Customer Shopping Behavior Analysis – Analytical Report

1. Project Overview

This report presents an in-depth analysis of customer shopping behavior using transactional retail data. The objective of this analysis is to identify purchasing patterns, customer segments, product preferences, and the impact of subscriptions and discounts on revenue. Insights derived from this study are intended to support data-driven business decisions related to marketing strategy, customer retention, and revenue optimization.

2. Dataset Summary

- Total Records: 3,900 customer purchase transactions
- Total Attributes: 18 variables describing customers and purchases
- Customer Attributes: Age, Gender, Location, Subscription Status
- Purchase Attributes: Item Purchased, Category, Purchase Amount, Season, Size, Color
- Behavioral Attributes: Discounts Applied, Purchase Frequency, Previous Purchases, Review Ratings, Shipping Type
- Data Quality: 37 missing values identified in the Review Rating column

3. Exploratory Data Analysis Using Python

Exploratory Data Analysis (EDA) was conducted using Python to prepare, clean, and enhance the dataset for advanced analysis. Pandas and NumPy libraries were primarily used for data manipulation and inspection.

- Data Loading and Inspection: Dataset imported into Pandas and reviewed using info() and describe() methods.
- Missing Value Treatment: Review Rating null values were imputed using the median rating per product category.
- Column Standardization: Column names were converted to snake_case to improve consistency and readability.
- Feature Engineering: Created age_group categories and derived purchase frequency metrics.
- Redundancy Removal: Promo code usage was removed after confirming overlap with discount indicators.
- Database Integration: Cleaned data was successfully loaded into PostgreSQL for structured SQL analysis.

4. Data Analysis Using SQL (PostgreSQL)

SQL queries were executed on the cleaned dataset to answer key business questions and generate actionable insights.

- Revenue distribution analysis by gender
- Identification of high-spending customers who used discounts
- Top five products based on average customer review ratings
- Comparison of average purchase amounts by shipping type
- Spending behavior comparison between subscribers and non-subscribers
- Detection of products most dependent on discounts
- Customer segmentation into New, Returning, and Loyal groups

- Top three products within each product category
- Relationship analysis between repeat purchases and subscription adoption
- Revenue contribution analysis by age group

5. Data Visualization and Dashboarding

An interactive Power BI dashboard was developed to visually communicate analytical findings. The dashboard enables stakeholders to explore revenue trends, customer segments, product performance, and subscription behavior through filters and drill-down features.

6. Business Recommendations

- Increase subscription adoption by offering exclusive deals and early-access benefits.
- Implement structured customer loyalty programs to convert repeat customers into loyal segments.
- Optimize discount strategies to maintain profitability while driving sales volume.
- Leverage top-rated and best-selling products in targeted marketing campaigns.
- Focus marketing efforts on high-revenue age groups and customers preferring express shipping.