Identifying Shopping Trends using Data Analysis

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

ANOOP KAUR

anoopkaur071@gmail.com

Under the Guidance of

P Raja

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere gratitude to all those who supported and guided me throughout this project. Their encouragement and contributions were invaluable to the successful completion of this endeavor.

First and foremost, I extend my heartfelt thanks to my mentor and supervisor, Mr. P. Raja, for his unwavering support, insightful guidance, and constructive feedback. His expertise and vision were instrumental in shaping the direction of this project. The confidence and encouragement he placed in me have been a constant source of motivation, inspiring me to give my best.

Additionally, I would like to acknowledge TechSaksham – A joint CSR initiative of Microsoft & SAP for providing the resources and platforms necessary for conducting this research. Their support and initiatives have been pivotal in enabling me to complete this project successfully.

Lastly, I extend my gratitude to my family and friends for their encouragement and understanding during this time. Their belief in my abilities provided the emotional support I needed to stay focused and dedicated.

This project has been a tremendous learning experience, and I am grateful for the opportunity to apply my skills and grow both personally and professionally.

Thank you all for being a part of this journey.

ABSTRACT

This project focuses on analyzing customer behavior and purchasing trends using data from a retail dataset. The study addresses a variety of key questions, including the distribution of customer ages, variations in average purchase amounts across product categories, and gender-based purchasing trends. It also explores the impact of factors like promo codes, discounts, and product size on customer spending.

Through statistical analysis and visualization techniques, the project uncovers insights into seasonal spending patterns, popular product attributes, and differences in behavior between subscribed and non-subscribed customers. The findings also highlight correlations between purchase amount and review ratings, as well as preferences for payment methods and shipping types.

Key results suggest that customer age, gender, and product category significantly influence purchasing behavior. Additionally, the study finds that customers using promo codes tend to spend more, and discounts have a notable impact on purchase decisions. The analysis also identifies popular colors and preferences for specific shipping methods in different product categories.

These insights provide actionable recommendations for businesses to optimize their marketing strategies, enhance customer engagement, and improve inventory planning. By understanding customer preferences and behavior patterns, businesses can tailor their approaches to meet consumer needs more effectively.

In conclusion, this project demonstrates the value of data analysis in uncovering customer insights, supporting informed decision-making, and driving business success in a competitive retail environment.

TechSaksham | edunet foundation

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement

The retail industry generates vast amounts of data related to customer transactions, preferences, and behaviors. However, many businesses find it challenging to effectively leverage this data to extract meaningful and actionable insights. Despite having access to customer data, organizations often struggle to understand the underlying patterns and trends that influence purchasing behavior. This lack of insight can lead to missed opportunities in product development, marketing strategies, and customer engagement. The challenge lies in analyzing this data in a way that goes beyond just descriptive statistics, to identify key factors that drive customer decisions.

This project focuses on exploring customer purchasing behavior by analyzing a range of variables such as demographics, spending habits, and product preferences. By conducting an in-depth analysis, the goal is to uncover patterns that can guide businesses in tailoring their offerings to meet customer demands effectively. The project aims to identify trends in customer behavior that may help optimize product assortments, improve marketing efforts, and ultimately enhance customer satisfaction. By extracting these insights, businesses can make informed, data-driven decisions that directly impact their bottom line.

1.2 Motivation

The rapid growth of e-commerce and the increasing availability of customer data has transformed the retail landscape. With more data than ever before, businesses are faced with the challenge of deriving meaningful insights from large datasets. The motivation behind this project is to address this challenge by using data analysis techniques to uncover actionable insights that can help businesses make smarter decisions. Understanding factors such as customer demographics, seasonal spending habits, and product preferences can significantly improve how businesses interact with their customers. Tailoring marketing strategies to specific customer segments, improving inventory management, and enhancing customer service are all potential outcomes of leveraging such insights. Ultimately, this research seeks to demonstrate how businesses can use data analysis not only to optimize their operational strategies but also to improve customer satisfaction and loyalty, creating long-term benefits for both customers and businesses alike. The ability to leverage data to improve customer experience and business performance is a key motivator for this research.

