

# Customer Shopping Behavior Analysis

Uncovering actionable insights through data cleaning, SQL querying, and interactive visualization to inform strategic business decisions.





## DATASET OVERVIEW

# Foundation of Analysis

3,900

Total Rows

Customer transactions  
analyzed

18

Data Columns

Comprehensive feature  
set

37

Missing Values

In Review Rating  
column

Dataset includes customer demographics, purchase details, and shopping behavior metrics. Missing values were imputed using median ratings by product category.

# Exploratory Data Analysis

01

## Data Loading & Exploration

Loaded dataset using Pandas and checked structure and summary statistics.

02

## Missing Data Handling

Imputed 37 missing Review Rating values with median ratings by category.

03

## Data Consistency

Verified redundancy between Discount Applied and Promo Code Used columns.

04

## Feature Engineering

Created age\_grouped and mapped\_frequency\_of\_purchase columns for deeper analysis.

05

## Database Integration

Connected Python to MySQL and loaded cleaned data for SQL-based querying.

# Key Findings at a Glance

## Revenue Drivers

Male customers and Young-Adults lead revenue generation.

## Quality Concerns

High-revenue products show low ratings, creating retention risks.

## Loyalty Paradox

Strong customer loyalty (80.5%) but low subscription adoption among high-frequency buyers.

## Shipping Impact

Shipping experience meaningfully affects customer satisfaction.

## Category Leaders

Clothing and Accessories dominate revenue streams.

## Strategic Discounts

Discounts don't reduce order value—use them strategically.



# Key Performance Indicators



Total Revenue

\$233,081

Average Order Value

\$59.76

Male Revenue Share

67.7%

Loyal Customers

80.5%

Subscription Revenue

26.9%

Top Category

Clothing (44.7%)

# Customer Demographics & Value

## Gender Revenue Split

Male customers generate **\$157,890 (67.7%)** of total revenue, while female customers contribute **\$75,191 (32.3%)**, indicating strong male-driven revenue skew.

## Age Group Performance

**Young-Adults** lead with \$62,143, followed by Mid-Age (\$59,197), Adults (\$55,978), and Seniors (\$55,763). Revenue is balanced with slight advantage among younger customers.

## The Subscription Gap

Non-subscribers contribute **73.1% of revenue** despite identical AOV. Among high-frequency buyers, **72.4% are non-subscribers**—a critical opportunity.



