Project: Ecommerce Sales & Customer Insights

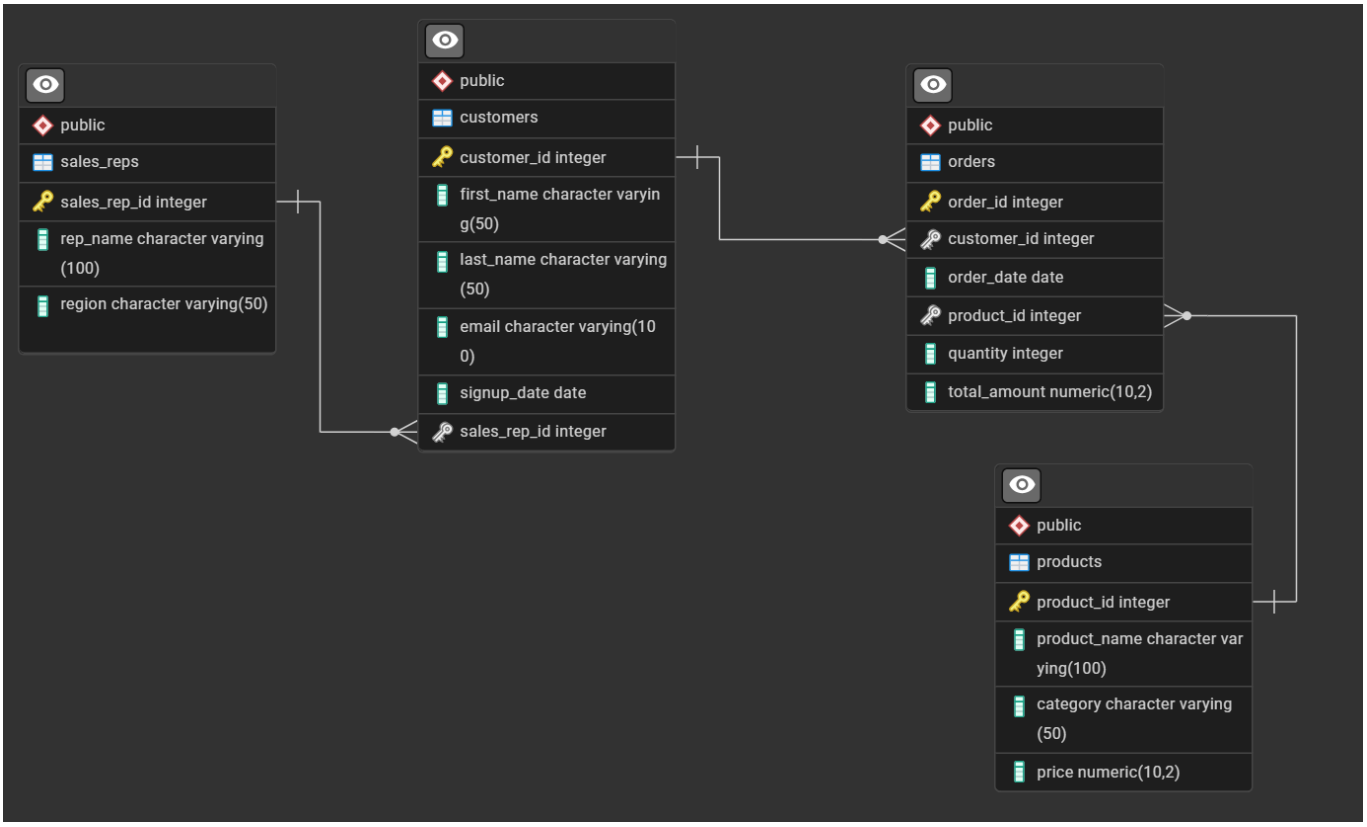
Project Description:

SQL-based project analyzing e-commerce sales and customer data to uncover actionable insights. The analysis identifies top-selling product categories, high-spending customers and regions, and segments customers by purchasing power (High, Medium, and Low Value).

What was the Scenario of the project?

I have been hired as a Data Analyst for MegaMart Online, a global e-commerce platform. My task was to analyze customer behavior, sales trends, and product performance to provide actionable insights for the marketing and operations teams.

The ERD Diagram for the Project



About the dataset:

