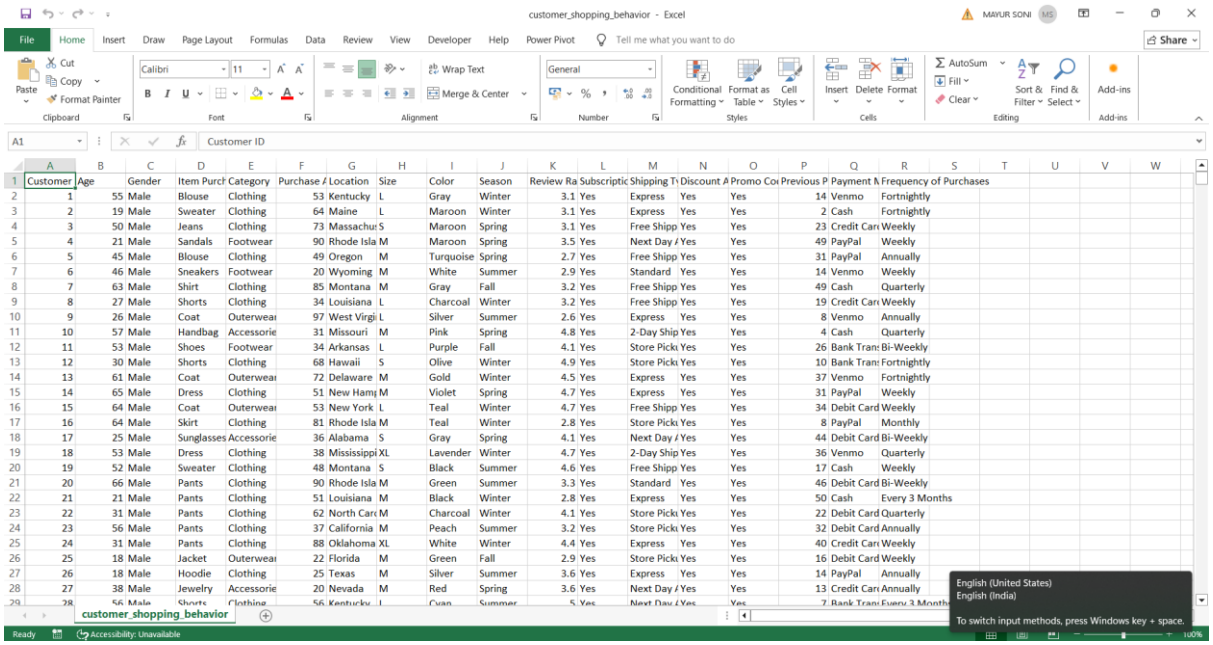
Customer Shopping Behaviour Analysis Report

