UNIVERSIDAD DON BOSCO

Materia: Dataware y minería de Datos

Docente: Ing. Karens Medrano

Tema: Proyecto Catedra fase 2

Presentan: Grupo #1

Nombre	Carnet
Henry Alejandro Martínez Guerra	MG202016
Erick Samuel García Rivas	GR181865
Ángel Vladimir Del Cid Rivera	DR220494
José Manuel Figueroa Aguilar	FA200209

Fecha de entrega: 2 de junio de 2024

INDICE

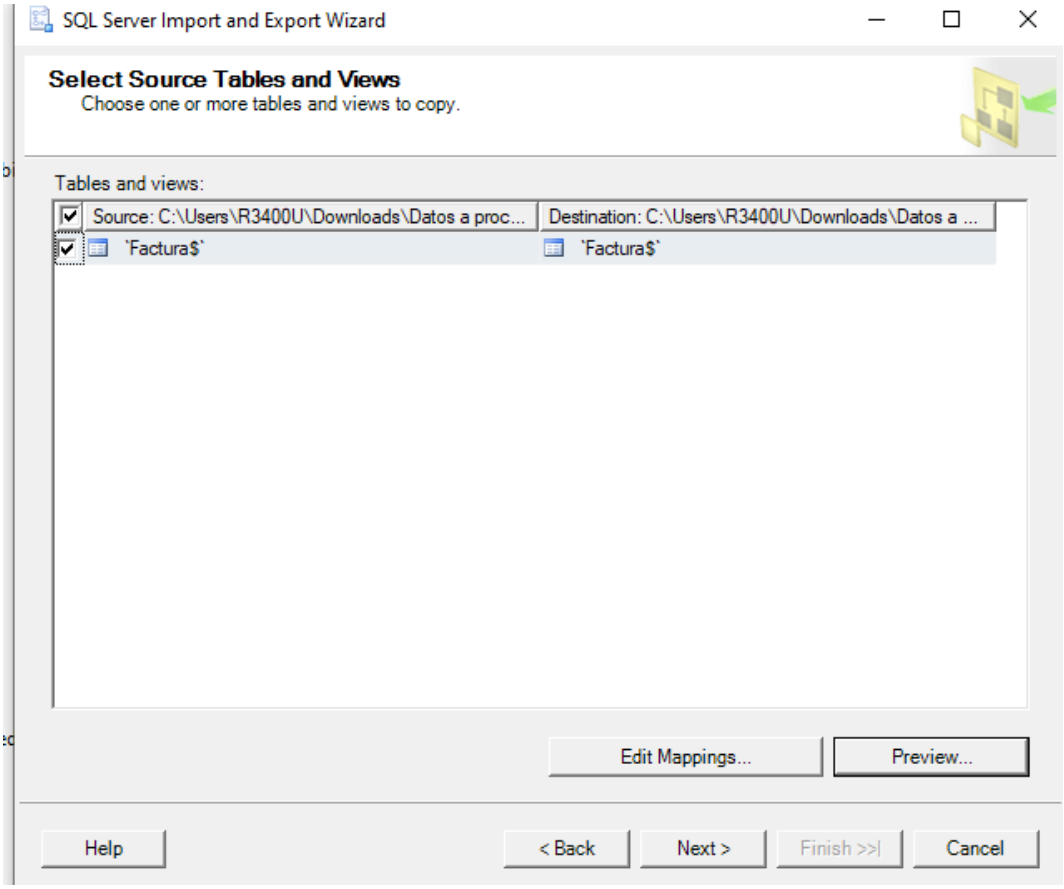
Contenido

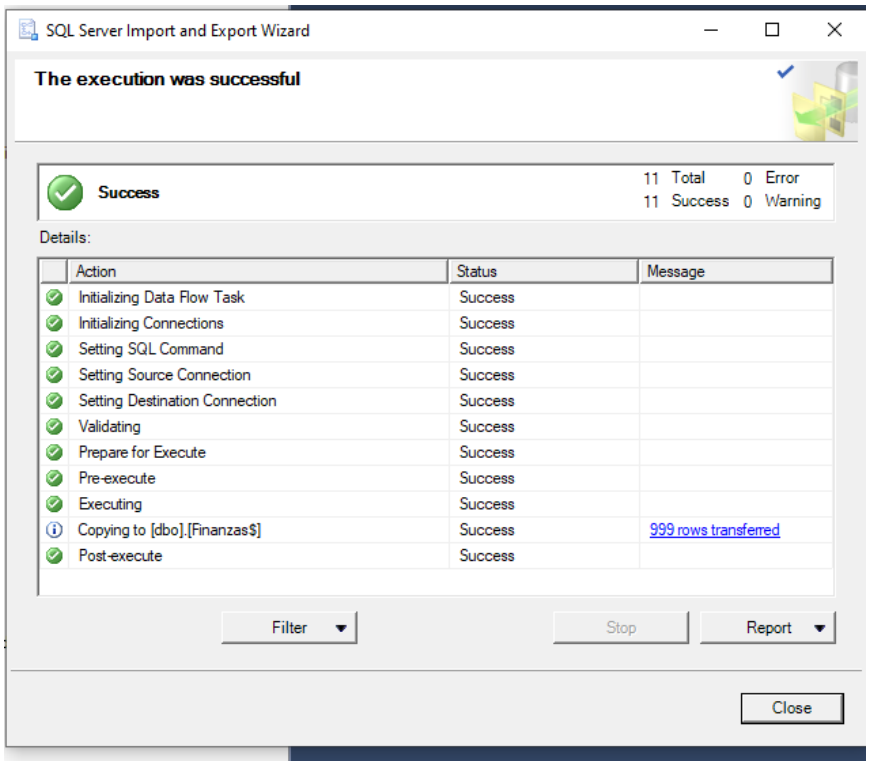
INDICE 2
PREPROCESAMIENTO DE DATOS 3
TRANSFORMACION DE DATOS 10
CREACIÓN DE MEDIDAS, COLUMNAS Y KPI 20
INFORMES Y ANALISIS DE RESULTADOS 21
UTILIZACION DE 2 METODOLOGIAS 25
GIT Y GITHUB 26

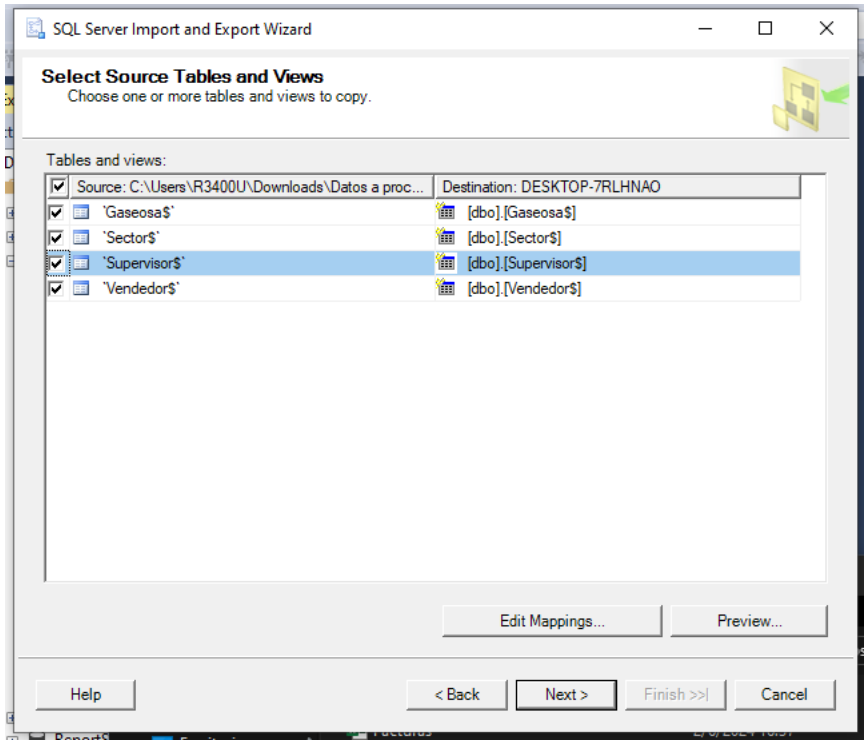
PREPROCESAMIENTO DE DATOS

1. Creamos una nueva base de datos en ssms.

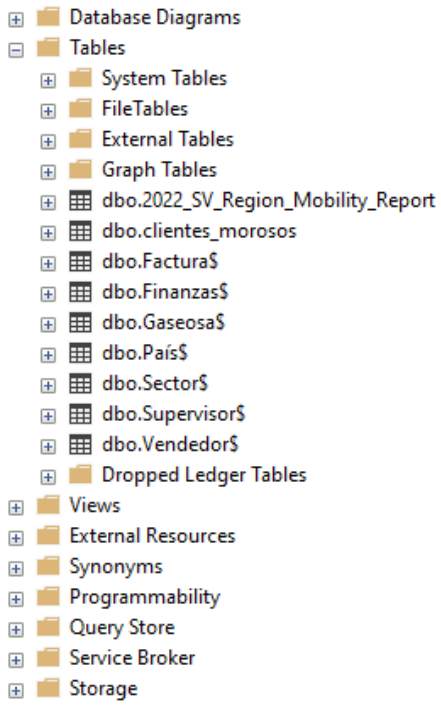
2. Ahora insertamos todas los Excel, con import data.



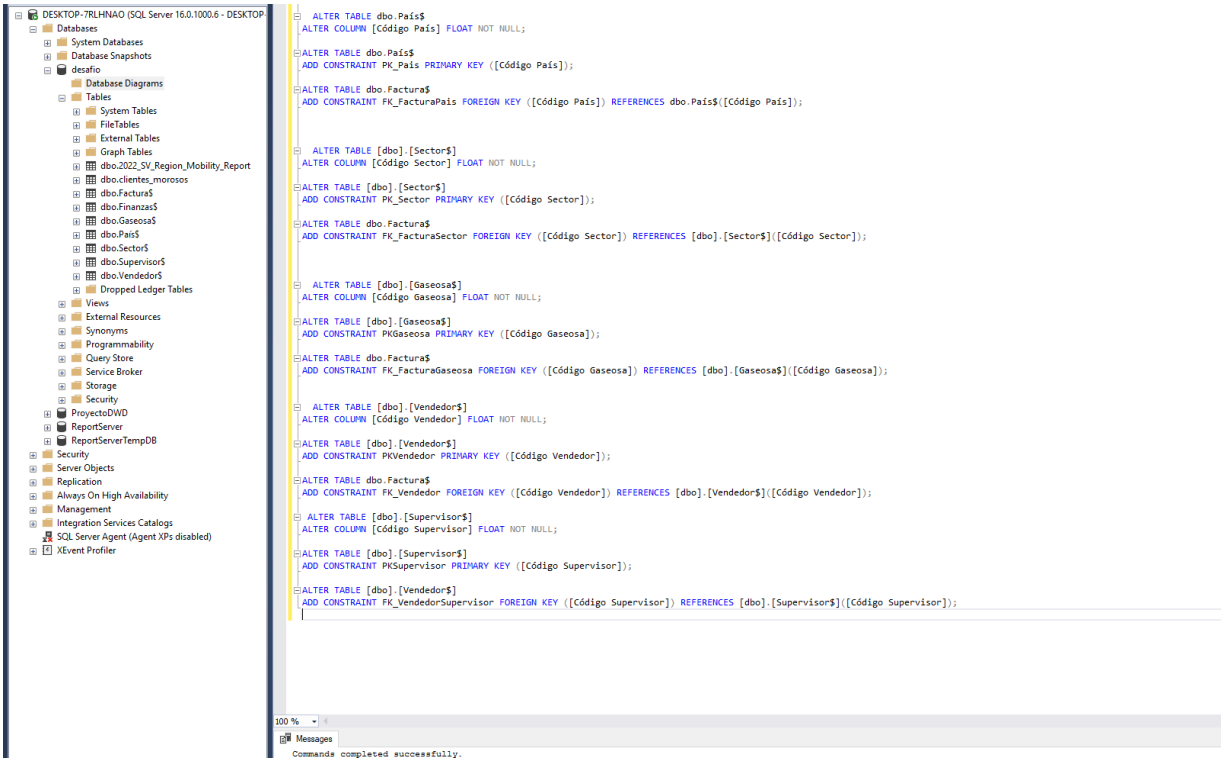

SQL Server Import and Export Wizard
The execution was successful
Success
11 Total 0 Error
11 Success 0 Warning
Details:
Action
Status
Message
Initializing Data Flow Task
Initializing Connections
Setting SQL Command
Setting Source Connection
Setting Destination Connection
Validating
Prepare for Execute
Pre-execute
Executing
Copying to [dbo].[Finanzas$]
999 rows transferred
Post-execute
Success
Filter
Stop
Report
Close


