

Customer Shopping Behavior Analysis

Uncovering insights from transactional data to guide strategic business decisions.



Project Overview

1

Data Analysis

Analyzed 3,900 purchases across various product categories.

2

Key Insights

Focused on spending patterns, customer segments, and product preferences.

3

Strategic Guidance

Aimed to inform business decisions and optimize strategies.



Dataset Summary

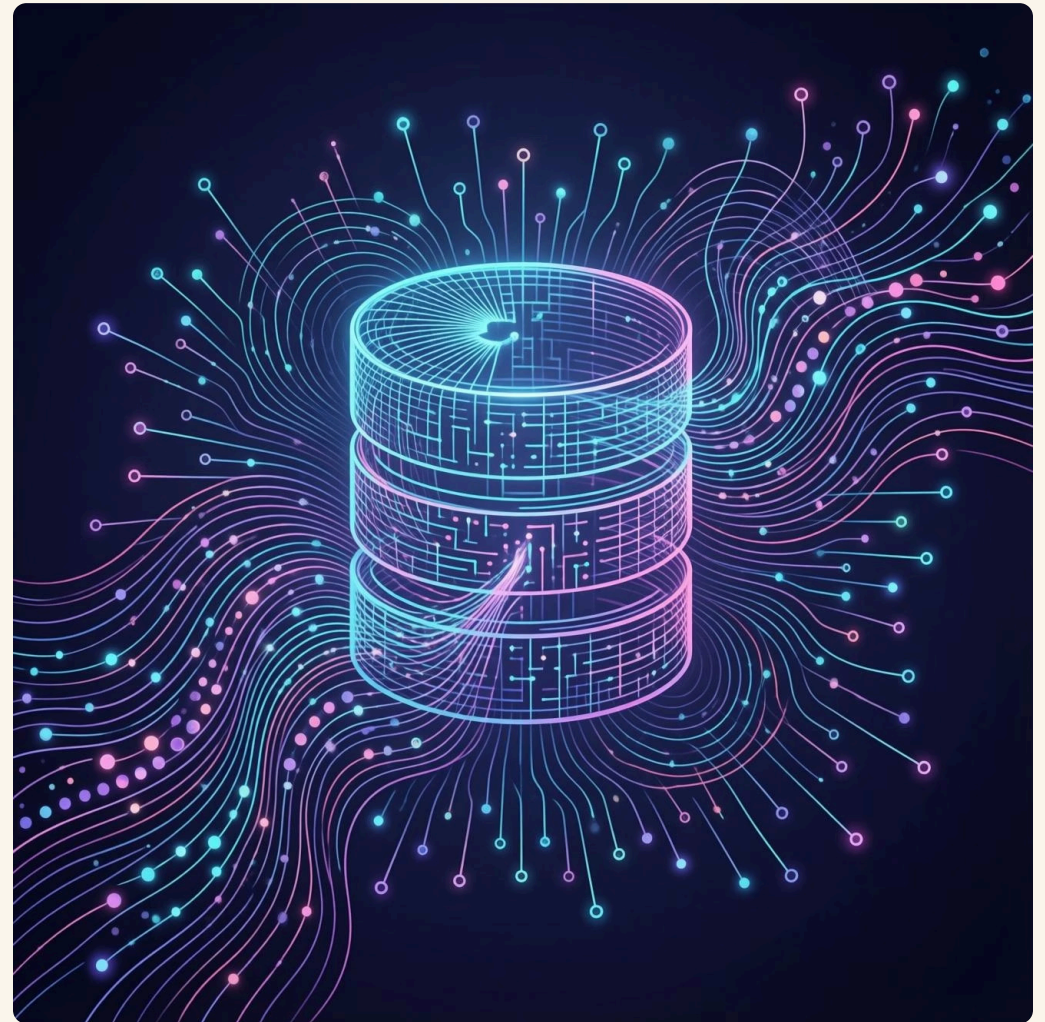
Rows: 3,900

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 Code, Previous Purchases, Frequency, Review, Shipping)



Exploratory Data Analysis (Python)

01

Data Loading & Exploration

Imported dataset using `pandas`; checked structure and summary statistics.

02

Missing Data Handling

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

03

Column Standardization

Renamed columns to snake case for readability.

04

Feature Engineering

Created 'age_group' and 'purchase_frequency_days' columns.

05

Database Integration

Loaded cleaned data into PostgreSQL for SQL analysis.

