**17/septiembre/2025**

**Proyecto #1: Introducción a la ciencia de datos – Documentación**

**Juan Camilo Gutierrez Viveros – 2159874**

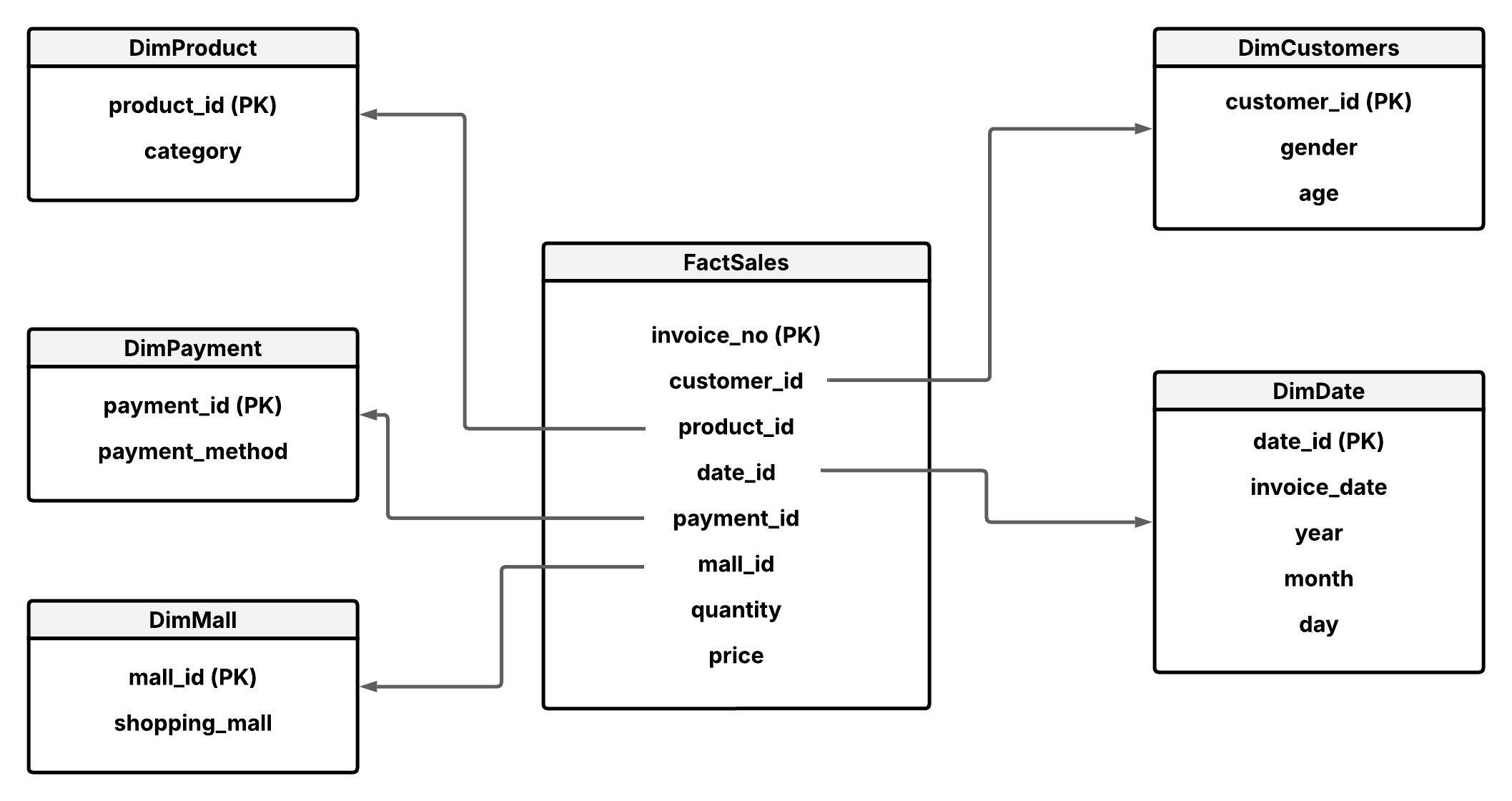
**Andrés Felipe Rojas – 2160328**

**Punto #1 Diseño del Modelo de la Bodega de Datos**

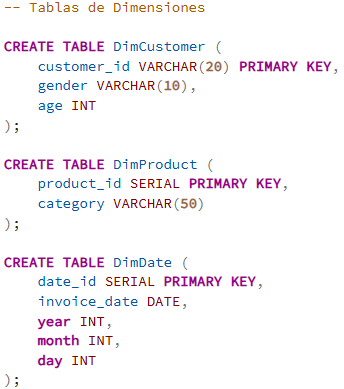
Se puede ver que el dataset de *customer\_shopping\_data.csv* contiene información transaccional de clientes en distintos centros comerciales.

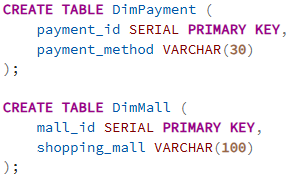
El dataset está orientado a transacciones de compras y ventas, por lo cual se optó por usar el modelo estrella para la bodega de datos. Este modelo es más sencillo de usar y más rápido para consultas de análisis de negocio.

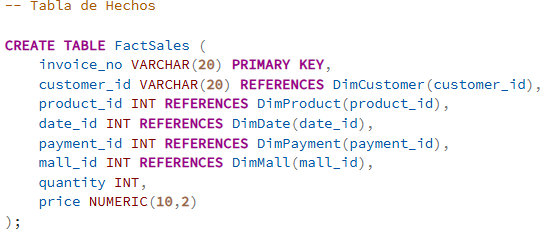
**Diagrama de tablas con modelo estrella:**



Posteriormente se incluyen las tablas en postgres sql para la creación de la bodega con los datos:







**Punto #2**

**Creación de tablas .csv**

### **a) DimCustomer**

* Campos: customer\_id, gender, age
* Se eliminan duplicados para asegurar que cada cliente aparezca una sola vez.
* Archivo generado: dim\_customers.csv

### b) **DimProduct**

* Campos: product\_id, category
* Se genera una **clave sustituta** product\_id asignando un identificador incremental a cada categoría de producto.
* Archivo generado: dim\_products.csv

### c) **DimPayment**

* Campos: payment\_id, payment\_method
* Se asigna un identificador único a cada método de pago.
* Archivo generado: dim\_payments.csv

### d) **DimMall**

* Campos: mall\_id, shopping\_mall
* Se asigna un identificador único a cada centro comercial.
* Archivo generado: dim\_malls.csv

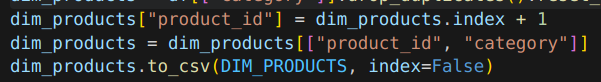
### e) **DimDate**

* Campos: date\_id, invoice\_date, year, month, day
* Se convierten las fechas al tipo datetime.
* Se genera un identificador numérico date\_id con el formato YYYYMMDD.
* Archivo generado: dim\_dates.csv

