**What I did in sql**

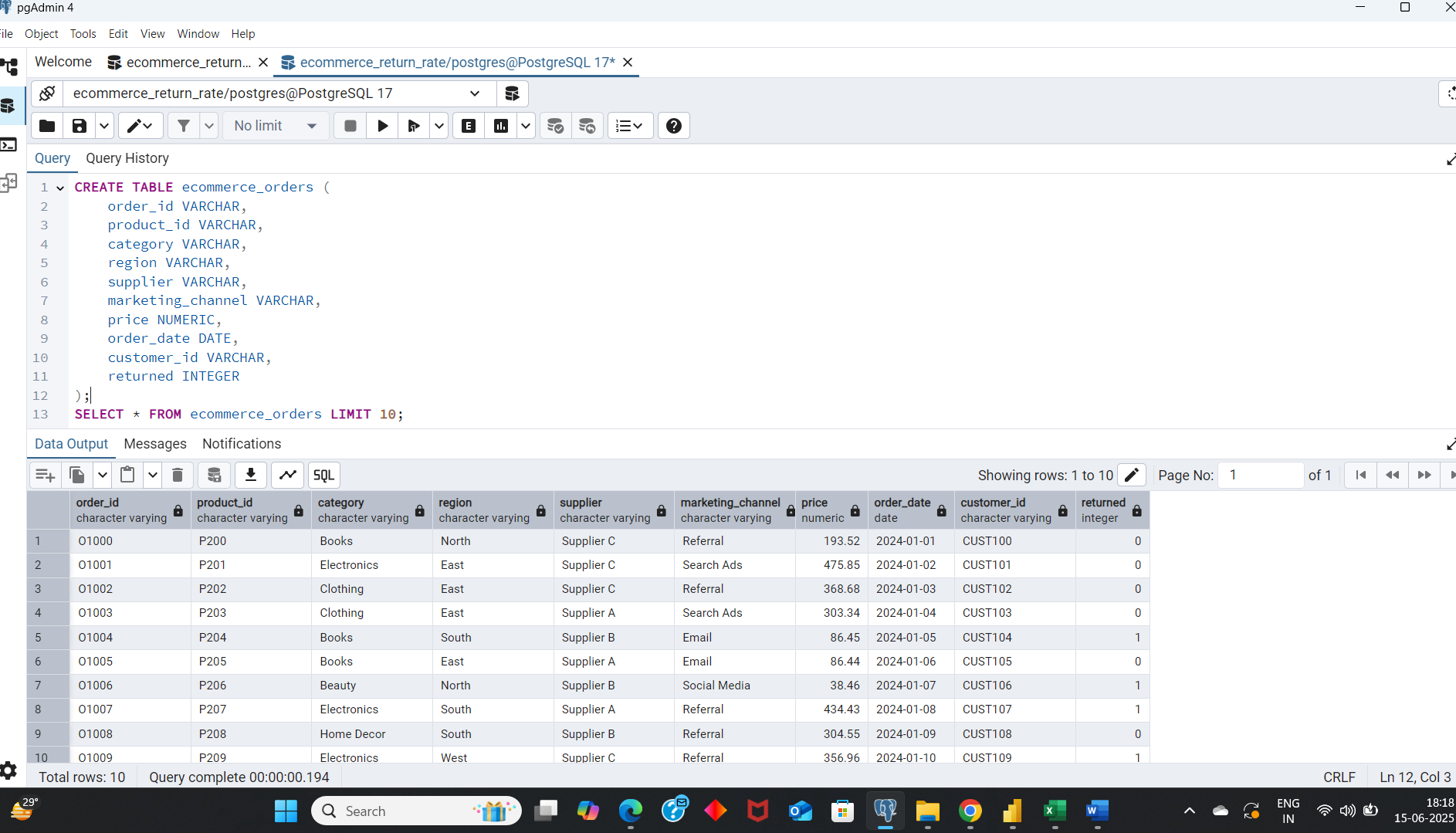
**Step 1: Data Preparation (SQL)**

Created the ecommerce\_orders table in PostgreSQL.