# Product & Category Performance

## Clothing Dominates

**\$104,264** revenue leader, featuring top sellers: Blouses, Pants, and Shirts.

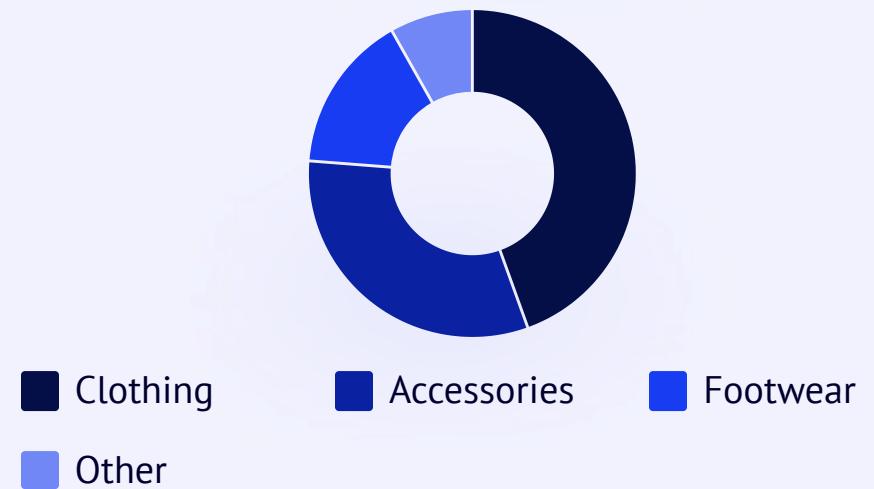
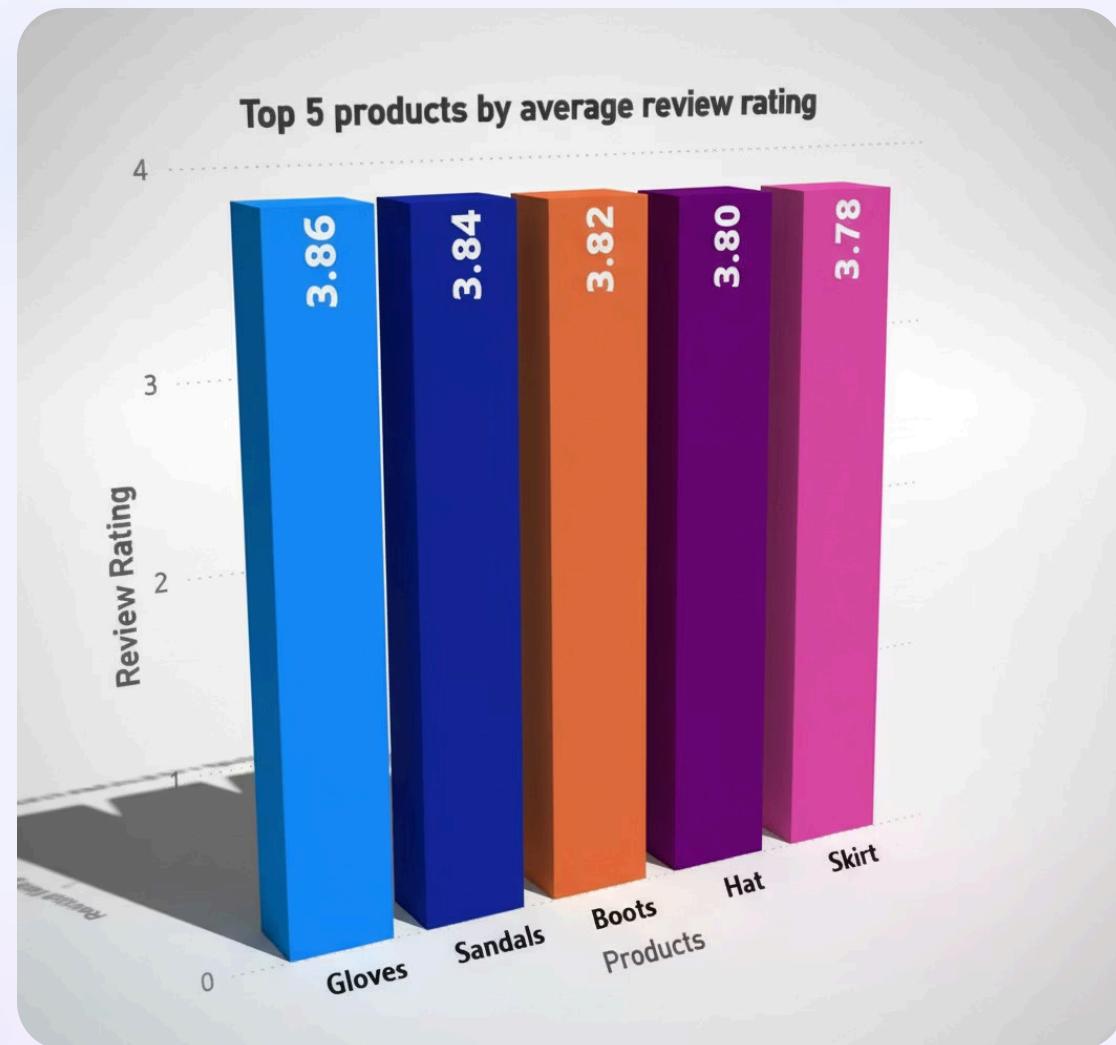
## Accessories Strong

**\$74,200** revenue with Jewelry and Sunglasses leading the category.

## Footwear Solid

**\$36,093** revenue with Sandals as category leader and high ratings.

- ☐ **Risk Alert:** Coats and Skirts show high purchase values (\$97 and \$81) but low ratings (2.6–2.8), indicating price-expectation mismatch.



🖨️ REVENUE BREAKDOWN

# Category Revenue Distribution

The donut chart illustrates the significant contribution of Clothing and Accessories, which together account for over 75% of total revenue. Footwear holds a substantial share, while other categories make up the remaining portion. This distribution highlights the primary focus areas for product development and marketing efforts.

# Location & Shipping Insights

## Geographic Performance

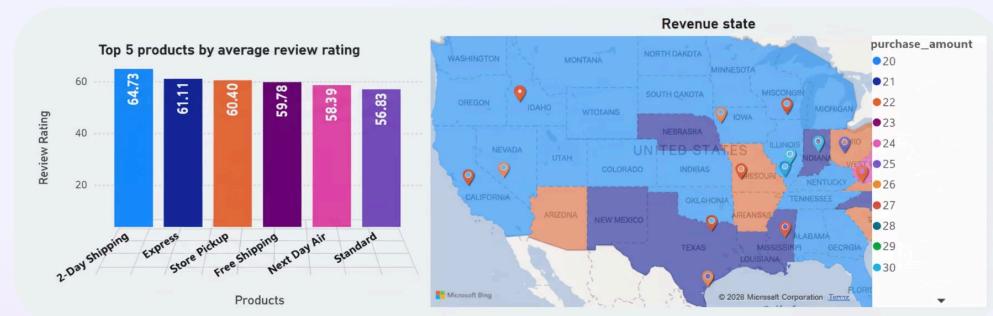
Montana, Illinois, California, and Idaho each generate approximately **\$5.6k** in revenue. Alaska records highest AOV at **\$67.60**.

