Customer Shopping Behavior Analysis - Project Report

Table of Contents

1. Executive Summary

This project analyzes customer shopping behavior using Python, SQL, and Power BI. The goal is to uncover trends related to customer demographics, purchase preferences, discount behavior, loyalty, and seasonal patterns.

2. Dataset Overview

The dataset comprises 3,900 customer transactions, featuring 18 columns with fields such as customer ID, age, gender, category, purchase amount, review ratings, subscription status, shipping type, and previous purchases.

3. Data Cleaning & Preparation

- Removed unnecessary columns
- Ensured consistent lowercase column names
- Created age groups using quantiles
- Converted categorical frequency-of-purchase text into numeric days
- The verified dataset contains no missing values
- Created PostgreSQL Database Using SQL.

4. Exploratory Data Analysis (Python)

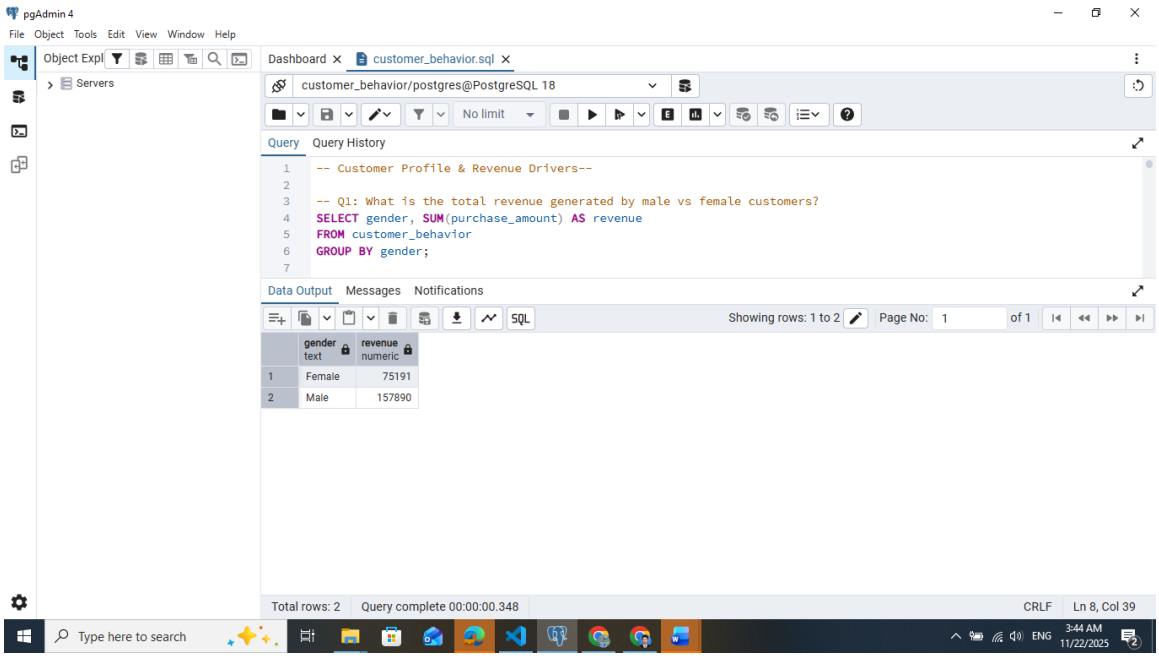
Basic profiling was performed using pandas (df.describe). Key statistics include:

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Review Rating
count	3900	3900	3900	3900	3900	3900	3900
unique	-	-	2	25	4	-	-
top	-	-	Male	Blouse	Clothing	-	-
freq	-	-	2652	171	1737	-	-
mean	1950.50	44.07	-	-	-	59.76	3.75
std	1125.98	15.21	-	-	-	23.69	0.72
min	1	18	-	-	-	20	2.5
25%	975.75	31	-	-	-	39	3.1
50%	1950.50	44	-	-	-	60	3.7
75%	2925.25	57	-	-	-	81	4.4
max	3900	70	-	-	-	100	5.0

