



Customer Shopping Behavior Analysis

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

Project Overview & Dataset Summary



Project Goal

Uncover insights into spending patterns, customer segments, product preferences, and subscription behavior.



Dataset Size

3,900 purchases across various product categories with 18 columns.



Key Features

Demographics, purchase details, shopping behavior, and subscription status.



Missing Data

37 values in the Review Rating column were identified.



Exploratory Data Analysis in Python



Data Loading

Imported dataset using `pandas` for initial processing.



Initial Exploration

Used `df.info()` and `.describe()` for structural and statistical summaries.



Missing Data Handling

Imputed missing values in 'Review Rating' using the median rating per product category.



Feature Engineering

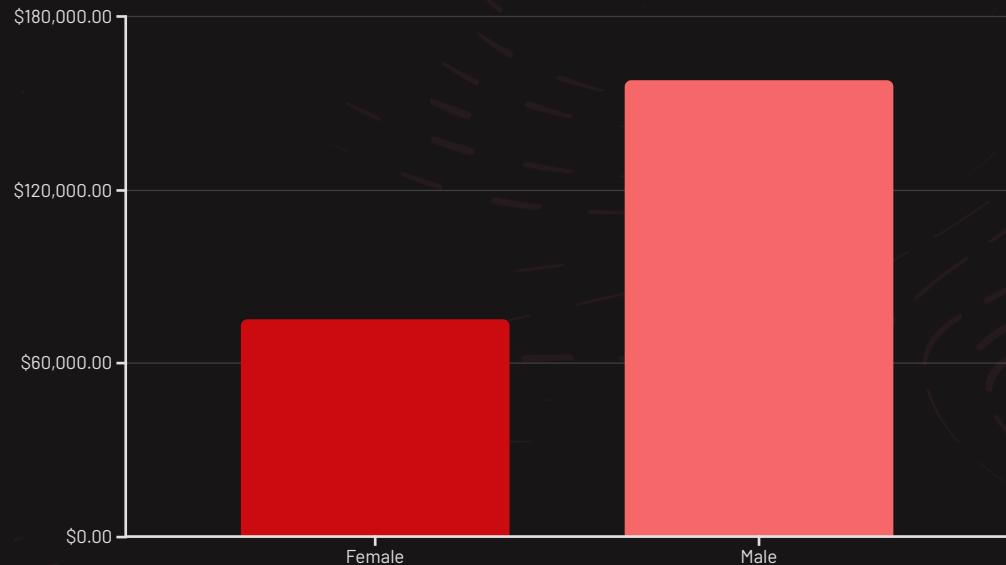
Created 'age_group' and 'purchase_frequency_days' columns.



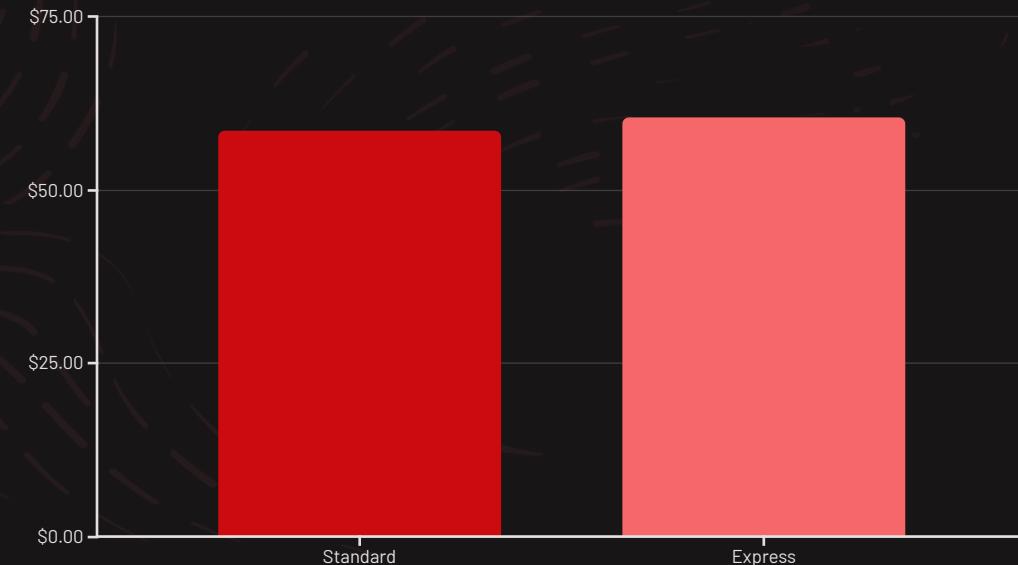
Database Integration

Loaded cleaned DataFrame into PostgreSQL for SQL analysis.

Revenue by Gender & Shipping Type



Male customers generated significantly higher revenue than female customers.



Express shipping users have a slightly higher average purchase amount.



