



Microsoft Power BI

Aviso

Se recuerda que esta sesión será grabada, salvo el listado de asistentes y las preguntas del chat, que no serán objeto de grabación. Si no has prestado tu consentimiento, recuerda desactivar tu cámara y micro, ya que si realizas una participación verbal esta constituirá tu consentimiento para la grabación

Cosas nuevas

Tablas de look up (el problema de los meses que deben aparecer)

Agrupaciones en PowerQuery

Medidas en tablas

Problema de los N primeros elementos (2 versiones)

DAX

Combinación de tablas

Problema/ejercicio

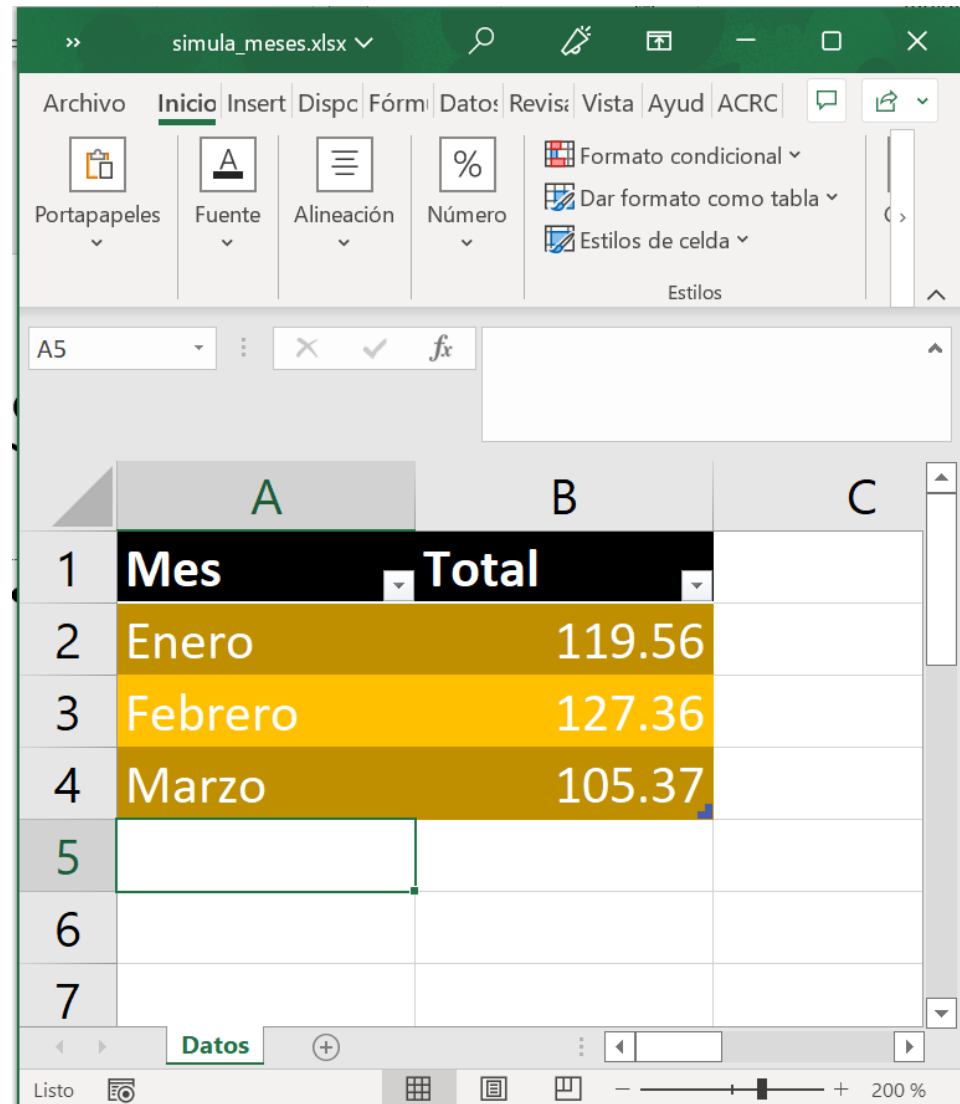
- Tenemos un informe con datos por meses que se va rellorando mensualmente; el primer mes solo aparece enero, el segundo enero y febrero, etc.
- Sin embargo, al hacer el informe queremos que siempre aparezca de enero a diciembre
- ¿Cómo solucionarlo?



Simulación

Creemos un documento Excel nuevo con una sola hoja representando esta situación y grabémoslo por ejemplo con nombre de “simula_meses.xlsx” con dos columnas, **Mes** y **Total**

Tras grabarlo podemos dejarlo abierto, lo vamos a modificar simulando la llegada de un nuevo mes



The screenshot shows an Excel window titled "simula_meses.xlsx". The "Inicio" ribbon is active, showing options like "Portapapeles", "Fuente", "Alineación", "Número", "Formato condicional", "Dar formato como tabla", and "Estilos de celda". The spreadsheet has two columns, A and B, and rows 1 through 7. Column A is labeled "Mes" and column B is labeled "Total". The data is as follows:

	A	B
1	Mes	Total
2	Enero	119.56
3	Febrero	127.36
4	Marzo	105.37
5		
6		
7		

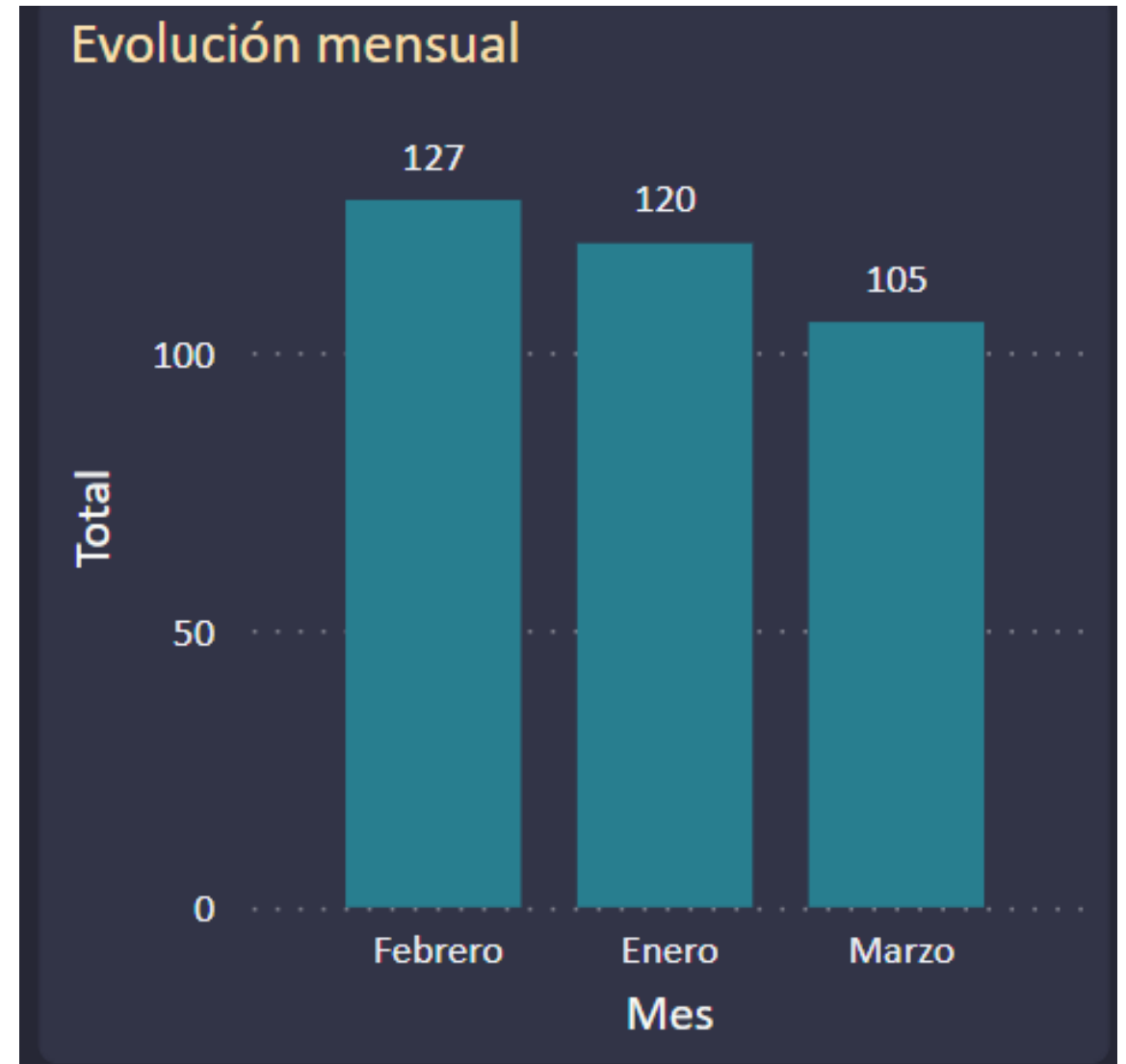
The status bar at the bottom shows "Listo" and a zoom level of 200%.

Simulación (II)

En Power BI Desktop, carguemos este fichero y creemos por ejemplo un gráfico de barras con **Mes** en las x y **Total** en las y

Descubrimos que no tenemos un problema, tenemos 2

- No salen todos los meses
- ¡Las barras están desordenadas!



Solución (I)

Elegimos “Especificar Datos” y creamos una tabla de meses numerados (podríamos hacerlo también en Excel, pero por probar otra funcionalidad...)

La llamamos **Meses**

Crear tabla

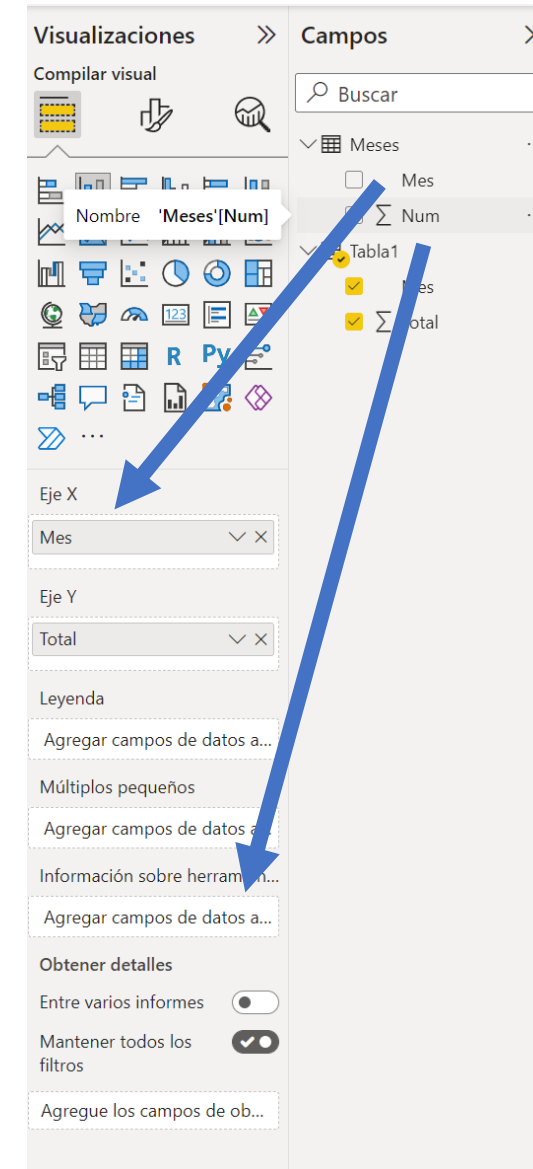
	Num	Mes	+
1	1	Enero	
2	2	Febrero	
3	3	Marzo	
4	4	Abril	
5	5	Mayo	
6	6	Junio	
7	7	Julio	
8	8	Agosto	
9	9	Septiembre	
10	10	Octubre	
11	11	Noviembre	
12	12	Diciembre	
+			

Nombre:

Solución (II)

Seleccionamos el diagrama de barras y añadimos en “información sobre herramientas” el campo numérico de la tabla nueva

Además reemplazamos en el eje X el Mes de la hoja Excel por el de la tabla nueva



¡Hecho!