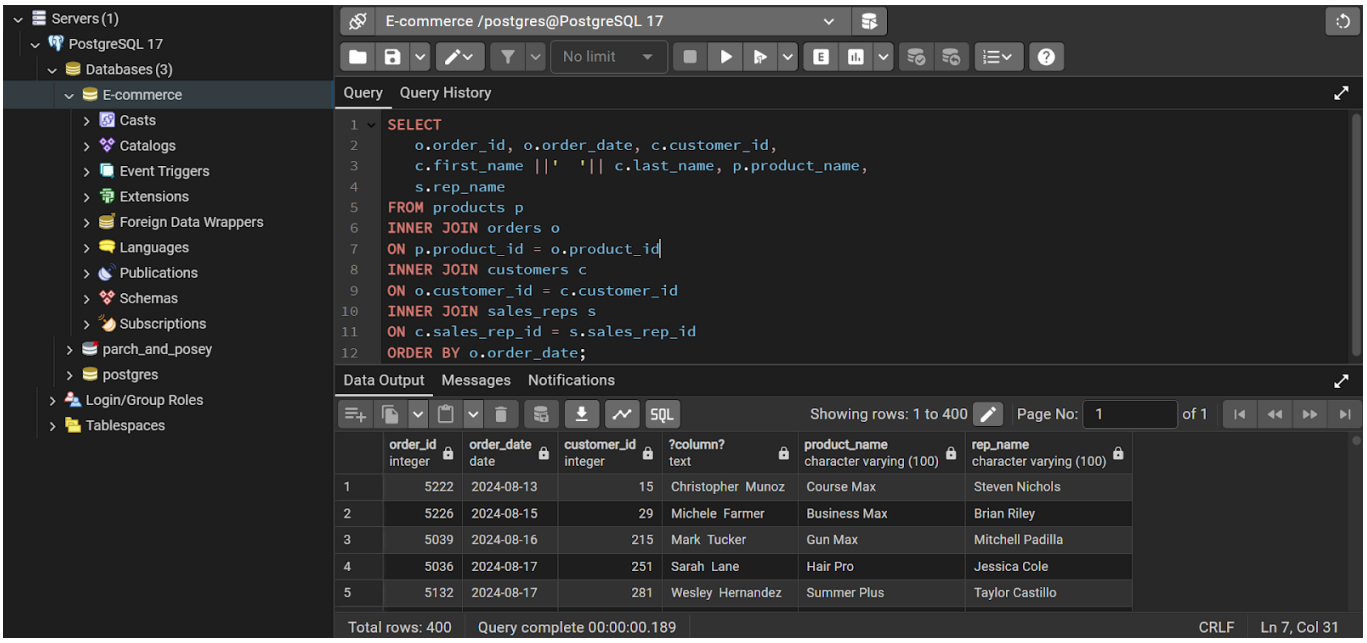
1. customers — Information about registered customers.
2. orders — Records of orders placed.
3. products — Information about the items sold.
4. sales_reps — List of sales representatives managing certain customers

Project Questions:

- Which product categories generate the most revenue?
- Who are the highest-spending customers and which regions manage them?
- How can customers be segmented based on purchasing power?

The Codes from the Analysis

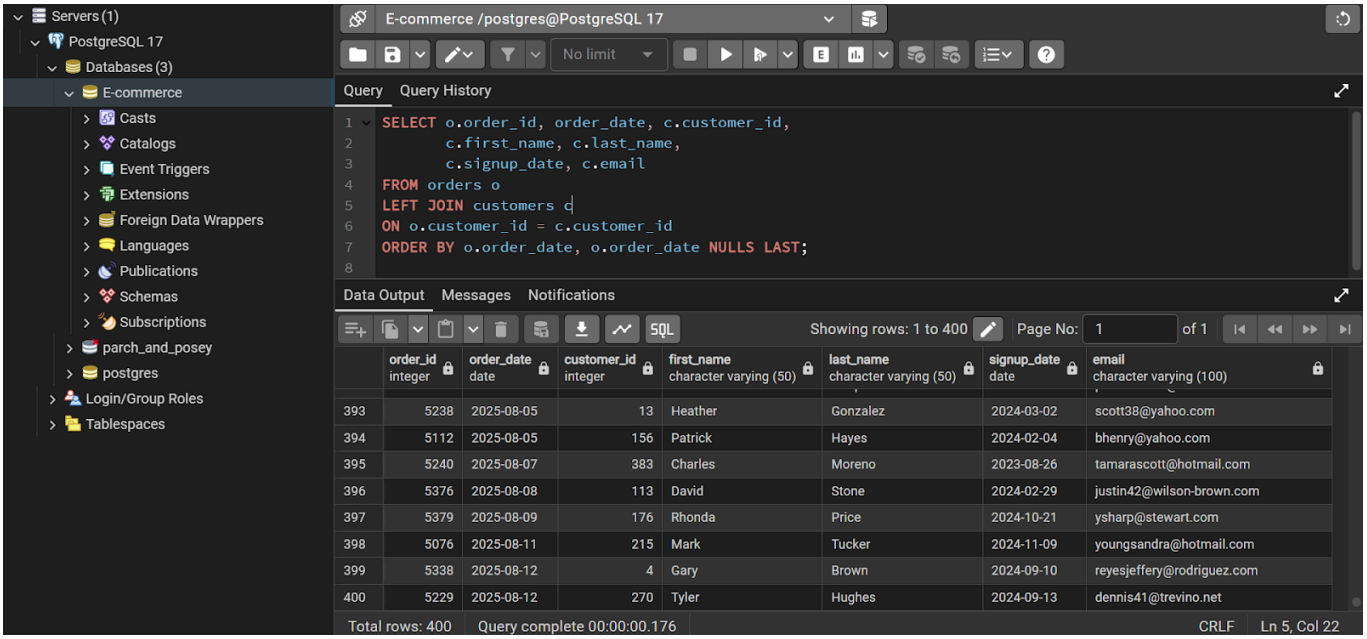
1. **INNER JOIN:** Retrieve all orders with customer names, product names, and sales rep names, and the highest grossing product.



