

~~HENRY~~



# Power Query

Data Science





# Agenda



- Power Query
- Herramientas de Power Query



# **OBJETIVOS DE LA CLASE**

***Al finalizar esta lecture estarás en la capacidad de...***

- Aplicar las principales herramientas de Power Query



Al **finalizar** cada uno de los temas,  
tendremos un **espacio de consultas**.



Hay un **mentor** asignado para  
responder el **Q&A**.

¡Pregunta, pregunta, pregunta! :D



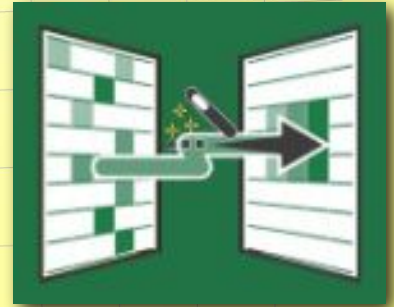
# Power Query





# ¿Qué es?

- Es una herramienta ETL.
- Permite agregar o modificar consultas, administrar consultas agrupando o agregando descripciones a los pasos o visualizar las consultas y su estructura con diferentes vistas.





# Transformaciones

- Remover filas o columnas
- Cambiar el tipo de dato
- Renombrar columnas
- Crear nuevas consultas (Merge & Append)
- Identificar y corregir errores



# Limpieza de datos







# Limpieza

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**.

Las más comunes son:

- Dimensiones y medidas
- Filtros
- Campos calculados
- Divisiones
- Renombrar columnas
- Alias



# Dimensiones y medidas

Es importante detectar las columnas que contienen datos descriptivos (dimensiones) y qué columnas son medidas.

- **Dimensiones**: cadena, fecha, índice, etc.
- **Medidas**: número entero, decimal.



# Filtros

¿Queremos el 100% de la información?

Power BI lee una base de datos, entonces... ¿vamos a analizar toda la información disponible? Podemos aplicar filtros y dejar solamente la relevante.

Esto impactará en la performance de la herramienta.



# Campos calculados

Se pueden crear nuevas columnas que surgen a partir de la aplicación de funciones.

Las funciones dependen del tipo de dato que toman para el cálculo, como ser: cadena, fecha, número, lógica, etc.

Ejemplo: si tengo **precio unitario** y **cantidad**, puedo generar campo **total**...

$$\text{Total} = \text{Precio unitario} * \text{Cantidad}$$



# Divisiones

Permiten separar los datos de una columna en varias columnas nuevas.

Esto se puede realizar con:

- **División**: la herramienta interpreta el delimitador.
- **División personalizada**: se puede configurar el carácter delimitador de columnas.



# Renombrar columnas

Power BI permite renombrar columnas.

Esto no modifica el origen de datos, sólo genera un nombre visible para Power BI lo que resulta útil cuando queremos mapear columnas con nombres codificados a nombres interpretables por el negocio.



# Alias

Podemos cambiar determinados valores de los datos, esto nos permite mapear valores incorrectos a una opción prolija y limpia.

Esto tampoco modifica el origen de los datos.

Ejemplo: Reemplazar cada una de las ocurrencias de "Bs As" por "Buenos Aires".



