

Esta clase va a ser  
● grabada

## Módulo 3

# Realizar cálculos de datos

# Temario



## M2

### Agregación de datos para analizar

- Concepto de agregación
- BUSCARV en Excel
- JOIN en SQL

## M3

### Realizar cálculos de datos

- Validación de datos
- Funciones condicionales
- Funciones escalares
- Tablas temporales

## M4

### Principios de diseño esenciales

- Criterios de diseño
- Elección de visualizaciones
- Audiencia

# **Validación de datos**

# Validación de datos

Es importante validar la integridad de nuestros datos para evitar:

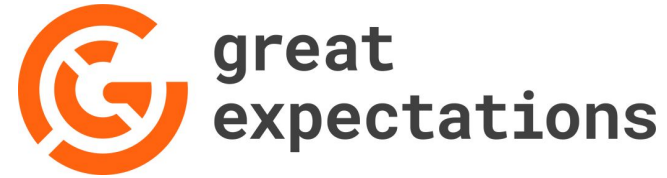
- Toma de decisiones incorrectas
- Entrenamiento de modelos con información errónea.
- Fallas en productos basados en datos



En los modelos de machine learning se testea tanto el **input** como el **output**

# Validación de datos

- Tipos de datos
- Rango de datos
- Limitaciones de datos
- Coherencia de los datos
- Estructura de los datos



# Funciones condicionales

# Funciones condicionales

Realizan una operación al cumplirse una condición predefinida

- CONTAR.SI()
- SUMAR.SI()
- CONTAR.SI.CONJUNTO()
- SI()

¿Cómo armaríamos estas mismas funciones en SQL?



# Funciones escalares

# Funciones escalares

Trabajan con **operadores aritméticos**. El resultado devuelve la misma cantidad de filas (no las agrega):

- Suma

- Resta

- Multiplicación

- División

Column A

A1

A2

A3

Column B

B1

B2

B3

Column X

Sum of A1 + B1

Sum of A2 + B2

Sum of A3 + B3

¿Cómo armaríamos estas mismas funciones en SQL?

# Tablas temporales

# Common table expressions (CTEs)

Son resultados de una consulta guardados con un nombre de manera temporal:

- El resultado se guarda durante el tiempo de ejecución de la query
- Sirven para **modularizar** el código y hacerlo más **mantenible**
- Nos ayudan a resolver **consultas más complejas** (usar el resultado de una consulta en otra consulta).

Las CTEs son especialmente útiles cuando queremos dividir el código en **bloques lógicos**

# Common table expressions (CTEs)

```
EJECUTAR GUARDAR COMPARTIR PROGRAMACIÓN
```

```
1 WITH trips_over_1_hr AS (  
2   SELECT *  
3   FROM `bigquery-public-data.new_york_citibike.citibike_trips`  
4   WHERE tripduration >= 60  
5 )  
6 SELECT COUNT(*) AS count  
7 FROM trips_over_1_hr
```

## Resultados de la consulta

INFORMACIÓN DEL TRABAJO		RESULTADOS	JSON	DETALLES DE LA
Fila	count			
1	53108721			

# SELECT INTO

Permite crear **tablas temporales** desde una consulta. La diferencia con la CTE es que permanecen luego de que termina la consulta:

- La tabla generalmente se borra cuando se cierra la sesión de la base de datos.
- Si queremos que la tabla se mantenga deberíamos crear una tabla nueva.

```
SELECT
*
INTO
AfricaSales
FROM
GlobalSales
WHERE
Region = "Africa"
```

# Demo guiada

**¿Preguntas?**



# Resumen de la clase de hoy

- Aprendimos la importancia de validar los datos de un modelo y los criterios generales.
- Revisamos la aplicación de las funciones escalares.
- Aprendimos a usar tablas temporales y CTEs para modularizar consultas



**Recuerden realizar la  
práctica de la clase**

**¡Gracias!**