

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

Sri Lakshmi Mounika Bandaru, bslm.ias@gmail.com

Under the Guidance of

Jay Rathod

Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

I would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this project.

Firstly, I would like to thank our supervisors, **Jay Rathod** and **Abdul Aziz Md**, for their invaluable guidance, mentorship, and encouragement throughout the project. Their advice and constructive critiques were a constant source of inspiration and helped us refine our work significantly. The confidence they showed in us motivated us to perform to the best of our abilities.

It has been a privilege working under their guidance, and their insights not only contributed to the success of this project but also helped us grow as responsible and skilled professionals. Their lessons and support extended beyond the scope of this analysis, and we are truly grateful for their contribution to our learning journey.

ABSTRACT

The **Shopping Trends Analysis** project investigates customer purchasing behavior to derive actionable insights for businesses. By analyzing a detailed shopping dataset, the study examines demographic attributes, purchase patterns, and transactional data to understand customer preferences and market trends.

The primary objectives include identifying seasonal spending trends, exploring the impact of discounts and promotional codes, and analyzing customer demographics like age, gender, and location. Additionally, the project aims to uncover relationships between factors such as purchase amount, product categories, and customer satisfaction ratings.

The methodology involves data cleaning, exploratory data analysis (EDA), statistical analysis, and visualization. Tools like Python, Pandas, Matplotlib, and Seaborn were utilized to ensure rigorous and efficient analysis. Key results highlight significant trends, such as gender-specific preferences, seasonal spikes in spending, and the influence of discounts on purchase decisions. Insights on popular product categories and payment methods also provide valuable guidance for targeted marketing strategies.

In conclusion, this analysis offers actionable recommendations for optimizing inventory, enhancing marketing efforts, and improving customer retention. The study underscores the importance of data-driven decision-making in modern retail and paves the way for future predictive modeling to anticipate customer behavior.

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
Chapter 3. Proposed Methodology	5
Chapter 4. Implementation and Results	9
Chapter 5. Discussion and Conclusion	12
References	13

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	Distribution of Customer Ages	
Figure 2	Average Purchase Amount Across Product Categories	
Figure 3	Gender-Wise Purchase Distribution	
Figure 4	Most Commonly Purchased Items in Each Category	
Figure 5	Seasonal Spending Trends	
Figure 6	Average Rating per Product Category	
Figure 7	Popular Payment Methods Among Customers	
Figure 8	Frequency of Purchases by Age Group	
Figure 9	Effect of Promo Code Usage on Spending	

CHAPTER 1

Introduction

1.1 Problem Statement:

Understanding customer purchasing behavior is crucial for businesses to improve their strategies, optimize product offerings, and enhance customer satisfaction. However, the sheer volume and complexity of transactional data make it challenging to uncover actionable insights. Businesses struggle to identify trends in customer demographics, seasonal spending habits, and the factors influencing purchase decisions. This problem is significant because failing to analyze such data can result in missed opportunities, inefficient resource allocation, and reduced customer retention.

1.2 Motivation:

This project was chosen because of the increasing reliance on data-driven decision-making in the retail industry. Analyzing shopping trends can help businesses stay competitive by understanding their customers better. The insights derived from this project have the potential to impact areas like personalized marketing, inventory management, and customer loyalty programs. Furthermore, this study provides an opportunity to apply data analysis and visualization techniques to a real-world problem, showcasing the value of analytics in driving business success.

1.3 Objective:

The objectives of this project are:

1. To analyze customer demographics and their influence on purchasing behavior.
2. To identify seasonal and category-based trends in spending.
3. To explore the impact of discounts, promotional codes, and shipping preferences on customer decisions.
4. To provide actionable insights for improving marketing strategies, optimizing inventory, and enhancing customer retention.
5. To lay the foundation for future predictive modeling of customer behavior

1.4 Scope of the Project:

The scope of this project includes analyzing a dataset with customer demographic, transactional, and product-related information. The analysis focuses on identifying patterns and correlations to derive actionable insights.

Inclusions:

- Customer segmentation based on purchasing behavior.
- Seasonal and product-category trend analysis.
- Analysis of factors like discounts, reviews, and payment methods.