**Eliminación de duplicados**

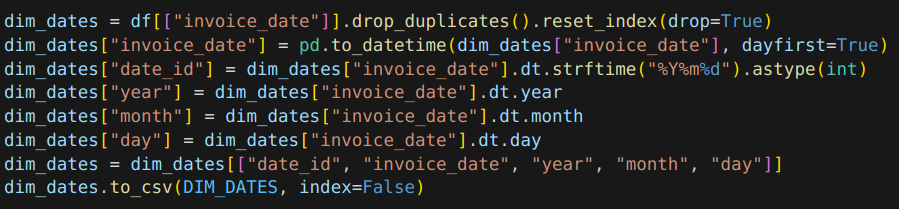
ademas de la generación de archivos .csv por separado para las tablas, dentro de estas mismas se hizo la adaptación de fechas y la eliminación de duplicados, todas tienen lineas con el siguiente formato:

con esto, se asegura de que solo haya una fila por entidad:



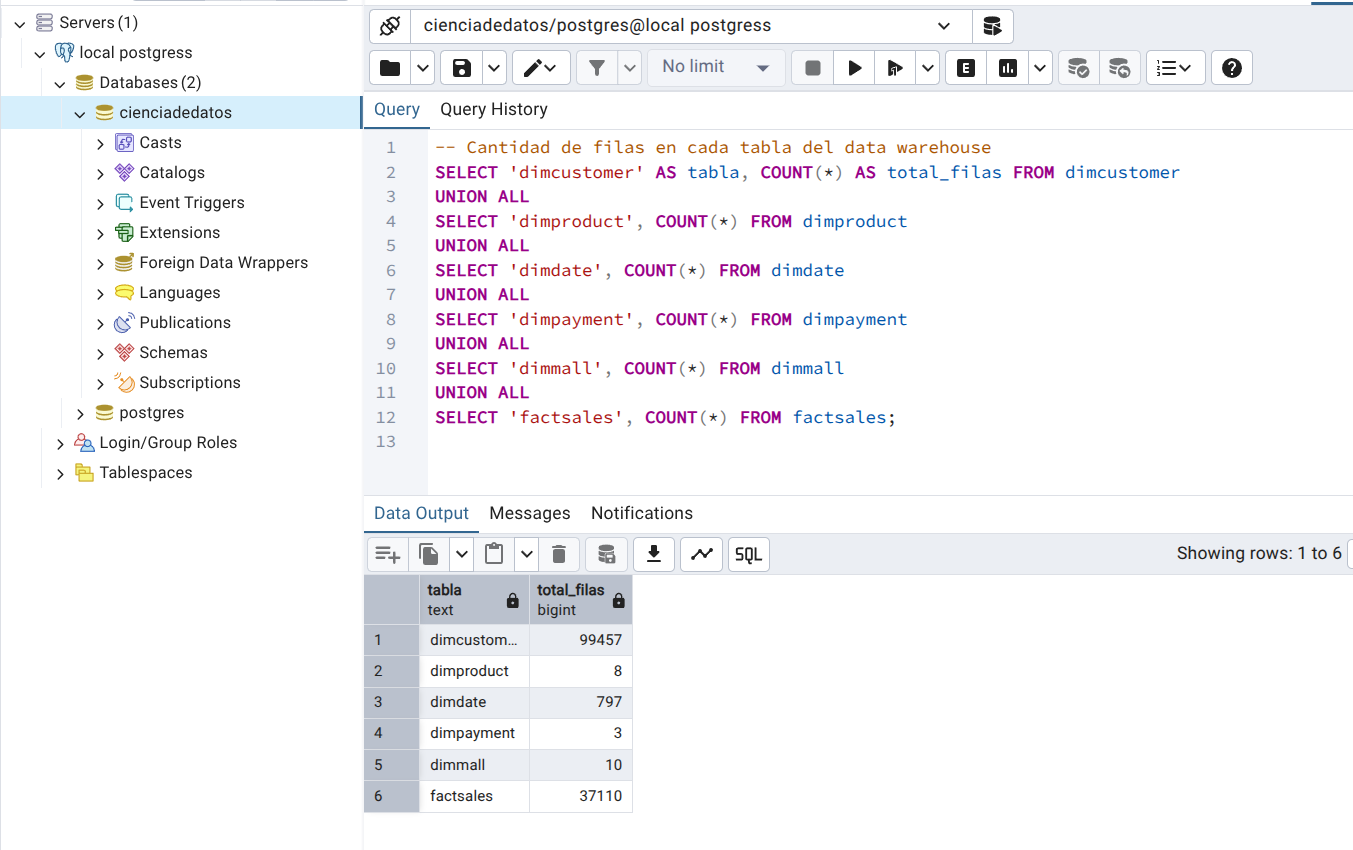
**Estandarización de fechas**

en este bloque de codigo las fechas pasaron de texto plano a formato “fecha”/Datetime en formato año/mes/dia lo que permite usarse la dimensión en analisis a futuro:

****

**Comprobación de carga:**

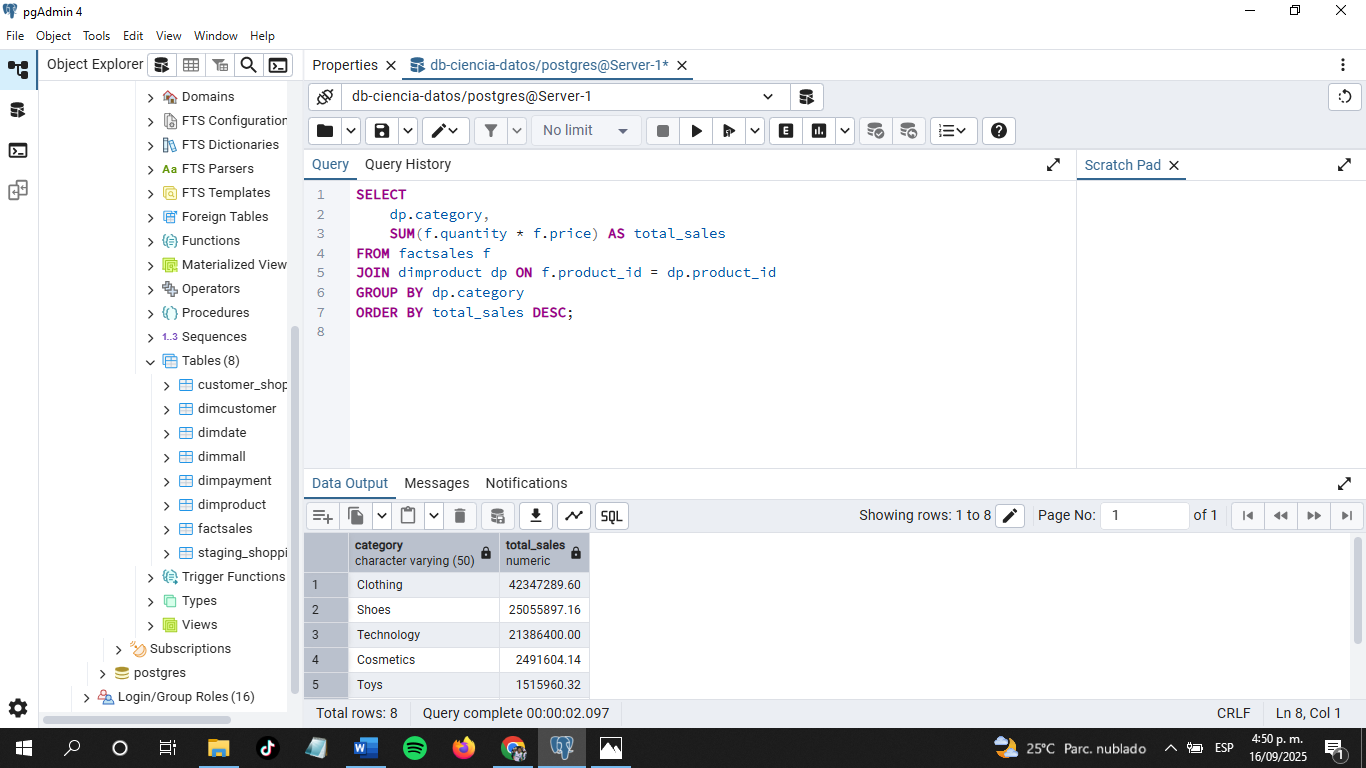
con los datos ya insertados en la BD, una simple consulta nos permite ver la cantidad de filas y asi asegurarnos de que los datos estén correctamente insertados

