Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes 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

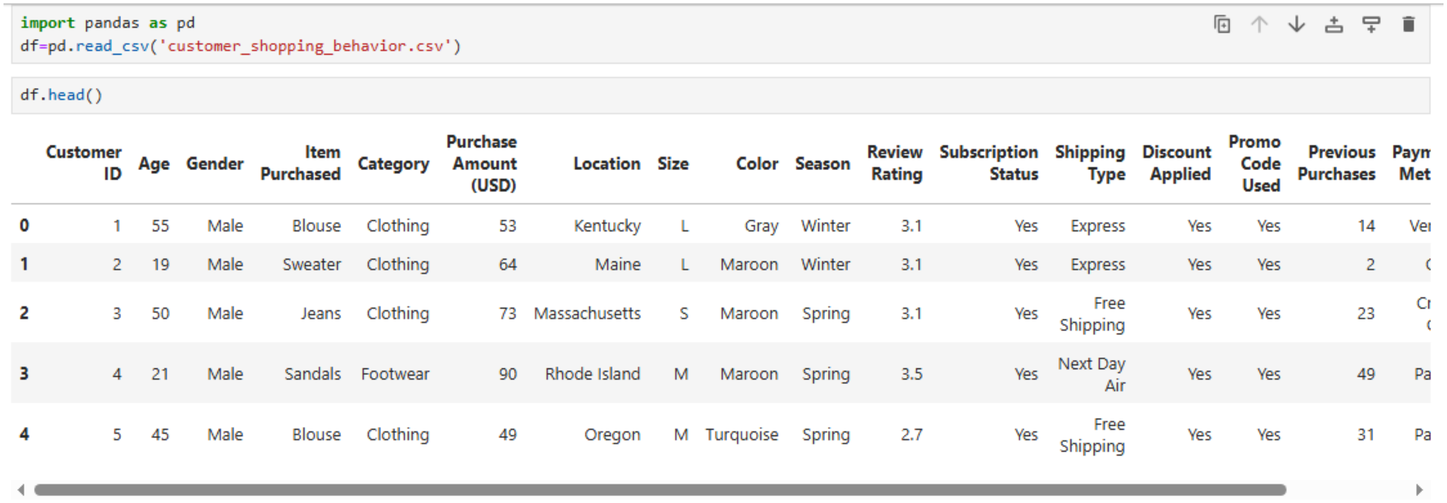
- Rows: 3,900
- Columns: 18
- Key Features:
- Customer demographics (Age, Gender, Location, Subscription Status)
- Purchase details (Item Purchased, Category, Purchase Amount, Season, Size, Color)
- Shopping behavior (Discount Applied, Promo Code Used, Previous Purchases, Frequency of Purchases, Review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

3. Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

● Data Loading: Imported the dataset using pandas.

● Initial Exploration: Used df.info() to check structure and df.describe() for summary statistics.



```
import pandas as pd
df=pd.read_csv('customer_shopping_behavior.csv')
```

```
df.head()
```

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Paym Met
0	1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Ver
1	2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	C
2	3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Cr C
3	4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	Pa
4	5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	Pa

Missing Data Handling: Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

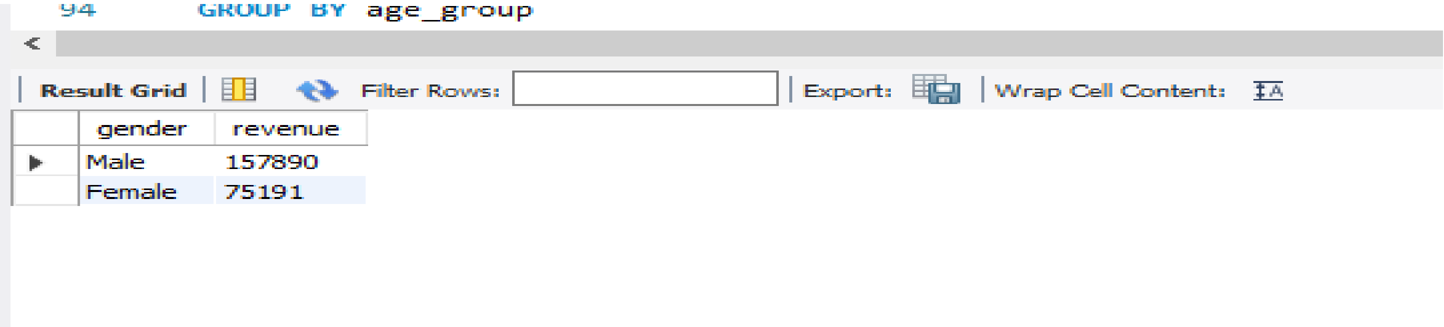
- Column Standardization: Renamed columns to **snake case** for better readability and documentation.
- Feature Engineering:
  - Created **age_group** column by binning customer ages.
  - Created **purchase_frequency_days** column from purchase data.
- Data Consistency Check: Verified if discount_applied and promo_code_used were redundant;

dropped promo_code_used.

