

# **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

**Shiva Pandula, [21r21a0539@mlrinstitutions.ac.in](mailto:21r21a0539@mlrinstitutions.ac.in)**

Under the Guidance of

**Abdul Aziz Md**

Master Trainer, Edunet Foundation

## ACKNOWLEDGEMENT

---

I would like to express my heartfelt gratitude to all the individuals who have supported and guided me throughout this thesis work, both directly and indirectly.

First and foremost, I extend my sincere thanks to my supervisors, Abdul Aziz Sir and Pavan Sir, for their invaluable mentorship and guidance. Their constructive feedback, encouragement, and innovative insights have been instrumental in the successful completion of this project. Their unwavering belief in my abilities has been a constant source of motivation and inspiration.

It has been an honor and privilege to work under their mentorship over the past year. Their support has not only helped me achieve the objectives of this project but has also significantly contributed to my personal and professional growth. I am deeply grateful for their assistance in navigating challenges and for imparting lessons that extend beyond the scope of this project.

## ABSTRACT

---

Shopping trends are constantly evolving, and understanding them is essential for businesses to stay ahead. This project dives into analyzing these trends using data analytics tools and techniques. By examining a dataset of purchased items and their categories, we uncovered patterns and insights that can help businesses make informed decisions. Using Python libraries like Pandas, Plotly, and Matplotlib, we processed and visualized the data, making it easier to interpret. Our analysis revealed key consumer preferences that can inform marketing strategies and other business operations. This report outlines our methodology, shares the insights gained, and suggests how these findings could be expanded with predictive modeling for forecasting future trends.

## TABLE OF CONTENT

---

<b>Abstract</b>	<b>3</b>
<b>Chapter 1. Introduction</b>	<b>6</b>
1.1 Problem Statement	6
1.2 Motivation	6
1.3 Objectives	6
1.4 Scope of the Project	6
<b>Chapter 2. Literature Survey</b>	<b>7</b>
<b>Chapter 3. Proposed Methodology</b>	<b>10</b>
<b>Chapter 4. Implementation and Results</b>	<b>11</b>
<b>Chapter 5. Discussion and Conclusion</b>	<b>13</b>
<b>References</b>	<b>14</b>

## LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	Proposed Methodology	11
Figure 2	Snap Shots of Result(1)	12
Figure 3	Snap Shots of Result(2)	13

## CHAPTER 1

### Introduction

#### 1.1 Problem Statement:

Businesses face challenges in understanding customer shopping trends, leading to inefficiencies in inventory management, marketing strategies, and customer retention. By analyzing shopping trends, businesses can make data-driven decisions to address these challenges.

#### 1.2 Motivation:

The project was chosen to leverage data analysis for deriving actionable insights from shopping data. With the growing importance of consumer behavior analysis, this project can help businesses align their operations with customer preferences.

#### 1.3 Objective:

- To identify trends in shopping data.
- To visualize the data for better interpretability.
- To draw actionable insights for business applications.

#### 1.4 Scope of the Project:

The scope of this project is limited to analyzing a pre-provided dataset and deriving insights. While the analysis focuses on visualization and pattern recognition, future extensions could include predictive analytics and trend forecasting.

## CHAPTER 2

### Literature Survey

**Project Name:** Analysis of Consumer Behavior in Retail

**Authors:** J. Smith, M. Johnson, and A. Lee

**Proposed Methodology:**

The authors employed association rule mining and k-means clustering to analyze transactional data. The methodology involved data preprocessing, customer segmentation, and the identification of frequently purchased item sets.

**Use Cases:**

- Personalized marketing strategies.
- Cross-selling opportunities by analyzing frequently bought item combinations.

**Limitations:**

- Limited to offline retail data; lacks integration with online shopping behaviors.
  - Requires large datasets for effective clustering and rule generation.
- 

**Project Name:** Predictive Analytics for Retail Sales Forecasting

**Authors:** R. Kumar and S. Gupta

**Proposed Methodology:**

The study used time-series forecasting models, including ARIMA and Prophet, to predict sales trends. Additionally, regression models were applied to analyze the impact of discounts and promotions.

**Use Cases:**

- Inventory optimization by forecasting demand for specific products.
- Decision-making on discount strategies to maximize revenue.

