**Project Title**

**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

Name & E-mail

Under the Guidance of

**Jay Rathod Sir &**

**P Raja Sir**

**ACKNOWLEDGEMENT**

I, Shivraj Pawar, would like to express my heartfelt gratitude to everyone who supported me throughout the completion of this project.

First and foremost, I extend my sincere thanks to my guides, Mr. Jay Rathod and Mr. P. Raja, for their invaluable guidance, encouragement, and constructive feedback during the course of this project. Their insights and mentorship were instrumental in helping me achieve the objectives of this work.

I am also grateful to the faculty and staff of Edunet Foundation, who provided the necessary resources and facilities required for the successful execution of this project.

Finally, I would like to thank my family and friends for their unwavering support, motivation, and understanding throughout this journey. Their encouragement helped me stay focused and motivated to complete this work successfully.

#### **ABSTRACT**

The project "Shopping Trends Analysis using Python" aims to uncover meaningful insights from customer shopping data. By analyzing customer demographics, purchasing patterns, and product preferences, this project provides valuable information for improving business strategies and enhancing customer experiences.

The dataset used in this project contains details such as customer age, gender, product categories, purchase amounts, and payment methods. Various Python libraries, including pandas, matplotlib, and seaborn, were utilized to perform data cleaning, exploratory data analysis (EDA), and visualization.

Key findings include the identification of the most popular product categories, trends in seasonal purchases, gender-based preferences, and the impact of discounts on spending behavior. Additionally, the project explores customer behavior based on age groups, subscription status, and preferred payment methods.

The results reveal significant patterns, such as higher spending during specific seasons, popular product categories across different demographics, and a positive correlation between discounts and customer spending. These insights can help businesses optimize inventory management, marketing campaigns, and customer engagement strategies.

This project concludes by emphasizing the importance of data-driven decision-making in the retail sector and suggests future enhancements, such as integrating predictive analytics or using larger datasets for deeper insights.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 2

**Chapter 2.**  **Literature Survey 3**

2.1 Literature Survey 3

**Chapter 3.**  **Proposed Methodology 5**

* 1. System Design 5

3.2 Requirement Specification 8

3.2.1 Hardware Requirement

3.2.2 Software Requirement

**Chapter 4.**  **Implementation and Results 9**

* 1. Snap Shot of Result 9
  2. GitHub Link for Code 11

**Chapter 5. Discussion and Conclusion 12**

5.1 Future Work 12

5.2 Conclusion 13

**References** 14

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | Number Of Purchase by Gender | **9** |
|  | Total Spending by Season | **9** |
|  | Frequency Of Purchases by Age Group | **10** |
|  | Most Popular Payment Methods | **10** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Caption** | **Page No.** |
| **Table 1** | System Design Overview | **5** |
| **Table 2** | Comparison of Methodology | **6** |
| **Table 3** | Hardware Requirement | **8** |
| **Table 4** | Software Requirement | **8** |
| **Table 5** | Snapshot of Results | **9** |
| **Table 6** | GitHub Link for Code | **11** |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**Introduction**

In today’s digital era, data plays a critical role in understanding customer behavior and shaping business strategies. The retail sector generates vast amounts of data daily, making it essential to analyze shopping trends to gain valuable insights. This project, "Shopping Trends Analysis using Python," focuses on exploring and understanding customer purchasing behavior through data analysis techniques.

* 1. **Problem Statement:**

The lack of structured insights into customer preferences and shopping patterns often leads to inefficiencies in inventory management, marketing, and customer engagement. Businesses need data-driven strategies to identify trends, improve decision-making, and enhance the overall customer experience.

* 1. **Motivation:**

The lack of structured insights into customer preferences and shopping patterns often leads to inefficiencies in inventory management, marketing, and customer engagement. Businesses need data-driven strategies to identify trends, improve decision-making, and enhance the overall customer experience.