Why I used INNER JOIN:

I used an INNER JOIN to combine the `products` table with `orders`, `customers`, and `sales_reps`, so that instead of just IDs, I could see the actual customer names, product names, and sales rep names. This gives a complete picture of each order and makes the data more useful for reporting and analysis, like tracking top products, top customers, or sales performance by rep.

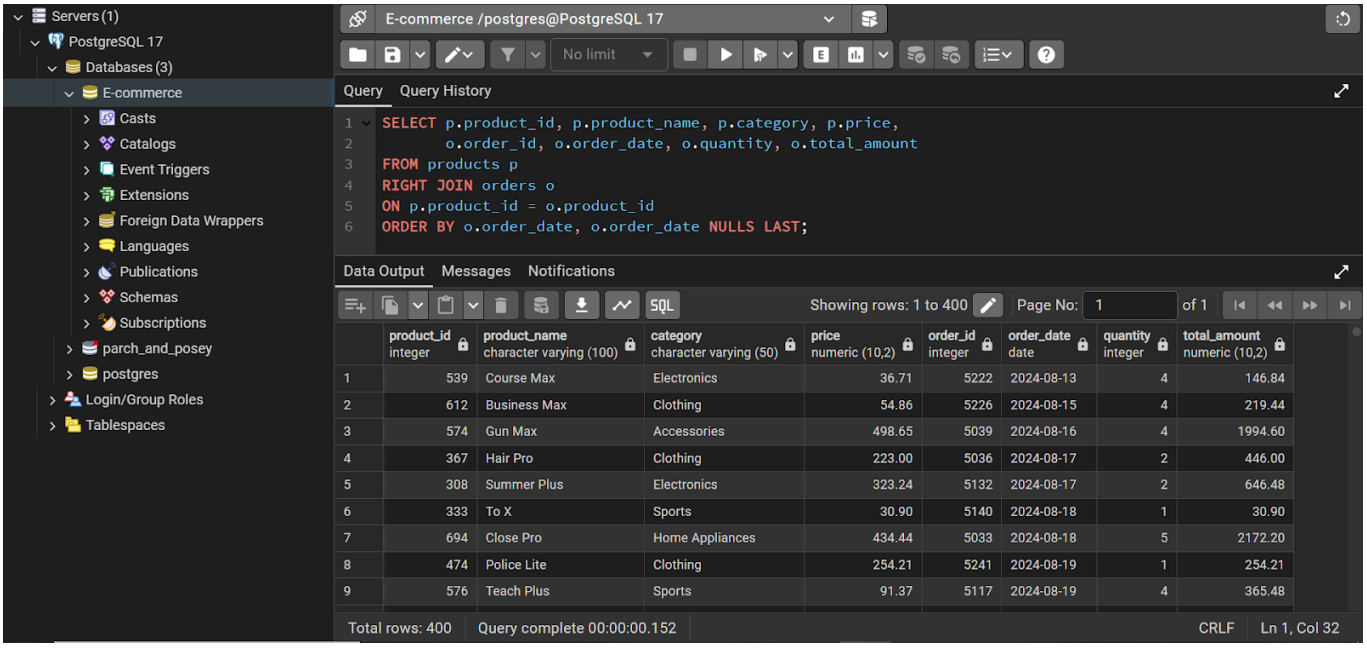
2. **LEFT JOIN:** List all customers with their order details, including customers who have never



Why I used LEFT JOIN:

I used a **LEFT JOIN** to make sure I could list **all customers** along with their order details, while also including those customers who have **never placed an order**. This way, I don't lose any customers in my analysis. I can see both active and inactive ones, which is useful for understanding customer engagement and identifying those who may need follow-ups or promotions.