Data Analysis using SQL: Key Business Insights

1

Revenue by Gender

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

2

High-Spending Discount Users

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

3

Top 5 Products by Rating

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

SQL Insights: Spending & Subscriptions

1

Shipping Type Comparison

Express shipping users had slightly higher average purchase amounts (\$60.48) than Standard (\$58.46).

2

Subscribers vs. Non-Subscribers

Non-subscribers (2847) generated more total revenue (\$170,436) than subscribers (1053) (\$62,645).

3

Discount-Dependent Products

Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), and Pants (47.37%) had highest discount rates.



SQL Insights: Customer & Product Segmentation

1

Customer Segmentation

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

2

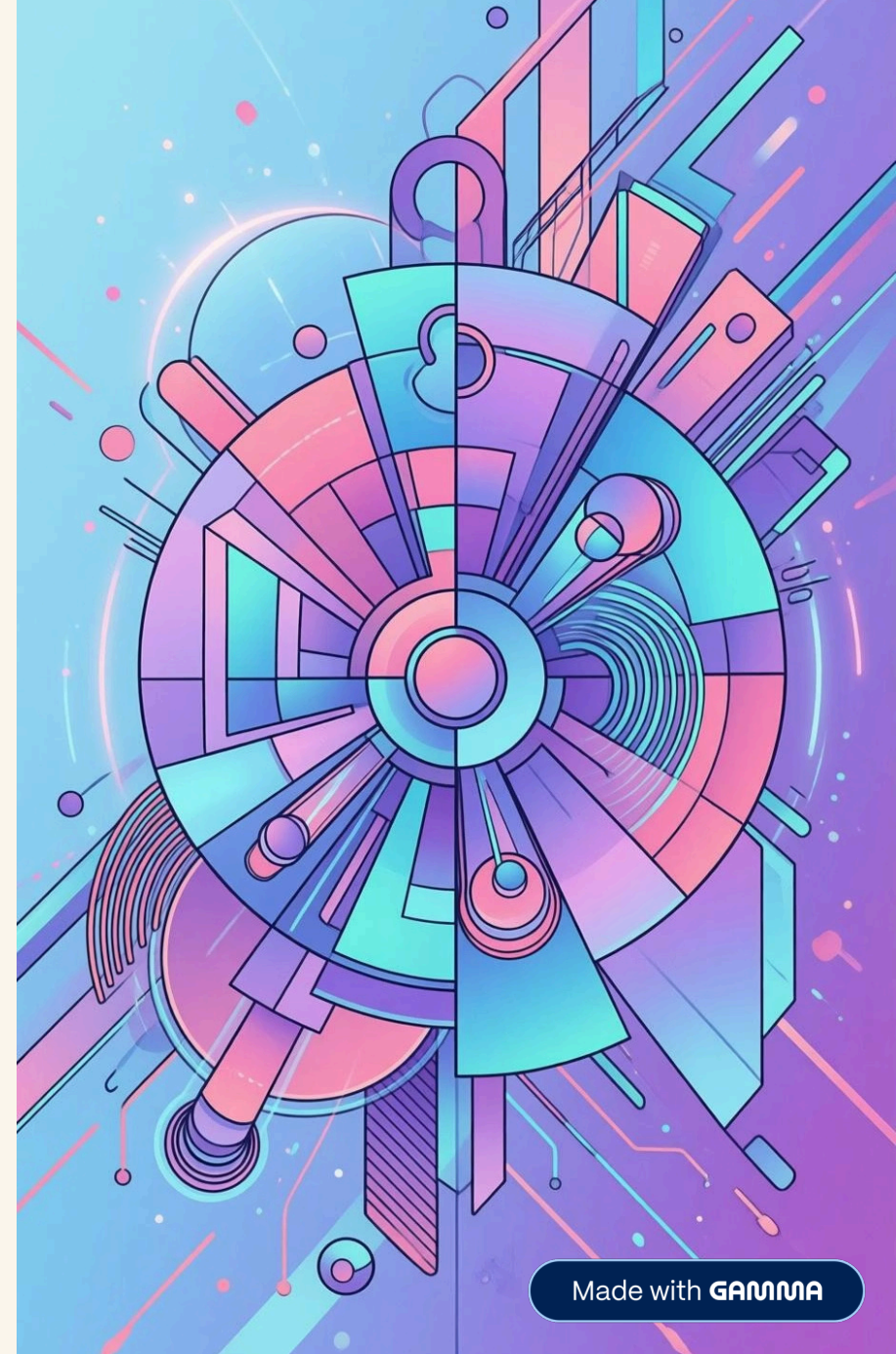
Top 3 Products per Category

Identified most purchased items within Accessories, Clothing, Footwear, and Outerwear.

