

Customer Shopping Behavior Analysis

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



Project Overview & Dataset Summary

Project Goal

Uncover insights from transactional data to inform business strategy.

Dataset Details

- Rows: 3,900
- Columns: 18
- Missing Data: 37 values in Review Rating



Key Features

- Customer demographics (Age, Gender, Location, Subscription)
- Purchase details (Item, Category, Amount, Season, Size, Color)
- Shopping behavior (Discount, Promo, Previous Purchases, Frequency, Review, Shipping)

Exploratory Data Analysis (Python)

We initiated data preparation and cleaning using Python, ensuring data quality for analysis.

1

Data Loading & Exploration

Imported dataset with pandas; checked structure and summary statistics.

2

Missing Data Handling

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

3

Column Standardization

Renamed columns to snake_case for improved readability.

4

Feature Engineering

Created 'age_group' and 'purchase_frequency_days' columns.

Data Analysis using SQL: Key Insights

Structured analysis in PostgreSQL answered critical business questions.



Revenue by Gender

Male customers generated \$157,890, while females generated \$75,191.



High-Spending Discount Users

Identified 839 customers who used discounts but spent above average.



Top 5 Products by Rating

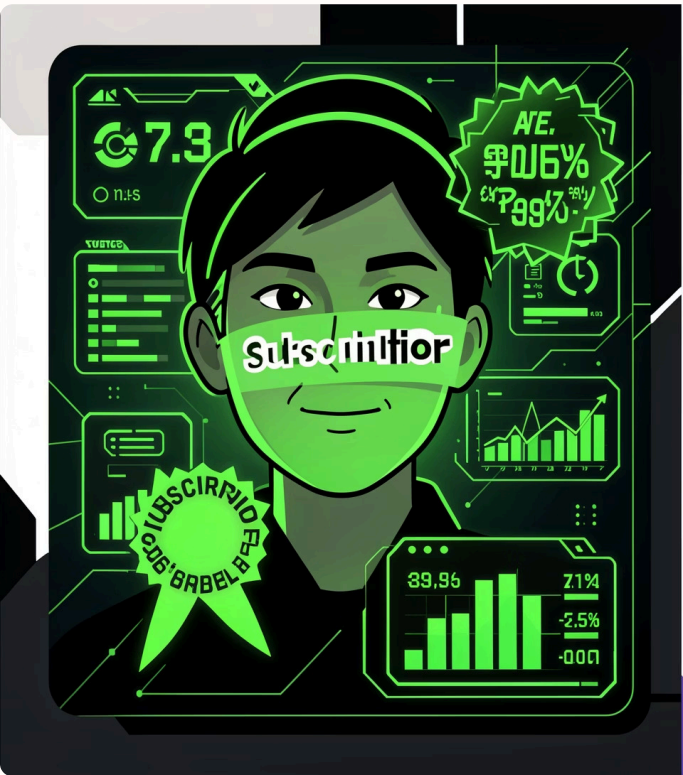
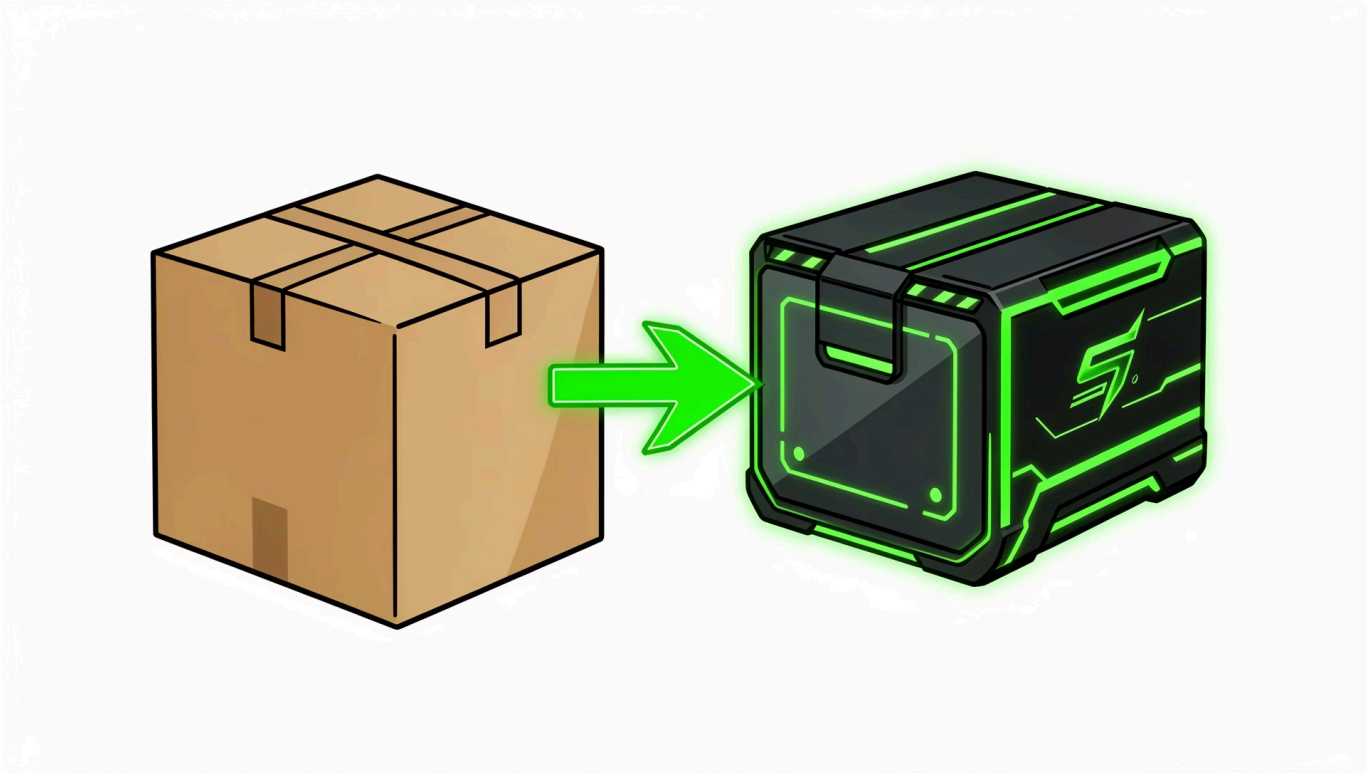
Gloves (3.86), Sandals (3.84), Boots (3.82), Hat (3.80), Skirt (3.78).