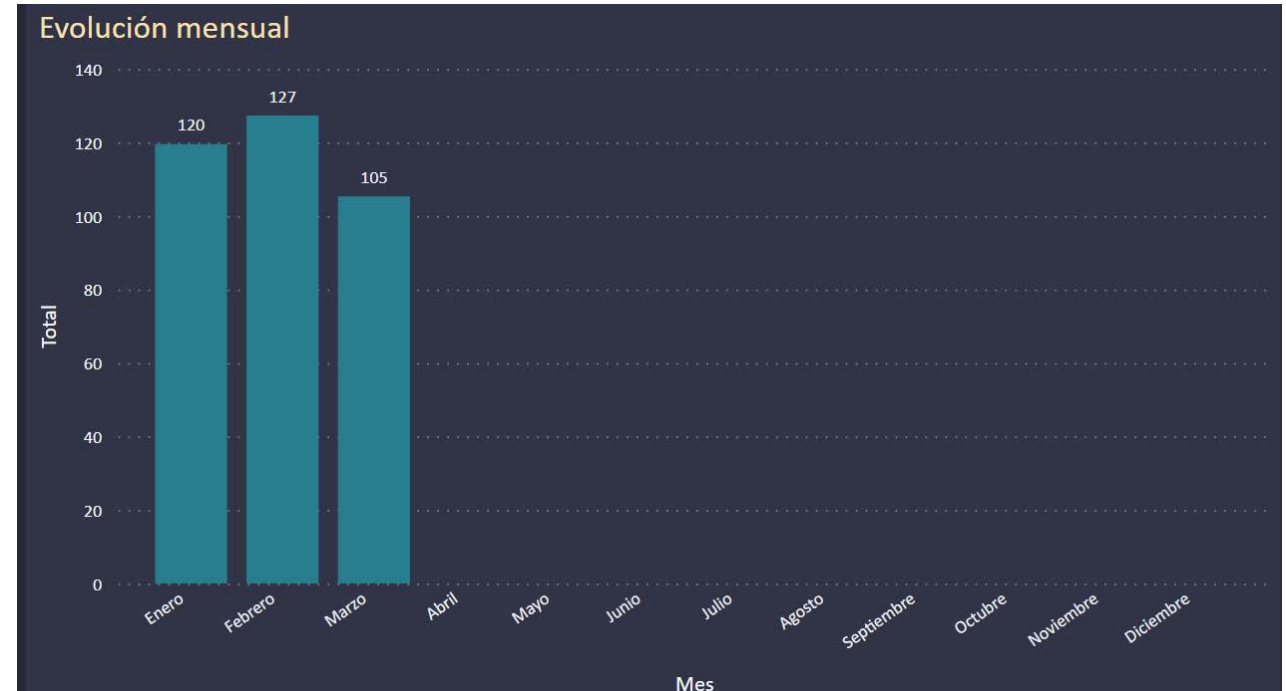
Solo nos falta ordenar (los ... del gráfico, arriba a la derecha) por Sum Num

Ya tenemos el gráfico como queríamos

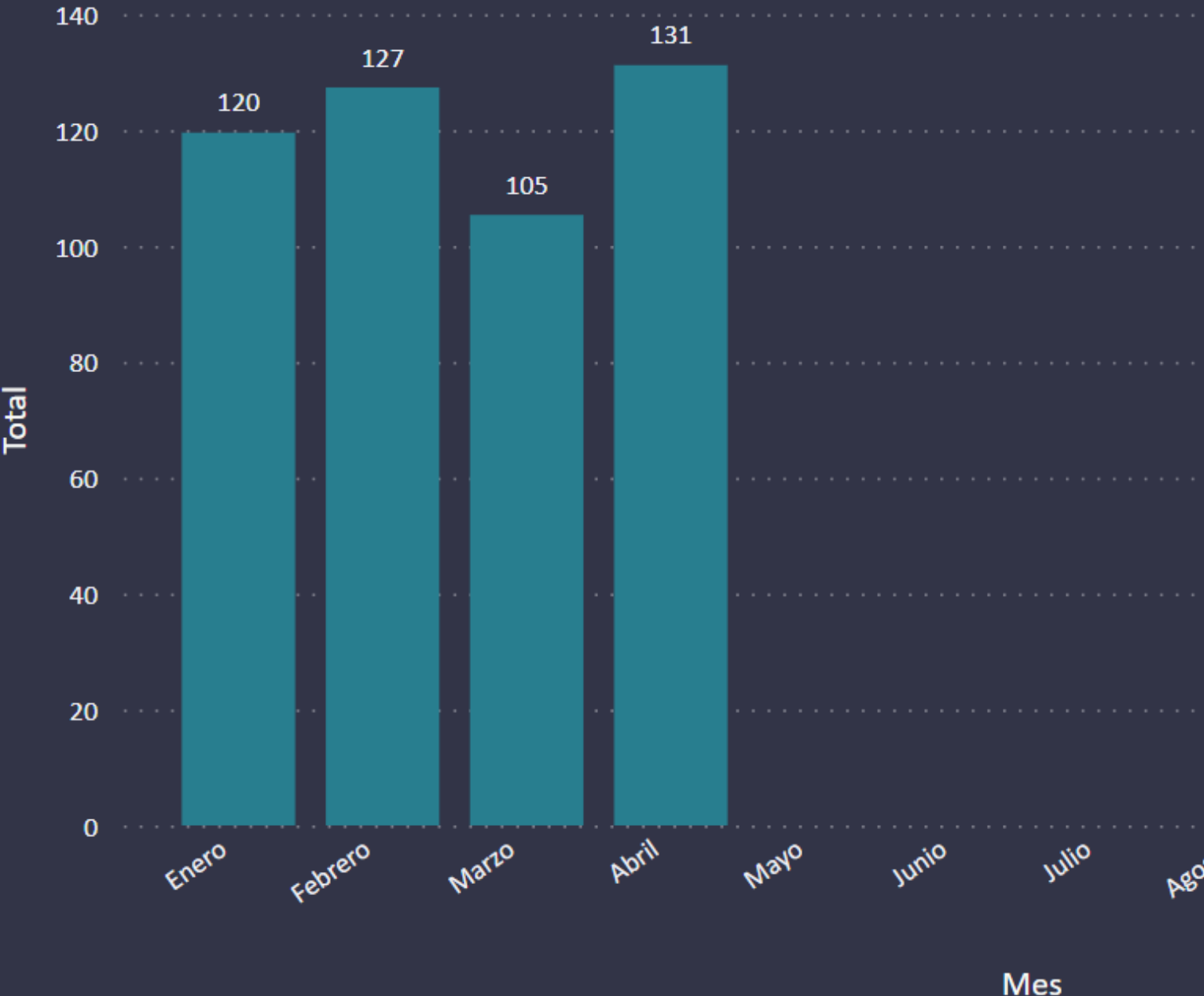
En modelo vemos que las tablas se han enlazado automáticamente

Este “truco” se conoce como “tabla de look-up”

Podemos probar a añadir un valor nuevo a la hoja Excel, grabar, y dar actualizar en Power BI



Evolución mensual



Mes	Total
Enero	119.56
Febrero	127.36
Marzo	105.37
Abril	131.21

The background is a dark blue gradient with a subtle grid of small dots. A large, stylized butterfly made of blue and white triangles is positioned in the upper right. Below it, a trail of small blue squares leads towards the bottom center. In the lower left, a portion of a laptop is visible, also rendered in a low-poly, geometric style.

Transformación y modelado en PowerBI

Fases de procesamiento de datos en PowerBI

Al principio hablamos de 3 fases

- 1.- Carga de datos (para lograr tablas)
- 2.- Transformación de datos (preparar las tablas para poder hacer los informes)
- 3.- Visualización (informe)

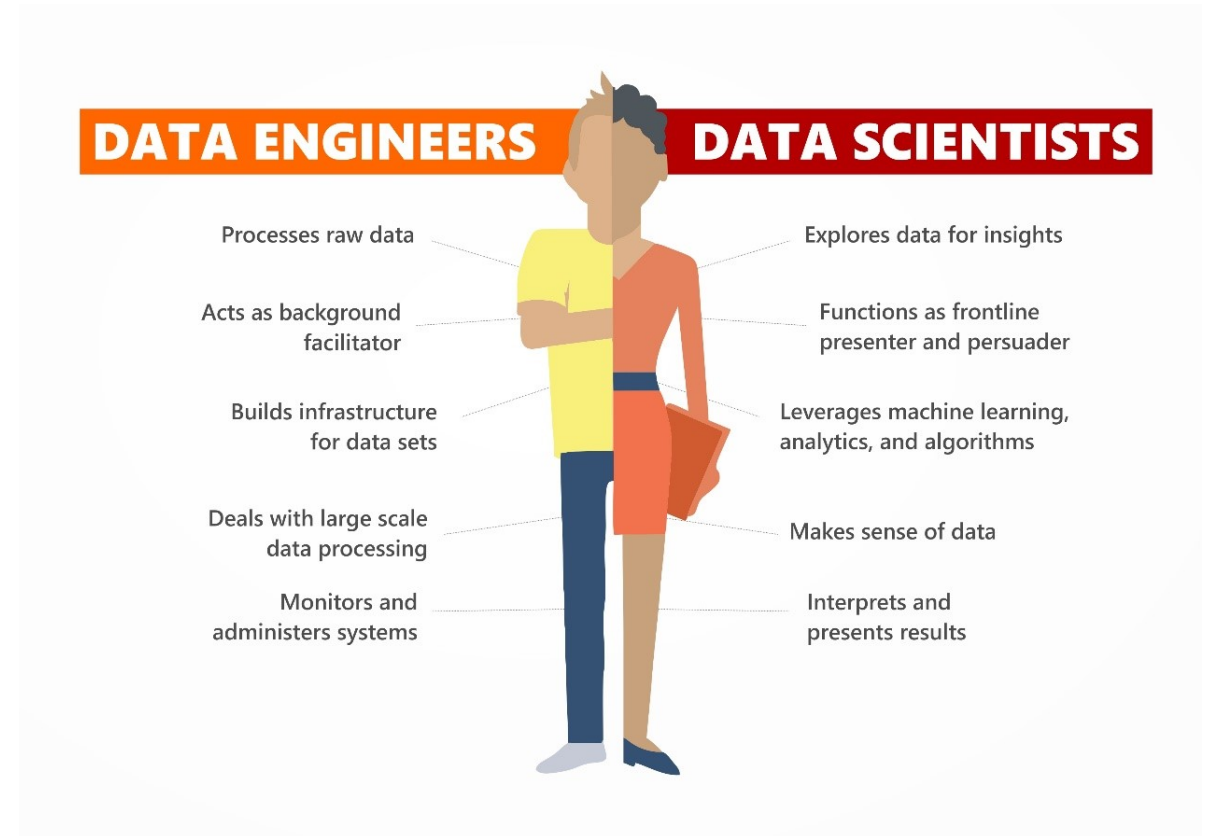
Transformación en detalle

Sin embargo la fase de “transformación” parece que a su vez se divide en Power Bi en varias

- 1) Una primera realizadas con Power Query Editor (transformar datos)
 - Más bajo nivel
 - Renombramiento de columnas, tipos, eliminación de caracteres raros, división agrupamiento de columnas, pivotar, trasponer, etc
- 2) Otra en la parte de datos e incluso durante la visualización
 - Más alto nivel nivel
 - Relaciones entre tablas (modelizado), formas de resumir, renombrar y formato, etc

Nueva division lógica

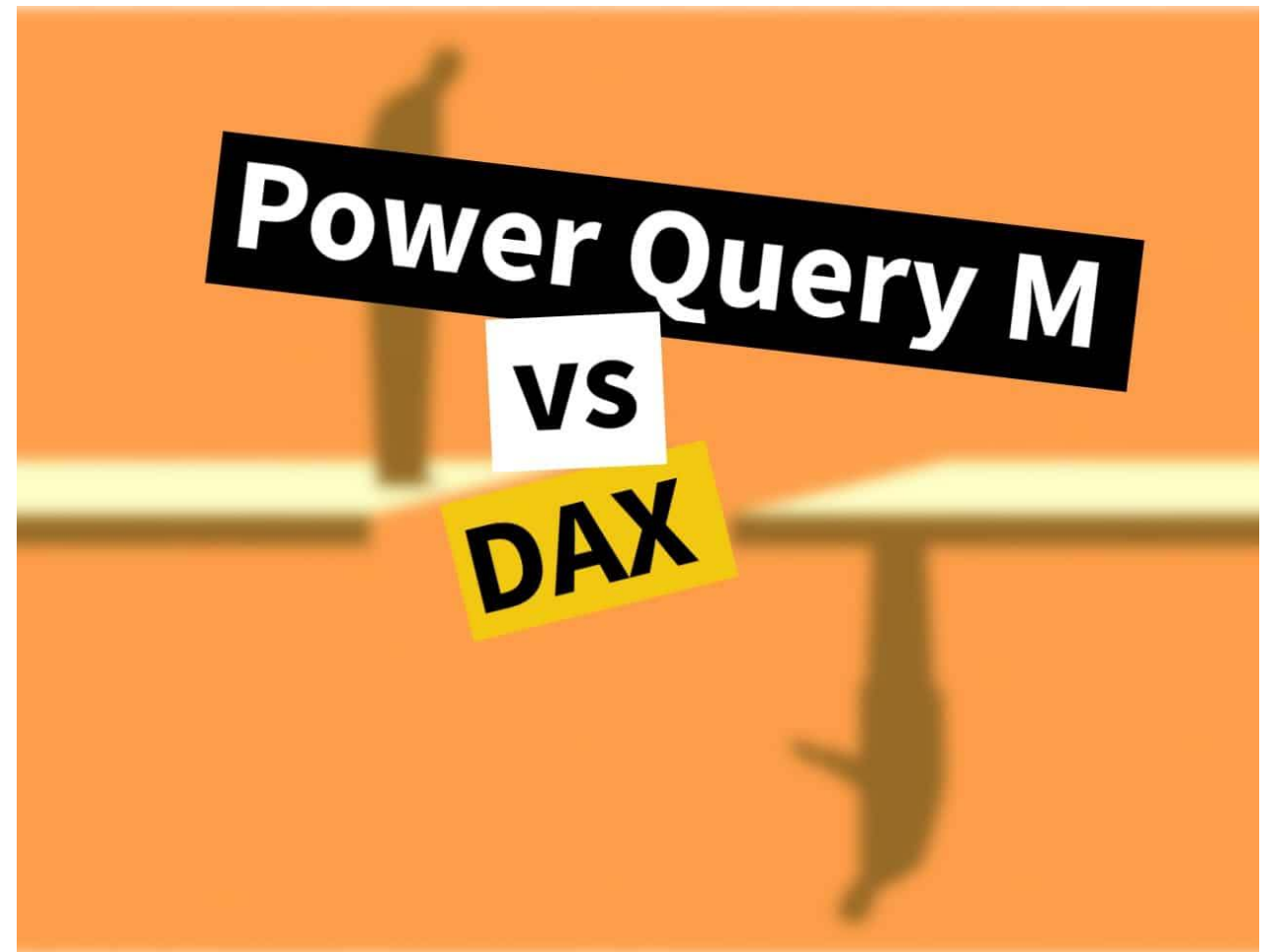
- 1) ETL (Extract Transform Load) en Power Query Editor (transformar datos) → **ingeniería de datos** (*data prep*)
- 2) **Análisis de datos**: modelización, informes, etc