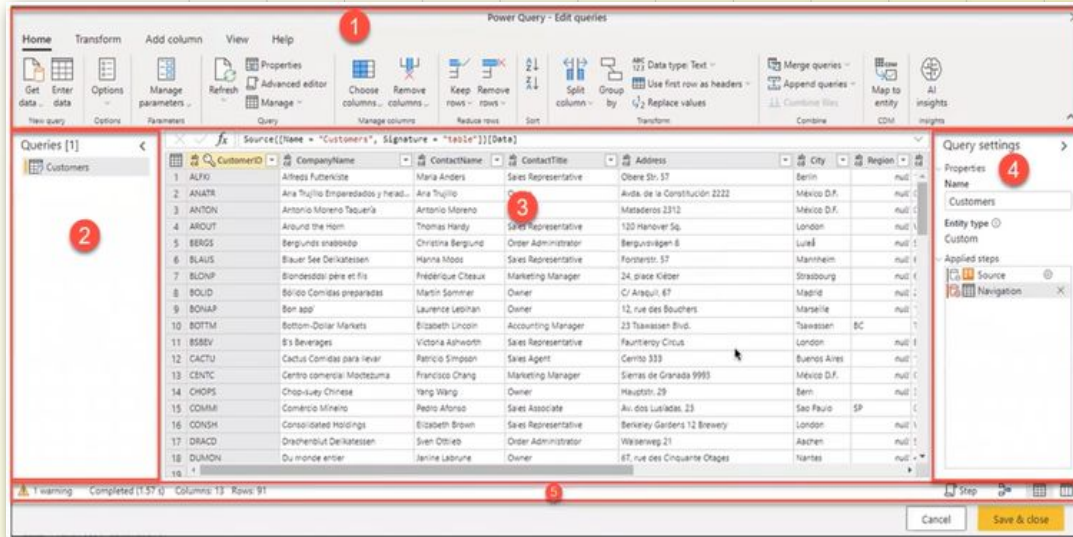


# Interfaz del editor





# Editor



- 1) **cinta de opciones** (proporciona varias pestañas para agregar transformaciones, seleccionar opciones para su consulta y acceder a botones de la cinta para completar tareas).
- 2) **Panel de consultas** (una vista de todas las consultas disponibles).
- 3) **Vista actual** (vista de trabajo principal).
- 4) **Configuración de consulta** (vista de la consulta seleccionada actualmente con información relevante, como el nombre de la consulta, los pasos y varios indicadores).
- 5) **Barra de estado** (muestra información relevante sobre la consulta, como el tiempo de ejecución, total de columnas y filas, y estado de procesamiento).



# Cinta de opciones



Es el componente donde se encuentran la mayoría de las transformaciones y acciones que se pueden realizar en Power Query.



# Lista de pasos

Query Editor interface showing a table with columns and a list of applied steps.

Formula Bar: `= Table.RenameColumns(#"Changed Type",{{"Column1", "NewColumnName"}})`

	NewColumnName	Column2	Column3	Column4	Column5
1	22c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
2	24c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
3	26c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
4	28c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
5	2ac09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
6	2cc09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
7	2ec09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
8	30c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
9	32c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	
10	34c09a13-84ab-ea11-a812-000d3a31a763	6/11/2020 1:38:02 AM	6/11/2020 1:38:02 AM	0	

Query Settings

PROPERTIES

Name: 2020

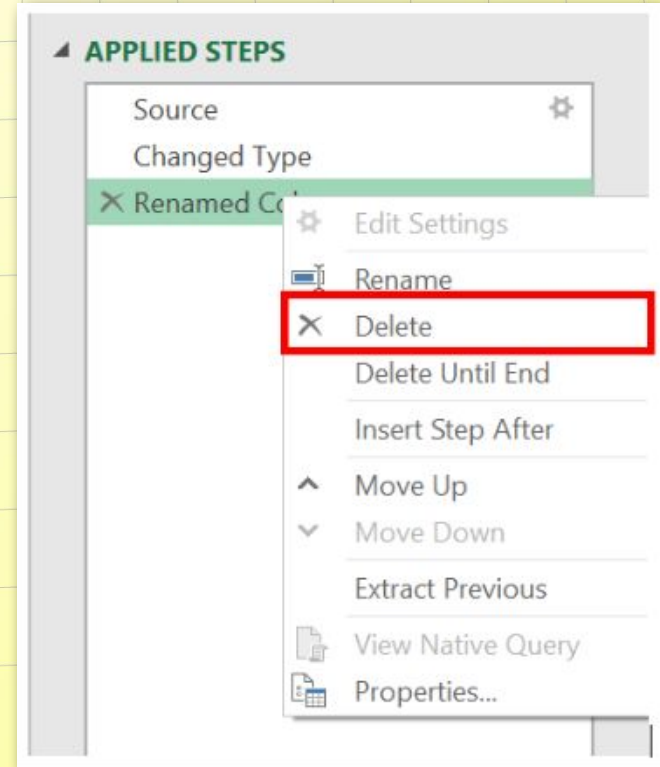
APPLIED STEPS

- Source
- Changed Type
- X Renamed Columns**



# Pasos

Se pueden borrar pasos, cambiar el orden, renombrar, entre otras opciones.





