



Customer Shopping Behavior Analysis

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.

Dataset Overview

3,900

Total Purchases

Comprehensive transaction records analyzed

18

Data Columns

Rich feature set for deep analysis

50

Locations

Geographic diversity across customer base

25

Product Types

Wide variety of items purchased

Key Features Captured

- 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)

Data Quality

The dataset demonstrates excellent completeness with only 37 missing values in the Review Rating column out of 3,900 records. This represents less than 1% missing data, ensuring robust analysis capabilities.

Missing values were strategically imputed using median ratings by product category to maintain data integrity.

Data Preparation & Python Analysis

We began with comprehensive data preparation and cleaning in Python to ensure data quality and analytical readiness. The process involved multiple stages of transformation and validation.

01

Data Loading & Exploration

Imported the dataset using pandas and performed initial exploration with `df.info()` to check structure and `.describe()` for summary statistics

02

Missing Data Handling

Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category

03

Column Standardization

Renamed columns to snake_case for better readability and documentation consistency

04

Feature Engineering

Created `age_group` column by binning customer ages and `purchase_frequency_days` column from purchase data

05

Data Consistency Check

Verified if `discount_applied` and `promo_code_used` were redundant; dropped `promo_code_used` to eliminate duplication

06

Database Integration

Connected Python script to PostgreSQL and loaded the cleaned DataFrame into the database for SQL analysis

Revenue Insights by Gender & Discount Behavior

Revenue by Gender



Male customers generate significantly more revenue, accounting for **68% of total revenue** compared to female customers at 32%. This represents a substantial opportunity for targeted marketing strategies.


High-Spending Discount Users

We identified **839 customers** who used discounts but still spent above the average purchase amount of \$59.76. This represents a valuable segment that responds to promotional offers while maintaining high purchase values.


These customers demonstrate that strategic discounting doesn't necessarily mean reduced revenue—they're willing to spend more when incentivized. Purchase amounts in this segment ranged from \$62 to \$97, well above the dataset average.

- Key insight: Discount strategies can attract high-value customers without sacrificing revenue quality.