Shipping & Subscription Analysis

Shipping Type Comparison

Average purchase amounts:

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



Subscribers vs. Non-Subscribers

Subscribers (1053 customers) had an average spend of \$59.49 and total revenue of \$62,645. Non-subscribers (2847 customers) averaged \$59.87 with \$170,436 revenue.

Discount & Loyalty Insights



Discount-Dependent Products

Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), Pants (47.37%) had highest discounted purchases.



Customer Segmentation

Loyal (3116), Returning (701), and New (83) customers based on purchase history.



Repeat Buyers & Subscriptions

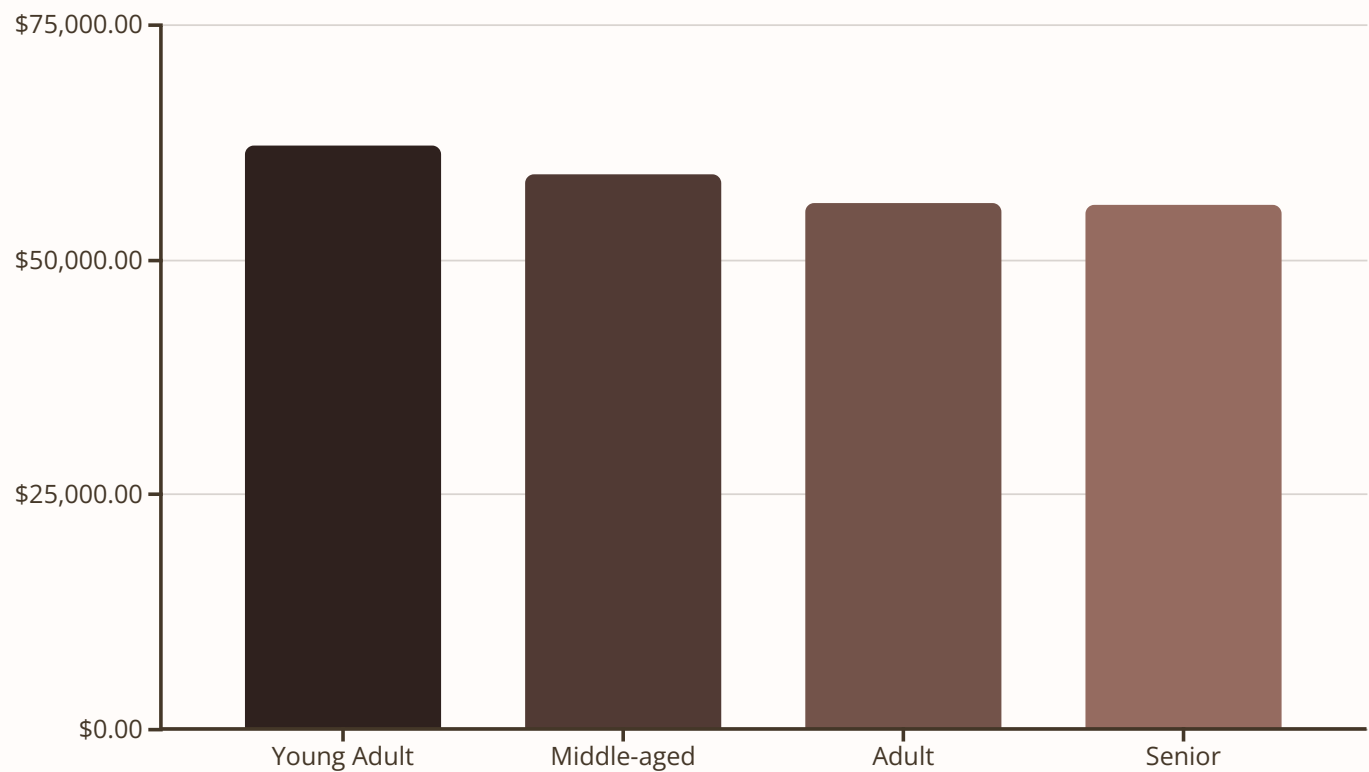
958 repeat buyers (more than 5 purchases) are subscribers, compared to 2518 non-subscribers.