Power BI: las dos visiones

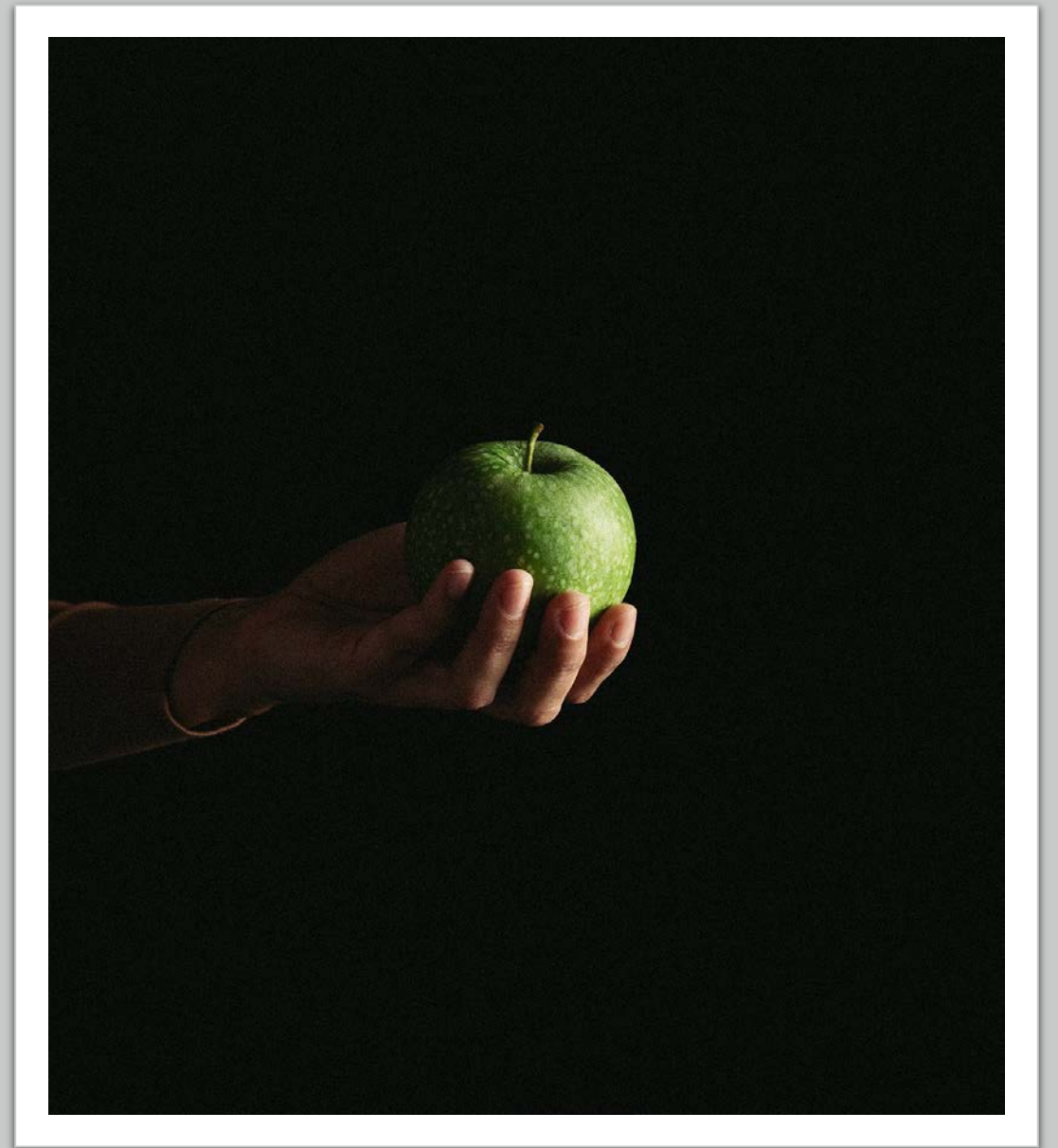
Power Query Editor tiene su propio lenguaje (**M**) para la parte “ingenieril”

Fuera de Power Query Editor tenemos otro lenguaje (**DAX**) para la parte de análisis



Los claroscuros...

- Hay aspectos, como la limpieza de datos, que podrían caer en ambos apartados y se pueden hacer de dos formas
- ¿Y qué más da?
- Sí que da, hacerlo en un lado o en otro puede cambiar la funcionalidad de la solución



ETL: Power Query Editor

Arranca con “Transformar Datos”

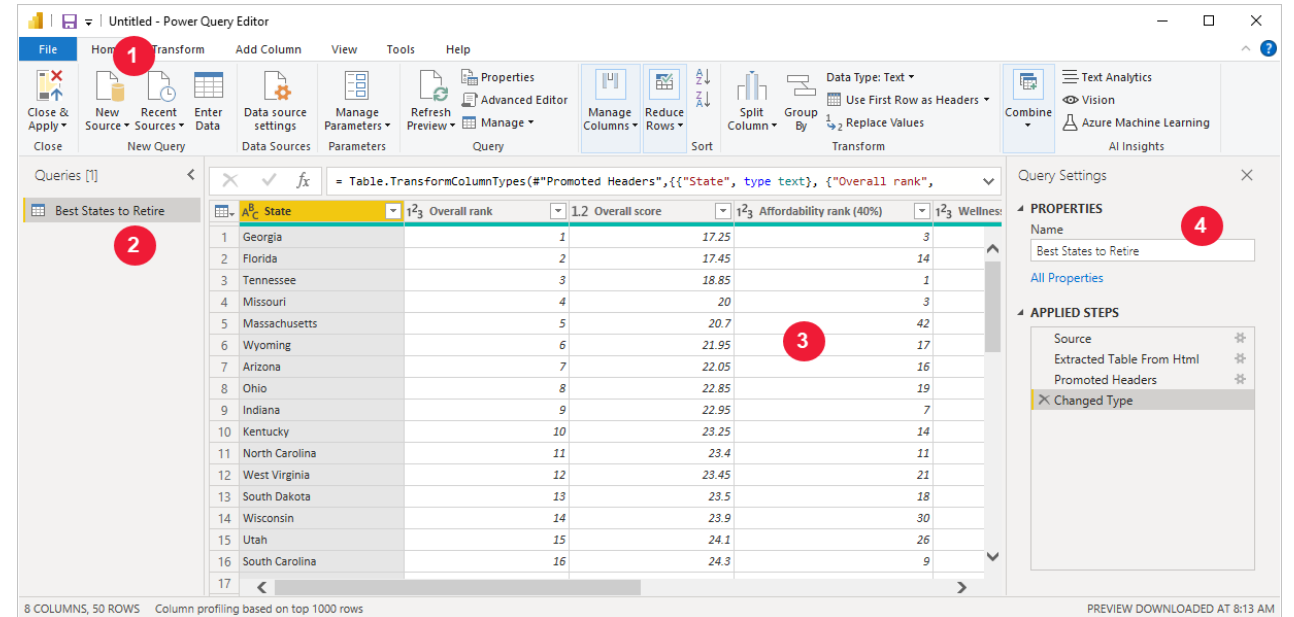
4 partes

1 Cinta de menús y botones

2 Fuentes de datos (consultas)

3 Datos de la consulta seleccionada

4 Configuración de la consulta



Cinta de opciones

Pestañas

Inicio: tareas comunes

Transformar: transformaciones más habituales

Agregar columna: columnas nuevas

Vista: Apariencia al mostrar la tabla y estadísticas (nulos, etc)

Herramientas: diagnóstico

Ejemplo: limitar el número de valores que se muestran en un diagrama de barras (V1)

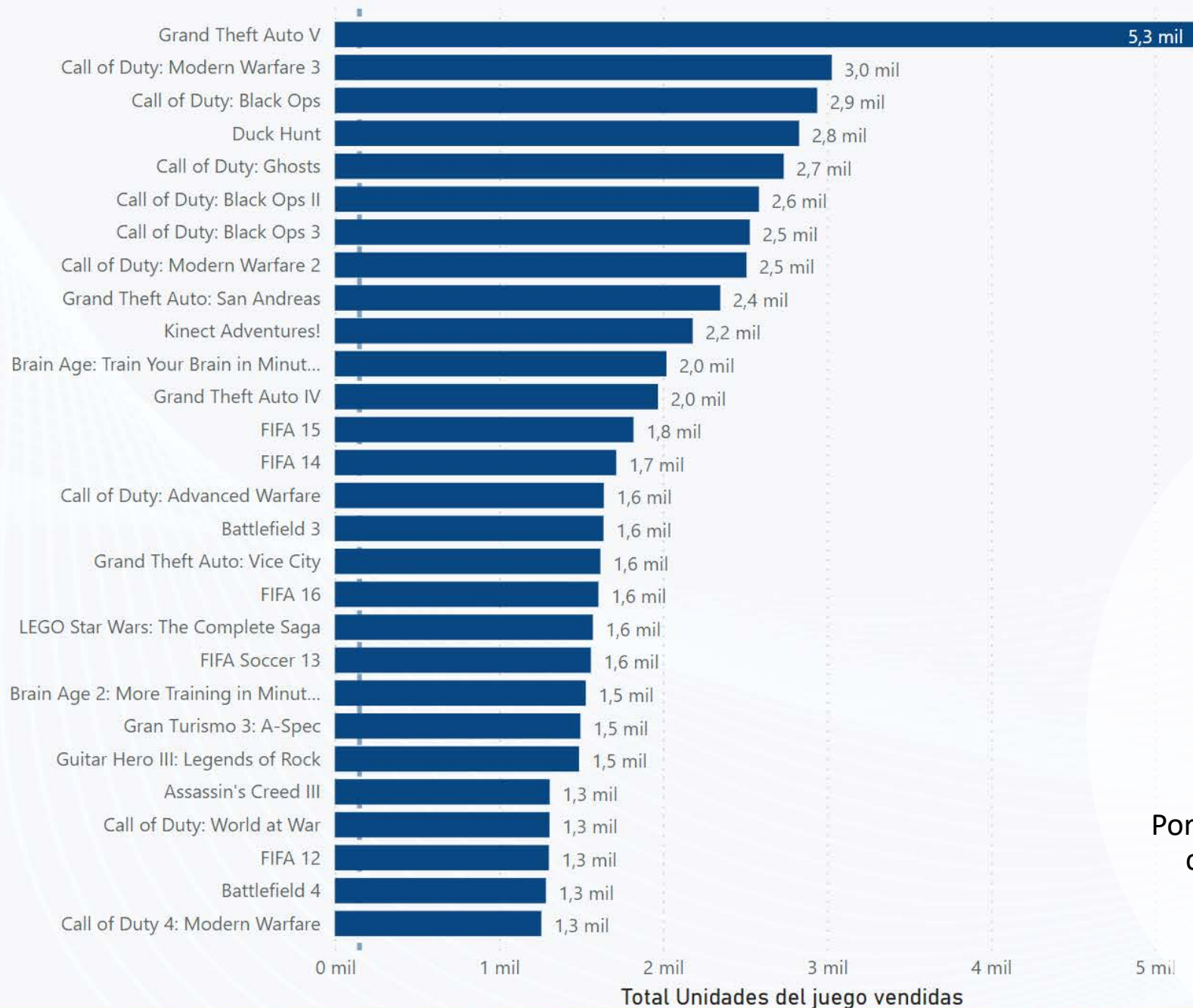
Partimos del fichero de videojuegos ya limpio y preparado

<https://github.com/RafaelCaballero/tdm/blob/master/powerbi/videojuegos.csv>

Queremos mostrar los títulos de los 10 videojuegos más vendidos



Nombre



30,00 mil

Total de juegos

Ejemplo: la triste
realidad

Por desgracia actualmente Power BI no incluye un control del número de elementos mostrados

Cerrar y aplicar

Cerrar

Nuevo origen

Nueva consulta

Orígenes recientes

Especificar datos

Configuración de origen de datos

Orígenes de datos

Administrar parámetros

Parámetros

Actualizar vista previa

Consulta

Propiedades

Editor avanzado

Administrar

Elegir columnas

Administrar columnas

Quitar columnas

Conservar filas

Reducir filas

Quitar filas

Ordenar

Dividir columna

Agrupar por

Reemplazar los valores

Tipo de datos: Número entero

Usar la primera fila como encabezado

Transformar

Consultas [3]

pinguinos

videojuegos

videojuegos (2)

123 Año

AB C

1	2006	Nar
2	2006	Nar
3	2006	Nar
4	2006	Nar
5	2006	Nar
6	2006	Nar
7	2006	Nar
8	2006	Nar
9	2006	Nar
10	2006	Nar
11	2006	Nar
12	2006	Nar

Agrupar por

Especifique la columna por la que quiera realizar la agrupación y la salida deseada.