Product Performance & Customer Ratings




Gloves
Rating: **3.86**
Top-rated product across all categories




Sandals
Rating: **3.84**
Strong customer satisfaction in footwear



Boots
Rating: **3.82**
Consistent quality performance

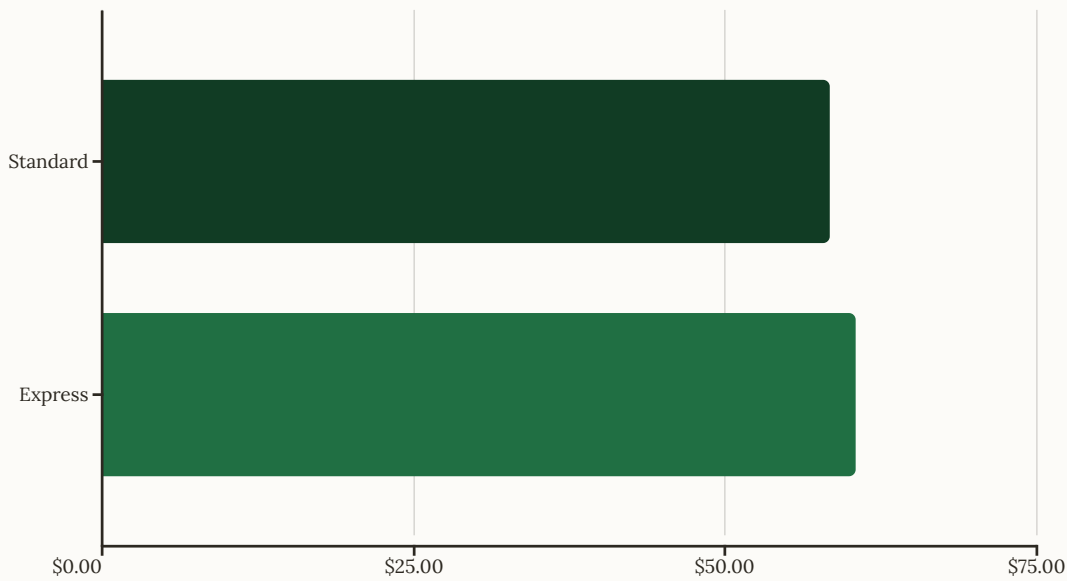


Hat
Rating: **3.80**
Popular accessory choice



Skirt
Rating: **3.78**
Strong clothing category performer

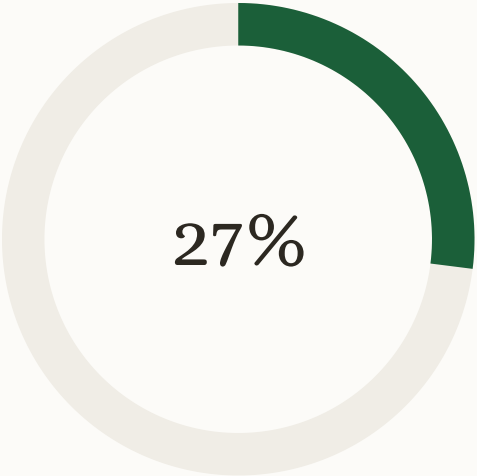
Shipping Type Analysis



Customers who choose Express shipping spend an average of **\$60.48** compared to **\$58.46** for Standard shipping—a difference of \$2.02 per transaction.

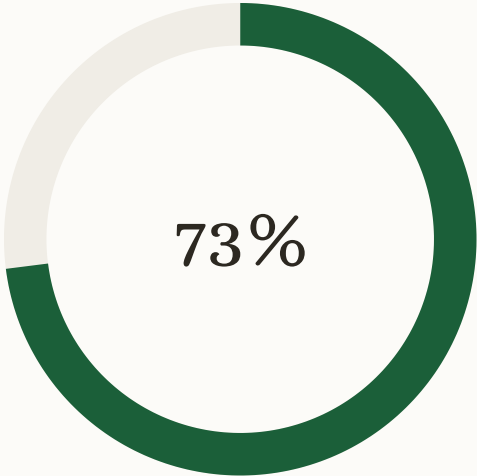
This suggests that express shipping customers may be more valuable overall, potentially representing a segment willing to pay premium prices for convenience and speed.

Subscription Status Deep Dive



Subscription Rate

1,053 out of 3,900 customers are subscribers



Non-Subscribers

2,847 customers without active subscriptions

Average Spend

Subscribers: \$59.49

Non-Subscribers: \$59.87

Surprisingly, non-subscribers spend slightly more per transaction, suggesting subscription value may lie in frequency rather than transaction size.

Total Revenue

Subscribers: \$62,645

Non-Subscribers: \$170,436

Non-subscribers generate significantly more total revenue due to their larger population size.

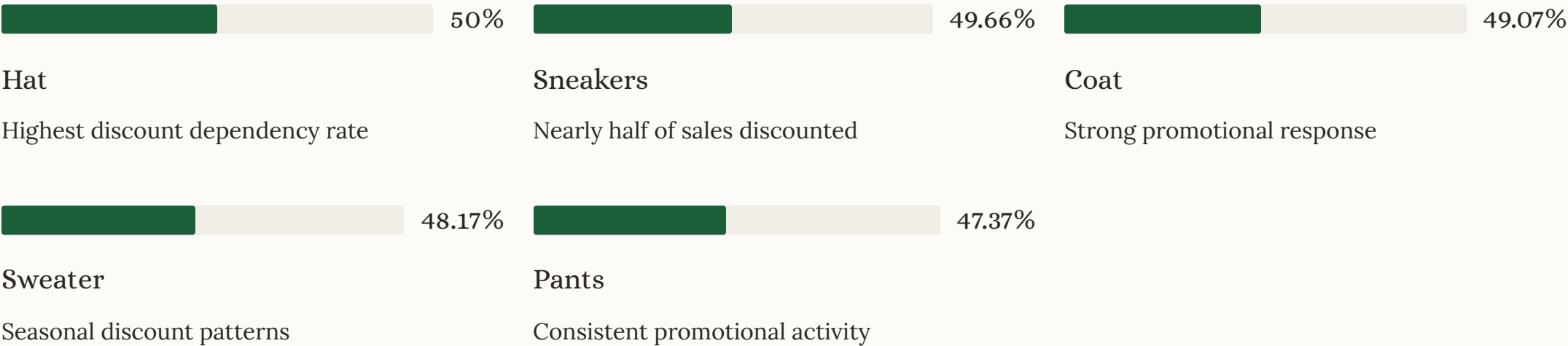
Strategic Opportunity

With 73% of customers not subscribed, there's **massive growth potential** in converting even a small percentage of non-subscribers.

A 10% conversion would add 285 subscribers to the base.