SQL Server Import and Export Wizard
Select Source Tables and Views
Choose one or more tables and views to copy.
Tables and views:
Source: C:\Users\R3400U\Downloads\Datos a proc...
Destination: DESKTOP-7RLHNAO
`Gaseosa$`
[dbo].[Gaseosa$]
`Sector$`
[dbo].[Sector$]
`Supervisor$`
[dbo].[Supervisor$]
`Vendedor$`
[dbo].[Vendedor$]
Edit Mappings...
Preview...
Help
< Back
Next >
Finish >>|
Cancel

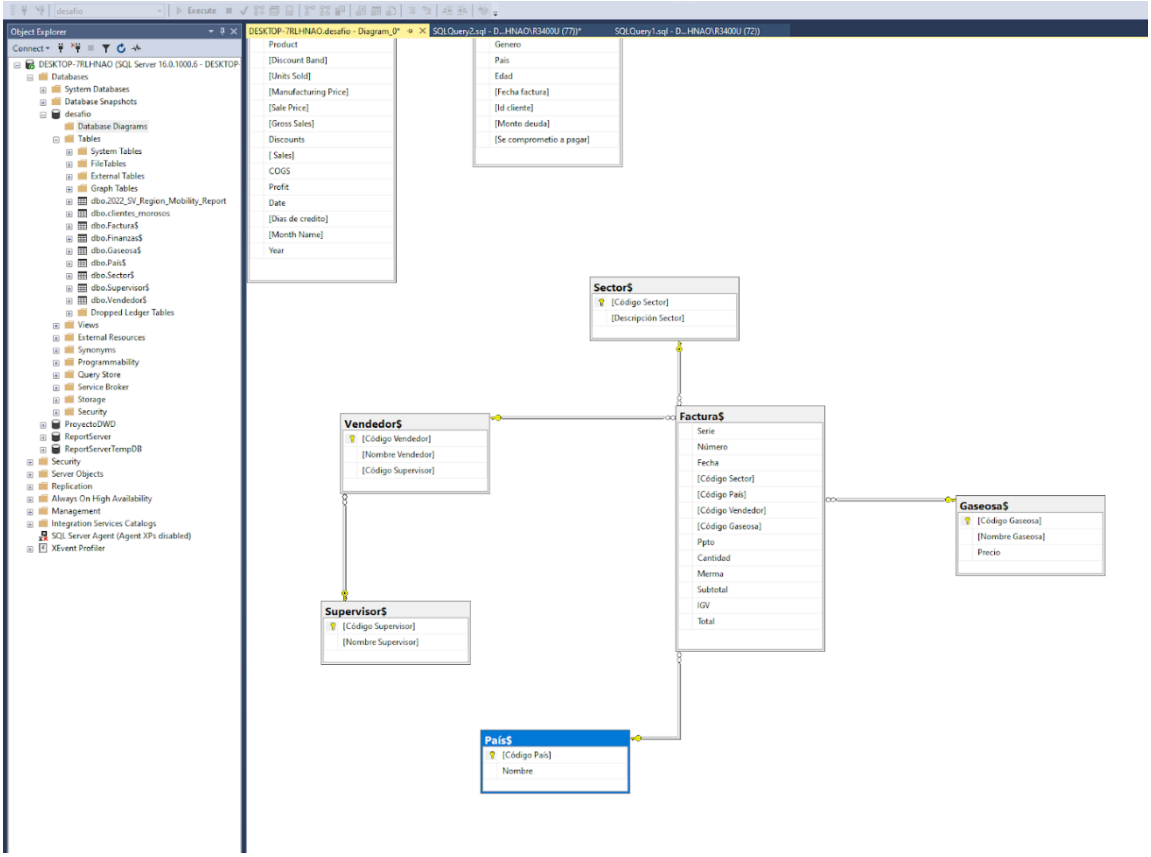
3. Cuando ya se inserten los excel nos aparecerá de esta manera. Y solo Ocuparemos las tablas Factura$, Gaseosa$, país$, Vendedor$, Sector$ y Supervisor$.



4. Ahora, ejecutamos las query para generar las llaves primarias y foráneas.



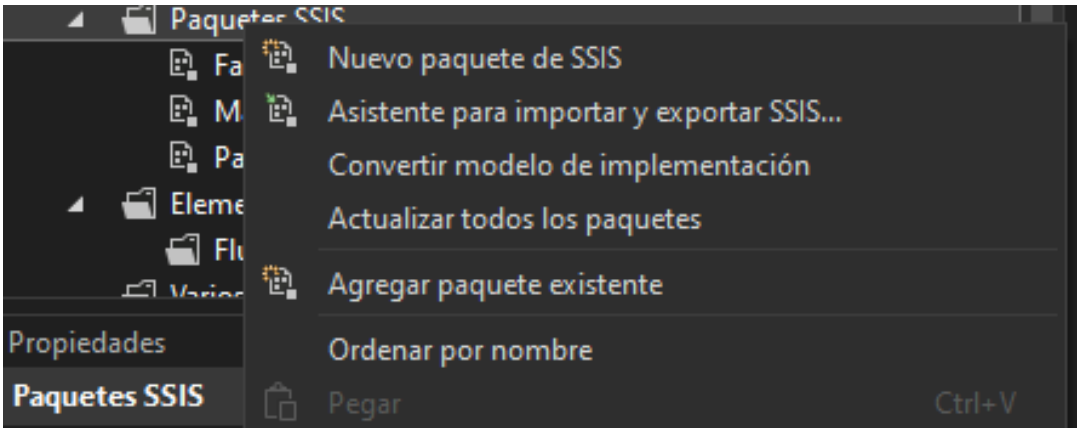
5. En ssms, generamos el diagrama.



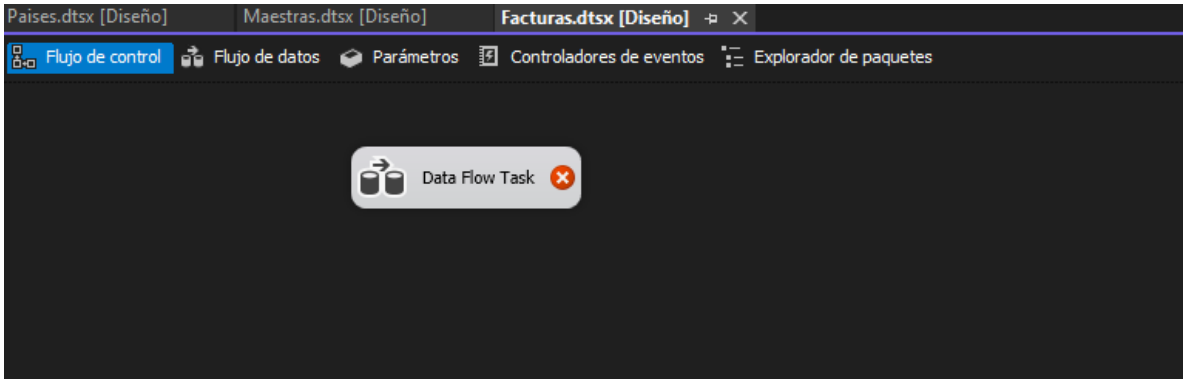
OTRA MANERA DE PREPROCESAR DATOS ESTA VEZ PARA VISUAL.

1. Creamos una nueva base de datos en ssms.

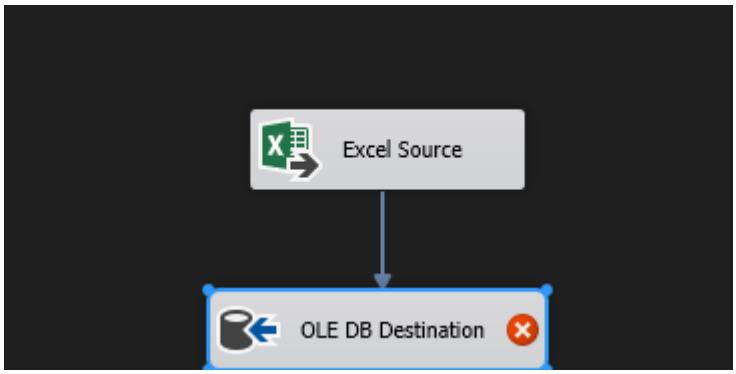
2. Ahora en visual studio creamos un proyecto de integration services. Y creamos 3 paquetes de SSIS, ya que utilizaremos 3 distintos Excel y así se agreguen en la misma BD.



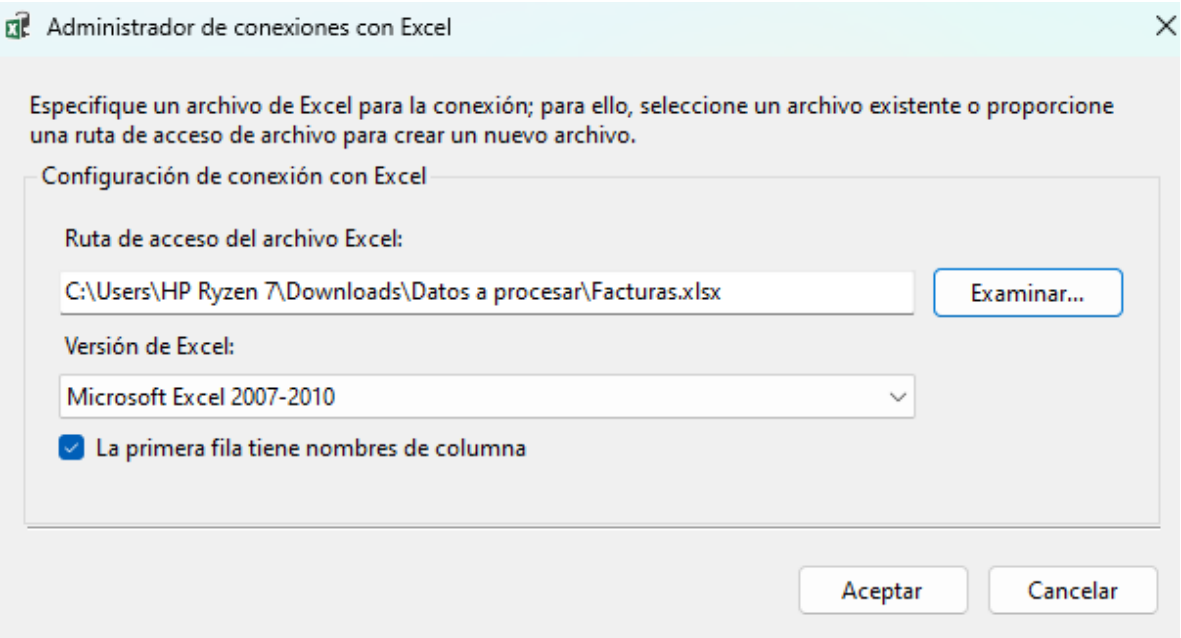
3. Agregaremos un Data Flow para cada paquete.