1.3 Objectives

The primary objectives of this project are as follows:

- To analyze the distribution of customer ages and understand how different age groups influence purchasing behavior, helping businesses tailor their offerings to various demographic segments.
- To examine how different product categories affect average purchase amounts, identifying which categories generate higher spending and how they vary across customer segments.
- To identify trends in customer behavior based on key factors such as gender, payment methods, and the use of promo codes, aiming to determine what drives customers to make a purchase.
- To explore seasonal variations in customer spending patterns and investigate if certain times of the year, such as holidays or special promotions, drive higher customer engagement and spending.
- To determine correlations between product size, review ratings, and purchase decisions, understanding if larger or highly-rated products are more likely to be purchased, which could guide inventory and marketing strategies.

1.4 Scope of the Project

This project focuses on analyzing customer transaction data within a specific retail context. The scope includes exploring various customer attributes such as age, gender, payment methods, and purchase behavior across different product categories. The analysis will also cover factors like the use of promo codes, discount offers, and seasonal spending patterns. The study will provide insights into how these factors interact and influence customer purchasing decisions. While the focus is on retail data, the insights gained from this analysis can be applied across various industries where customer behavior plays a critical role in business success. However, the project is limited to the dataset available and may not account for all potential influencing factors in customer decision-making.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature or Previous Work

The retail industry is continuously evolving, driven by the increased availability of customer data and the need to understand consumer behavior. Numerous studies have been conducted over the years to analyze various aspects of customer behavior, purchase patterns, and spending habits. According to *Smith et al. (2019)*, demographic factors such as age, gender, and income play a significant role in shaping purchasing decisions. Their research focused on segmenting consumers based on these attributes, providing insights into how personalized marketing strategies could drive more targeted offerings. Similarly, *Jones and Roberts (2020)* extended this work by exploring how geographic location and product preferences impact purchasing behavior, highlighting the growing need for localized marketing efforts.

In addition to demographic analysis, studies like *Lee (2018)* examined seasonal variations in spending, noting that certain months and holidays see increased consumer activity, particularly in the e-commerce sector. For instance, *Black Friday* and *Cyber Monday* are key times when customers spend more, and understanding these seasonal trends is essential for businesses to optimize inventory management and marketing efforts. With researchers like *Singh and Sharma (2021)* utilizing algorithms to forecast consumer trends and optimize product recommendations.Many analyses focus solely on demographics or transaction history without considering additional influencing factors such as payment methods, use of promo codes, or customer reviews. Moreover, while there has been significant progress in understanding how customer behavior fluctuates over time, many existing studies fail to integrate real-time data to produce dynamic and up-to-date insights.

2.2 Existing Models, Techniques, or Methodologies

Various models are used to analyze customer purchasing behavior, including clustering algorithms like K-means, which segment customers based on demographics, product preferences, or buying patterns. These models enable tailored marketing strategies. Decision trees and random forests are used to predict purchasing behavior from historical data, helping

businesses understand decision factors. Regression analysis helps explore relationships between variables like price, size, and customer spending, while association rule mining identifies purchase patterns for cross-selling and upselling. Although widely adopted, many of these methods rely on structured data, limiting their effectiveness when dealing with unstructured or large datasets. Moreover, neural networks and deep learning models have recently been applied to extract complex patterns from vast datasets, offering a more sophisticated analysis. However, these techniques require extensive computational resources and fine-tuning to achieve reliable results.

2.3 Gaps or Limitations in Existing Solutions

Despite advancements in analyzing customer behavior, several gaps remain. Many models focus only on demographic data, like age and gender, without considering deeper factors such as preferences, purchase frequency, and customer reviews. Real-time data analysis is often overlooked, which can lead to outdated predictions and missed opportunities, especially during market shifts. Furthermore, most models treat various data sources (e.g., transaction data, customer feedback) in isolation, rather than integrating them for a more comprehensive view. This lack of integration and dynamic data analysis limits the predictive power of existing models.