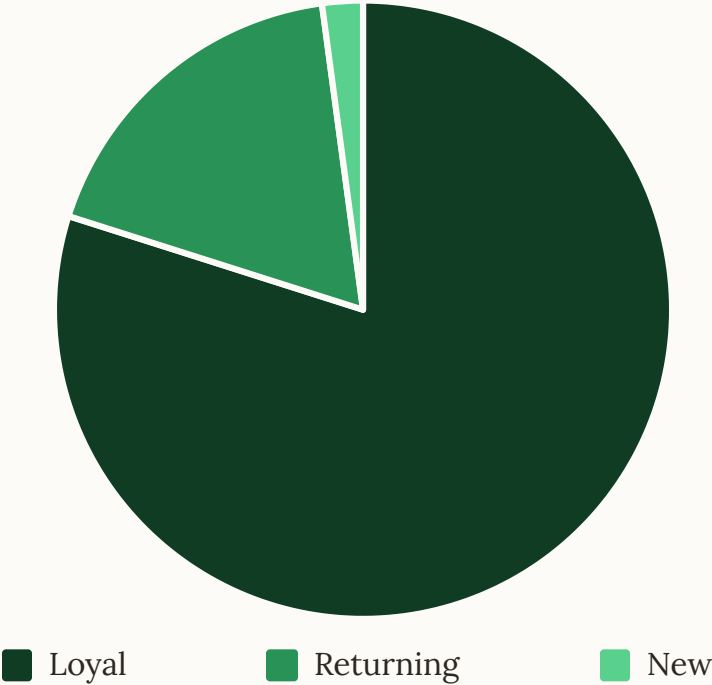
Discount-Dependent Products

Analysis reveals which products rely most heavily on discounts to drive sales. Understanding these patterns helps optimize promotional strategies and maintain healthy profit margins.



❏ These products show discount rates near 50%, indicating that approximately half of all purchases occur with promotional pricing. This suggests strong price sensitivity in these categories and opportunities to test alternative value propositions beyond discounting.

Customer Segmentation Analysis



Segment Breakdown

Loyal customers dominate at 79.9% of the customer base, demonstrating strong retention and repeat purchase behavior.

Returning customers represent 18.0%, showing engagement but not yet reaching loyal status.

New customers account for only 2.1%, suggesting either strong retention or limited new customer acquisition.

Repeat Buyers & Subscription Correlation

Among customers with more than 5 purchases, we examined subscription patterns to understand loyalty program effectiveness:

Non-Subscribed Repeat Buyers

2,518 customers

High-frequency purchasers without subscriptions represent a prime conversion opportunity