****

**Punto#3 Consultas Analíticas en SQL**

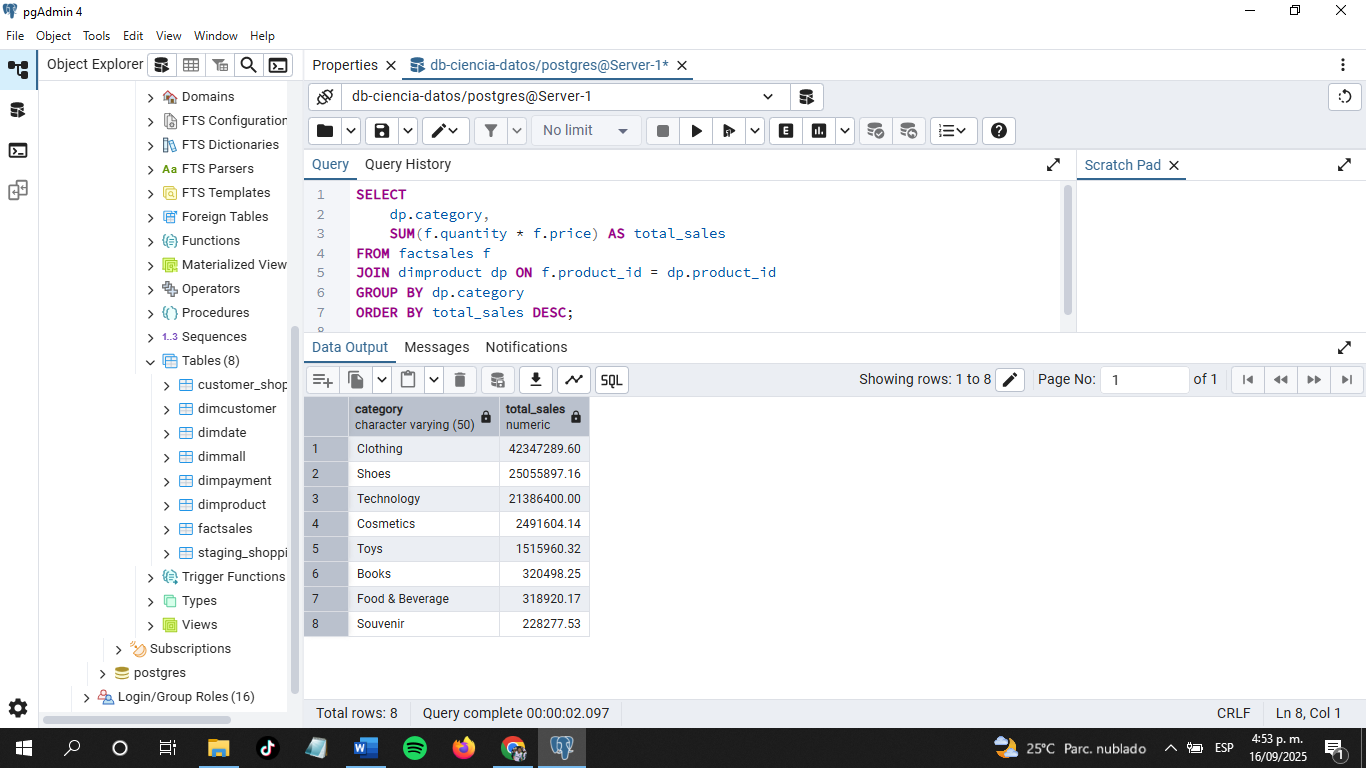
Se diseñas consultas en SQL para responder las siguientes preguntas de negocio:

**Total de ventas por categoría de producto:**

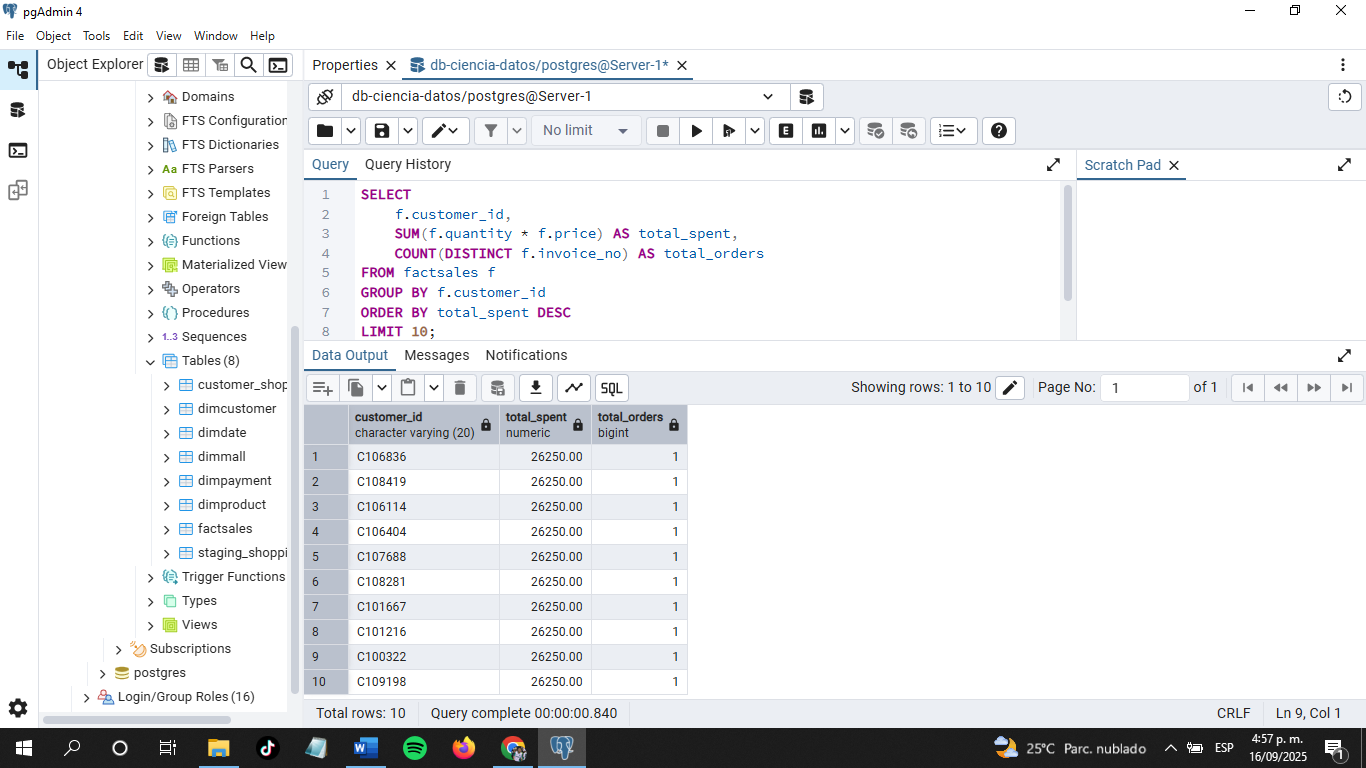


En esta consulta se calcula las ventas totales (precio por cantidad) agrupadas por categoría de producto y también organiza de forma descendente las categorías que generan más ingresos.