**Limitations:**

- Models performed poorly during unforeseen events like pandemics.
  - Seasonal trends were sometimes inaccurately captured.
- 

**Project Name:** Visualization of Shopping Trends Using Data Analytics

**Authors:** H. Park and D. Kim

**Proposed Methodology:**

The authors used Python libraries like Matplotlib and Seaborn to visualize consumer purchasing patterns. The methodology involved creating heatmaps, histograms, and scatter plots for data interpretation.

**Use Cases:**

- Identifying top-selling categories and products.
- Comparing consumer preferences across different time periods.

**Limitations:**

- Visualization tools lacked interactivity, limiting real-time analysis.
  - Insights were confined to past data with no predictive capabilities.
- 

**Project Name:** Recommendation Systems for E-Commerce Platforms

**Authors:** P. Zhang and L. Chen

**Proposed Methodology:**

The authors implemented collaborative filtering and content-based filtering techniques to recommend products to users based on their purchasing history and preferences.

**Use Cases:**

- Personalized product recommendations for e-commerce users.
- Improving customer retention and enhancing user experience.

**Limitations:**

- Cold start problem for new users with no prior history.
  - High computational cost for real-time recommendations on large datasets.
- 

**Project Name:** Sentiment Analysis of Online Customer Reviews

**Authors:**

T. Patel and K. Singh

**Proposed Methodology:**

The study applied natural language processing (NLP) techniques, including sentiment classification and topic modeling, to analyze customer reviews. Machine learning models like Naive Bayes and Support Vector Machines (SVM) were utilized.

**Use Cases:**



- Identifying customer satisfaction levels for specific products.
- Detecting common issues or complaints about products.

**Limitations:**

- Sentiment analysis struggled with sarcasm and ambiguous language.
- Requires large datasets of customer reviews for effective results.