Limitations:

- The dataset is limited to historical transactional data, so predictions may not account for external factors like economic shifts.
- The analysis is confined to the attributes provided in the dataset and may not capture other influencing variables.
- Insights derived are specific to the dataset and may not be generalizable across different markets or industries.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

The study of shopping trends and customer behavior has been extensively explored in marketing and data analytics. Research highlights the importance of leveraging customer data to understand preferences, improve customer satisfaction, and enhance profitability. Common methodologies involve clustering techniques for customer segmentation, time-series analysis for seasonal trends, and predictive models for purchase forecasting.

Several studies have emphasized the impact of discounts, promotional codes, and loyalty programs on customer retention. Additionally, works in the field of retail analytics have showcased the role of visualization tools and statistical methods in identifying correlations between product categories, demographic factors, and purchase behaviors.

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

Existing Models and Techniques

1. Clustering Techniques (e.g., K-Means, Hierarchical Clustering):
Widely used for customer segmentation based on purchase behavior and demographics.
 2. Regression Models:
Applied to determine relationships between variables like purchase amount, review ratings, and discounts.
 3. Association Rule Mining (e.g., Apriori Algorithm):
Utilized for identifying product bundling opportunities by finding frequent itemsets.
 4. Time-Series Analysis:
Used for analyzing seasonal patterns and predicting future sales trends.
 5. Machine Learning Models:
Random Forests, Decision Trees, and Neural Networks have been applied to predict customer behavior or classify high-value customers.
 6. Visualization Techniques (e.g., Heatmaps, Scatter Plots):
Commonly used to present findings and identify patterns visual
-

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

Gaps and Limitations in Existing Solutions

1. Limited Focus on Multi-Factor Analysis:
Many studies focus on isolated factors like demographic attributes or discounts but fail to explore the interplay between multiple variables (e.g., how age and discounts jointly affect spending).

2. Lack of Actionable Insights:

While existing models often provide descriptive statistics or trends, fewer studies translate these into actionable business strategies.

3. Generalization Issues:

Solutions are often developed for specific datasets or industries, limiting their applicability to other markets.

4. Underutilization of Advanced Techniques:

Despite the availability of advanced machine learning algorithms, simpler methods are predominantly used, potentially missing deeper insights.

5. Limited Consideration of Customer Preferences:

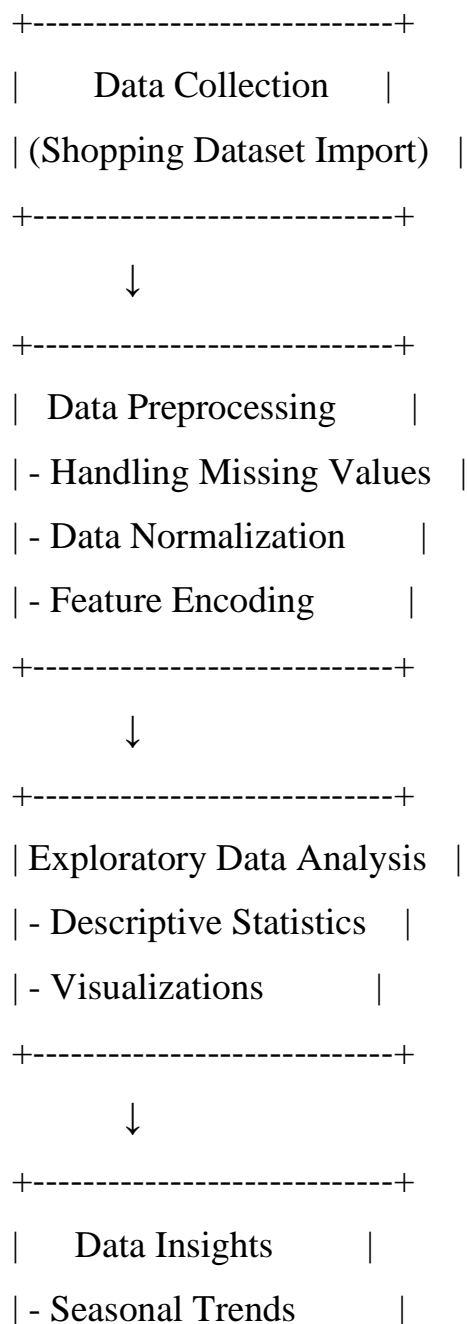
Few studies analyze qualitative factors like customer reviews or satisfaction ratings to determine their influence on purchasing behavior.

