
BIG DATA – POWER BI

PROYECTO 2

EDUARD LARA

INDICE

1. Proyecto 2
2. Solución Proyecto 2

1. PROYECTO

- Nos vamos a poner en el papel de analistas de una importante empresa de producción de conservas de frutas que está en pleno desarrollo de expansión
- Nuestro trabajo consiste en explorar distintos orígenes de datos que nos han proporcionado con información valiosa sobre la actividad de la empresa en distintas ubicaciones geográficas
- Esta empresa se encuentra en la provincia de Río Negro, Argentina y desde allí distribuye sus deliciosas conservas de frutas a todo el país.
- No obstante, han vivido un crecimiento exponencial reciente, lo cual, lejos de ser un problema, presenta un nuevo desafío.

1. PROYECTO

- El crecimiento ha traído consigo un incremento en los costos de distribución y en ciertas dificultades para satisfacer la demanda de manera oportuna y eficiente en las distintas provincias.
- Por este motivo se han propuesto la creación de un centro de distribución que centralice las actividades logísticas en una sola región única y estratégica del país
- La ubicación actual de su fábrica hace que todos los puntos de consumo queden bastante lejos.
- Nuestro trabajo es ayudarles a tomar una decisión basada en la información brindada
- Combinando las tablas de ventas con las tablas de referencia de provincias y de productos y analizarlas de manera integral.

1. PROYECTO

Preguntas fundamentadas

1) ¿donde deberían ubicar su nuevo centro de distribución para optimizar sus operaciones?

Esta decisión no debería basarse simplemente en ubicarla en un lugar del centro del país, sino en los lugares donde se concentren los mayores niveles de consumo.

2) ¿cuál es el producto que concentra las mayores ventas en cantidad de unidades

3) ¿Cuál es el mes en el cual se produce la mayor cantidad de ventas?

2. SOLUCION PROYECTO

Paso 1. Comenzamos viendo los datos que tenemos. El archivo ventas.CSV contiene información sobre las ventas: código de la provincia, el mes de la venta, la cantidad de productos que se vendieron en cada uno de esos meses por provincia y luego los códigos de los productos, solamente el código.

A	B	C	D
1	Cod_Prov	Mes	Cantidad de Producto
2	BUE	ene-23	6782 MF001
3	BUE	feb-23	11080 MF001
4	BUE	mar-23	17994 MF001
5	BUE	abr-23	34340 MF001
6	BUE	may-23	28825 MF001
7	BUE	jun-23	48054 MF001
8	BUE	jul-23	86016 MF001
9	BUE	ago-23	103320 MF001
10	BUE	sep-23	133479 MF001
11	BUE	oct-23	120910 MF001
12	BUE	nov-23	57486 MF001
13	BUE	dic-23	94788 MF001
14	CAT	ene-23	666 MF001
15	CAT	feb-23	1442 MF001
16	CAT	mar-23	4059 MF001

2. SOLUCION PROYECTO

Paso 2. En el archivo Excel tenemos 3 hojas:

- Datos geográficos. Tenemos las provincias, con su código, nombre y capital. Hay 2 líneas a eliminar
- Tabla capitales, con las capitales con sus ubicaciones
- Productos, asocia código de producto con la descripción, la forma y el volumen de cada presentación. Hay una columna a eliminar

Cod_Pro	Provincia	Capital_Prov
Datos elaborados por		
FRUTACOR SA		
BUE	Buenos Aires	La Plata
CAT	Catamarca	San Fernando del Valle de Catamarc
CHA	Chaco	Resistencia
CHU	Chubut	Rawson

Capital	Latitud	Longitud
La Plata	-34,9207	-57,9533
San Fernando del Valle de Catamarc	-28,4696	-65,7852
Resistencia	-27,4512	-58,9869
Rawson	-43,2992	-65,1046
Córdoba	-31,4201	-64,1888
Corrientes	-27,4692	-58,8309
Paraná	-31,7413	-60,5116

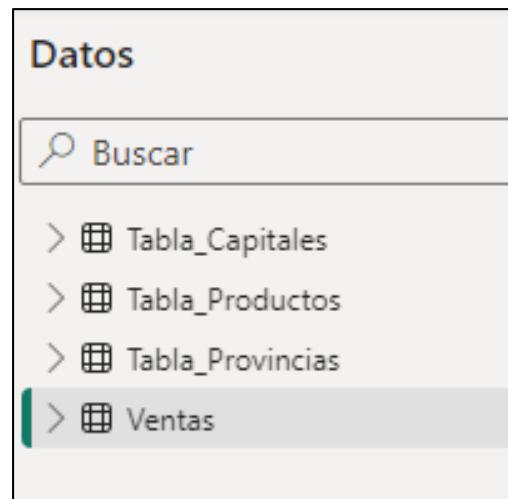
Código Produc	Descripción Producto	Presentació	Volumen (cc)	Columna
MF001	Mermelada de Frambuesa	Frasco	500	
MN002	Mermelada de Naranja	Frasco	500	
MD005	Mermelada de Durazno	Frasco	500	
CF001	Conserva Frambuesa	Lata	680	
CM002	Conserva Manzana	Lata	680	
CM059	Conserva Peras	Lata	1500	

2. SOLUCION PROYECTO

Paso 3. Importamos los ficheros en PowerBI.