Básico

Uso avanzado

Nombre

Nuevo nombre de columna

Operación

Columna

Ventas totales

Suma

Unidades Vendidas

Aceptar

Cancelar

Ejemplo: la solución
Power Query

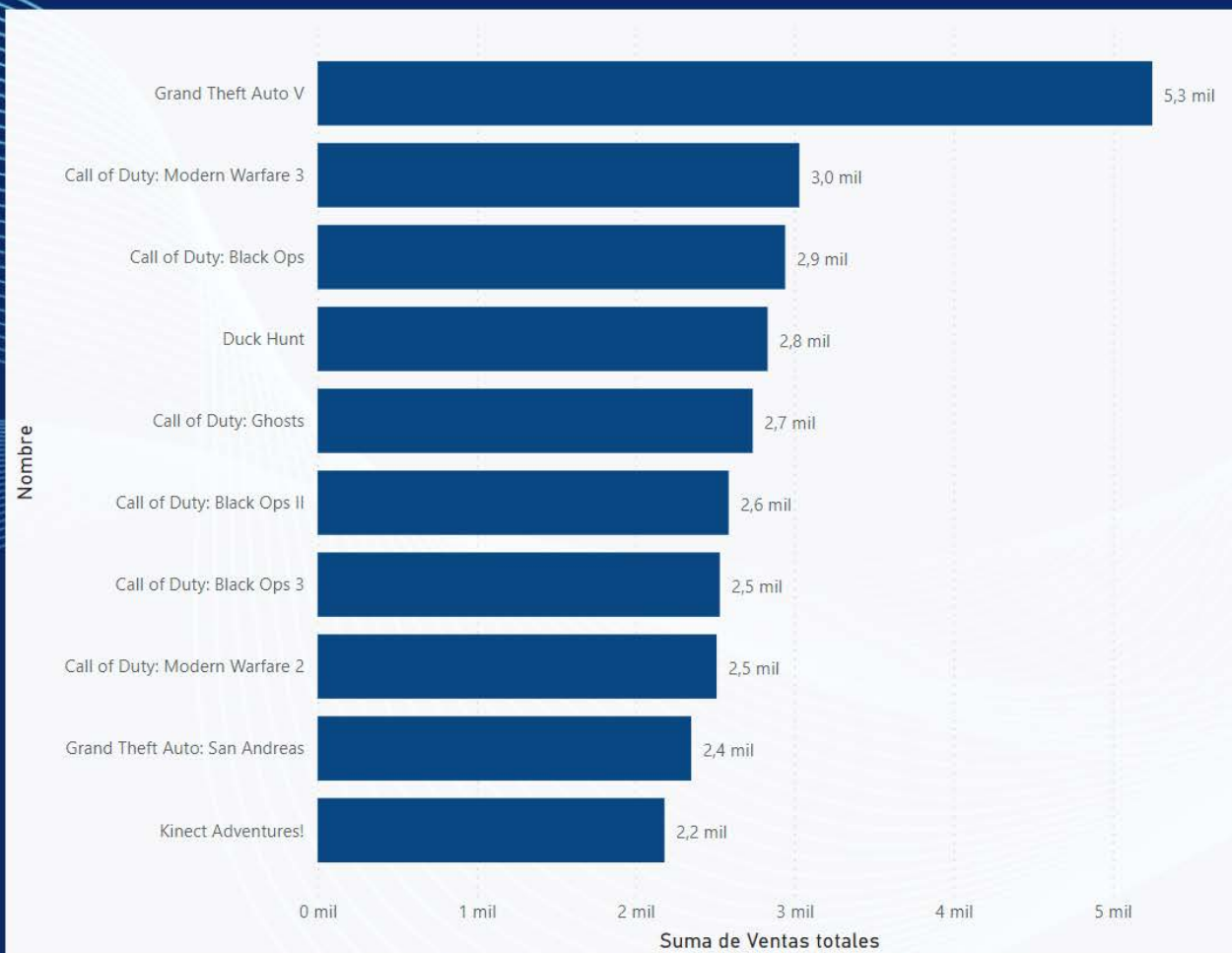
Transformar datos
Duplicamos la tabla para comparar resultados
Damos agrupar por para calcular los totales

Ejemplo: recortando filas en power query

Ahora ya sabemos hacerlo:

- Dar ordenar por total de ventas descendente
- Quedarse con las 10 filas superiores (conservar filas superiores)
- Hacer el informe como siempre

	ABC Nombre	1.2 Ventas totales
1	Grand Theft Auto V	5250
2	Call of Duty: Modern Warfare 3	3030
3	Call of Duty: Black Ops	2941
4	Duck Hunt	2831
5	Call of Duty: Ghosts	2737
6	Call of Duty: Black Ops II	2586
7	Call of Duty: Black Ops 3	2530
8	Call of Duty: Modern Warfare 2	2510
9	Grand Theft Auto: San Andreas	2350
10	Kinect Adventures!	2182



10
Recuento de Nombre

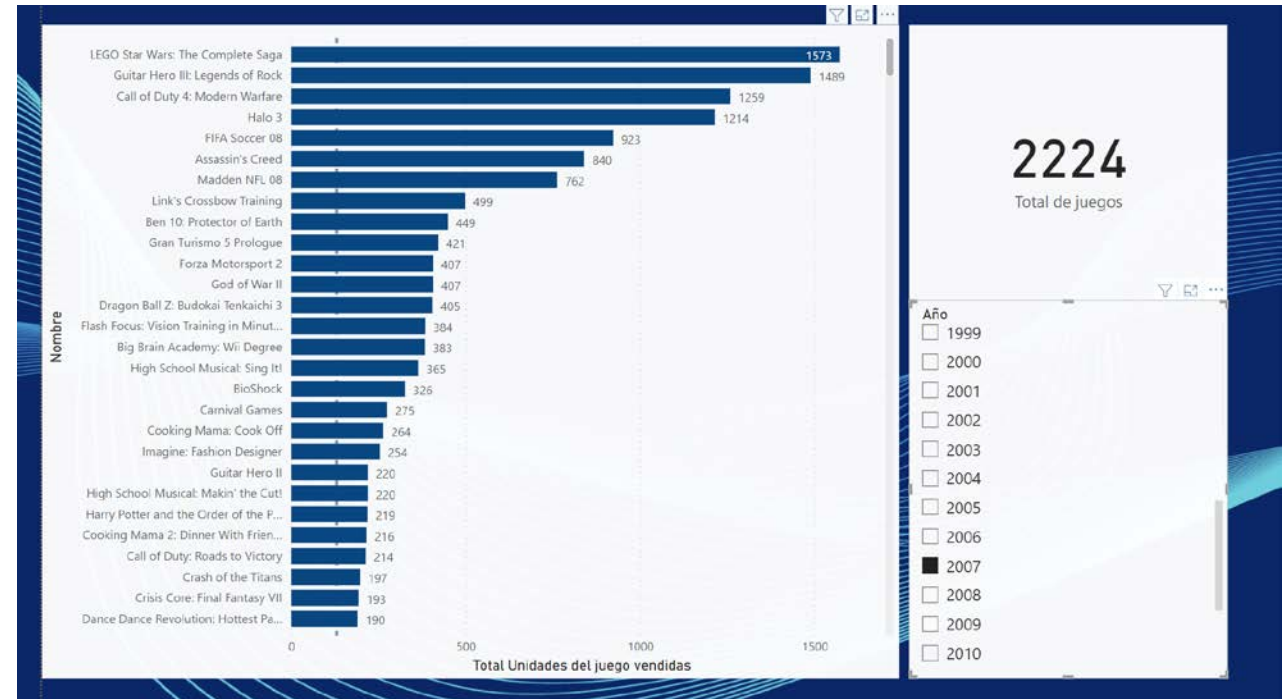


Problemas y desventajas

¿Qué pasa si queremos añadir un filtro de año? Hacerlo en las dos páginas

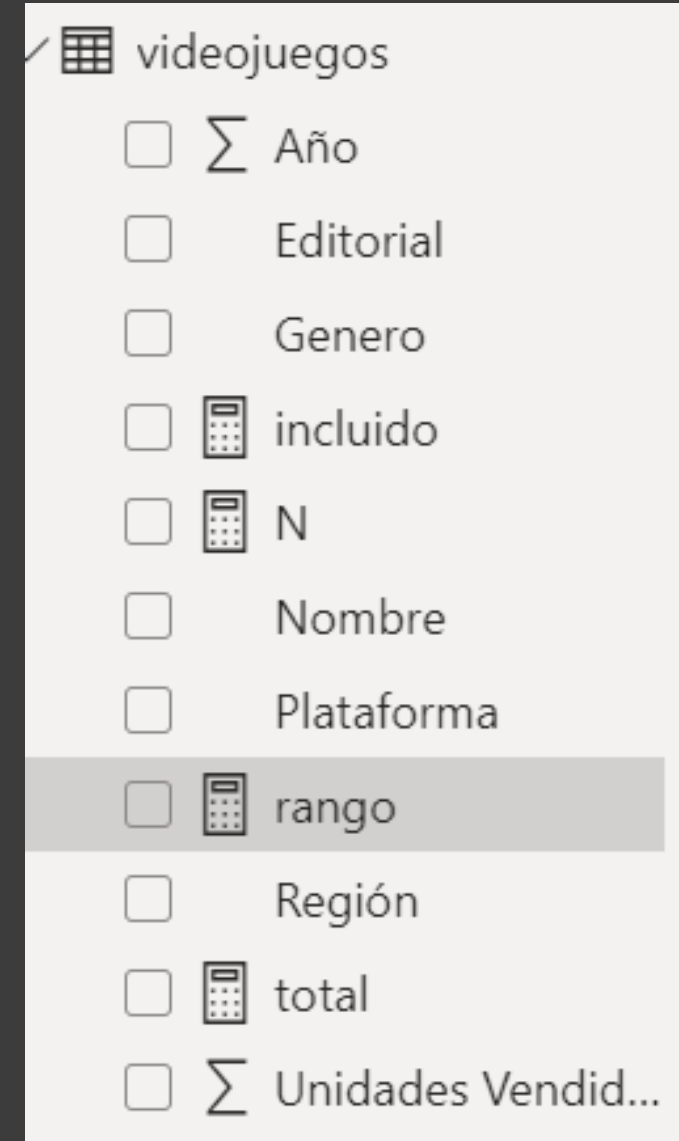
¿Qué se puede hacer para arreglarlo?