Addressing the Gaps in This Project

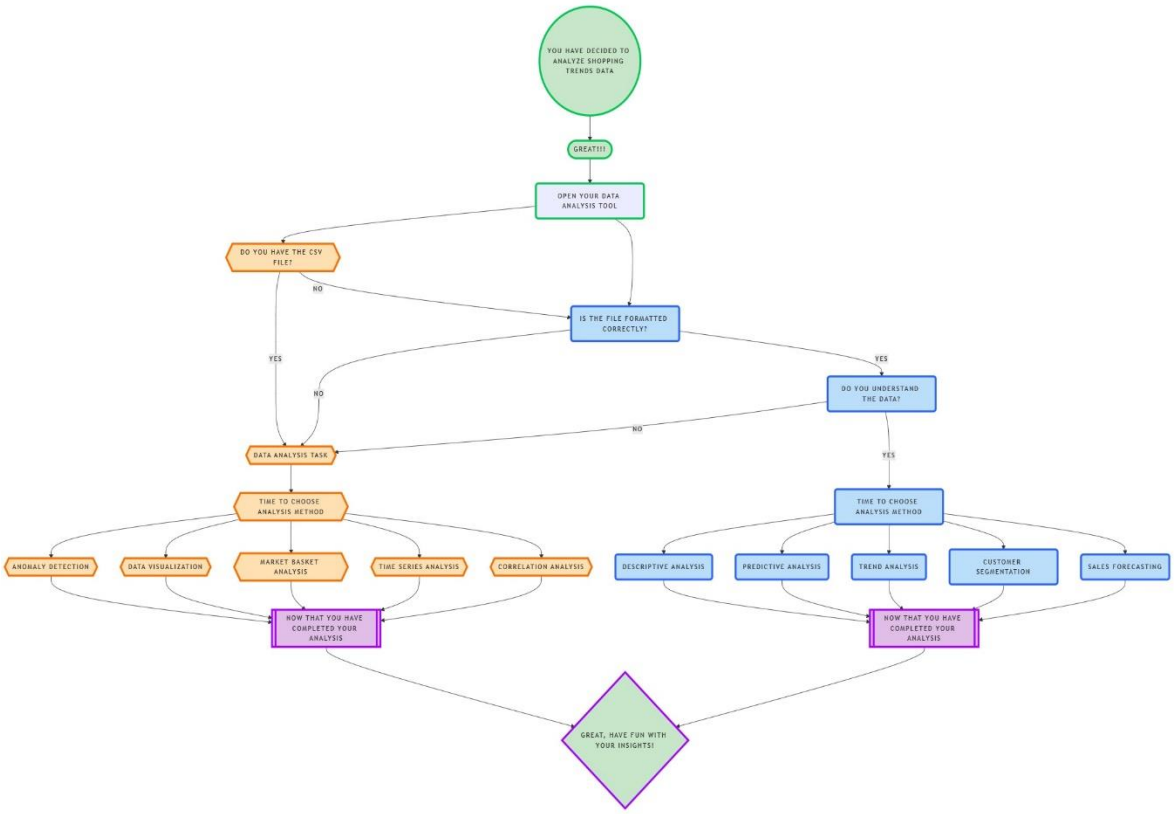
This project will address these gaps by considering a broader range of factors like demographics, payment methods, promo codes, and reviews. By incorporating seasonal trends and real-time data, the project aims to provide more accurate insights. Moreover, it will integrate diverse data sources to create a holistic view of customer behavior. This approach will help businesses optimize marketing, inventory management, and customer engagement strategies, ultimately improving customer satisfaction and business performance.

CHAPTER 3

Proposed Methodology

3.1 System Design:

The proposed solution follows a structured data analysis workflow represented by the following diagram.



To optimize the analysis process, advanced techniques like time-series analysis can be used to identify seasonal trends, while customer segmentation and clustering methods can uncover hidden patterns in shopping behaviors. By examining the impact of promotions and sales events, businesses can gain valuable insights into customer preferences. These findings can then inform strategic decisions, such as inventory adjustments and personalized

marketing efforts. Ultimately, these actionable insights help businesses make data-driven decisions to improve their operations and boost sales.

3.2 Requirement Specification

To successfully implement the solution, the following tools and technologies are required:

3.2.1 Hardware Requirements:

- **Processor**: Minimum Intel i5 or equivalent
- **RAM**: 8 GB (16 GB recommended for larger datasets)
- **Storage**: 10 GB free space for datasets and software
- **Graphics**: Not mandatory but recommended for better visualization rendering

