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

BRITHIKSHA NV

brithikshanv@gmail.com

Under the Guidance of

Mr. P Raja

Mr. Jay Rathod

Mr. Abdul Aziz

ACKNOWLEDGEMENT

 I would like to take this opportunity to express my heartfelt gratitude to everyone who supported me, directly or indirectly, throughout the course of this project on identifying shopping trends using data analysis.

First and foremost, I extend my sincere thanks to my supervisor, Mr. P Raja whose mentorship and guidance have been invaluable. His insightful advice, constructive feedback, and constant encouragement have played a pivotal role in shaping this project. His unwavering confidence in my abilities has been a tremendous source of motivation, inspiring me to strive for excellence. Collaborating with him over the past year has been a truly enriching experience. His support extended beyond the scope of this project, influencing not just my academic pursuits but also my personal and professional growth.

I am deeply grateful Mr. Jay Rathod and Mr. Abdul Aziz, whose assistance and insights contributed significantly to the success of this project. Their contributions, whether through technical advice, resources, or moral support, have been instrumental in helping me achieve my objectives.

I also acknowledge the support of my friends, family, and colleagues, who provided encouragement and understanding throughout this journey. Their belief in me helped me stay focused and determined, even during challenging times.

Lastly, I thank everyone who contributed, directly or indirectly, to the completion of this project. This experience has not only enhanced my technical skills but also taught me the importance of collaboration, dedication, and resilience in achieving my goals.

ABSTRACT

This project aims to identify shopping trends by analysing customer purchase data. The primary goal is to uncover patterns in customer behaviour, preferences, and purchasing habits to provide actionable insights for targeted marketing and business strategy optimization.

The dataset comprises 3,900 records with 18 features, including demographic details, product attributes, purchase amounts, and customer interactions. The problem statement focuses on understanding how factors such as age, gender, location, and discounts influence purchasing behaviour, seasonal trends, and product popularity.

The methodology involved exploratory data analysis (EDA) to identify patterns and correlations, followed by data visualization techniques to highlight trends. Statistical analysis was conducted to determine significant factors affecting purchases, and clustering techniques were used to segment customers into distinct groups based on shared characteristics.

- Certain product categories and colours were more popular in specific seasons.

- Discount availability and promo codes significantly influenced purchase decisions.

- Subscription status and payment methods were strong indicators of customer loyalty and frequency of purchases.

The findings suggest tailored recommendations for inventory management, personalized promotions, and customer segmentation strategies to enhance sales and customer satisfaction. This analysis underscores the importance of data-driven decision-making in understanding and predicting shopping trends, ultimately driving business growth.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1   Problem Statement:

•   Retail businesses accumulate vast amounts of shopping data from multiple channels (in-store, online, etc.), but struggle to effectively analyze this data to identify emerging trends, customer preferences, and seasonal buying patterns.

•   Failure to identify and act on shopping trends could result in lost revenue, overstock or stockouts, ineffective marketing strategies, and an inability to maintain competitive advantage in the retail market.

•   The challenge primarily affects retail operations, marketing teams, and inventory management, as they rely on accurate and timely trend analysis to drive decision-making.

1.2   Motivation:

•   The project was chosen to help businesses adapt to rapidly changing consumer behaviors in a competitive retail market.

•   It addresses the impact of e-commerce growth, social media influence, and emerging technologies like AI and AR on shopping patterns.

•   Potential applications include optimizing inventory, personalizing marketing campaigns, and enhancing customer experiences.

- It aims to identify key trends such as seasonal demand, popular product categories, and preferred shopping channels.

- The project can drive innovation, improve operational efficiency, and foster customer loyalty.

- Its insights support data-driven decision-making to boost sales and ensure long-term business growth.

1.3 Objective:

- Understand and target the right customers.
- Optimize store locations and performance.
- Develop effective marketing and operational strategies.
- Mitigate risks and adapt to market changes.
- Measure and improve business performance.

1.4 Scope of the Project:

- **Predictive Analytics**: Implement machine learning algorithms to predict future shopping trends based on historical data.

- **Real-time Data Analysis**: Integrate real-time data streams for dynamic trend analysis and immediate decision-making.

- **Customer Segmentation**: Use advanced clustering techniques to segment customers based on shopping behaviors for targeted marketing.

- **Integration with Business Intelligence Tools**: Incorporate the analysis into Bi tools like Tableau or Power BI for broader accessibility and real-time dashboards.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

- Analyzing shopping trends is a well-researched area in retail analytics, consumer behavior studies, and data science.

- Prior work has focused on leveraging machine learning techniques, such as clustering and predictive modeling, to segment customers, forecast sales, and identify preferences.