## Shipping Preferences

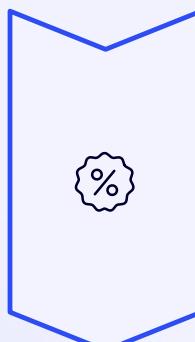
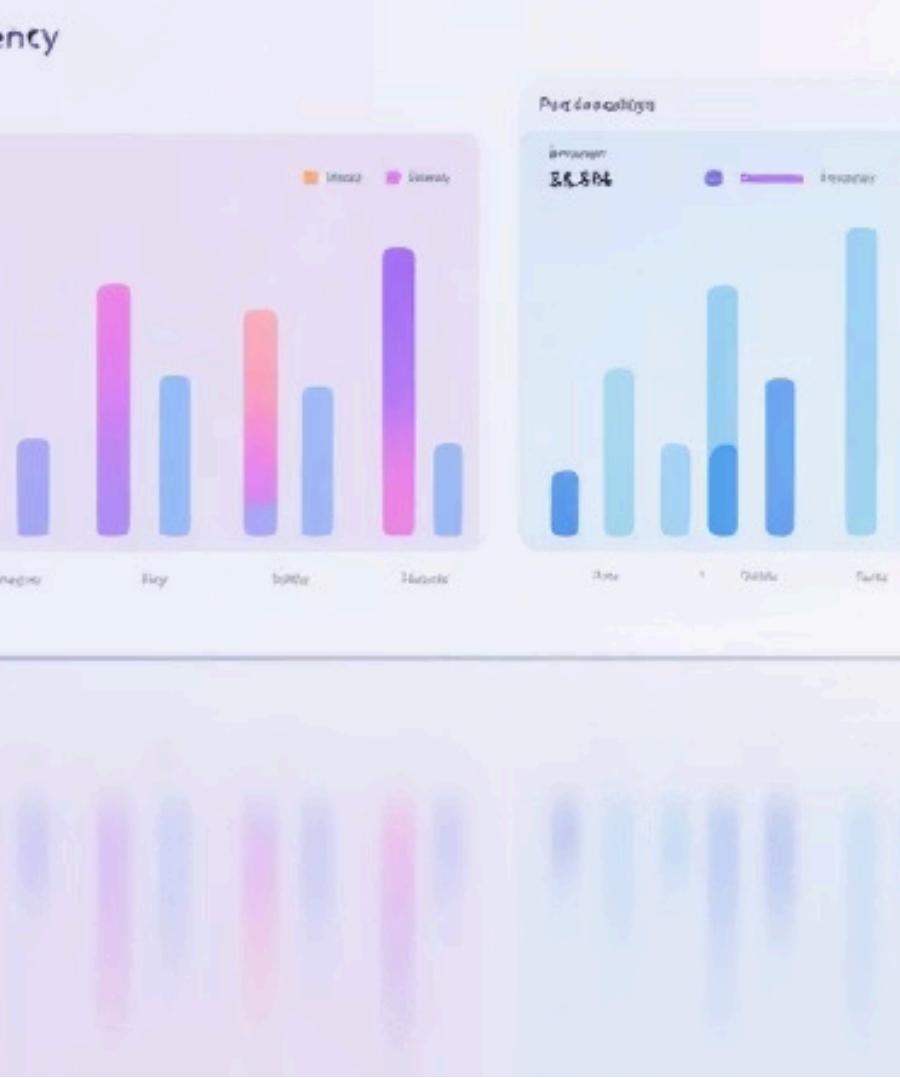
Preferences vary by location: **Store Pickup** dominates Illinois while **Free Shipping** is favored in Montana.

## Satisfaction Winner

**Standard Shipping** achieves highest customer rating (**3.82**), outperforming faster fulfillment methods.



# Discounts, Payments & Seasonality



## Smart Discounting

Discounted orders exceed overall AOV (\$59.76). Highest usage in Rhode Island (39.7%), Vermont (38.8%), Virginia (37.7%).



