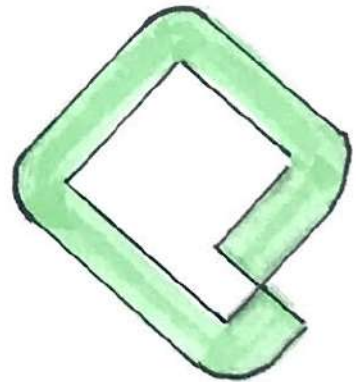
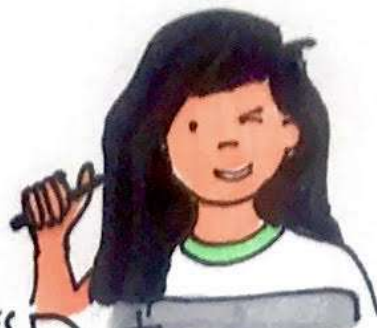


# Análisis de Negocios

para Ciencia de Datos



# ¿Qué es **big data**?



“Datos grandes”

Las empresas tienen una gran cantidad de información traducida en clientes, empleados, ventas, precios...

Las personas que ayudan a convertir los datos en acciones son los **especialistas de datos**.

## ¿Qué **tipos de datos** se usan?

### ≡ TIPOLOGIA DE DATOS ≡

- **Personas**: datos que generamos nosotros mismos al, por ejemplo, dar likes, etiquetar a amigos. De esta manera almacenamos información de preferencias de ciudades, calles, edificios, individuos.



- **Transacciones**: pueden ser monetarias o no monetarias.

→ **Monetarias**: tienen un flujo económico de por medio: transacciones con crédito o débito. Dejan un registro de quien lo hizo, el monto y donde. A partir de esto se realizan sugerencias según las preferencias.

→ **No monetarias**: Pueden ser, por ejemplo, compañías telefónicas que identifican tu patrón habitual: a quién llamas, cuánto duran tus llamadas, etc. A partir de esto pueden detectar si no te comportas 'habitualmente' y enviarte alertas o sugerencias.

- **Navegación web**: las 'cookies', las cuales advierten del registro, qué información recodean. De esta manera, se conoce quien accede a la página, de qué país proviene, qué género, etc.

- **Machine 2 Machine**: Pueden ser los dispositivos GPS, que identifican donde estas y donde hay un conductor cercano. Realiza la locación de dispositivo según distancia.

- **Biométricos**: te identifican como ser único: sangre, saliva, huella dactilar.

# Cultura data-driven

"Dejar de basarnos en información no digital"

Crear una cultura data-driven:

- 1: crear la cultura: entenderla, hacer que los empleados tomen decisiones basadas en datos. Explicar qué son los datos.
- 2: recolectar los datos, almacenarlos, procesarlos.
- 3: medir los datos: entender la cantidad de registros, la velocidad de crecimiento de la información.
- 4: datos relevantes y precisos: identificar qué es relevante y basarnos en datos que sean útiles. Identificar que sean datos precisos.
- 5: Testear y crear hipótesis a partir de estos datos. Hacer preguntas concretas.
- 6: tomar acción basada en los datos.
- 7: cumplir las regulaciones de datos: la ética es importante.
- 8: automatizar: para no realizar el estudio habitualmente, sino automatizarlo y replicarlo.



- **inteligencia artificial**: máquinas inteligentes.
- **machine learning**: aprendizaje automatizado.

# machine learning

¿Para qué se utiliza?

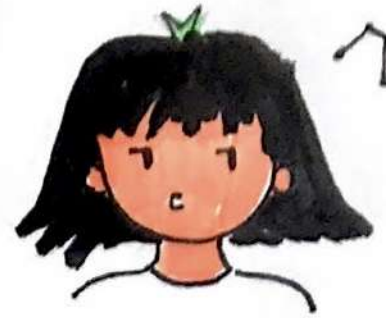


- Detectar Fraudes: mayormente son de ventas virtuales. El algoritmo detecta patrones, quién realizó una acción fraudulenta.
- Búsqueda web: sugerencias a partir de búsquedas. Según horas, días, etc.
- Anuncios a tiempo real: sugerencias de compra cuando estamos en el carrito de la tienda online.
- Análisis de textos: se analiza información categórica.
- Next best action: consiste en saber en qué está el usuario y detectar cual será su siguiente paso. Ser proactivos con la información a partir de la evolución de los usuarios.