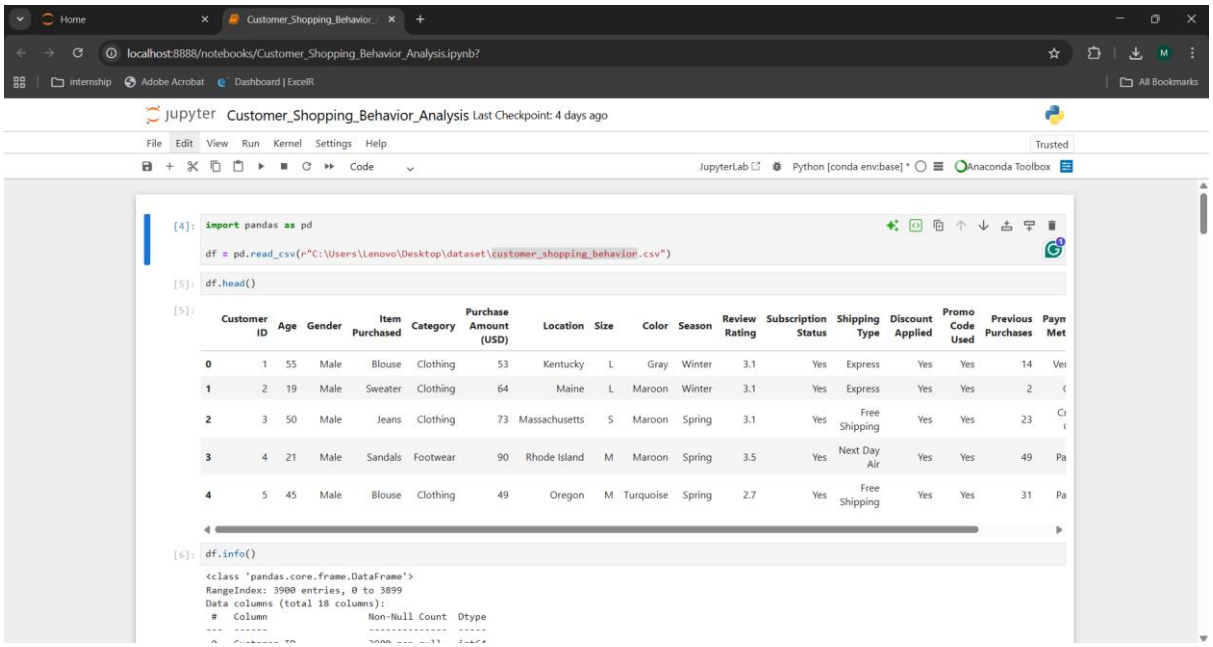
1. Project Goal

I looked closely at **3,900 customer purchases** to understand how people shop. The main goal was to find out who spends what, what they buy most often, and if they're subscribing. These insights will help us make smarter business decisions.



2. The Data Snapshot

- **Total Purchases:** 3,900



- **Details Tracked:** 18 pieces of information per purchase (e.g., age, what they bought, price, if they used a discount, shipping type, and review rating).

```
[6]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3900 entries, 0 to 3899
Data columns (total 18 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Customer ID             3900 non-null   int64
 1   Age                     3900 non-null   int64
 2   Gender                  3900 non-null   object
 3   Item Purchased          3900 non-null   object
 4   Category                3900 non-null   object
 5   Purchase Amount (USD)   3900 non-null   int64
 6   Location                3900 non-null   object
 7   Size                    3900 non-null   object
 8   Color                   3900 non-null   object
 9   Season                  3900 non-null   object
 10  Review Rating           3863 non-null   float64
 11  Subscription Status     3900 non-null   object
 12  Shipping Type           3900 non-null   object
 13  Discount Applied        3900 non-null   object
 14  Promo Code Used         3900 non-null   object
 15  Previous Purchases      3900 non-null   int64
 16  Payment Method          3900 non-null   object
 17  Frequency of Purchases  3900 non-null   object
dtypes: float64(1), int64(4), object(13)
memory usage: 548.6+ KB
```

- **Small Fix Needed:** I had to fill in **37 missing review scores**.

3. Cleaning Up the Data (Using Python)

I started by getting the data ready for analysis:

- **Load and Check:** I opened the data and checked its basic structure.
- **Missing Scores:** I filled in the missing product review scores by using the **median rating** for that specific product type.