Conclusión: hemos actuado a nivel de ETL pero es demasiado “pronto” debería ser a nivel de análisis



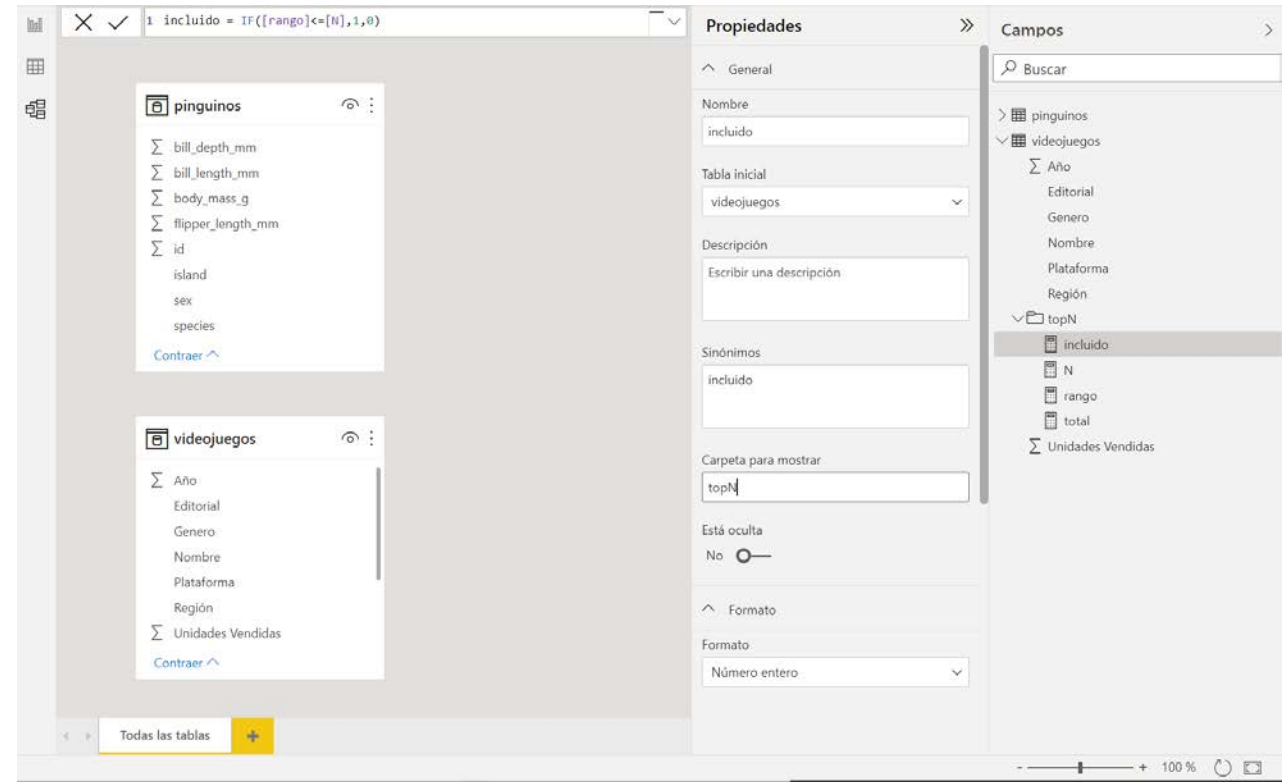
Medidas

- Se pueden añadir en la parte de análisis de datos: vista de informe, de datos o de modelo
- Se definen mediante el lenguaje DAX
- NO son columnas nuevas aunque se alojan en una tabla
- Similares a las medidas de Excel
- Se muestran como campos de la tabla pero con un icono de calculadora



Medidas organizadas

- Se pueden agrupar en carpetas para mejor organización
- Elegir “Modelo” a la izq. y en propiedades escribir el nombre de la carpeta
- Se puede incluir la misma medida en varias carpetas poniendo los nombres separados con ; (y crear subcarpetas)

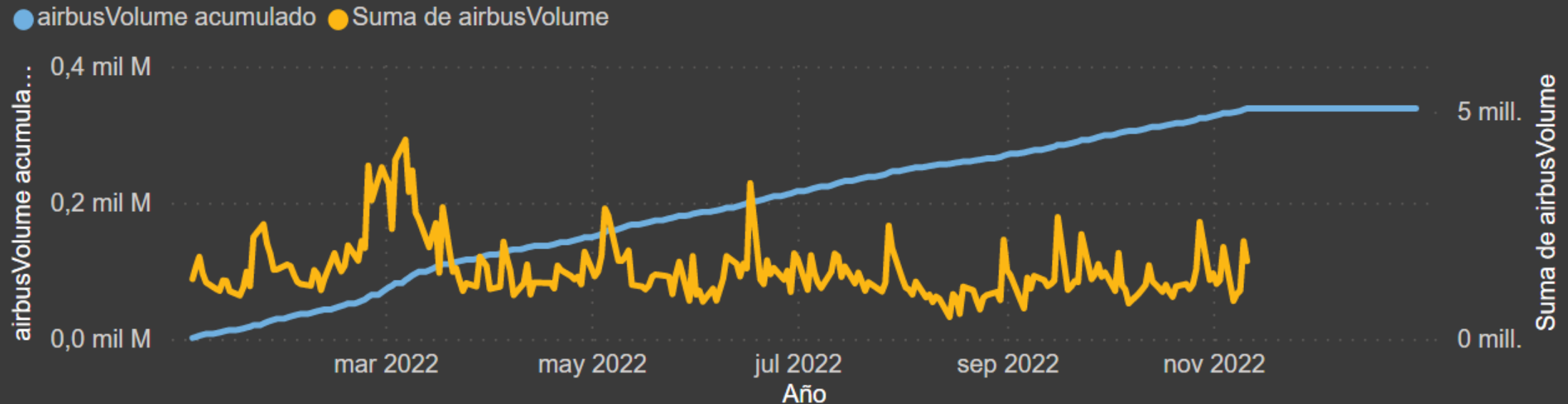


Medidas automáticas y rápidas

- En realidad todas las columnas que tienen un “resumen” tienen asociada una medida
- Si queremos ahorrarnos teclear el código DAX podemos usar las medidas rápidas
- Ejemplo: En el fichero <https://raw.githubusercontent.com/RafaelCaballero/tdm/master/datos/stockp22.csv>

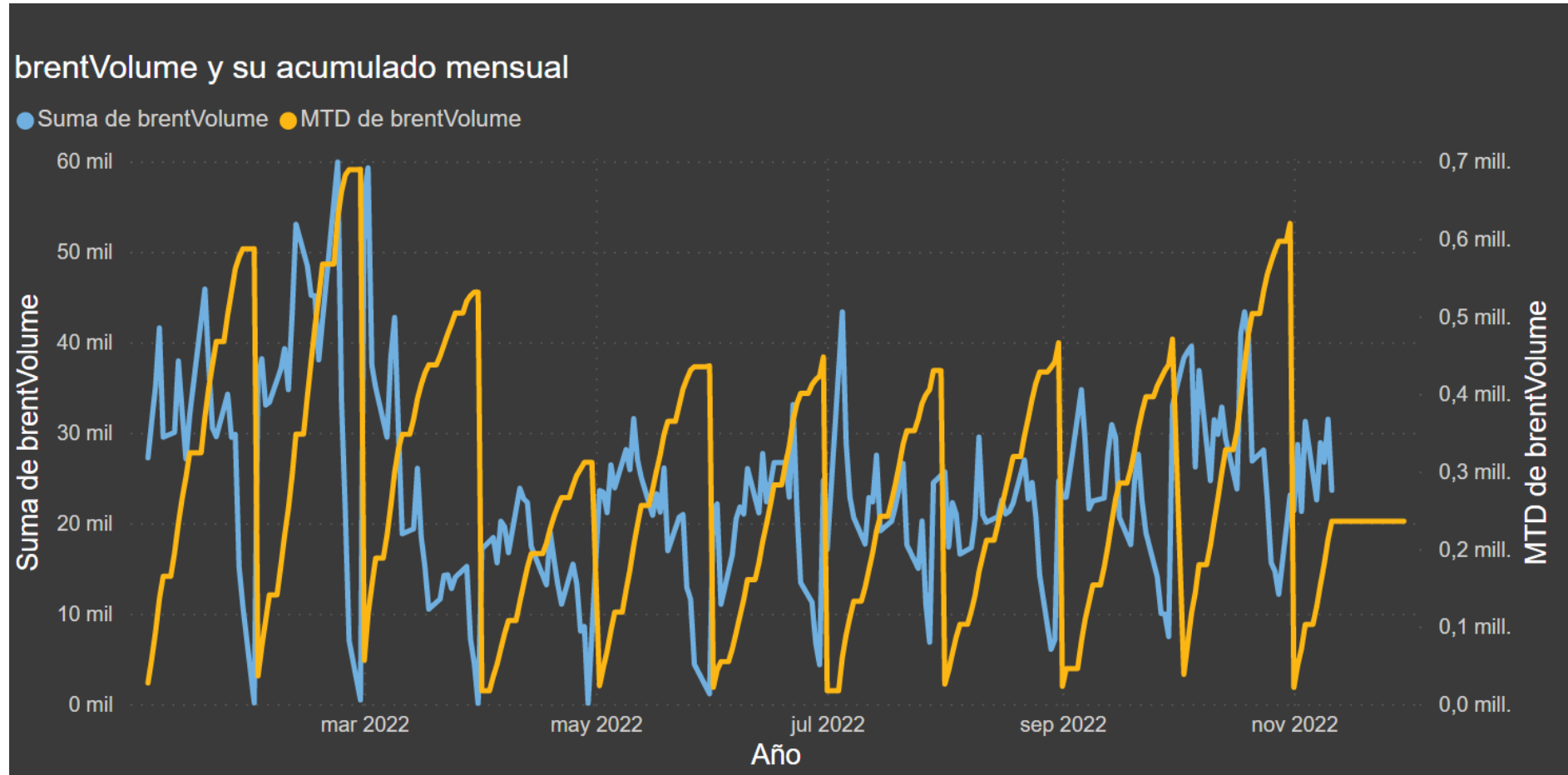
Queremos representar el valor airbusVolume junto con la suma anual acumulada de este mismo valor

airbusVolume con su suma acumulada



Ejercicio

Hacer algo similar pero para el volumen de Brent y su acumulado por meses



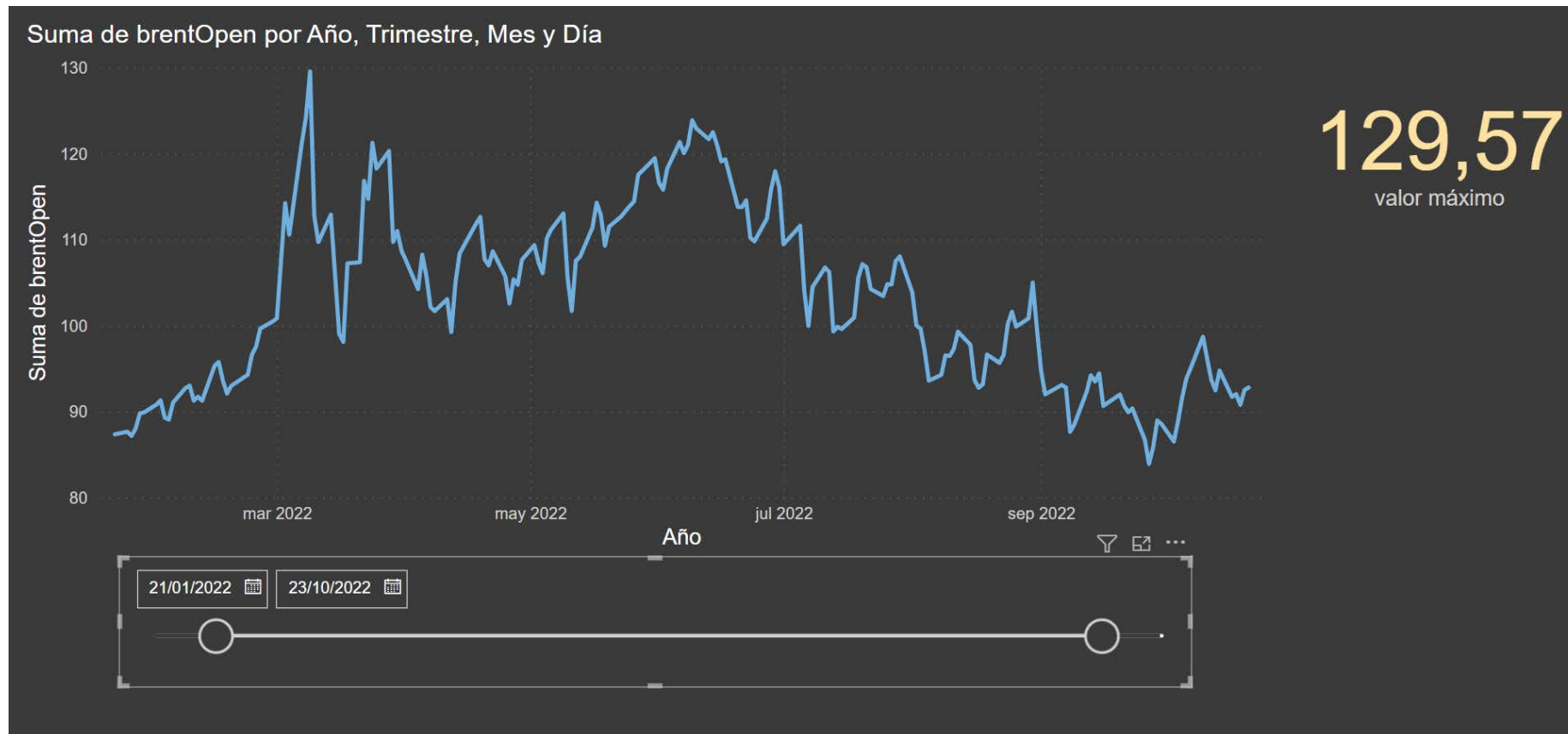
Medidas calculadas

Ejemplo:

Empezamos reproduciendo algo que ya sabemos: el máximo

```
mayorBrent = max(stockp22[brentOpen])
```