# Ejemplos de transformaciones





# Promover encabezados



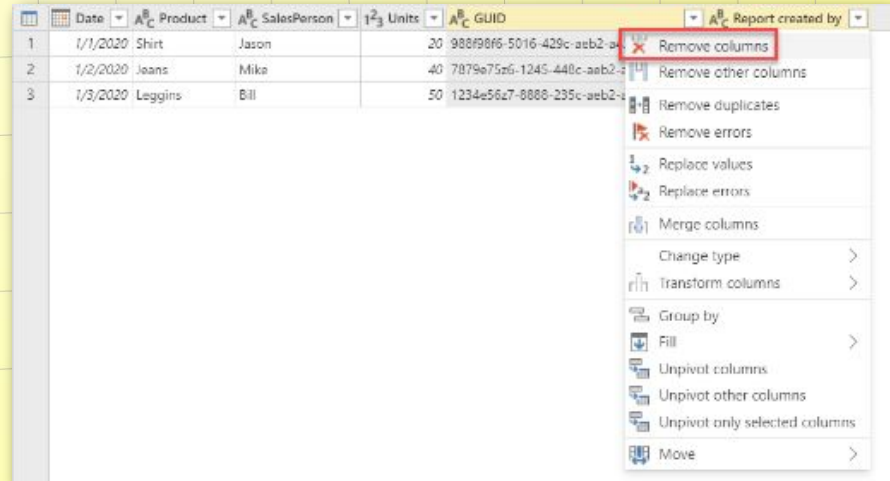
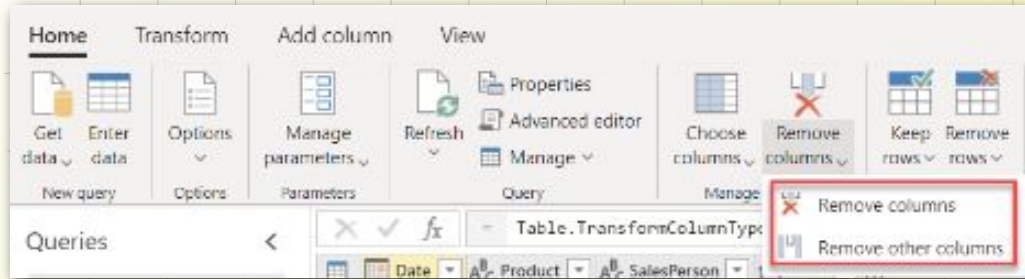
Column1	Column2	Column3	Column4
Country	Total Units	Total Revenue	
Panama	556	1917.35	
USA	926	4208.45	
Canada	157	72.25	
Panama	334	932.44	
USA	434	1142.87	
Canada	407	5651.3	
Mexico	806	4614.86	

- Use first row as headers
- Add custom column
- Add conditional column
- Index column
- Choose columns
- Keep top rows
- Keep bottom rows
- Keep range of rows
- Keep errors
- Remove top rows
- Remove bottom rows
- Remove alternate rows
- Remove duplicates
- Remove errors
- Merge queries
- Append queries





# Remover columnas





# Tipo de datos

1.2 Decimal Number    \$ Fixed Decimal Number    1<sup>2</sup> Whole Number    % Percentage    Date/Time    Date

Home   Transform   Add column   View

Group by   Use first row as headers   Transpose   Reverse rows   Count rows   Replace values

1   2

Data type

- Decimal number
- Currency
- Whole number
- Percentage
- Date/time
- Date
- Time
- Date/Time/Zone
- Duration
- Text
- True/False
- Binary
- Using locale...

Rename   Pivot columns   Unpivot columns   Convert to list   Any column

table", ({"column1", "Date"})

Queries

	ABC	Date
1	1/1/2019	
2	1/2/2019	
3	1/3/2019	
4	1/4/2019	
5	1/5/2019	
6	1/6/2019	
7	1/7/2019	
8	1/8/2019	
9	1/9/2019	
10	1/10/2019	
11	1/11/2019	
12	1/12/2019	
13	1/13/2019	
14	1/14/2019	
15	1/15/2019	