Del libro de Excel Tablas auxiliares, vamos a trabajar con las 3 tablas existentes: Productos, Provincias, Capitales

Del archivo ventas.CSV cargamos su tabla, lo que al final supone 4 tablas para trabajar.



2. SOLUCION PROYECTO

Paso 4. Vamos a revisar las tablas una por una y a hacer limpieza de los datos en caso necesario, en la vista tabla.

Tabla_Capitales

- Capital: Verificamos que sea de tipo texto y la categorizamos como ciudad.
- Latitud: Verificamos que el tipo de dato sea número decimal y que la categoría sea latitud
- Longitud: Verificamos que sea número decimal y longitud

Capital	Latitud	Longitud
La Plata	-34,9207	-57,9533
San Fernando del Valle de Catamarca	-28,4696	-65,7852
Resistencia	-27,4512	-58,9869
Rawson	-43,2992	-65,1046
Córdoba	-31,4201	-64,1888
Corrientes	-27,4692	-58,8309

2. SOLUCION PROYECTO

Paso 5. Tabla_productos

- Eliminamos la columna 1, la seleccionamos, hacemos clic derecho y Eliminar. No es necesario ir a Power Query
- Verificamos que código de producto, descripción y presentación sea de tipo texto
- Verificamos que volumen sea un número entero y en sumatorio resumen indicamos no resumir. No corresponde que nos haga una suma de los cm3

Código Producto	Descripción Producto	Presentación	Volumen (cc)
MF001	Mermelada de Frambuesa	Frasco	500
MN002	Mermelada de Naranja	Frasco	500
MD005	Mermelada de Durazno	Frasco	500
CF001	Conserva Frambuesa	Lata	680
CM002	Conserva Manzana	Lata	680
CM059	Conserva Peras	Lata	1500

2. SOLUCION PROYECTO

Paso 6. Tabla_provincias

- Tenemos dos primeras filas en blanco. Las eliminaremos yendo al editor de Power Query. Abrimos el filtro de cualquier columna donde estén los nulos, los deseleccionamos y hacemos clic en Cerrar y Aplicar
- Verificamos que Cod_Prov sea de texto.
- Verificamos que Provincia sea de texto, y su categoría de datos corresponda a Estado o provincia
- Verificamos que Capital_Prov sea de texto y su categoría de ciudad

Cod_Prov	Provincia	Capital_Prov
BUE	Buenos Aires	La Plata
CAT	Catamarca	San Fernando del Valle de Catamarca
CHA	Chaco	Resistencia
CHU	Chubut	Rawson
COR	Córdoba	Córdoba

2. SOLUCION PROYECTO

Paso 7. Tabla_ventas

- Verificamos que Cod_Prov se de tipo texto
- Verificamos que Mes sea de tipo fecha, con formato de mostrar de solamente año y mes
- Verificamos que Cantidad de unidades sea un n° entero
- Verificamos que Producto sea texto.

