Customer Shopping Behavior Analysis

1. Project Overview

This project analyses customer shopping behavior using transactional data from 3,900 purchases across various product categories. The goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

2. Dataset Summary

- Rows: 3,900
- Columns: 18
- Key Features:
    - Customer demographics (Age, Gender, Location, Subscription Status)
    - Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
    - Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review Rating, Shipping Type)
- Missing Data: 37 values in Review Rating column

3. Exploratory Data Analysis using PostgreSQL

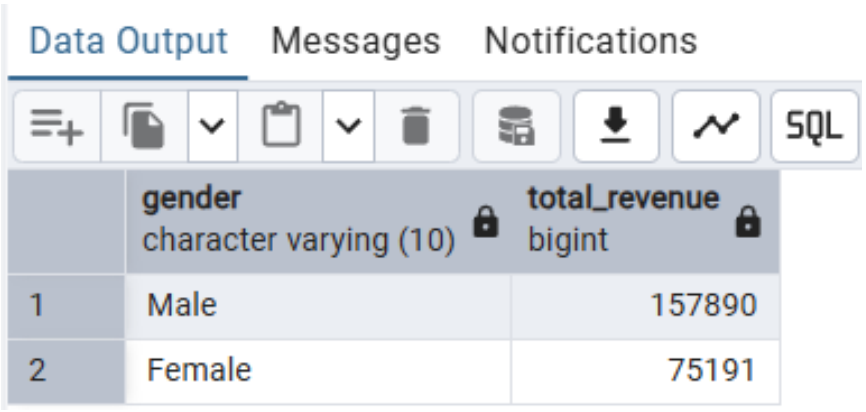
I began with data preparation and cleaning in PostgreSQL:

- **Data Loading:** Import CSV in PostgreSQL database.
- **Handle Missing Data:** Checked for null values and imputed missing values in the review_rating column using the median rating of each product category.
- **Column Standardization:** Renamed columns to snake case for better readability and documentation.
- **Feature Engineering:**
    - Created age_group column by binning customer ages.
    - Created purchase_frequency_days column from purchase data.
- **Data Consistency Check:** Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.
- **Database Integration into Power BI:** Connect PostgreSQL to the Power BI for dashboard creation and insight generation.

4. Data Analysis using PostgreSQL

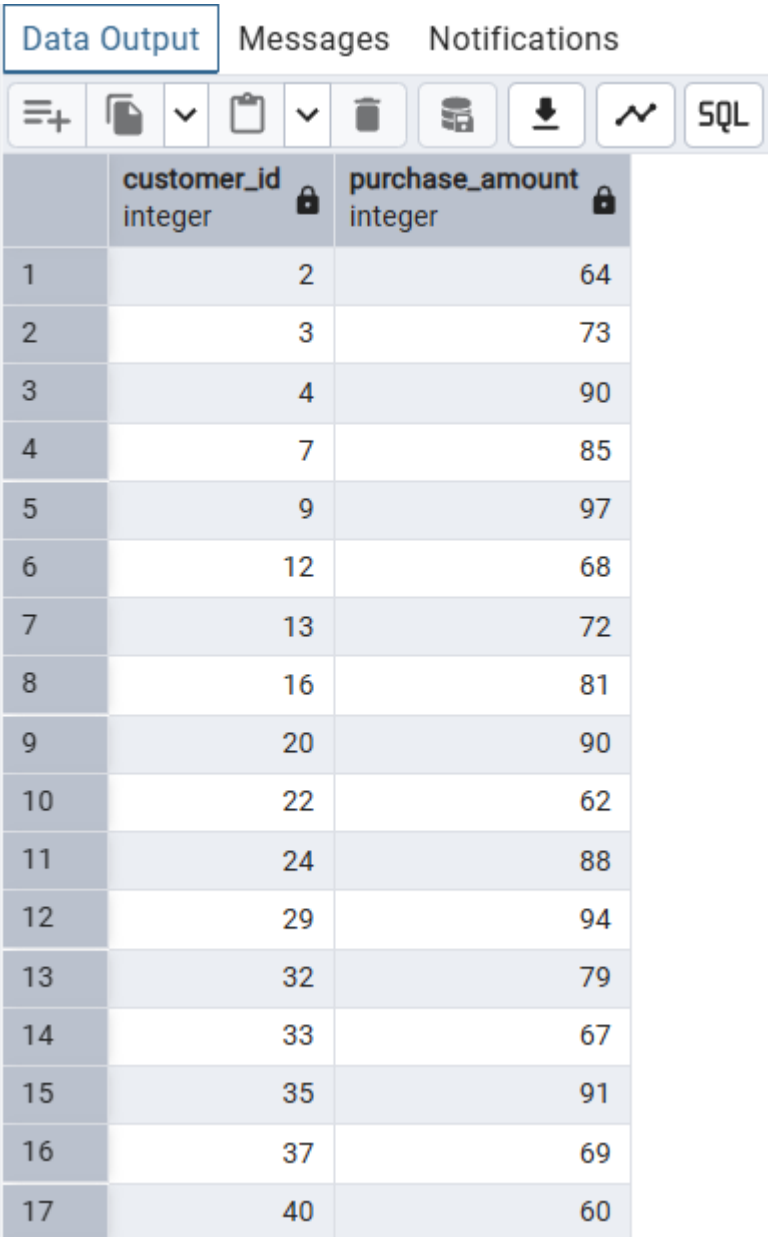
We performed structured analysis in PostgreSQL to answer key business questions:

1. **Revenue by Gender –** Compared total revenue generated by male vs. female customers.



	gender character varying (10)	total_revenue bigint
1	Male	157890
2	Female	75191

2. **High-Spending Discount Users –** Identified customers who used discounts but still spent above the average purchase amount.



Data Output | Messages | Notifications

SQL

	customer_id integer	purchase_amount integer
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	88
12	29	94
13	32	79
14	33	67
15	35	91
16	37	69
17	40	60

3. **Top 5 Products by Rating –** Found products with the highest average review ratings.

Data Output | Messages | Notifications

	item_purchased character varying (50)	average_review_rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

4. **Shipping Type Comparison –** Compared average purchase amounts between Standard and Express shipping.

Data Output | Messages | Notifications

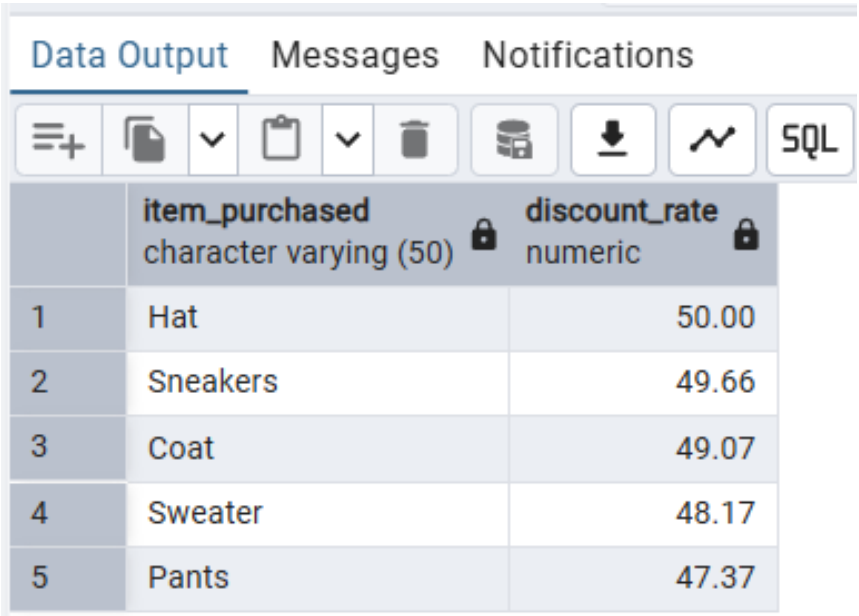
	shipping_type character varying (20)	average_purchase_amount numeric
1	Standard	58.46
2	Express	60.48

5. **Subscribers vs. Non-Subscribers –** Compared average spend and total revenue across subscription status.