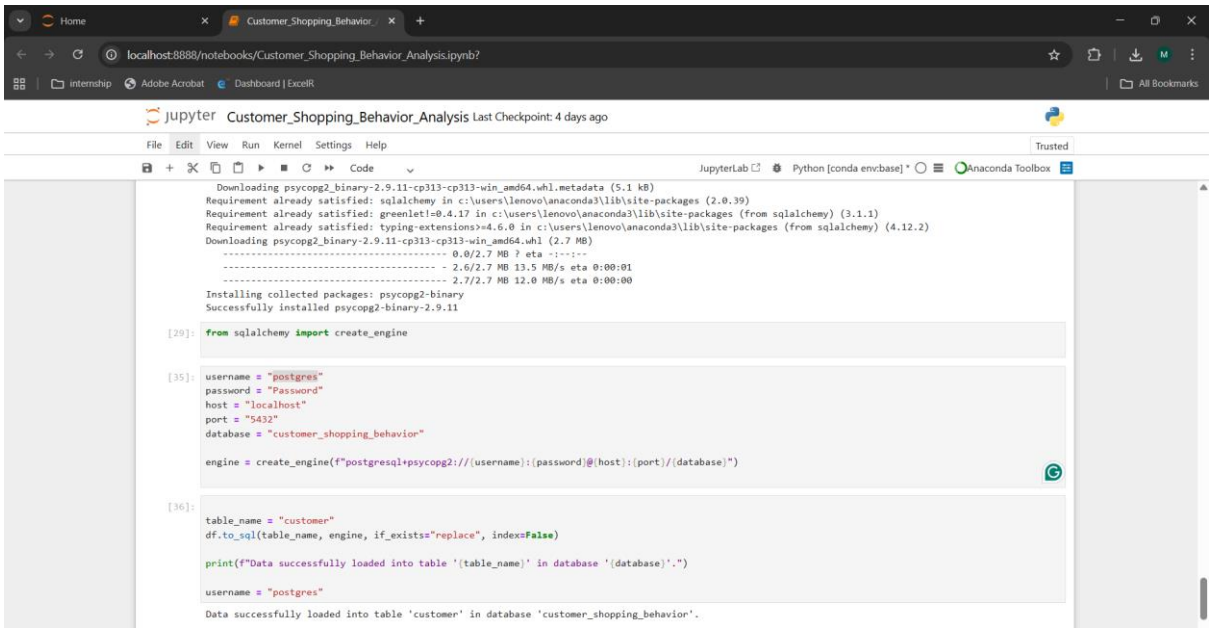
```
[8]: df.isnull().sum()

[8]: Customer ID                0
     Age                        0
     Gender                     0
     Item Purchased             0
     Category                   0
     Purchase Amount (USD)      0
     Location                   0
     Size                       0
     Color                      0
     Season                     0
     Review Rating             37
     Subscription Status        0
     Shipping Type              0
     Discount Applied           0
     Promo Code Used            0
     Previous Purchases         0
     Payment Method             0
     Frequency of Purchases     0
     dtype: int64
```

```
[9]: df['Review Rating' ] = df.groupby('Category') ['Revi

     df.isnull().sum()

[9]: Customer ID                0
     Age                        0
     Gender                     0
     Item Purchased             0
     Category                   0
     Purchase Amount (USD)      0
     Location                   0
     Size                       0
     Color                      0
     Season                     0
     Review Rating              0
     Subscription Status        0
     Shipping Type              0
     Discount Applied           0
     Promo Code Used            0
     Previous Purchases         0
     Payment Method             0
     Frequency of Purchases     0
     dtype: int64
```

- **Tidying Up:** I renamed columns to be simple and easy to read (like item purchased instead of ItemPurchased).
- **New Insights Created:**
    - I grouped customers into **Age Groups** (like 'Young Adults', 'Middle Aged', etc.).

- I figured out how many days pass between purchases.

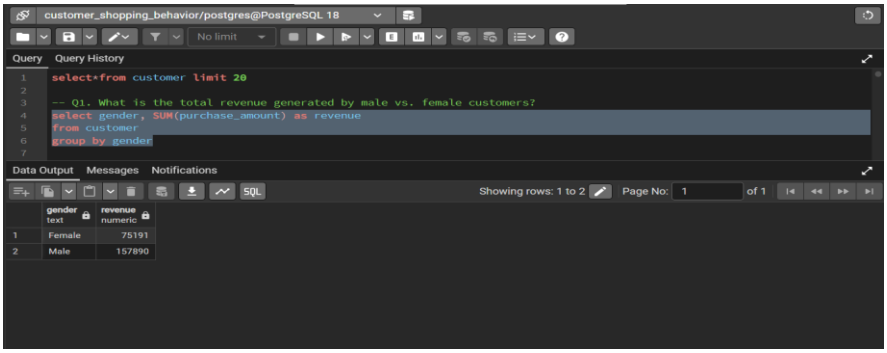
- **Removing Duplicates:** I found that the 'promo code' column was basically the same as the 'discount applied' column, so I only kept the **'discount applied'** one.
- **Ready for Deep Dive:** I saved the cleaned data into a **PostgreSQL database** for My detailed analysis using SQL.