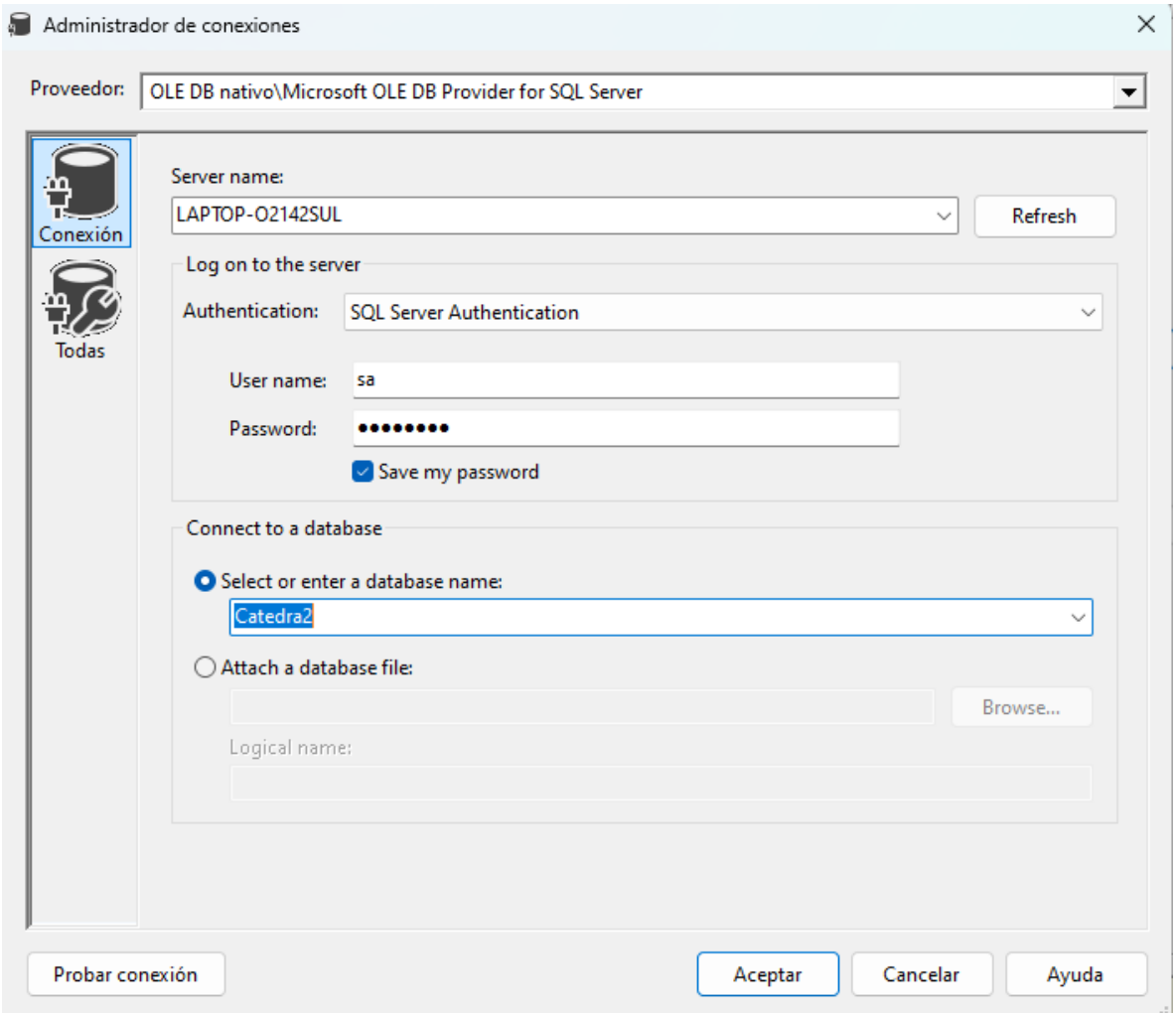
4. Agregaremos dentro del data Flow un Excel source y un ole DB destination. Para cada paquete.



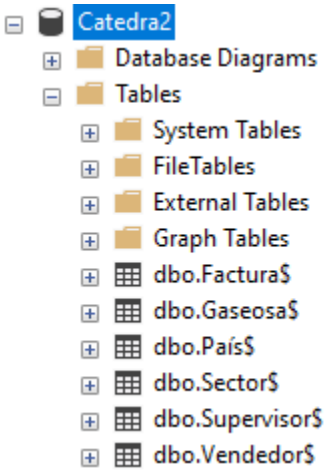
5. Ahora en el Excel pondremos cada Excel así sea finanzas, maestras y países.



6. Ahora conectamos con la base de datos.



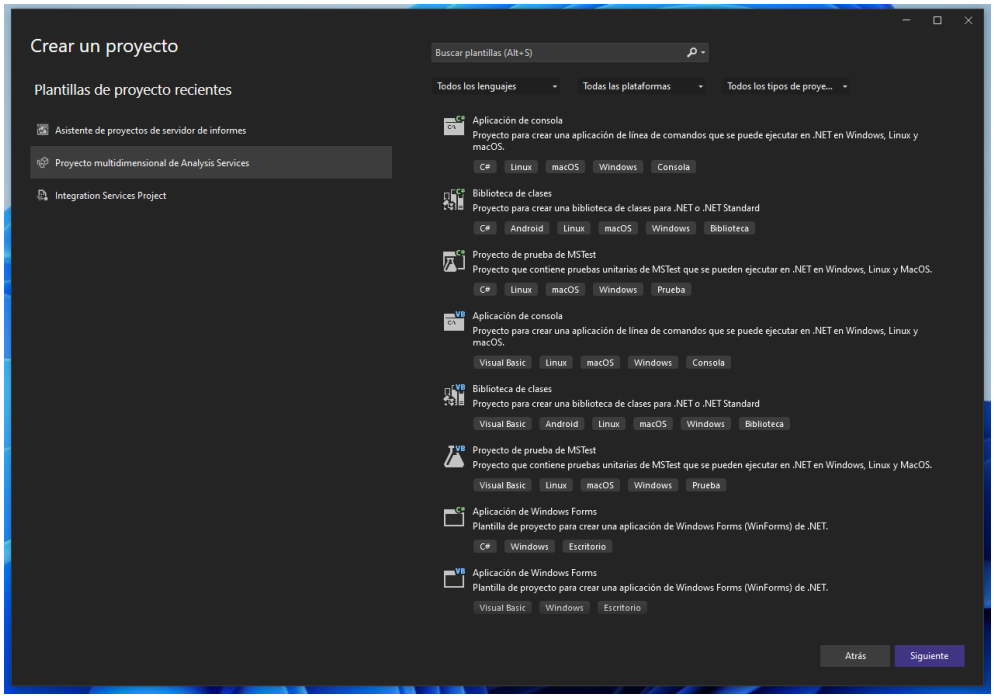
7. Y luego le damos iniciar a cada uno y nos saldrá en verde luego vamos a ssms y ya podemos ver las tablas agregadas con sus datos.