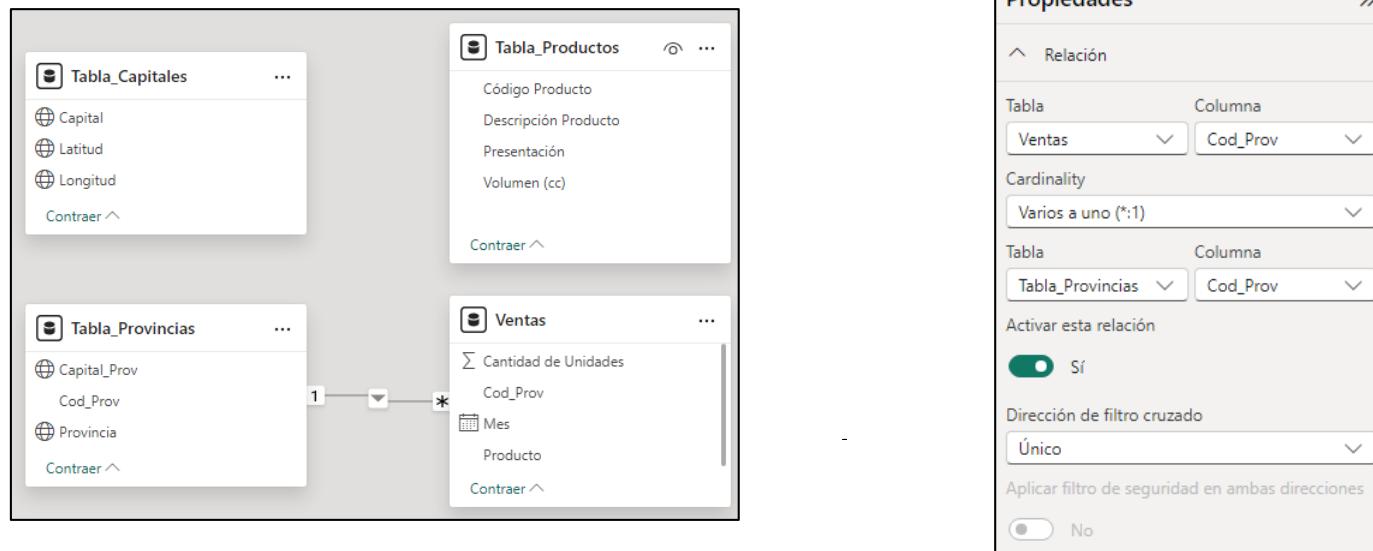
Cod_Prov	Mes	Cantidad de Unidades	Producto
BUE	2024-01	6782	MF001
BUE	2024-02	11080	MF001
BUE	2024-03	17994	MF001
BUE	2024-04	34340	MF001

Esta es parte de limpieza y modelación de los datos para asegurarnos de que estamos en condiciones de empezar a trabajar con ellos.

2. SOLUCION PROYECTO

Paso 8. Conectaremos nuestras tablas. Si vamos a vista de modelo, vemos que Power BI automáticamente estableció una relación entre las columnas Cod_Prov de las tablas Provincias y Ventas, de 1 a muchos.

Por cada vez que aparece el código de provincias en la tabla Provincias aparecen muchas veces estos códigos según las ventas que se hayan hecho en cada mes del año.



2. SOLUCION PROYECTO

Paso 9. Ahora estableceremos manualmente otras relaciones que Power BI no ha detectado. Estableceremos una relación entre las columnas con la misma información:

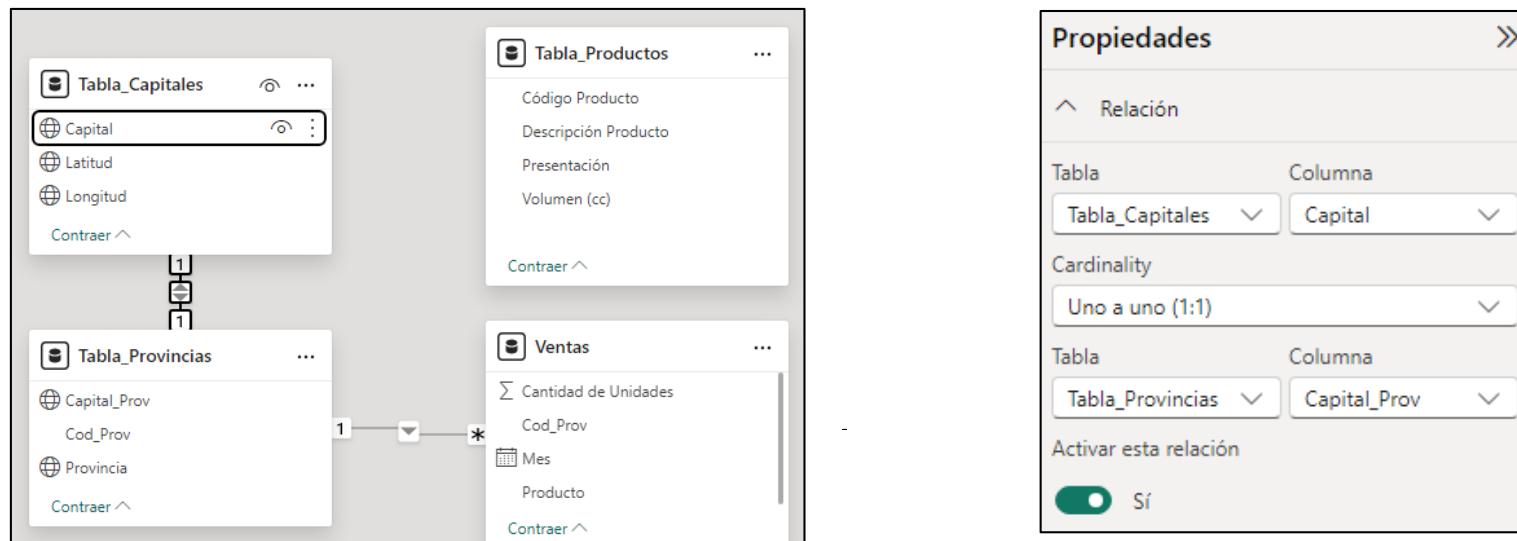
Tabla_Provincias

Columna Capital_Prov

Tabla_Capitales

Columna Capital

Arrastramos una columna contra la otra y establecemos automáticamente una relación 1 a 1