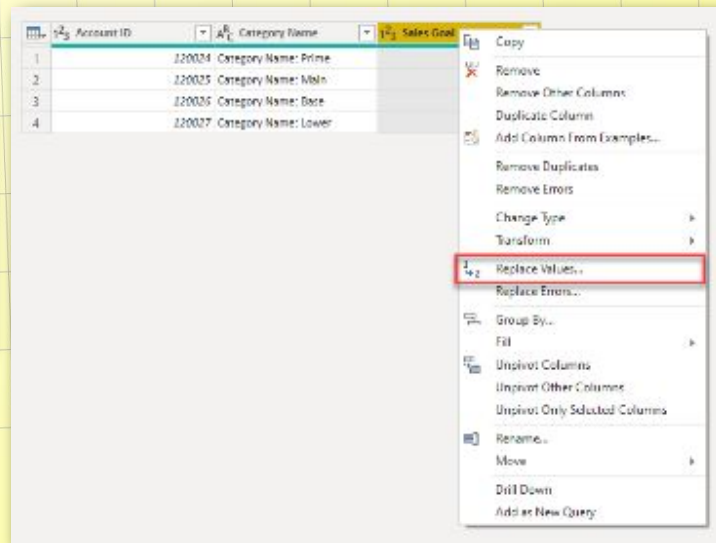
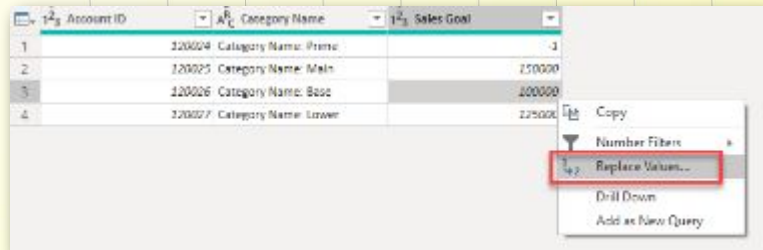
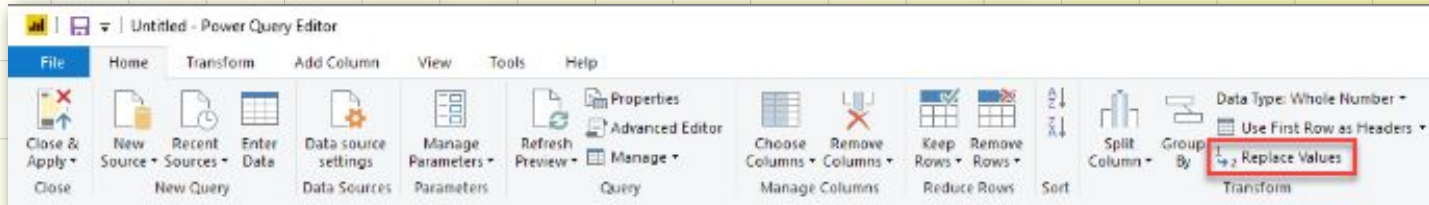
ABC Date

- 1.2 Decimal number
- \$ Currency
- 1<sup>2</sup> Whole number
- % Percentage
- Date/Time
- Date
- Time
- Date/Time/Zone
- Duration
- Text
- True/False
- Binary
- Using locale...

1/16/2019

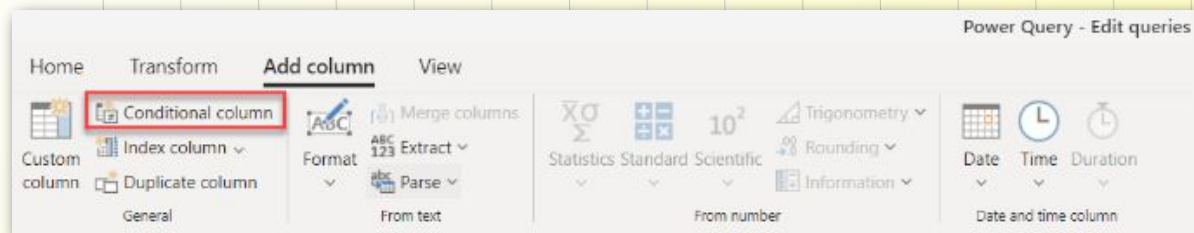


# Reemplazar valores





# columnas condicionales



Add conditional column

Add a conditional column that is computed from the other columns or values.

1 New column name  
Final Price

2

	Column name	Operator	Value	Output
if	CustomerGroup	equals	1	Tier 1 Price

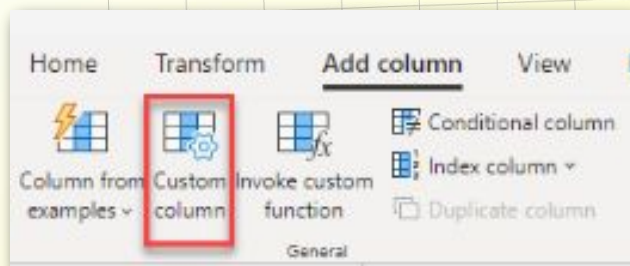
Add clause

3 Else

	Tier 3 Price
--	--------------

OK Cancel

# columnas personalizadas



**Custom column** ?

Add a column that is computed from other columns or values.

