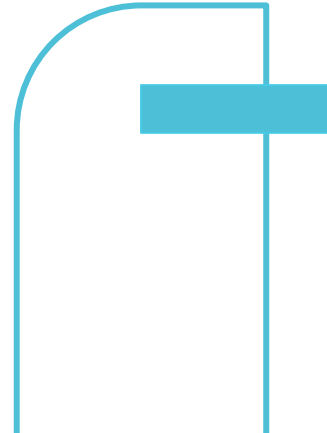


# Módulo 4

## **Tipos y estructuras de datos**



## AGENDA DE LA CLASE

- ✓ Decisiones y datos
- ✓ Formatos de datos
- ✓ Tipo de datos
- ✓ Factores a considerar cuando se recolectan datos
- ✓ Transformación de datos
- ✓ Como informa los datos (Informes y Dashboard)
- ✓ Datos y métricas
- ✓ Pensamiento matemático y Big y Small Data



# 01 - Decisiones y datos



Datos



Análisis de datos



Ayudar a **tomar decisiones** más informadas

Formas para tomar decisiones



La toma de decisiones **basada en datos**: Las empresas y organizaciones utilizan datos para tomar mejores decisiones.

*Por ejemplo: elegir a qué restaurante ir a partir de buscar "restaurantes cercanos"*

La toma de decisiones **inspirada en datos**: explora diferentes fuentes de datos para averiguar qué tienen en común (buscar patrones).

*Por ejemplo: Buscar datos de distintos orígenes para buscar qué tienen en común*

## 02 - Formatos de datos

### 1 Primarios o Secundarios

**Datos primarios:** Recogidos por un investigador a partir de **fuentes de primera** mano.

*Por ejemplo, datos de una encuesta realizada a 20 participantes.*

**Datos secundarios:** Recogidos por **otras** personas o a partir de otras investigaciones.

*Por ejemplo, datos comprados de los perfiles de clientes de una empresa local.*

### 2 Internos o Externos

**Datos internos:** Alojados en los **sistemas propios** de una empresa.

*Por ejemplo, datos de ventas por ubicación de la tienda.*

**Datos externos:** Datos que viven **fuera** de una empresa u organización.

*Por ejemplo, salarios medios nacionales para los distintos puestos de su organización.*

## 02 - Formatos de datos

### 3 Continuos o Discretos

**Datos continuos:** Datos que se **miden** y que pueden tener casi cualquier valor numérico.

*Por ejemplo, altura de los niños en las clases de tercer grado (52,5 pulgadas; 65,7 pulgadas).*

**Datos discretos:** Datos que se **cuentan** y tienen un número limitado de valores.

*Por ejemplo, entradas vendidas en el mes en curso.*

### 4 Cualitativos o cuantitativos

**Datos cualitativos:** Medidas subjetivas y explicativas de **cualidades y características**.

*Por ejemplo, actividad física que más se disfruta.*

**Datos cuantitativos:** Medidas específicas y objetivas de **hechos numéricos**.

*Por ejemplo, población de elefantes en África.*

## 02 - Formatos de datos

### 5 Nominales u Ordinales

**Datos nominales:** Un tipo de datos **cualitativos** que no se categoriza con un **orden establecido**.

*Por ejemplo, cliente nuevo, cliente que vuelve, cliente habitual.*

**Datos ordinales:** Un tipo de datos **cualitativos** con un **orden o escala preestablecidos**.

*Por ejemplo, calificaciones de películas (cantidad de estrellas: 1 estrella, 2 estrellas, 3 estrellas).*

### 6 Estructurados o No estructurados

**Datos estructurados:** Datos **organizados de cierta forma**, por ejemplo, en filas y columnas.

*Por ejemplo, inventario de tiendas.*

**Datos no estructurados:** Datos que **no se organizan de forma** que sea **fácil identificarlos**.

*Por ejemplo, mensajes de redes sociales.*

## 03 - Tipos de datos

El **tipo de datos** es una **clase específica de atributo de datos** que nos dice **qué tipo de valor tiene ese dato**.

En otras palabras, un tipo de datos te indica con qué clase de datos estás trabajando



### Números

Contiene cualquier número entero o decimal  
Con estos podemos hacer cálculos

### Textos

Es una **secuencia de caracteres y puntuación**  
que contiene información textual

### Booleanos

Es un tipo de datos con **solo dos valores posibles**, verdadero (TRUE) o falso (FALSE).