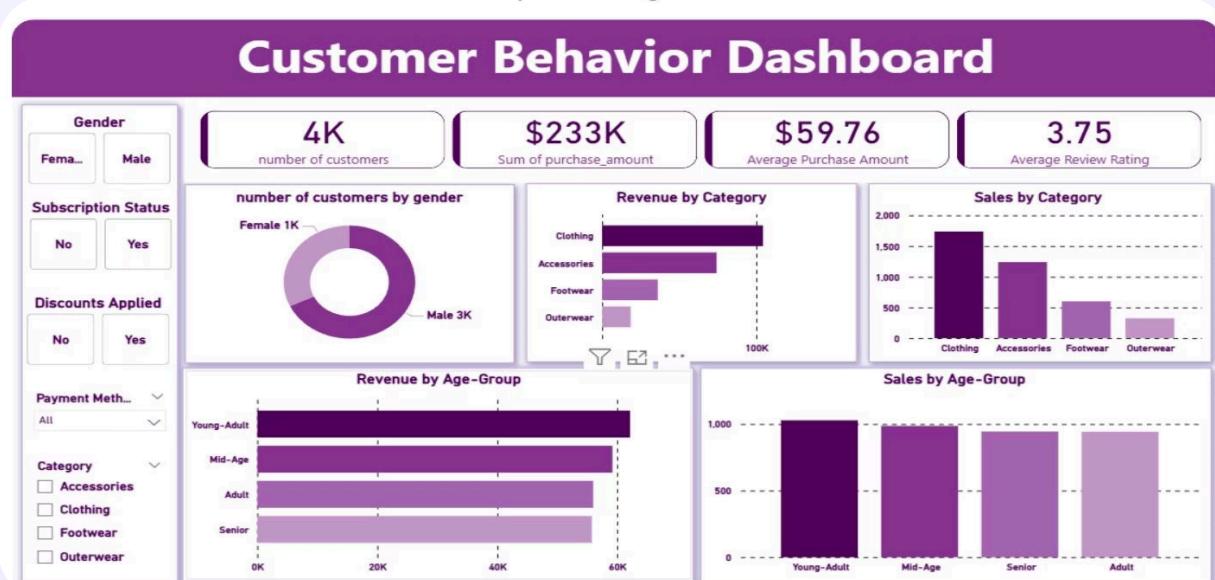
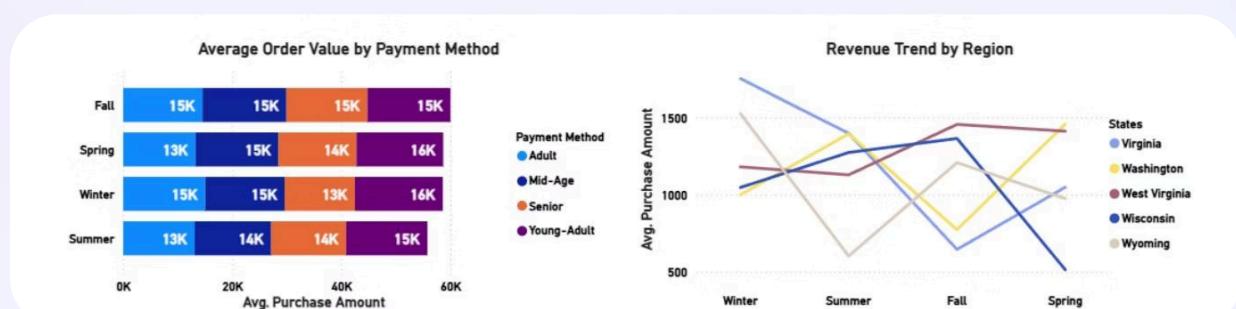
## Payment Patterns

Debit card transactions show highest AOV (\$60.92). Quarterly purchasers most promotion-responsive (44.2%).



## Seasonal Trends

Revenue peaks in Fall for Colorado, Spring for Ohio. Winter favors Young-Adults/Adults; Fall favors Mid-Age/Seniors.



# The Path Forward

## Convert High-Frequency Buyers

- 1 Redesign subscription offerings with value-led incentives targeting the 72.4% of repeat buyers who aren't subscribers.

## Address Quality Perception

- 2 Improve quality and value perception for high-revenue, low-rated products (Coats, Skirts) to reduce churn risk.

## Optimize Discount Strategy

- 3 Target discounts toward mid-frequency shoppers (quarterly purchasers) to maximize incremental revenue without eroding margins.

## Promote Standard Shipping

- 4 Make Standard Shipping the default option to enhance customer satisfaction (3.82 rating) and reduce fulfillment costs.

## Align with Seasonality

- 5 Synchronize marketing campaigns and inventory planning with age-based and regional seasonal demand patterns.

## Leverage Top Performers

- 6 Cross-sell and bundle high-repeat products (Dresses, Scarves, Boots) to strengthen loyalty programs and boost revenue.