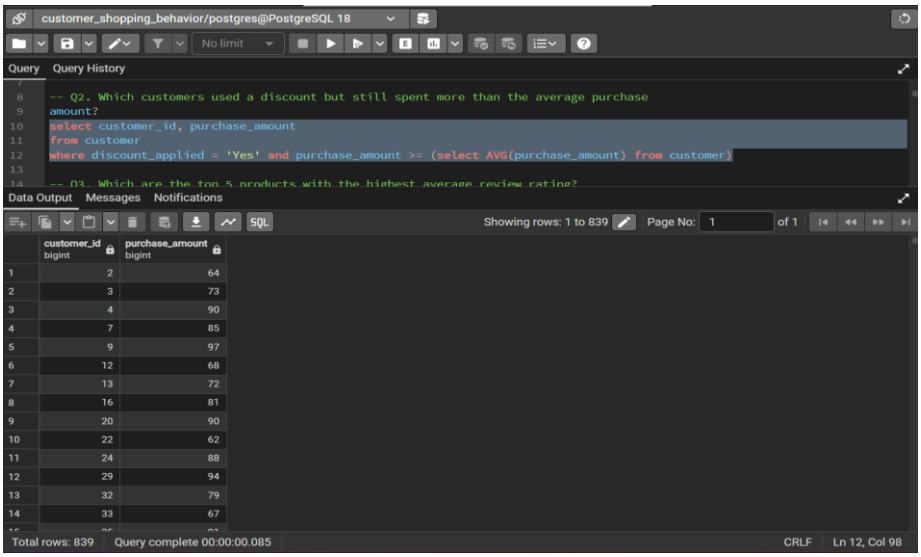
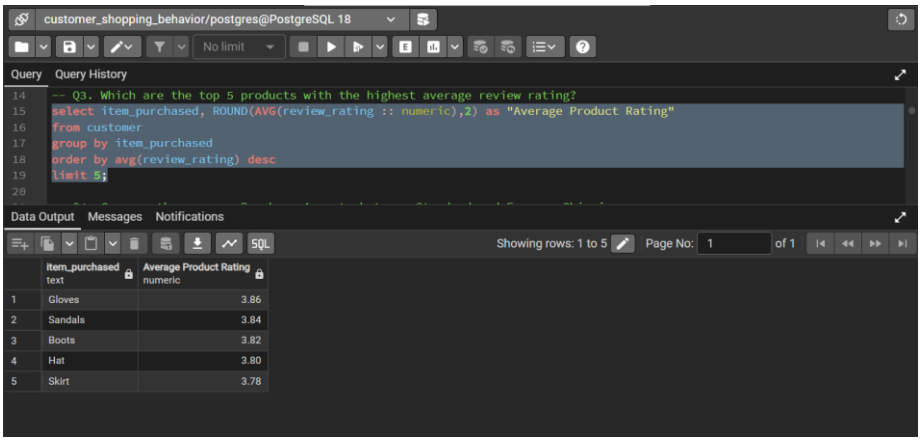
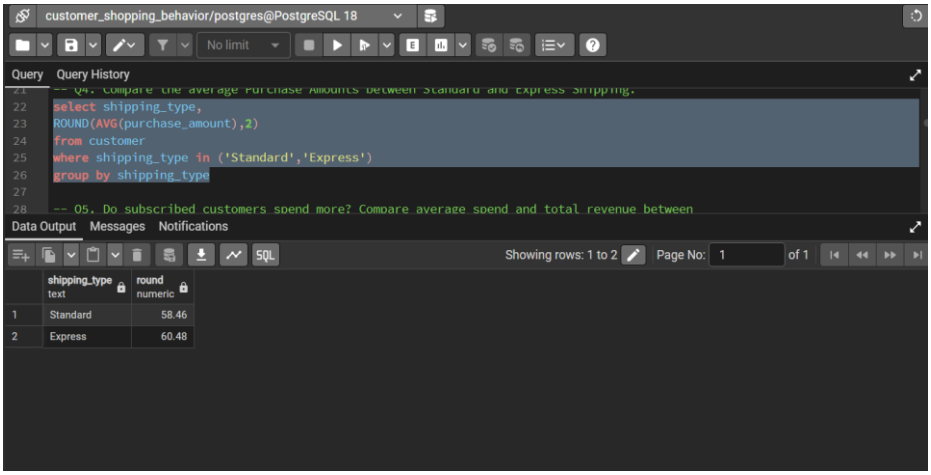