## 03 - Tipos de datos

|    | A                   | B      | C           | D      | E           | F         | G           | H                    | I |
|----|---------------------|--------|-------------|--------|-------------|-----------|-------------|----------------------|---|
| 1  | Inicio de la semana | Helado | Mayor a 10? | Helado | Mayor a 10? | Caramelos | Mayor a 10? | Elección mas popular |   |
| 2  | 28/7/2019           | 14     | VERDADERO   | 5      | FALSO       | 19        | VERDADERO   | Caramelos            |   |
| 3  | 4/8/2019            | 5      | FALSO       | 8      | FALSO       | 9         | FALSO       | Caramelos            |   |
| 4  | 11/8/2019           | 8      | FALSO       | 31     | VERDADERO   | 7         | FALSO       | Helado               |   |
| 5  | 18/8/2019           | 25     | VERDADERO   | 26     | VERDADERO   | 3         | FALSO       | Helado               |   |
| 6  | 25/8/2019           | 2      | FALSO       | 19     | VERDADERO   | 11        | VERDADERO   | Helado               |   |
| 7  |                     |        |             |        |             |           |             |                      |   |
| 8  |                     |        |             |        |             |           |             |                      |   |
| 9  |                     |        |             |        |             |           |             |                      |   |
| 10 |                     |        |             |        |             |           |             |                      |   |
| 11 |                     |        |             |        |             |           |             |                      |   |
| 12 |                     |        |             |        |             |           |             |                      |   |
| 13 |                     |        |             |        |             |           |             |                      |   |
| 14 |                     |        |             |        |             |           |             |                      |   |
| 15 |                     |        |             |        |             |           |             |                      |   |

Dato numerico.  
Observar que se ubica a la derecha excepto le cambiemos la alineación.

Dato Booleano.  
Para calcularlo hicimos la fórmula:  
`=SI(B2>10;"VERDADERO";"FALSO")`

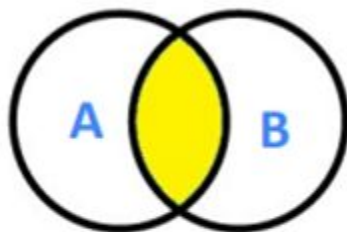
Dato texto.  
Para calcularlo hicimos la fórmula:  
`=SI(MAX(B2;D2;F2)=B2;$B$1;SI(MAX(B2;D2;F2)=D2;$D$1;$F$1))`



## 03 - Tipos de datos - Entendiendo la lógica booleana

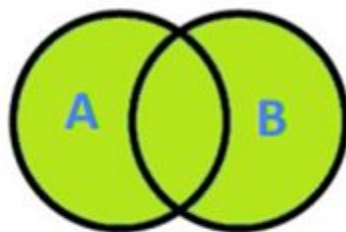
Los tipos de datos **booleanos** utilizan **operadores** que nos permiten crear **múltiples expresiones lógicas** que filtran los resultados

### AND



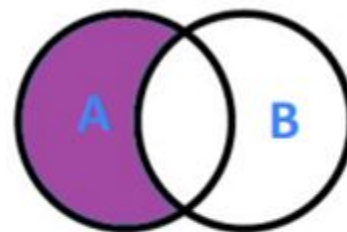
**True:** Se cumple A y B

### OR



**True:** Se cumple A, B, A y B

### NOT



**True:** Se cumple A pero no B

## 04 - Factores a considerar cuando se recolectan datos

Algunas **consideraciones** sobre la **recopilación** de datos que hay que tener en cuenta para el análisis:

### Cómo se recogerán los datos

¿Datos propios o comprados?

### Fuentes de datos

**Primaria:** datos recogidos por uno mismo

**Secundaria:** recogidos directamente por otros grupos y vendidos

**Terciaria:** vendidos por un proveedor que no los ha recogido por sí mismo

### Resolver el problema empresarial

Elegir datos que realmente puedan ayudar a resolver tu problema

### Cuántos datos hay que recoger

Toma decisiones razonables sobre el tamaño de la muestra

### Marco temporal

Decide cuánto tiempo necesitarás para recopilarnos. (Nuevos datos / datos históricos)

## 05 - Transformación de datos

La **transformación de datos** es el **proceso** de **cambiar el formato**, la **estructura** o los **valores** de los datos

- Añadir, copiar o replicar datos
- Eliminar campos o registros
- Estandarizar los nombres de las variables
- Cambiar el nombre, mover o combinar columnas en una base de datos
- Unir un conjunto de datos con otro
- Guardar un archivo en un formato diferente. Ejemplo: Excel a CSV

## 05 - Transformación de datos - Objetivos

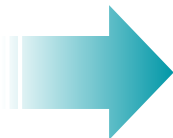


## 05 - Transformación de datos - Ejemplo organización de datos

Para facilitar la creación de gráficos, puede ser necesario transformar los datos de **formato largo** en datos en **formato ancho**.