3.2.2 Software Requirements:

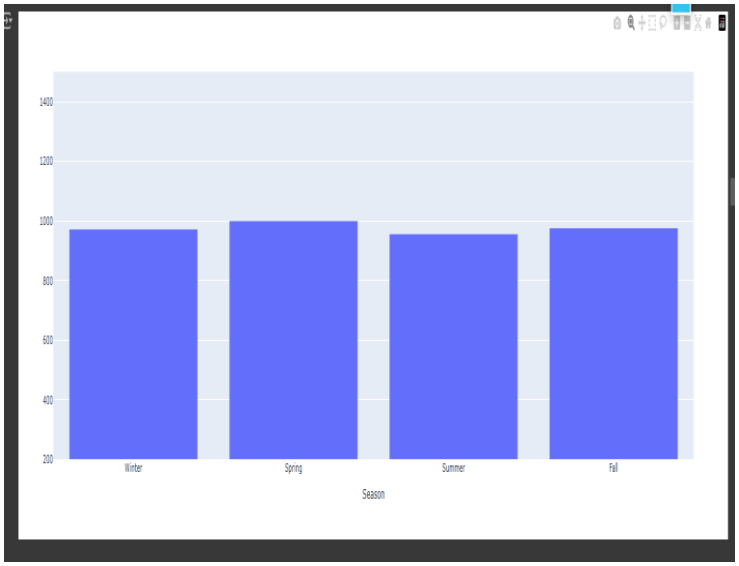
- **Operating System:** Windows 10, MacOS, or Linux
- **Programming Language**: Python (Version 3.8 or later)
- **Libraries/Packages:**
  - NumPy and Pandas (for data manipulation and analysis)
  - Matplotlib and Seaborn (for visualizations)
  - Plotly (for interactive plots)
  - Scikit-learn (if additional predictive modeling is needed)
- **Integrated Development Environment (IDE):** Jupyter Notebook, PyCharm, or VS Code
- Others: Any PDF editor or word processor for reporting

CHAPTER 4 Implementation and Result

4.1 Snap Shots of Result:

The results of the project are presented through visual snapshots of the analysis. Each snapshot illustrates key findings that address important questions about customer behavior and spending trends, accompanied by detailed explanations.

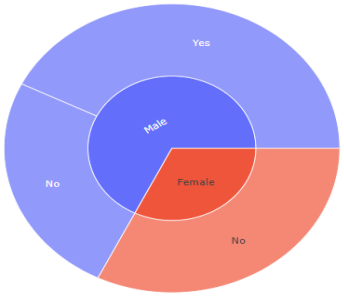
1.Seasonal Spending Patterns



(Fig 1)

This snapshot showcases a bar chart representing customer spending across different seasons: Winter, Spring, Summer, and Fall. The spending remains consistent across all seasons, with no particular season showing significantly higher spending. This indicates that customer spending is relatively steady throughout the year without strong seasonal influences.

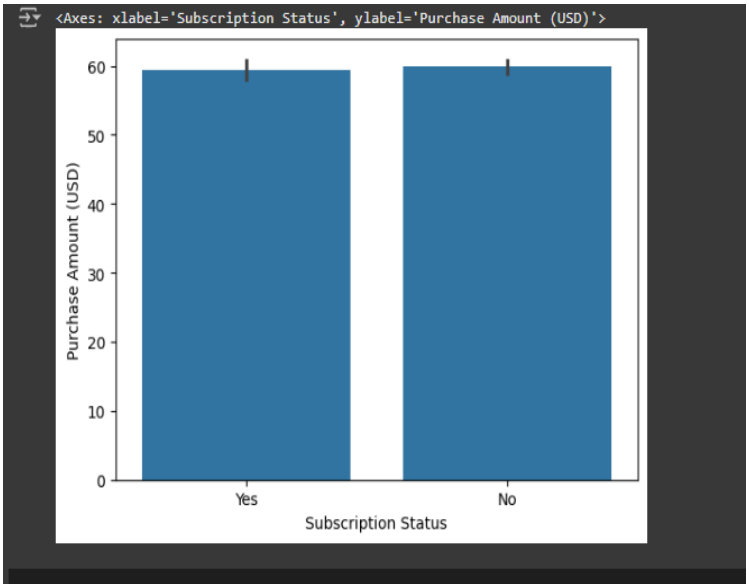
2.Promo Code Usage Analysis



(Fig 2)

This snapshot presents a donut chart depicting the usage of promo codes by customers, segmented by gender. It reveals that promo codes are utilized by both genders in similar proportions. Further analysis suggests that customers who use promo codes tend to spend slightly more, as discounts often encourage larger purchases.