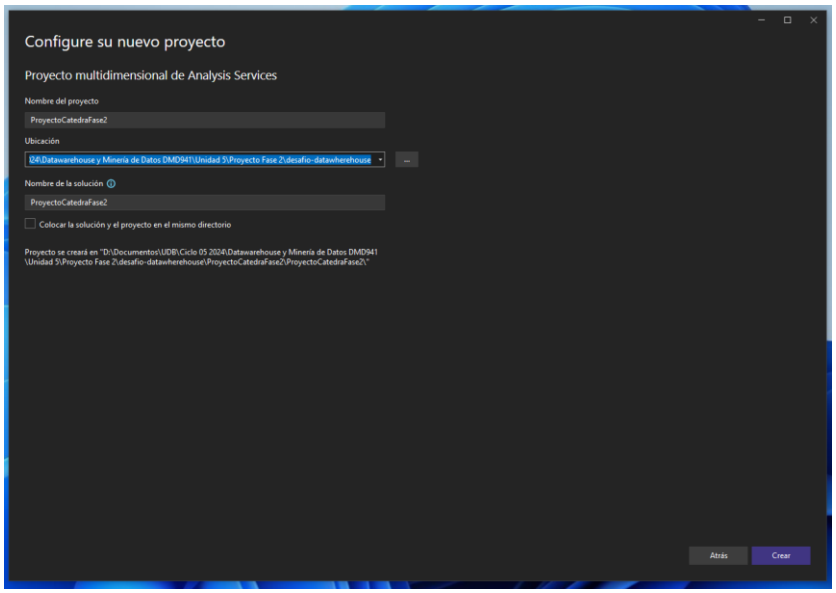


TRANSFORMACION DE DATOS

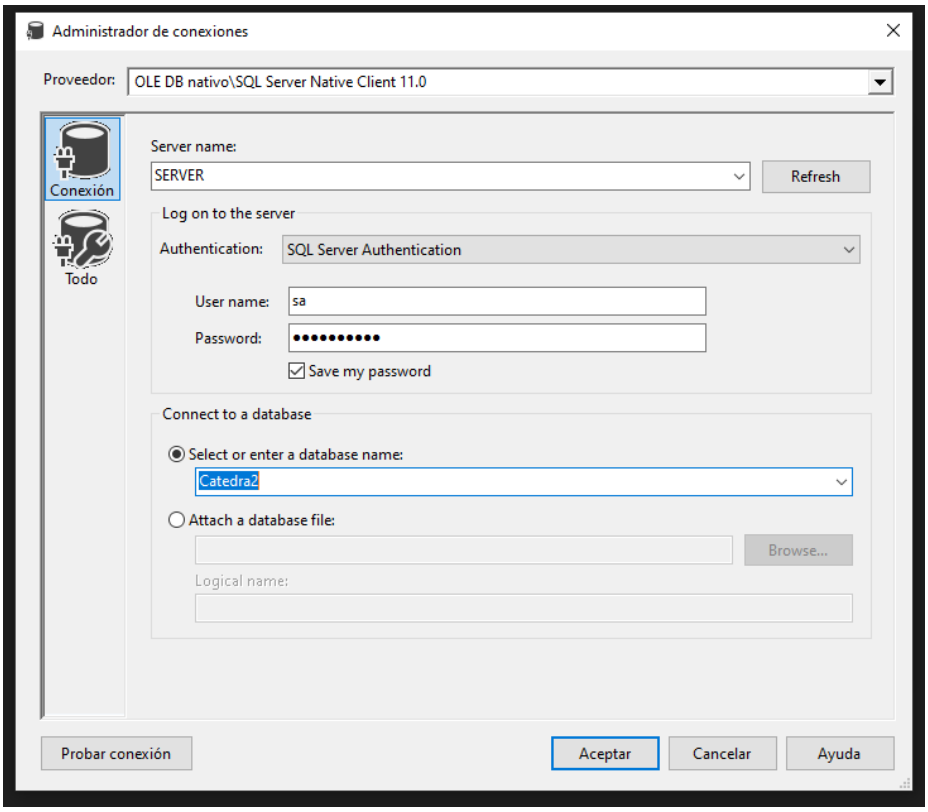
CUBO OLAP

1. Para crear cubo primero creamos un proyecto multidimensional.

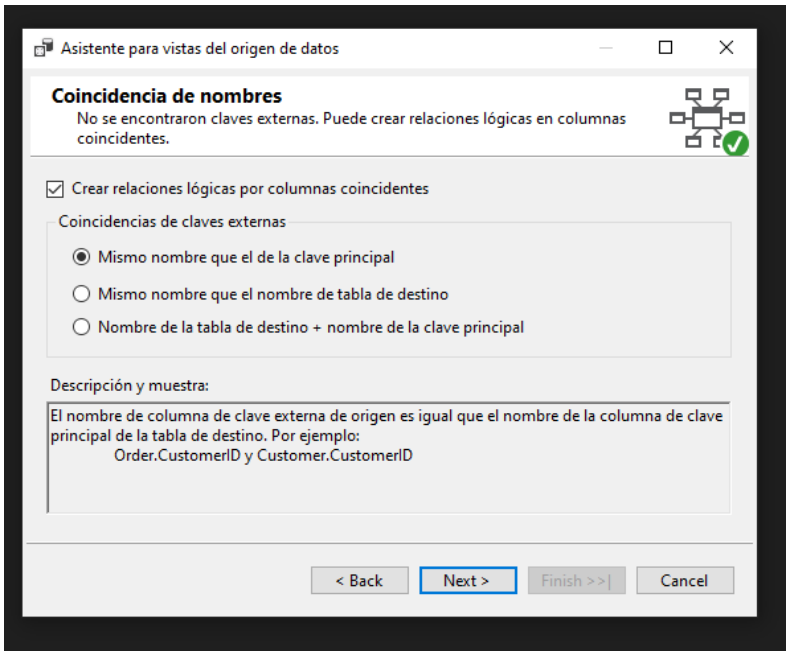




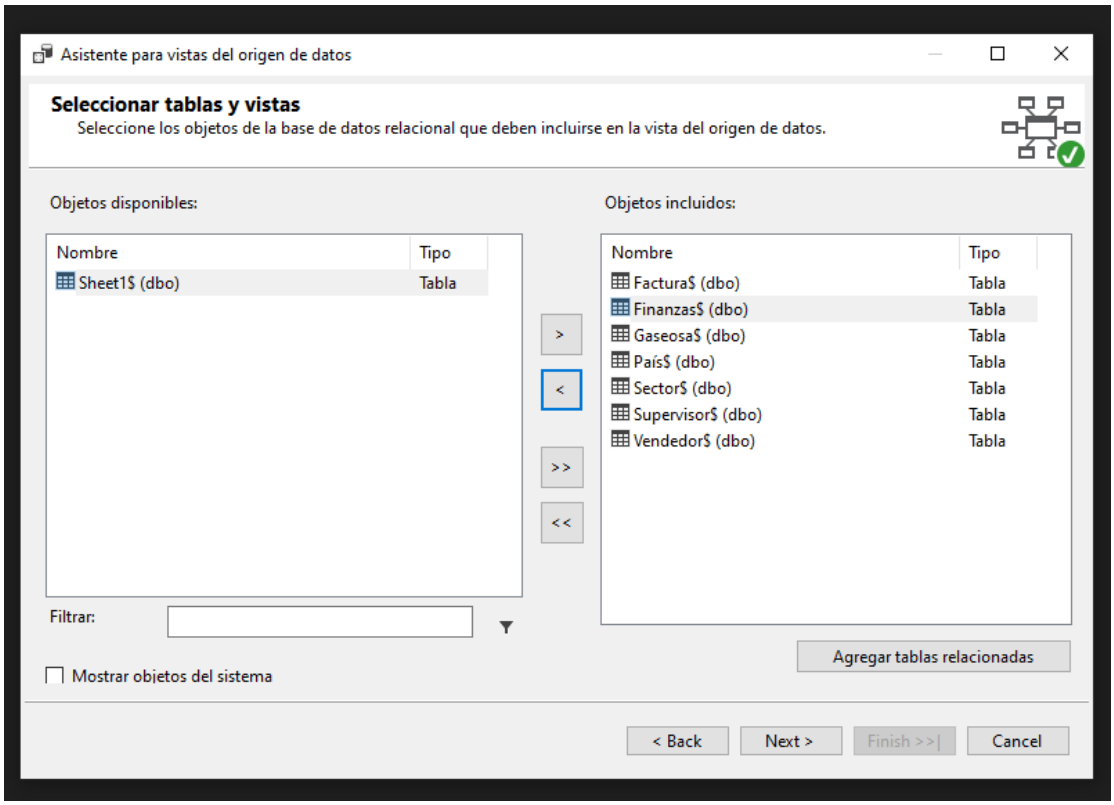
2. Se conecta a la BD.



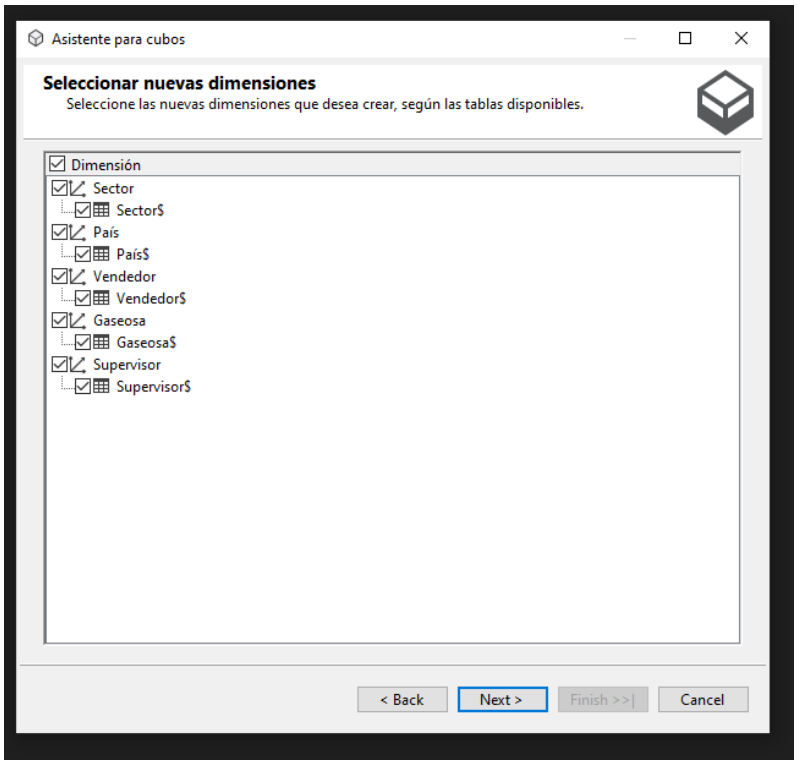
3. En el asistente para vista de origen de datos lo dejamos de la siguiente forma.



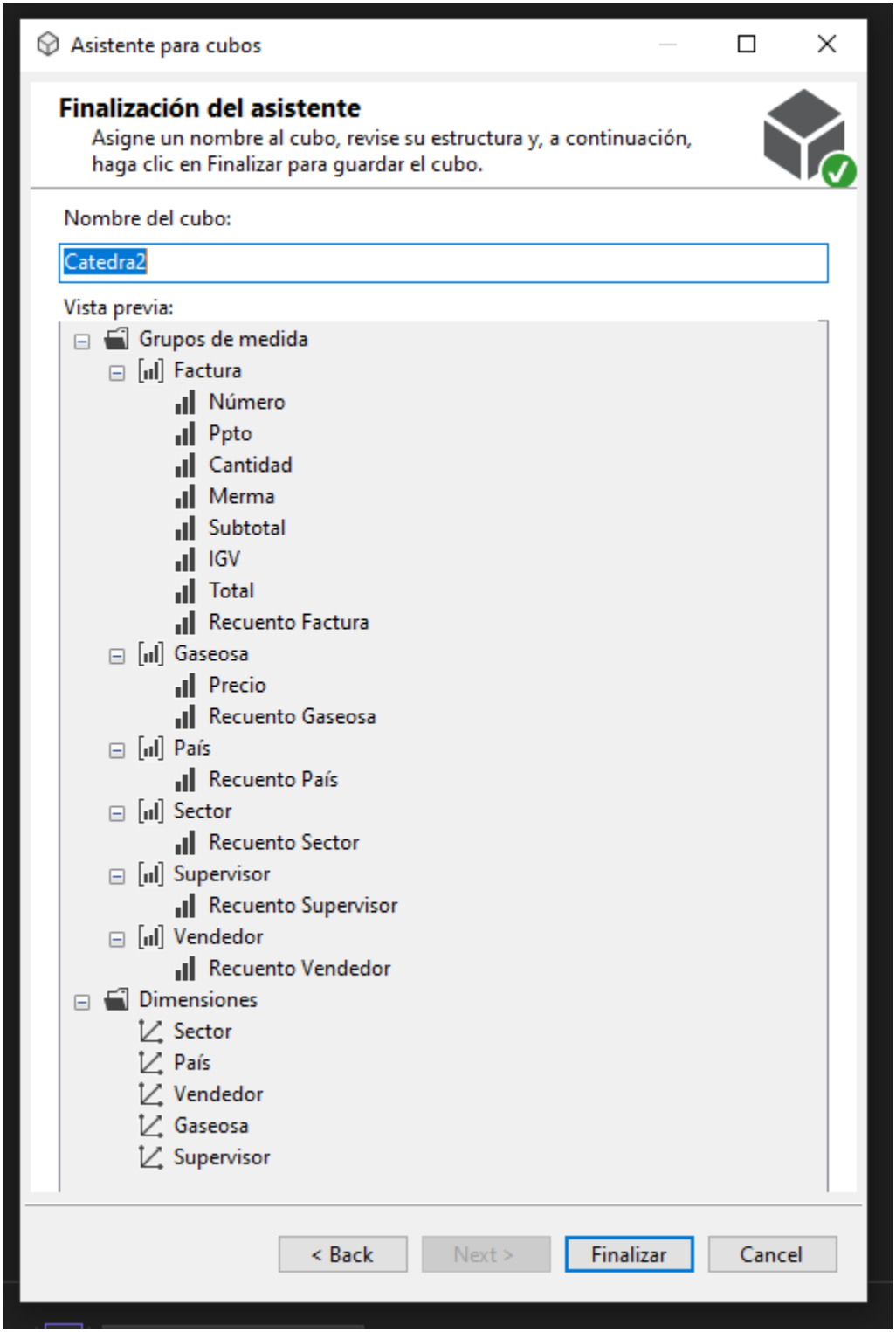
4. Escogemos las tablas necesarias.



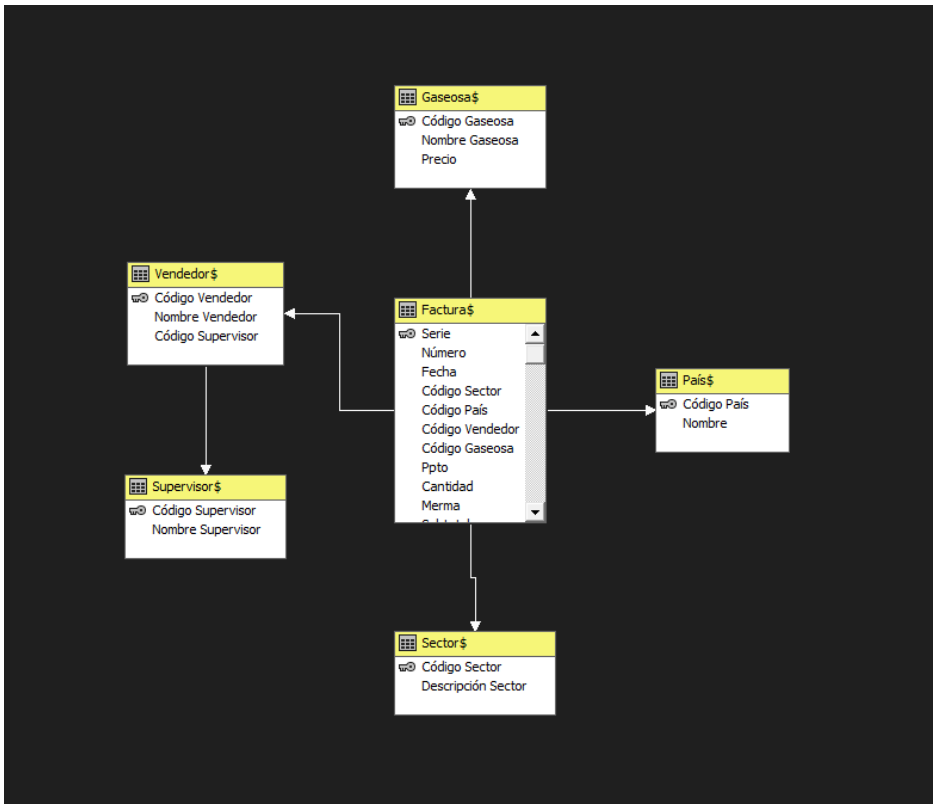
5. Así nos quedarían las dimensiones.