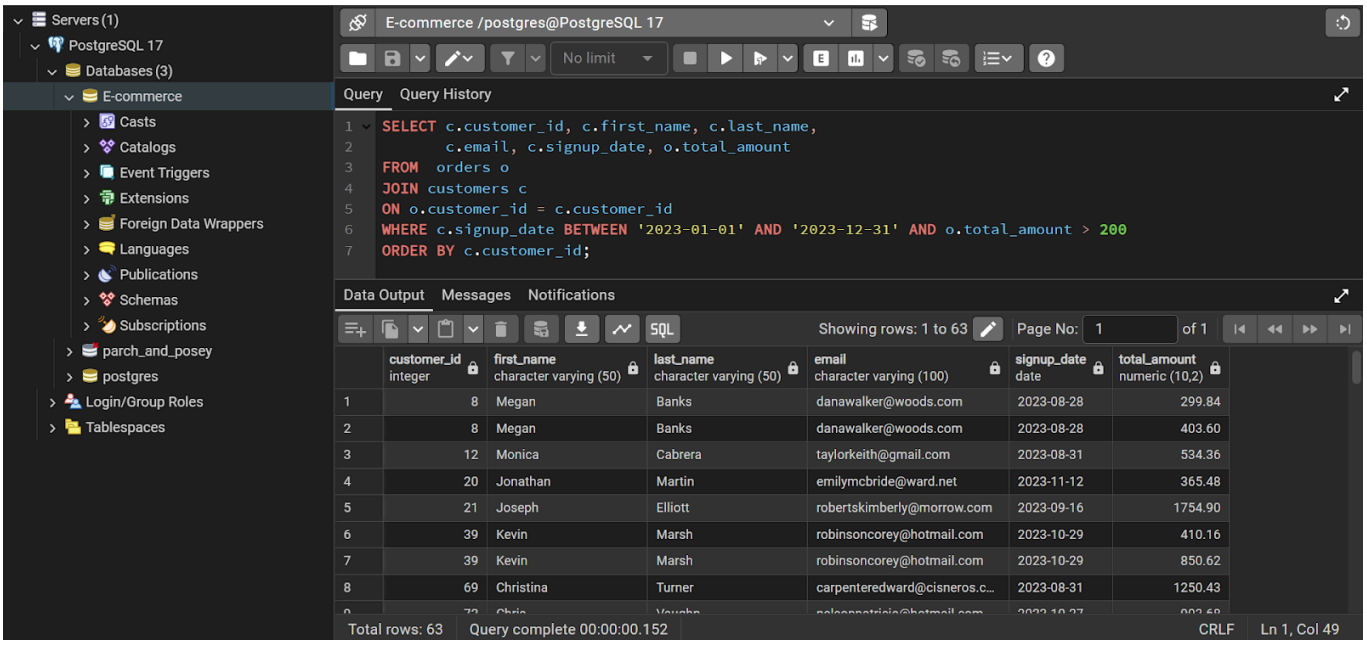
3. **RIGHT JOIN:** Show all products and the orders placed for them (including products never sold).



Why I used RIGHT JOIN:

I used a RIGHT JOIN to display **all products** along with any orders linked to them, while also making sure products that have **never been sold** still appear. This is important for analyzing product performance, because it highlights not only the top-selling items but also those with little or no sales, which can guide inventory and marketing decisions.

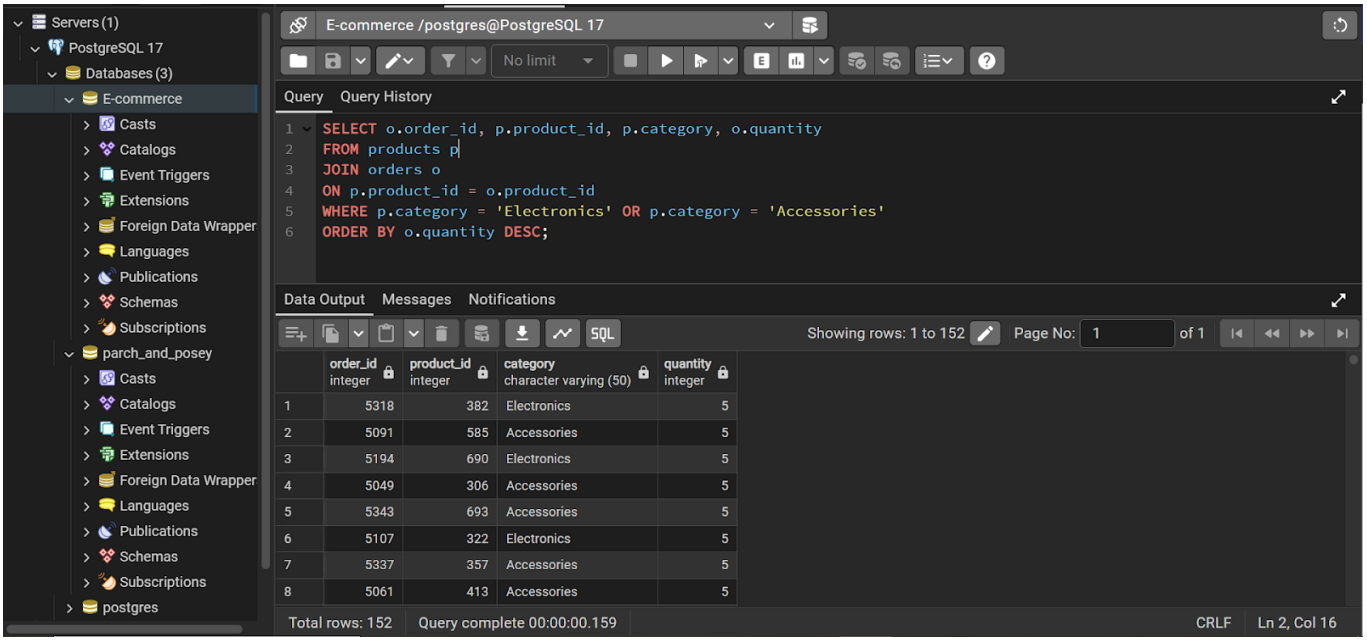
4. Find customers who **signed up in 2023 AND** have placed at least one order worth more than $200



Why I wrote this query:

I used this query to find customers who **signed up in 2023** and also made at least one **high-value order above $200**. Filtering by signup date ensures I'm only looking at new customers, and the order amount condition helps me identify those who quickly became valuable. This is useful for analyzing the quality of recent signups and targeting early high-spending customers for loyalty programs.