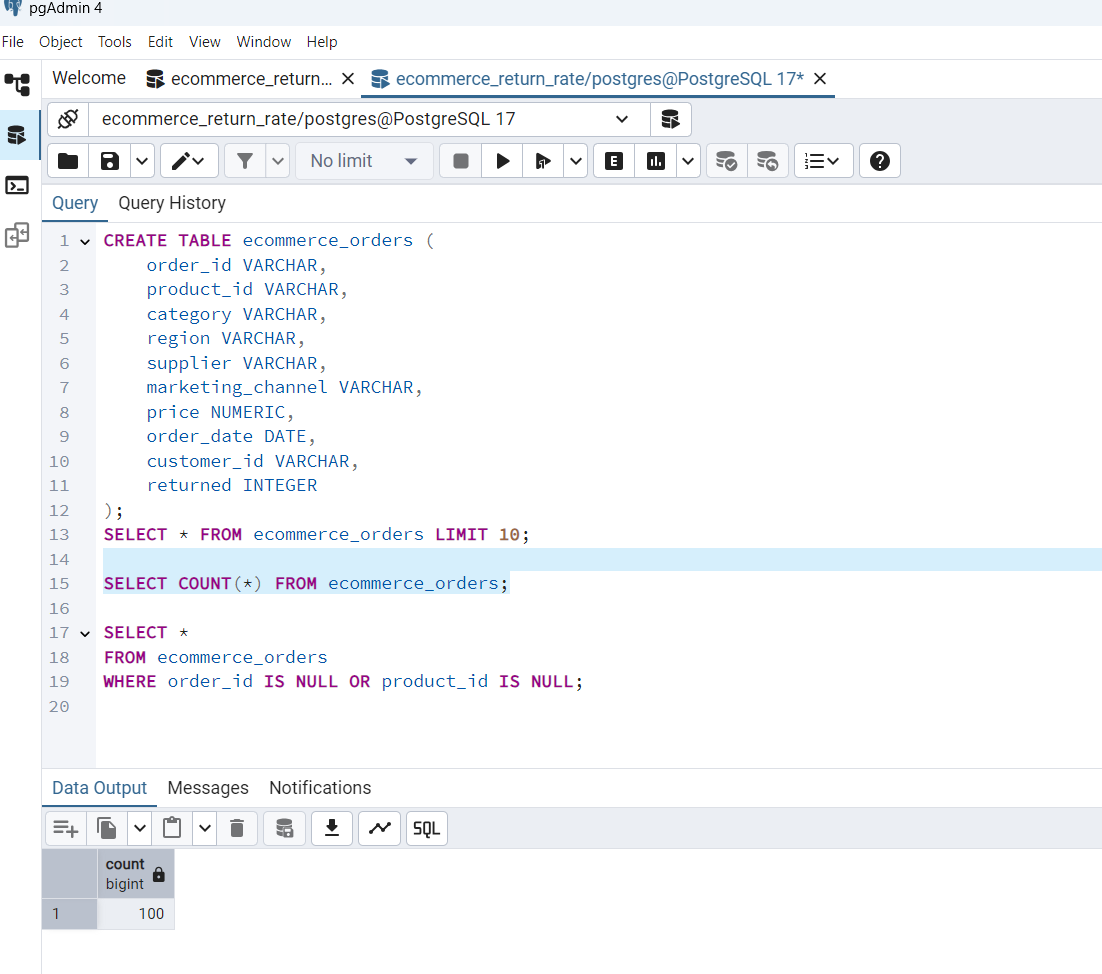
Defined key fields like order\_id, product\_id, category, region, supplier, marketing\_channel, price, order\_date, customer\_id, and returned.

Inserted sample data with a mix of return and non-return cases (0 = Not Returned, 1 = Returned).

Verified data using SELECT \* FROM ecommerce\_orders LIMIT 10.



**Step 2: Data Validation**  
Validated the dataset by checking the total number of records (100) and confirming there are no null values in key fields like order\_id and product\_id. The dataset is clean and ready for analysis.



**Step 3: Data Quality Check**  
Verified that there are no missing values in critical fields (order\_id, product\_id). This ensures the data is clean and reliable for further analysis.