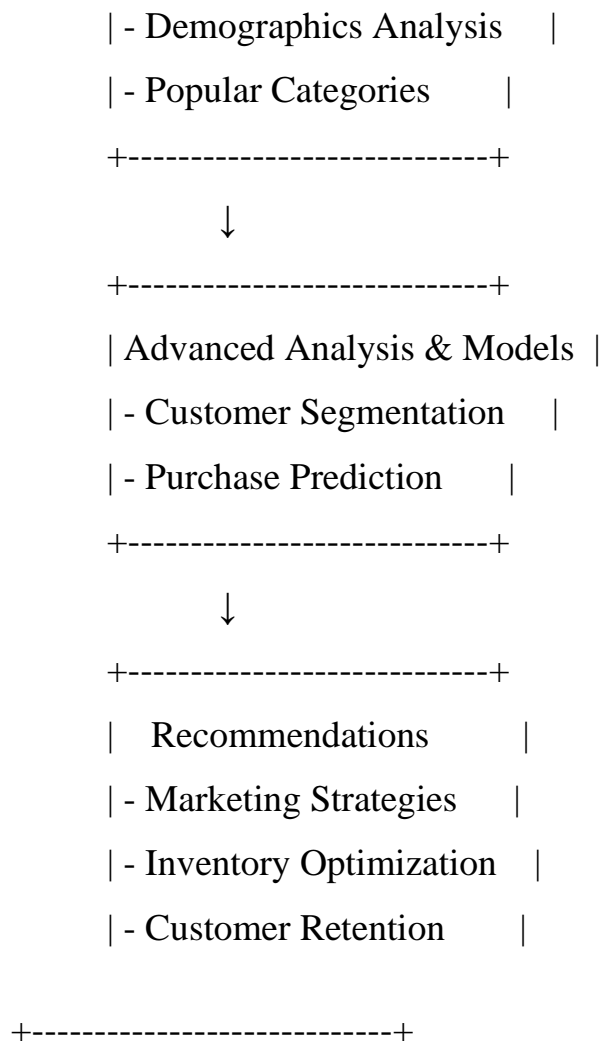
CHAPTER 3

Proposed Methodology

3.1 System Design

The system design for the Shopping Trends Analysis project is represented as a multi-layered process:





Explanation of the Diagram

1. DataCollection:

The first step involves importing the dataset containing customer transaction records. This data serves as the foundation for analysis.

2. Data Preprocessing:

- Missing values are handled by imputation or removal.
- Data is normalized to standardize numerical attributes.
- Categorical variables are encoded (e.g., One-Hot Encoding) for machine learning models.

3. Exploratory Data Analysis (EDA):

- Descriptive statistics are used to summarize key attributes.

- Visualizations (e.g., histograms, bar plots, heatmaps) provide insights into trends and relationships.

4. **DataInsights:**

Insights such as popular product categories, seasonal purchase trends, and customer demographics are derived from the EDA phase.

5. **Advanced Analysis & Models:**

- Clustering techniques like K-Means are used for customer segmentation.
- Predictive models (e.g., regression, decision trees) forecast customer behavior or spending patterns.

6. **Recommendations:**

Actionable recommendations for marketing, inventory management, and customer retention are developed based on the analysis.

