Abstract

This report presents a comprehensive analysis of customer shopping behavior based on 3,900 transactions. The dataset reveals critical insights into spending patterns, customer segments, product preferences, and subscription trends, and provides actionable recommendations to drive revenue growth and customer retention.

Executive Summary

This analysis is based on 3,900 customer transactions with 18 features, including demographics, purchase details, and behavioral attributes. The total revenue analyzed is $233,081, with an average order value of approximately $59.70 and subscription penetration of 27% (1,053 subscribers).
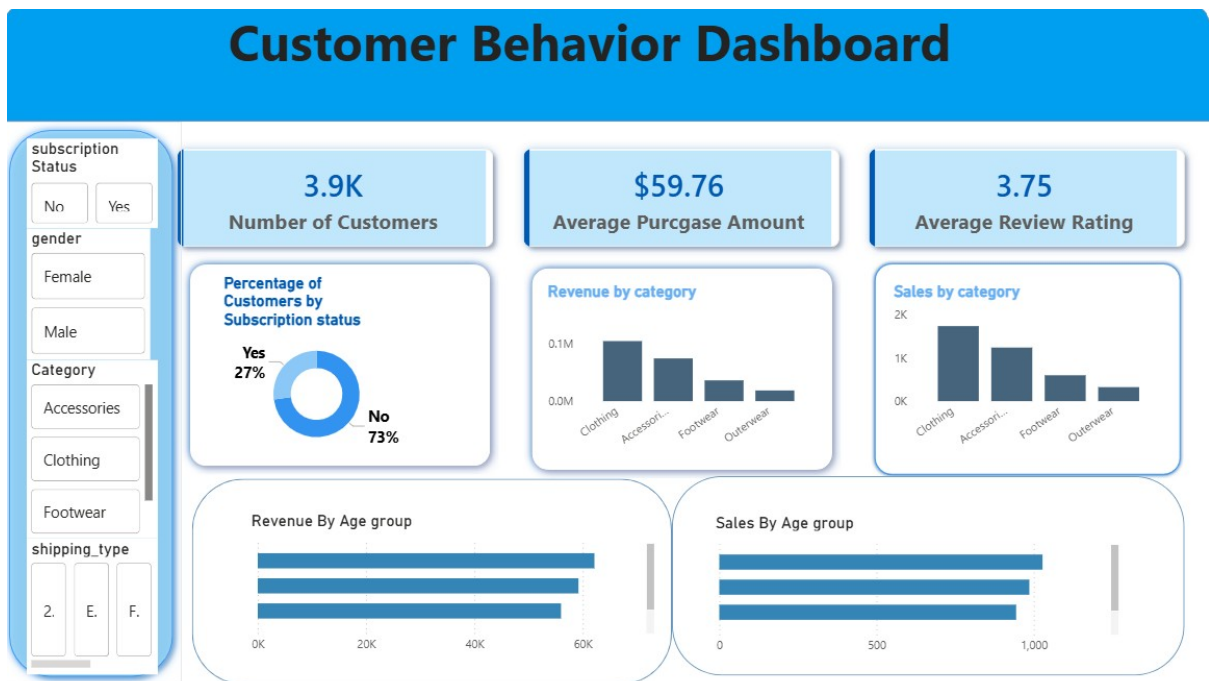

Figure 1: Customer Behavior Dashboard with key KPIs: number of customers, average purchase amount, average review rating, subscription distribution, revenue and sales by category, and age-group performance.

1 Dataset Overview

1.1 Dataset Composition

The analysis uses comprehensive transactional data with the following properties:

- Rows: 3,900 transactions.

- Columns: 18 features describing customer, product, and transaction characteristics.

- Data Quality: 37 missing values in the `review_rating` column, addressed during preprocessing.

1.2 Key Features Analyzed

- Demographics: age, gender, location.

- Purchase Details: item, category, purchase amount.

- Customer Behavior: subscription status, previous purchases, purchase frequency.

- Transaction Features: shipping type, discount applied, review rating.

2 Data Processing and Methodology

2.1 Data Preparation

The following steps were applied to prepare the dataset:

1. **Data Loading and Exploration**: Data was imported using `pandas` and explored with methods such as `df.info()` and `df.describe()` to understand structure and distributions.

2. **Missing Data Handling**: Missing `review_rating` values (37 records) were imputed using the median rating within each product category.

3. **Column Standardization**: Column names were converted to `snake_case` to improve readability and consistency in the analysis code.

4. **Feature Engineering**: New features, including `age_group` and `purchase_frequency_days`, were created to enable deeper behavioral segmentation.

5. **Data Integration**: A redundant `promo_code_used` column was dropped and the cleaned dataset was loaded into PostgreSQL for SQL-based aggregation and analysis.

2.2 Analytical Tools

- Data processing and cleaning: Python with `pandas`.

- Storage and querying: PostgreSQL.

- Methods: exploratory data analysis, SQL aggregation, customer segmentation.

3 Revenue Analysis

3.1 Revenue by Customer Gender

The analysis shows a clear disparity in revenue by gender, with male customers generating substantially higher total revenue than female customers. This pattern is consistent across multiple product categories and shipping preferences, suggesting that gender-specific marketing strategies could unlock additional growth.

3.2 Discount Strategy and High-Value Customers

A total of 839 customers used discounts while still spending above the overall average transaction value. This indicates that discounts are effective at attracting and retaining price-sensitive yet high-value customers, and that carefully designed promotions do not necessarily erode average order values.