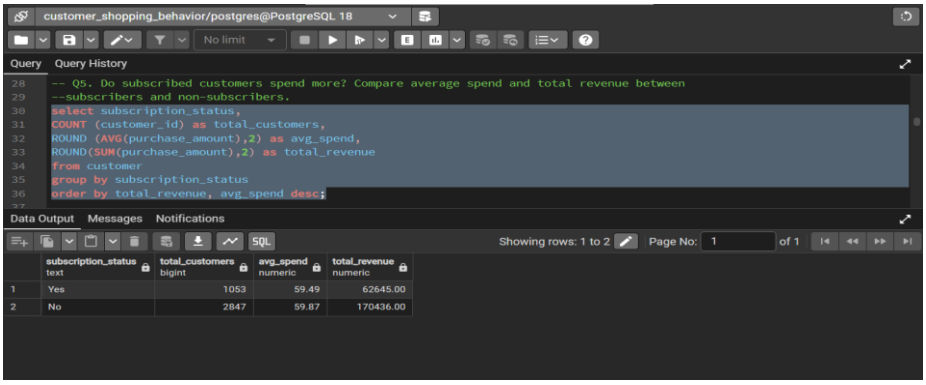
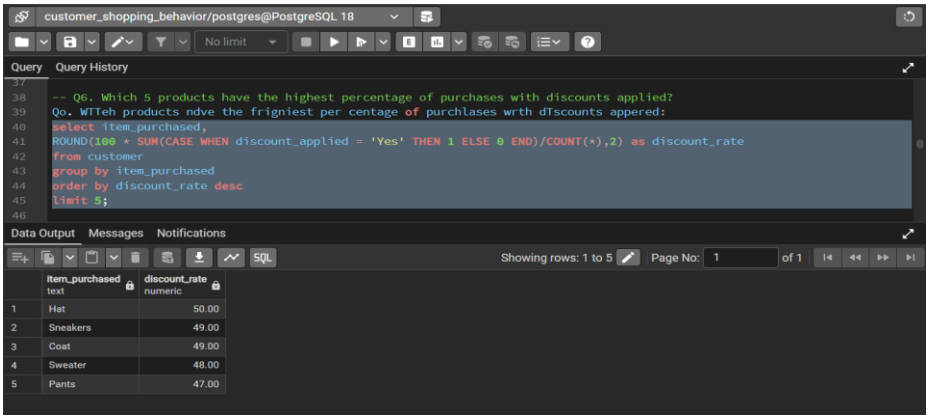
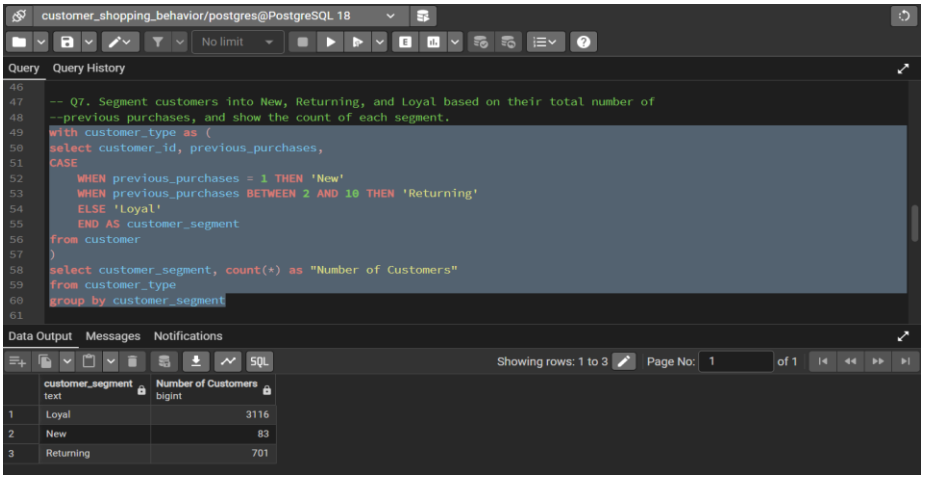
---