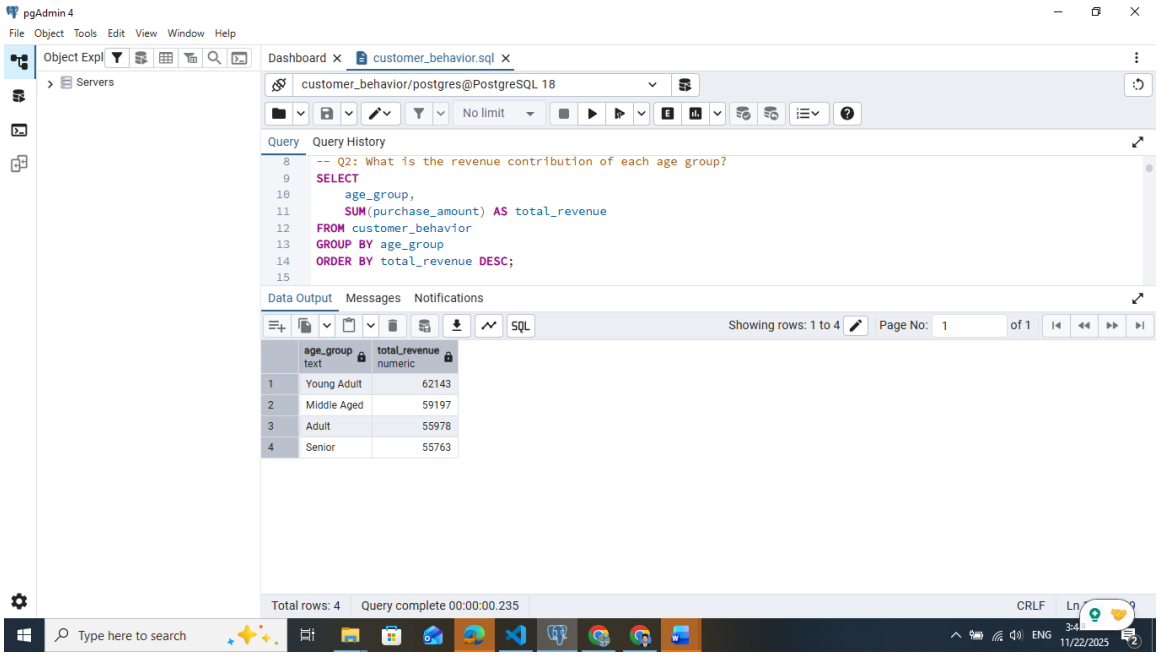
5. SQL Analysis & Insights

Multiple SQL queries were executed to analyze revenue drivers, customer loyalty, discount sensitivity, and product performance. Key insights include:

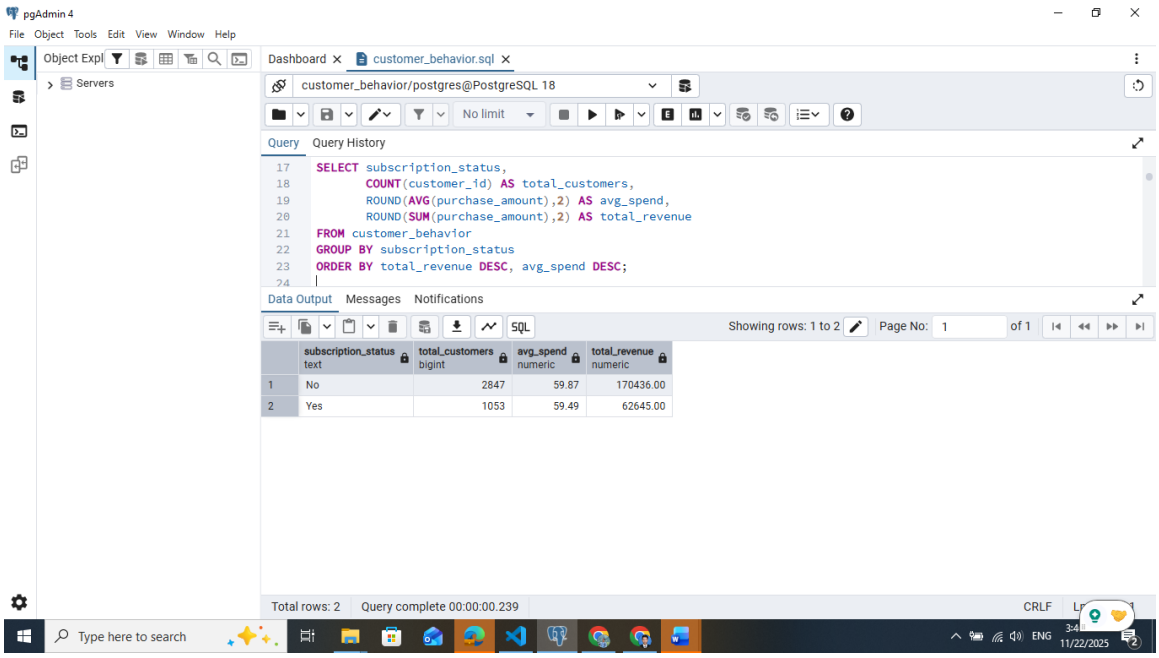
Q1: What is the total revenue generated by male vs female customers?



