Customer Behavior Analysis

1. Project Overview

This project focuses on analyzing customer purchasing behavior to understand sales patterns, customer segmentation, and subscription trends. The analysis was carried out using SQL for data querying and Power BI for data visualization.

The main aim of the project is to extract meaningful insights from customer data and present them through an interactive dashboard that can support basic business decision-making.

2. Dataset Summary

The dataset contains customer-level transactional data with the following key attributes:

Rows: 3,900

Columns: 18

Key features:

- Customer demographics (Age, age group, gender, location, subscription status)
- Purchase details (Item purchased, category, previous purchases, frequency of purhases, purchase amount, season, size, colour)
- Review ratings
- Shipping type
- Discount-related information

The dataset was initially reviewed and cleaned manually in Microsoft Excel before being loaded into SQL for analysis.

3. Data Cleaning (Performed in Excel)

Data cleaning was performed manually in Excel to ensure consistency and accuracy before analysis.

The following steps were taken:

- Handling missing values
    - The review rating column contained null values.
    - These null values were replaced with the average review rating of the respective product category, ensuring minimal distortion of data.
- Removing duplicate records
    - Duplicate rows were identified and removed to avoid double-counting customers or transactions.
- Text standardization
    - Text values were standardized (for example, consistent casing for categories and gender).
    - Additional helper columns were created where necessary to manage data efficiently.
- Column name formatting

- Spaces in column names were replaced with underscores to ensure smooth querying in SQL.

- Age group creation
  - A new age group column was created to classify customers into meaningful segments for analysis.
- Removing unnecessary columns
  - Columns such as promo code applied were removed since they provided redundant information already covered by the discount applied column.
- Final data preparation
  - The cleaned dataset was saved and uploaded into SQL for further analysis.

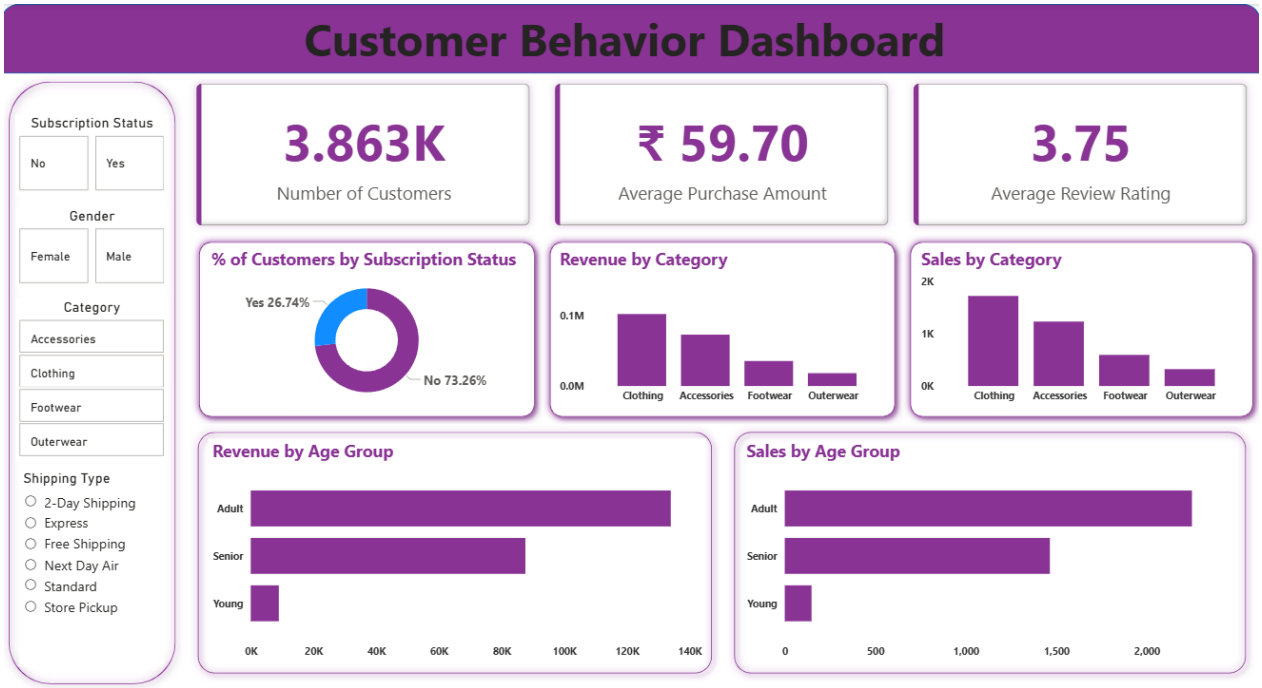
4. Data Analysis Using SQL

SQL was used to perform data aggregation and analysis. The following SQL concepts and functions were applied:

- GROUP BY for category-wise and age-group-wise analysis
- COUNT() to calculate total customers and transactions
- SUM() to calculate total sales and revenue
- AVG() to calculate average purchase amount and ratings
- CASE WHEN for conditional grouping and segmentation
- Window functions (ROW_NUMBER()) to rank and analyze records within specific groups

The results from SQL queries were then connected to Power BI for visualization.

5. Dashboard and Results



The Power BI dashboard provides a consolidated view of customer behavior through interactive visuals.

Key results displayed on the dashboard include:

- Total number of customers
- Average purchase amount
- Average review rating
- Subscription vs non-subscription customer distribution
- Category-wise sales and revenue
- Age-group-wise sales and revenue
- Interactive slicers for:
    - Gender
    - Product category
    - Shipping type
    - Subscription status

The dashboard allows users to filter and explore customer trends dynamically.

6. Key Business Insights

Based on the dashboard analysis, the following insights were observed:

- A large portion of customers are non-subscribers, indicating an opportunity to increase subscription adoption.
- Clothing is the highest-performing category in terms of both sales and revenue.
- Adult customers contribute the most to overall purchases and revenue.
- Younger age groups show comparatively lower purchasing activity.
- The average customer rating suggests moderate satisfaction, highlighting scope for service and product improvements.

7. Business Recommendations

Based on the analysis, the following recommendations can be considered:

- Boost subscriptions - Introduce subscription-based discounts or exclusive offers to convert non-subscribers.
- Customer loyalty programs - Reward repeat customers with loyalty points or special benefits to increase retention.
- Targeted marketing - Focus marketing campaigns on high-performing categories like Clothing.
- Improve customer satisfaction - Analyze feedback and ratings to enhance product quality and service experience.