- Extensive research has been conducted on the role of seasonality, pricing strategies, and digital marketing in shaping shopping behavior.

- With the advent of big data and advanced analytics, there remains a significant opportunity to refine these methodologies for better accuracy and scalability.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

- **Customer Segmentation:** Techniques like K-Means clustering and hierarchical clustering are commonly used to group customers based on behavior**.**

- **Demand Forecasting:** Time series models such as ARIMA, Prophet, and LSTM are widely adopted for predicting future sales trends.

- **Recommendation Systems**: Collaborative filtering and content-based filtering techniques are utilized to provide personalized product recommendations.

- **Visual Analytics:** Tools like Tableau and Seaborn help in identifying patterns and trends through graphical representations.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

- **Data Silos:** Many studies lack integration across multiple datasets (e.g., transactional, demographic, and web traffic data).

- **Limited Real-Time Insights:** Existing models often fail to provide actionable, real-time insights.

- **Scalability Issues:** Traditional methodologies struggle with the scale and complexity of modern e-commerce datasets**.**

- **-Personalization Challenges:** Many solutions lack the ability to offer highly personalized recommendations beyond broad categories.

How This Project Addresses These Gaps?

This project integrates advanced machine learning techniques and exploratory data analysis to provide a holistic view of shopping trends. By combining diverse data sources and emphasizing actionable insights, it aims to bridge the gap between theoretical research and practical applications.

CHAPTER 3

Proposed Methodology

3.1   System Design

To address the problem of identifying shopping trends, the following data-driven solution is proposed. This solution will help businesses a stay competitive by identifying and responding to shopping trends in a timely, efficient and data-driven manner.
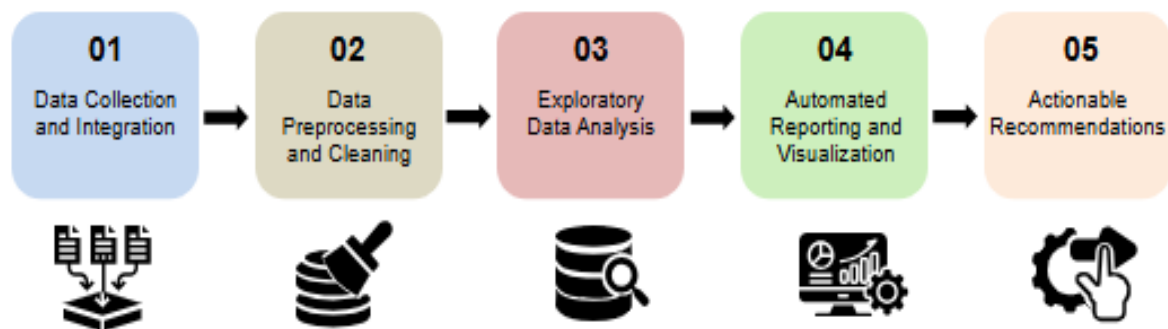
1.Data Collection and Integration: This is the initial stage where relevant data is gathered from various sources. The data may be in different formats and structures, so it needs to be integrated into a unified format for further analysis.

2.Data Preprocessing and Cleaning: The collected data often contains inconsistencies, errors, and missing values. This step involves cleaning and transforming the data to ensure its quality and consistency.

3.Exploratory Data Analysis: In this stage, the cleaned data is analyzed using various techniques to gain insights into its underlying patterns,

trends, and relationships. This helps to understand the data better and identify potential areas of interest.

4.Automated Reporting and Visualization: The results of the analysis are presented in a clear and concise manner through automated reports and visualizations. This helps to communicate the findings effectively to stakeholders.

5.Actionable Recommendations: Based on the insights gained from the analysis, actionable recommendations are made. These recommendations can be used to inform decision-making and drive business outcomes.

3.2   Requirement Specification

3.2.1   Hardware Requirements:

- **Processor:** Intel i5 or higher

- **RAM:** 8GB or more

- **Storage:** Minimum 500GB HDD/SSD

3.2.2   Software Requirements:

- **Python 3.x:** The programming language used for data analysis.

- **NumPy:** For numerical data operations and array manipulation

- **Pandas:** For data manipulation and analysis

- **Seaborn:** For statistical data visualization.