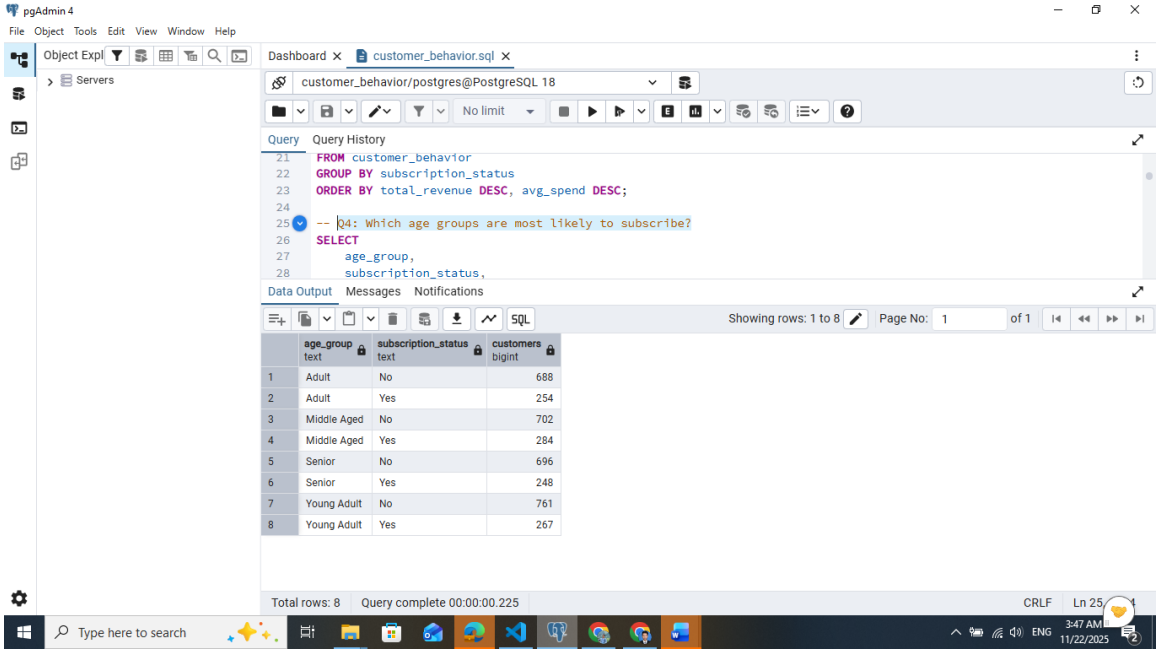
-- Q2: What is the revenue contribution of each age group?



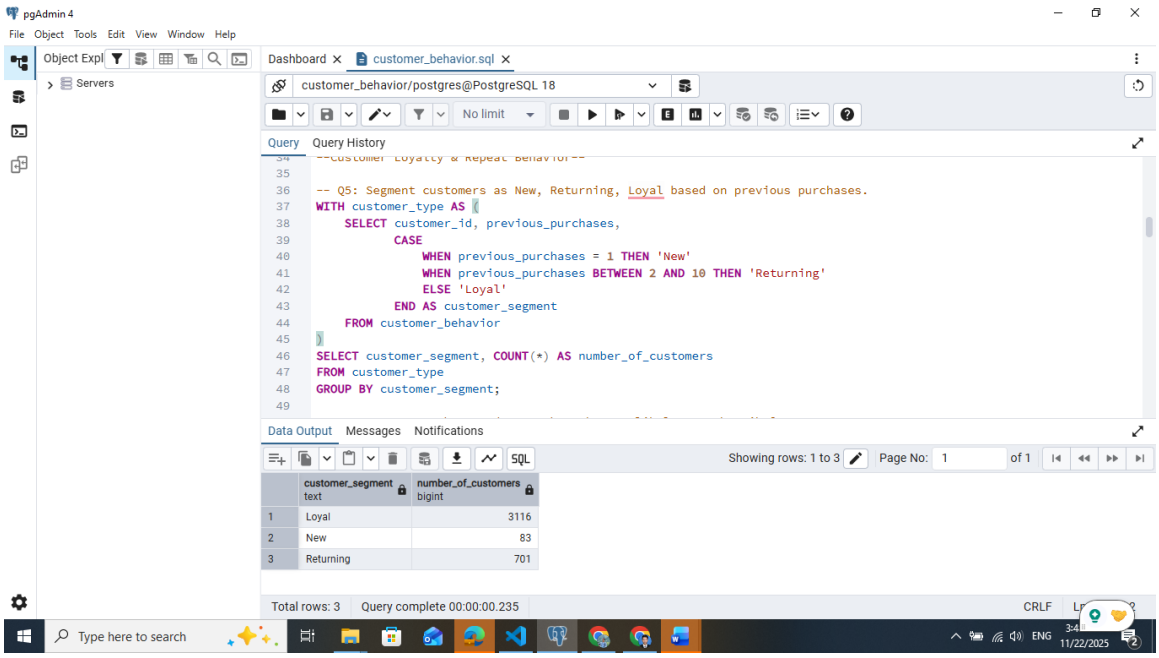
Q3: Do subscribed customers spend more? Compare avg spend & total revenue.



