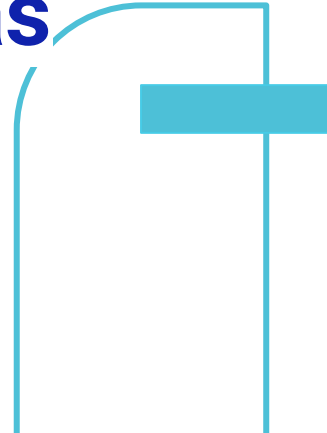


Módulo 2

El mundo de los datos.

Configura tu caja de herramientas



AGENDA DE LA CLASE

- ✓ Hojas de cálculo
 - ✓ Ordenar y filtrar
 - ✓ Fórmulas y funciones
- ✓ Lenguaje de consulta estructurado (SQL)
 - ✓ Sintaxis básica
 - ✓ Mayúsculas, sangría, punto y coma
 - ✓ Condición WHERE
 - ✓ Elegir todas las columnas
 - ✓ Comentarios
 - ✓ Alias
 - ✓ Prácticas recomendadas en SQL
- ✓ Visualización de datos
 - ✓ Herramientas de viz
- ✓ Selección de las herramientas



01 - Hojas de cálculo

La **hoja de cálculo** organiza y almacena la información en un área de trabajo en forma de matriz.



Columnas

Filas

Celdas

	A	B	C	D
1			Esta	
2	Esta es una Celda		es	
3	Esta es una Celda	es	una	fila.
4			columna	
5				

Observación o **registro** es una **fila** de un conjunto de datos e incluye todos los atributos de lo que contenga esa fila.

Encabezado es la primera **fila**, donde se presentan las **etiquetas** de los atributos o características de cada observación.

	A	B	C
1	Nombre	Apellido	Edad
2	Martin	Perez	26
3	Carolina	Gonzales	50
4	Analia	Martinez	32
5			

01 - Hojas de cálculo

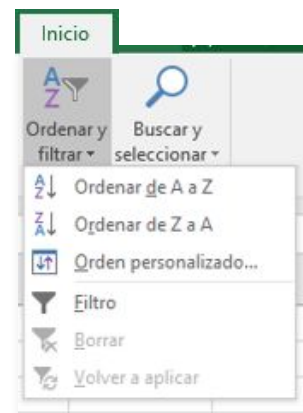
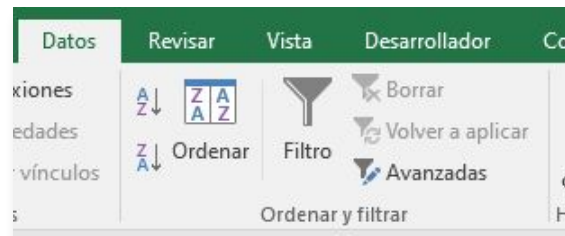
Cuatro operaciones fundamentales

1 Ordenar

- Ordenar de A a Z o Ascendente
- Ordenar de Z a A o Descendente
- Orden personalizado

2 Filtrar

- Para seleccionar los **grupos** de información específicos
- La información que ya no se ve solo está **oculta** y no se borra ni se pierde



	A	B	C
1	Nombre	Apellido	Edad
3	Carolina	Gonzales	50
4	Analía	Martinez	32

01 - Hojas de cálculo

3 Fórmula

Es un **conjunto de instrucciones** que realiza un **cálculo específico** con los datos de una hoja de cálculo

- Operaciones básicas (+, -, *, /)
- Operaciones combinadas

	A	B	C	D	E
1	Nombre	Apellido	Edad		
2	Martin	Perez	26		Edad promedio
3	Carolina	Gonzales	50		$=(C2+C3+C4)/3$
4	Analia	Martinez	32		
5					