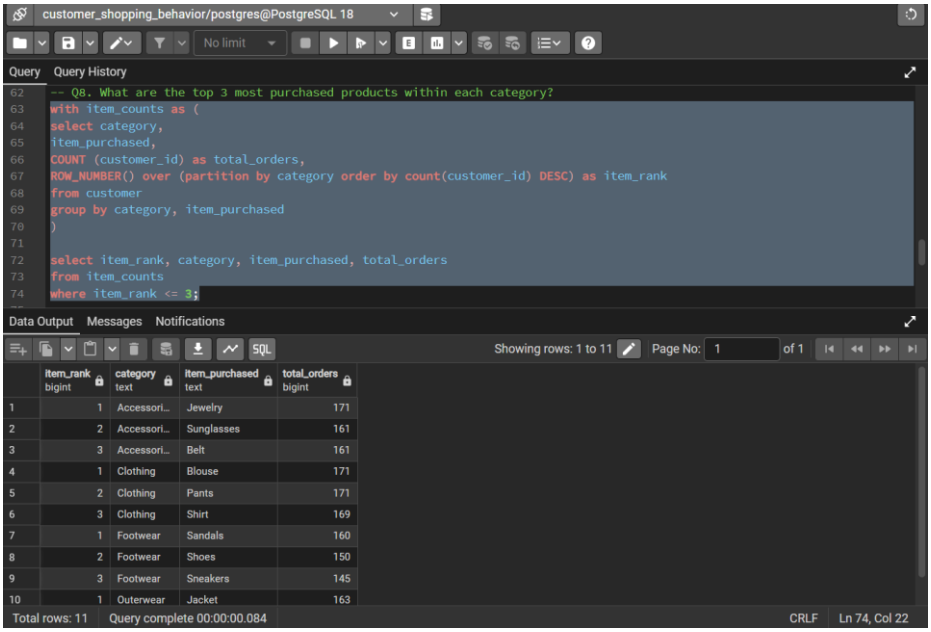
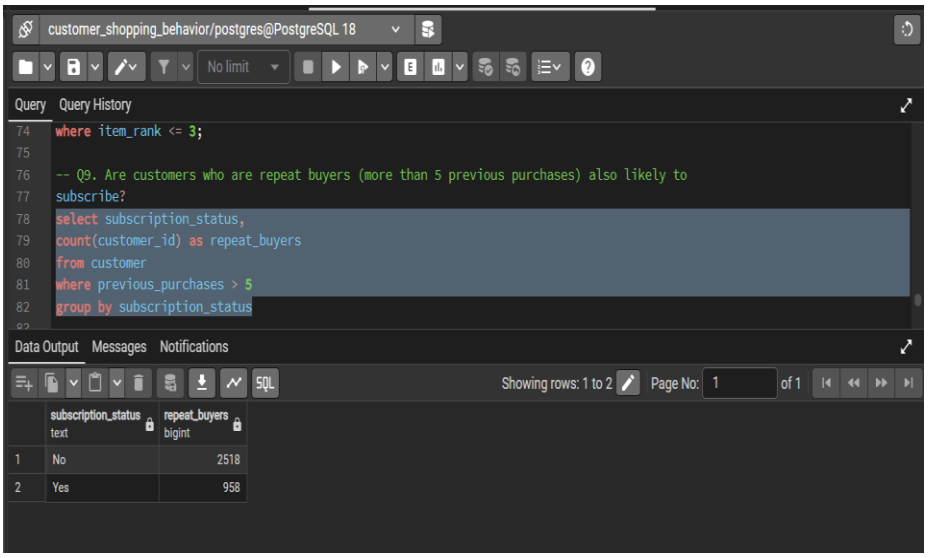
4. Key Findings from the Analysis (Using SQL)