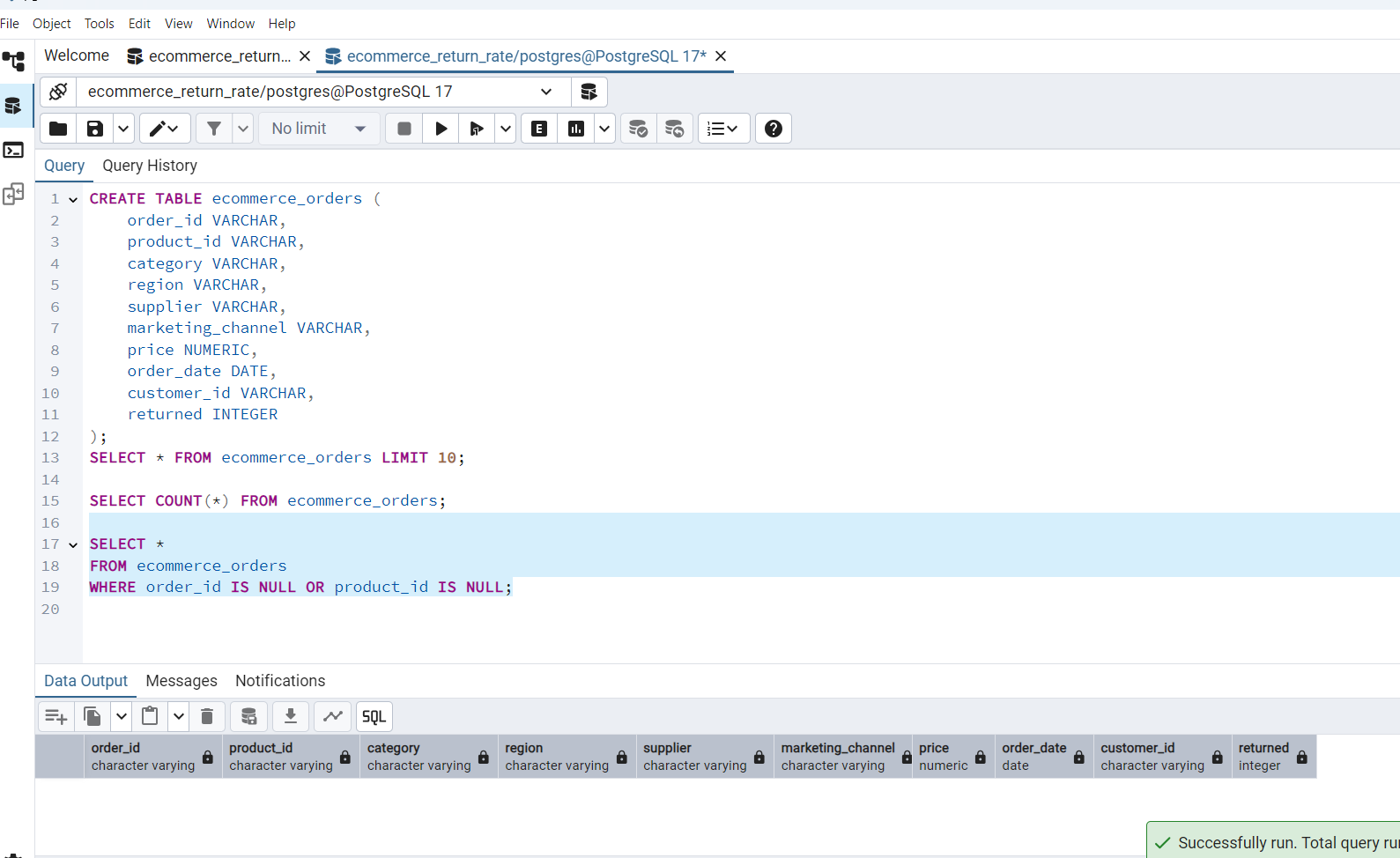
Ran a query to check if any rows have missing values in key fields order\_id or product\_id:

SELECT \*

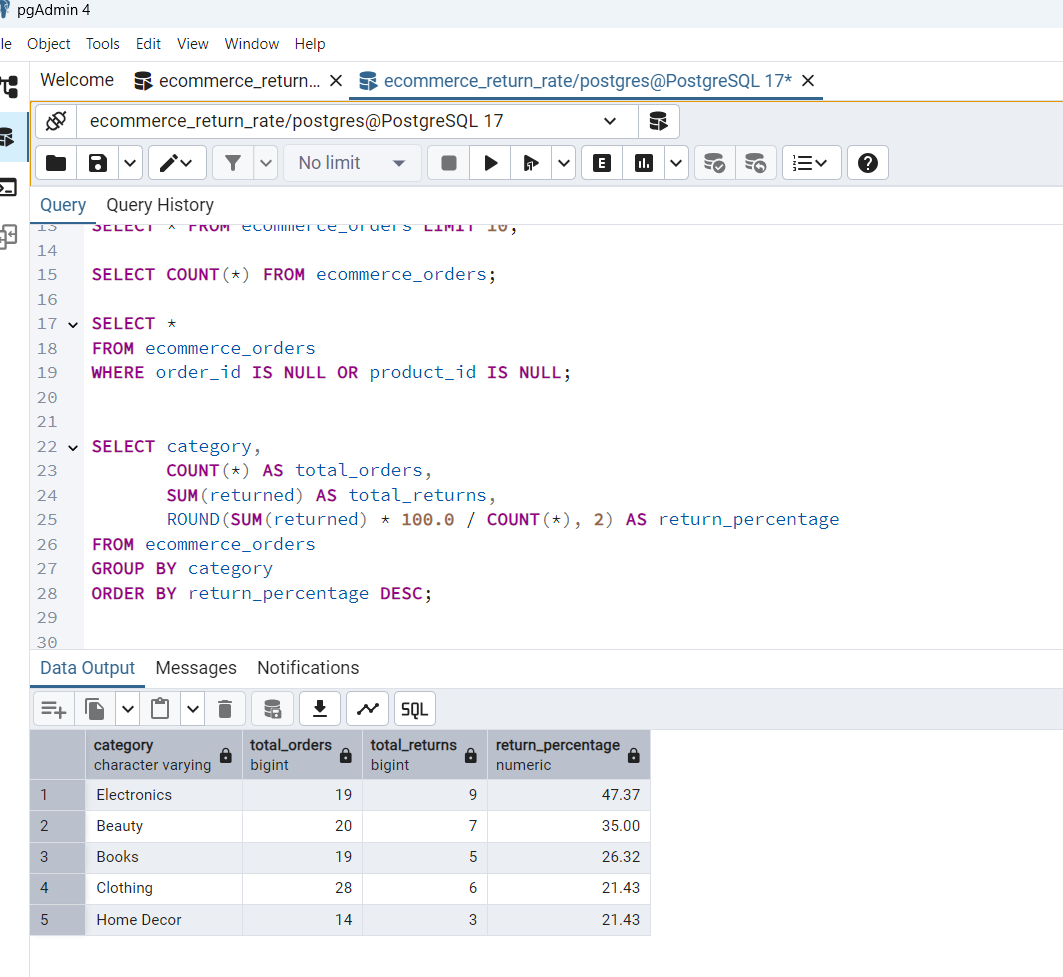
FROM ecommerce\_orders

WHERE order\_id IS NULL OR product\_id IS NULL;

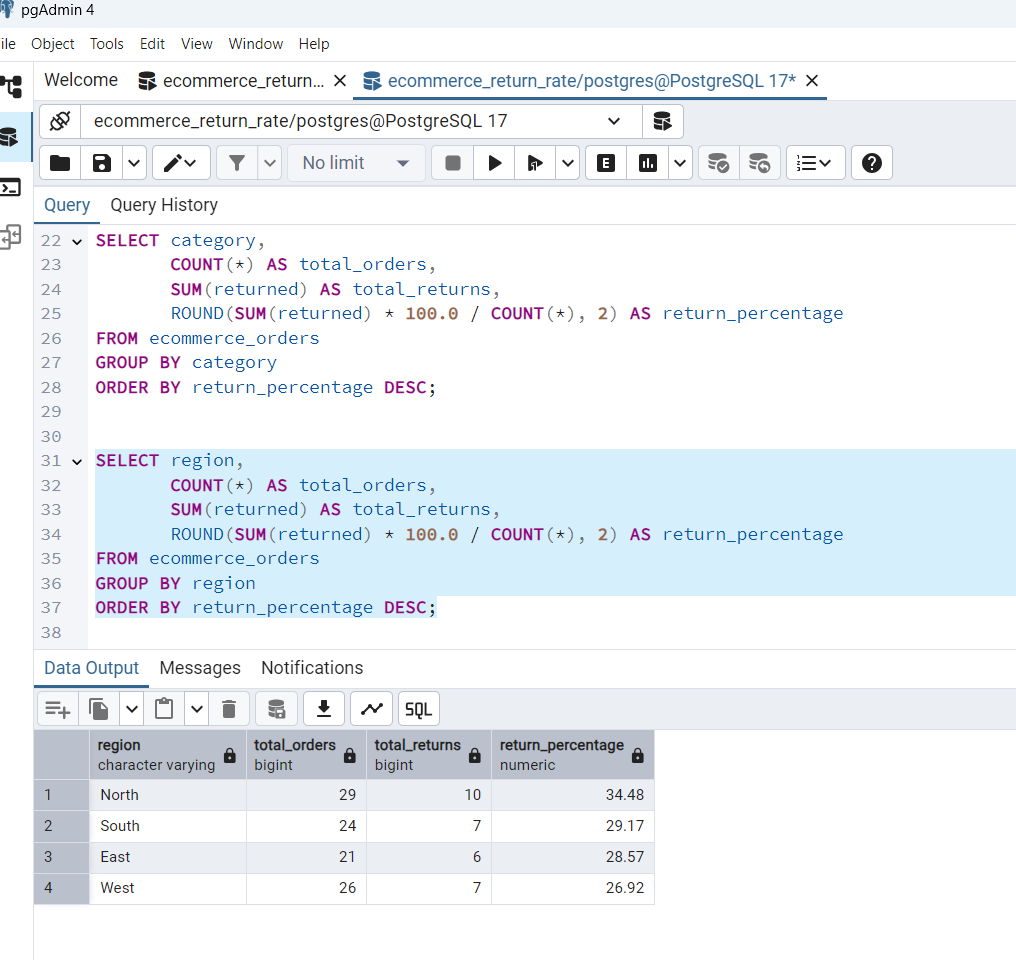
**Result:** No null values found – dataset is clean and complete.