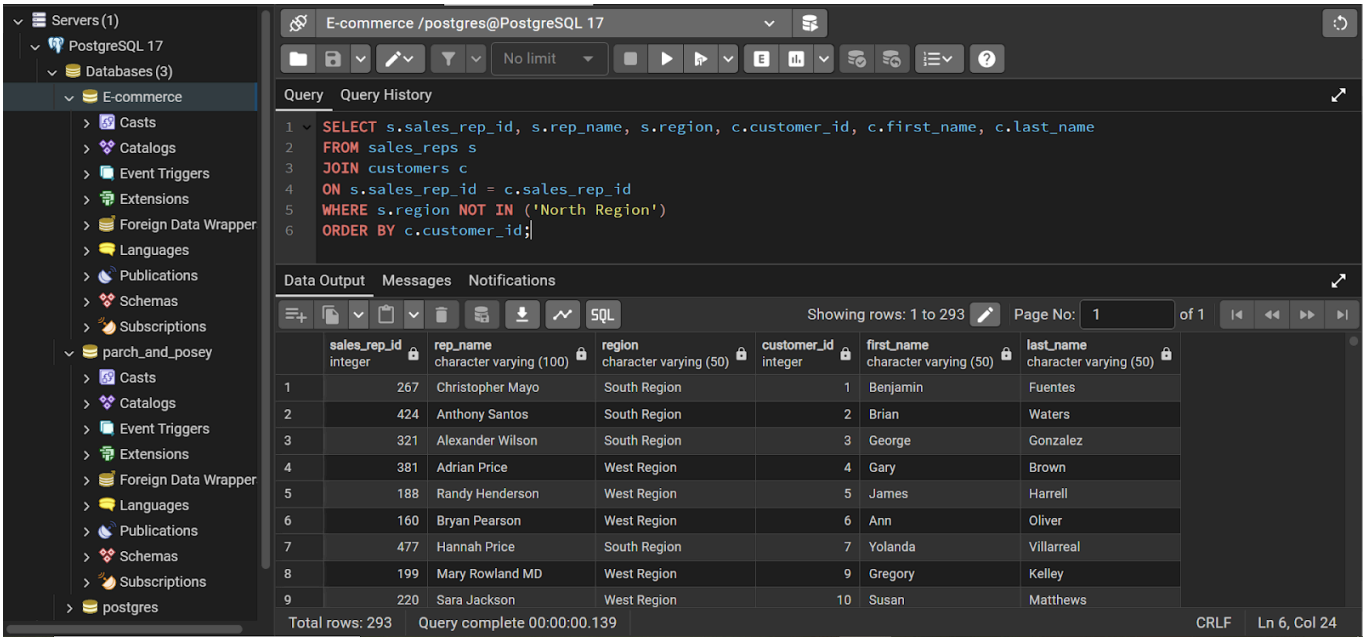
5. List orders where the product category is "Electronics" **OR** "Accessories".



Why I wrote this query:

I used this query to list orders where the products fall under **Electronics or Accessories**, because these categories are key revenue drivers. Filtering this way makes it easier to track performance, compare demand between the two, and identify which items sell the most within these popular categories