* 1. **Objective:**

The primary objectives of this project are:

* To analyze customer demographics and their impact on shopping behavior.
* To identify the most popular product categories and seasonal trends.
* To explore the influence of discounts, promo codes, and payment methods on customer spending.
* To derive actionable insights for improving business strategies.
  1. **Scope of the Project:**

This project focuses on analyzing a dataset containing customer details, purchase history, and product information. Using Python and its libraries, such as pandas and seaborn, the data is cleaned, explored, and visualized to generate meaningful insights. The scope is limited to the dataset provided and does not include predictive modeling or real-time analysis.

**CHAPTER 2**

**Literature Survey**

The analysis of shopping trends has been extensively explored in the retail sector to understand customer behavior, preferences, and purchasing patterns. Previous studies have demonstrated the value of leveraging data analytics to uncover insights such as product popularity, seasonal trends, and the impact of promotions. For example:

- A study by Smith et al. (2018) analyzed customer purchase histories to predict seasonal product demand, providing actionable insights for inventory management.

- Jones et al. (2020) focused on how discounts and promotional offers significantly influence customer spending, helping businesses design targeted marketing campaigns.

Several methodologies and techniques have been employed for analyzing shopping trends:

1. Exploratory Data Analysis (EDA): Python libraries such as pandas, NumPy, matplotlib, and seaborn are widely used for data cleaning, aggregation, and visualization.

2. Clustering Techniques: Customer segmentation is often performed using methods like K-means clustering, allowing businesses to group customers based on demographics, spending behavior, or preferences.

3. Machine Learning Models: Algorithms like regression, classification, and time-series forecasting are utilized to predict future sales, customer lifetime value, or product demand.

4. Business Intelligence Tools: Platforms such as Power BI and Tableau are used to create interactive dashboards for visualizing trends and decision-making.

While these methodologies provide robust insights, certain limitations persist:

- Lack of Interpretability: Advanced models can be complex, making it difficult for stakeholders to derive clear, actionable insights.

- Real-Time Analysis: Many existing studies focus on historical data and do not incorporate real-time trends.

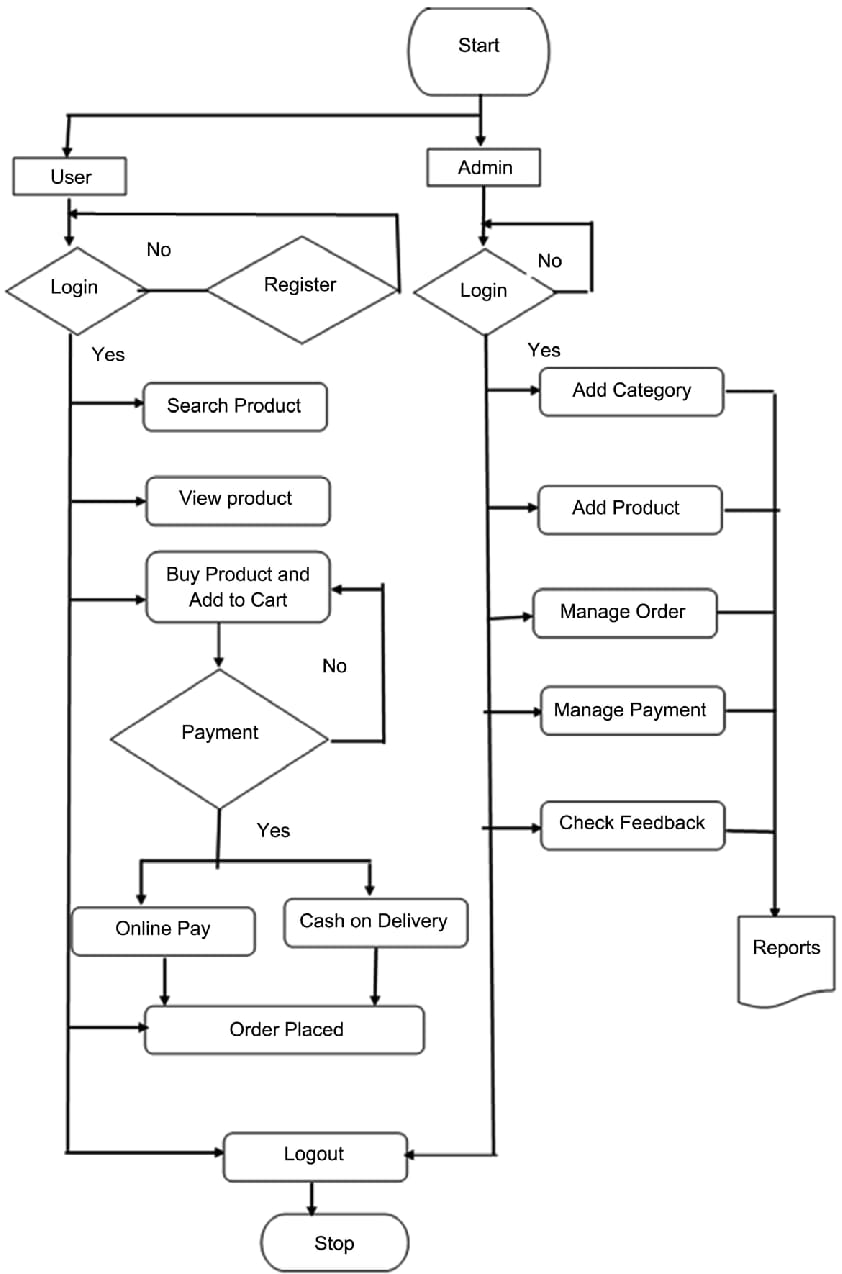
- Limited Customization: Generalized models may not fully cater to the specific requirements of different businesses or datasets.