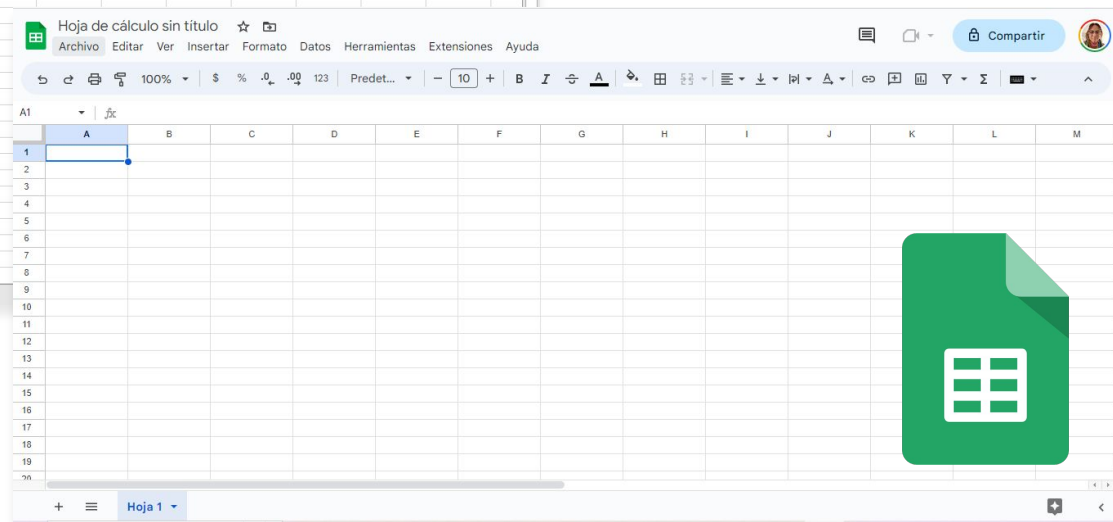
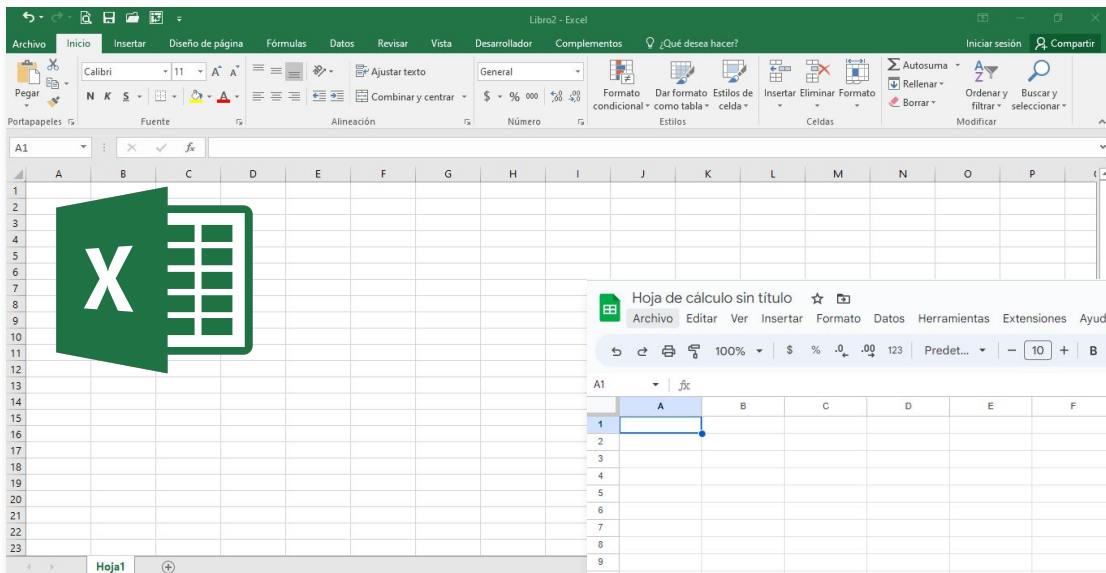
4 Función

Es un **comando preestablecido** que realiza **automáticamente** un **proceso o tarea específicos** con los datos de una hoja de cálculo.

