Customer Shopping Behavior Analysis

Analyzing 3,900 purchases to uncover insights into spending patterns, customer segments, and product preferences that drive strategic business decisions.

Dataset Overview

3,900
Total Purchases
Transactions analyzed across all categories

18
Data Columns
Comprehensive customer and purchase attributes

50
Locations
Geographic diversity in customer base

25
Product Types
Items across multiple categories

Key features include customer demographics (age, gender, location, subscription status), purchase details (item, category, amount, season, size, color), and shopping behavior (discounts, promo codes, previous purchases, review ratings, shipping type).

Data Preparation & Cleaning

01
Data Loading

Imported dataset using pandas for analysis

02
Initial Exploration

Used df.info() and .describe() for structure review

03
Missing Data Handling

Imputed 37 missing review ratings using category medians

04
Column Standardization

Renamed columns to snake_case format

Feature Engineering

- Created age_group column by binning customer ages
- Generated purchase_frequency_days from purchase data
- Verified discount_applied and promo_code_used redundancy
- Connected to PostgreSQL for SQL analysis

Revenue Analysis by Gender

Male Customers

$157,890 total revenue

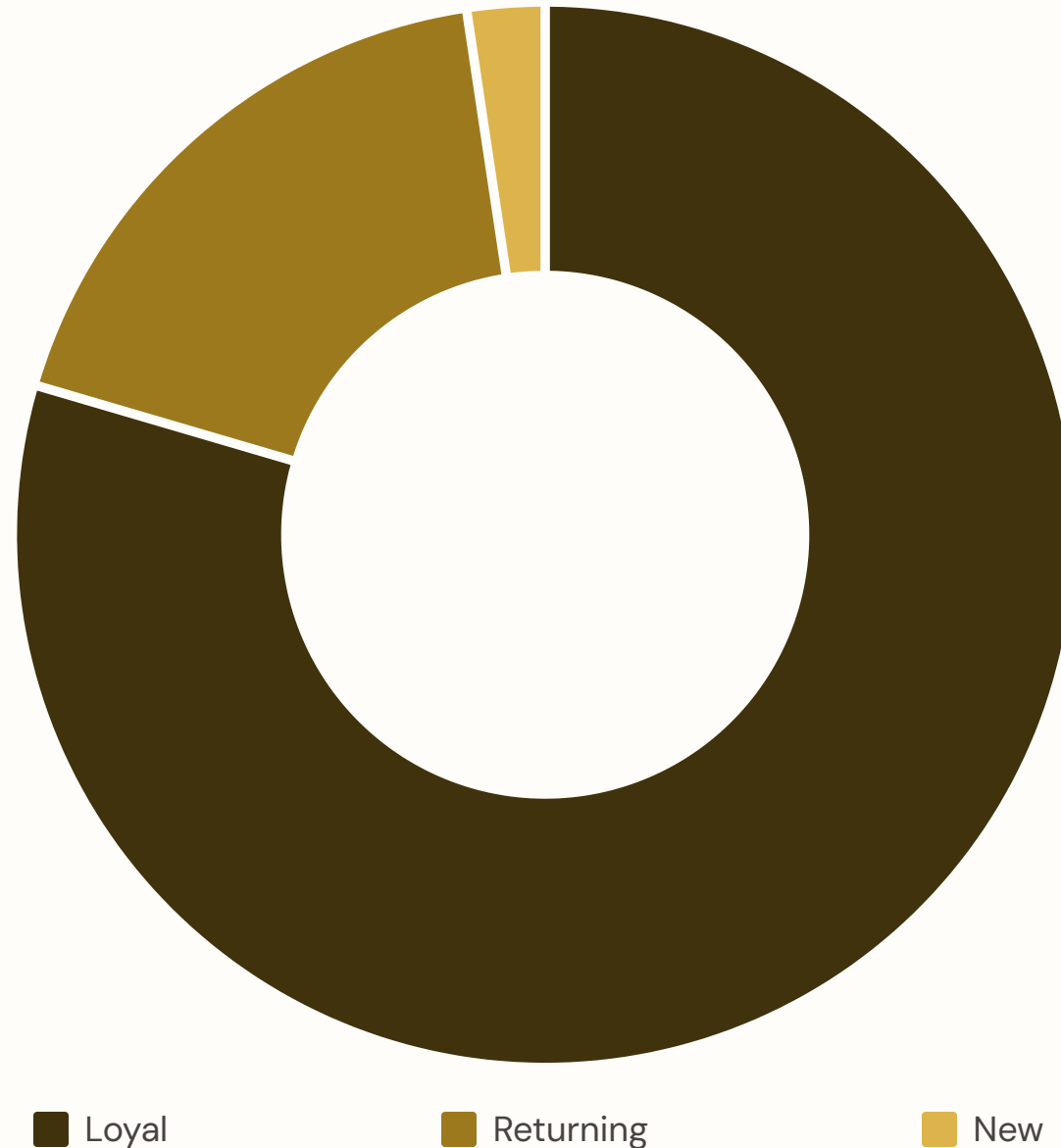Representing 68% of total sales

