Customer Shopping Behavior Analysis

# 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

* 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

# 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.

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

* **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.

# Data Analysis using SQL (Business Transactions)

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

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

A screenshot of a social media account

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a number of products

Description automatically generated

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

A screenshot of a phone

Description automatically generated

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

A screenshot of a phone

Description automatically generated

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

A screenshot of a shopping list

Description automatically generated

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

A screenshot of a phone

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

# Dashboard in Tableau

Finally, we built an interactive dashboard in **Tableau** to present insights visually.

A screenshot of a data dashboard

Description automatically generated

Suggestions

Here are some strategic suggestions based on this customer behavior dashboard:

**Key Insights & Recommendations**

**1. Low Average Purchase Amount ($9.76)**

* This is quite low. Consider implementing:
  + Bundle deals or "frequently bought together" promotions
  + Minimum purchase thresholds for free shipping
  + Volume discounts to encourage larger orders

**2. Subscription Opportunity (73% Not Subscribed)**

* Huge untapped potential here with only 26.88% subscribed
* Create compelling subscription incentives:
  + 10-15% discount for subscribers
  + Exclusive early access to new products
  + Free shipping for subscription members
* Target your Middle-age and Senior segments (they show high purchase volume)

**3. Young Adult Segment Dominance**

* Young Adults generate the most revenue but may have growth limits
* Diversify by targeting Adult and Senior segments more aggressively
* These demographics typically have higher disposable income

**4. Payment Method Insights**

* Venmo shows lowest revenue despite being available
* Cash and Credit Card dominate - ensure these processes are seamless
* Consider promoting digital wallets for faster checkout

**5. Seasonal Revenue Pattern**

* Clear peak in Fall season
* Prepare inventory and marketing campaigns around this trend
* Create promotions during slower seasons (Spring appears weakest)

**6. Top 10 Items Strategy**

* Revenue concentration in specific items (Blouses, Dress, Jewelry)
* Cross-sell complementary products
* Analyze why these items perform well and replicate those features

**7. Review Rating (3.75/5)**

* Good but not excellent - address common complaints
* Incentivize reviews from satisfied customers
* This could be holding back conversion rates