Customer Shopping Behavior Analysis

This project analyzes customer shopping behavior using transactional data from 3,900 purchases. Our goal is to uncover insights into spending patterns, customer segments, product preferences, and subscription behavior to guide strategic business decisions.

Dataset Summary

Our dataset comprises 3,900 rows and 18 columns, offering a comprehensive view of customer interactions.

- **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 the Review Rating column were imputed using the median rating per product category.

Exploratory Data Analysis: Python

Data Loading & Exploration

Imported dataset with pandas, checked structure with `df.info()` and summary statistics with `.describe()`.

Missing Data Handling

Imputed missing 'Review Rating' values using the median rating of each product category.

Column Standardization

Renamed columns to **snake_case** for improved readability and consistency.

Feature Engineering

Created **age_group** by binning ages and **purchase_frequency_days** from purchase data.

Data Analysis: SQL Insights

Structured analysis in PostgreSQL answered key business questions, revealing critical patterns in customer behavior.

Revenue by Gender

Male customers generated significantly more revenue ($157,890) than female customers ($75,191).

High-Spending Discount Users

Identified 839 customers who used discounts but still spent above the average purchase amount.

Top 5 Products by Rating

Gloves (3.86), Sandals (3.84), Boots (3.82), Hat (3.80), and Skirt (3.78) received the highest average ratings.

Shipping & Subscription Impact

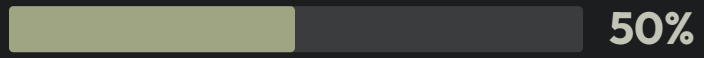
Shipping Type Comparison

Express shipping users had a higher average purchase amount ($60.48) compared to Standard shipping ($58.46).

Subscribers vs. Non-Subscribers

Non-subscribers (2,847 customers) contributed more total revenue ($170,436) than subscribers (1,053 customers, $62,645).

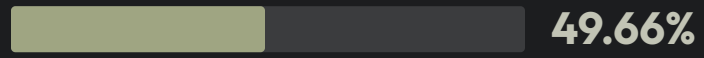
Discount & Product Insights



50%

Hat

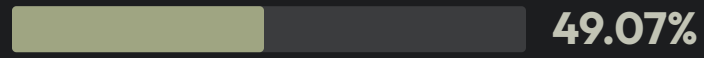
Highest percentage of discounted purchases.



49.66%

Sneakers

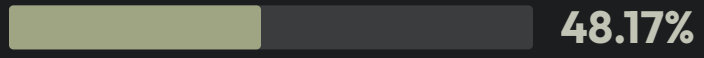
Second highest discount dependency.



49.07%

Coat

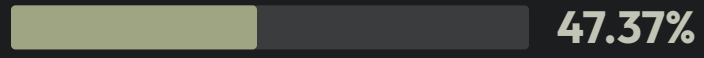
Significant portion of sales with discounts.



48.17%

Sweater

High reliance on discounts for sales.



47.37%

Pants

Nearly half of purchases involved discounts.

These products show the highest dependency on discounts for sales, indicating potential for strategic pricing adjustments.

Customer Segmentation

Customers were classified into three segments based on their purchase history.

Loyal

3,116 Customers

Frequent and consistent buyers.

New

83 Customers

First-time purchasers.

Returning

701 Customers

Customers with multiple, but not frequent, purchases.

Top Products by Category

1	Accessories	Jewelry	171
1	Clothing	Blouse	171
1	Footwear	Sandals	160
1	Outerwear	Jacket	163

This table highlights the most purchased items within each product category, providing insights for inventory and marketing strategies.

Repeat Buyers & Age Group Revenue

Repeat Buyers & Subscriptions

Customers with more than 5 purchases are more likely to subscribe. 958 repeat buyers are subscribers, compared to 2,518 non-subscribers.

Revenue by Age Group

Young Adults (62,143) and Middle-aged (59,197) groups contribute the most revenue, followed by Adults (55,978) and Seniors (55,763).

Business Recommendations

1

Boost Subscriptions

Promote exclusive benefits to increase subscriber base and engagement.

2

Customer Loyalty Programs

Reward repeat buyers to foster loyalty and transition them into the "Loyal" segment.

3

Review Discount Policy

Strategically adjust discount offerings to balance sales boosts with profit margins.

4

Product Positioning

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

5

Targeted Marketing

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