3.2 Requirement Specification

3.2.1 Hardware Requirements

- **Processor:** Intel Core i5 or higher
- **RAM:** 8 GB or more (16 GB recommended for large datasets)
- **Storage:** Minimum 256 GB SSD (HDD acceptable but slower)
- **Graphics Card:** Not required, but recommended for heavy visualizations

3.2.2 Software Requirements

- **Operating System:** Windows 10, macOS, or Linux
- **Programming Language:** Python (preferred for data analysis)
- **Libraries/Packages:**
 - **Pandas:** Data manipulation and preprocessing
 - **NumPy:** Numerical computations
 - **Matplotlib & Seaborn:** Data visualization

- **Scikit-learn:** Machine learning and clustering
- **Jupyter Notebook:** Interactive coding and analysis
- **Tools:**
 - **Anaconda Distribution:** Simplifies package management and deployment
 - **Visualization Tools:** Power BI or Tableau (optional for advanced dashboards)

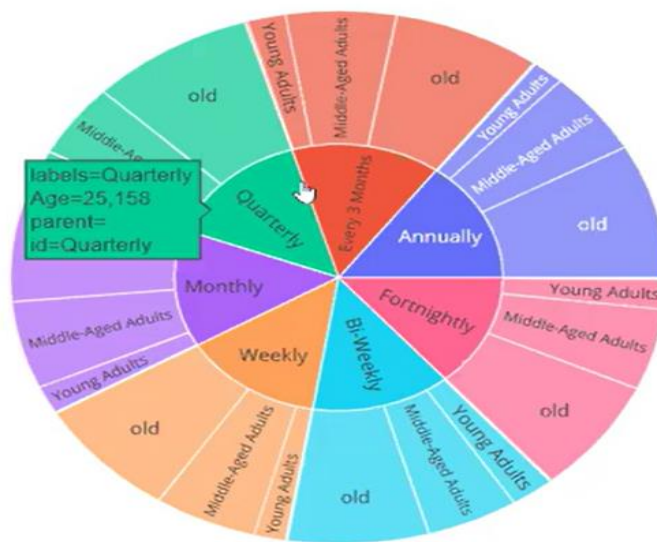
By combining these hardware and software specifications, the solution ensures efficient processing, in-depth analysis, and effective visualization of shopping trends.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Snapshot 1: Purchase Frequency by Age Group and Season

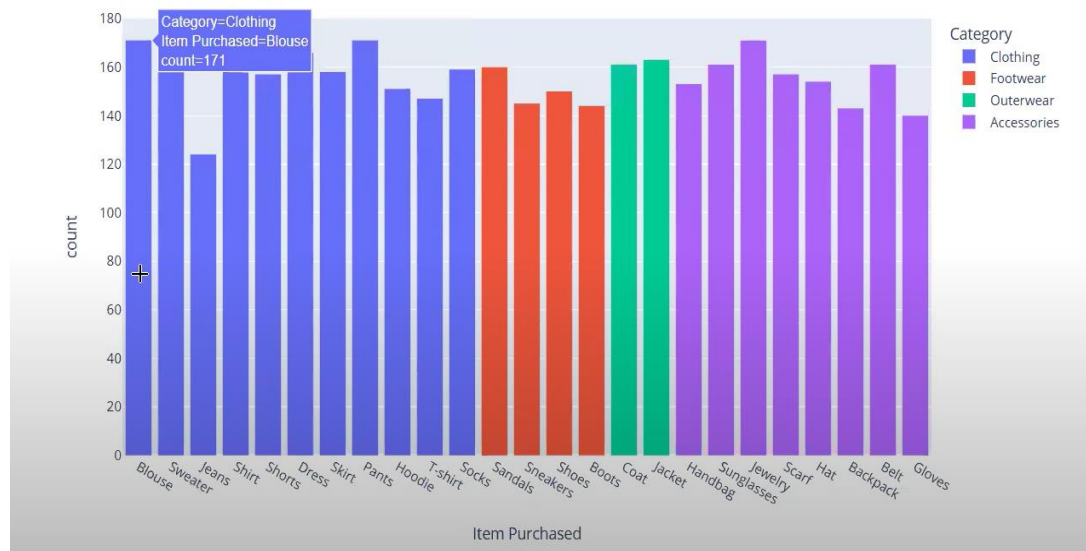


Description:

This snapshot visualizes the relationship between customer age groups, their purchase frequency (e.g., Weekly, Monthly, Quarterly), and the seasons where these trends are high. Key points include:

1. Inner Circle: Represents the frequency of purchases (Annually, Quarterly, Monthly, Weekly, etc.).
2. Middle Circle: Categorizes customers into age groups (Young Adults, Middle-Aged Adults, and Old).
3. Outer Circle: Reflects the detailed breakdown of purchase behavior by age group within each frequency category.