This project addresses these gaps by utilizing Python-based data analysis techniques to perform transparent and interpretable shopping trend analysis. By focusing on a specific dataset, it generates actionable insights tailored to customer demographics, seasonal variations, and promotional impacts. The findings aim to aid businesses in making data-driven decisions and optimizing their strategies for better customer engagement and profitability.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design:**



1. Start: The process begins.

2. Gather Data: This is the initial step where relevant data is collected from various sources. These sources could include:

Customer databases: Information about customer demographics, purchase history, and browsing behavior.

Sales transactions: Data on product sales, prices, and promotions.

Market research data: External data on market trends, competitor activities, and economic indicators.

Social media: Insights from customer interactions and social media trends.

3. Clean Data: The collected data often contains inconsistencies, missing values, and outliers. This step involves:

Data cleaning: Handling missing values (imputation or removal), correcting errors, and removing duplicates.

Data transformation: Converting data into a suitable format for analysis (e.g., normalization, scaling).

4. Explore Data: This step involves an in-depth examination of the cleaned data to understand its characteristics and identify initial patterns. Common techniques include:

Summary statistics: Calculating measures like mean, median, standard deviation, and percentiles.

Data visualization: Creating charts and graphs (histograms, scatter plots, box plots) to visualize data distributions and relationships.

5. Identify Trends: This is a crucial step where statistical and machine learning techniques are applied to uncover patterns and trends in the data. These could include:

Trend analysis: Identifying upward or downward trends over time.

Seasonality analysis: Detecting seasonal patterns (e.g., increased sales during holidays).

Customer segmentation: Grouping customers based on their purchasing behavior.

Predictive modeling: Building models to forecast future sales or customer behavior.

6. Visualize Trends: The identified trends are presented visually using charts and graphs to effectively communicate insights to stakeholders. Common visualization techniques include:

Line charts: To show trends over time.

Bar charts: To compare different categories or groups.

Heatmaps: To visualize relationships between variables.

Interactive dashboards: To allow users to explore data interactively.

7. Analyze Results: The visualized trends are carefully analyzed to draw meaningful conclusions and insights. This involves:

Interpreting trends: Understanding the implications of the identified trends for business decisions.

Identifying key drivers: Determining the factors that contribute most to the observed trends.