Data Output | Messages | Notifications

	subscription_status character varying (10)	average_spend numeric	total_revenue bigint
1	Yes	59.49	62645
2	No	59.87	170436

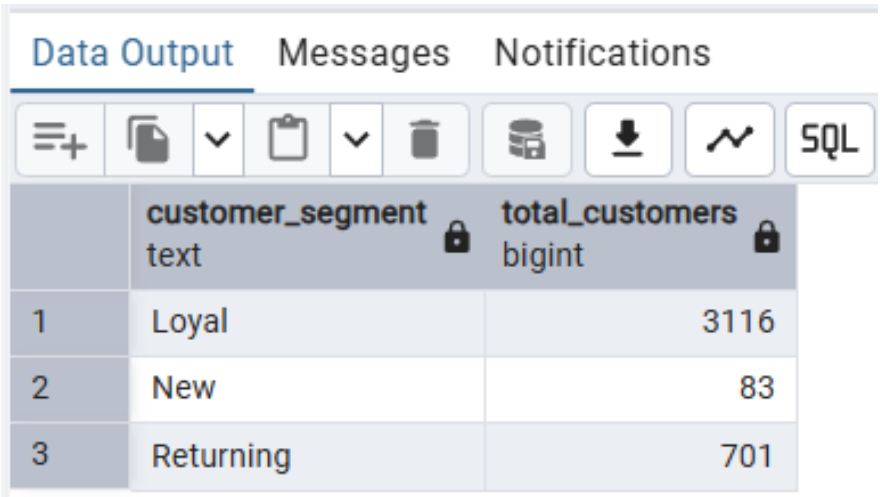
6. **Discount-Dependent Products –** Identified 5 products with the highest percentage of discounted purchases.



Data Output | Messages | Notifications

	item_purchased character varying (50)	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. **Customer Segmentation –** Classified customers into New, Returning, and Loyal segments based on purchase history.



Data Output | Messages | Notifications

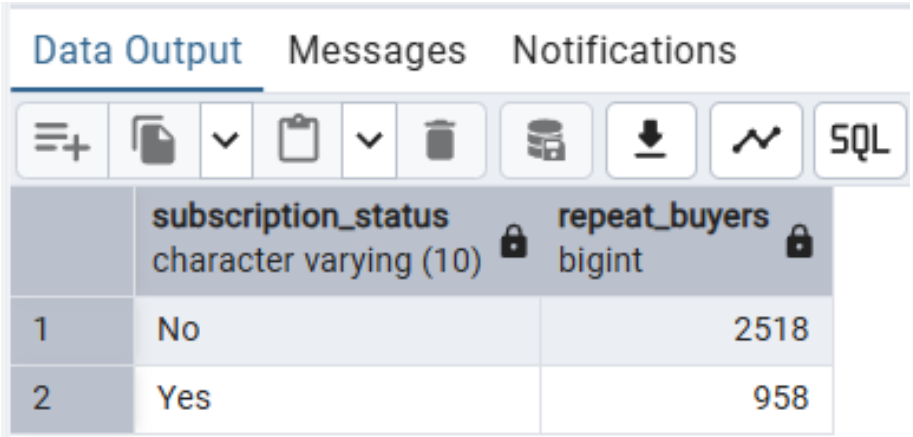
	customer_segment text	total_customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

8. **Top 3 Products per Category –** Listed the most purchased products within each category.

Data Output | Messages | Notifications

	item_rank bigint	category character varying (30)	item_purchased character varying (50)	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. **Repeat Buyers & Subscriptions –** Checked whether customers with >5 purchases are more likely to subscribe.



	subscription_status character varying (10)	repeat_buyers bigint
1	No	2518
2	Yes	958

10. **Revenue by Age Group –** Calculated total revenue contribution of each age group.

Data Output | Messages | Notifications

SQL

	age_group character varying (30)	revenue_contribution bigint
1	Middle-aged	58864
2	Young Adult	58705
3	Adult	57991
4	Senior	57521

5. Business Recommendations

- **Boost Subscriptions –** Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs** – Reward repeat buyers to move them into the “Loyal” segment.
- **Review Discount Policy –** Balance sales boosts with margin control.
- **Product Positioning –** Highlight top-rated and best-selling products in campaigns.
- **Targeted Marketing –** Focus efforts on high-revenue age groups and express-shipping users.