Customer Shopping Behavior Analysis

Comprehensive Data Insight & Strategic Recommendations

◎ 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.

⊟ Dataset Summary

3,900	18
TOTAL ROWS	COLUMNS

37

MISSING RATINGS

👤 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

</> Exploratory Data Analysis (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.
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.

▦ SQL Analysis (Business Transactions)

Structured analysis performed in PostgreSQL to answer key business questions:

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

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

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

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

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

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

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

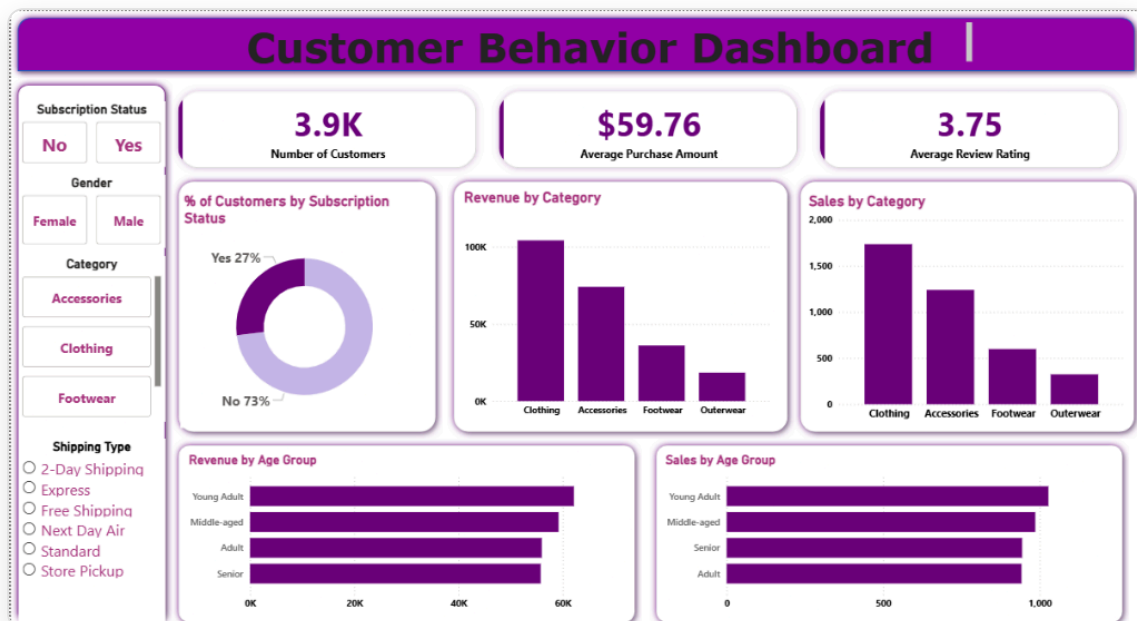8. **Top 3 Products per Category** – Listed the most purchased products within each category.

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

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

📊 Dashboard in Power BI

Interactive Visualization

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

💡 Business Recommendations

GROWTH STRATEGY	RETENTION
Boost Subscriptions	**Customer Loyalty Programs**

Promote exclusive benefits for subscribers to increase recurring revenue.

Reward repeat buyers to move them into the "Loyal" segment.

FINANCIALS

Review Discount Policy

Balance sales boosts with margin control to ensure profitability.

MARKETING

Product Positioning

Highlight top-rated and best-selling products in marketing campaigns.

SEGMENTATION

Targeted Marketing

Focus efforts on high-revenue age groups and express-shipping users for maximum ROI.