



# Customer Shopping Behaviour Analysis

Uncovering insights from 3,900 purchases to guide strategic business decisions through data-driven customer segmentation and spending patterns.

PROJECT SCOPE

# Dataset Overview

**3.9K**

## Total Purchases

Transactions analyzed across all categories

**18**

## Data Points

Comprehensive attributes per transaction

**37**

## Missing Values

Only in review rating column

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

# Data Preparation Process

01

## Data Loading & Exploration

Imported dataset using pandas, checked structure with `df.info()` and summary statistics

03

## Column Standardization

Renamed columns to snake\_case for better code readability

05

## Consistency Check

Verified `discount_applied` and `promo_code_used` alignment, dropped redundant column

02

## Data Cleaning

Handled 37 null values by imputing median review ratings per product category

04

## Feature Engineering

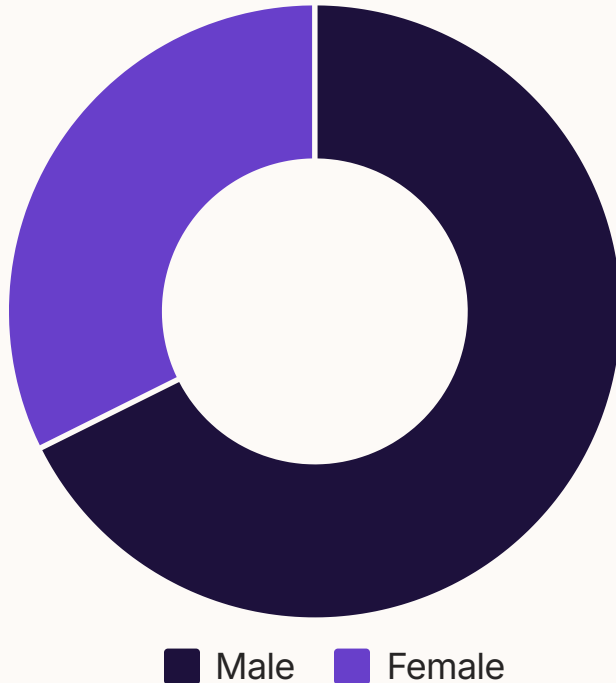
Created `age_group` and `purchase_frequency_days` columns for deeper analysis

06

## Database Integration

Connected to MySQL and loaded clean dataframe for SQL analysis

# Revenue Insights by Gender



## Key Finding

Male customers generate **\$157,890** in revenue compared to **\$75,191** from female customers—representing a 2.1x difference in spending.

This significant gap suggests opportunities for targeted marketing strategies.

# Product Performance Analysis

1

## Gloves

Average rating: **3.86**

2

## Sandals

Average rating: **3.84**

3

## Boots

Average rating: **3.82**

4

## Hat

Average rating: **3.80**

5

## Handbag

Average rating: 3.78

# Shipping & Spending Patterns

## Shipping Type Comparison

Express shipping customers spend an average of **\$60.48** per purchase, while standard shipping averages **\$58.46**.

The \$2.02 difference suggests premium shipping correlates with slightly higher purchase values.



---

## High-Value Discount Users

Identified customers who used discounts but still spent above average—10 customers spent between \$62-\$97 despite applying discounts, indicating price-conscious yet high-value shoppers.