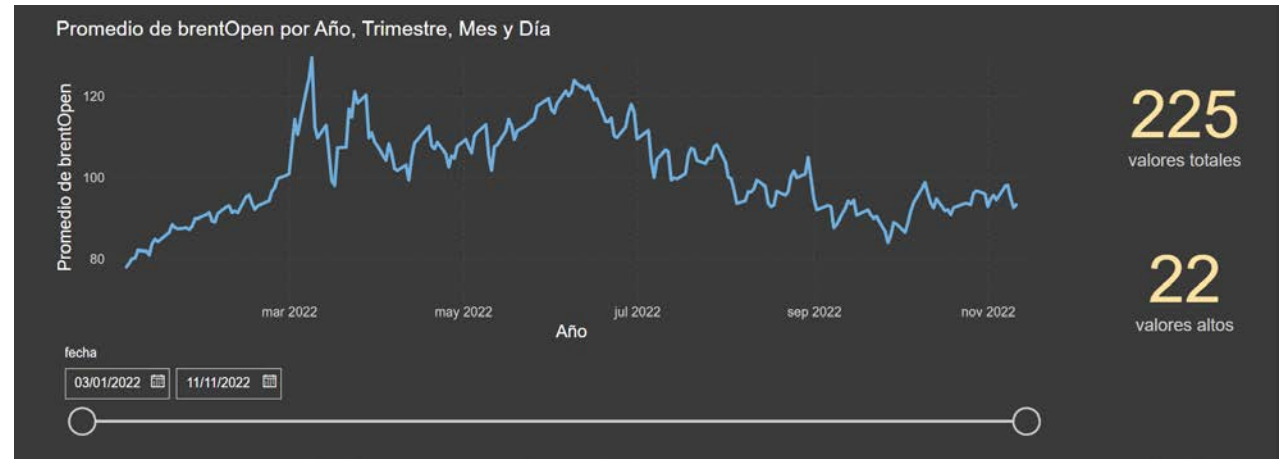
Podemos utilizar esta medida para crear una tarjeta



Medidas calculadas

Queremos saber superan el
90% del valor máximo

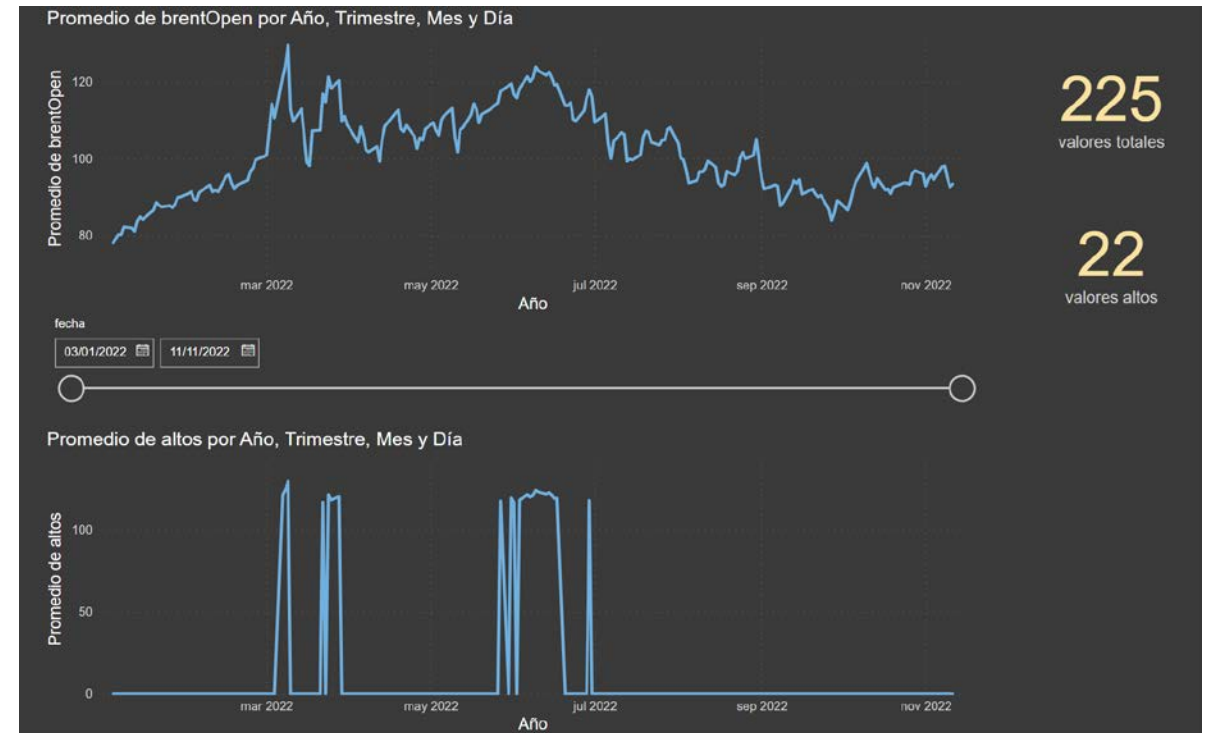
```
outlier = CALCULATE(Countrows(stockp22),  
    stockp22[brentOpen]>=max(stockp22[brentOpen])*0.9)
```



Columnas calculadas

Análogas a las medidas pero con un valor para cada fila

```
altos =  
IF(stockp22[brentOpen]>=max(stockp22[brentOpen])  
*0.9,stockp22[brentOpen],0)
```



Ejemplo: limitar el número de valores que se muestran en un diagrama de barras (V2)

Partimos del fichero de videojuegos ya limpio y preparado

<https://github.com/RafaelCaballero/tdm/blob/master/powerbi/videojuegos.csv>

Queremos mostrar los títulos de los 10 videojuegos más vendidos

Ahora usando DAX y medidas



Ejemplo V2: medidas

N: constante. Veremos,
como mucho, N elementos

total: unidades vendidas

rango: posición que ocupa la
fila actual en el orden de
ventas

incluido: Indica si la fila
actual debe ser incluida



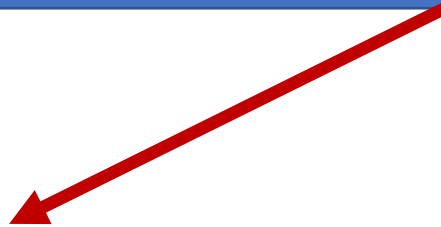
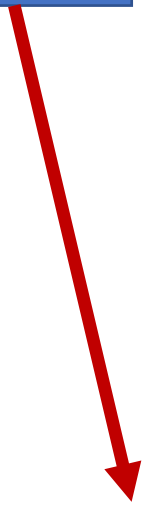
Definición de las medidas

N = 10

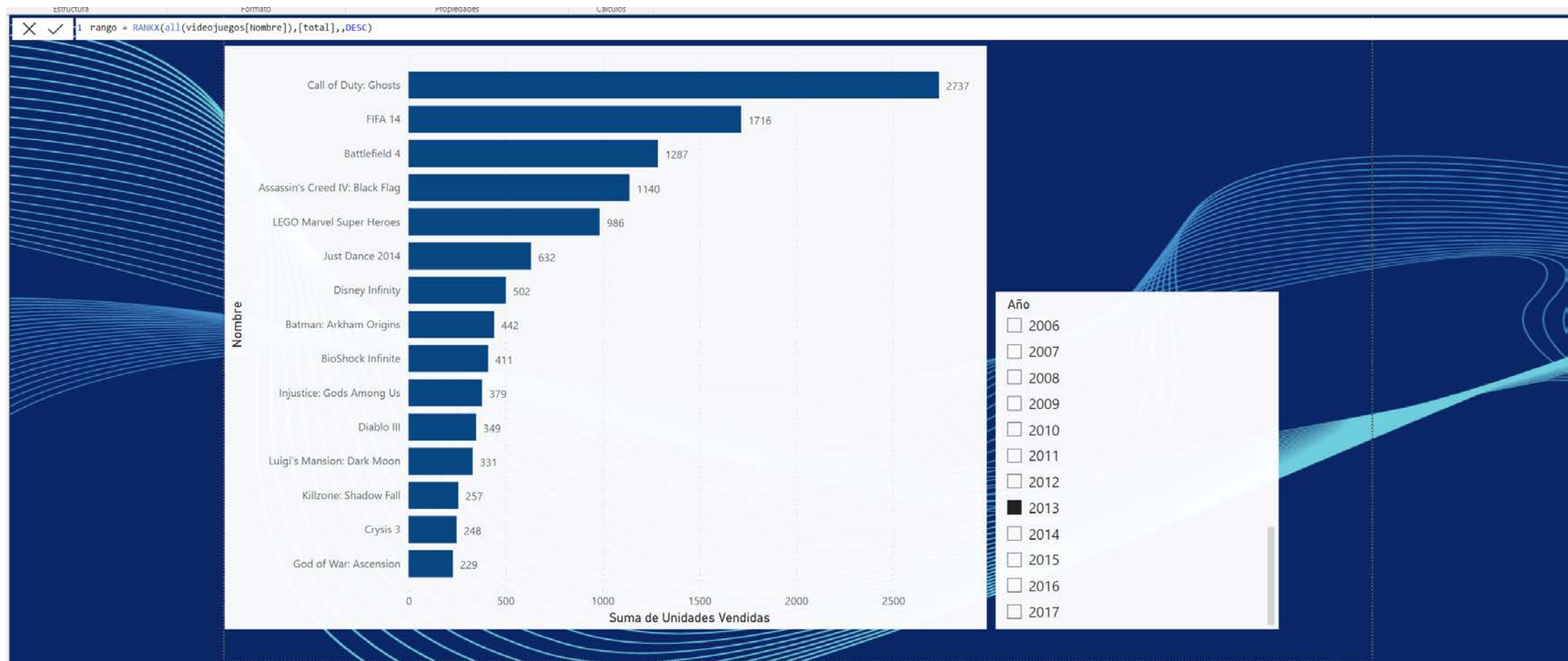
total = sum(videojuegos[Unidades Vendidas])

rango = RANKX(all(videojuegos[Nombre]),[total],,DESC)

incluido = IF([rango]<=[N],1,0)



Ejemplo V2: resultado



Combinar tablas

Table 1

Table 2

Combinación de tablas

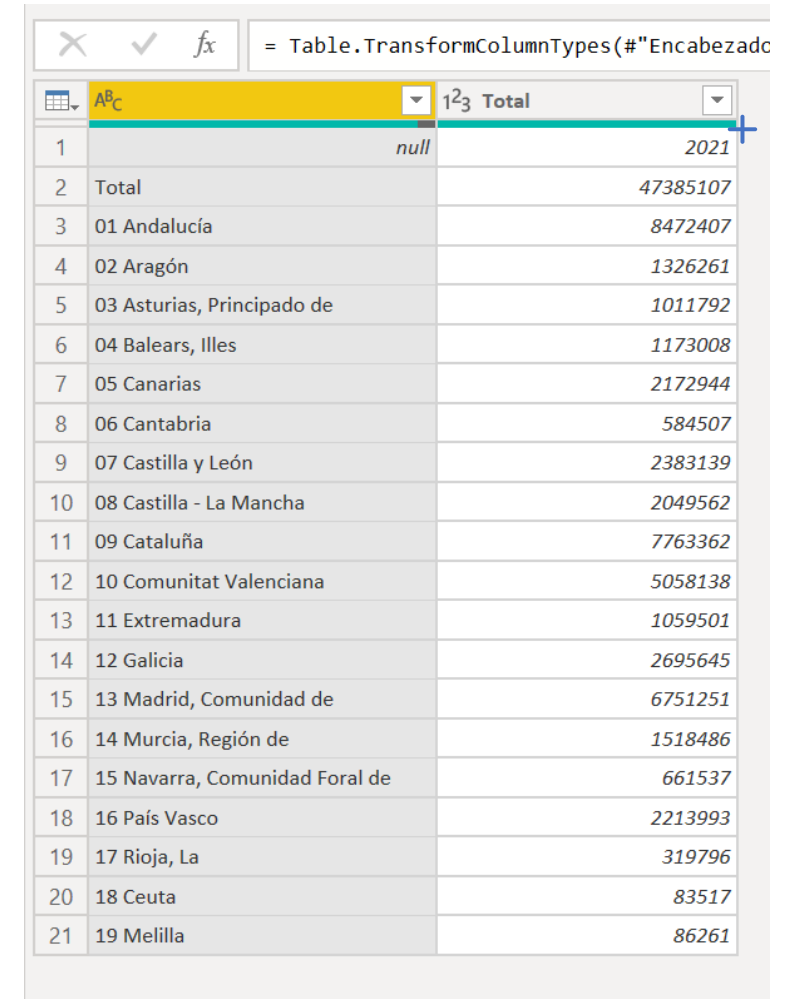
- En algunos problemas queremos combinar dos tablas en una sola
- Esto no siempre tiene sentido, a menudo es mejor mantener las dos tablas separadas
- Ejemplo: partimos de [este fichero](#) con número y tipo de pensiones por comunidad autónoma
- Lo cargamos y vamos a “Inicio” + “transformar datos”

ANDALUCÍA	JUBILACIÓN	939516
ARAGÓN	JUBILACIÓN	201574
Asturias, Principado de	JUBILACIÓN	183302
Balears, Illes	JUBILACIÓN	133044
CANARIAS	JUBILACIÓN	197285
CANTABRIA	JUBILACIÓN	89638
CASTILLA - LA MANCHA	JUBILACIÓN	224788
CASTILLA Y LEÓN	JUBILACIÓN	396392
CATALUÑA	JUBILACIÓN	1155755
COMUNITAT VALENCIANA	JUBILACIÓN	640987
EXTREMADURA	JUBILACIÓN	133997
GALICIA	JUBILACIÓN	482310
Madrid, Comunidad de	JUBILACIÓN	814590
Murcia, Región de	JUBILACIÓN	149421
Navarra, Comunidad Foral de	JUBILACIÓN	96452
PAÍS VASCO	JUBILACIÓN	376035
Rioja, La	JUBILACIÓN	48970
Ceuta	JUBILACIÓN	4458
Melilla	JUBILACIÓN	4031
ANDALUCÍA	VIUDEDAD	392378
ARAGÓN	VIUDEDAD	73625
Asturias, Principado de	VIUDEDAD	78813
Balears, Illes	VIUDEDAD	44980
CANARIAS	VIUDEDAD	82268