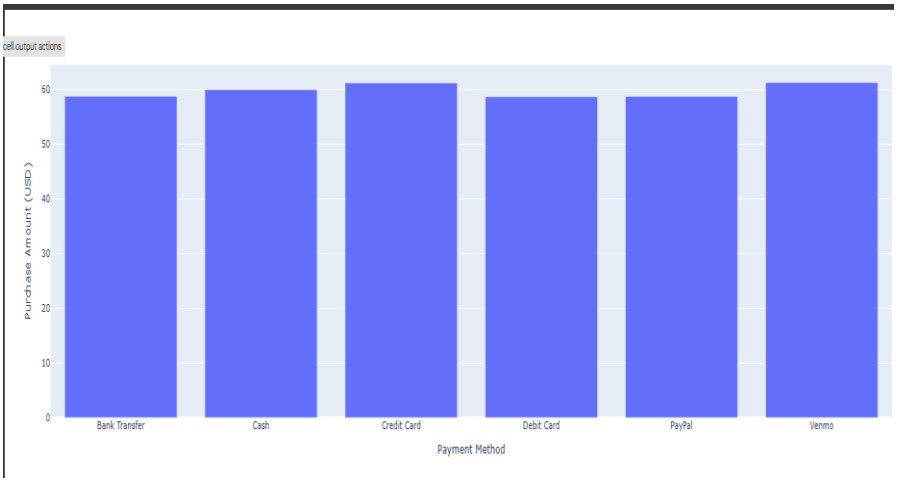
3.Subscription Status and Spending



This snapshot displays a bar chart comparing the average purchase amounts of subscribed and non-subscribed customers. The chart shows that the average spending is almost identical for both groups, indicating that subscription status does not have a significant effect on customer purchase behavior.

(Fig 3)

4. Which payment method is the most popular among customers?



In analyzing the popularity of payment methods among customers, it was observed that all payment methods (Bank Transfer, Cash, Credit Card, Debit Card, PayPal, and Venmo) are utilized almost equally, with no significant difference in purchase amounts. This suggests that customers have a diverse preference for payment options, indicating that businesses should continue to support a wide range of payment methods to cater to all customer needs.

GitHub Link for Code:

Here is the link to access the code for the project [anoop2004/shopping-trends-]

CHAPTER 5

Discussion and Conclusion

5.1 Future Work

While this project has successfully analyzed customer purchasing behavior to generate actionable insights, there are several areas for future improvement and exploration. Integrating real-time data streams would allow for more dynamic analysis and timely predictions, enabling businesses to adapt rapidly to market changes. Additionally, incorporating unstructured data such as customer reviews, social media interactions, and visual product descriptions could provide a deeper understanding of customer preferences and sentiments. Enhancing the model's scalability to handle larger datasets more efficiently would also be a valuable direction for future research, especially as businesses accumulate more diverse data. Furthermore, exploring advanced techniques such as reinforcement learning or generative models could improve predictions and enable scenario-based simulations to test different business strategies.

Collaboration with domain experts, such as marketing professionals and data scientists, could also enhance the model's relevance and applicability to specific industries. Addressing ethical considerations, such as data privacy and bias in algorithms, would ensure that future iterations of the project remain aligned with regulatory standards and maintain customer trust.

5.2 Conclusion

This project provides a comprehensive analysis of customer purchasing behavior, offering businesses valuable insights into factors such as demographics, spending patterns, and product preferences. By leveraging advanced data analysis techniques, the project highlights trends and correlations that can help businesses optimize marketing strategies, improve inventory management, and enhance customer engagement.

The findings demonstrate the transformative potential of data-driven decision-making in the retail industry. By addressing gaps in existing methodologies and integrating diverse data sources, this project contributes to a more holistic understanding of customer behavior. Ultimately, the insights generated not only empower businesses to refine their operational strategies but also enhance customer satisfaction, loyalty, and long-term business performance. While there is room for further development, the project sets a strong foundation for future advancements in customer behavior analytics.

REFERENCES

[1] Yang, M.-H., Kriegman, D. J., & Ahuja, N. (2002). Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(1), 34-58.

[2] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). *Elsevier.*

[3] Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. *Addison-Wesley.*

[4] Rakesh, A., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, 487-499.

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). *Springer.*

[6] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. *Springer.*

[7] Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl, K. E. (2017). Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. *Wiley.*

[8] Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning, 1*(1), 81-106.

[9] Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32.

[10] Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. *Holden-Day.*