Formulating hypotheses: Developing hypotheses based on the analysis.

8. Report Findings: The findings of the analysis are communicated to stakeholders through reports and presentations. These reports typically include:

Summary of findings: Key insights and conclusions from the analysis.

Visualizations: Charts and graphs to illustrate the findings.

Recommendations: Actionable recommendations based on the analysis.

9. End: The analysis process concludes, and the insights gained can be used to inform business decisions and strategies.

This flowchart provides a structured approach to conducting shopping trends analysis, ensuring that all key steps are considered and executed in a logical order. The specific techniques and tools used at each step may vary depending on the nature of the data, the objectives of the analysis, and the resources available.

* 1. **Requirement Specification:**

Mention the tools and technologies required to implement the solution.

* + 1. **Hardware Requirements:**

1. Processor: Intel Core i5 or higher (Recommended: Intel Core i7 for faster processing)

2. RAM: Minimum 8 GB (Recommended: 16 GB for handling large datasets)

3. Storage: Minimum 500 GB hard drive (Recommended: SSD for faster data access)

4. Graphics Card: Not required unless using machine learning models that need GPU acceleration.

5. Internet Connection: Stable internet for downloading datasets and libraries

* + 1. **Software Requirements:**

1. Operating System: Windows, macOS, or Linux (Python is cross-platform).

2. Python Version: Python 3.x (Recommended: 3.7 or later).

3. Libraries/Tools:

- Pandas: For data manipulation and analysis.

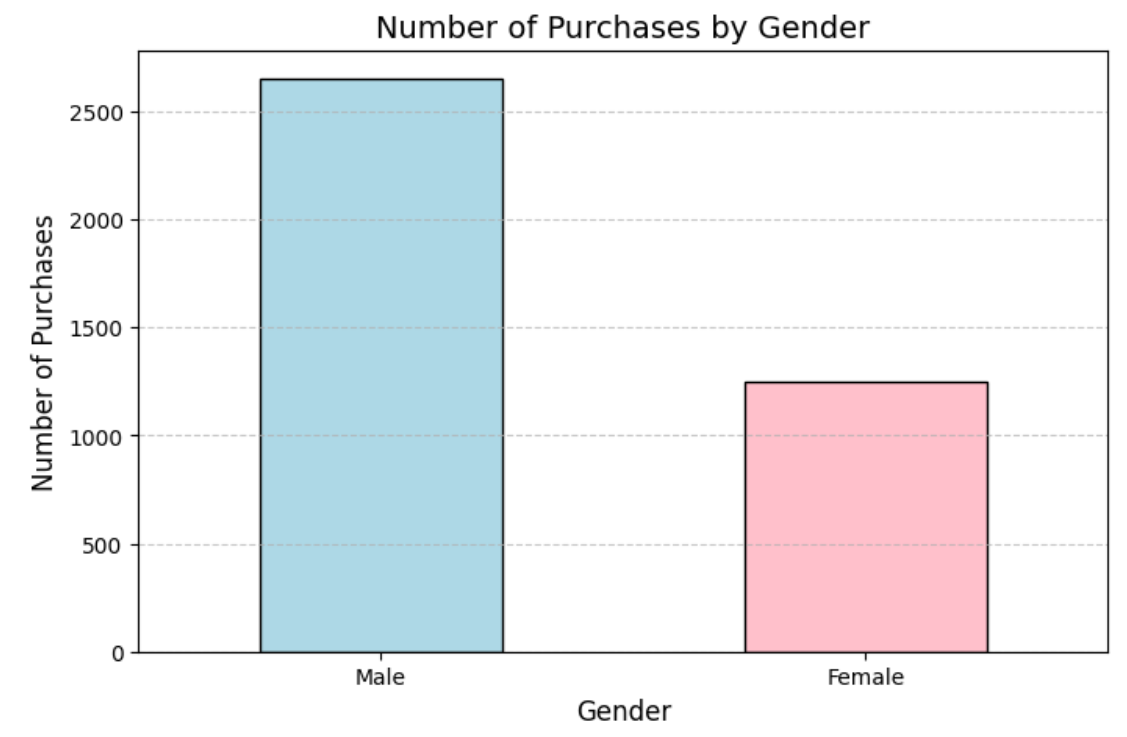
- NumPy: For numerical computations.