Top Products by Category

Identifying the most purchased items within each product category.

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

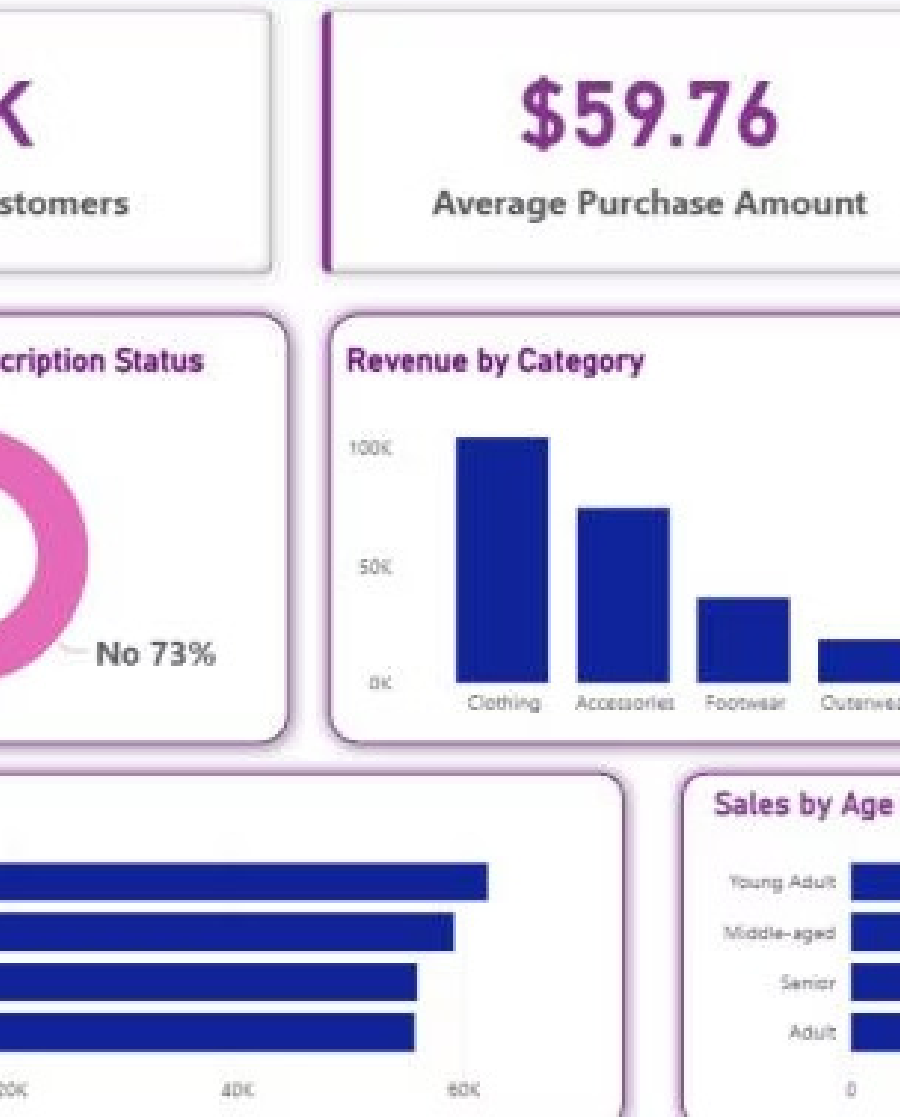
Revenue by Age Group



Young Adults contribute the highest revenue, followed closely by Middle-aged customers. This highlights the importance of targeted marketing strategies for each demographic.



Customer Behavior Dashboard



Power BI Dashboard: Visualizing Insights

The interactive Power BI dashboard provides a comprehensive visual overview of customer behavior, including subscription status, revenue by category, and sales by age group.

3.9K

Customers

\$59.76

Avg. Purchase

3.75

Avg. Rating



Business Recommendations



Boost Subscriptions

Promote exclusive benefits to increase subscriber base.



Customer Loyalty Programs

Reward repeat buyers to foster loyalty and retention.



Review Discount Policy

Optimize discounts to balance sales and profit margins.



Targeted Marketing

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