

Customer Shopping Behavior Analysis

A comprehensive analysis of 3,900 customer transactions using Python, SQL, and Power BI to uncover behavioral patterns and drive data-informed business decisions.



Project Overview & Objectives

This multi-phase analysis combines technical expertise with business intelligence to transform raw transaction data into actionable strategic insights.



Project Objective

Analyze customer shopping behavior across demographics, purchases, and engagement to support data-driven decision making and revenue optimization.



Dataset Scale

3,900 rows representing real customer transactions with 18 distinct data fields spanning demographics, purchase details, and behavioral attributes.



Technical Stack

Python for data preparation, SQL for structured analysis, and Power BI for interactive visualization and dashboard creation.

Data Preparation & Python Engineering



Data Loading

Initial exploration and quality assessment of raw customer data



Data Cleaning

Median imputation for 37 missing review ratings by category



Standardization

Consistent column naming and data type alignment



Feature Engineering

Created age groups and purchase frequency calculations



SQL Loading

Final dataset loaded into PostgreSQL for analysis

Feature engineering and data cleaning ensured consistency and accuracy for downstream SQL queries and Power BI visualization. The cleaned dataset maintains integrity while adding valuable derived metrics for deeper analysis.

SQL Analysis & Business Questions

Revenue Segmentation

Break down revenue by gender, age groups, subscription status, and shipping preferences

Product Performance

Identify top-rated products, discount-dependent items, and category-specific trends

Behavioral Patterns

Analyze purchase frequency, repeat buying behavior, and customer lifetime value

Comparative Analysis

Compare subscriber vs. non-subscriber spending and standard vs. express shipping behavior



Key Revenue Insight: Gender Split

\$157,890

Male Customers

2.1× higher revenue than female customers

\$75,191

Female Customers

\$82,699 difference in total revenue



Interpretation

- Product mix favors male demographics
- Potential pricing or category differences
- Marketing opportunity to rebalance segments

Customer Segmentation & Loyalty

1

Loyal Customers

3,116 customers represent strong retention and brand loyalty. This segment should be prioritized for reward programs and exclusive offers.

2

Returning Buyers

701 customers show intermediate engagement. Target with re-engagement campaigns and personalized recommendations.

3

New Customers

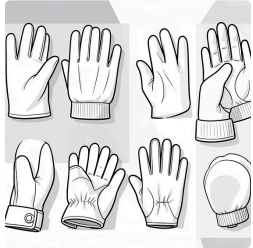
83 customers represent 2% of total database. Focus on conversion optimization and first-purchase experience.



Insight: The business has a **strong loyal customer base**, representing 79% of the customer database. Retention strategies should focus on rewarding loyalty and converting returning buyers.

Product & Rating Insights

Top-rated products provide opportunities for marketing campaigns and inventory optimization. These items receive consistently positive customer feedback.



Gloves

3.86 average rating

Top of category, often purchased with boots and jackets



Sandals

3.84 average rating

Seasonal high performer, excellent summer revenue driver



Boots

3.82 average rating

Premium pricing, pairs well with glove recommendations



Hat

3.80 average rating

High purchase frequency, low individual price point

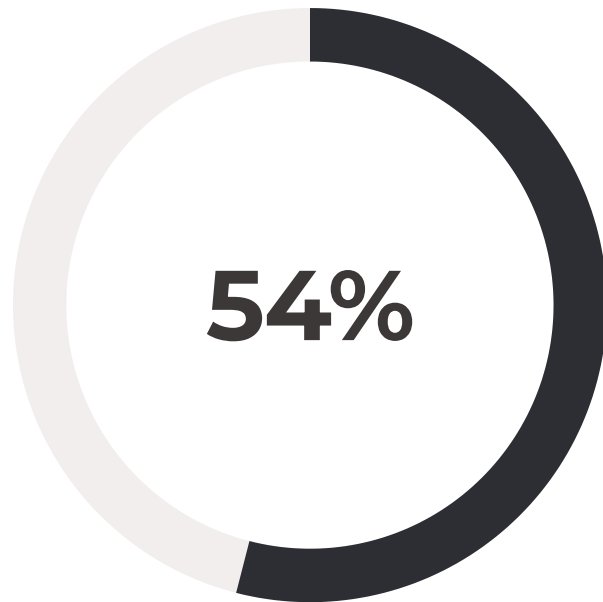


Skirt

3.78 average rating

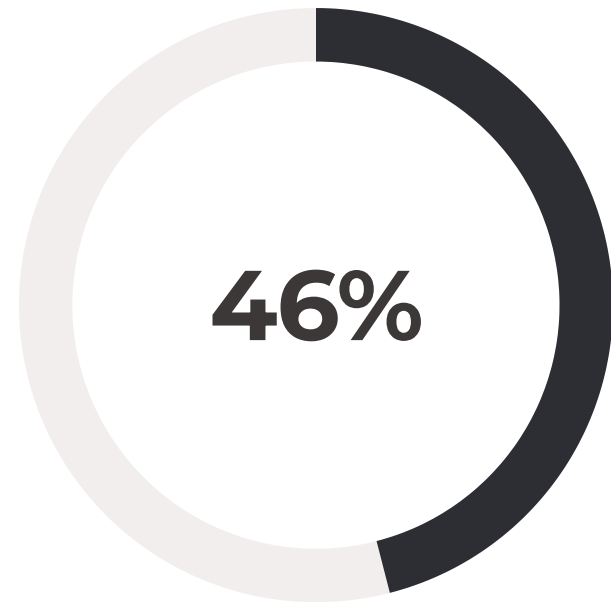
Mid-range category with steady year-round demand

Shipping Behavior Analysis



Standard Shipping

Average spend: **\$58.46**



Express Shipping

Average spend: **\$60.48**



Key Insight

Express shipping users spend **\$2.02 more per purchase** (3.5% increase). While the difference is modest, this segment may represent higher-value customers willing to pay for convenience.

Recommendation: Test express shipping upsell opportunities at checkout for high-margin products.

Subscription Analysis

Subscribers

- 652 customers (17% of base)
- \$59.49 average spend
- \$62,645 total revenue
- \$141.78 average value

Non-Subscribers

- 3,248 customers (83% of base)
- \$59.87 average spend
- \$170,436 total revenue
- \$87.30 average value

Market Opportunity

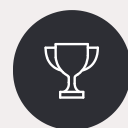
Non-subscribers generate **2.7× more revenue** due to higher volume, but subscribers have nearly double the average customer value. This represents significant growth potential through subscription program optimization.

Power BI Dashboard & Business Recommendations



Interactive Dashboard

Visualize revenue, sales by category, age group performance, shipping filters, and gender distribution



Performance KPIs

\$59.76 avg purchase • 3.75 avg rating • 3.9K customers • 18 data fields

Strategic Recommendations

- 1

Boost Subscription Conversion

Promote exclusive benefits to increase subscriber count from current 17% baseline
- 2

Strengthen Loyalty Programs

Reward the 3,116 loyal customers to reinforce retention and increase lifetime value
- 3

Optimize Discount Strategy

Balance discount incentives with margin protection using category-level analysis
- 4

Highlight Top Products

Leverage high-rated gloves, sandals, boots, hats, and skirts in marketing campaigns
- 5

Target High-Value Segments

Focus marketing spend on express-shipping customers and high-revenue age groups