New column name \*

Custom

Data type

Available column(s)

Date

Country

Units

Unit Price

Discount

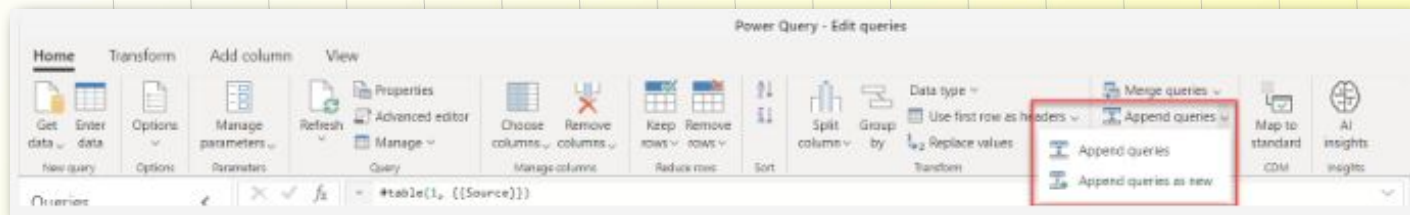
Custom column formula \*

Insert column

Learn more about Power Query formulas

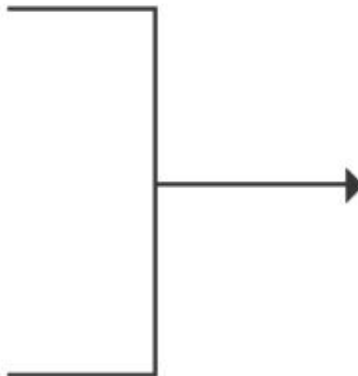
OK Cancel

# Anexar



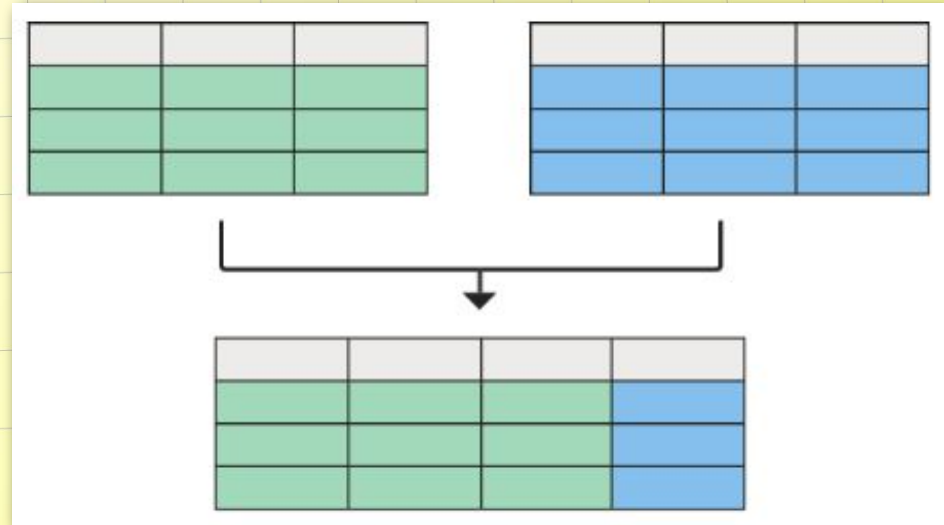
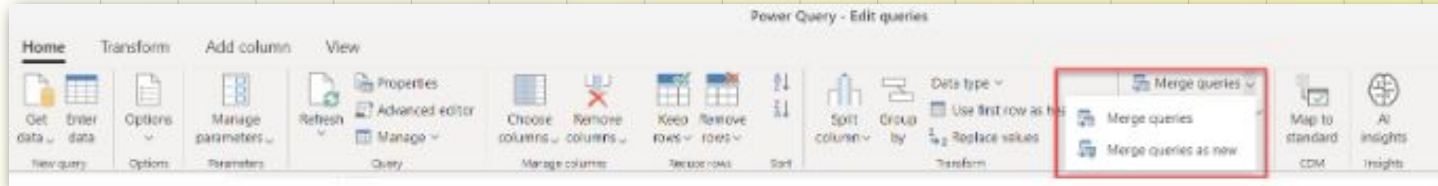
A	B	C
1	1	1
2	2	2
3	3	3

A	B	D
4	4	4
5	5	5



A	B	C	D
1	1	1	null
2	2	2	null
3	3	3	null
4	4	null	4
5	5	null	5

# Combinar



# Combinar



## Combinar

Seleccione una tabla y las columnas coincidentes para crear una tabla combinada.

