

# Limpieza de datos

# Limpieza de datos

Los datos no siempre se encuentran en un formato adecuado para el análisis. Por eso debemos realizar transformaciones que impacten dentro de la herramienta sin modificar el origen de datos.

Cada uno de los cambios que realicemos será en pos de aumentar la calidad de los datos.

# Limpieza de datos

En Tableau hay dos puntos donde podemos hacer limpieza de datos:

- Al inicio una vez que se conecta con el origen de datos.
- Luego, una vez que se tiene un mayor conocimiento de la fuente de datos y se detectan transformaciones necesarias para ciertas visualizaciones. En esos casos, no hace falta volver al inicio, se puede realizar desde la visualización.

# Dimensiones y medidas

**Limpieza:** Detectar las columnas que son dimensiones y su formato (Cadena, fecha, geográfico, índice, etc.) y medidas (números enteros y decimal).

¿Hay que crear alguna nueva métrica o dimensión?

# Filtros

**¿Queremos leer el 100% de la información?** Cuando el origen de datos es un Excel o un archivo local esta pregunta no es tan importante.

Pero si Tableau está leyendo una base de datos ¿vamos a analizar toda la información disponible? ¿O podemos hacer filtros y dejar solamente lo relevante?

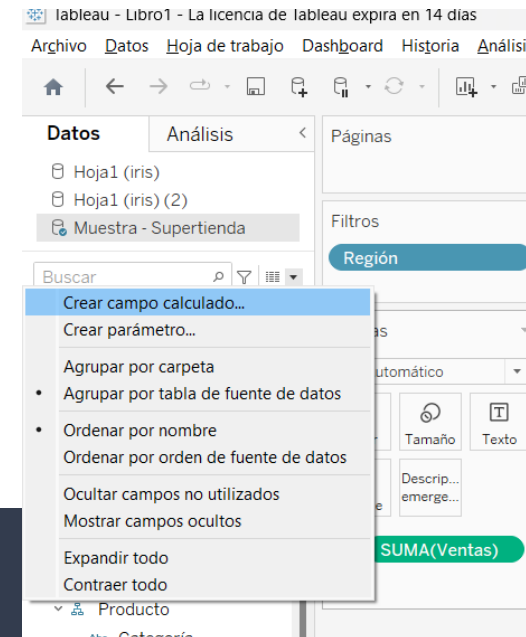
Esto impactará en la performance de la herramienta.

# Campos calculados

Tableau nos permite crear nuevas columnas que surgen a partir de la aplicación de funciones.

Las funciones se dividen por el tipo de dato que toman para el cálculo:

- Cadena
- Fecha
- Número
- Conversión de tipo
- Lógica



# Campos calculados

	A	B	C	D	E	F
1						
2		Simple Profit and Loss Statement (P&L)				
3		(\$ in millions)				2021A
4						
5		Revenue				\$100
6		Less: COGS				(40)
7		Gross Profit				\$60
8		Less: SG&A				(20)
9		EBIT				\$40
10		Less: Interest Expense				(5)
11		EBT				\$35
12		Less: Taxes	% Rate	30.0%		(11)
13		Net Income				\$25
14						

# Campos calculados

KPI VS BPI



# Divisiones

Las divisiones permiten separar los datos de una columna en varias columnas nuevas.

Esto se puede realizar con la división o división personalizada. La diferencia entre ambas es que en la primera, Tableau interpreta el delimitador en cambio en la personalizada se puede configurar qué carácter deseamos que sea el delimitador de las columnas.

# Renombrar columnas y Alias

Tableau nos permite renombrar las columnas. Esto no modifica el origen de datos. Solo genera un nuevo nombre visible para Tableau lo cual resulta una opción útil cuando queremos mapear columnas con nombres codificados a nombres interpretables por el negocio

# Importante

Cuando la fuente de datos se actualice y crezca el volumen de datos, para todos los nuevos valores, Tableau va a realizar la limpieza de datos realizada para los anteriores valores.