- Database Integration: Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis.

4. Data Analysis using SQL (Business Transactions)

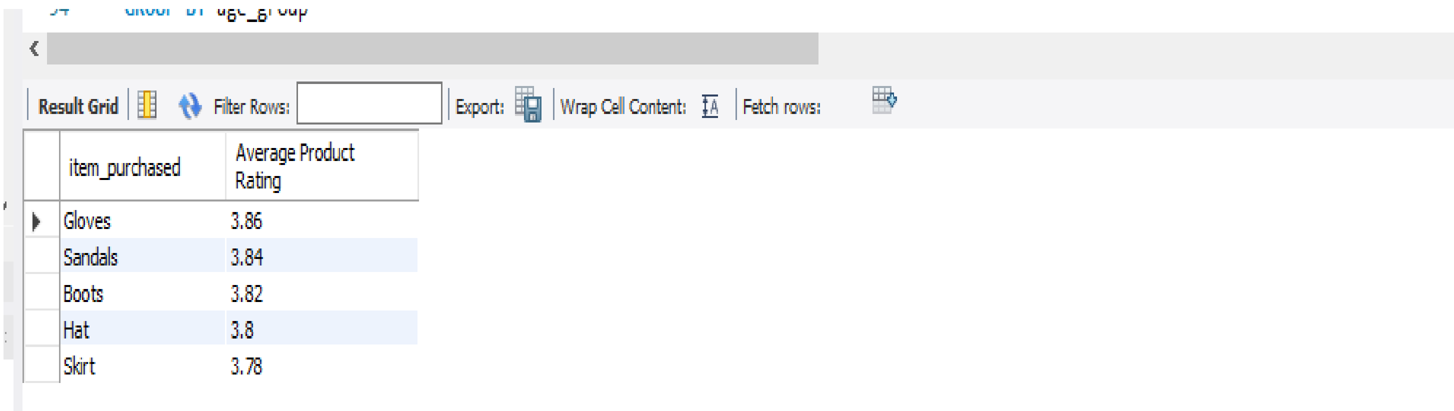
We performed structured analysis in PostgreSQL to answer key business questions:



```
94        GROUP BY age_group
```

gender	revenue
Male	157890
Female	75191

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

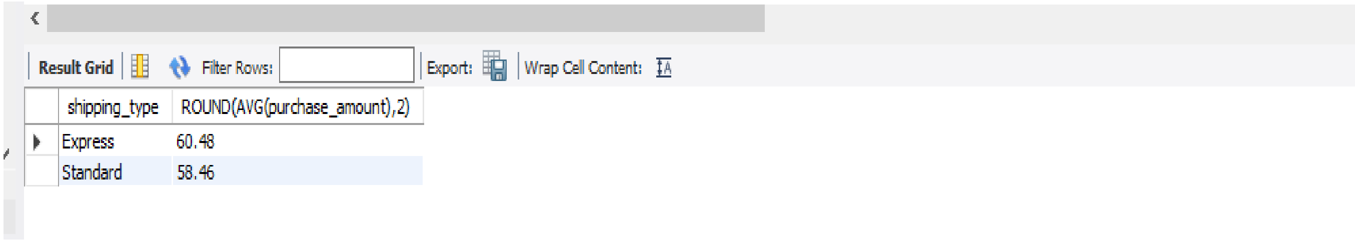


```
94        GROUP BY age_group
```

item_purchased	Average Product Rating
Gloves	3.86
Sandals	3.84
Boots	3.82
Hat	3.8
Skirt	3.78