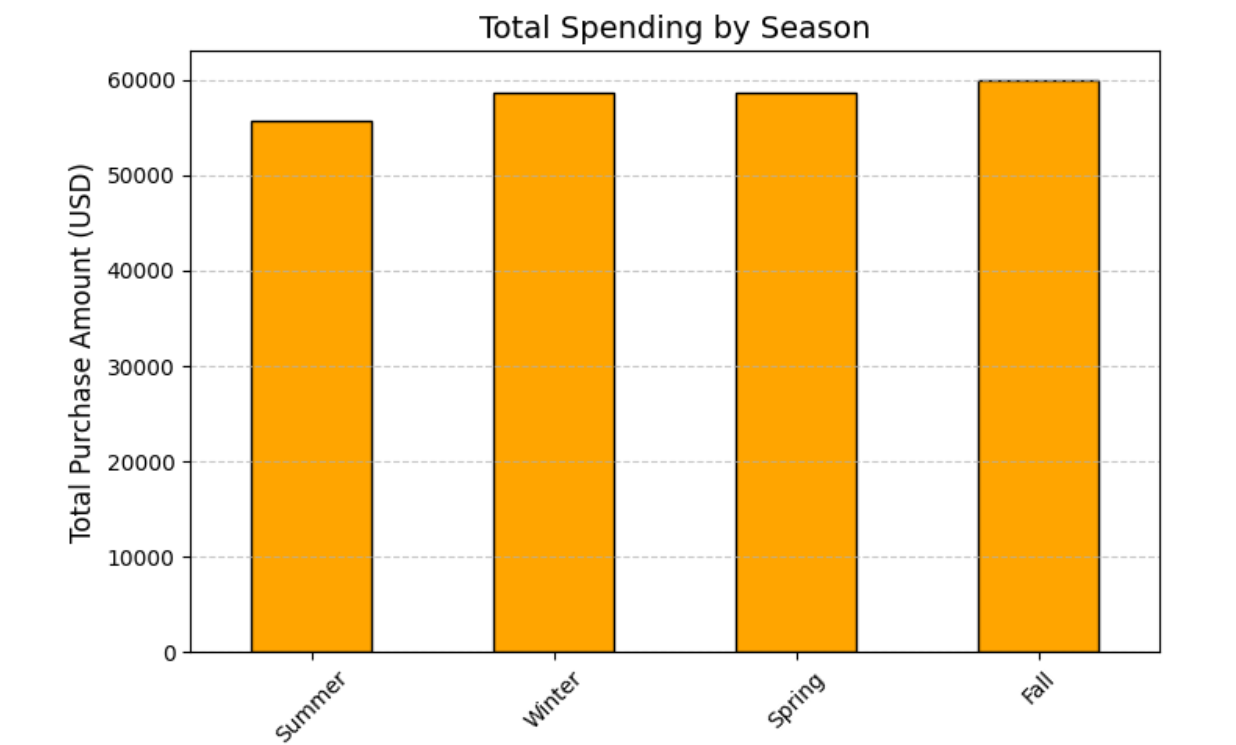
- Matplotlib and Seaborn: For data visualization.