- **Jupyter Notebook:** Python IDE

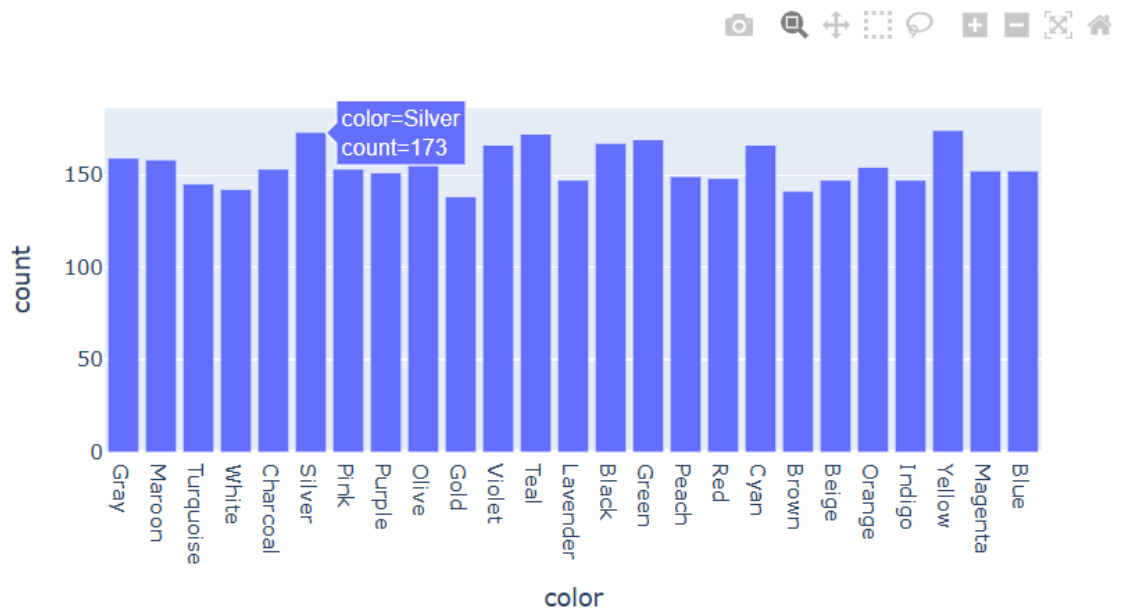- **Matplotlib:** For additional plotting capabilities.

CHAPTER 4

Implementation and Result

### 4.1 Snap Shots of Result:

13. Are there any specific colors that are more popular among customers?

```
px.histogram(df , x = 'color')
```


4.1.1 Color Distribution:

- The bar chart shows the frequency of different colors.
- Blue, Silver, Purple, and Teal are the most frequent colors.
- Maroon, Gray, and Turquoise are the least frequent.
- This distribution likely reflects color preferences within the dataset's context.

Promo Code Usage by Preferred Payment Method
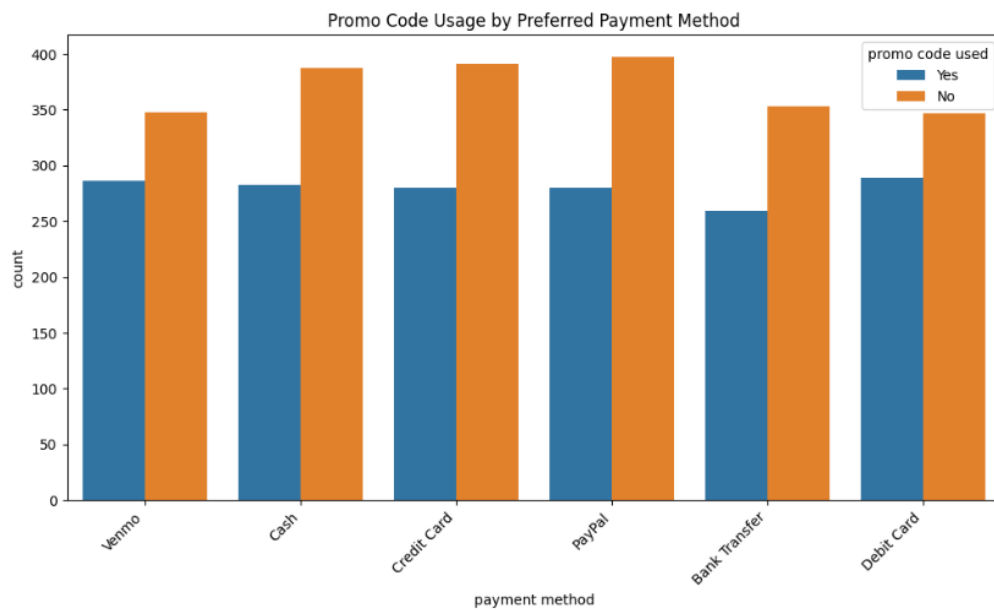
4.1.2 Promo Code Usage by Preferred Payment Method:

- The bar chart compares promo code usage across different payment methods.

- Overall, a higher number of customers used promo codes regardless of the payment method.

- Credit Card and Debit Card users had the highest promo code usage, while Venmo and Cash users had the lowest.