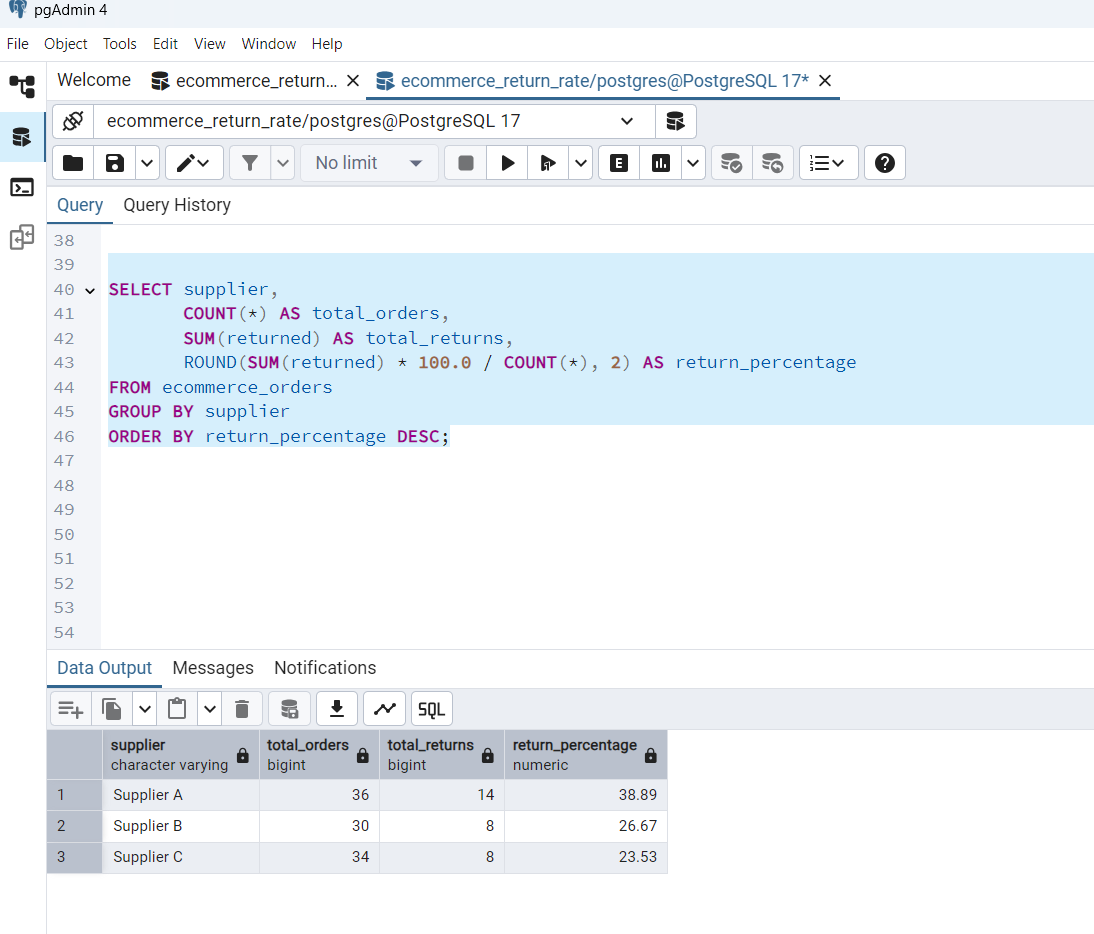
**Step 4: Return Rate by Category**  
Analyzed return percentages per product category. Electronics showed the highest return rate (47.37%), indicating a need for deeper review or improvement in this category.



**Step 5: Return Rate by Region**  
Return percentages were analyzed by region. North showed the highest return rate (34.48%), suggesting a need to investigate product or delivery issues in that area.



**Step 6: Return Rate by Supplier**  
Analyzed return rates across suppliers. Supplier A showed the highest return percentage (38.89%), indicating potential issues in product quality or logistics.



**Step 7: Return Rate by Marketing Channel**  
Referral-based orders showed the highest return percentage (43.75%), suggesting the need to assess the quality of referrals or the target audience reached through this channel.