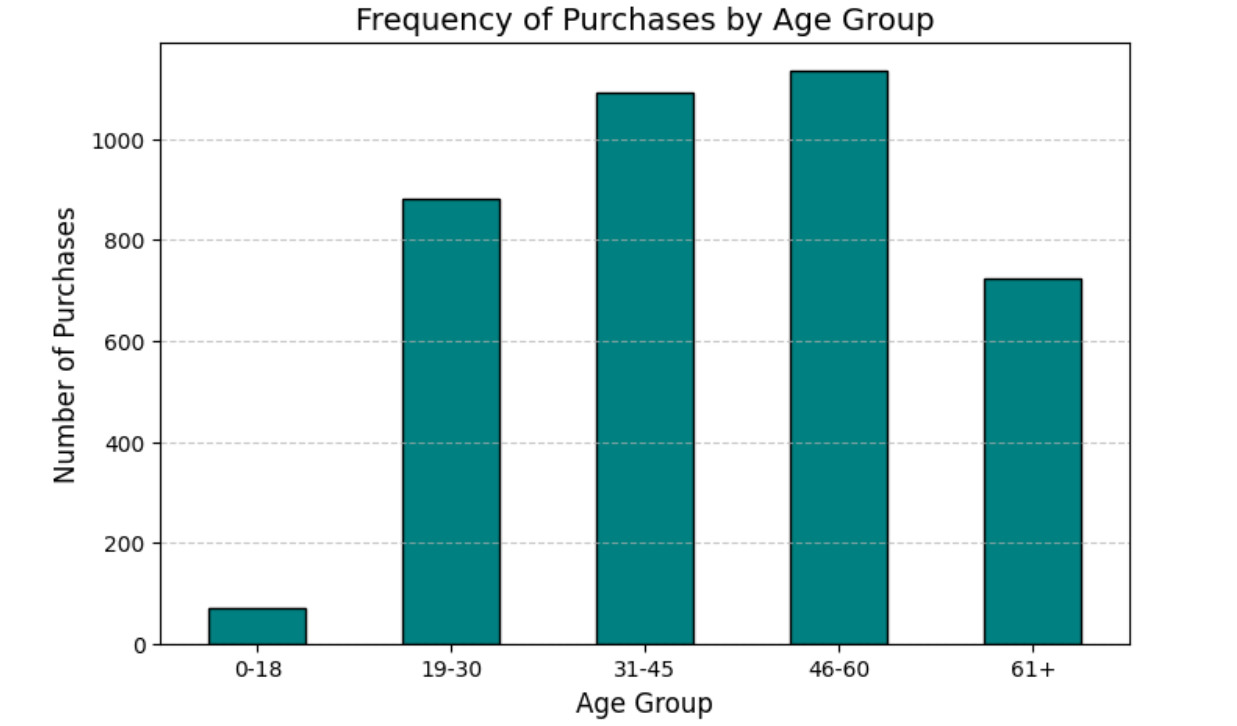
**CHAPTER 4**

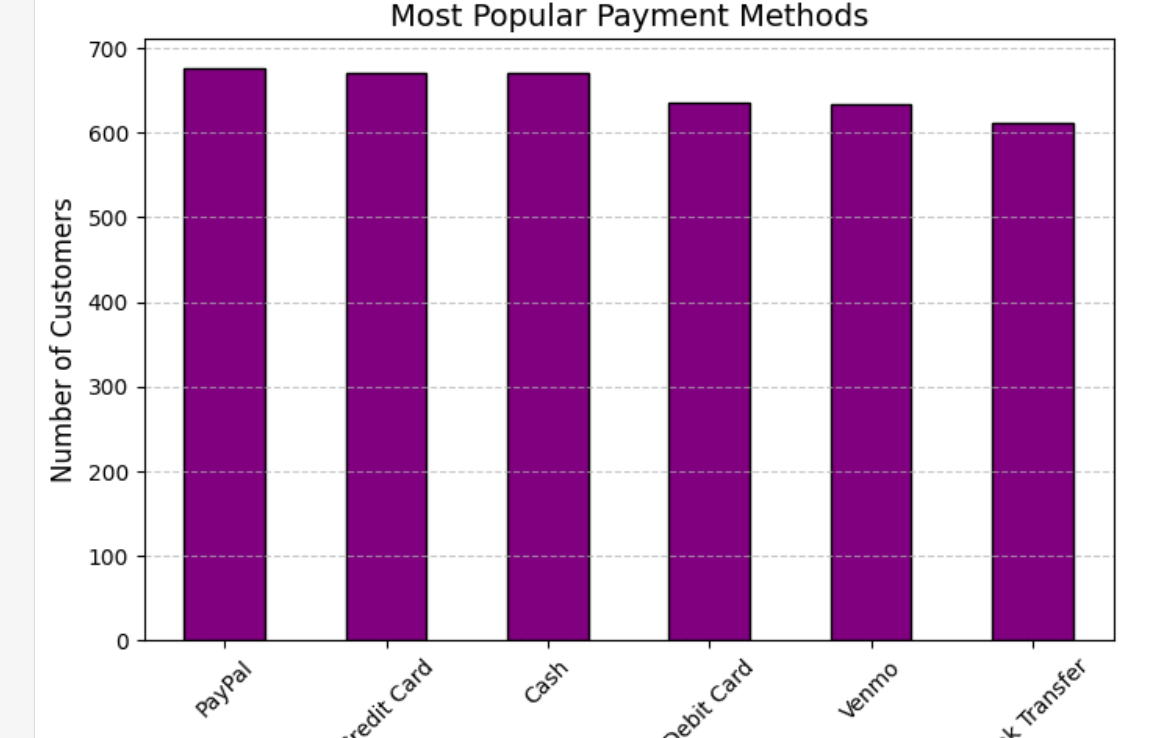
**Implementation and Result**

* 1. **Snap Shots of Result:**









* 1. **GitHub Link for Code:**

[GitHub](https://github.com/Shivraj1019/Shopping-Trends-Analysis-Using-Python) :

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

1. \*Predictive Analytics\*:

- Implement machine learning models like \*ARIMA\* or \*LSTM\* to predict future shopping trends based on historical data. This can help businesses forecast product demand and optimize inventory.

2. \*Sentiment Analysis\*:

- Integrate \*Natural Language Processing (NLP)\* to analyze customer reviews or social media posts, identifying consumer sentiments. This can enhance trend analysis by capturing the mood and preferences of shoppers.

3. \*Real-time Trend Analysis\*:

- Develop a real-time system to track live shopping trends using data scraping or APIs from major e-commerce platforms (like Amazon). This would allow businesses to quickly adapt to changing market demands.

4. \*Personalized Recommendations\*:

- Use machine learning to build a recommendation system that suggests products based on customer shopping behavior and trends, offering personalized shopping experiences.

5. \*Market Basket Analysis\*:

- Implement \*market basket analysis\* to discover which products are often bought together. This can help in understanding consumer purchasing patterns and improve product bundling strategies.

* 1. **Conclusion:**