- Operaciones más eficientes
- Operaciones complejas

	A	B	C	D	E	F
1	Nombre	Apellido	Edad			
2	Martin	Perez	26		Edad promedio	
3	Carolina	Gonzales	50		36	
4	Analia	Martinez	32		$=promedio(C2:C4)$	
5						

01 - Hojas de cálculo



02 - Lenguaje de Consulta Estructurado

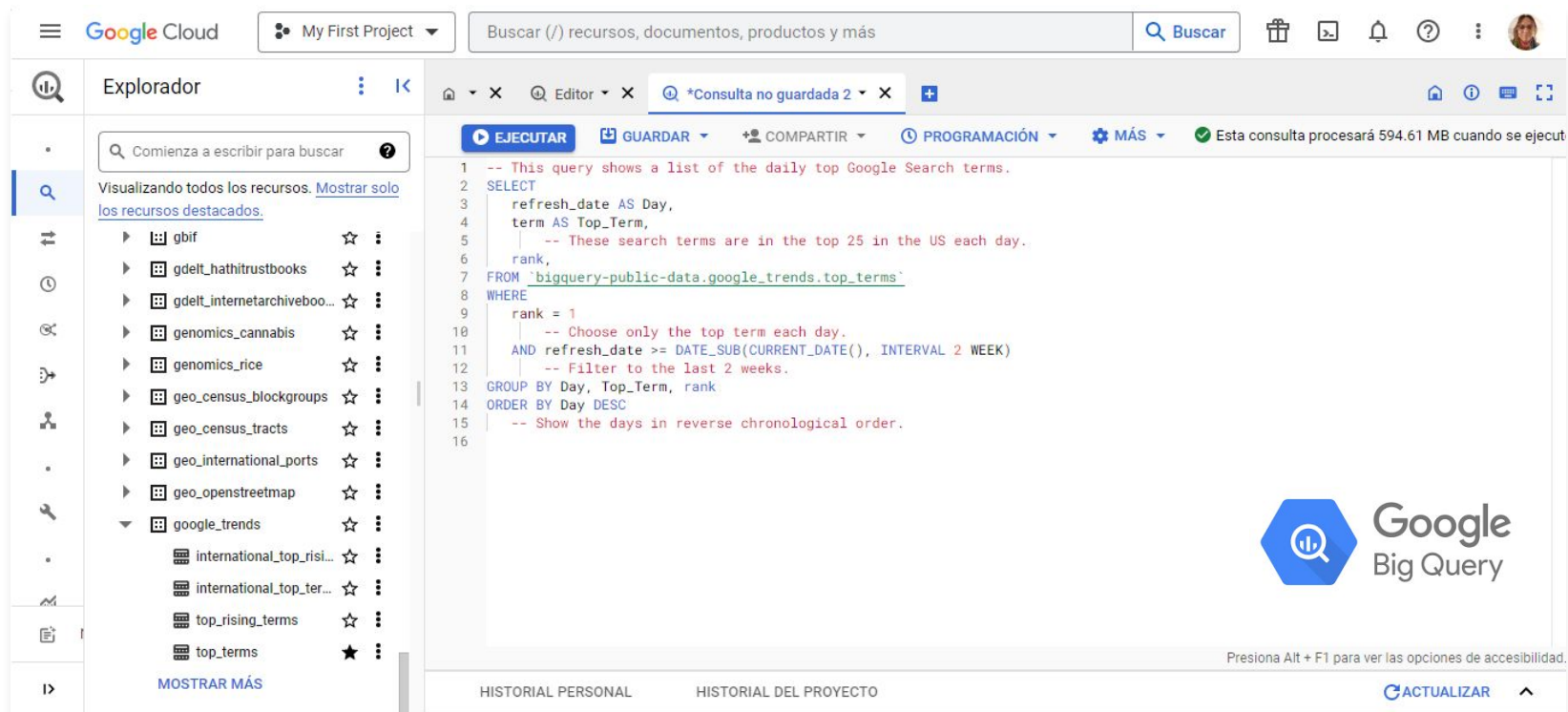
El **lenguaje de consulta estructurado** (SQL) permite a los analistas de datos **hablar** con sus **bases de datos**.

Una **consulta** es una **solicitud** de datos o información a una base de datos.