Female Customers

$75,191 total revenue

Representing 32% of total sales

Male customers generate more than double the revenue of female customers, indicating a significant opportunity for targeted marketing campaigns to balance gender representation.

Customer Segmentation Insights

■ Loyal   ■ Returning   ■ New

Loyal customers dominate at 80% of the customer base, while new customers represent only 2%. Focus on converting returning customers (18%) into loyal segments through targeted retention programs.

Subscription Status Comparison

Subscribers

1,053 customers

Average spend: **$59.49**

Total revenue: **$62,645**

Non-Subscribers

2,847 customers

Average spend: **$59.87**

Total revenue: **$170,436**

Despite similar average spending, non–subscribers generate significantly more total revenue due to their larger population. Subscription conversion represents a major growth opportunity.

Top Products & Categories

Clothing Category

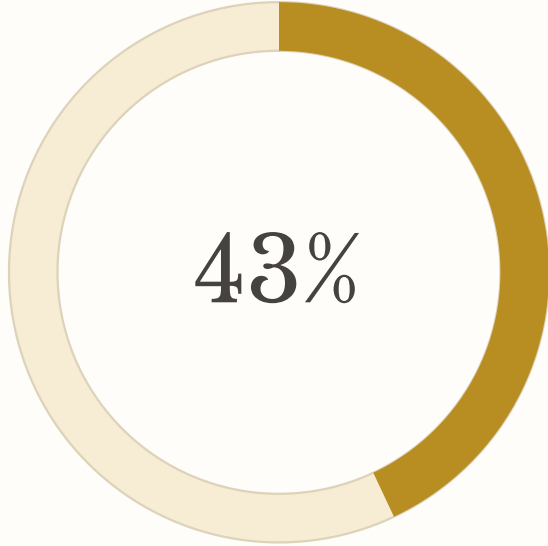**Blouse** leads with 171 orders, followed by Pants (171) and Shirt (169)

Accessories Category

Jewelry tops at 171 orders, with Sunglasses (161) and Belt (161) close behind

Footwear Category

Sandals lead with 160 orders, followed by Shoes (150) and Sneakers (145)
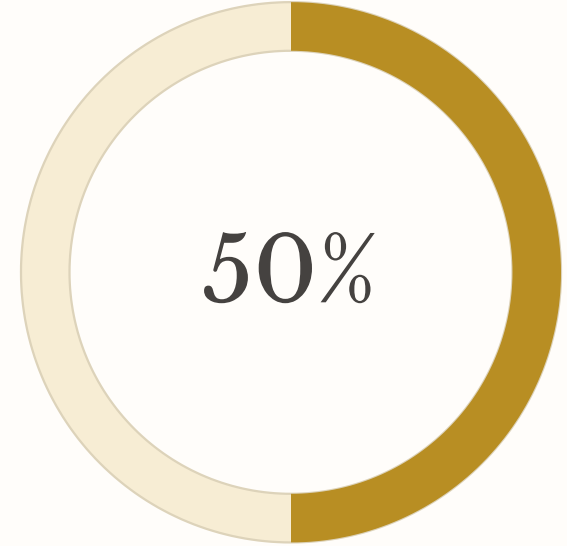
Discount Strategy Analysis