3.3 Revenue by Age Group

Age Segment	Revenue Contribution
Young Adults	Highest
Middle-Aged	Second Highest
Adults	Third Highest
Other Segments	Supporting contribution

Table 1: Revenue distribution across age demographics.

Young Adults and Middle-Aged customers are the primary revenue drivers, indicating that campaigns tailored to these demographics are likely to yield the strongest returns.

4 Product Performance and Category Analysis

4.1 Highest-Rated Products

Customer satisfaction analysis highlights several consistently high-performing products in terms of average review rating:

- Gloves

- Sandals

- Boots

- Hat

- Skirt

These products demonstrate strong product–market fit and can be leveraged in marketing and bundling strategies.

4.2 Top Products by Category

Category	Top Products
Accessories	Jewelry, Sunglasses, Belt
Clothing	Blouse, Pants, Shirt
Footwear	Sandals, Shoes, Sneakers
Outerwear	Jacket, Coat

Table 2: Best-performing products by major category.

4.3 Top Discounted Products

The products with the highest proportion of discounted purchases are:

- Hat

- Sneakers

- Coat

- Sweater

- Pants

These items are central to the discount strategy and should be monitored for margin impact and promotional effectiveness.

5 Shipping and Fulfilment Insights

5.1 Shipping Type and Average Spend

Shipping Type	Average Purchase Amount ($)
Express Shipping	60.48
Standard Shipping	58.46

Table 3: Average purchase amount by shipping type.

Express shipping customers exhibit an average order value around 3.5% higher than standard shipping customers, indicating a segment willing to pay both higher product and shipping prices in exchange for speed.

6 Subscription Status and Customer Segmentation

6.1 Subscribers vs. Non-Subscribers

Metric	Subscribed	Not Subscribed
Customer Count	1,053	2,847
Average Purchase ($)	59.49	59.87
Total Revenue ($)	62,645	170,436

Table 4: Comparison of key metrics by subscription status.

Non-subscribers represent approximately 73% of the base and deliver a similar proportion of revenue, while average purchase values between subscribers and non-subscribers are nearly identical. This suggests that the current subscription offer does not yet provide a strong transactional advantage.

6.2 Customer Segmentation Framework

The customer base is segmented into three main groups:

- **Loyal Customers**: 3,116 customers who purchase frequently and provide a stable revenue foundation.

- **Returning Customers**: 701 customers with multiple purchases but lower frequency than the loyal segment.

- **New Customers**: 83 recently acquired customers who require nurturing and onboarding to reach higher value tiers.

6.3 Repeat Purchase Behavior and Subscriptions

Among customers with more than five purchases, there are 958 subscribers and 2,518 non-subscribers. The subscription conversion rate in this high-frequency cohort is therefore similar to the overall base, indicating that repeat purchase behavior alone is not a strong driver of subscription uptake.

7 Average Order Value Analysis

Average order value remains stable at approximately $59.70 across demographics and customer segments. This stability simplifies revenue forecasting and suggests that initiatives such as product bundling, cross-selling, and upselling could increase order values without disrupting established purchasing habits.

8 Key Findings

1. Male customers generate significantly higher revenue, implying untapped potential in better engaging female customers.

2. Average order value is stable at around $60 across segments, providing a predictable baseline for financial modeling.

3. Subscription penetration is 27%, with limited correlation to repeat purchase frequency, indicating that the current subscription value proposition may have plateaued.

4. Discount users (839 customers) maintain above-average spending levels, supporting the case for a structured, margin-aware discount strategy.

5. Young Adults and Middle-Aged customers contribute the majority of revenue, making them priority targets for campaigns and product design.

6. High-performing products such as Gloves, Sandals, and Boots, as well as heavily discounted items like Hats and Sneakers, provide clear levers for merchandising.

7. Express shipping users show a 3.5% higher average order value, pointing to a viable premium service segment.

8. The customer base is mature, with roughly 80% loyal customers and only a small share of new customers, highlighting the importance of both retention and acquisition initiatives.

9 Strategic Recommendations

9.1 Boost Subscription Growth

Enhance subscription benefits (exclusive products, early access, premium support) and introduce tiered plans to increase penetration beyond 27%. Target high-frequency non-subscribers with tailored campaigns that emphasize concrete monetary and experiential value.

9.2 Strengthen Customer Loyalty Programs

Implement a structured loyalty program (for example, points-based rewards and tiered status levels) to increase repeat purchase frequency and customer lifetime value. Provide personalized incentives based on customer segment and product category preferences.

9.3 Optimize Discount Strategy

Refine discount policies by analyzing margin impact at the product and category level and by introducing dynamic discounts based on customer history. Focus promotions on products that already show high discount responsiveness while testing lower discount levels for price-insensitive segments.

9.4 Leverage Product Positioning

Promote highest-rated and best-selling items prominently in marketing channels and on-site placements. Create bundles that pair these top performers with lower-velocity items to lift overall basket value and inventory turnover.

9.5 Target Key Demographics

Design gender- and age-specific campaigns, particularly for Young Adult and Middle-Aged segments, and address the current revenue gap among female customers. Encourage standard-shipping customers to try express services through time-limited promotions, capitalizing on higher observed order values.

9.6 Accelerate New Customer Acquisition

Allocate additional budget to acquisition channels, introduce first-purchase incentives, and deploy referral programs that mobilize existing loyal customers. These actions can increase the proportion of new customers and refresh the long-term revenue funnel.