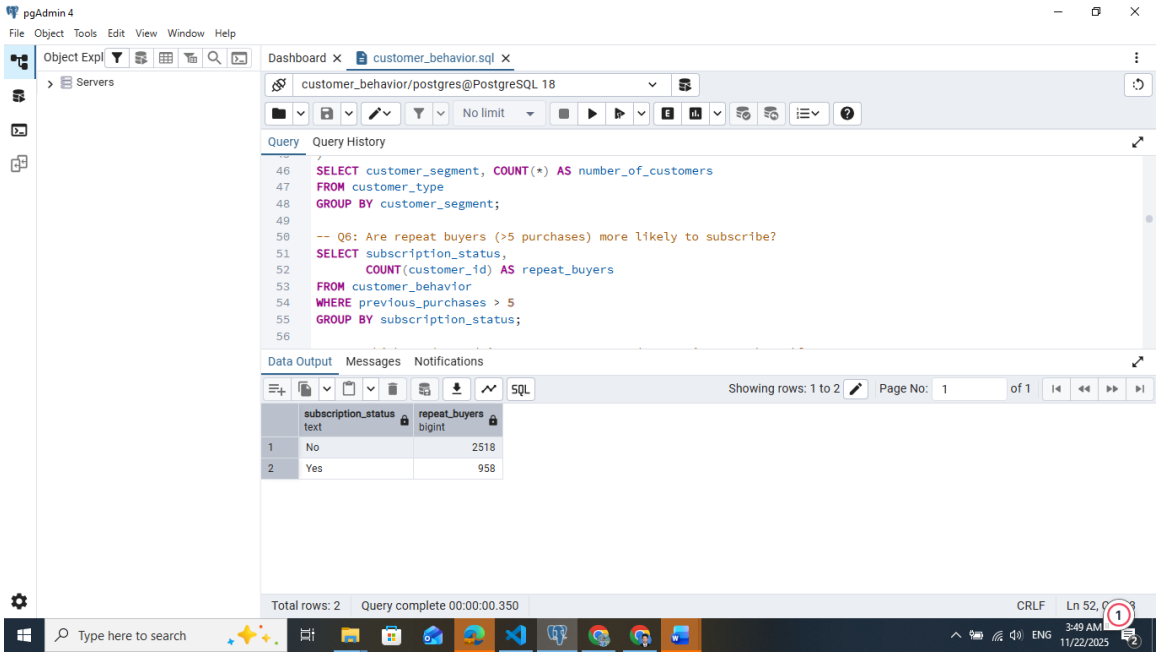
Q4: Which age groups are most likely to subscribe?



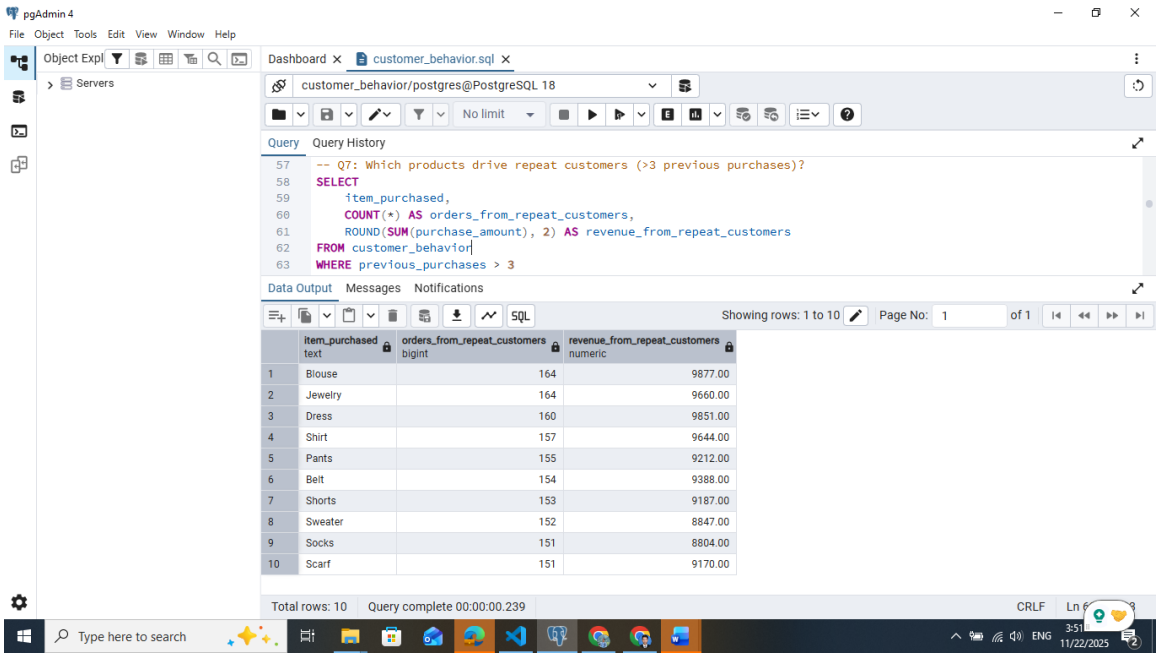
Q5: Segment customers as New, Returning, Loyal based on previous purchases.