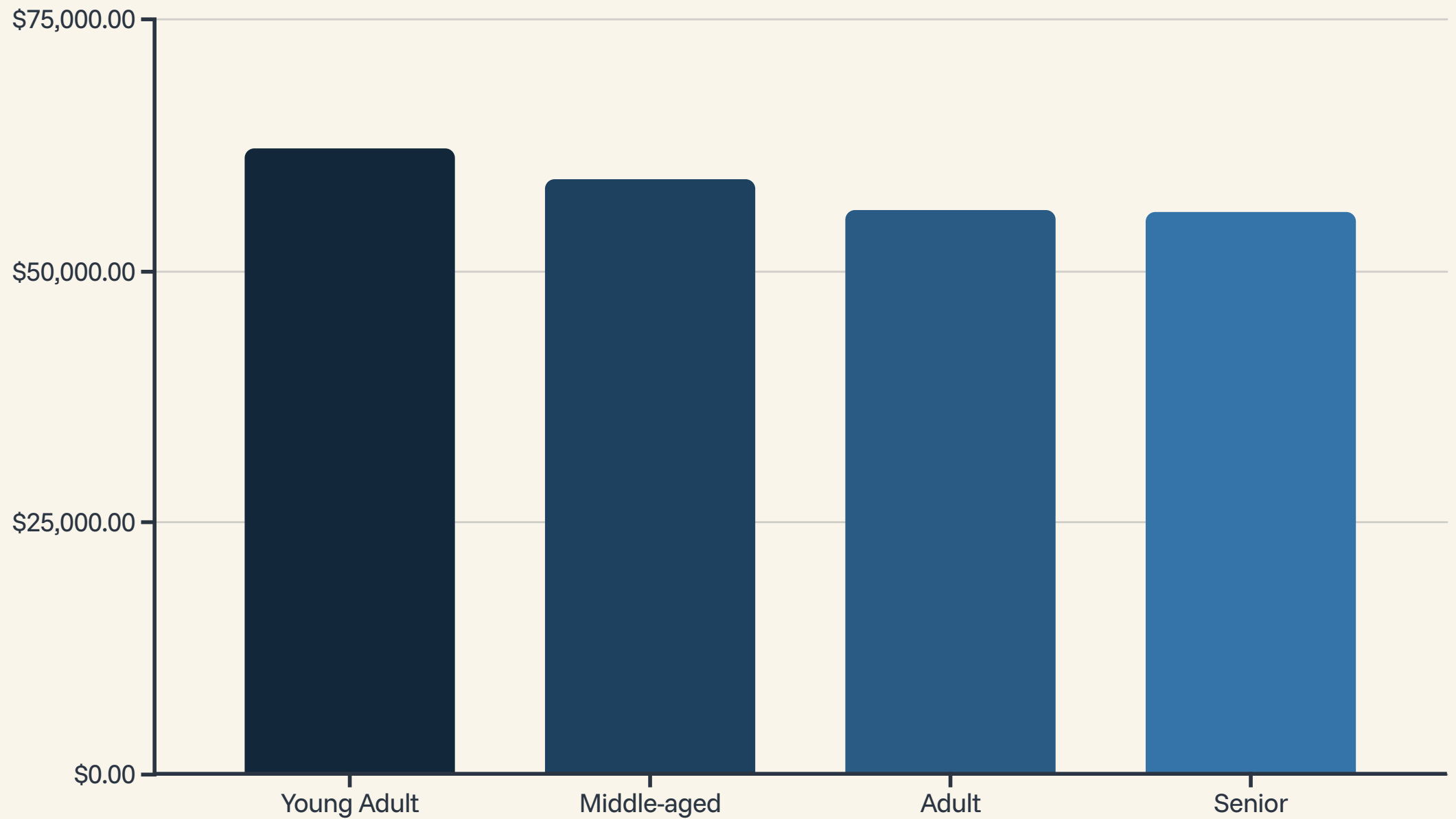
3

Repeat Buyers & Subscriptions

Customers with >5 purchases are more likely to subscribe (958 subscribers vs. 2518 non-subscribers).



Revenue by Age Group



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

Power BI Dashboard: Visualizing Insights

An interactive dashboard was built in Power BI to present key findings visually.



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

Business Recommendations



Boost Subscriptions

Promote exclusive benefits to increase subscriber base.



Customer Loyalty

Implement programs to reward repeat buyers and foster loyalty.



Review Discount Policy

Optimize discount strategies to balance sales and profit margins.



Targeted Marketing

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