**Resultado:**

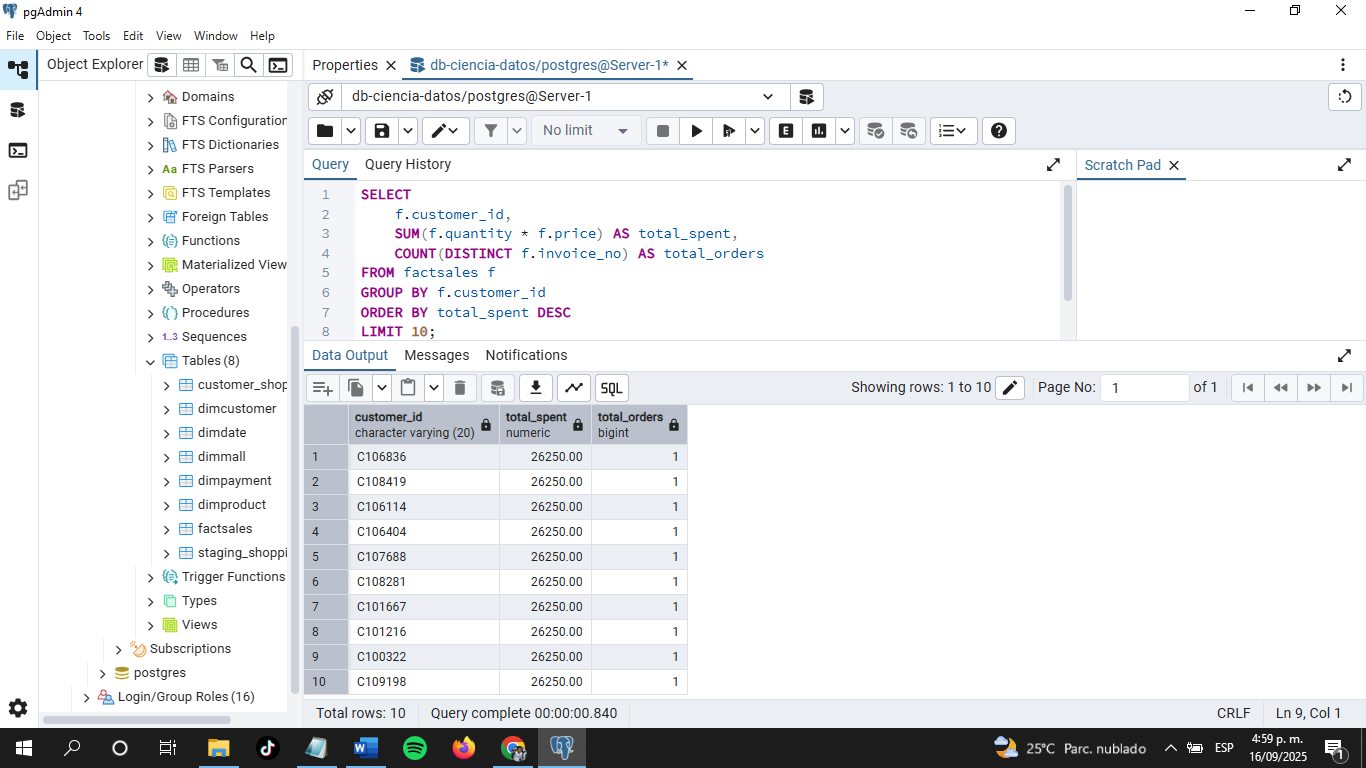


**Clientes con mayor volumen de compra:**

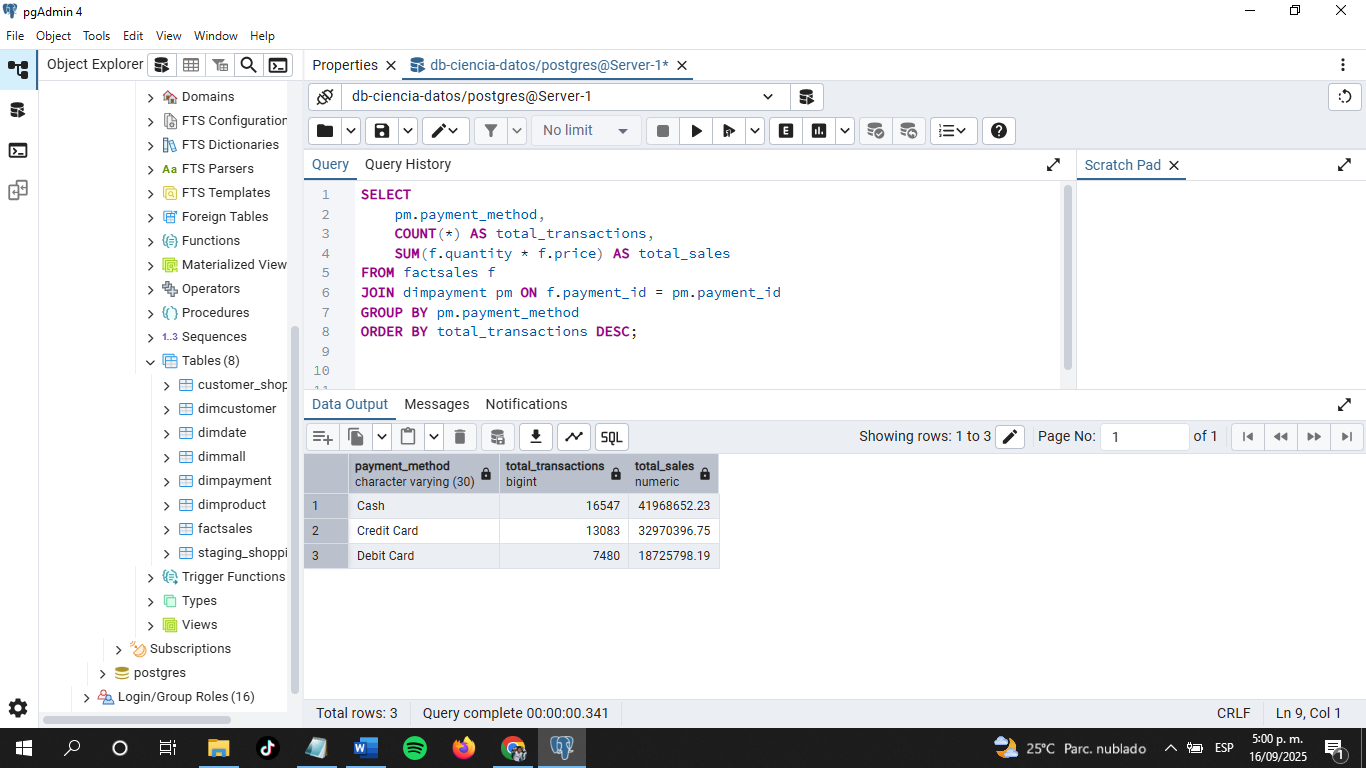


La consulta identifica los clientes más valiosos por monto gastado. Además, muestra cuántas compras hicieron (facturas distintas).