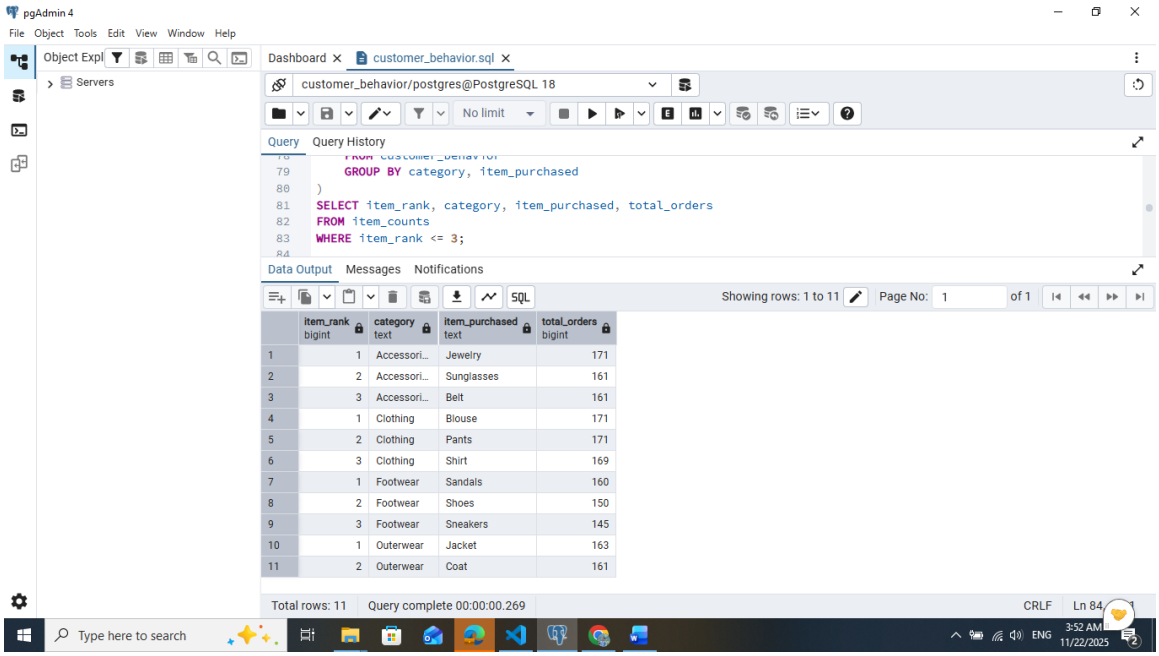
Q6: Are repeat buyers (>5 purchases) more likely to subscribe?



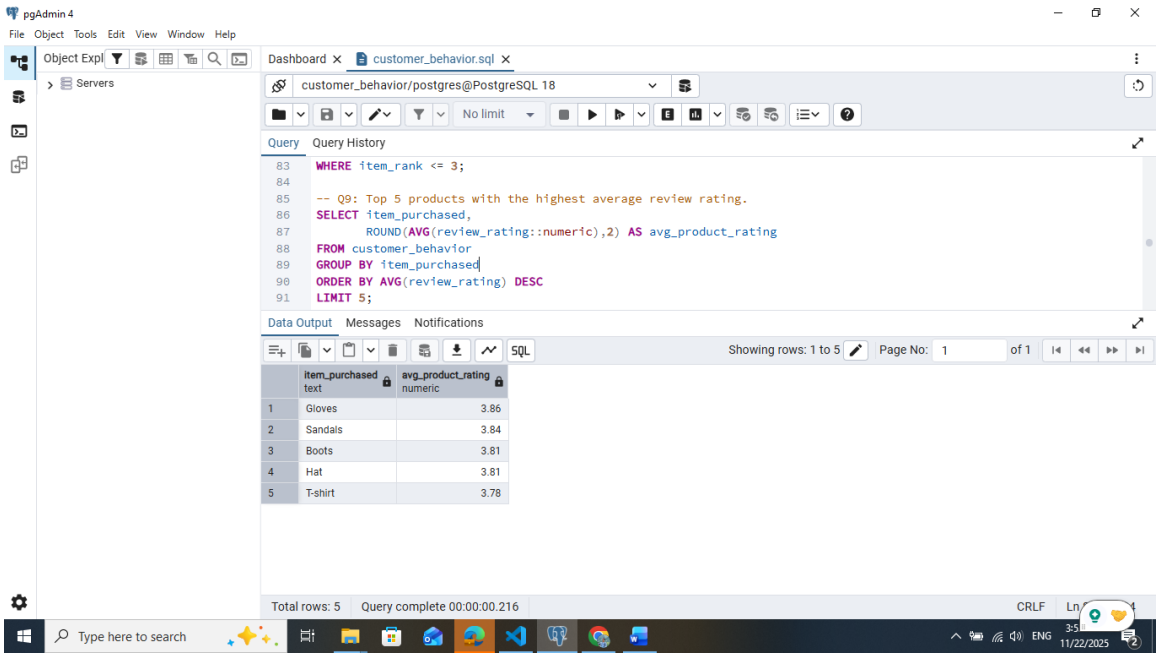
Q7: Which products drive repeat customers (>3 previous purchases)?