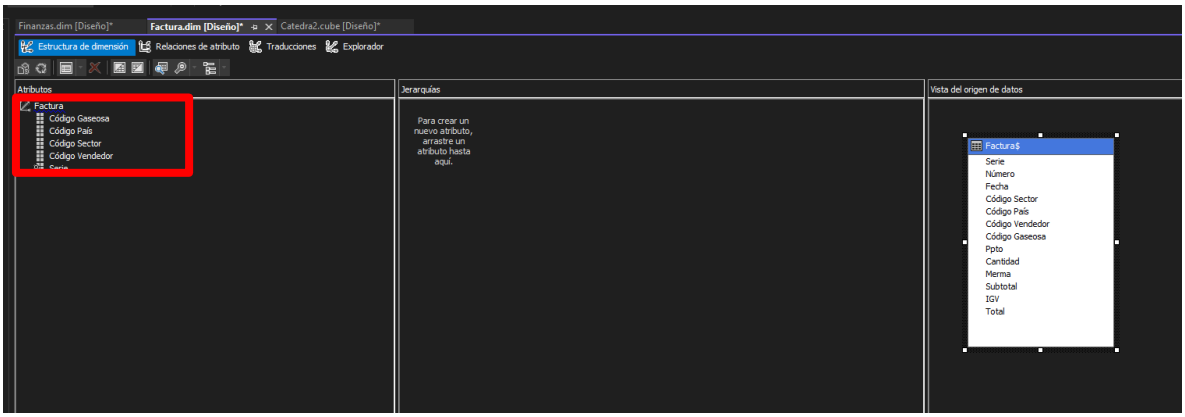
6. Así nos quedarían las medidas y dimensiones.

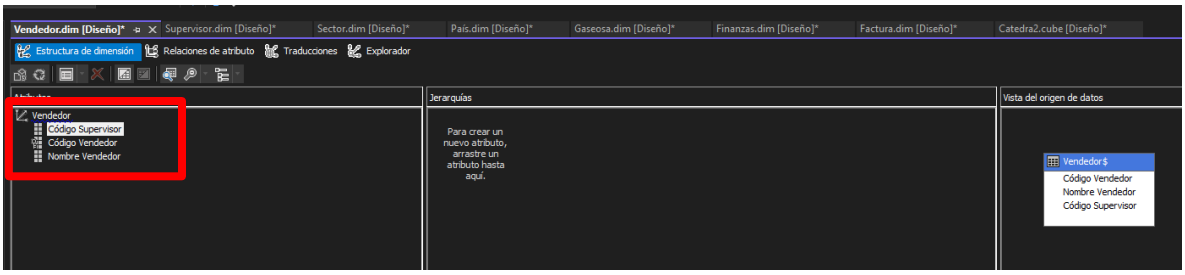


7. Y nos aparecerá el cubo de esta manera.



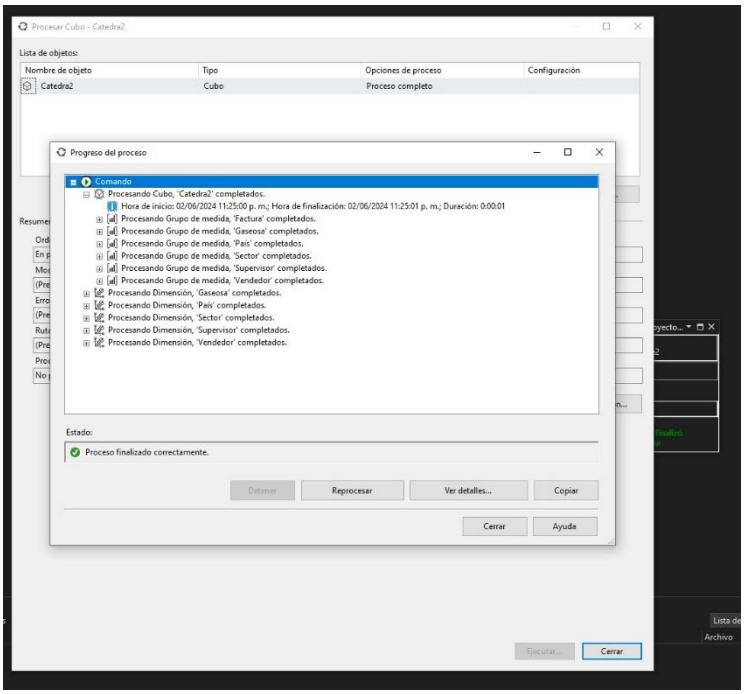
8. Y si no nos aparece así debemos agregar los atributos que tienen relación.



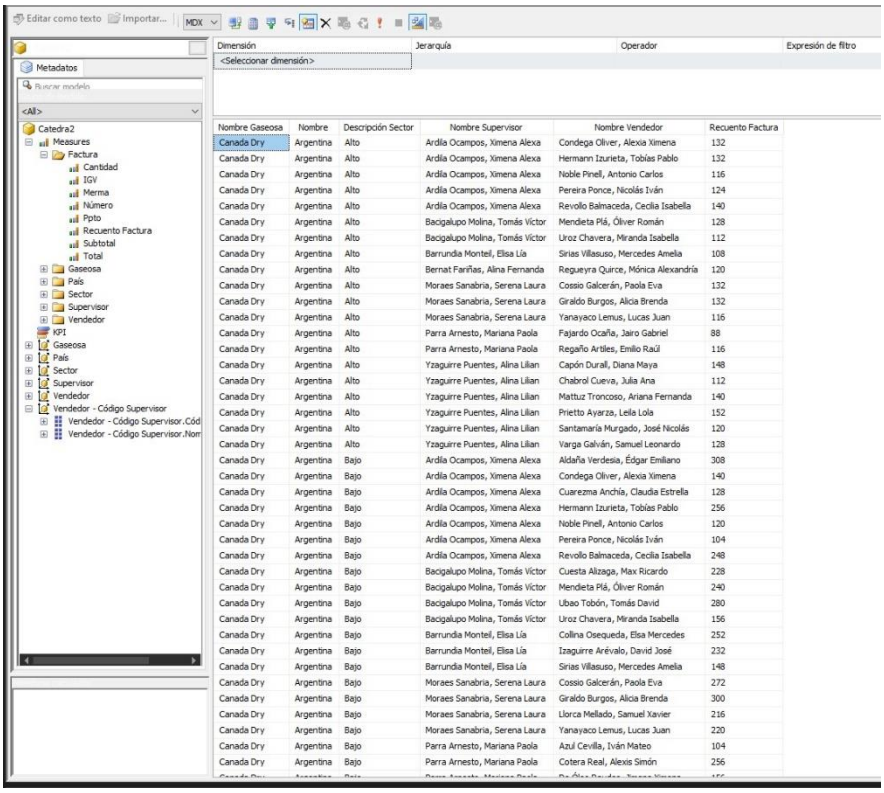


En la parte de la izquierda las que están en blanco son los atributos que se agregaron para que las tablas tengan la relación entre sí, en todo caso serían los códigos, que la tabla factura tiene relación con las demás.

9. Procesamos el cubo para ver si todo nos salió bien.



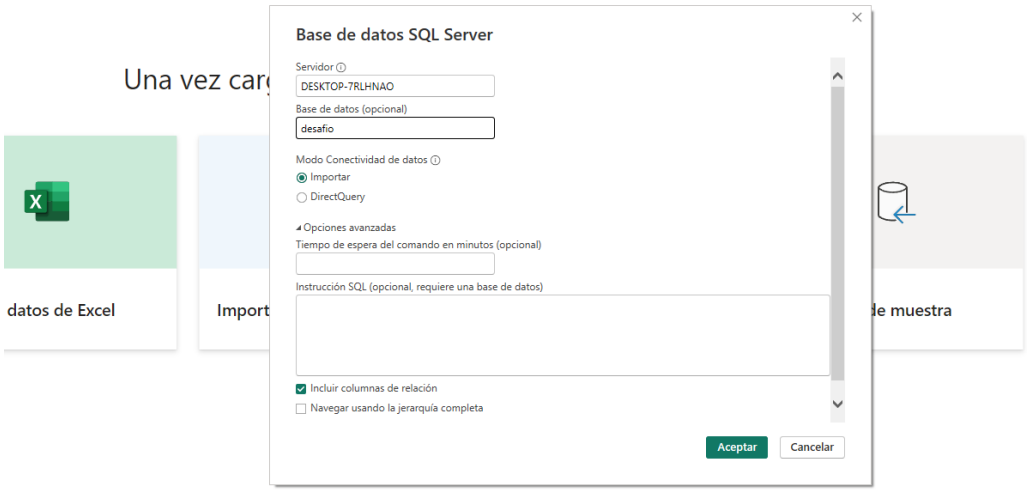
10. Todo salió en verde así que bien hacemos consulta para asegurarnos que funciona el cubo. Y como podemos observar el cubo funciona perfecto.