**Resultado:**

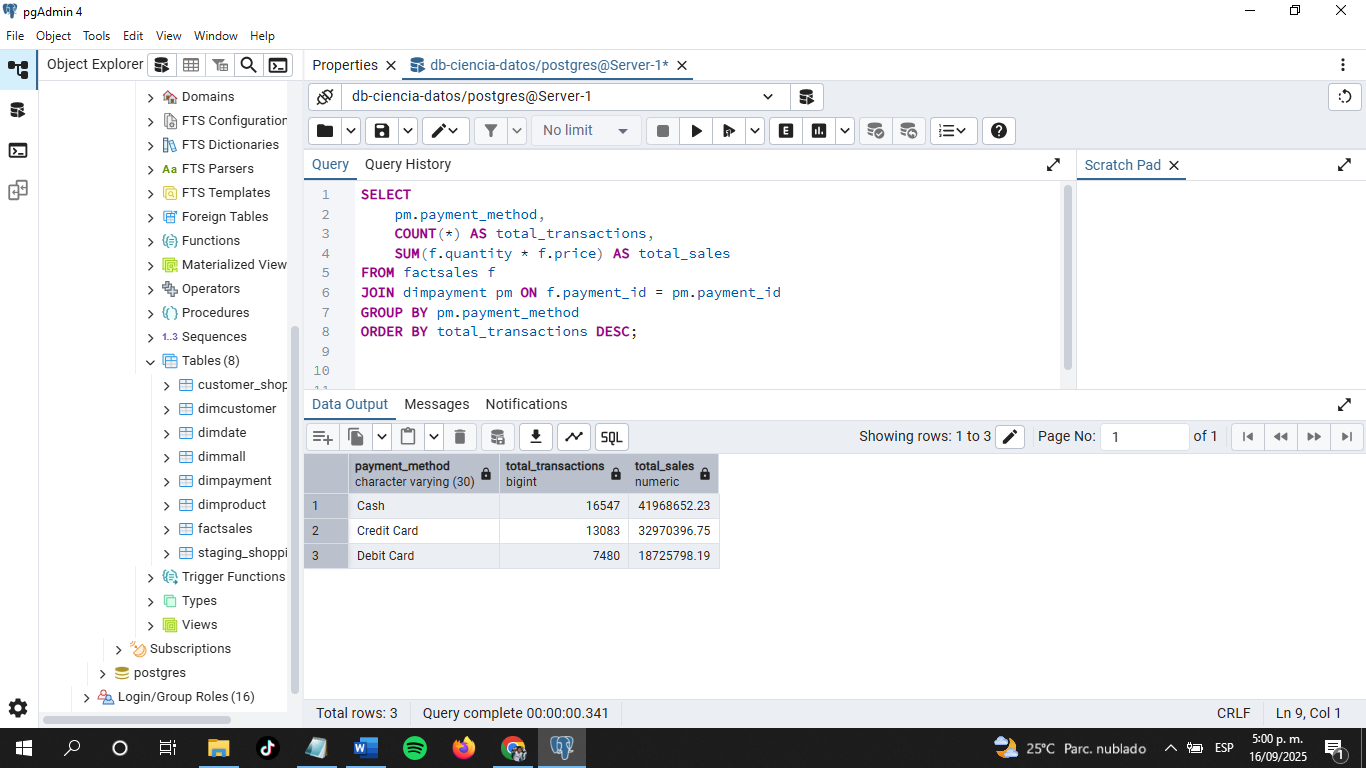


**Métodos de pago más utilizado:**

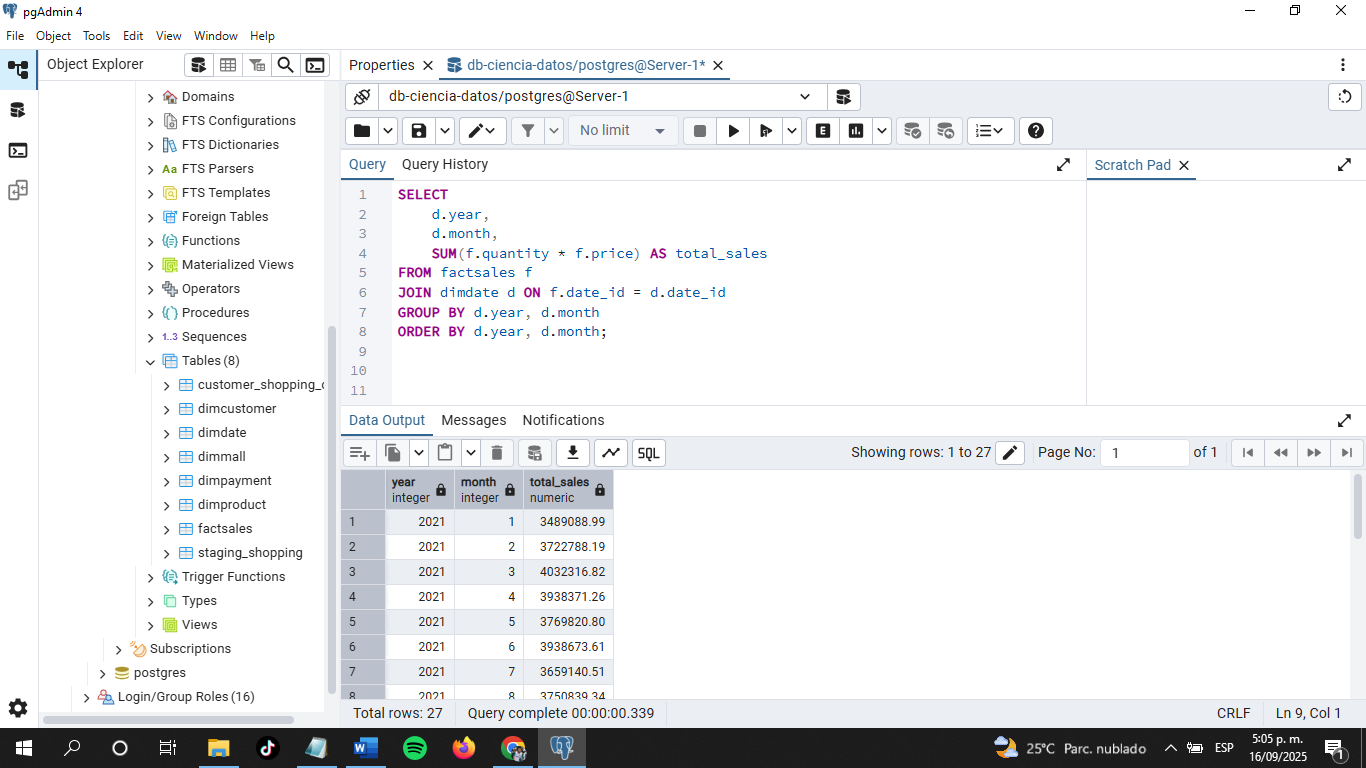


Esta consulta posiciona los métodos de pago más usado y su total en ventas.

**Resultado:**

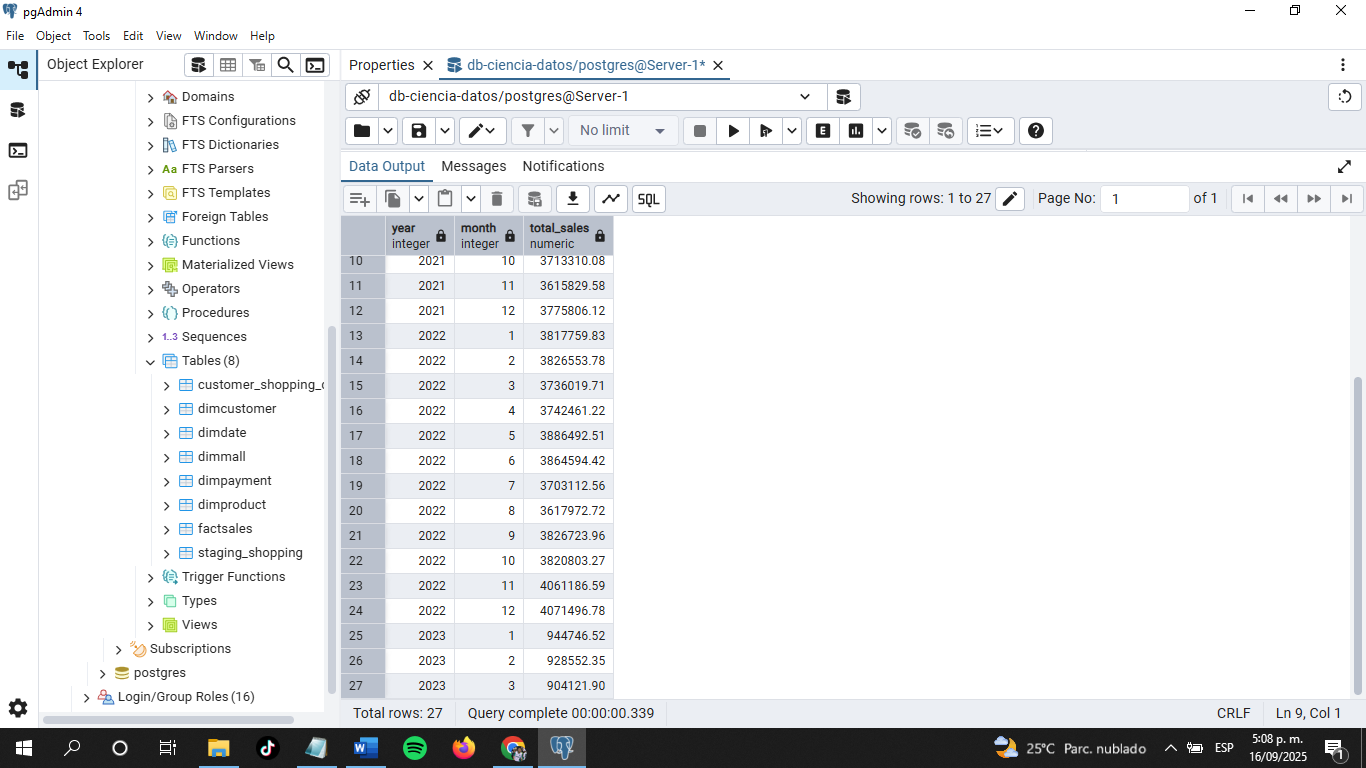
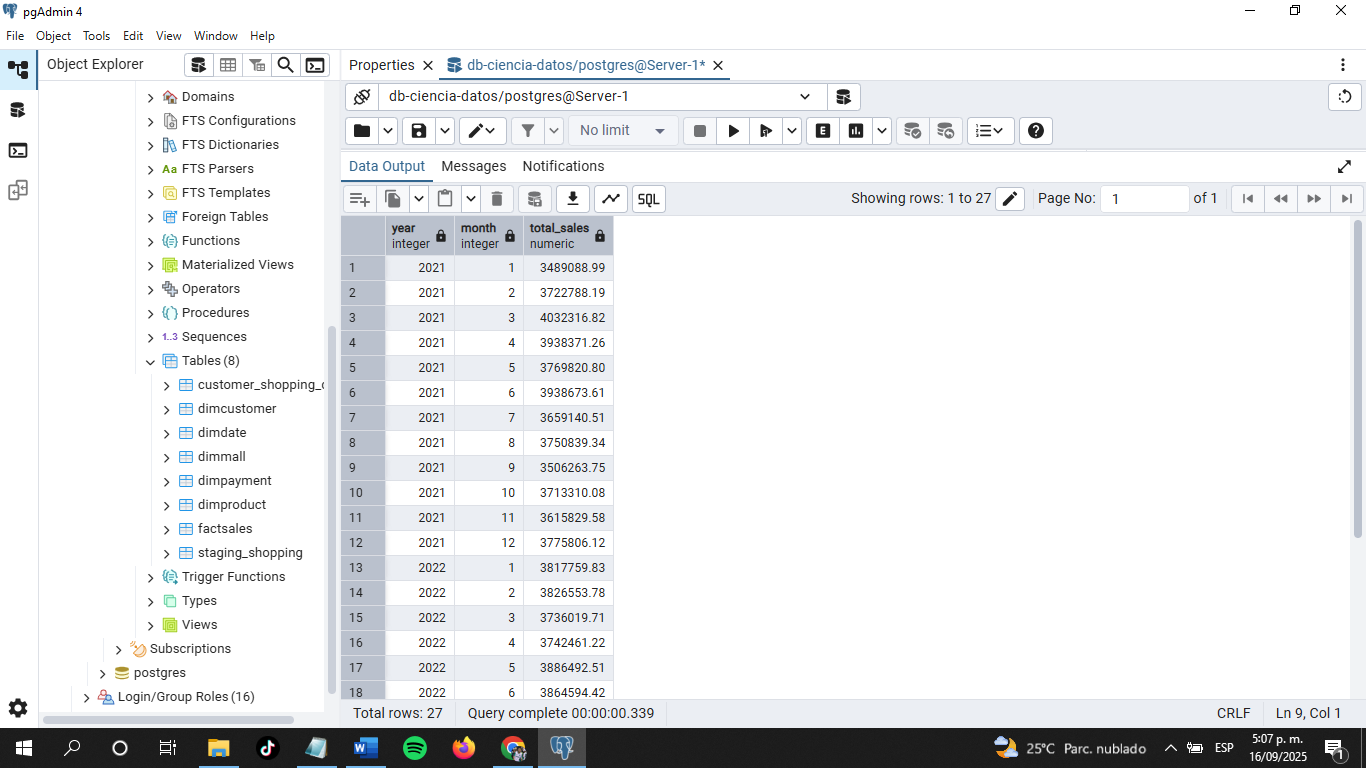