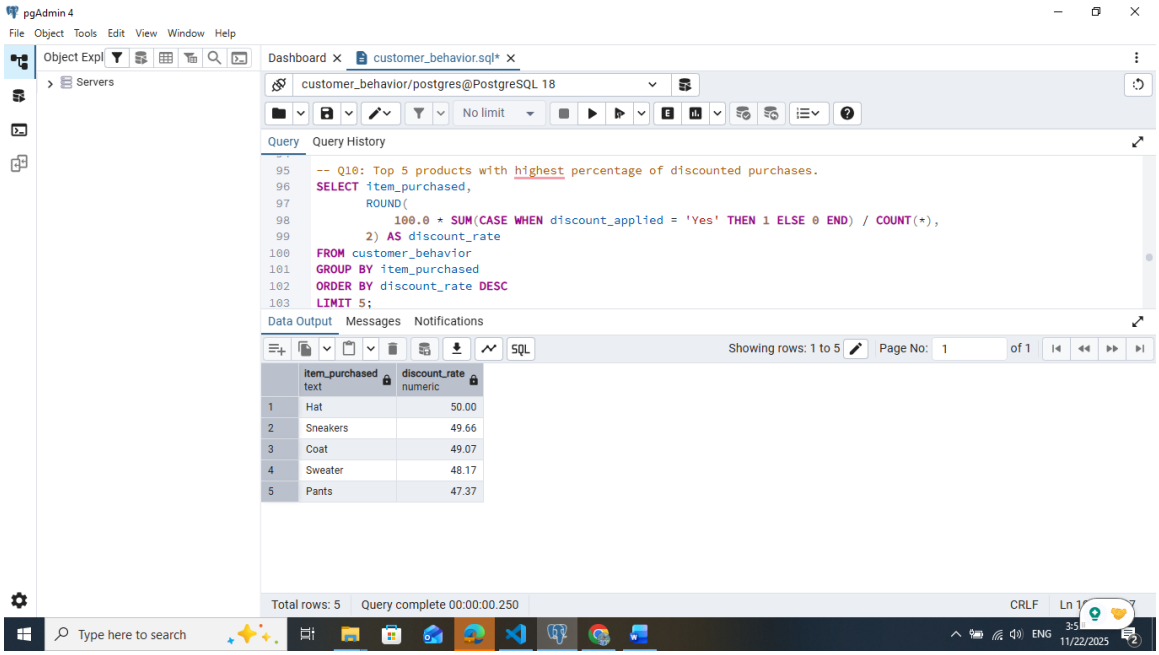
Q8: Top 3 most purchased products within each category.



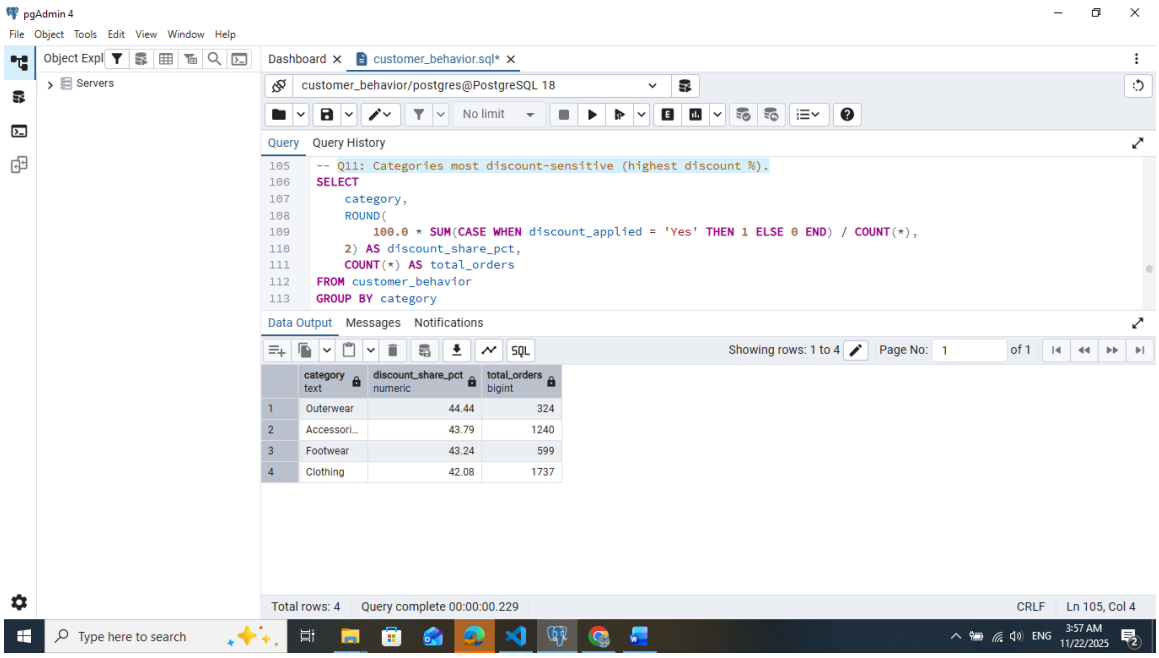
Q9: Top 5 products with the highest average review rating