43%

Overall Discount Rate

1,677 of 3,900 purchases used discounts

839

High-Value Discount Users

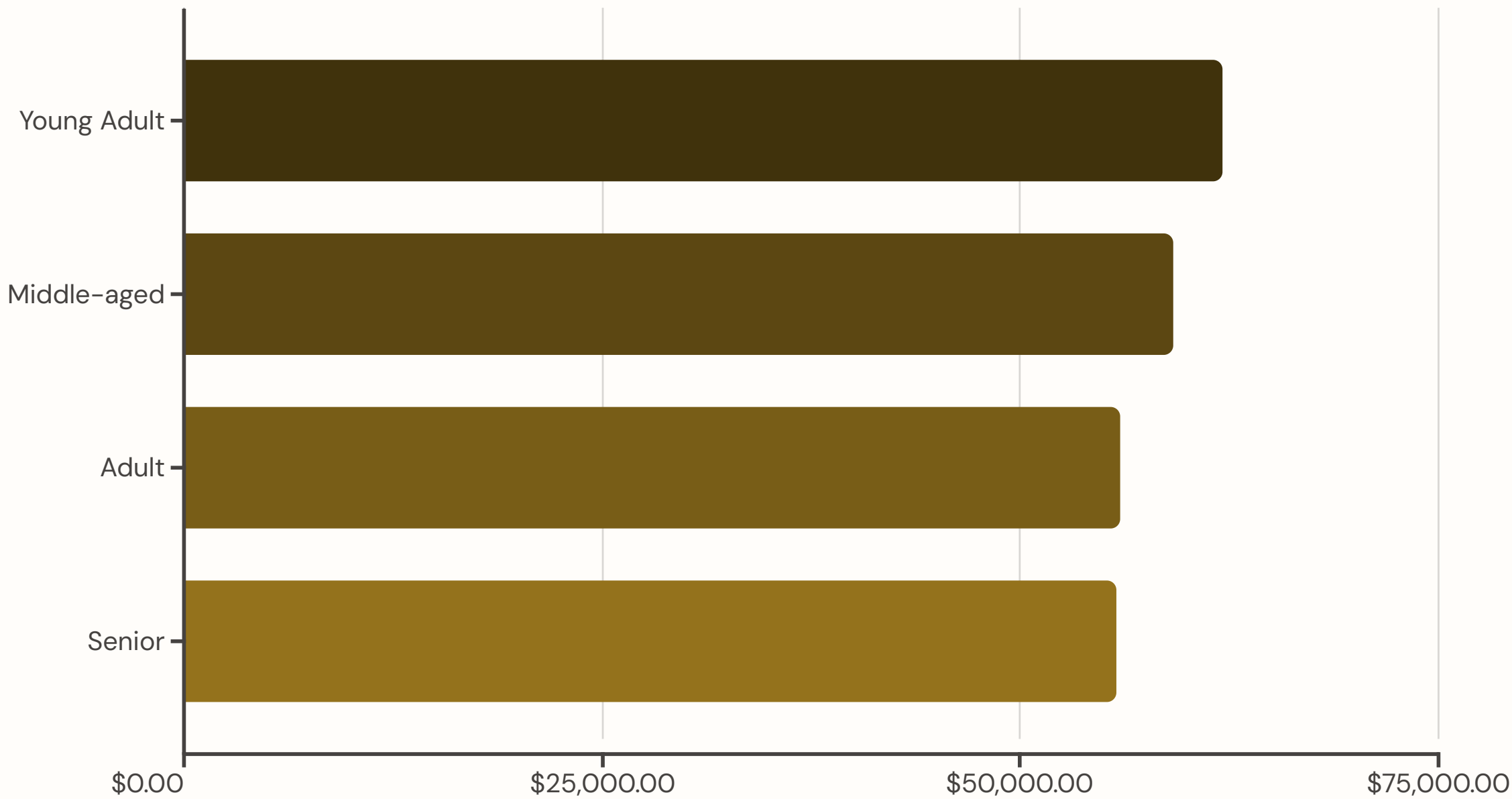Customers who used discounts but spent above average

50%

Hat Discount Rate

Highest discount dependency among products

Top discount–dependent products: Hat (50%), Sneakers (49.66%), Coat (49.07%), Sweater (48.17%), and Pants (47.37%). Balance promotional strategies to maintain healthy margins.

Made with GAMMA

Revenue by Age Group

Age Group	Revenue
Young Adult	~$62,143
Middle-aged	~$59,197
Adult	~$56,000
Senior	~$57,000

(x-axis: $0.00, $25,000.00, $50,000.00, $75,000.00)

Young adults generate the highest revenue at $62,143, followed closely by middle-aged customers at $59,197. Revenue distribution is relatively balanced across age groups.

Made with GAMMA

Strategic Recommendations

Boost Subscriptions

Promote exclusive benefits to convert non-subscribers and increase recurring revenue

Loyalty Programs

Reward repeat buyers to accelerate movement into the loyal customer segment

Review Discount Policy

Balance promotional sales boosts with margin control for sustainable growth

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