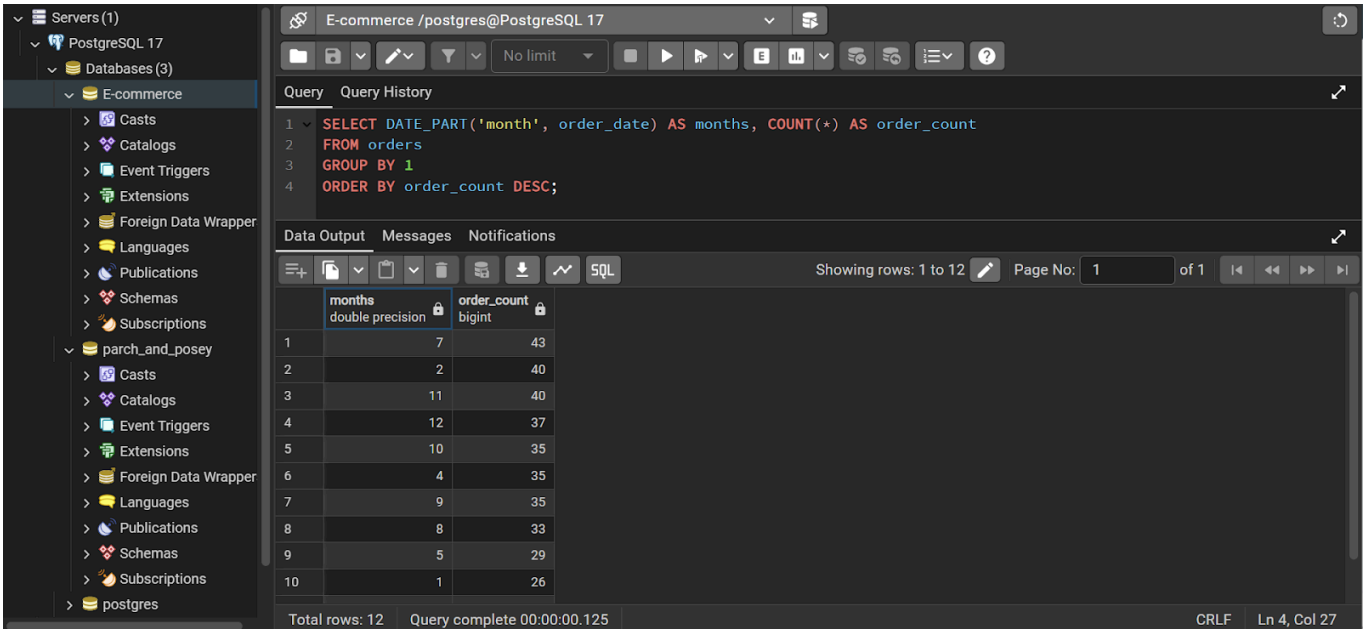
6. Show customers whose sales reps are **NOT** assigned to the "North Region".



Why I wrote this query:

I used this query to find customers whose sales reps are **not assigned to the North Region**. This helps compare customer distribution and sales activity across other regions, ensuring that analysis isn't limited to just one area and making it easier to spot differences in performance between regions.

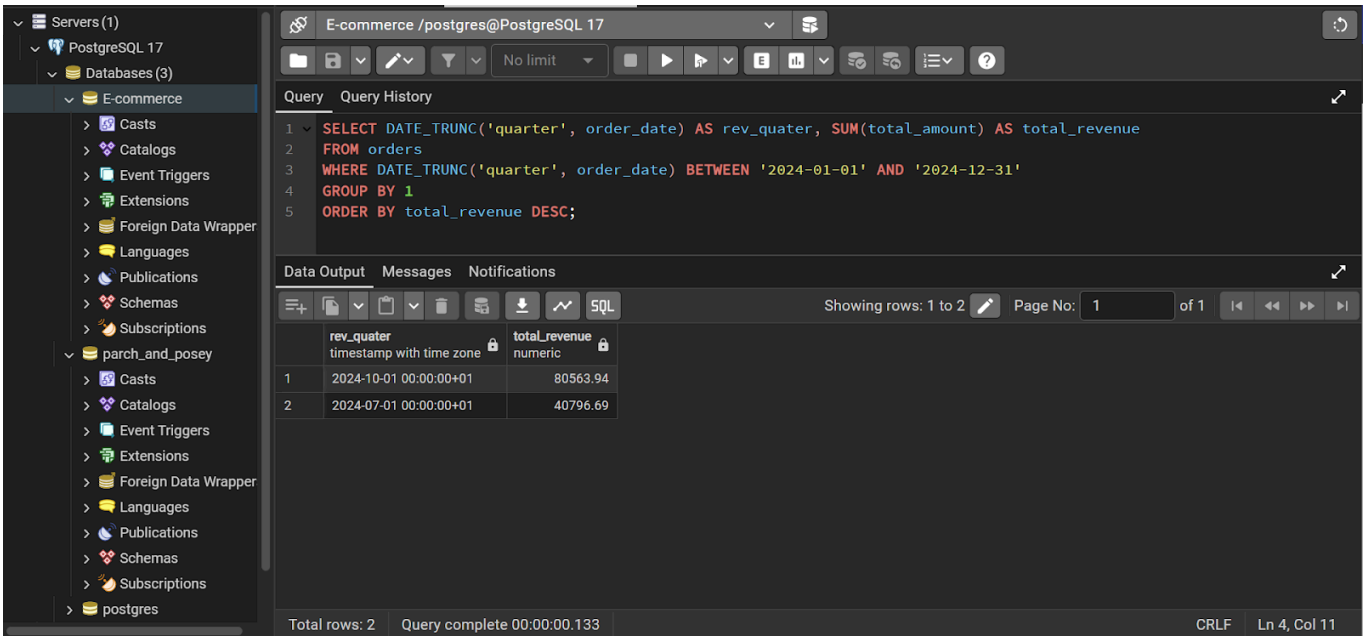
7. **DATE_PART:** Count how many orders were placed in each **month** of 2024.



Why I wrote this query:

I used **DATE_PART** to count how many orders were placed in each **month of 2024**. This breaks down sales activity month by month, making it easier to spot seasonal trends, peak periods, and months with low performance that may need attention.