Ejemplo de transformación de los precios de las acciones:

| Empresa | FECHA      | CIERRE |
|---------|------------|--------|
| EDENOR  | 09/29/2021 | 62,6   |
| EDENOR  | 09/28/2021 | 62,45  |
| ALUAR   | 09/29/2021 | 63,6   |
| ALUAR   | 09/28/2021 | 62,7   |
| BBVA    | 09/29/2021 | 244    |
| BBVA    | 09/28/2021 | 242,5  |



| EMPRESA | 09/28/2021 | 09/29/2021 |
|---------|------------|------------|
| ALUAR   | 62,7       | 63,6       |
| BBVA    | 242,45     | 244        |
| EDENOR  | 62,45      | 62,6       |

**Formato largo** son datos en los que **cada fila** contiene un **único punto de datos** para un elemento concreto

**Formato ancho** son en los que **cada fila** contiene **múltiples puntos de datos** para los elementos particulares identificados en las **columnas**

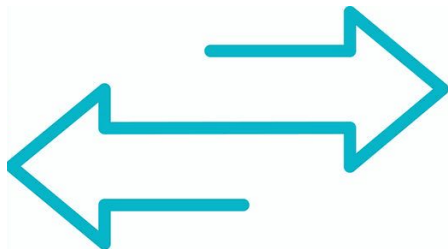
## 05 - Transformación de datos - Ejemplo organización de datos

Cada uno de estos formatos se prefieren para determinados **objetivos**

### Formato largo

Crear tablas y gráficos con pocas variables sobre cada tema.

Comparar gráficos lineales sencillos



### Formato ancho

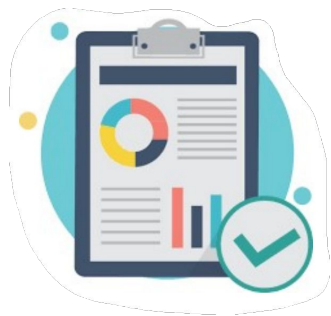
Almacenar muchas variables sobre cada tema.

*Por ejemplo, 60 años de tasas de interés para cada banco*

Realizar análisis estadísticos avanzados o gráficos

## 06 - Como informa los datos (Informes y Dashboard)

Existen todo tipo de herramientas para ayudarte a visualizar y compartir tus análisis de datos con los interesados. Dos de estas herramientas son:



**Informes:** es una **recopilación estática** de datos que se entrega **periódicamente** a los interesados



**Paneles (dashboard):** es una **recopilación dinámica** que monitorea los datos entrantes en **vivo**

## 06 - Como informa los datos - Informes ventajas y desventajas

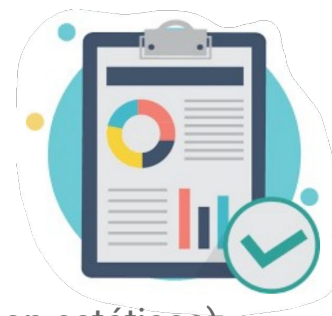


Se pueden diseñar y enviar periódicamente

Información organizada y de fácil referencia

Son rápidos de diseñar y fáciles de usar

Reflejan datos que ya se han limpiado y ordenado (porque son estáticos)