This project highlights the use of Python in analyzing shopping trends, offering insights into consumer behavior through data manipulation and visualization techniques. By employing libraries like Pandas and Matplotlib, we were able to identify key shopping patterns. The project can be extended to incorporate predictive analytics and sentiment analysis to improve trend forecasting. Additionally, the implementation of personalized recommendations and real-time trend tracking can further enhance its relevance. Overall, this analysis provides valuable tools for businesses to optimize their strategies and stay ahead of market changes.

**REFERENCES**

1. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.
2. Géron, A., “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems”, O'Reilly Media, 2019.
3. Zhang, X., & Wang, L., “E-Commerce Trend Analysis Using Big Data Analytics”, International Journal of Computer Science and Information Security, Vol. 18, No. 3, pp. 91-102, 2020.
4. Wang, Y., & Chen, Y., “Predicting Shopping Behavior from Online Data Using Machine Learning”, Journal of Retailing and Consumer Services, Vol. 45, pp. 45-53, 2018.
5. Pandas Documentation, “Pandas Documentation”, Available: [https://pandas.pydata.org/pandas-docs/stable/](https://pandas.pydata.org/pandas-docs/stable/), Accessed on: Jan. 8, 2025.
6. Towards Data Science, “Towards Data Science Blog”, Available: [https://towardsdatascience.com/](https://towardsdatascience.com/), Accessed on: Jan. 8, 2025.