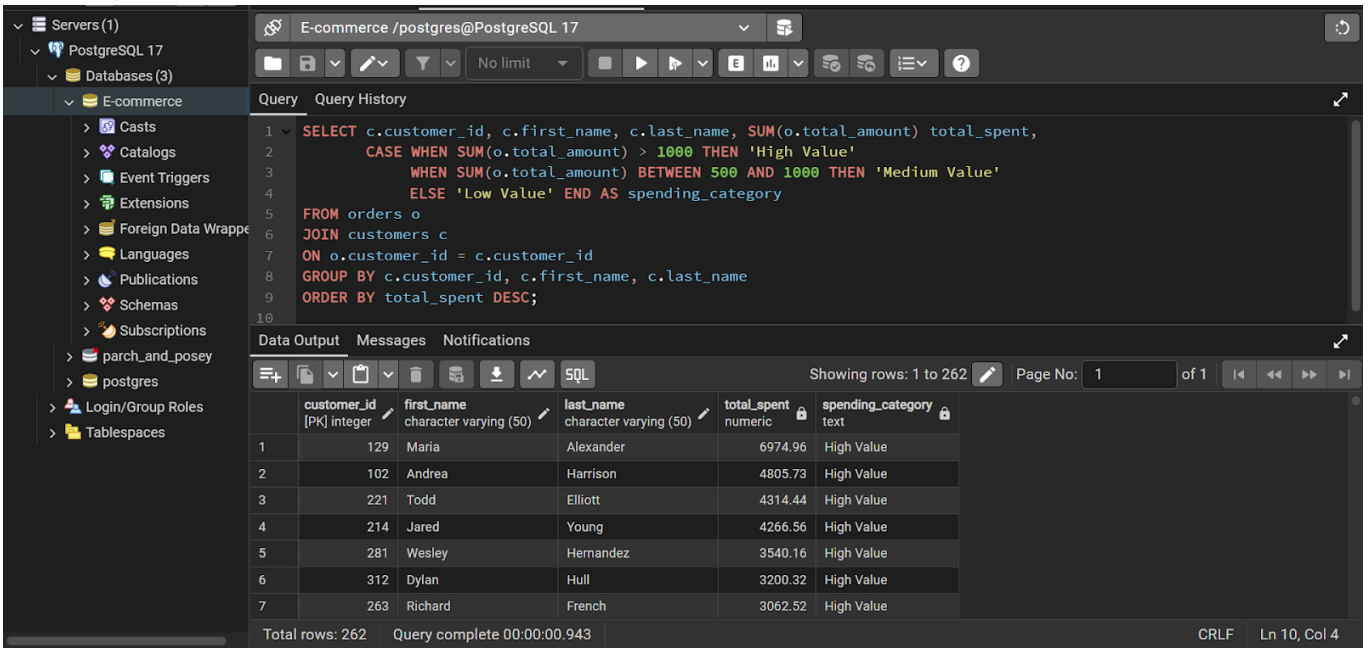
8. **DATE_TRUNC:** Calculate total revenue for each **quarter** in 2024.



Why I wrote this query:

I used **DATE_TRUNC** to calculate the total revenue for each **quarter in 2024**. Looking at revenue by quarter gives a clearer view of broader sales performance, highlights growth patterns over time, and helps compare business results across different parts of the year.

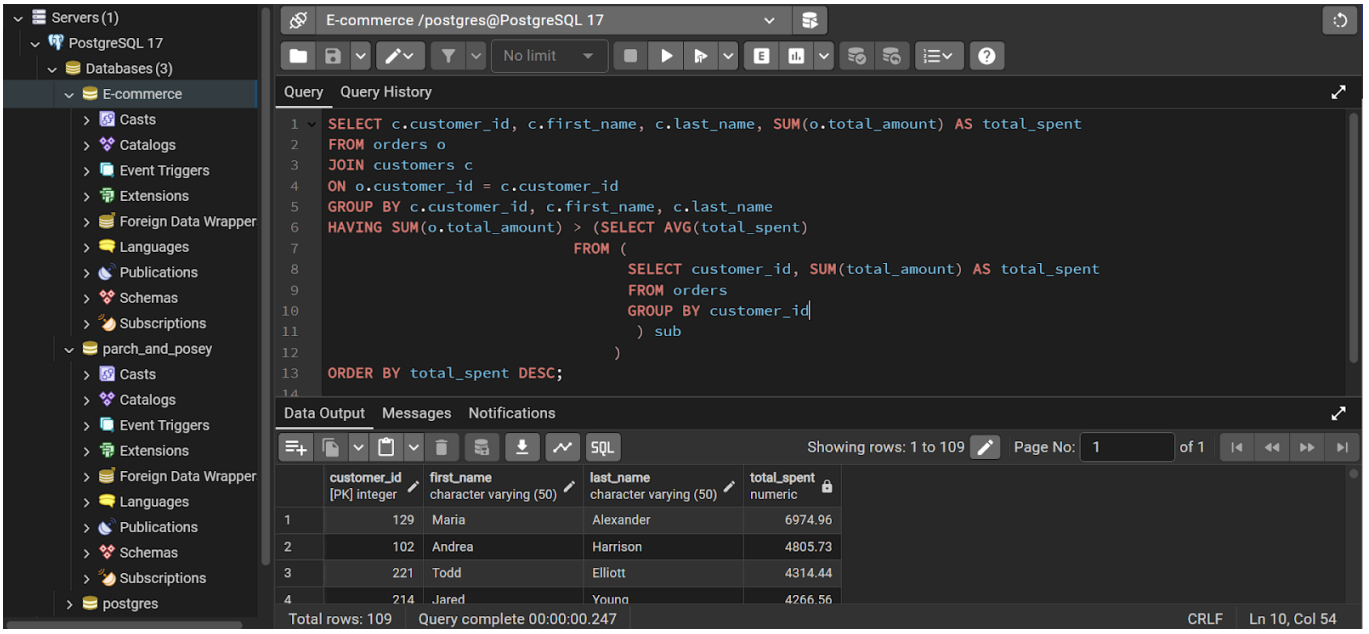
9. Create a column **spending_category** to classify customers based on their total spending:



