

Customer Shopping Behavior Analysis

I. Project Overview

Uncovering Shopping Insights from Transactional Data

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

Dataset Size

- 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:

01	02	03
Data Loading	Initial Exploration	Missing Data Handling
Imported the dataset using pandas.	Used <code>df.info()</code> to check structure and <code>.describe()</code> for summary statistics.	Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.
04	05	06
Column Standardization	Feature Engineering	Data Consistency Check
Renamed columns to snake case for better readability and documentation.	<ul style="list-style-type: none">Created age_group column by binning customer ages.Created purchase_frequency_days column from purchase data.	Verified if <code>discount_applied</code> and <code>promo_code_used</code> were redundant; dropped <code>promo_code_used</code> .
07		
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:

1. Revenue by Gender

Compared total revenue generated by male vs. female customers.

Female	Male
\$75,191	\$157,890

2. High-Spending Discount Users

Identified customers who used discounts but still spent above the average purchase amount.

Total rows: 839 customers used discounts while spending above average.

3. Top 5 Products by Rating

Found products with the highest average review ratings.

- Gloves - 3.86
- Sandals - 3.84
- Boots - 3.82
- Hat - 3.80
- Skirt - 3.78

4. Shipping Type Comparison

Compared average purchase amounts between Standard and Express shipping.

- Standard: \$58.46
- Express: \$60.48

5. Subscribers vs. Non-Subscribers

Compared average spend and total revenue across subscription status.

Subscribers (Yes)	Non-Subscribers (No)
<ul style="list-style-type: none">Total Customers: 1,053Avg Spend: \$59.49Total Revenue: \$62,645	<ul style="list-style-type: none">Total Customers: 2,847Avg Spend: \$59.87Total Revenue: \$170,436

6. Discount-Dependent Products

Identified 5 products with the highest percentage of discounted purchases.

- Hat - 50.00%
- Sneakers - 49.66%
- Coat - 49.07%
- Sweater - 48.17%
- Pants - 47.37%

7. Customer Segmentation

Classified customers into New, Returning, and Loyal segments based on purchase history.

3116	701	83
Loyal	Returning	New

8. Top 3 Products per Category

Listed the most purchased products within each category.

Accessories	Clothing	Footwear	Outerwear
<ol style="list-style-type: none">Jewelry - 171Sunglasses - 161Belt - 161	<ol style="list-style-type: none">Blouse - 171Pants - 171Shirt - 169	<ol style="list-style-type: none">Sandals - 160Shoes - 150Sneakers - 145	<ol style="list-style-type: none">Jacket - 163Coat - 161

9. Repeat Buyers & Subscriptions

Checked whether customers with >5 purchases are more likely to subscribe.

- No subscription: 2,518 repeat buyers
- Yes subscription: 958 repeat buyers

10. Revenue by Age Group

Calculated total revenue contribution of each age group.

<div><div></div><div>27%</div></div> <div>Young Adult</div> <div>\$62,143</div>	<div><div></div><div>25%</div></div> <div>Middle-aged</div> <div>\$59,197</div>
<div><div></div><div>24%</div></div> <div>Adult</div> <div>\$55,978</div>	<div><div></div><div>24%</div></div> <div>Senior</div> <div>\$55,763</div>

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.