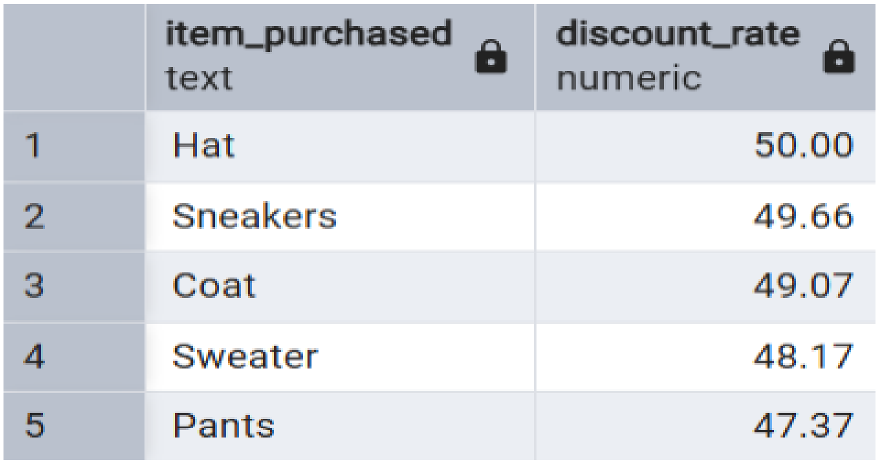
2. **Top 5 Products by Rating** – Found products with the highest average review ratings

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



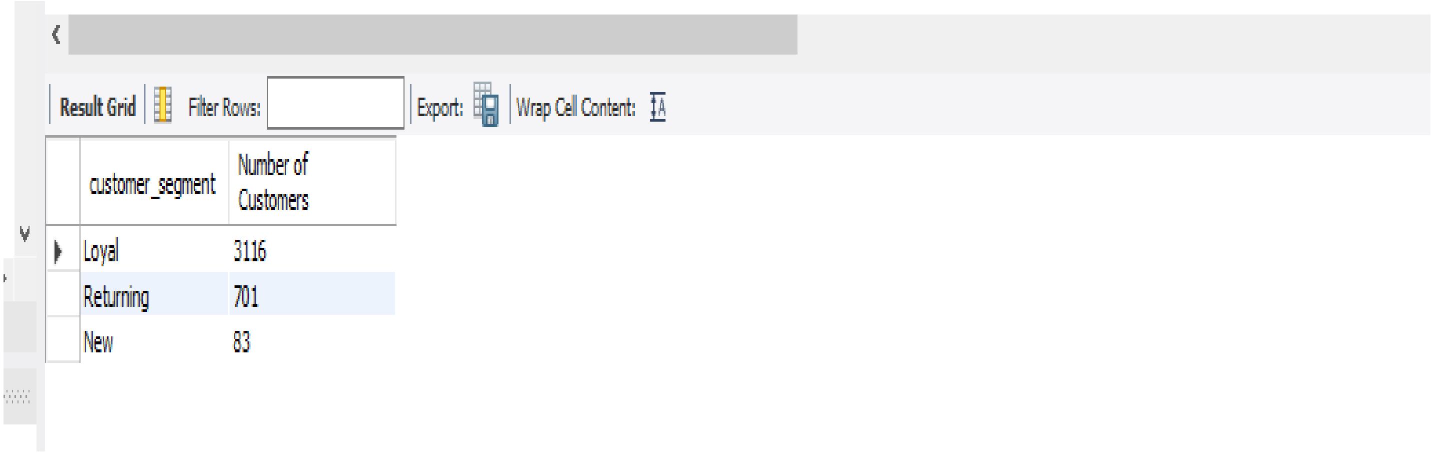
shipping_type	ROUND(AVG(purchase_amount),2)
Express	60.48
Standard	58.46