Subscribed Repeat Buyers

958 customers

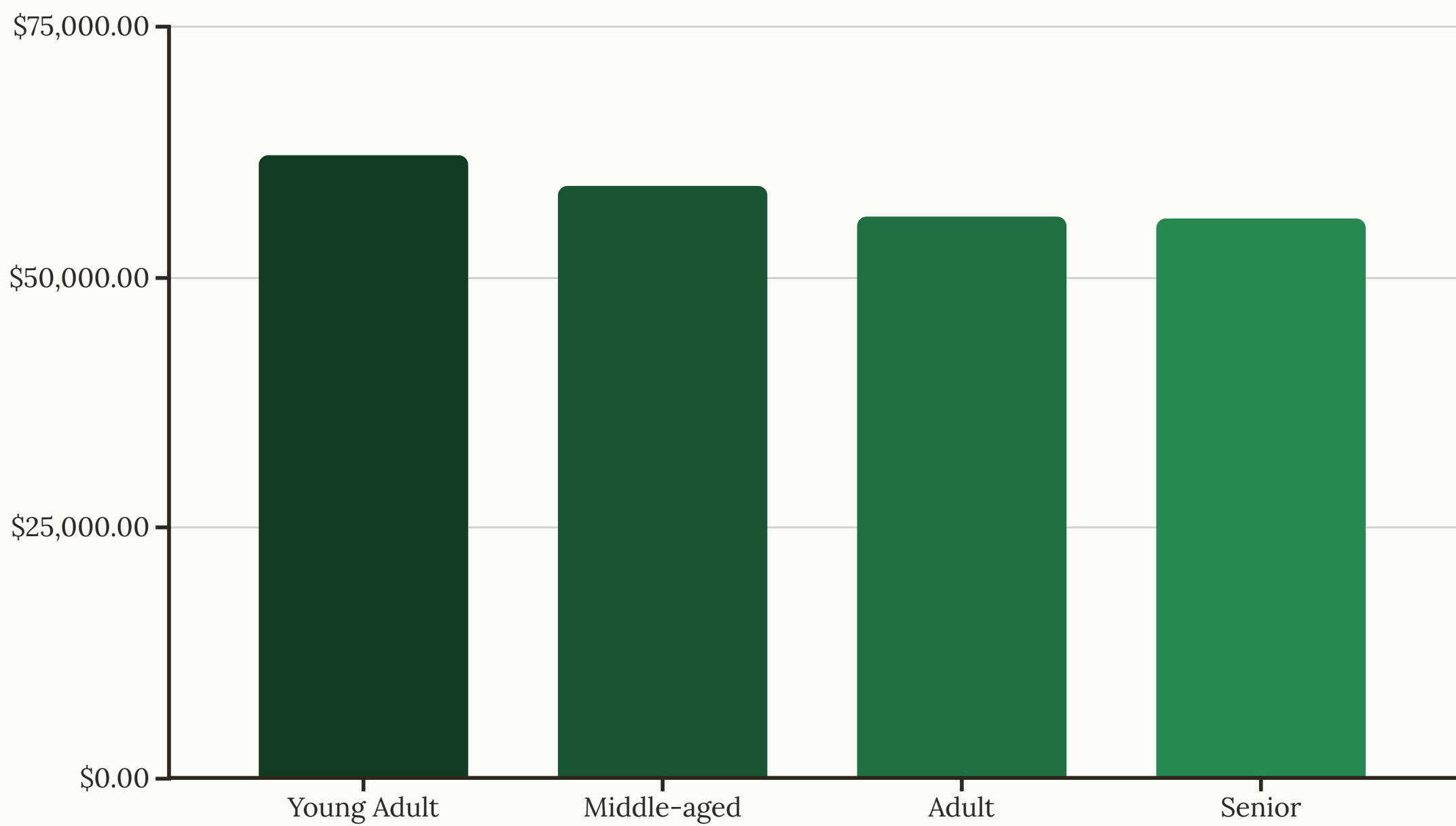
Loyal customers who have formalized their relationship through subscription

This data reveals that **72% of repeat buyers are not subscribed**, indicating significant untapped potential for subscription program growth among the most engaged customer segment.

Top Products by Category

1	2
<p>Accessories</p> <p>Top 3: Jewelry (171 orders), Sunglasses (161 orders), Belt (161 orders)</p> <p>Accessories show strong, balanced demand across multiple product types with jewelry leading the category.</p>	<p>Clothing</p> <p>Top 3: Blouse (171 orders), Pants (171 orders), Shirt (169 orders)</p> <p>Core clothing items demonstrate consistent popularity with minimal variation in order volume across top products.</p>
3	4
<p>Footwear</p> <p>Top 3: Sandals (160 orders), Shoes (150 orders), Sneakers (145 orders)</p> <p>Footwear category shows clear preference hierarchy with sandals leading, possibly influenced by seasonal factors.</p>	<p>Outerwear</p> <p>Top 3: Jacket (163 orders), Coat (161 orders)</p> <p>Outerwear maintains strong demand with jackets slightly outperforming coats in purchase frequency.</p>

Revenue by Age Group



Young Adults generate the highest revenue at **\$62,143**, followed closely by Middle-aged customers. Revenue distribution is relatively balanced across age groups, with only an 11% difference between highest and lowest performing segments.

Strategic Business Recommendations



Boost Subscriptions

With 72% of repeat buyers not subscribed and only 27% overall subscription rate, promote exclusive benefits for subscribers. Target the 2,518 non-subscribed repeat buyers with personalized subscription offers emphasizing convenience and exclusive perks.



Customer Loyalty Programs

Reward repeat buyers to accelerate their transition into the "Loyal" segment. With 79.9% already loyal, focus on converting the 18% returning customers through tiered rewards and recognition programs.



Review Discount Policy

Balance sales boosts with margin control, especially for products with 50% discount rates (Hat, Sneakers, Coat). Test alternative value propositions and consider strategic pricing adjustments to reduce discount dependency while maintaining volume.



Product Positioning

Highlight top-rated products (Gloves 3.86, Sandals 3.84, Boots 3.82) and best-sellers (Jewelry, Blouse, Pants with 171 orders each) in marketing campaigns. Leverage social proof and customer reviews to drive conversion.



Targeted Marketing

Focus efforts on high-revenue segments: Young Adults (\$62,143 revenue), male customers (68% of revenue), and express-shipping users (\$60.48 average spend). Develop persona-specific campaigns that resonate with these valuable customer segments.