**Comparación de ventas por mes:**

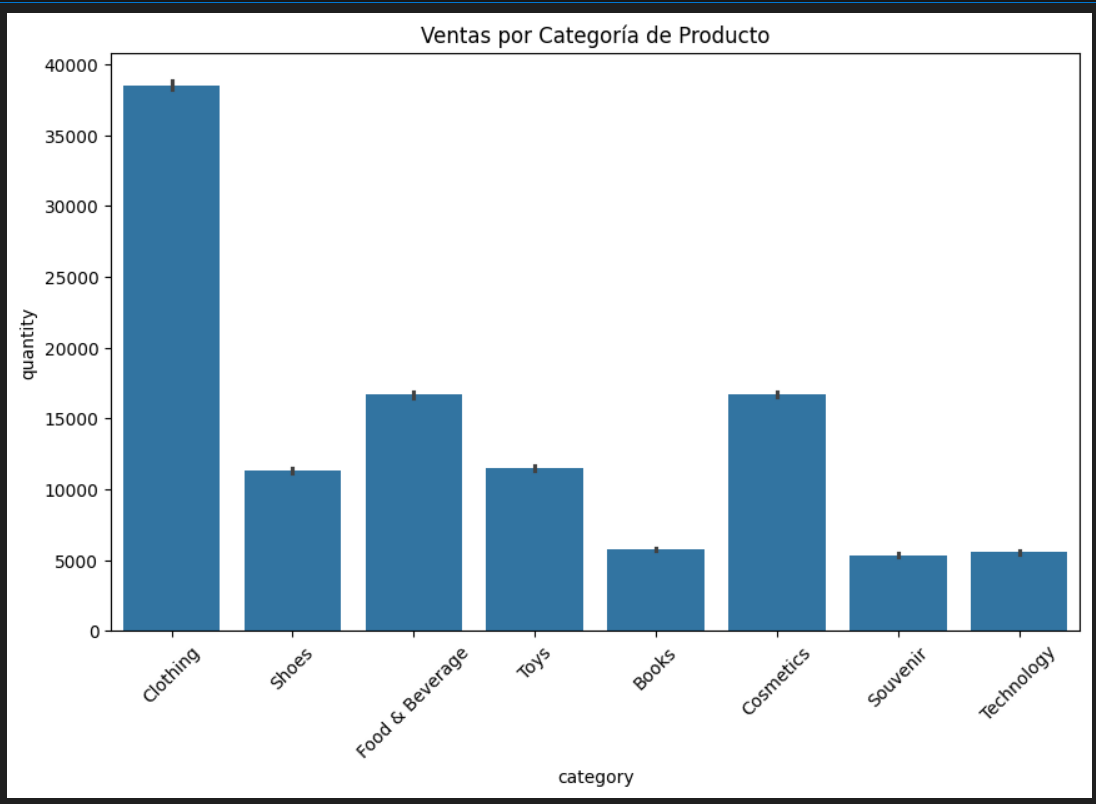


La consulta agrupa las ventas por año y mes permitiendo analizar la evolución temporal y detectar tendencias.