Combinación de tablas

- Para valorar si el número de pensionistas es más o menos elevado deberíamos conocer cuántos habitantes tiene la comunidad autónoma
- Desde dentro de la transformación de datos cargamos los datos del INE para población de comunides

<https://www.ine.es/jaxiT3/Datos.htm?t=2853>



	AB_C	123 Total
1	null	2021
2	Total	47385107
3	01 Andalucía	8472407
4	02 Aragón	1326261
5	03 Asturias, Principado de	1011792
6	04 Balears, Illes	1173008
7	05 Canarias	2172944
8	06 Cantabria	584507
9	07 Castilla y León	2383139
10	08 Castilla - La Mancha	2049562
11	09 Cataluña	7763362
12	10 Comunitat Valenciana	5058138
13	11 Extremadura	1059501
14	12 Galicia	2695645
15	13 Madrid, Comunidad de	6751251
16	14 Murcia, Región de	1518486
17	15 Navarra, Comunidad Foral de	661537
18	16 País Vasco	2213993
19	17 Rioja, La	319796
20	18 Ceuta	83517
21	19 Melilla	86261

Combinación de tablas

Algunos cambios rápidos

- Renombrar la tabla y las cabeceras
- Quitar las dos primeras filas
- Quitar los dos primeros espacios
- En ambos casos quitar blancos por el principio y pasar a mayúsculas

	AB ^C Comunidad	1 ² ₃ Habitantes
1	Andalucía	8472407
2	Aragón	1326261
3	Asturias, Principado De	1011792
4	Balears, Illes	1173008
5	Canarias	2172944
6	Cantabria	584507
7	Castilla Y León	2383139
8	Castilla - La Mancha	2049562
9	Cataluña	7763362
10	Comunitat Valenciana	5058138
11	Extremadura	1059501
12	Galicia	2695645
13	Madrid, Comunidad De	6751251
14	Murcia, Región De	1518486
15	Navarra, Comunidad Foral De	661537
16	País Vasco	2213993
17	Rioja, La	319796
18	Ceuta	83517
19	Melilla	86261

Combinación de tablas

- Dentro de Power Query vamos a **Inicio+combinar**
- Nos preguntará por la columna que queremos usar para combinar en ambas tablas
- Nos puede poner algún mensaje de seguridad ya que estamos mezclando orígenes diferentes (damos aceptar)
- Debemos elegir el tipo de combinación

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)

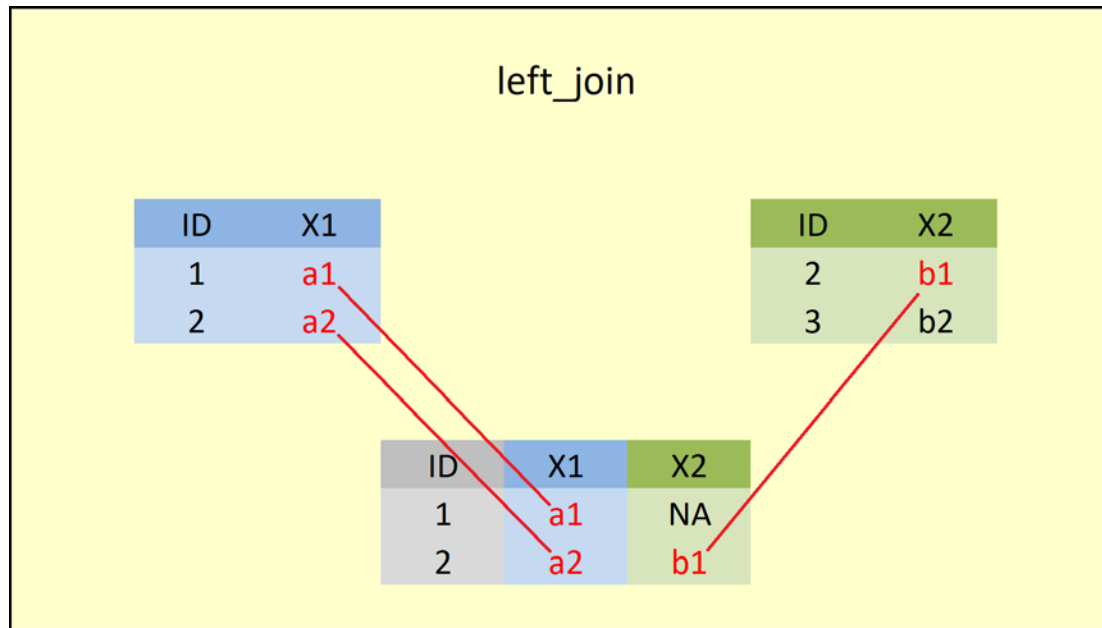
Interna (todas las filas coincidentes)

Anti izquierda (solo filas de la primera)

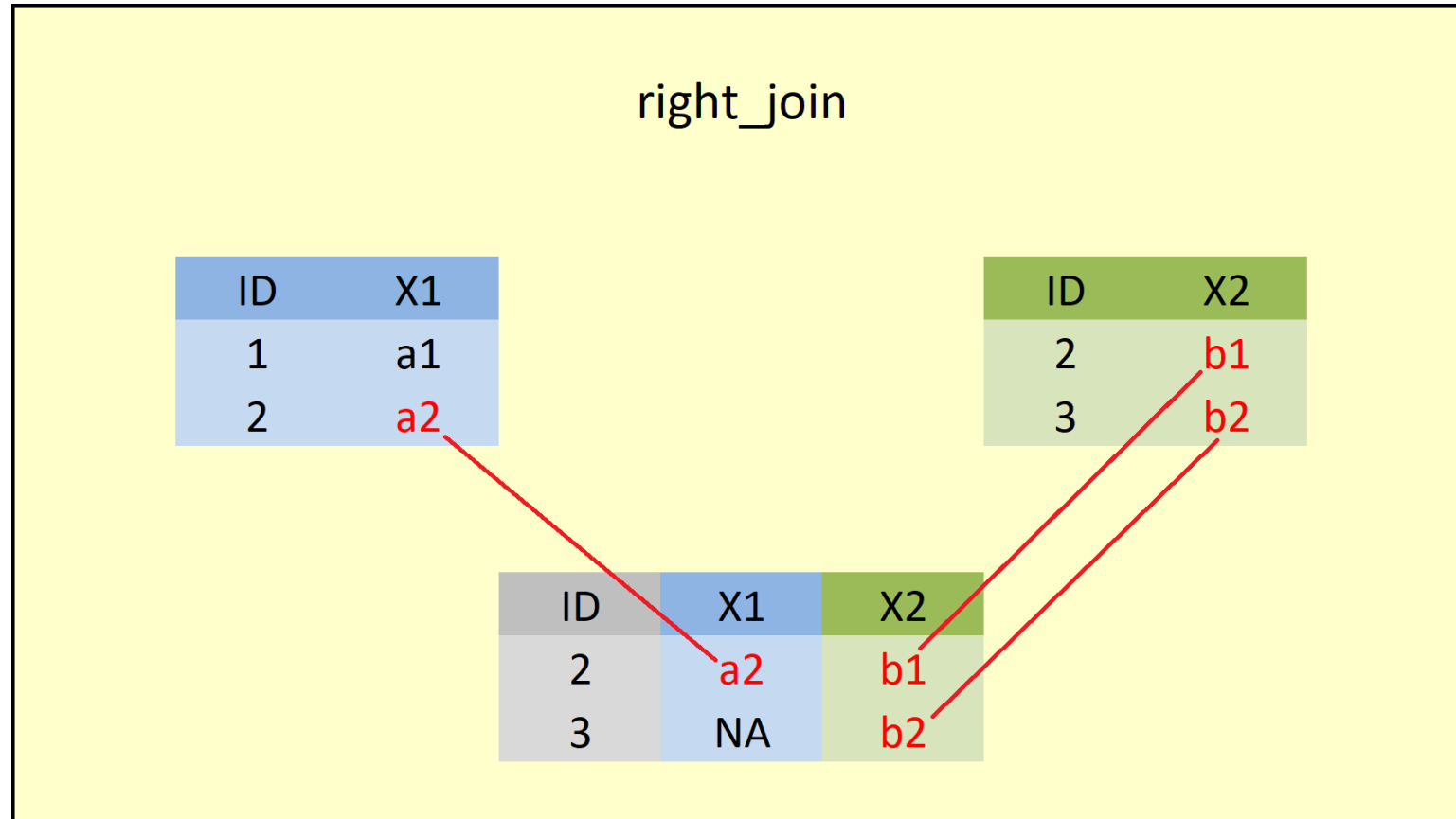
Anti derecha (solo filas de la segunda)

Externa izquierda

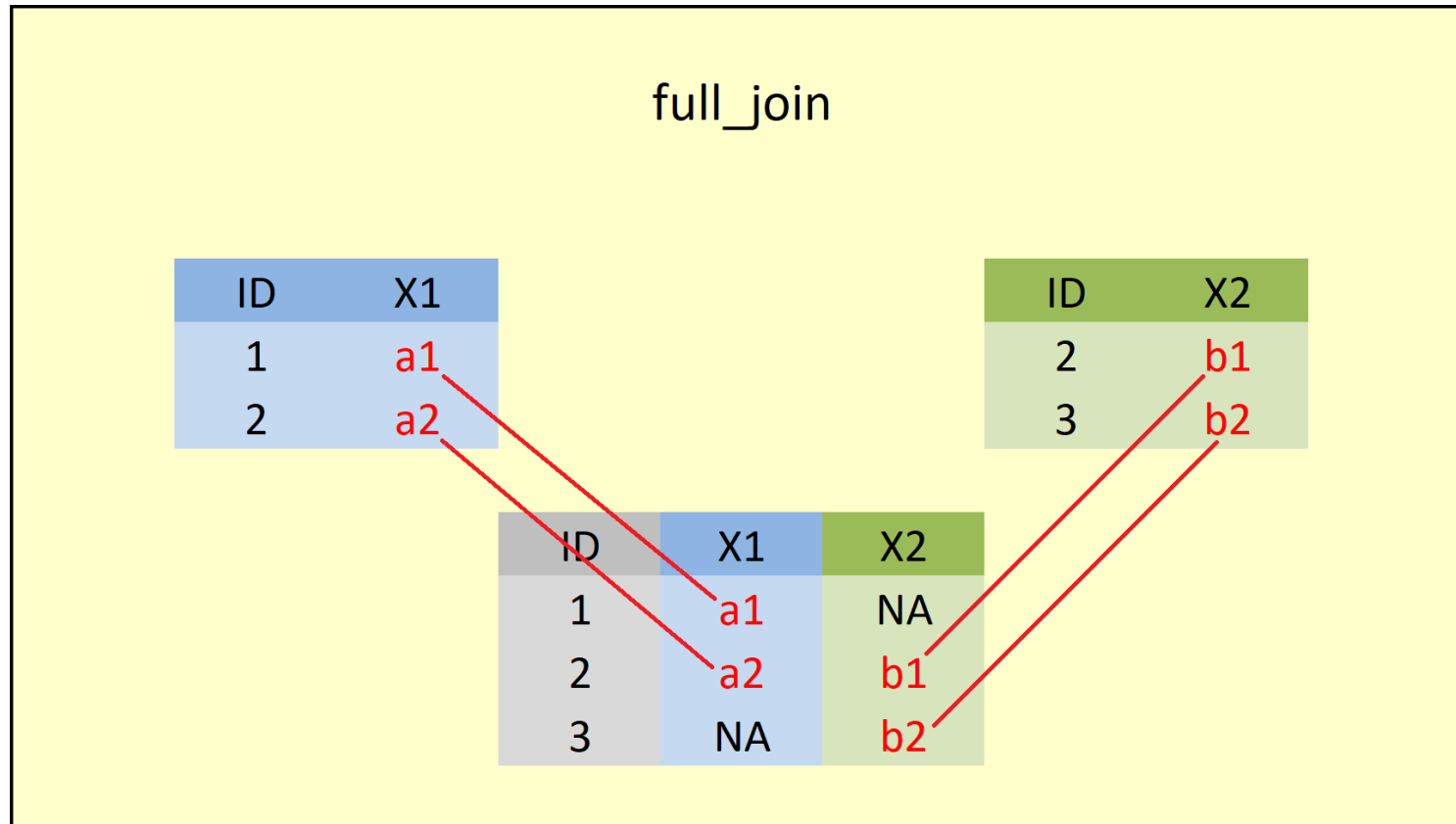
- No se pierde ninguna fila de la izq.



