Customer Shopping Behavior Analysis

1. Project Overview

This project analyzes customer shopping patterns using transactional data from **3,900 purchases** across multiple product categories. The objective is to examine spending trends, customer segmentation, product performance, and subscription patterns to support data-driven business decisions.

The analysis was performed using **Python** for data cleaning and preparation, **SQL** for structured business analysis, and **Power BI** for interactive visualization. Key insights were obtained by analyzing customer demographics, purchase history, discount usage, shipping preferences, and review ratings. The final outcome includes an analytical report and an interactive dashboard that presents actionable insights for improving marketing strategies, customer engagement, and revenue optimization.

2. Business Problem Statement

Retail businesses generate large volumes of transactional data; however, without proper analysis, valuable insights into customer behavior remain underutilized. Understanding differences in spending habits, product preferences, discount usage, and subscription status is essential for improving sales performance and customer retention.

The objective of this project was to analyze customer shopping data to identify key behavioral trends and customer segments, and to evaluate factors influencing purchasing and repeat buying decisions. The insights derived aim to support data-driven decisions related to marketing strategies, product positioning, and customer engagement.

3. Dataset Summary

This project utilizes a **customer transactions table** containing **3,900 records** and **18 columns**, where each row represents an individual customer purchase. The table includes customer demographic details, purchase-related attributes, and shopping behavior indicators such as subscription status, discount application, previous purchases, review ratings, and shipping type.

An initial data quality assessment identified a small number of missing values in the review rating column. These were addressed during data preprocessing to ensure data consistency and reliable analysis.

4. Data Analysis using SQL (Business Transactions)

The project followed a structured data analytics workflow to transform raw transactional data into meaningful business insights.

Data Preparation (Python):
Python was used to load, clean, and preprocess the dataset. This included handling missing values, standardizing column names, and performing feature engineering such as creating age groups and customer segments to support analysis.

Data Analysis (SQL):
The cleaned data was stored in a relational database, where SQL queries were executed to analyze customer behavior, spending patterns, product performance, discount usage, and subscription trends.

Data Visualization (Power BI):
Power BI was used to design an interactive dashboard that visually presents key metrics and insights, enabling easy exploration of customer trends and supporting data-driven decision-making.

5. Key Insights

The analysis of customer transactional data revealed the following key insights:

- **Revenue by Gender:**
  Male customers generated significantly higher total revenue compared to female customers, indicating stronger spending contribution from male shoppers.
- **High-Spending Discount Users:**
  A large number of customers used discounts while still spending above the average purchase amount, suggesting that discounts are effective in encouraging higher-value purchases rather than only attracting low spenders.
- **Top-Rated Products:**
  Products such as *Gloves, Sandals, Boots, Hats, and Skirts* recorded the highest average review ratings, reflecting strong customer satisfaction and preference for these items.
- **Shipping Type and Spending Behavior:**
  Customers opting for **Express Shipping** showed a slightly higher average purchase amount compared to those using **Standard Shipping**, indicating a correlation between urgency and higher spending.
- **Subscribers vs. Non-Subscribers:**
  While non-subscribers contributed higher total revenue due to a larger customer base, subscribed customers showed comparable average spending, highlighting the long-term value of subscription-based customers.
- **Discount-Dependent Products:**
  Items such as *Hats, Sneakers, Coats, and Sweaters* had the highest percentage of discounted purchases, indicating that sales of these products are highly price-sensitive.
- **Customer Segmentation:**
  The majority of customers fell into the **Loyal** segment, followed by **Returning** customers, with a smaller proportion of **New** customers. This indicates strong repeat purchasing behavior within the customer base.
- **Top Products by Category:**
  Each product category had clear bestsellers, such as *Jewelry* in Accessories, *Blouses and Pants* in Clothing, *Sandals* in Footwear, and *Jackets* in Outerwear.
- **Repeat Buyers and Subscription Behavior:**
  Most repeat buyers (customers with more than five previous purchases) were non-subscribers, indicating an opportunity to convert loyal customers into subscribers.
- **Revenue Contribution by Age Group:**
  Young Adults and **Middle-aged** customers contributed the highest share of total revenue, making them key target segments for marketing efforts.
- **Subscription Distribution by Gender:**
  Male customers had a higher number of active subscriptions compared to female customers, suggesting gender-based differences in subscription adoption.

6. Dashboard Overview

An interactive dashboard was developed using **Power BI** to visually present the key insights derived from the analysis. The dashboard provides a consolidated view of customer behavior, enabling stakeholders to quickly understand trends and patterns without reviewing detailed data tables.

The dashboard includes key performance indicators such as total customers, average purchase amount, and average review rating. It also presents visual breakdowns of revenue by category and age group, customer distribution by subscription status, and sales patterns across product categories and shipping types. Interactive filters allow users to explore the data dynamically and gain deeper insights into specific customer segments.

7. Business Recommendations

Based on the analysis of customer shopping behavior, the following business recommendations are proposed:

- Strengthen Subscription Programs:
  Introduce exclusive offers and personalized benefits for subscribers to improve retention and encourage non-subscribers, especially repeat buyers, to adopt subscriptions.
- Leverage Discount Strategies Strategically:
  Use targeted discounts on price-sensitive products such as hats, sneakers, and outerwear to boost sales while monitoring margins to avoid over-discounting.
- Focus on High-Performing Products:
  Promote top-rated and best-selling products across categories in marketing campaigns to capitalize on strong customer preferences and satisfaction.
- Target High-Revenue Customer Segments:
  Prioritize marketing efforts toward young adult and middle-aged customers, as they contribute the highest share of revenue.
- Encourage Premium Shipping Options:
  Highlight the benefits of express shipping, as customers choosing this option tend to have higher average purchase values.

8. Conclusion

This project successfully analyzed customer shopping behavior using transactional data to uncover meaningful insights into spending patterns, customer segments, product performance, and subscription behavior. By combining Python for data preparation, SQL for structured analysis, and Power BI for visualization, the analysis translated raw data into actionable business insights.
The findings highlight key opportunities for improving customer engagement, optimizing discount strategies, and strengthening subscription-based retention. Overall, the project demonstrates how data-driven analysis can support informed decision-making and enhance retail business performance.