La **sintaxis** es la **estructura predeterminada** de un lenguaje, que incluye todas las **palabras**, los **símbolos** y la **puntuación** requeridos, así como su correcta **colocación**



02 - Lenguaje de Consulta Estructurado



The screenshot displays the Google Cloud BigQuery interface. At the top, the Google Cloud logo and 'My First Project' are visible. A search bar contains the text 'Buscar (/) recursos, documentos, productos y más'. Below this, the 'Explorador' (Explorer) panel on the left shows a list of datasets under the 'google_trends' project, including 'international_top_ris...', 'international_top_ter...', 'top_rising_terms', and 'top_terms'. The main editor area shows a SQL query for selecting top search terms from the 'bigquery-public-data.google_trends.top_terms' table. The query is as follows:

```
1 -- This query shows a list of the daily top Google Search terms.
2 SELECT
3   refresh_date AS Day,
4   term AS Top_Term,
5   -- These search terms are in the top 25 in the US each day.
6   rank,
7 FROM `bigquery-public-data.google_trends.top_terms`
8 WHERE
9   rank = 1
10  -- Choose only the top term each day.
11  AND refresh_date >= DATE_SUB(CURRENT_DATE(), INTERVAL 2 WEEK)
12  -- Filter to the last 2 weeks.
13 GROUP BY Day, Top_Term, rank
14 ORDER BY Day DESC
15 -- Show the days in reverse chronological order.
16
```

Below the query editor, the 'Google Big Query' logo is displayed. At the bottom, there are links for 'HISTORIAL PERSONAL' and 'HISTORIAL DEL PROYECTO', and a button labeled 'ACTUALIZAR'.

02 - Lenguaje de Consulta Estructurado

1

Sintaxis básica

- **SELECT** para elegir las columnas que deseas devolver
- **FROM** para elegir las tablas donde se encuentran las columnas que deseas
- **WHERE** para filtrar determinada información



Consejo: escribe una consulta SQL desde cero.

Es mucho mejor iniciar una consulta escribiendo las palabras clave SELECT, FROM y WHERE, una palabra por cada línea y asegúrate de agregar sangría cuando agregues nuevas líneas dentro de estos comandos.

```
EJECUTAR  GUARDAR  + COI
1 SELECT las columnas que necesitas
2 FROM la tabla donde estan los datos
3 WHERE una condición
```

```
EJECUTAR  GUARDAR  COMPARTIR  PROGRAMACIÓN
1 SELECT nombre, ciudad
2 FROM 'my-first-project-370223.customer_data.customer_address'
```

02 - Lenguaje de Consulta Estructurado

- El mismo comando **SELECT** puedes elegir varias columnas que se pueden sangrar (usar sangría) y agrupar
- Cada columna está separada por una coma

```
SELECT Columna A
      ,Columna B
      ,Columna C
FROM mi tabla de datos
WHERE condición
```

EJECUTAR GUARDAR COMPARTIR PROGRAMAR

```
1 SELECT id_cliente
2      ,nombre
3      ,ciudad
4 FROM my-first-project-370223.customer_data.customer_address
5 WHERE ciudad='Santa Clara'
```

- Si tienes varias condiciones en tu cláusula **WHERE**

EJECUTAR GUARDAR

```
1 SELECT Columna A
2      ,Columna B
3      ,Columna C
4 FROM la tabla de datos
5 WHERE Condición 1
6      AND Condición 2
7      AND Condición 3
```

EJECUTAR GUARDAR COMPARTIR PROGRAMAR

```
1 SELECT id_cliente
2      ,nombre
3      ,ciudad
4 FROM my-first-project-370223.customer_data.customer_address
5 WHERE ciudad='Santa Clara'
6      OR ciudad='Palm City'
7      OR ciudad='Bonita Springs'
```

02 - Lenguaje de Consulta Estructurado



Cómo usar

*SELECT
FROM
WHERE*

en una consulta

Una vez que hayas practicado cómo escribir tus propias consultas SQL las consultas con varios campos serán más sencillas