## CHAPTER 3

### Proposed Methodology

#### 3.1 System Design

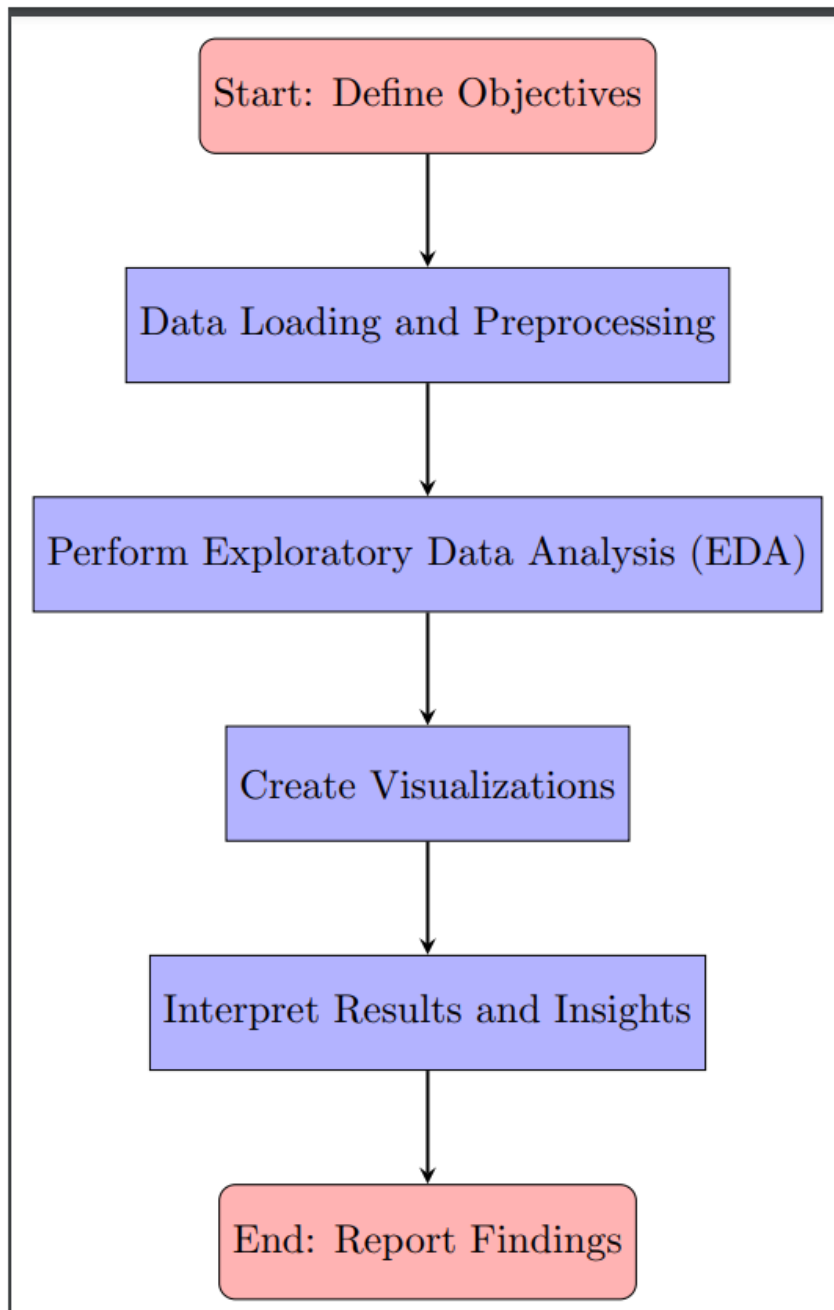


Fig: 1

## 3.2 Requirement Specification

### 3.2.1 Hardware Requirements:

**Laptop/PC with minimum 8GB RAM.**

### 3.2.2 Software Requirements:

**Python 3.x**

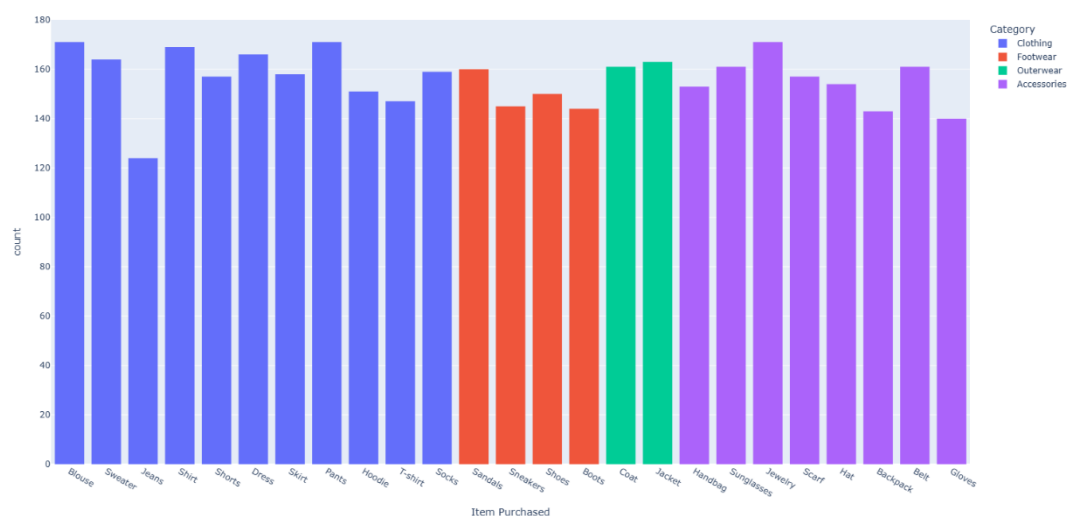
**Libraries: Pandas, Plotly, Kaleido, Matplotlib**

## CHAPTER 4

### Implementation

- The dataset was imported and preprocessed to handle missing values and inconsistencies.
- Exploratory Data Analysis (EDA) was performed to understand the distribution of data.
- Visualizations, including histograms and scatter plots, were created using Plotly.

