Customer Shopping Behavior Analysis

1. Project Overview

This project provides a comprehensive end-to-end analysis of retail transactional data, covering 3,900 purchases across diverse product categories. By integrating **Python** for data engineering, **SQL** for deep-dive business logic, and **Power BI** for visual storytelling, the study uncovers the underlying drivers of customer spending, subscription stickiness, and product performance.

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 .describe() for summary statistics.

● **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.

● **Data Consistency Check:** Verified if discount_applied and promo_code_used were redundant; dropped promo_code_used.

● **Database Integration:** Connected Python script to MySQL and loaded the cleaned DataFrame into the database for SQL analysis

4. Data Analysis using SQL (Business Transactions)

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

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

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

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

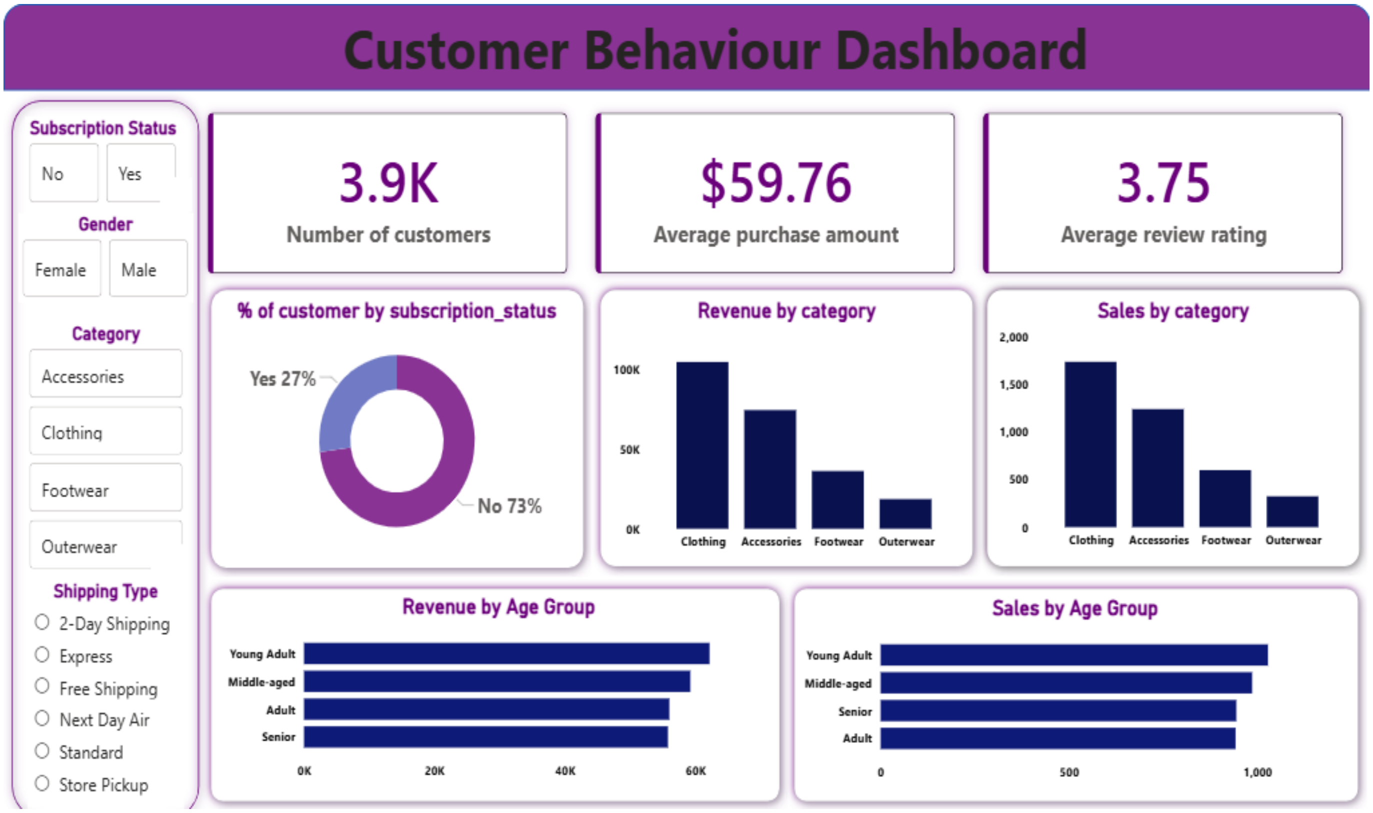
4. **Discount-Dependent Products –** Identified 5 products with the highest percentage discounted purchases.

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

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

5. Dashboard in Power BI

Finally, we built an interactive dashboard in Power BI to present insights visually



6. 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.