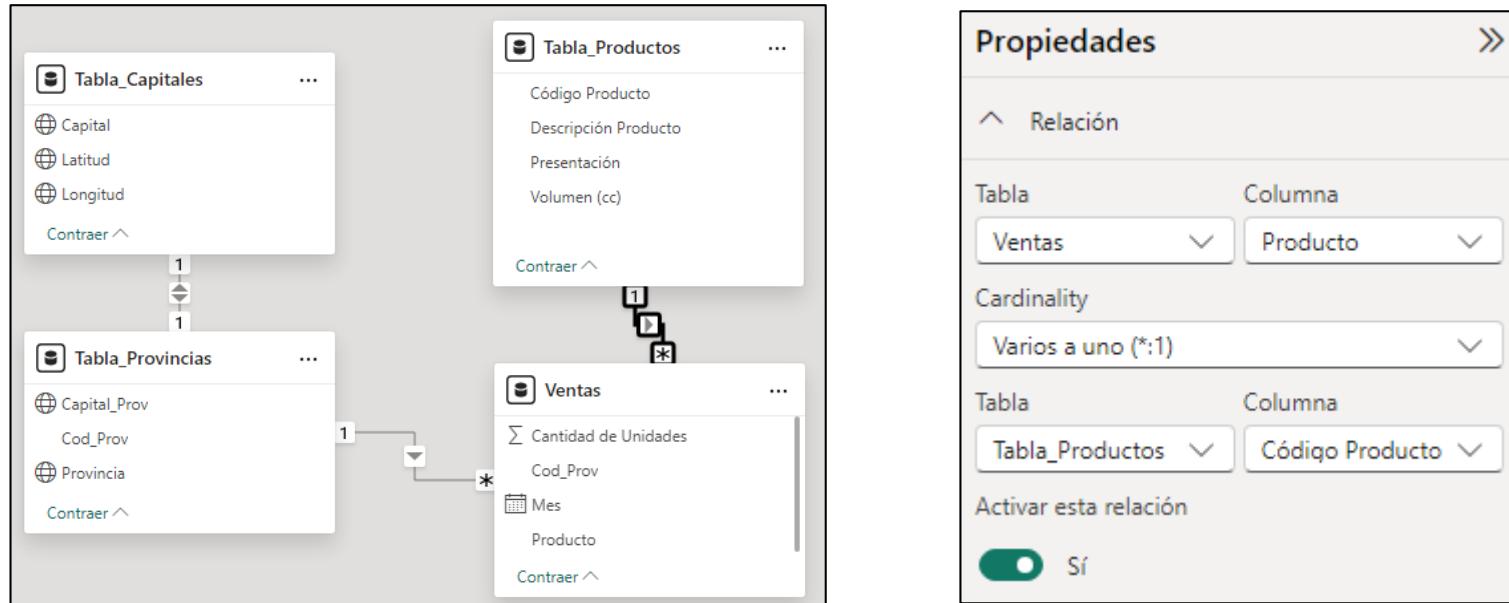


2. SOLUCION PROYECTO

Paso 10. Otra relación que podemos establecer:

Tabla de productos tabla de ventas
Código_Producto Producto

Arrastramos una columna contra la otra y se establece una relación 1 a muchos. Ahora ya tenemos todas las tablas vinculadas entre si



2. SOLUCION PROYECTO

Paso 11. Vamos a la vista de informes y respondemos la primera pregunta: **En qué lugar del país ubicar una planta de logística o distribución que se relacione con cómo están distribuidas las ventas en el país?**

Creamos un mapa, donde dentro de Ubicación colocamos la columna Provincia de la Tabla de provincias. En Latitud y Longitud arrastramos las columnas de tabla_capitales



Ubicación

Provincia

Leyenda

Agregar campos de datos aquí

Latitud

Promedio de Latitud

Longitud

Promedio de Longitud

This screenshot shows the "Ubicación" (Location) section of a Power BI report. It displays a dropdown menu for "Provincia" and two data fields: "Promedio de Latitud" and "Promedio de Longitud". There is also a "Leyenda" (Legend) section with a placeholder "Agregar campos de datos aquí" (Add data fields here).