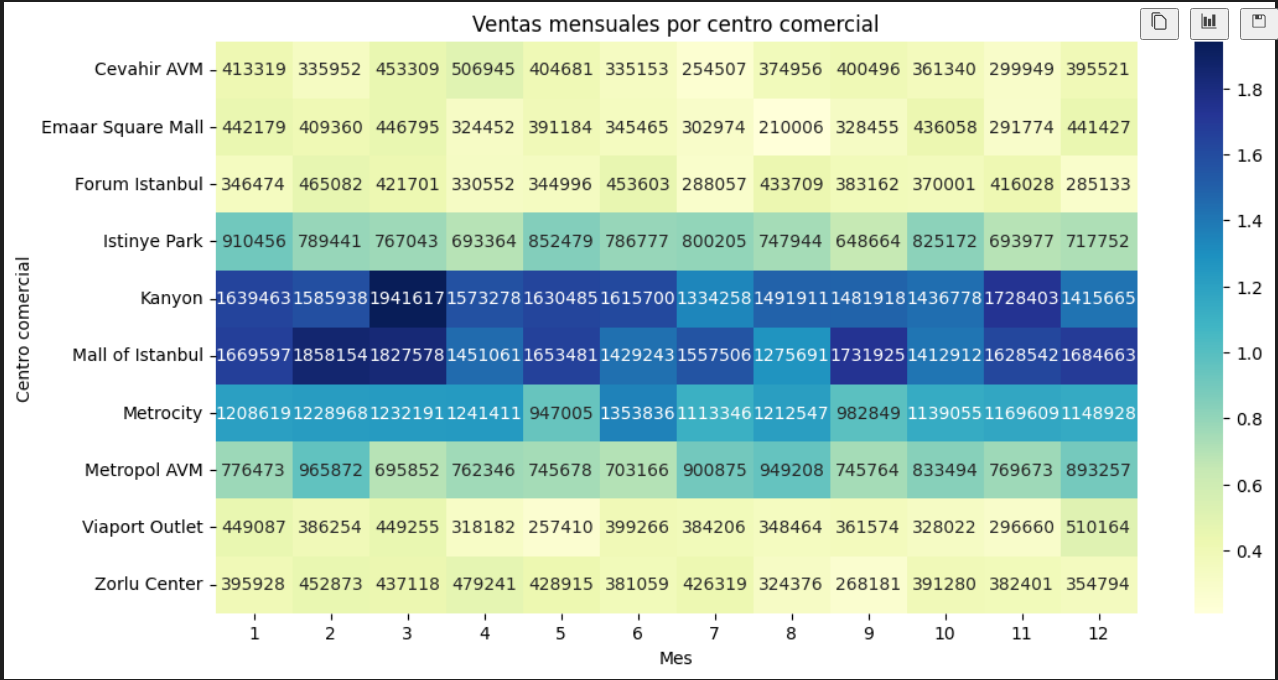
**Resultado:**



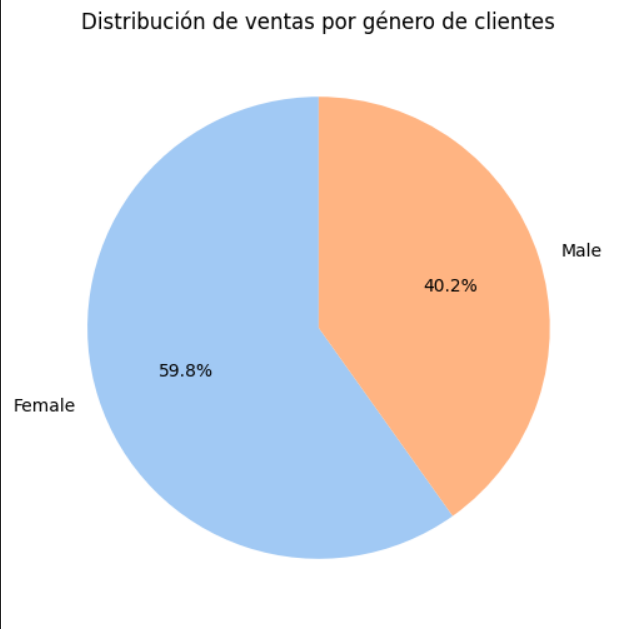
**#4 Análisis descriptivo**

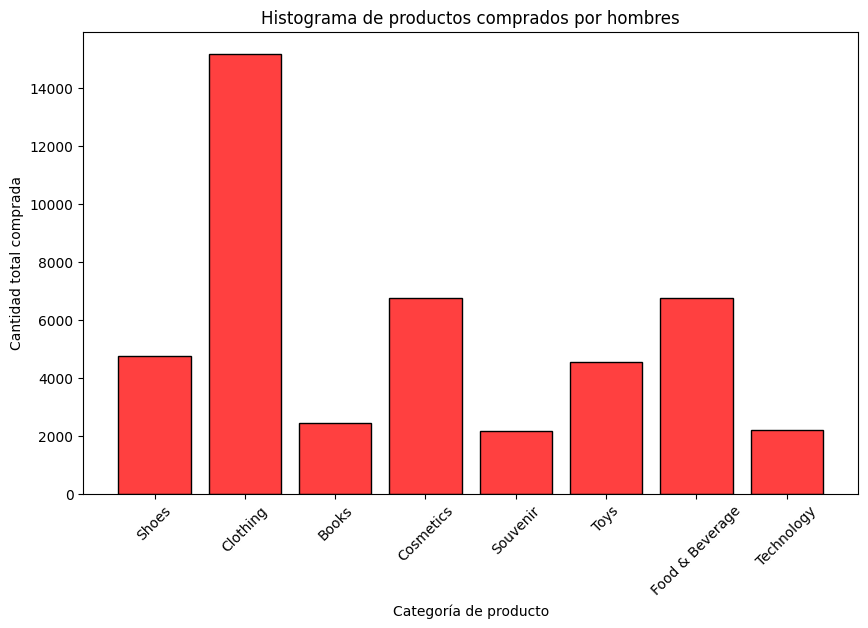
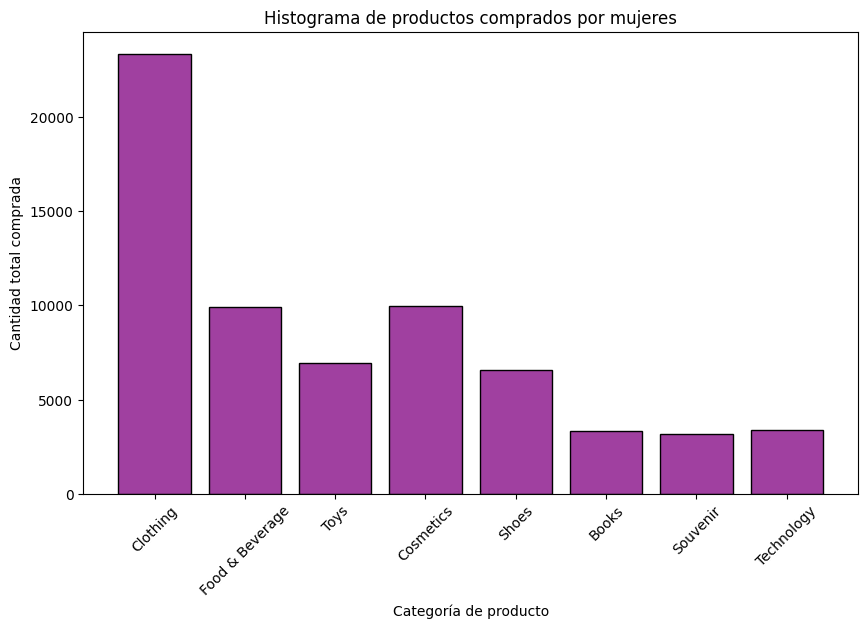
**- Gráfica en base a la cantidad de productos vendidos:**en esta gráfica podemos ver que la ropa de todo tipo es la mas vendida mientras que la tecnología y souvenir son los menos relevantes

**- Grafico de calor ventas mensuales por centro comercial**

mapa de calor diseñado para ver que centro comercial vende mas por mes en toda la base de datos, en este caso, un color mas fuerte significa un mayor éxito y volumen de ventas

**- Diagrama de pastel comparativo por genero**

****

****

para el analisis de estos datos no se llegan a demasiados resultados en cuanto a gustos por diferencia de genero, respetando los porcentajes.

**- Conclusiones y mejora:**

- Con los anteriores analisis se pudo notar que la diferencia de genero no modifica realmente los numeros vendidos, por los que podrian concentrarse en potenciar su producto estrella a sabiendas de que sus clientes estan interesados.

- Las 3 categorias menos vendidas que son “books” , “souvenir” y “technology” podrian desaparecer sin afectar de forma significativa sus ventas y mejorar las ganancias generales al sacar de stock productos que no llaman la atencion

- Sus ventas se centran principalmente en 3 establecimientos principales, por lo que campañas de marketing para promocionar otras localizaciones podrian ayudar a las ventas en los demas con menos volumen de clientes. Sobretodo con el conocimiento adquirido de los productos mas solicitados.