- This suggests promo codes are effective, but usage varies by payment method.

```
Purchases by Gender:
gender
1    2652
0    1248
Name: count, dtype: int64
```

Number of Purchases by Gender

4.1.3 Purchases by Gender:

- The bar chart shows the number of purchases made by two genders, likely male and female.
- Males (represented by "1") made significantly more purchases than females (represented by "0").
- This suggests a potential difference in purchasing behavior between the two genders.

Distribution of Customer Ages
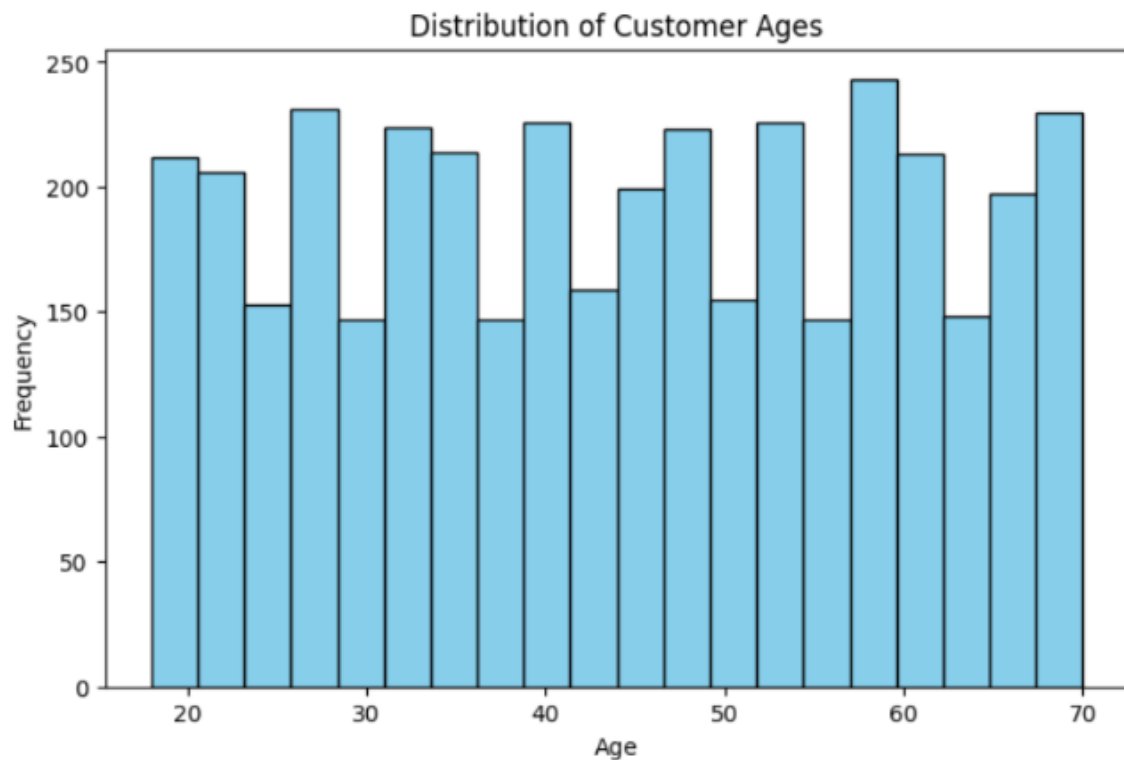
4.1.4 Distribution of Customer Ages:

- The histogram displays the distribution of customer ages.
- There's a high frequency of customers in the 20-30 age range.
- The frequency gradually decreases as age increases, with a slight uptick in the 60-70 age group.
- This indicates the business attracts a younger customer base with a segment of older customers.

4.2 GitHub Link for Code:

https://github.com/brithikshanv/Shopping-Trends-Analysis

CHAPTER 5

Discussion and Conclusion

5.1   Future Work:

- Explore real-time data integration to enable dynamic trend analysis.

- Incorporate deep learning techniques for more accurate demand forecasting.

- Expand the analysis to include unstructured data, such as customer reviews and social media posts.

- Enhance recommendation systems with hybrid approaches combining collaborative and content-based filtering.

5.2   Conclusion:

- Analyzing shopping trend data using Python's robust libraries provides a powerful approach to understanding customer behavior and market dynamics.
- By extracting actionable insights from the data, businesses can make informed decisions that enhance operational efficiency and drive growth.
- The future scope of this analysis includes expanding into predictive analytics and real time processing further enriching the decision-making process.

REFERENCES

1.Kaggle:
https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset

2.Techsaksham study material:
https://techsaksham.org/view-study-material/215

3.Session Recordings (Tech Saksham):
https://techsaksham.org/session-recording