Retail Customer Behavior Analysis

Data Analytics Project Report

1. Executive Summary

This report presents an analytical study of customer shopping behavior for a retail company using transactional data.
The primary objective is to uncover purchasing patterns, customer loyalty drivers, and product performance insights that can support strategic business decisions.

The analysis was conducted using Python for data preparation, SQL for exploratory analysis, and Power BI for visualization and reporting.

2. Business Objectives

The key objectives of this project are to:

- Understand customer purchasing behavior across demographics
- Identify loyal and high-value customer segments
- Evaluate the impact of discounts and subscriptions on spending
- Analyze product and category performance
- Provide actionable insights to improve marketing and customer engagement strategies

3. Dataset Overview

The dataset consists of **3,900 customer transaction records** and includes:

- Customer demographics (age, gender, location)
- Purchase details (purchase amount, discounts, shipping type)
- Product information (item purchased, category, season)
- Loyalty indicators (previous purchases, purchase frequency)
- Subscription status and payment methods

The dataset was cleaned, standardized, and enhanced through feature engineering before analysis.

4. Methodology & Tools

Tools Used

- **Python (Pandas)** for data cleaning and feature engineering
- **SQL (SQLite via SQLAlchemy)** for exploratory data analysis
- **Power BI** for data visualization and dashboard creation

Data Preparation Steps

- Standardized column naming conventions
- Handled missing review ratings using category-level medians
- Removed redundant features
- Created analytical features such as:
    - Customer age groups
    - Numerical representation of purchase frequency

A final, stable version of the dataset was prepared for reporting and visualization.

5. Key Analytical Insights

5.1 Customer Demographics & Revenue

- Male customers generate higher total revenue primarily due to a larger customer base.
- Average spending per purchase is similar across genders.
- Middle-aged and adult customers contribute the highest share of total revenue.

5.2 Customer Loyalty & Repeat Purchases

- Loyal customers represent the largest customer segment.
- Repeat buyers demonstrate higher engagement and purchasing consistency.
- Older customers tend to show stronger loyalty behavior.

5.3 Discounts & Promotions

- A significant portion of purchases involves discounts.
- Customers who use discounts often spend more than the average purchase amount.
- Certain products show higher discount dependency, indicating price sensitivity.

5.4 Subscription Behavior

- Subscribed customers do not significantly outspend non-subscribers on a per-purchase basis.
- However, subscribers represent a stable and predictable revenue segment.
- Repeat buyers are more likely to subscribe than occasional customers.

5.5 Product & Category Performance

- Clothing is the strongest-performing category in terms of sales volume.
- Several products consistently rank among top-selling items across categories.
- Products with higher repeat purchases also tend to receive higher review ratings.

6. Business Recommendations

Based on the analysis, the following recommendations are proposed:

1. **Strengthen Loyalty Programs**
   Focus on rewarding loyal and repeat customers with personalized offers and exclusive benefits.
2. **Optimize Discount Strategy**
   Apply discounts strategically on products where they increase sales volume without lowering average spending.
3. **Enhance Subscription Value**
   Improve subscription offerings to increase customer retention and long-term engagement.
4. **Prioritize High-Performing Products**
   Allocate marketing and inventory resources toward consistently high-performing products.
5. **Segment-Based Marketing**
   Design targeted campaigns for middle-aged and adult customer segments, which contribute the highest revenue.

7. Limitations & Future Work

Limitations

- The analysis is based on a static dataset and does not include time-series data.
- Seasonal trends could not be analyzed in depth.

Future Work

- Incorporate time-based data for trend and seasonality analysis.
- Develop predictive models to estimate:
    - Customer lifetime value (CLV)
    - Subscription likelihood
    - Repeat purchase probability

8. Conclusion

This project demonstrates how retail transaction data can be transformed into actionable business insights using Python, SQL, and business intelligence tools. The findings provide a solid foundation for improving customer engagement, optimizing marketing strategies, and supporting data-driven decision-making.

Author

Qusay
Data Analytics Portfolio Project