Customer Shopping Behavior – Data Analysis

Submitted by: Durgesh Kumar
College: Government Engineering College, Vaishali, Bihar
Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Power BI

Abstract

This project aims to analyze customer shopping behavior using data analytics and visualization techniques. The analysis focuses on identifying key patterns, spending habits, and customer preferences that influence purchasing decisions. Using Python for data preprocessing and exploratory data analysis (EDA), followed by visualization in Power BI, this study provides insights to enhance marketing strategies, product placement, and overall customer experience.

1. Introduction

In today's data-driven retail environment, understanding customer behavior is essential for business growth. Companies collect vast amounts of data on customer demographics, transactions, and preferences. Analyzing this data enables businesses to identify high-value customers, predict future purchases, and tailor marketing campaigns accordingly. This project provides a systematic analysis of customer shopping data using Python and Power BI to uncover actionable insights.

2. Objective

The main objectives of this project are: • To study customer demographics and purchase behavior. • To identify trends in spending patterns and product preferences. • To segment customers based on behavior and value. • To provide insights for improving marketing and sales strategies.

3. Dataset Description

The dataset includes customer demographic information, transaction history, and product details. Attributes such as customer ID, gender, age, location, product category, purchase amount, and payment method were analyzed. Data preprocessing involved handling missing values, removing duplicates, and ensuring consistency in formats.

4. Methodology

The methodology followed a structured data analysis lifecycle: 1. Data Cleaning: Missing data handled using imputation and removal of anomalies. 2. Exploratory Data Analysis (EDA): Statistical summaries and correlation analysis conducted using Python libraries such as Pandas and NumPy. 3. Visualization: Trends visualized using Matplotlib and Seaborn; interactive dashboards created in Power BI. 4. Insight Extraction: Identified behavior patterns and customer segmentation based on spending and demographics.

5. Results and Insights

The analysis revealed several key insights: • Young adults (ages 18–35) show higher spending in electronics and fashion categories. • Female customers tend to spend more on fashion and beauty products, while male customers prefer electronics and sports items. • Seasonal trends show peak spending during festivals and holiday seasons. • High-value customers contribute significantly to total revenue, emphasizing the importance of loyalty programs. • Urban customers demonstrated a higher purchase frequency compared to rural customers. These findings suggest that targeted marketing and personalized recommendations can improve conversion rates.

6. Conclusion and Recommendations

This project demonstrates the power of data analytics in understanding customer behavior and improving business strategies. Through systematic EDA and visualization, businesses can segment customers effectively and enhance decision-making. Future enhancements may include predictive modeling using machine learning to forecast customer churn or lifetime value, and integration of real-time analytics dashboards.

7. Tools and Technologies Used

• Python (Pandas, NumPy, Matplotlib, Seaborn) • Power BI • Excel for data preprocessing • Statistical and visual analysis methods

8. References

1. Python Data Science Handbook – Jake VanderPlas 2. Power BI Documentation – Microsoft Learn 3. Towards Data Science articles on EDA and Visualization 4. Kaggle Datasets for Retail Customer Analysis