High-Spending Discount Users & Top Products

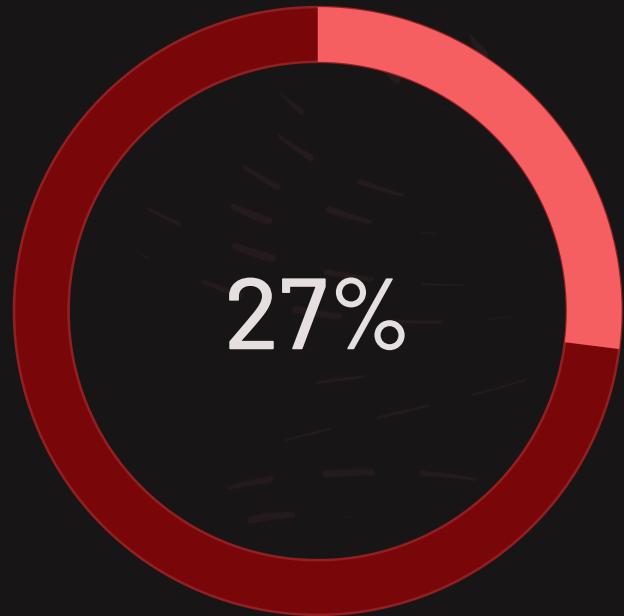
High-Spending Discount Users

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

Top 5 Products by Rating

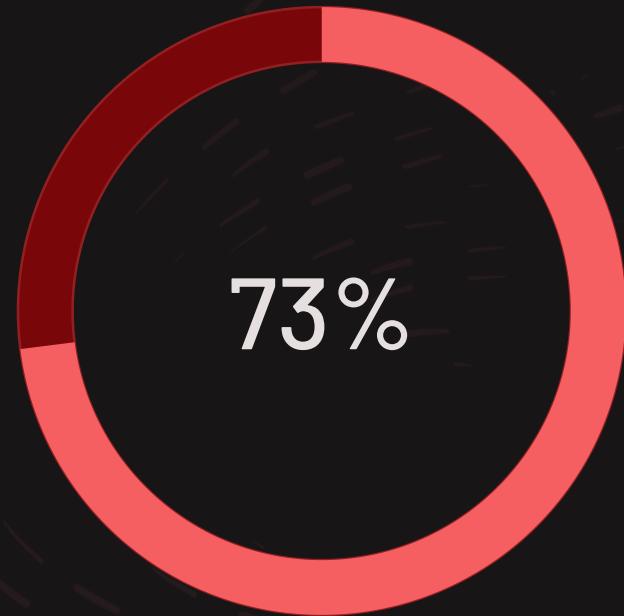
Gloves, Sandals, Boots, Hat, and Skirt received the highest average review ratings.

Subscribers vs. Non-Subscribers



Subscribers

1,053 customers with an average spend of \$59.49,
generating \$62,645 in total revenue.



Non-Subscribers

2,847 customers with an average spend of \$59.87,
generating \$170,436 in total revenue.

Non-subscribers represent a larger customer base and higher total revenue, though average spend is similar.



Discount-Dependent Products & Customer Segmentation

Discount-Dependent Products

- Hat (50.00% discounted purchases)
- Sneakers (49.66%)
- Coat (49.07%)
- Sweater (48.17%)
- Pants (47.37%)

These products show the highest reliance on discounts for sales.

Customer Segmentation

- Loyal: 3,116 customers
- Returning: 701 customers
- New: 83 customers

A large loyal customer base, with opportunities to convert new and returning buyers.



Top Products per Category & Repeat Buyers

Top 3 Products per Category

- Accessories:** Jewelry, Sunglasses, Belt
- Clothing:** Blouse, Pants, Shirt
- Footwear:** Sandals, Shoes, Sneakers
- Outerwear:** Jacket, Coat

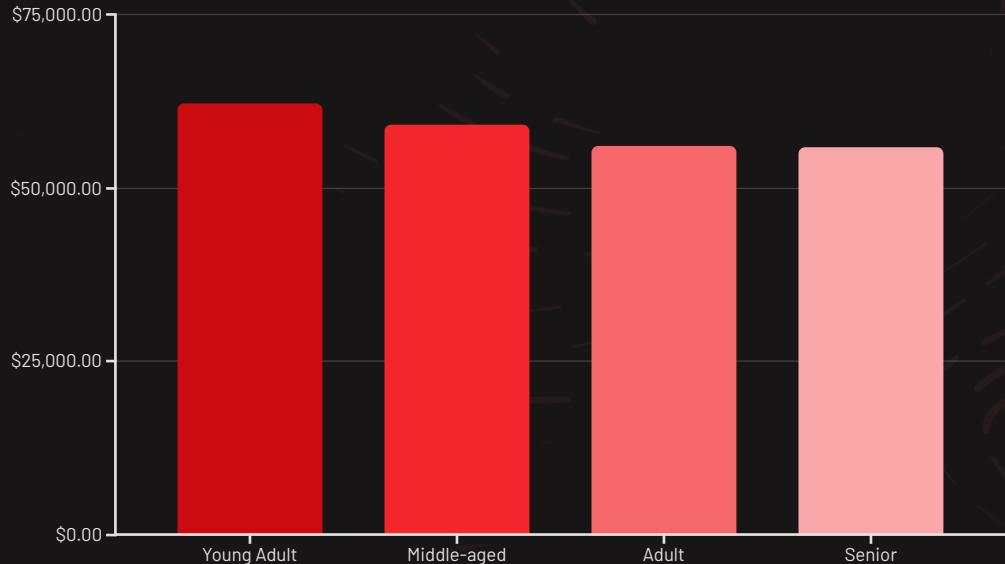
Repeat Buyers & Subscriptions

Customers with more than 5 purchases:

- No Subscription: 2,518
- Yes Subscription: 958

Repeat buyers are more likely to subscribe, but a significant portion remains unsubscribed.

Revenue by Age Group & Dashboard Preview



Young Adults contribute the highest revenue, followed closely by Middle-aged customers.

Customer Behavior Dashboard



Interactive Power BI dashboard for visual insights into customer behavior.

Strategic Business Recommendations



Boost Subscriptions

Promote exclusive benefits to increase subscriber base.



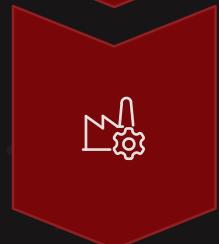
Customer Loyalty Programs

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



Review Discount Policy

Balance sales boosts with margin control, especially for discount-dependent products.



Product Positioning

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



Targeted Marketing

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