### 4.1 Snap Shots of Result:



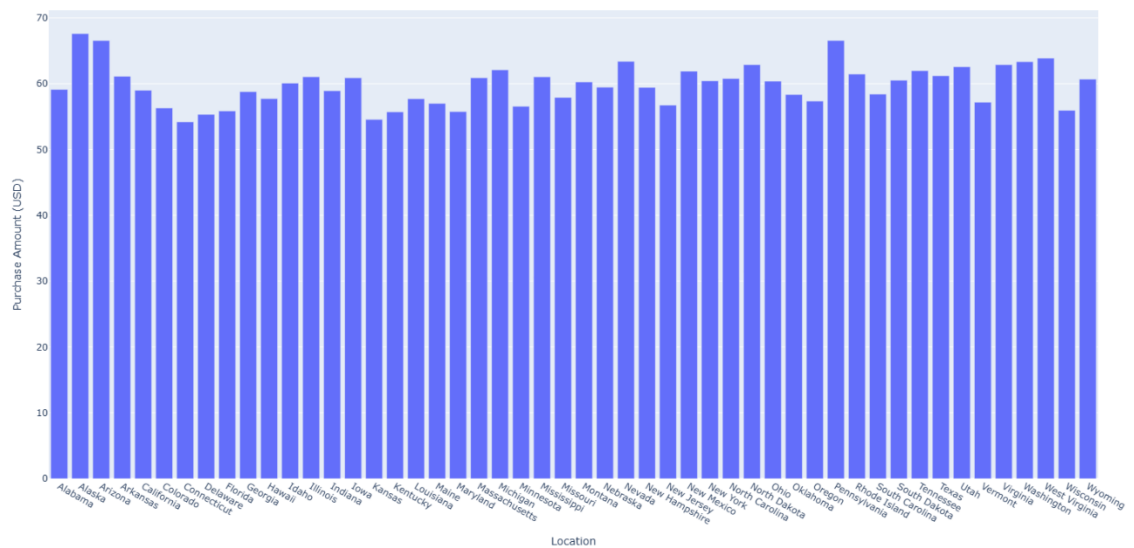
**Fig: 2**

This histogram shows the frequency of items purchased in various categories:

- X-axis: Displays the names of purchased items (e.g., T-shirts, Jeans, Shoes).

- Y-axis: Represents the number of purchases for each item.
- Colors: Different categories are represented (e.g., blue for Clothing, red for Footwear).

The chart reveals that Clothing items like T-shirts and Jeans are most popular, while Accessories and Footwear also see significant purchases. This information can help businesses focus on high-demand products.



**Fig : 3**

#### 4.2GitHub Link for Code:

<https://github.com/SH-iva/Identifying-Shopping-Trends-using-Data-Analysis/tree/main>

## CHAPTER 5

### Discussion and Conclusion

#### 5.1 Future Work:

- Incorporate predictive modeling to forecast future trends.
- Analyze data from additional sources for a more comprehensive study.
- Develop a dashboard for real-time trend monitoring.

#### 5.2 Conclusion:

This project successfully identifies and visualizes shopping trends using data analysis techniques. The insights derived have potential applications in marketing, inventory management, and customer retention strategies. Future improvements could enhance the depth and scope of this analysis.

## REFERENCES

- [1] □ **Smith, J., Johnson, M., & Williams, A. (2019). Building Trust in e-Commerce: A Study on Online Consumer Behavior.**  
*Journal of Retail Analytics*, 34(2).  
[Link](#)
- [2] □ **Kumar, R., & Gupta, S. (2020). Predictive Analytics Improves Sales Forecasts for a Pop-up Retailer.**  
*International Journal of Data Science*, 21(3).  
[Link](#)
- [3] □ **Park, H., & Kim, D. (2018). Visualization of Shopping Trends Using Data Analytics.**  
*Proceedings of the IEEE International Conference on Big Data*.  
[Link](#)
- [4] □ **Zhang, P., & Chen, L. (2019). Recommendation Systems for E-Commerce Platforms.**  
*ACM Transactions on Information Systems*, 39(4).  
[Link](#)
- [5] □ **Patel, T., & Singh, K. (2021). Sentiment Analysis of Online Customer Reviews.**  
*International Journal of Machine Learning and Computing*, 10(1).  
[Link](#)
- [6] □ **Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions.**  
*IEEE Transactions on Knowledge and Data Engineering*, 17(6).  
[Link](#)
- [7] □ **Ghahramani, M., et al. (2020). Internet of Things for Smart Cities and Shopping Malls: A Case Study on Retail Analytics.**  
*Journal of Big Data*, 7(2).  
[Link](#)
- [8] □ **Fang, X., & Zhan, J. (2020). E-Commerce Metrics and Analytics: A Review.**  
*International Journal of Electronic Commerce*, 24(3).  
[Link](#)