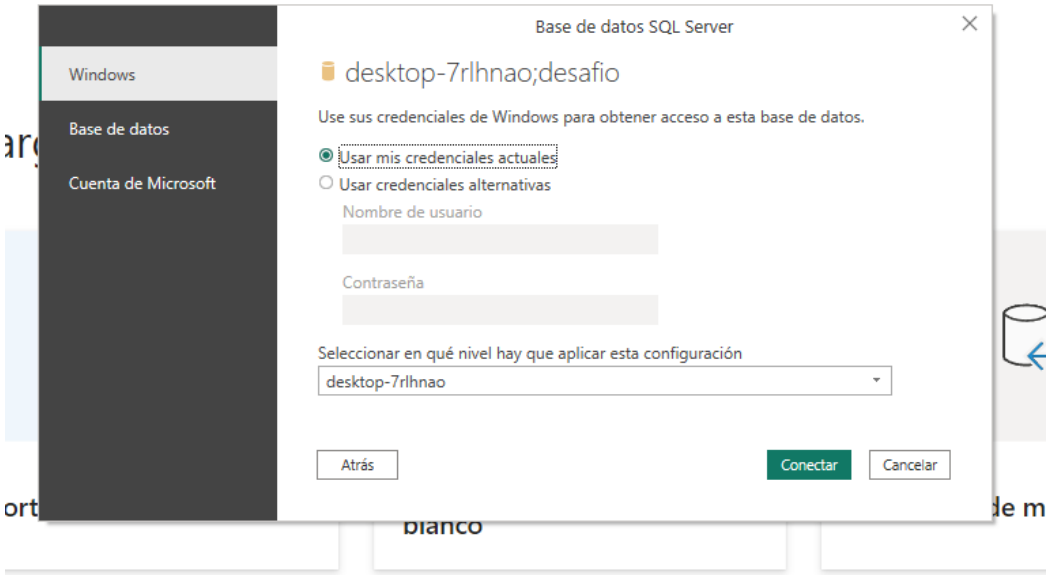
Nombre Gaseosa	Nombre	Descripción Sector	Nombre Supervisor	Nombre Vendedor	Recuento Factura
Canada Dry	Argentina	Alto	Ardila Ocampos, Ximena Alexa	Condega Oliver, Alexia Ximena	132
Canada Dry	Argentina	Alto	Ardila Ocampos, Ximena Alexa	Hermann Izurieta, Tobías Pablo	132
Canada Dry	Argentina	Alto	Ardila Ocampos, Ximena Alexa	Noble Pinell, Antonio Carlos	116
Canada Dry	Argentina	Alto	Ardila Ocampos, Ximena Alexa	Pereira Ponce, Nicolás Iván	124
Canada Dry	Argentina	Alto	Ardila Ocampos, Ximena Alexa	Revollo Balmaceda, Cecilia Isabella	140
Canada Dry	Argentina	Alto	Bacigalupo Molina, Tomás Víctor	Mendieta Plá, Óliver Román	128
Canada Dry	Argentina	Alto	Bacigalupo Molina, Tomás Víctor	Uroz Chavera, Miranda Isabella	112
Canada Dry	Argentina	Alto	Barrundia Monteil, Elisa Lía	Sirias Villasuso, Mercedes Amelia	108
Canada Dry	Argentina	Alto	Bernat Fariñas, Alina Fernanda	Regueyra Quirce, Mónica Alexandría	120
Canada Dry	Argentina	Alto	Moraes Sanabria, Serena Laura	Cossio Galcerán, Paola Eva	132
Canada Dry	Argentina	Alto	Moraes Sanabria, Serena Laura	Giraldo Burgos, Alicia Brenda	132
Canada Dry	Argentina	Alto	Moraes Sanabria, Serena Laura	Yanayaco Lemus, Lucas Juan	116
Canada Dry	Argentina	Alto	Parra Arnesto, Mariana Paola	Fajardo Ocaña, Jairo Gabriel	88
Canada Dry	Argentina	Alto	Parra Arnesto, Mariana Paola	Regaño Artiles, Emilio Raúl	116
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Capón Durall, Diana Maya	148
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Chabrol Cueva, Julia Ana	112
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Mattuz Troncoso, Ariana Fernanda	140
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Prietto Ayarza, Leila Lola	152
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Santamaría Murgado, José Nicolás	120
Canada Dry	Argentina	Alto	Yzagurre Puentes, Alina Lilian	Varga Galván, Samuel Leonardo	128
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Aldaña Verdesia, Édgar Emiliano	308
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Condega Oliver, Alexia Ximena	140
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Cuarezma Anchía, Claudia Estrella	128
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Hermann Izurieta, Tobías Pablo	256
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Noble Pinell, Antonio Carlos	120
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Pereira Ponce, Nicolás Iván	104
Canada Dry	Argentina	Bajo	Ardila Ocampos, Ximena Alexa	Revollo Balmaceda, Cecilia Isabella	248
Canada Dry	Argentina	Bajo	Bacigalupo Molina, Tomás Víctor	Cuesta Alizaga, Max Ricardo	228
Canada Dry	Argentina	Bajo	Bacigalupo Molina, Tomás Víctor	Mendieta Plá, Óliver Román	240
Canada Dry	Argentina	Bajo	Bacigalupo Molina, Tomás Víctor	Ubao Tobón, Tomás David	280
Canada Dry	Argentina	Bajo	Bacigalupo Molina, Tomás Víctor	Uroz Chavera, Miranda Isabella	156
Canada Dry	Argentina	Bajo	Barrundia Monteil, Elisa Lía	Collina Osequeda, Elsa Mercedes	252
Canada Dry	Argentina	Bajo	Barrundia Monteil, Elisa Lía	Izaguirre Arévalo, David José	232
Canada Dry	Argentina	Bajo	Barrundia Monteil, Elisa Lía	Sirias Villasuso, Mercedes Amelia	148
Canada Dry	Argentina	Bajo	Moraes Sanabria, Serena Laura	Cossio Galcerán, Paola Eva	272
Canada Dry	Argentina	Bajo	Moraes Sanabria, Serena Laura	Giraldo Burgos, Alicia Brenda	300
Canada Dry	Argentina	Bajo	Moraes Sanabria, Serena Laura	Llorca Mellado, Samuel Xavier	216
Canada Dry	Argentina	Bajo	Moraes Sanabria, Serena Laura	Yanayaco Lemus, Lucas Juan	220
Canada Dry	Argentina	Bajo	Parra Arnesto, Mariana Paola	Azul Cevilla, Iván Mateo	104
Canada Dry	Argentina	Bajo	Parra Arnesto, Mariana Paola	Cotera Real, Alexis Simón	256

POWER BI Y DONDE SE TRABAJARON LAS MEDIDAS, KPI, ETC.

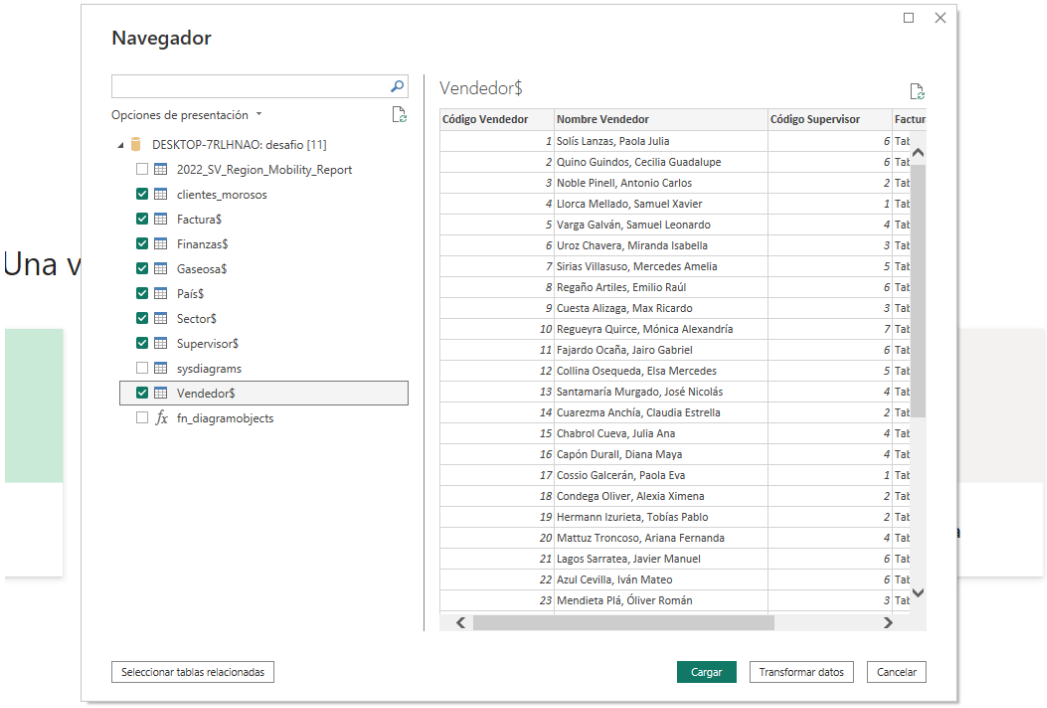
1. Nos conectamos a Power Bi. Y nos conectamos a la BD.



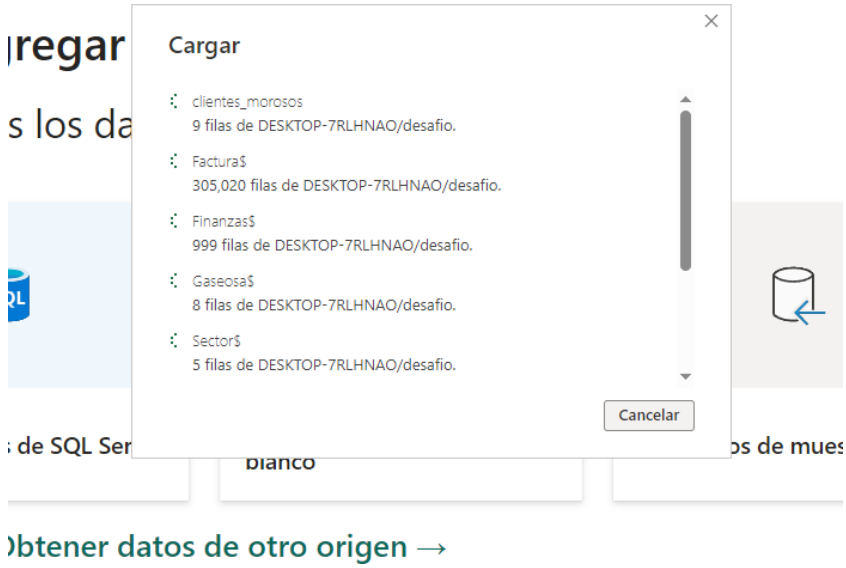
2. Usamos las credenciales actuales.



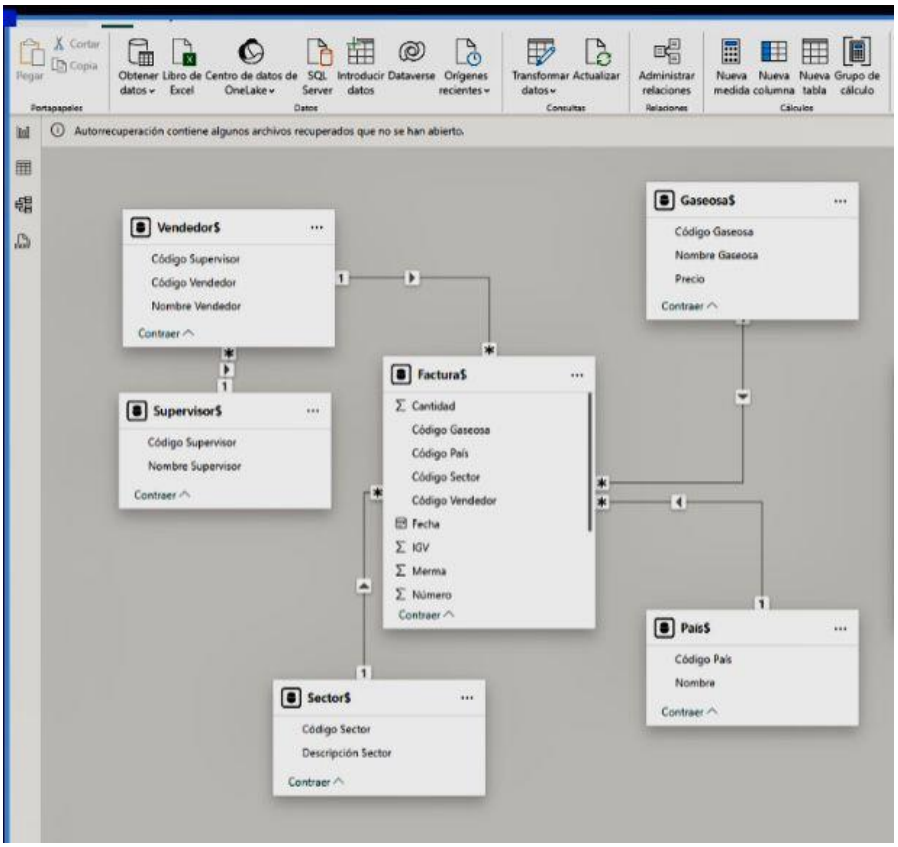
3. Seleccionamos las tablas a usar como se mencionó anteriormente.