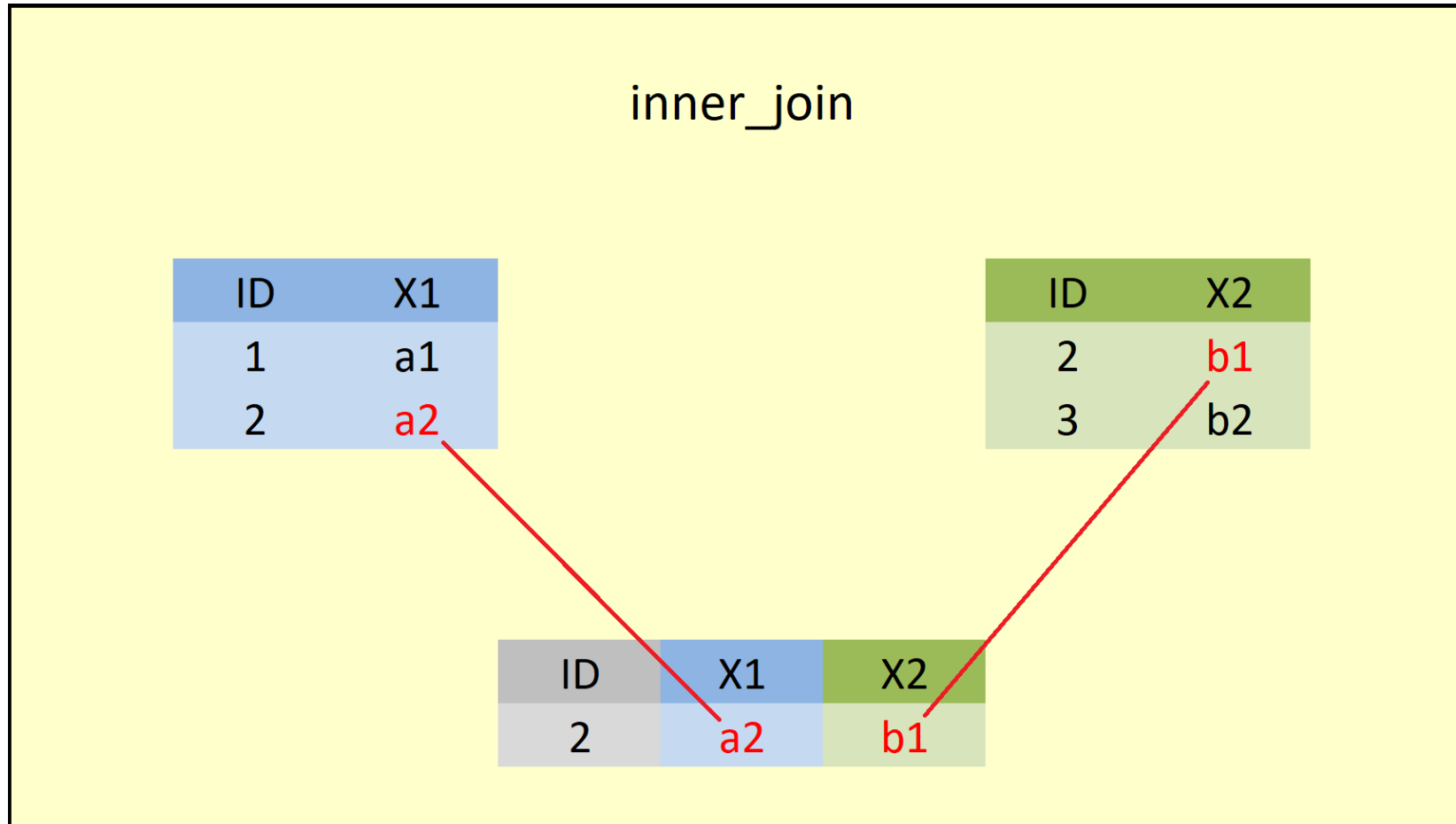
Externa derecha



Externa completa

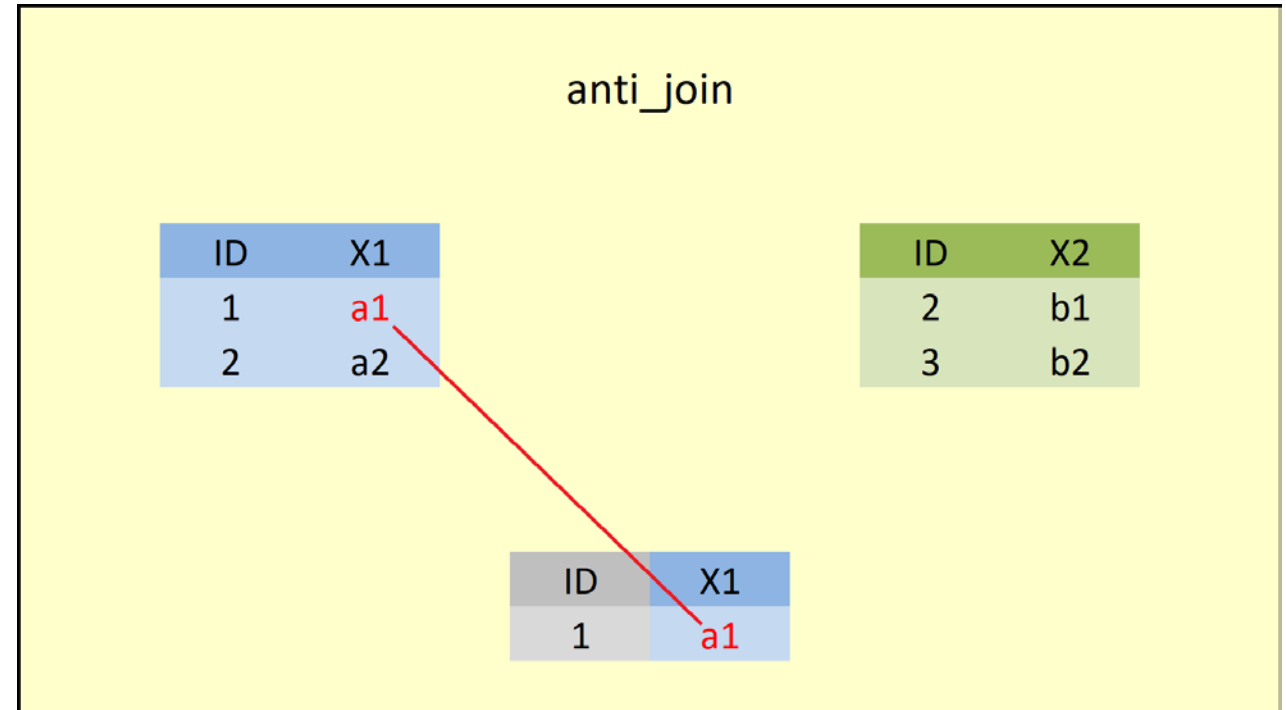


Inner Join



Anti izquierda (análogo antiderecha)

La segunda tabla se usa para
quitar filas de la primera, de
forma que solo queda la
primera reducida a las filas que
no cuadran con la segunda



Combinación

- En nuestro caso se trata de una externa izquierda
- Lo hacemos, expandimos el resultado y vemos que hay algunos nulos, parece que el caso de Navarra no ha encajado
- Deshacemos la combinación
- Como no vemos lo que está mal probamos en “Transformar+Formato” a dar recortar y limpiar para ambas columnas
- Ahora sí, para terminar podemos
 - Eliminar la columna Comunidad duplicada
 - Renombrar habitantes

Tras la combinación

A ^B _C Comunidad	A ^B _C tipo	1 ² ₃ Pensionistas	1 ² ₃ habitantes
Andalucía	JUBILACIÓN	939516	8472407
Andalucía	VIUDEDAD	392378	8472407
Aragón	JUBILACIÓN	201574	1326261
Asturias, Principado De	JUBILACIÓN	183302	1011792
Balears, Illes	JUBILACIÓN	133044	1173008
Canarias	JUBILACIÓN	197285	2172944
Cantabria	JUBILACIÓN	89638	584507
Castilla - La Mancha	JUBILACIÓN	224788	2049562
Castilla Y León	JUBILACIÓN	396392	2383139
Cataluña	JUBILACIÓN	1155755	7763362
Comunitat Valenciana	JUBILACIÓN	640987	5058138
Extremadura	JUBILACIÓN	133997	1059501
Galicia	JUBILACIÓN	482310	2695645
Madrid, Comunidad De	JUBILACIÓN	814590	6751251
Murcia, Región De	JUBILACIÓN	149421	1518486
Navarra, Comunidad Foral De	JUBILACIÓN	96452	661537
País Vasco	JUBILACIÓN	376035	2213993
Rioja, La	JUBILACIÓN	48970	319796
Ceuta	JUBILACIÓN	4458	83517
Melilla	JUBILACIÓN	4031	86261
Aragón	VIUDEDAD	73625	1326261
Asturias, Principado De	VIUDEDAD	78813	1011792
Balears, Illes	VIUDEDAD	44980	1173008
Canarias	VIUDEDAD	82268	2172944
Cantabria	VIUDEDAD	35144	584507

Creación de una nueva columna

- Añadir columna personalizada
- Obsérvese que estamos en M, no en DAX

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna

Fórmula de columna personalizada ⓘ

= [Pensionistas] / [habitantes]

Columnas disponibles

Comunidad
tipo
Pensionistas
habitantes

<< Insertar

[Información sobre fórmulas de Power Query](#)

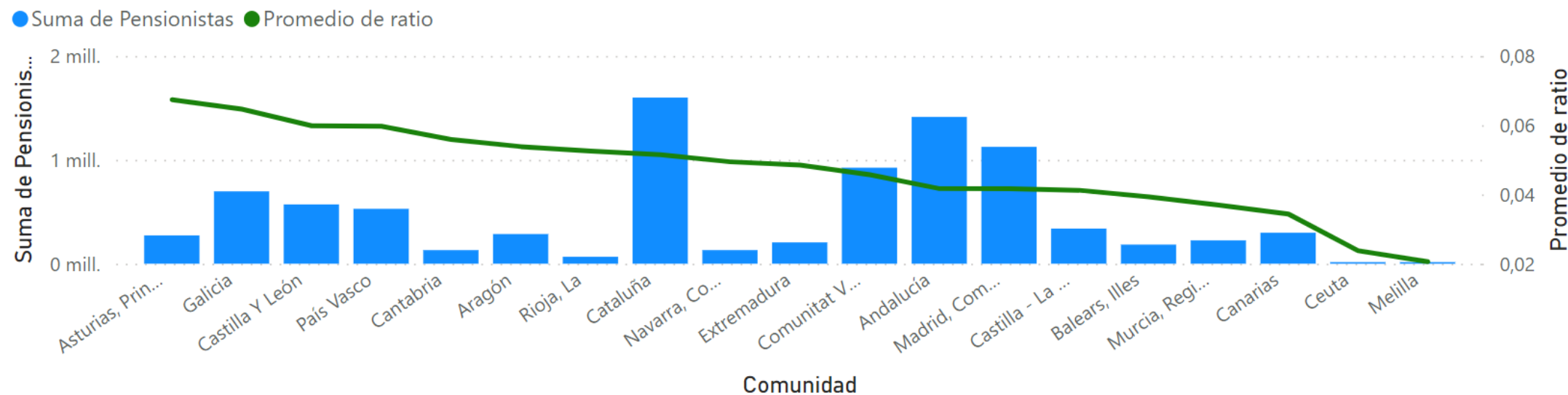
✓ No se han detectado errores de sintaxis.

Aceptar

Cancelar

Ejercicio: gráfica con ratio por habitante

Total de Pensionistas y ratio por habitante por Comunidad



Ejercicio/Ejemplo

Descargamos el fichero

<https://github.com/RafaelCaballero/tdm/blob/master/powerbi/facturas.xlsx?raw=true>

Contiene datos de empresas y de facturas

Empresa	Dirección	CIF	Provincia
Girasoles Solete	Rosal 5	Q55555	Sevilla
Aceites Olé	Jazmín 4	Q4444	Murcia
Envasadora Paketo	Clavel 33	Q33	Barcelona
Piensos Rúmíame	Petunia 8	Q8888	Toledo
Semillas Sidi	Amapola s/n	Q0000001	Córdoba
Invernaderos Invernalía	Cactus 27	Q272727	Soria
Vaquería La Leche	Margarita 1	Q1111111	Santander
Prestamos Arrieritos Somos	Desesperanza 66	Q6666	Madrid

CIF	Empresa Facturada	Total
Q55555	Q33	111,11
Q55555	Q4444	273,31
Q55555	Q8888	165,9
Q55555		1155
Q4444	Q33	52,9
Q33	Q55555	987,65
Q8888	Q33	213,87
Q0000001		511,11
Q272727	Q25255	1121
Q1111111	Q33	879,01
Q25255		899
Q55555		761,3
Q4444	Q55555	231
Q33	Q25255	578
Q8888	Q33	987,11
Q0000001	Q33	1999,99
Q272727	Q4444	2391
Q1111111	Q67890111	876,65
Q25252	Q33	1567
Q272727	Q25255	1121
Q1111111	Q33	879,01

Ejercicio/Ejemplo

Queremos conocer los datos de las empresas que no facturan a nadie

	ABC Empresa	ABC Dirección	ABC CIF	ABC Provincia
1	Prestamos Arrieritos Somos	Desesperanza 66	Q6666	Madrid