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



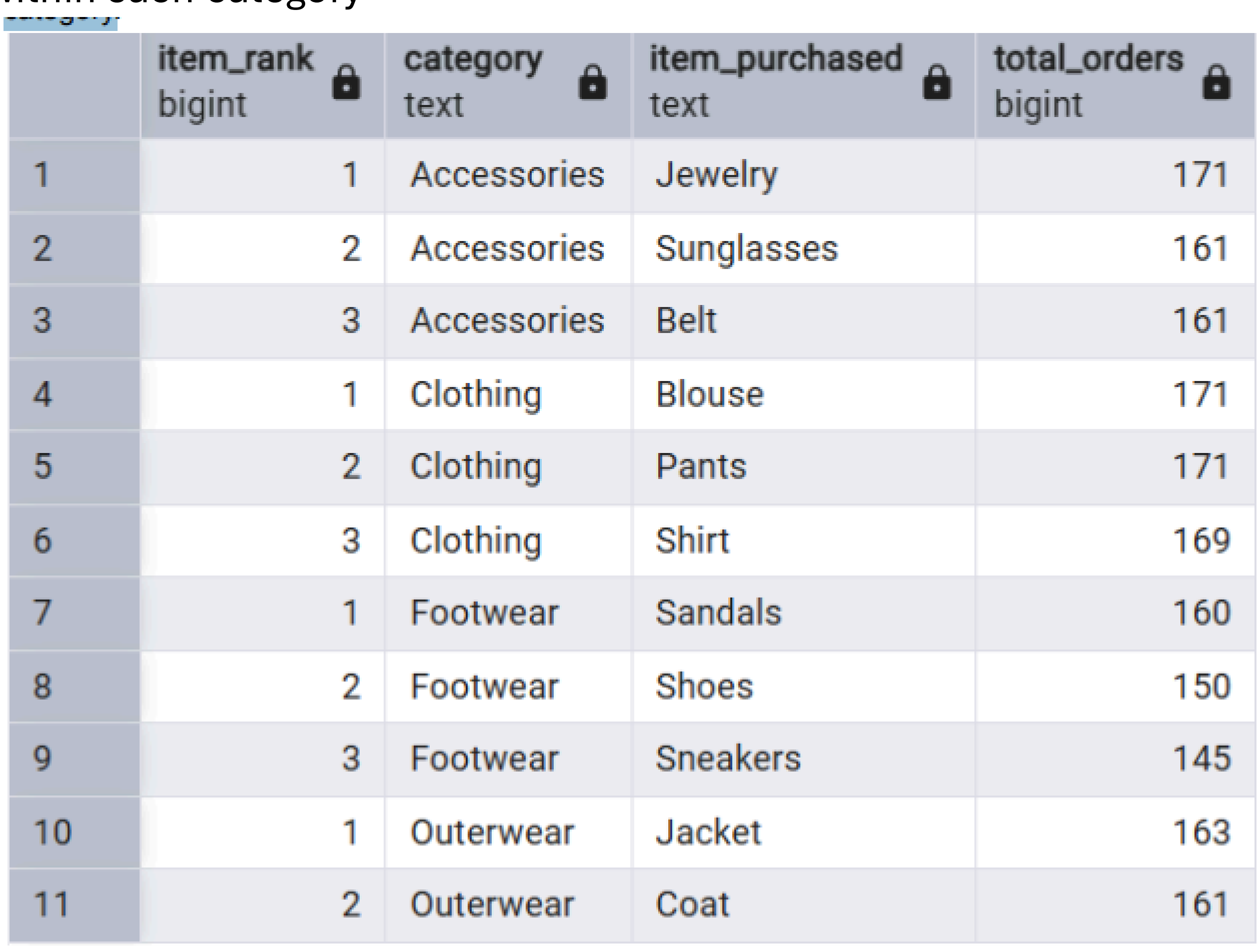
	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

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



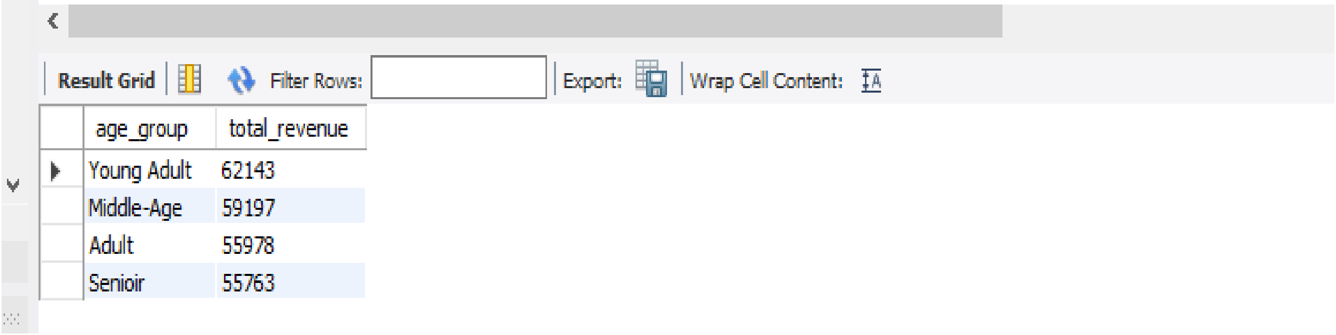
customer_segment	Number of Customers
Loyal	3116
Returning	701
New	83