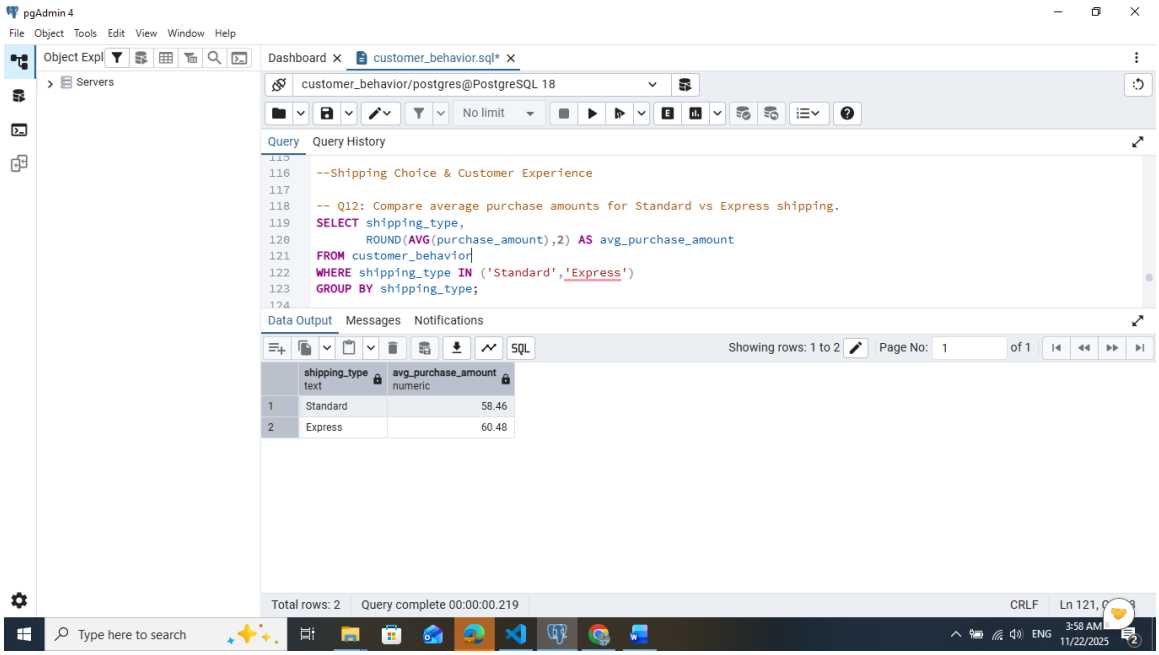
Q10: Top 5 products with highest percentage of discounted purchases.



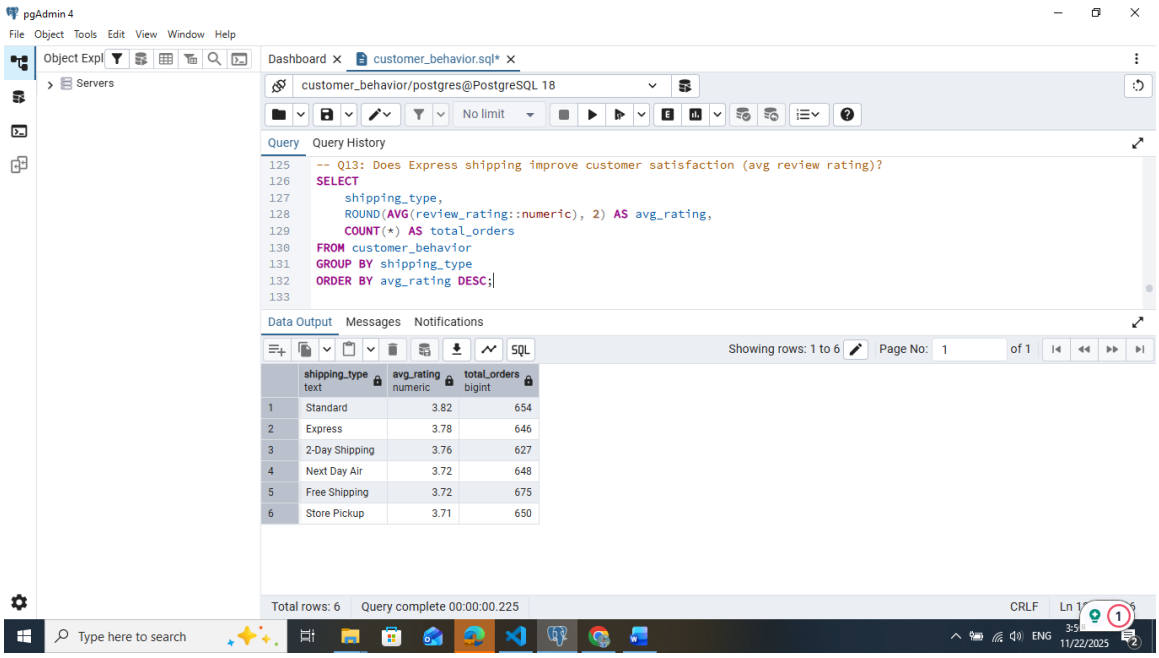
Q11: Categories most discount-sensitive (highest discount %).