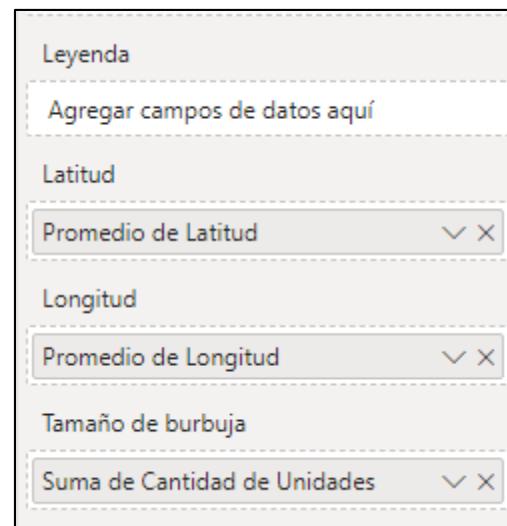
2. SOLUCION PROYECTO

Paso 12. Damos formato a las burbujas. Vamos a formato a su objeto visual y las haremos un poco más grandes y de color rojo.

The screenshot shows the Microsoft Power BI interface. On the left, there is a 'Format' pane with a title 'Dar formato a su objeto visual' and a subtitle 'Combinar visu...'. It includes icons for 'Formato de datos' (green), 'Formato de visual' (blue), and 'Formato de tabla' (orange). Below this is a map of South America (Chile, Argentina, Uruguay, Paraguay) with several red bubbles of varying sizes scattered across the continent, primarily concentrated in Argentina and Paraguay. To the right of the map is the 'Format' pane, which has tabs for 'Buscar', 'Objeto visual' (selected), and 'General'. Under the 'Objeto visual' tab, there are sections for 'Configuración del mapa', 'Leyenda' (disabled), and 'Burbujas'. The 'Burbujas' section is expanded, showing 'Tamaño' (Size) settings with a dropdown set to 'Automático' (Automatic) and a slider set to 17, and 'Colores' (Colors) settings with a dropdown set to 'Valor predeterminado' (Default value) and a color swatch set to red.

2. SOLUCION PROYECTO

Paso 13. Volvemos a la información del grafico. Vemos donde están distribuidas las ventas pero no cuáles son los que tienen más. Arrastramos al tamaño de la burbuja el campo cantidad de unidades vendidas de la tabla ventas. Nuestras burbujas cobran relevancia según la cantidad de ventas.



2. SOLUCION PROYECTO

- Si tenemos que aconsejar dónde colocar un centro de distribución, lo haríamos en el corredor central, entre las ciudades de Mendoza, Córdoba o lo más probable Buenos Aires.
- Desde ahí, podemos distribuir para el norte y para el sur.
- Esta sería la respuesta a la primera pregunta:
Ubicaremos el almacén principal en el eje central, en alguna de estas tres ubicaciones más grandes.

2. SOLUCION PROYECTO

Paso 14. La segunda pregunta a responder: **¿cuál es el producto más vendido a lo largo de todo el periodo de tiempo?** La resolveremos utilizando un gráfico que nos muestre cantidades en comparativa de los diferentes productos.

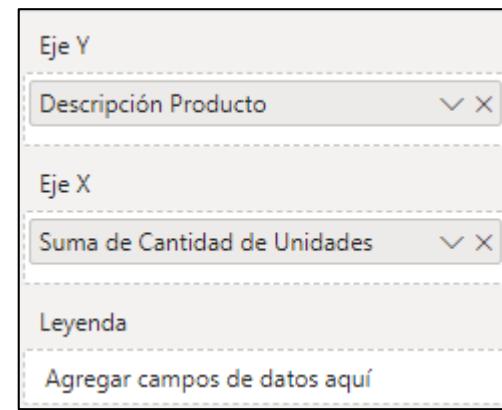
Lo podemos hacer con un gráfico de barras apiladas horizontales



2. SOLUCION PROYECTO

Paso 15. En el eje vertical Y pondremos los distintos productos, pero no con su código, sino con la descripción del producto que tengo en tabla productos.

