

Customer Shopping Behavior Analysis

Name: M Prathyusha

Phone: 9342594979

Email: santhoprathi26@gmail.com

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
- 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

- Data Loading: Imported dataset using pandas.
- Initial Exploration: Used `df.info()` and `df.describe()` for structure and summary statistics.
- Missing Data Handling: Imputed missing Review Rating values using median rating of each product category.
- Column Standardization: Renamed columns to snake_case format.
- Feature Engineering: Created `age_group` and `purchase_frequency_days` columns.
- Data Consistency Check: Verified `discount_applied` and `promo_code_used` redundancy and removed `promo_code_used`.
- Database Integration: Loaded cleaned data into PostgreSQL for SQL analysis.

4. Data Analysis using SQL

- Revenue by Gender: Female = 75,191 | Male = 157,890
- High-Spending Discount Users: Identified customers spending above average with discounts.
- Subscribers vs Non-Subscribers:
 - - Yes: 1,053 customers | Avg Spend: 59.49 | Total Revenue: 62,645
 - - No: 2,847 customers | Avg Spend: 59.87 | Total Revenue: 170,436
- Top 5 Highest Rated Products identified using average review rating.
- Top 3 Most Purchased Products within each category using window functions.
- Repeat Buyers Subscription Analysis performed for customers with more than 5 previous purchases.
- Revenue contribution analyzed across age groups.

5. Power BI Dashboard KPIs

- Number of Customers: 3.9K
- Average Purchase Amount: \$59.76
- Average Review Rating: 3.75
- Revenue by Category analysis performed.
- Sales by Age Group visualization created.
- Subscription percentage analysis (Yes 27% | No 73%).

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.