Necesitan un mantenimiento regular

No son visualmente atractivos

No muestran datos en vivo ni en evolución



## 06 - Como informa los datos - Dashboard ventajas y desventajas



Más acceso a la información registrada

Se puede interactuar a través de los datos con los filtros

Se puede acceder continuamente a la información

Son agradables a la vista



Requiere mucho tiempo para diseñarlos

Pueden ser menos eficientes que los informes si no se usan con mucha frecuencia

Necesita mucho mantenimiento si se rompe una tabla

Pueden abrumar a las personas con información

Si no sueles buscar datos en un panel, es posible que te pierdas en él

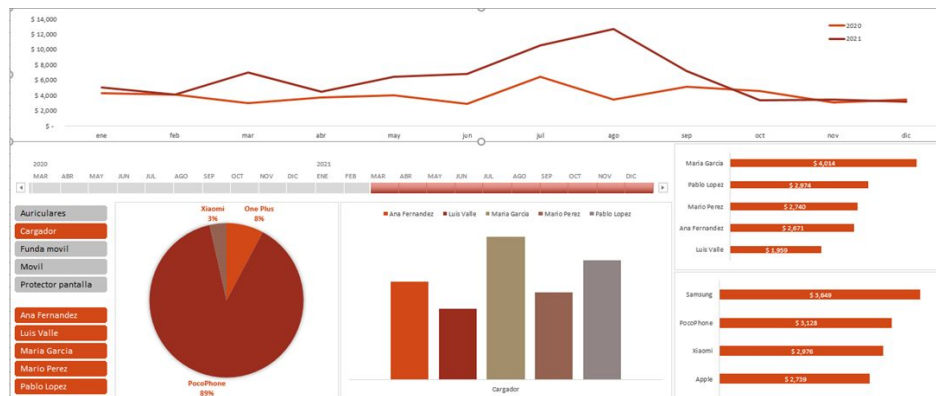
## 06 - Como informa los datos - Dashboard en Hojas de cálculo

**Tabla dinámica:** Es una **herramienta** de resumen de datos utilizada en el **procesamiento** de datos.

| FACTURA | FECHA    | COMERCIAL     | MARCA    | DESCRIPCION        | Can |
|---------|----------|---------------|----------|--------------------|-----|
| 1001    | 1/1/2020 | Maria Garcia  | Samsung  | Auriculares        |     |
| 1002    | 2/1/2020 | Luis Valle    | Samsung  | Movil              |     |
| 1003    | 3/1/2020 | Pablo Lopez   | One Plus | Funda movil        |     |
| 1004    | 3/1/2020 | Pablo Lopez   | One Plus | Movil              |     |
| 1005    | 4/1/2020 | Pablo Lopez   | Samsung  | Cargador           |     |
| 1006    | 4/1/2020 | Pablo Lopez   | Xiaomi   | Funda movil        |     |
| 1007    | 4/1/2020 | Luis Valle    | One Plus | Movil              |     |
| 1008    | 6/1/2020 | Ana Fernandez | One Plus | Protector pantalla |     |
| 1009    | 6/1/2020 | Luis Valle    | Xiaomi   | Funda movil        |     |

|    | A                        | B                    |
|----|--------------------------|----------------------|
| 1  |                          |                      |
| 2  |                          |                      |
| 3  | <b>Etiquetas de fila</b> | <b>Suma de Venta</b> |
| 4  | Ana Fernandez            | 44589.20513          |
| 5  | Luis Valle               | 44136.15385          |
| 6  | Maria Garcia             | 57373.87179          |
| 7  | Mario Perez              | 49415.79487          |
| 8  | Pablo Lopez              | 48216.74359          |
| 9  | <b>Total general</b>     | <b>243731.7692</b>   |
| 10 |                          |                      |

Las tablas dinámicas se usan para **resumir, ordenar, reorganizar, agrupar, contabilizar, totalizar** o **promediar** los datos almacenados en una base de datos

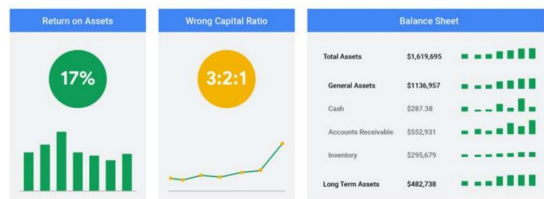


## 06 - Como informa los datos - Tipos de Dashboard

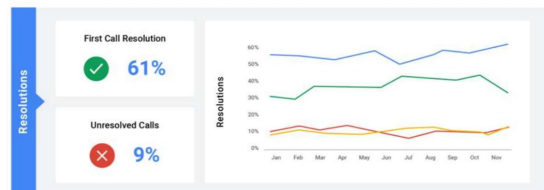
Revenue and Customer Overview - Q1



Financial Performance Dashboard



Customer Service Team Dashboard



**Estratégicos:** Se centra en **objetivos y estrategias a largo plazo** al más alto nivel de métricas.

Contienen información útil para la toma de decisiones en toda la empresa.

**Operativos:** Presenta un **seguimiento del desempeño a corto plazo y objetivos intermedios**.

Pueden proporcionar información sobre el desempeño casi en tiempo real.