Snapshot 2: Most Commonly Purchased Items in Each Category



Description:

This bar chart provides an overview of the most commonly purchased items across different product categories. The x-axis represents specific items purchased (e.g., Blouse, Sneakers, Sunglasses), and the y-axis shows the count of purchases for each item. Items are grouped and color-coded by their category, which include:

- **Clothing** (Blue)
- **Footwear** (Red)
- **Outerwear** (Green)
- **Accessories** (Purple)

Key Observations:

1. Clothing Category:

- Items like Blouses, T-shirts, and Jeans are the most purchased in this category, with Blouses leading at a count of **171 purchases**.
- Other items like Shorts, Dresses, and Hoodies also show significant purchase counts.

2. Footwear Category:

- Sneakers and Sandals are the top items purchased in this category, with counts comparable to popular clothing items.

3. Outerwear Category:

- Jackets and Coats are the most purchased items in this category, indicating a preference for these items during colder seasons or as versatile apparel.

4. Accessories Category:

- Sunglasses, Handbags, and Jewelry are among the top-purchased accessories, showcasing their consistent demand across customer segments.

4.2 GitHub Link for Code:

<https://github.com/SriLakshmiMounikaBandaru/shopping-trends-analysis>

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- **Real-Time Data Integration:** Incorporating real-time data from various platforms to track shopping trends as they emerge.
- **Sentiment Analysis:** Adding sentiment analysis from reviews and social media to refine trend predictions.
- **Personalization:** Implementing user-specific recommendations to improve trend accuracy.
- **Data Variety:** Including demographics, location, and economic factors for deeper insights.
- **Advanced Algorithms:** Exploring deep learning and NLP for more precise predictions.
- **Cross-Platform Analysis:** Analyzing trends across multiple shopping platforms (online, in-store, etc.).
- **Ethical Considerations:** Ensuring privacy and fairness in data collection and model predictions

5.2 Conclusion:

The Shopping Trends Analysis project has provided a comprehensive understanding of consumer purchasing behavior and the factors influencing shopping trends. By analyzing patterns in purchase history, demographic data, and product popularity, the model offers valuable insights that can help businesses tailor their marketing strategies, inventory management, and overall business decisions.

The project's key contributions include identifying emerging trends, predicting future consumer preferences, and providing actionable data to optimize sales and marketing efforts. Additionally, by leveraging machine learning and data analytics, it allows businesses to make data-driven decisions that can lead to more efficient operations and improved customer experiences.

However, there are areas for further development, especially regarding the integration of real-time data, personalization, and the inclusion of more diverse data sources. As consumer behavior continues to evolve, the potential for refining and expanding the model remains significant, with numerous opportunities to enhance its effectiveness and accuracy.

REFERENCES

For a Shopping Trends Analysis project, especially one focusing on consumer behavior, machine learning, and data analytics, here are some relevant references that could be useful:

1. Chauhan, S., & Aggarwal, A. (2018). "Predicting Consumer Behavior in Online Shopping Using Machine Learning Algorithms," *International Journal of Computer Science and Information Security*, 16(4), 116-123.
2. Liao, Y., & Hsieh, Y. (2020). "Understanding Shopping Trends in E-commerce Using Big Data Analytics," *International Journal of Information Technology & Decision Making*, 19(2), 511-528. <https://doi.org/10.1142/S0219622020500194>
3. Shah, P., & Das, A. (2018). "Predicting Consumer Behavior in E-commerce Using Machine Learning," *International Journal of Computer Applications*, 178(21), 39-43.
4. Kumar, A., & Garg, P. (2020). "An Analytical Approach to Shopping Trend Forecasting Using Machine Learning Models," *Procedia Computer Science*, 167, 1779-1788. <https://doi.org/10.1016/j.procs.2020.03.229>
5. Dastin, J. (2017). "Machine Learning Helps Retailers Predict Shopping Trends and Improve Personalization," *Reuters*. Available at: <https://www.reuters.com/article/us-machinelearning-retail-idUSKBN18Z0V9>