02 - Lenguaje de Consulta Estructurado

2

Mayúsculas, sangrías y punto y coma

- Puedes usar minúsculas, mayúsculas, ambas
- No tienes que preocuparte por los espacios adicionales entre palabras.
- Sin embargo, el uso de mayúsculas y sangría puede ayudarte a leer la información más fácilmente.
 - Mayúsculas para cláusulas y funciones
 - Minúsculas nombres de columnas
 - snake_case
 - CamelCase
- El **punto y coma** es un **terminador de instrucción** que forma parte de la norma SQL-92 del Instituto Nacional Estadounidense de Estándares (ANSI)



```
1 SELECT Campo1
2 FROM Tabla
3 WHERE Campo1='Condicion';
4
```

EJECUTAR GUARDAR

02 - Lenguaje de Consulta Estructurado

3

Condiciones WHERE

- La cláusula **WHERE** restringe la consulta para que la base de datos devuelva solo los datos con una **coincidencia** de **valor exacta**, o los datos que coincidan con una determinada **condición** que desees satisfacer.
- La cláusula **WHERE LIKE** permite decirle a la base de datos que busque un **patrón determinado**.
- Ejemplo: WHERE Campo1 LIKE 'Ch%' devolverá todo lo que en el Campo 1 empiece con Ch.

4

Elegir todas las columnas

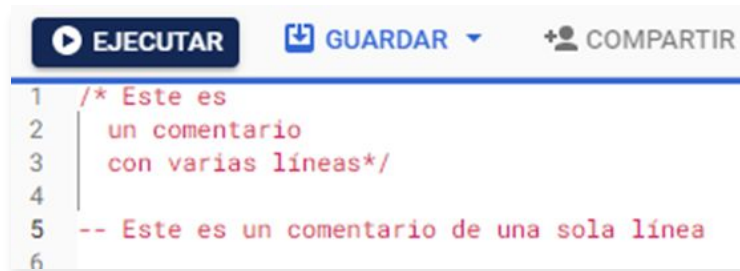
- **SELECT ***
- Usar con moderación

02 - Lenguaje de Consulta Estructurado

5

Comentarios

- Dos guiones (--) para una línea
- /* y */ para más de una línea
- Dentro o fuera de una instrucción



The screenshot shows a SQL editor interface with three buttons at the top: 'EJECUTAR' (Execute), 'GUARDAR' (Save), and 'COMPARTIR' (Share). Below the buttons, a code editor displays two types of SQL comments. The first is a multi-line comment starting with '/*' on line 1 and ending with '*/' on line 3, with the text 'Este es un comentario con varias líneas' in between. The second is a single-line comment on line 5 starting with '--' and containing the text 'Este es un comentario de una sola línea'. Line numbers 1 through 6 are visible on the left side of the code editor.

```
1  /* Este es
2     un comentario
3     con varias líneas*/
4
5  -- Este es un comentario de una sola línea
6
```

6

Alias

- **AS**
- Es una manera de **renombrar** una columna o tabla
- Son **válidos** sólo para la duración de la consulta.
- Un alias no cambia el nombre real de una columna o tabla de la base de datos.