Canal

CODIGO	DESCRIPCION
1	Telefónica
2	Online
3	Presencial

Venta

idVenta	Fecha	Fecha_Entrada	idCanal	idSocio	idSucursal	idEmpleado	idProducto	Precio
7	2/1/2018	1/1/2018	3	180	1	16/0	4081	10.0
49	31/12/2015	4/1/2016	2	1133	1	16/0	42737	16.0
65	13/11/2017	14/11/2017	1	639	1	16/0	42737	16.0
284	31/12/2015	4/1/2016	3	1133	1	1516	42737	16.0

Tipo de combinación

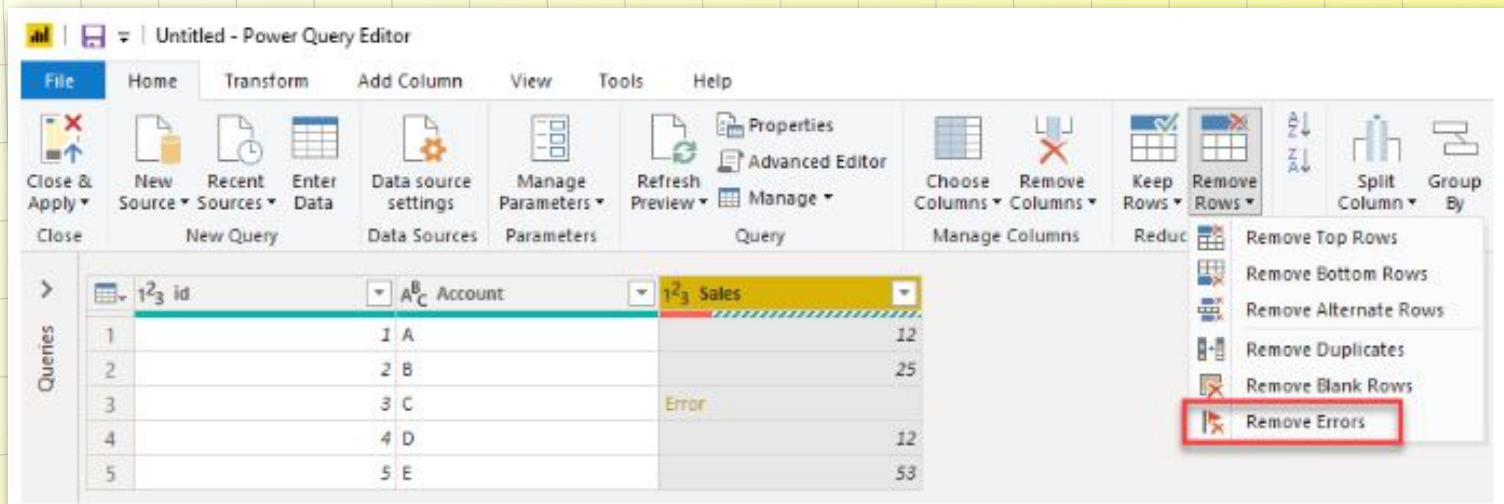
- Externa izquierda (todas de la primera, coincidencias de la segunda)
- Externa izquierda (todas de la primera, coincidencias de la segunda)
- Externa derecha (todas de la segunda, coincidencias de la primera)
- Externa completa (todas las filas de ambas)
- Interna (todas las filas coincidentes)
- Anti izquierda (solo filas de la primera)
- Anti derecha (solo filas de la segunda)

La selección coincide con 3 de 3 filas de la primera tabla.

Aceptar Cancelar



# Errores





**¿PREGUNTAS?**





**Vamos al programa...**



**Power BI**



# ¿Alguien dijo Homework?



~~HENRY~~



Próxima lecture

# **Power Query avanzado**





# ¡Feedback!

Click on me



Dispones de un **formulario** en:



Homeworks



Guías de clase



Slack

# HENRY