# Loyalty Distribution

## Loyal Customers

**2,721 customers**

Represent 70% of customer base with consistent purchase history

## Returning Customers

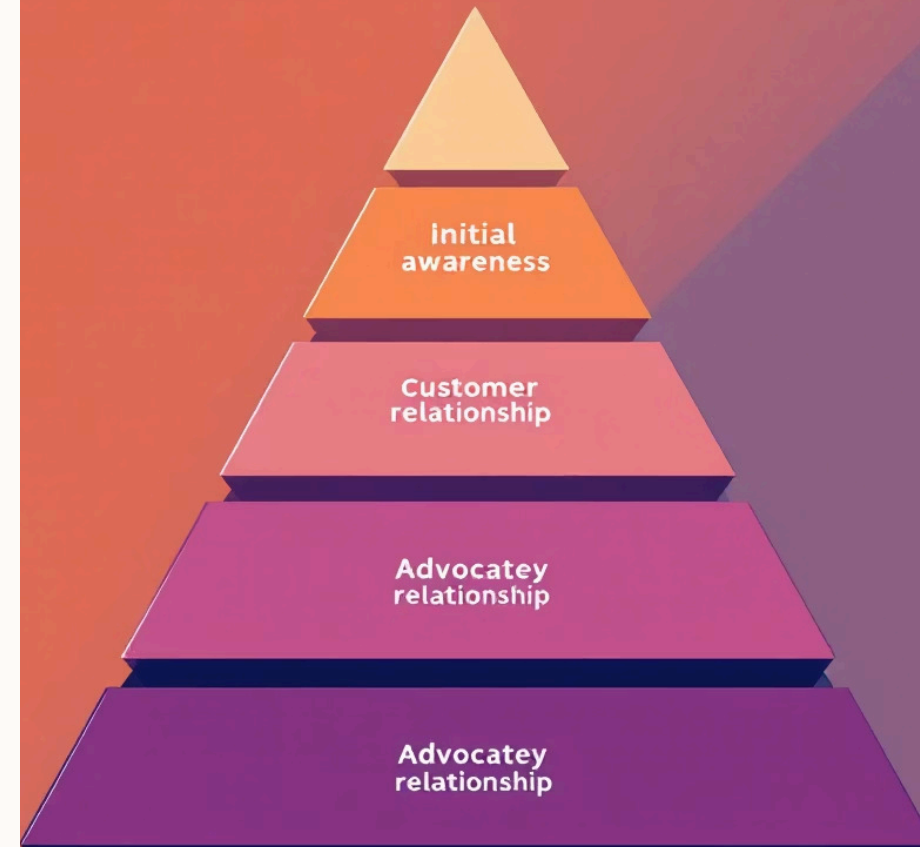
**1,096 customers**

28% showing repeat purchase behavior

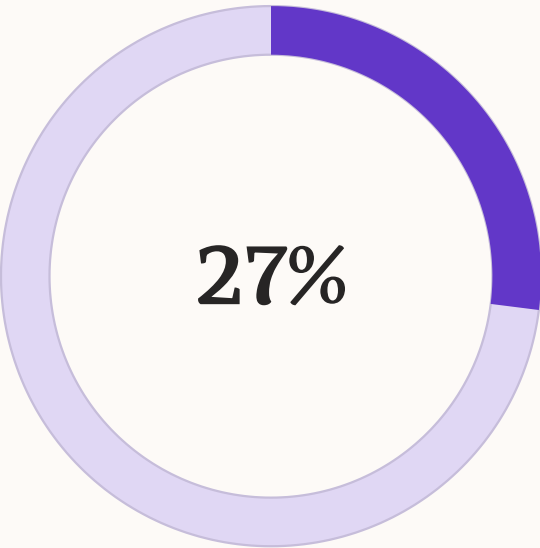
## New Customers

**83 customers**

2% representing growth opportunity

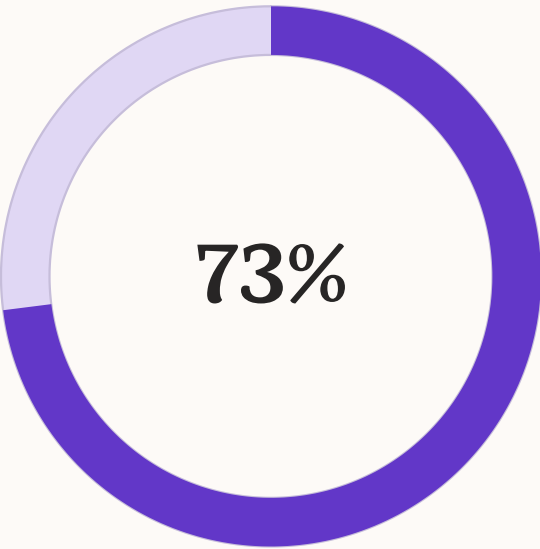


# Subscription Behavior Insights



Subscription Rate

1,050 subscribed customers



Non-Subscribers

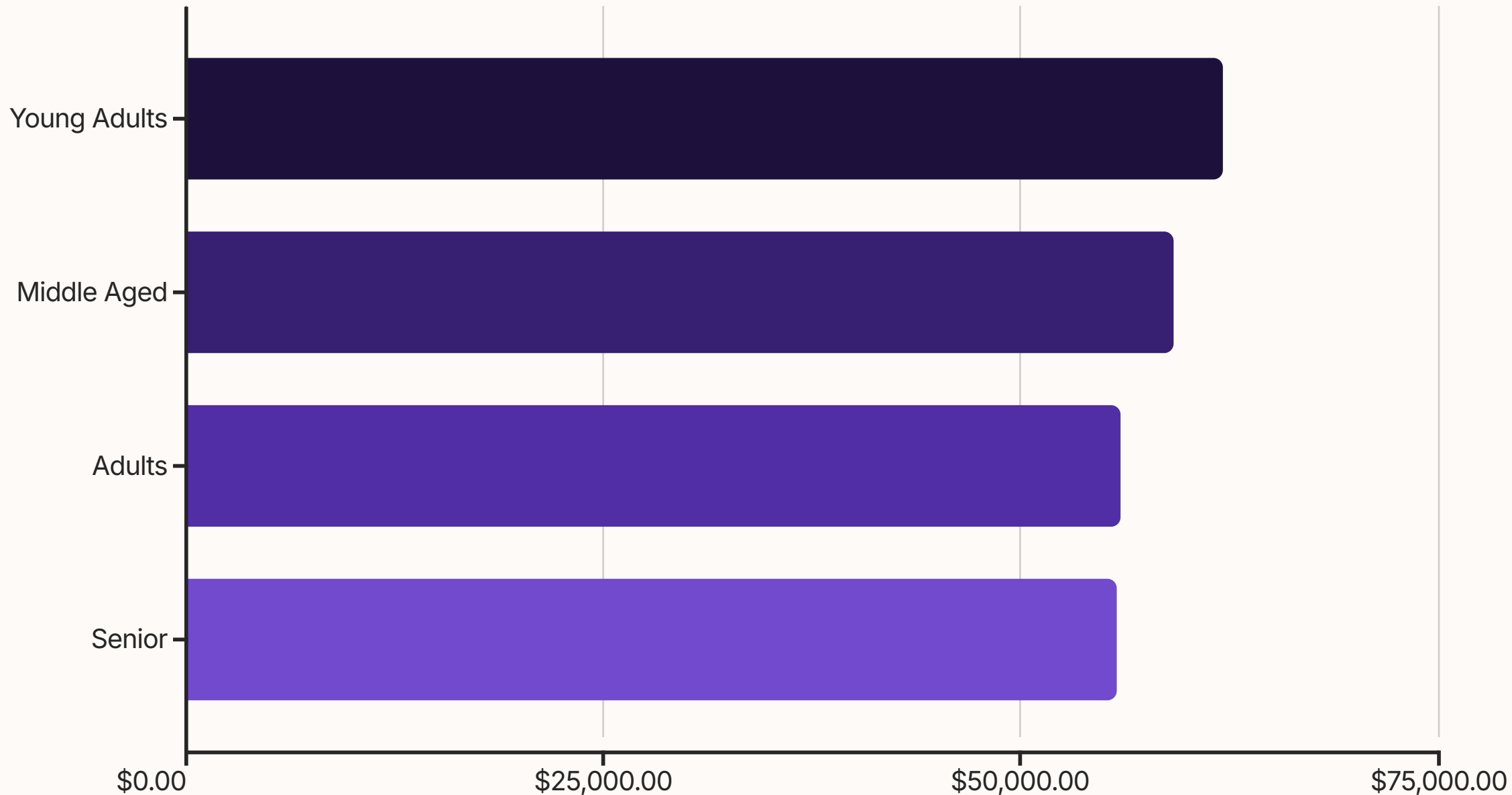
2,850 customers

## Repeat Buyer Analysis

Among customers with **5+ previous purchases**, only **958 are subscribers** while **2,518 are not**.

This reveals a significant opportunity: repeat buyers don't necessarily convert to subscriptions, suggesting room for improved subscription incentives.

# Revenue by Age Group



Young adults lead revenue generation at **\$62,143**, followed closely by middle-aged customers at **\$59,197**. Revenue distribution is relatively balanced across age groups, with only a \$6,380 difference between highest and lowest.



# Key Takeaways & Strategic Recommendations



## Gender-Focused Marketing

Male customers drive 68% of revenue—optimize campaigns while exploring female customer growth opportunities



## Product Excellence

Top-rated items (Gloves, Sandals, Boots) maintain 3.8+ ratings—leverage these for promotional campaigns



## Subscription Conversion

Only 27% subscribe despite high repeat purchases—enhance subscription value propositions



## Loyalty Optimization

70% loyal customer base provides stable foundation—focus on converting returning customers to loyal status