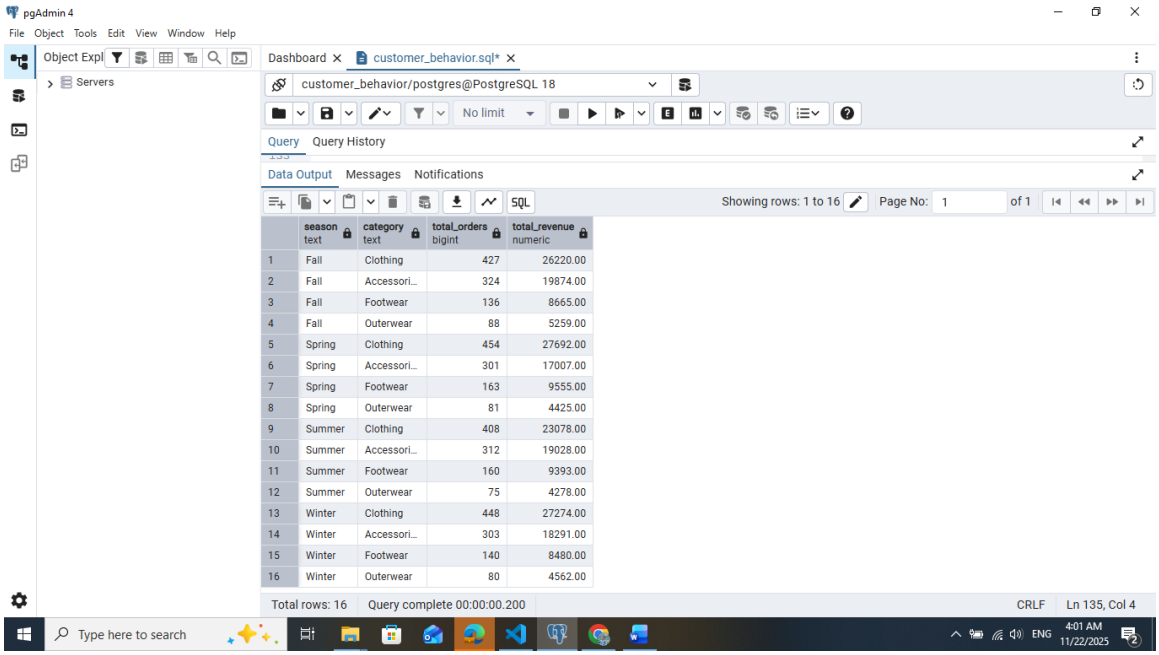
Q12: Compare average purchase amounts for Standard vs Express shipping.



Q13: Does Express shipping improve customer satisfaction (avg review rating)?



Q14: Seasonal demand – revenue by season and category.



6. Power BI Dashboard

An interactive dashboard was built in Power BI, including metrics such as:

- Total customers
- Total revenue
- Average purchase amount

- Revenue by category
- Sales by age group
- Subscription-based segmentation

7. Seasonality Insights

Across all seasons, **Clothing** consistently generates strong revenue, with peaks in **Spring and Winter**.
Insight: The business exhibits stable year-round performance with moderate seasonal variation.

8. Strategic Recommendations

Based on the findings from the SQL analysis, the following targeted recommendations are proposed:

8.1 Strengthen Marketing Toward High-Value Segments

- Prioritize campaigns targeting **Male** and **Young Adult** customers, the highest revenue contributors.
- Tailor product recommendations and email campaigns around these demographics.

8.2 Revamp the Subscription Program

Because non-subscribers currently:

- Spend slightly more,
- Generate significantly higher revenue,
- And include the majority of repeat buyers,

Implement improvements such as:

- Exclusive member-only offers
- Loyalty points multiplied
- Early access to new collections
- Personalized incentives for returning and Loyal customers (the 3,800+ most valuable users)

8.3 Optimize Discount Strategy

Since discount rates hover around 42–44% across all categories:

- Reduce blanket discounts that erode margins.
- Shift to **targeted** discounts for new or dormant customers only.
- Re-price products with ~50% discount reliance (e.g., Hat, Sneakers, Coat).

8.4 Capitalize on Repeat-Driving Products

- Promote **Blouse, Jewelry, Dress, Shirt, Pants** as anchor products in advertising.
- Bundle high-performing items (e.g., Blouse + Jewelry) to increase average order value.
- Feature top-rated products (Gloves, Sandals, Boots) on the homepage.

8.5 Improve Customer Experience Without Increasing Logistics Cost

- Keep **Standard Shipping** as the default since it has the highest satisfaction.
- Offer **Express** selectively as an upsell on high-value carts.

8.6 Leverage Seasonality While Maintaining Steady Inventory

- Prepare seasonal campaigns for **Spring and Winter**, when Clothing performs best.
- Maintain consistent inventory year-round to support stable demand.