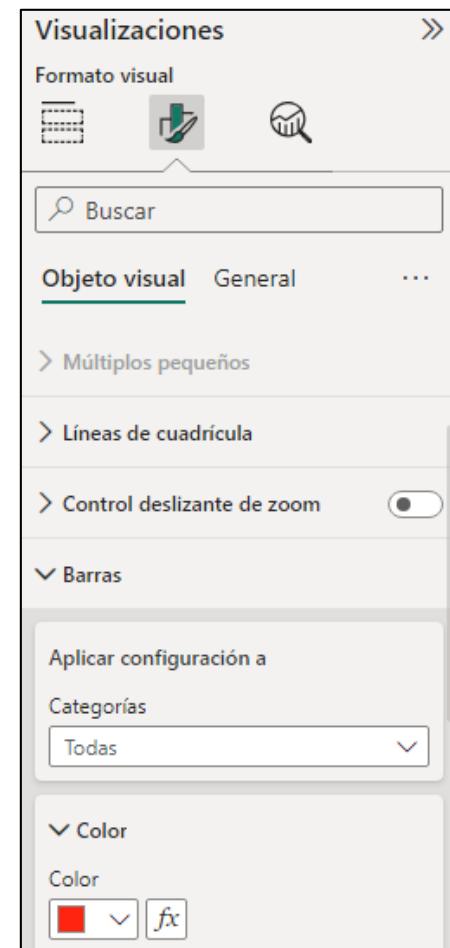
En el eje horizontal X, vamos a poner la cantidad de unidades vendidas que tenemos en ventas



El gráfico nos muestra que el producto que más se vendió fue mermelada de frambuesa → Respuesta a la pregunta²².

2. SOLUCION PROYECTO

Paso 16. Por una cuestión de estilo, vamos a hacer que las barras sean también de color rojo. Vamos a Dar formato a su objeto visual



2. SOLUCION PROYECTO

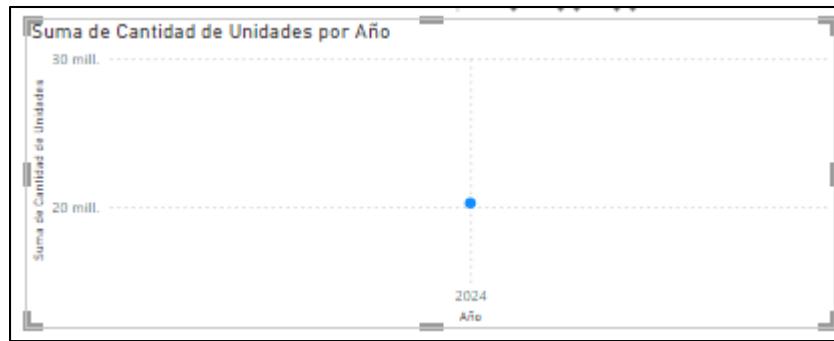
Paso 17. La última respuesta que queremos es: **¿cuál fue el mes de mayor ventas a lo largo del año?**

Necesitamos un gráfico que represente el paso del tiempo para poder ver una tendencia. Un gráfico de líneas sería una buena opción.

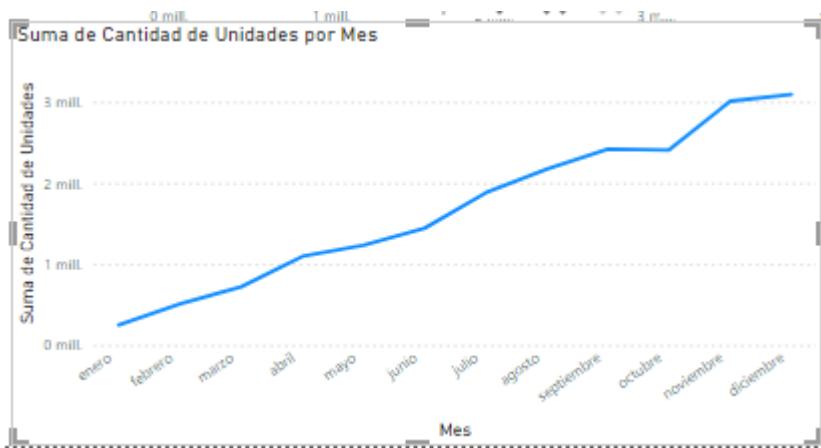


2. SOLUCION PROYECTO

Paso 18. En el eje de las X podremos los meses de Ventas. En el eje Y pondremos la cantidad de unidades vendidas