02 - Lenguaje de Consulta Estructurado

7

Comillas

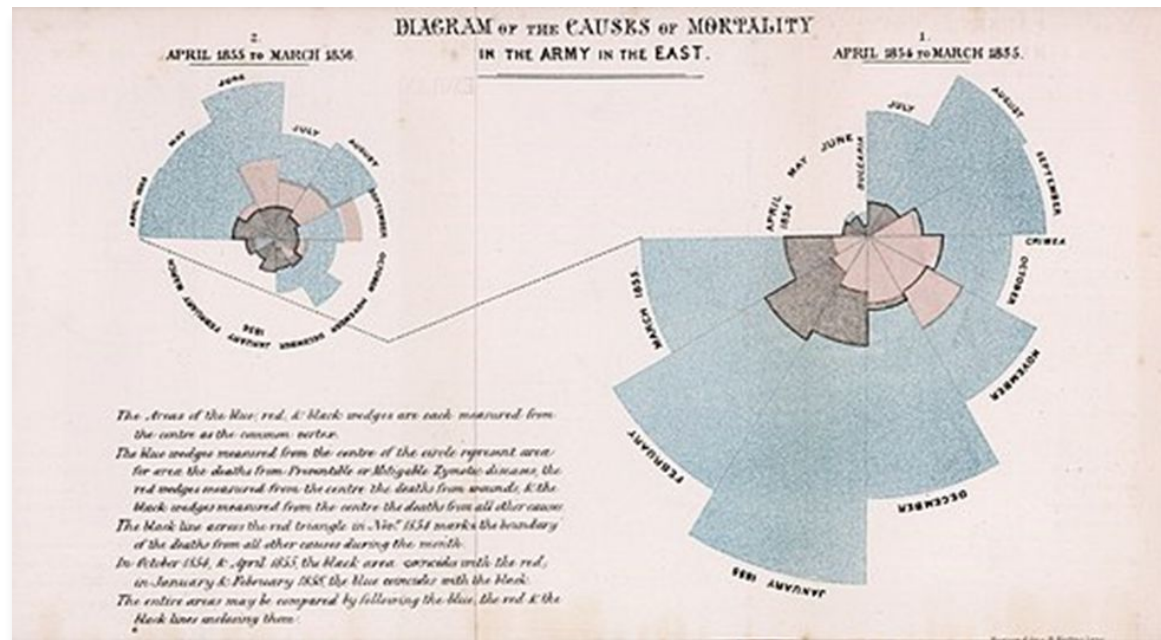
- Simples: `' '`
- Dobles: `" "`
- Es para **cadenas de texto**
- Pero hay dos situaciones en las que sí importa qué tipo de comillas utilices:
 - Cuando quieres que las cadenas sean identificables en cualquier dialecto SQL:
`WHERE country_code = 'US'`
 - Cuando tu cadena contiene un apóstrofo o comillas:
`WHERE Favorite_food = "Shepherd's pie"`

03 - Visualización de datos

La **visualización** de datos es la **representación gráfica** de cierta información.