Why I wrote this query:

I created a `spending_category` column using a CASE statement to classify customers based on their **total spending**. This makes it easy to separate customers into **High, Medium, and Low Value** groups, which is useful for targeted marketing, loyalty programs, and understanding which customer segments drive the most revenue.

10. Find all customers whose **total spending** is above the **average spending** of all customers.



Why I wrote this query:

I used a subquery to compare each customer's total spending against the **average spending of all customers**. This helps identify the **above-average spenders**, who are valuable for loyalty programs, upselling, or personalized offers since they contribute more than the typical customer.

INSIGHTS FROM MY PROJECT ANALYSIS

1. Top-Selling Categories

Analysis shows that **Accessories** and **Clothing** stand out as the key categories due to the large amount of revenue generated.

- **Accessories** dominate in terms of sales volume, making them the strongest revenue driver.
- **Clothing** and **Sports** products also contributed significantly, reinforcing their importance in overall sales performance.

2. Using INNER JOIN to Find Top-Grossing Products

I applied an INNER JOIN to connect the `products` table with `orders`, `customers`, and `sales_reps`.

- This allowed me to not only identify the **highest-grossing products** but also see **which customers purchased them** and **which sales reps managed those customers**, giving a complete view of product performance and customer reach.

3. High-Spending Customers & Regions

Using the CASE STATEMENT, I was able to identify the **top-spending customers** and the **regions** that manage them.

- Insight: The **East Region** consistently handles higher-value customers according to the top 5 high value customers in our company, suggesting either stronger customer relationships or greater purchasing power in that region.

Customer Segmentation (Based on Purchasing Power)

I categorized customers into three groups, which are the **High Value**, **Medium Value**, and **Low Value**, based on their total spending.

- This segmentation allows the company to design targeted strategies such as **discounts, loyalty rewards, or subscription plans** for different customer groups.
- Notably, the analysis revealed that the company has a relatively **large proportion of high-value customers**, which is a strong indicator of customer loyalty and spending potential.

IDEAS AND RECOMMENDATIONS FOR MORE SALES IMPROVEMENT

1. Double Down on Accessories & Clothing (Category Focus)

Since **Accessories** dominate in sales volume and **Clothing** contributes significantly to revenue, the company should:

- Expand product variety in these categories.
- Offer bundled deals (e.g., *"Buy Clothing + Accessory and get 10% off"*).
- Run targeted marketing campaigns highlighting best-sellers in these categories.

Benefit: Boosts sales in the strongest-performing categories while encouraging cross-sells.

2. Target High-Value Customers with Loyalty Programs

The analysis showed a **large proportion of high-value customers**. To retain and grow this segment:

- Introduce **exclusive loyalty rewards** (points, VIP discounts, early access to new products).
- Personalized offers via email or app notifications.

Benefit: Keeps big spenders engaged, increases repeat purchases, and reduces churn.

3. Strengthen Sales Strategy in the East Region

The **East Region** consistently manages higher-value customers, signaling strong purchasing power. The company should:

- Increase inventory allocation in this region.
- Invest in **regional promotions** (ads, influencer marketing, or local partnerships).
- Equip sales reps with additional support or incentives to grow accounts further.

Benefit: Maximizes revenue in a proven high-potential region.

4. Re-Engage Low & Medium Value Customers

Customer segmentation revealed that not all customers spend at the same level. To improve sales:

- Offer **discount vouchers** or "first repeat order" incentives for low-value customers.
- Upsell to medium-value customers with subscription plans (e.g., *"Subscribe & Save"* offers).
- Use remarketing ads to bring back inactive customers.

Benefit: Expands overall customer lifetime value by moving more customers into the medium/high-value brackets.