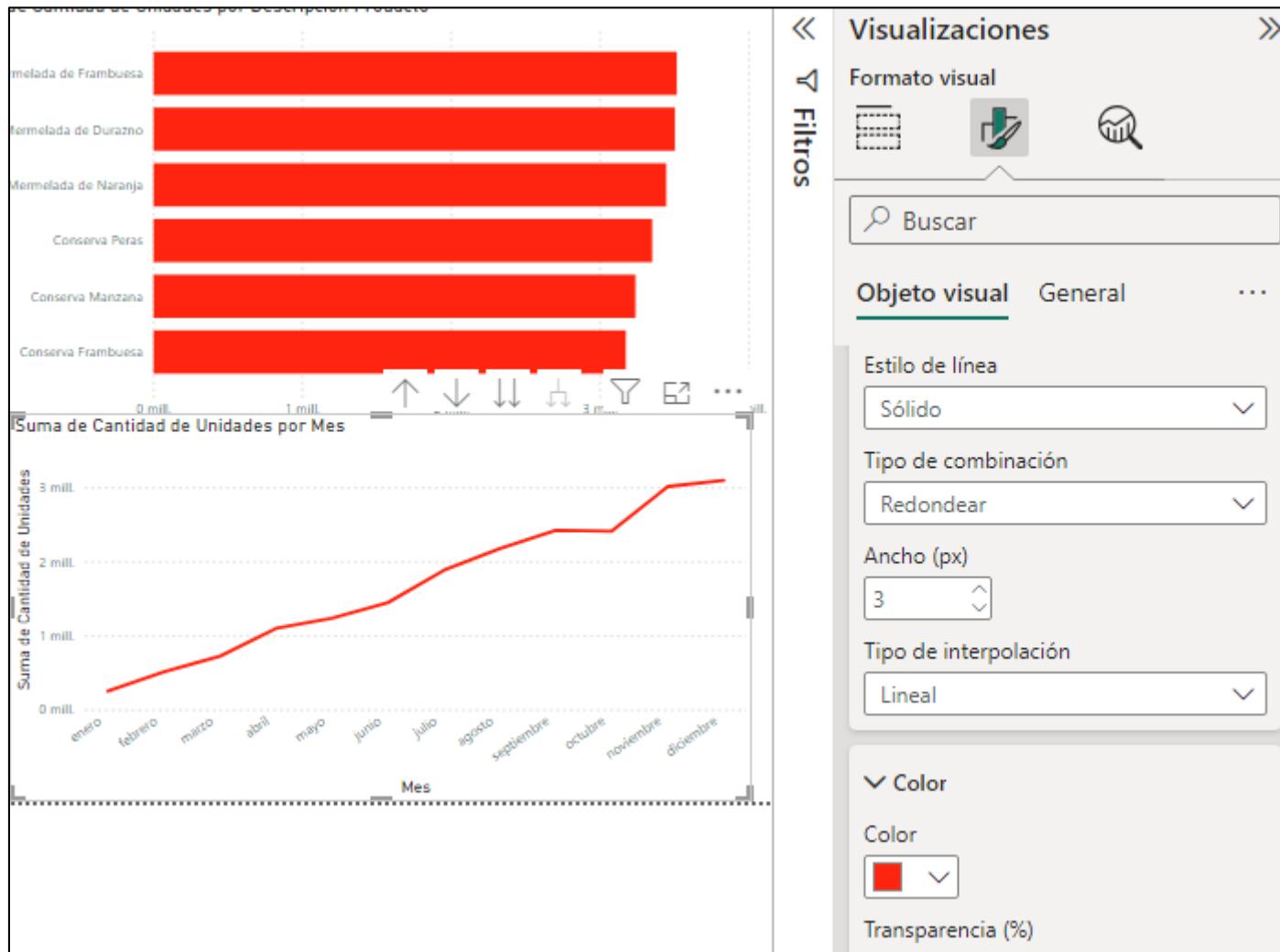
Vamos a ir bajando de jerarquías de fechas para pasar de año a trimestre y de trimestre a mes.



Podemos decir que el mes con más ventas ha sido diciembre,

2. SOLUCION PROYECTO

Paso 19. Cambiamos el color de las líneas del grafico a rojo.



2. SOLUCION PROYECTO

Paso 20. Podemos cambiar los títulos de los gráficos, por ejemplo el del mapa:

Titulo: Cantidad de unidades por provincia



Visualizaciones

Formato visual

Buscar

Objeto visual General ...