Permiten realizar un seguimiento y mantener sus procesos operativos.

**Analíticos:** Contienen los detalles relacionados con el **uso**, el **análisis** y las **predicciones** realizadas por los científicos de datos.

Rara vez se comparten con la alta dirección. Son los más técnicos

## 06 - Como informa los datos - ¿Cómo los Dashboard ayudan al AD?



Para los analistas

Para los interesados



Se comparte **una sola** cosa a los interesados

**Centralización**

Trabajar con una **visión integral de datos**, iniciativas, objetivos, proyectos, procesos y más

Mostrar y actualizar datos entrantes en vivo en **tiempo real**

**Visualización**

Detectar **patrones y tendencias** cambiantes más rápidamente

Obtener **información relevante** de diferentes conjuntos de datos

**Percepción**

Comprender la **historia de los datos**

Crear **vistas personalizadas** a una persona específica, un proyecto o una presentación

**Personalización**

**Profundizar** en áreas más específicas de preocupaciones o intereses especializados

## 06 - Como informa los datos - Pasos para diseñar un Dashboard

1

### Identificar a los interesados que necesitan ver los datos y cómo los usarán

Hacer preguntas efectivas

2

### Diseñar el panel

Usa un **encabezado** claro para etiquetar la información

Agrega **descripciones** de texto cortas a cada visualización

Muestra la **información más importante** en la parte **superior**

3

### Crea prototipos si lo deseas

Esbozar sus paneles antes de crearlos

4

### Selecciona las visualizaciones que utilizarás en el panel

Aquí tienes muchas opciones y todo depende de la historia de datos que cuentes

5

### Crea filtros según sea necesario

Esto puede ser de gran ayuda para identificar patrones mientras se mantienen intactos los datos originales.

## 07 - Datos y métricas

Una **métrica** es un **tipo único** y **cuantificable** de datos que pueden utilizarse **para la medición**

Las métricas suelen implicar **matemáticas simples**.

*Por ejemplo:  $VENTAS = PRECIO\ UNITARIO * PRODUCTOS\ VENDIDOS$*



Intentan alcanzar un **objetivo específico** midiendo datos.

Este **objetivo métrico** es un conjunto de **objetivos medibles** establecido por una empresa y evaluado mediante métricas.

*Por ejemplo: alcanzar una cierta cantidad de ventas mensuales o quizás un cierto porcentaje de clientes habituales*

## 08 - Pensamiento matemático y Big y Small Data

El **pensamiento matemático** es una habilidad poderosa que puedes usar para ayudarte a **resolver problemas** y **crear nuevas soluciones**.



Significa mirar un problema y **desglosarlo lógicamente** paso a paso, para que puedas ver la relación de los **patrones** en tus datos y usar eso para **analizar tu problema**



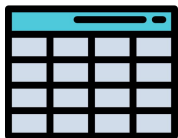
Puede ayudarte a encontrar las mejores **herramientas** para el análisis

Uno de los factores que influyen en la elección de las herramientas es el

**tamaño de tu conjunto de datos**



## 08 - Pensamiento matemático y Big y Small Data



### SMALL DATA Microdatos

Conjunto de datos **específicos** que cubren un periodo de **tiempo limitado**

Organizado y analizado en **hojas de cálculo**

Gestión **simple**

Para negocios de **mediano tamaño**

**Tamaño manejable** para el análisis

### BIG DATA Macrodatos



**Grandes bases de datos** con información no específica y un **amplio periodo de tiempo**

Acceso con **consultas**

**Mucho esfuerzo** para recolectar, gestionar, ordenar y visualizar la información

Utilizada por **grandes empresas**

Es necesario tomar solo una parte de los datos para el análisis



## 08 - Pensamiento matemático y Big y Small Data



### BENEFICIOS

Grandes volúmenes de datos ayudan a las empresas a **identificar formas más eficientes de hacer negocios**

Facilita la **detección** de **tendencias**, **condiciones** y **patrones** de mercado

### BIG DATA Macrodatos

### DESAFÍOS

**Sobrecarga** de datos innecesarios

Los datos importantes pueden quedar “perdidos” entre la gran cantidad de **datos irrelevantes**

No siempre los datos necesarios son de fácil acceso

Las tecnologías actuales todavía están **en desarrollo**

## 08 - Pensamiento matemático y Big y Small Data



¿PREGUNTAS?





**¡Muchas gracias!**