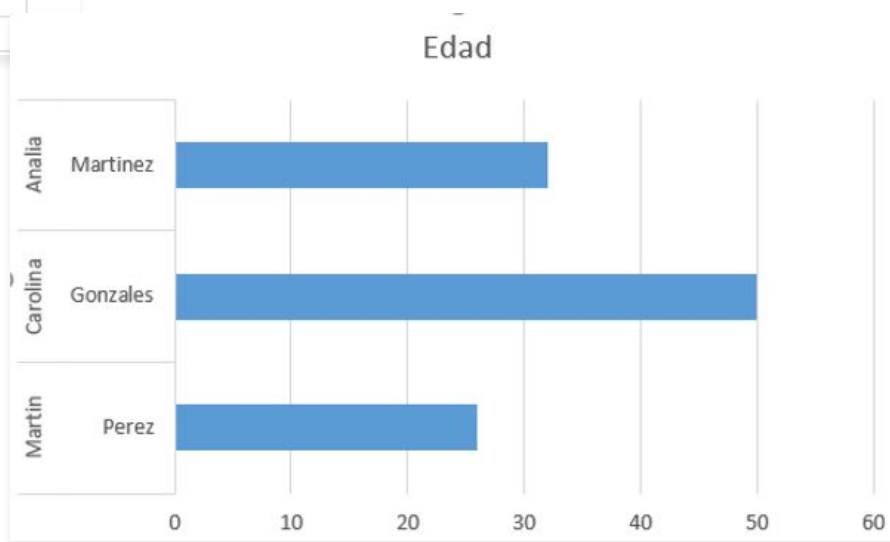
Florence Nightingale

para que los datos
fueran más **fáciles** de
entender para su
audiencia



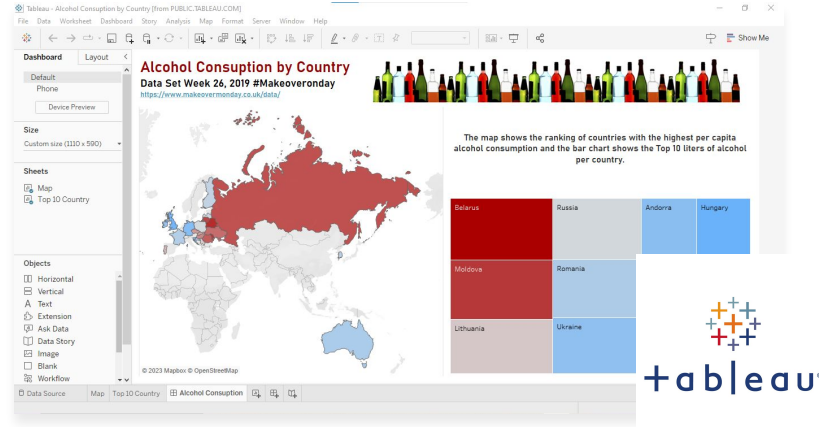
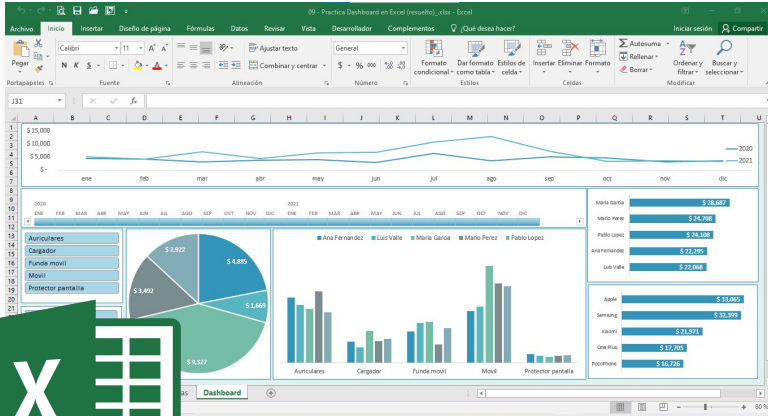
03 - Visualización de datos

	A	B	C	D	E
1	Nombre	Apellido	Edad	Color favorito	Postre Favorito
2	Martin	Perez	26	Azul	Helado
3	Carolina	Gonzales	50	Azul	Flan
4	Analia	Martinez	32	Verde	Torta



03 - Visualización de datos

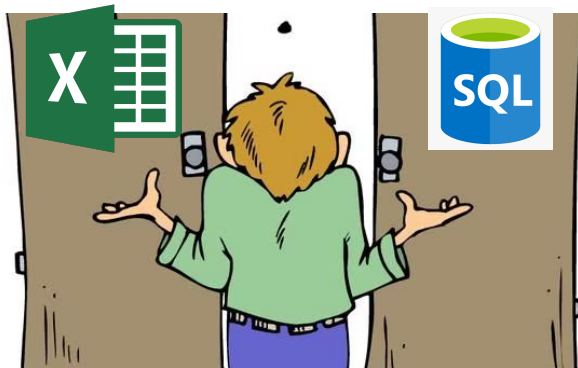
Herramientas de Visualización



04 - Selección de la herramienta

HOJAS DE CÁLCULO

- **Programas**
- La información se guarda en forma **local**
- **Límite** máximo de 1 millón de registros o filas.
- Los datos son ingresados **manualmente**.
- Los datos están **estructurados** en filas y columnas.
- Preferiblemente la debe usar un usuario a la vez para no generar **inconsistencias**.
- La base de datos es controlada por el usuario.
- Se pueden hacer **visualizaciones**.



SQL

- Es un **lenguaje**.
- **Interactúa** con bases de datos normalmente extensas.
- Bases de datos extensas no reduce la **velocidad** de consulta.
- Se ingresan a los datos a través de **Queries** (consultas, peticiones).
- Organiza la información con **relaciones entre tablas**.
- Pueden usarlos **varios usuarios simultáneamente**.
- La base de datos está controlada por un **sistema de gestión de bases de datos**.

¿PREGUNTAS?





¡Muchas gracias!