> Propiedades

▼ Título

▼ Título

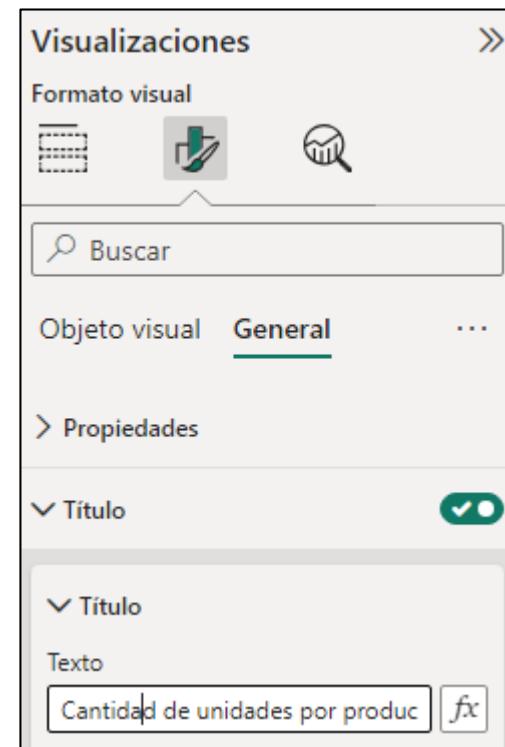
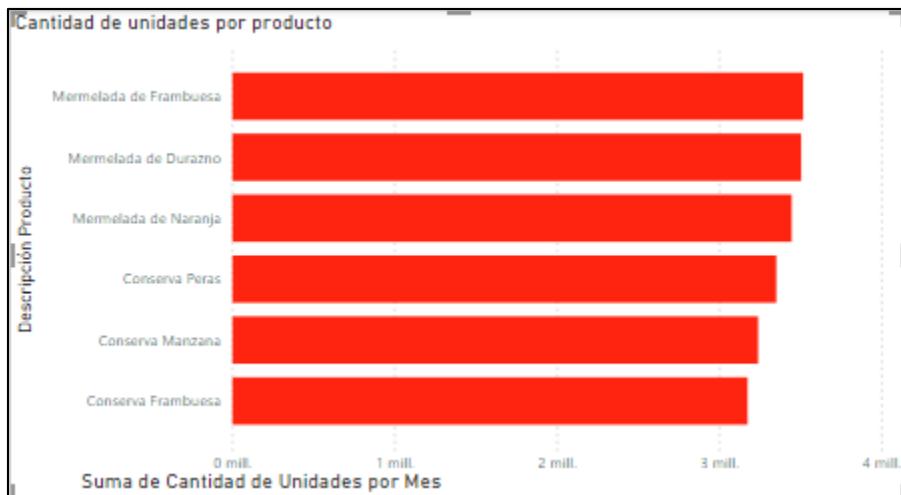
Texto fx

Título

2024 TomTom ©2024 Microsoft Corporation. [Acerca de Microsoft](#)

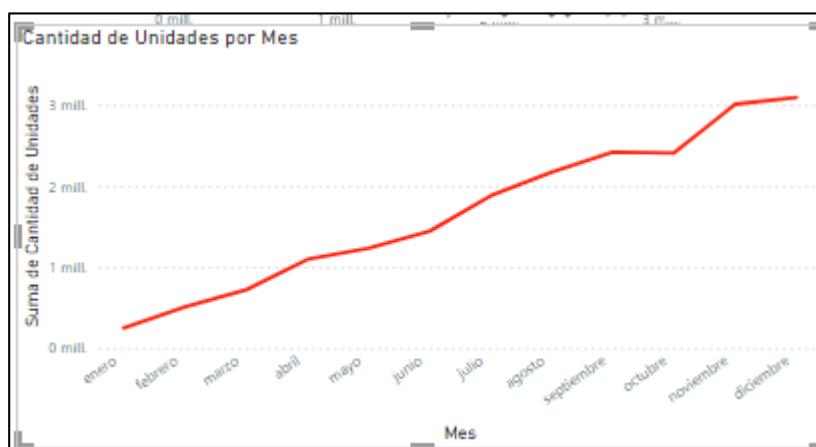
2. SOLUCION PROYECTO

Paso 21. El título del gráfico de las barras puede ser:
Titulo: Cantidad de unidades por producto



2. SOLUCION PROYECTO

Paso 22. El título del gráfico de las líneas puede ser:
Titulo: Cantidad de unidades por mes



The screenshot shows the "Visualizaciones" (Visualizations) pane in Power BI. The "Formato visual" (Visual Style) section includes icons for edit, search, and refresh. Below it is a search bar with the placeholder "Buscar". The "Objeto visual" (Visual Object) dropdown is set to "General". Under "Propiedades" (Properties), the "Título" (Title) section is expanded, showing the current title "Cantidad de Unidades por Mes" in the "Texto" (Text) input field. A green checkmark icon indicates the title is active.

2. SOLUCION PROYECTO

Conclusiones

- Es muy importante trabajar con la interacción entre las tablas.
- El hecho de haber permitido que las tablas se relacionen entre sí nos ha permitido mejorar el análisis de datos
- Nos permitió ver la descripción de los productos que no está en la tabla de ventas. No existe esa información en ventas
- Si no hubiéramos tenido relacionadas las tablas veríamos ceros en muchas graficas y esto dificulta el análisis
- Es importante inicialmente analizar la información y conectar las tablas adecuadamente.