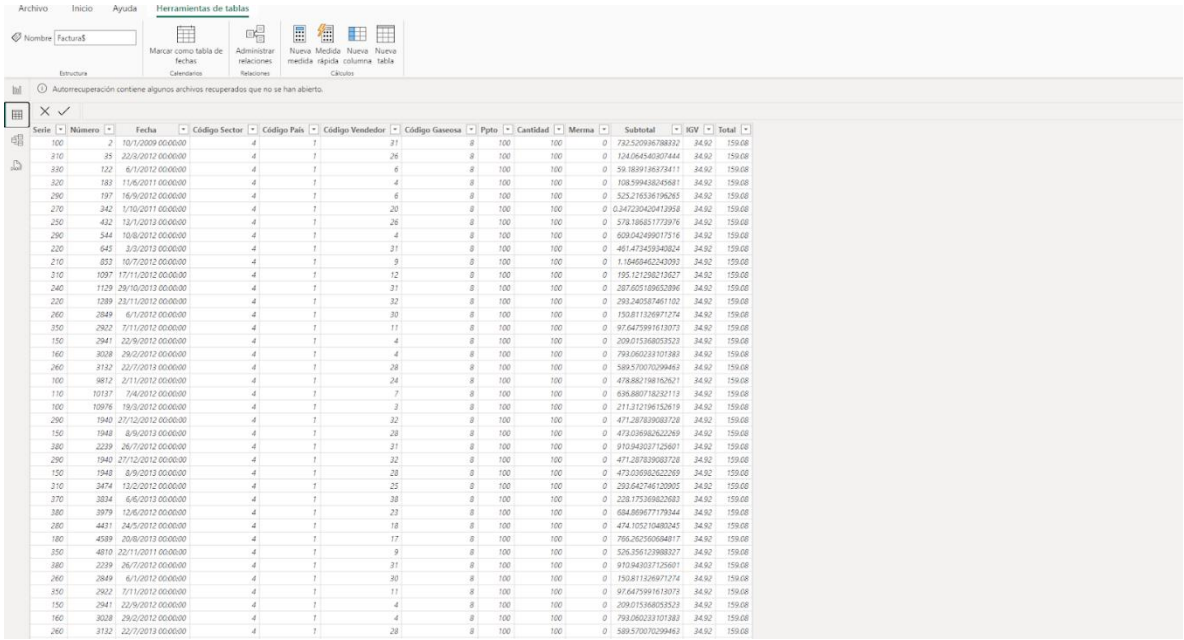
4. Se cargarán los datos.



5. Vista del modelo.



6. La visualización de datos.



Serie	Número	Fecha	Código Sector	Código País	Código Vendedor	Código Gaseosa	Ppto	Cantidad	Merma	Subtotal	IGV	Total
100	2	10/1/2009 00:00:00	4	1	31	8	100	100	0	732.520936788332	34.92	159.08
310	35	22/3/2012 00:00:00	4	1	26	8	100	100	0	124.064540307444	34.92	159.08
330	122	6/1/2012 00:00:00	4	1	6	8	100	100	0	59.1839136573411	34.92	159.08
320	183	11/6/2011 00:00:00	4	1	4	8	100	100	0	108.599438245681	34.92	159.08
290	197	16/9/2012 00:00:00	4	1	6	8	100	100	0	525.216536196265	34.92	159.08
270	342	1/10/2011 00:00:00	4	1	20	8	100	100	0	0.34723042041395 8	34.92	159.08
250	432	12/1/2013 00:00:00	4	1	26	8	100	100	0	578.186851773976	34.92	159.08
290	544	10/8/2012 00:00:00	4	1	4	8	100	100	0	609.042499017516	34.92	159.08
220	645	3/3/2013 00:00:00	4	1	31	8	100	100	0	461.473459340824	34.92	159.08
210	833	10/7/2012 00:00:00	4	1	9	8	100	100	0	1.16468462243093	34.92	159.08
310	1097	17/11/2012 00:00:00	4	1	12	8	100	100	0	195.121298213627	34.92	159.08
240	1129	29/10/2013 00:00:00	4	1	31	8	100	100	0	287.605189652896	34.92	159.08
220	1289	23/11/2012 00:00:00	4	1	32	8	100	100	0	293.240587461102	34.92	159.08
260	2849	6/1/2012 00:00:00	4	1	30	8	100	100	0	150.811326971274	34.92	159.08
350	2922	7/11/2012 00:00:00	4	1	11	8	100	100	0	97.6475991613073	34.92	159.08
150	2941	22/9/2012 00:00:00	4	1	4	8	100	100	0	209.015368053523	34.92	159.08
160	3028	29/2/2012 00:00:00	4	1	4	8	100	100	0	793.060233101383	34.92	159.08
260	3132	22/7/2013 00:00:00	4	1	28	8	100	100	0	589.570070299463	34.92	159.08
100	9812	2/11/2012 00:00:00	4	1	24	8	100	100	0	478.882198162621	34.92	159.08
110	10137	7/4/2012 00:00:00	4	1	7	8	100	100	0	635.880718232113	34.92	159.08
100	10976	19/3/2012 00:00:00	4	1	3	8	100	100	0	211.312196152619	34.92	159.08
290	1940	27/12/2012 00:00:00	4	1	32	8	100	100	0	471.287839088728	34.92	159.08
150	1948	8/9/2013 00:00:00	4	1	28	8	100	100	0	473.036982622269	34.92	159.08
380	2239	26/7/2012 00:00:00	4	1	31	8	100	100	0	910.943037125601	34.92	159.08
290	1940	27/12/2012 00:00:00	4	1	32	8	100	100	0	471.287839088728	34.92	159.08
150	1948	8/9/2013 00:00:00	4	1	28	8	100	100	0	473.036982622269	34.92	159.08
310	3474	13/2/2012 00:00:00	4	1	25	8	100	100	0	293.642746120905	34.92	159.08
370	3834	6/6/2013 00:00:00	4	1	38	8	100	100	0	228.175369822683	34.92	159.08
380	3979	12/6/2012 00:00:00	4	1	23	8	100	100	0	684.869677179344	34.92	159.08
280	4431	24/5/2012 00:00:00	4	1	18	8	100	100	0	474.105210480245	34.92	159.08
180	4589	20/8/2013 00:00:00	4	1	17	8	100	100	0	766.262560684817	34.92	159.08
350	4810	22/11/2011 00:00:00	4	1	9	8	100	100	0	526.356123988327	34.92	159.08
380	2239	26/7/2012 00:00:00	4	1	31	8	100	100	0	910.943037125601	34.92	159.08
260	2849	6/1/2012 00:00:00	4	1	30	8	100	100	0	150.811326971274	34.92	159.08
350	2922	7/11/2012 00:00:00	4	1	11	8	100	100	0	97.6475991613073	34.92	159.08
150	2941	22/9/2012 00:00:00	4	1	4	8	100	100	0	209.015368053523	34.92	159.08
160	3028	29/2/2012 00:00:00	4	1	4	8	100	100	0	793.060233101383	34.92	159.08
260	3132	22/7/2013 00:00:00	4	1	28	8	100	100	0	589.570070299463	34.92	159.08

CREACIÓN DE MEDIDAS, COLUMNAS Y KPI.

1. Estas medidas y columnas calculadas se utilizan para generar métricas clave (KPIs)

Datos

```
Categoría Cantidad =
IF([Cantidad] > 100, "Alta",
    IF([Cantidad] > 50, "Media", "Baja"))
```

2. Al categorizar los productos o servicios en "Alta", "Media" y "Baja" según su cantidad, puedes identificar cuáles son tus productos estrella (Alta), cuáles tienen un rendimiento aceptable (Media) y cuáles necesitan atención o posible discontinuación (Baja).

Datos

```
Comisiones = [Precio Total] * 0.1
```

Datos

```
Total Ventas = SUM([Precio Total])
```

3. Esta medida calcula el monto total de las ventas sumando el precio total de cada factura. El total de ventas es un KPI fundamental que muestra el rendimiento global de ventas de la empresa. Es crucial para evaluar la salud financiera y el éxito comercial.

Datos

```
Comisiones Totales = SUM([Comisiones])
```

4. Esta medida calcula el total de comisiones pagadas sumando el valor de las comisiones en cada factura.

Datos

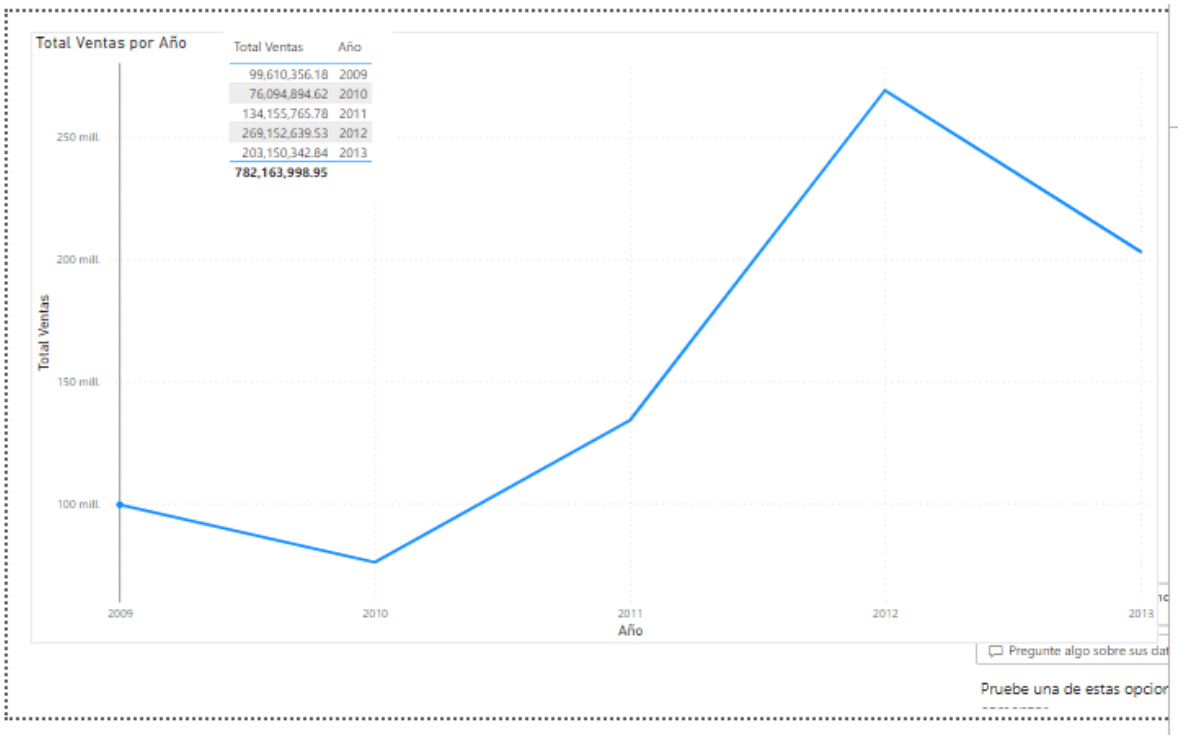
```
Promedio Ventas por Vendedor = AVERAGE([Precio Total])
```

5. Esta medida calcula el promedio de ventas por vendedor, tomando el precio total de las ventas y promediándolo. El promedio de ventas por vendedor es útil para comparar el rendimiento individual de los vendedores y para identificar a los vendedores que están superando o bajo rendimiento.