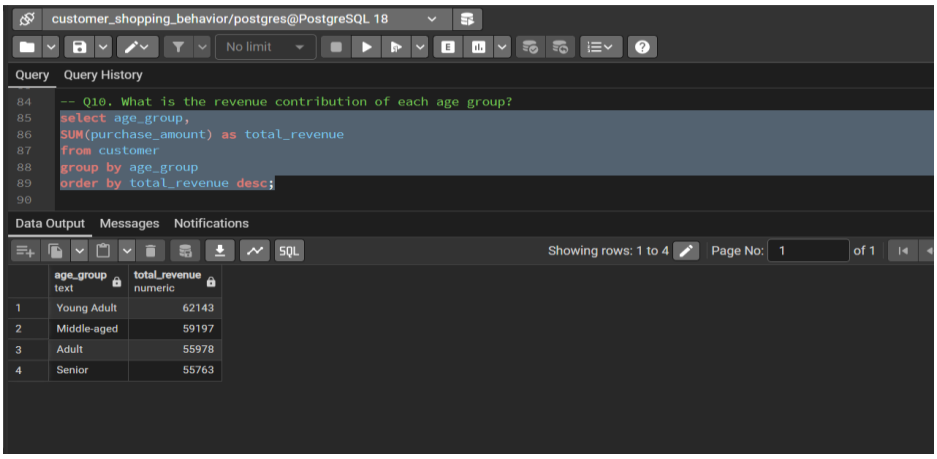
I ran specific checks to Answer important business questions:

Question	What I Looked For
Spending by Gender	Who brings in more total sales: **men or women**? 
Smart Discount Users	Which customers used a discount but still **spent more than the average amount**?

Question	What I Looked For
	
Best Loved Products	The **Top 5 products** based on their high average review score. 
Shipping Impact	Does the **purchase amount change** between **Standard** and **Express** shipping? 

Question	What I Looked For
Subscribers vs. Others	How much more do **subscribers** spend, and how much total revenue do they generate compared to non-subscribers? 
"Discount Only" Items	The **Top 5 products** that are most often bought *only* when a discount is offered. 
Customer Groups	How to categorize customers based on how much they've bought before: **New**, **Returning**, or **Loyal**? 

Question	What I Looked For
Top Items in Each Category	The **three most purchased products** within *each* product category (e.g., most purchased in "Apparel"). 
Loyalty & Subscriptions	Are customers who buy **more than 5 times** more likely to be subscribers? 
Revenue by Age	Which **age groups** bring in the most money?

Question	What I Looked For
	

5. Visualizing the Data (PoIr BI)

I created an **interactive dashboard** in Power BI to easily see all these findings in charts and graphs.

6. My Recommendations for the Business

Based on the data, here's how I should move forward:

- **Grow Subscriptions:** Offer **exclusive perks** to encourage more people to sign up for subscriptions.
- **Reward Loyalty:** Create a rewards program to turn **returning customers** into **loyal, high-value buyers**.
- **Fine-Tune Discounts:** Carefully check My discount strategy to ensure sales boosts don't hurt My profit margins too much.
- **Showcase the Best:** Run marketing campaigns that **highlight the top-rated and best-selling products**.
- **Focus your Ads:** Direct marketing efforts toward the **age groups** that spend the most, and consider special promotions for those who choose **Express shipping**.