El machine learning replica los movimientos, los perfecciona y entiende quién está atrás.

# roles en datos

2



<b>Ingeniero o arquitecto de datos</b>	Genera el entorno de información. Hace la conexión entre dispositivos y la creación de una base de datos. Usan SQL y NoSql.
<b>Analista BI</b>	Crea cuadros de control a partir de una base de datos. Automatiza procedimientos. Explica la situación de la empresa. Usan SQL, Excel.
<b>Data Scientist</b>	Científico de datos. Hace el rol del analista pero también sabe predecir. Explica el futuro con estadística y modelos. Usan R, Python, SQL.
<b>Data translator</b>	Proyecta el equipo. Puede ser un experto en negocios que conozca de datos. Guía en el procedimiento de datos.

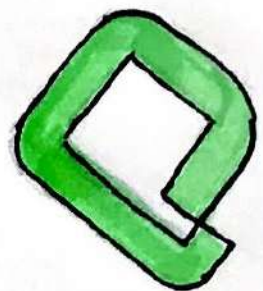


# herramientas para cada etapa del análisis de datos



SQL	Extracción de datos. Sintetiza los datos.
R Python	Herramientas predictivas. Se basan en modelos estadísticos/matemáticos. Para crear gráficos.  R: Enfoque más estadístico. Análisis descriptivo y exploratorio. Packages: ggplot2, dplyr.  Python: Enfoque con base de Ingeniería. Análisis descriptivo y exploratorio. Packages: Pandas, Numpy.

# funciones principales de SQL

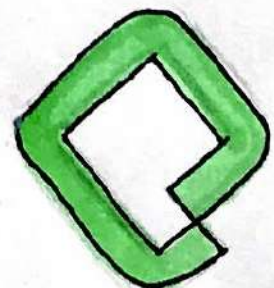
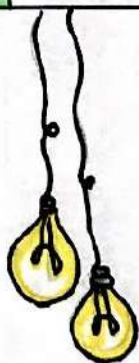


SQL: herramienta que nos ayuda a sintetizar la información, a extraer datos de las tablas.

CLÁUSULAS	COMANDO	Select	selecciona la tabla que quieres elegir. SELECT fecha
	From		tabla donde se almacena la información. FROM ventas
	Where		Especifica las condiciones. 'Solo del 2020' WHERE Fecha = 2020
	Group by		Campos de agrupación.
	Order by		Campos de ordenación.



OPERADORES LÓGICOS	and	Une condiciones que tienen que ser cumplidas para obtener resultados.
	or	Evalúa condiciones y obtienes resultados si una de ellas se cumple.
	not	Excluye un valor de la información a obtener.
FUNCIONES AGREGADAS	AVG	promedio de una columna/campo.
	COUNT	recuento de valores de una columna.
	DISTINCT	Encuentra valores únicos.
	SUM	Suma de valores de una columna.
	MAX	valor más alto de una columna.
	MIN	valor más bajo de una columna.



# Cómo estructurar un caso de negocio



Desglosar un problema:

QUÉ	El problema del negocio. Ej: Quejas de clientes.
POR QUÉ	Clasificar los motivos de las quejas/problemas.
CÓMO	<p>La estrategia de cómo vamos a diseñar y organizar nuestro análisis.</p> <ol style="list-style-type: none"><li>1. <b>IDENTIFICAR</b>: Encontrar la información y hacer una clasificación numérica. Definir y segmentar. Análisis cuantitativo.</li><li>2. <b>ENCONTRAR LAS CATEGORÍAS</b>: En función de texto. Análisis cualitativo.</li><li>3. <b>Juntar la información</b> a partir de una <b>matriz cuantitativa-cualitativa</b>.</li><li>4. <b>Tomar Acción</b> con la info que tenemos.</li><li>5. <b>Validar</b> que las acciones tomen efecto.</li></ol>