Datos | Consultas

```
Ventas por Pais = CALCULATE([Total Ventas], ALL('País'[Nombre]))
```

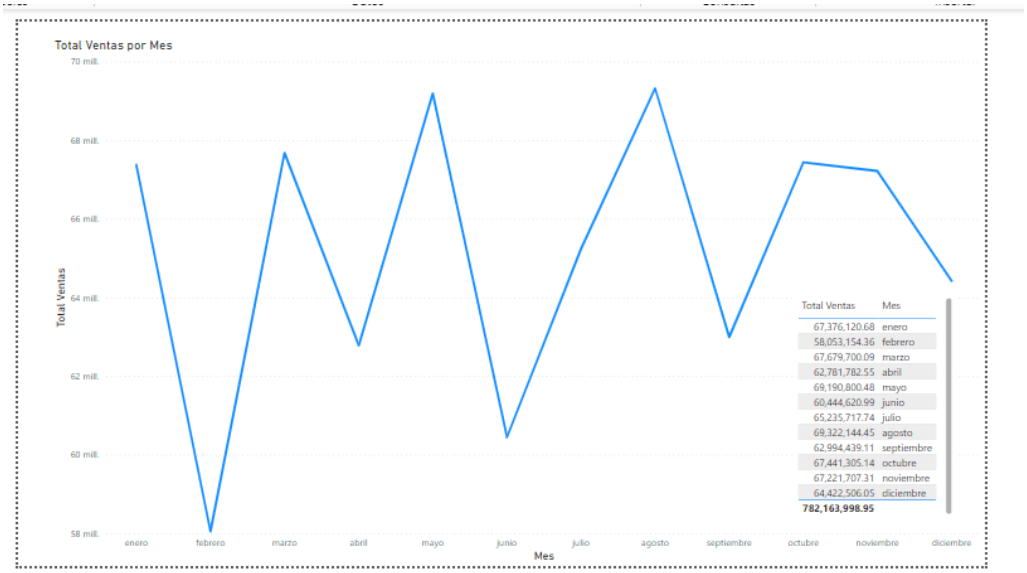
INFORMES Y ANALISIS DE RESULTADOS



El análisis del total de ventas por año proporciona una valiosa información para la toma de decisiones estratégicas dentro de una empresa. Algunos beneficios clave identificados incluyen:

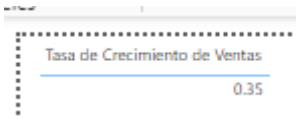
Evaluación del Rendimiento: Comparar las ventas anuales permite identificar tendencias de crecimiento y patrones estacionales, facilitando una mejor planificación de la producción y las campañas de marketing, como se puede observar en la gráfica esa tienda para el año 2013 tuvo un decremento con respecto al año anterior, pero ha ido creciendo los últimos años.

Planeación y Presupuestación: Las ventas históricas sirven de base para elaborar presupuestos realistas y proyecciones futuras, estableciendo metas alcanzables para los próximos períodos.



El análisis de las ventas mensuales proporciona información valiosa para la toma de decisiones empresariales, ilustrada con un ejemplo concreto:

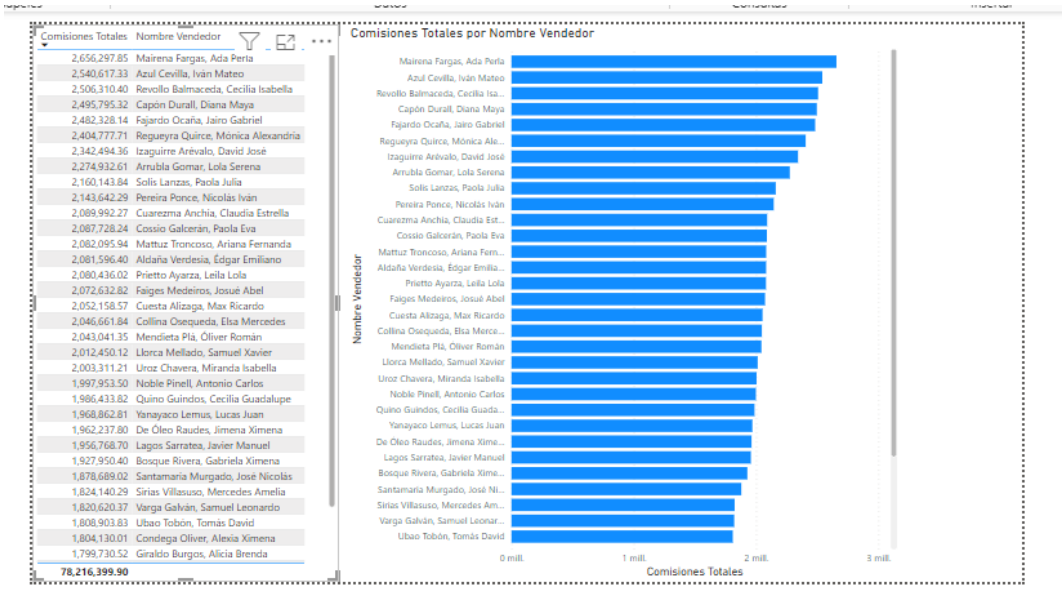
Al observar las ventas mensuales durante el año, podemos identificar claramente patrones que ayudan a la empresa a tomar decisiones estratégicas. Por ejemplo, si representamos las ventas mensuales en una gráfica, podríamos notar que los meses de mayo y agosto presentan picos más altos en comparación con otros meses del año. Estos picos podrían ser indicativos de temporadas de mayor demanda, posiblemente debido a festividades, promociones especiales o cambios estacionales en las preferencias de los consumidores.



Una tasa de crecimiento de ventas del 0.35 indica un aumento del 35% a lo largo del período de 2009 al 2013 implica un aumento constante en las ventas de la empresa durante esos cinco años. Esto puede tener varios impactos positivos en la toma de decisiones empresariales:

Indicador de Rendimiento Sostenido: Una tasa de crecimiento constante indica que la empresa ha mantenido un buen desempeño y ha sido capaz de aumentar sus ventas de manera consistente a lo largo del tiempo. Esta información puede brindar confianza a la dirección y a los inversores sobre la estabilidad y la salud financiera de la empresa.

Planificación a largo plazo: Con un historial de crecimiento sostenido durante cinco años, la empresa puede realizar una planificación a largo plazo con mayor confianza. Esto incluye decisiones como la expansión de instalaciones, la diversificación de productos o la entrada en nuevos mercados.



En la gráfica de comisiones totales por vendedor, podemos observar varias tendencias importantes que van en el periodo de los 5 años:

Identificación de Vendedores Sobresalientes:

La vendedora Mairena Fargas destaca claramente con comisiones totales de 2,656,297.85. Esta vendedora supera significativamente a sus compañeros, lo cual indica un rendimiento excepcional. Otros vendedores como el vendedor Ivan Mateo y la vendedora Cecilia Isabella también muestran un buen desempeño con comisiones arriba de los 2,500,000 respectivamente.

Evaluación del Rendimiento General:

Mientras que algunos vendedores están obteniendo comisiones cercanas o superiores a los 2,000,000, hay otros que están por debajo de esta marca, como el vendedor Tomas Davis que está por debajo de los 1,900,000.
Este contraste en las comisiones sugiere que hay una variabilidad significativa en el rendimiento de los vendedores.

Los datos sugieren que sería beneficioso revisar los planes de incentivos. Podría ser útil implementar programas de recompensas adicionales para motivar a los vendedores que están cerca del rendimiento medio a mejorar aún más.

Para los vendedores con comisiones más bajas, sería recomendable ofrecer capacitación adicional y apoyo para mejorar sus técnicas de ventas.

Revisar los planes de incentivos y ofrecer capacitación adicional permite a la empresa motivar a sus vendedores a mejorar su rendimiento, lo que puede aumentar las ventas y la eficiencia general del equipo.

Categoria Cantidad	Suma de Cantidad	Nombre Gaseosa	Nombre	Año
Media	200	Coca Cola	Colombia	2012
Media	200	Crush	Bolivia	2009
Media	200	Pepsi	Colombia	2013
Media	200	Seven Up	Bolivia	2010
Media	200	Seven Up	Colombia	2013
Baja	150	Seven Up	Argentina	2010
Media	100	Canada Dry	Colombia	2012
Media	100	Canada Dry	Colombia	2013
Media	100	Coca Cola	Colombia	2011
Media	100	Crush	Colombia	2010
Media	100	Crush	Colombia	2013
Media	100	Fanta	Colombia	2011
Media	100	Inca Kola	Colombia	2010
Media	100	Pepsi	Colombia	2009
Media	100	Pepsi	Colombia	2010
Media	100	Pepsi	Colombia	2011
Media	100	Pepsi	Colombia	2012
Media	100	Sprite	Colombia	2009
Media	100	Sprite	Colombia	2011
Media	100	Sprite	Colombia	2013
Total	**363220837**			

Como se puede observar en este informe, se categorizaron los productos en tres categorías: alta, media y baja, y también se hizo una suma de cantidad por categoría para poder analizar qué productos son los que menos se venden y en qué país, y como se puede observar hay una tendencia significativa en Colombia con varias productos pero más con Pepsi y Sprite que está en una media de 100 por año y no ha aumentado en el periodo de los 5 años, este análisis puede ayudar a la empresa a tomar mejores estrategias de marketing con dicho producto en un país en específico, según el informe se puede observar que ha seguido cerca de esta tendencia hasta el 2013 con una cantidad de 200 productos vendidos.

UTILIZACION DE 2 METODOLOGIAS

Power BI es una herramienta de análisis y visualización de datos desarrollada por Microsoft. Su propósito principal es ayudar a las organizaciones a convertir datos sin procesar en información útil y visualmente atractiva, facilitando la toma de decisiones basada en datos. Power BI se compone de varios componentes:

- **Power BI Desktop:** Una aplicación de escritorio que permite a los usuarios conectarse a diversas fuentes de datos, transformarlos y crear visualizaciones interactivas. Aquí, los usuarios pueden combinar datos de múltiples orígenes y crear modelos de datos complejos.
- **Power BI Service:** Una plataforma en la nube donde los usuarios pueden publicar y compartir los informes y dashboards creados en Power BI Desktop. Esta plataforma permite la colaboración en tiempo real y la actualización automática de los datos.
- **Power BI Mobile:** Aplicaciones móviles disponibles para iOS, Android y Windows, que permiten a los usuarios acceder e interactuar con sus dashboards e informes desde cualquier lugar.
- **Power BI Report Server:** Una solución on-premises que permite a las organizaciones mantener sus datos y reportes dentro de su propia infraestructura de TI, sin necesidad de subirlos a la nube.

Power BI soporta una amplia gama de fuentes de datos, como bases de datos SQL, Excel, servicios web y más. Ofrece herramientas poderosas para el modelado de datos, incluyendo Power Query para la transformación de datos y DAX (Data Analysis Expressions) para cálculos avanzados. Los usuarios pueden crear visualizaciones interactivas y dashboards personalizados que incluyen gráficos de barras, líneas, mapas geográficos, matrices y mucho más.

Cubos OLAP es una tecnología que facilita el análisis multidimensional de grandes volúmenes de datos en el ámbito de la inteligencia de negocios (BI). Los cubos OLAP son estructuras de datos diseñadas para permitir consultas rápidas y eficientes desde múltiples perspectivas o dimensiones. Los componentes principales de un cubo OLAP incluyen:

- **Dimensiones:** Categorías de datos, como tiempo, geografía o producto, que permiten filtrar y segmentar los datos. Cada dimensión puede tener jerarquías que permiten desglosar los datos en niveles más detallados.
- **Medidas:** Datos numéricos que se analizan, como ventas, ingresos o costos. Estas medidas se calculan y se agregan en función de las dimensiones.
- **Jerarquías:** Estructuras dentro de las dimensiones que permiten a los usuarios profundizar o resumir datos, por ejemplo, años, trimestres y meses dentro de la dimensión de tiempo.

Los cubos OLAP son conocidos por su capacidad para manejar grandes volúmenes de datos y responder rápidamente a consultas complejas. Esto se logra mediante la pre-agregación de datos y el almacenamiento en una estructura multidimensional optimizada para la consulta.

GIT Y GITHUB

Git es un sistema de control de versiones distribuido que permite a los desarrolladores rastrear y gestionar cambios en el código fuente durante el desarrollo de software. Facilita la colaboración al permitir que múltiples desarrolladores trabajen en diferentes partes de un proyecto simultáneamente sin interferencias.

GitHub es una plataforma basada en la web que utiliza Git para alojar repositorios de código. Proporciona herramientas adicionales para la colaboración, como la revisión de código, seguimiento de problemas y wikis. GitHub facilita el trabajo en equipo, permitiendo a los desarrolladores compartir, revisar y mejorar el código de manera conjunta en un entorno centralizado y accesible desde cualquier lugar.

ENLACE DE GITHUB: https://github.com/iAmRivard/desafio-datawherehouse