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



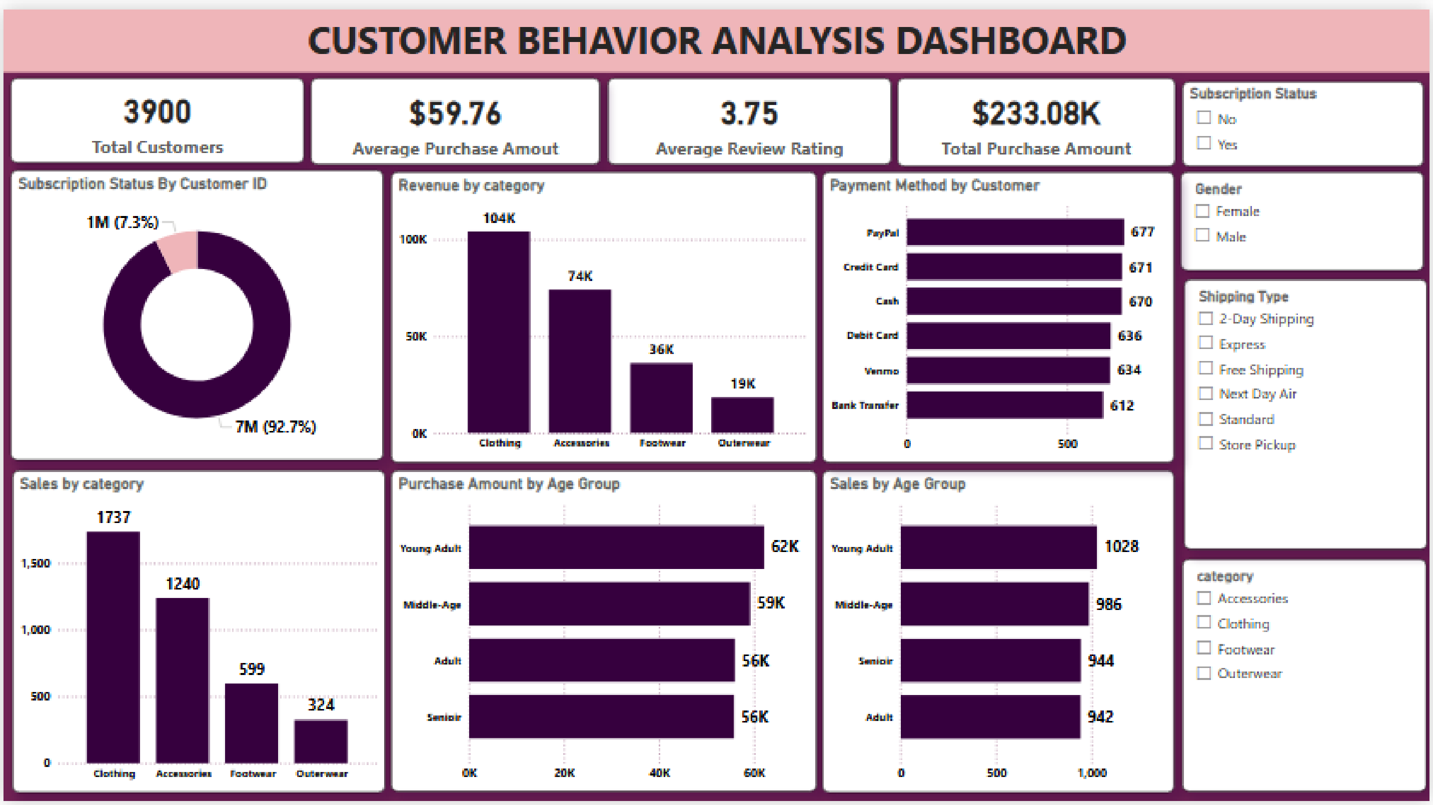
	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10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

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



age_group	total_revenue
Young Adult	62143
Middle-Age	59197
Adult	55978
Senioir	55763

8. **Dashboard in Power BI**- Finally, we built an interactive dashboard in Power BI to present insights visually.



5. <u>Business Recommendations-</u>

- **Boost Subscriptions** – Promote exclusive benefits for subscribers.
- **Customer Loyalty Programs** – Reward repe-at 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

~Muhammad Samad Qureshi