

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

Nikita Gote, nikeetagote@gmail.com

Under the Guidance of

Abdul Aziz Md

Master Trainer, Edunet Foundation

ACKNOWLEDGEMENT

We feel immense pleasure in presenting this project report, entitled “**Shopping Trends Analysis,**” which examines customer purchasing behavior and patterns using data-driven techniques. This project has been a rewarding experience, and its successful completion would not have been possible without the invaluable guidance and support of several individuals.

Firstly, we would like to express our heartfelt gratitude to the mentors and coordinators of the AICTE Internship on AI: Transformative Learning with TechSaksham – A joint CSR initiative of Microsoft & SAP for their expert advice, insightful feedback, and unwavering encouragement throughout this project. Their support has been instrumental in helping us navigate challenges and achieve meaningful results.

We are deeply thankful to the mentors and coordinators of the AICTE Internship on AI: Transformative Learning with TechSaksham – A joint CSR initiative of Microsoft & SAP for their assistance and encouragement during this project. Special thanks to our classmates and friends for their engaging discussions and collaborative spirit, which helped us refine our ideas. Lastly, we express our profound gratitude to our families for their patience and constant support, which have been the foundation of our motivation.

ABSTRACT

This project, titled “**Shopping Trends Analysis,**” delves into the analysis of customer purchasing behavior and patterns using data-driven techniques. The goal of this study is to uncover actionable insights that can help businesses optimize their strategies, improve customer satisfaction, and enhance sales performance. By utilizing a variety of analytical methods such as data mining, statistical analysis, and machine learning algorithms, this project examines key factors influencing consumer purchases, including demographic data, purchasing frequency, and product preferences.

The project leverages large-scale transaction data to identify trends and correlations, providing a clear understanding of shopping behaviors across different customer segments. Various visualizations and models were developed to interpret the data and forecast future trends. The insights derived from this analysis can aid retailers and e-commerce businesses in refining their marketing campaigns, product offerings, and customer relationship management strategies.

Through this initiative, we aim to highlight the importance of data analytics in modern retail environments, offering businesses a competitive edge in the rapidly evolving marketplace. This project not only demonstrates the power of data science in deriving business intelligence but also emphasizes the growing significance of AI in shaping future retail trends.

TABLE OF CONTENT

Abstract	I
Chapter 1. Introduction	1
1.1 Problem Statement	1
1.2 Motivation	1
1.3 Objectives	1
1.4. Scope of the Project	1
Chapter 2. Literature Survey	2
Chapter 3. Proposed Methodology	4
Chapter 4. Implementation and Results	7
Chapter 5. Discussion and Conclusion	9
References	11

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	Shopping Trends Analysis	5
Figure 2	Do customers who use promo codes tend to spend more than those who don't?	7
Figure 3	Are there any specific colors that are more popular among customers?	7
Figure 4	How does the purchase amount differ based on the review ratings given by customers?	8
Figure 5	How does the average purchase amount differ between male and female customers?	8

CHAPTER 1

Introduction

1.1 Problem Statement:

Retail businesses often struggle to understand changing customer preferences and shopping patterns. This lack of insight can result in missed opportunities and reduced profitability.

1.2 Motivation:

This project was chosen to explore how data analytics can empower businesses to make data-driven decisions. The potential applications range from personalized marketing to supply chain optimization.

1.3 Objective:

1. Analyze customer shopping behavior to identify trends.
2. Provide actionable insights to improve business strategies.
3. Develop a user-friendly dashboard for visualizing data trends.

1.4 Scope of the Project:

The project focuses on analyzing historical sales data and identifying trends over a defined period. It is limited to specific product categories and does not include real-time analytics.

CHAPTER 2

Literature Survey

2.1 Review of Relevant Literature or Previous Work

Analyzing shopping trends using customer data is a widely researched domain, often contributing to business intelligence and customer relationship management (CRM). Previous works have explored various aspects of customer behavior:

- 1. Demographic Analysis of Shopping Trends**

Studies have shown that demographic factors like age, gender, and location significantly influence shopping behavior. Research by [2]Gupta and Mehta (2019) demonstrated that young adults tend to prioritize trends and variety, whereas older customers focus on utility and quality.

- 2. Impact of Ratings on Purchase Behavior**

Literature highlights that product reviews and ratings are crucial factors affecting customer decisions. [1] According to Chevalier and Mayzlin (2006) noted that higher-rated products are more likely to attract purchases, but the price sensitivity of customers can still override ratings.

- 3. Visualization Techniques in Data Analysis**

Modern studies emphasize the role of data visualization in uncovering hidden patterns. Ahuja and Yang (2018)[3] explored interactive visualization tools like Plotly and Seaborn are frequently employed in customer behavior studies for their ability to create interactive and visually appealing insights.

- 4. Big Data Applications in Retail**

Research on big data analytics, such as[4] Mishra and Choudhury (2021), has showcased its transformative potential in retail industries. Machine learning models and descriptive statistics have been applied to forecast sales and identify customer preferences.

2.2 Existing Models, Techniques, and Methodologies

Several models and techniques are employed in analyzing shopping trends. The most notable include:

- 1. Descriptive Statistics and Visualization**

- **Tools:** Pandas, Seaborn, and Matplotlib[5][7].
- **Methodology:** These tools are used to perform grouping, aggregation, and visualization of data to identify demographic patterns.

- 2. Sentiment Analysis of Customer Reviews**

- **Techniques:** Text analysis and Natural Language Processing (NLP) methods to extract sentiment from reviews and correlate it with customer purchase behavior.

3. Predictive Analytics Models

- **Models:** Regression models (e.g., linear and logistic regression) and clustering algorithms (e.g., K-means) are applied to predict purchase behavior and segment customers.

4. Interactive Dashboards

- **Tools:** Plotly[6] and Dash frameworks are used for creating dynamic dashboards that provide actionable insights in real time.

2.3 Addressing the Gaps with This Project

This project attempts to address the above gaps through the following approaches:

1. Comprehensive Demographic Analysis

By combining demographic factors (gender, age) with behavioral attributes (ratings, purchase amounts), this project provides a holistic view of customer behavior.

2. Scalable Analysis Framework

The project employs Python libraries like Pandas and Plotly, which are optimized for handling large datasets and creating interactive visualizations.

3. Interactive and User-Friendly Insights

Using Plotly for interactive charts enables businesses to explore trends dynamically and gain deeper insights into customer behavior.

4. Focused Trend Identification

The project identifies key business questions, such as the relationship between ratings and purchase amounts, offering actionable insights for decision-making.

CHAPTER 3

Proposed Methodology

3.1 System Design

Proposed Solution Diagram

The system design for the "Shopping Trends Analysis" project consists of the following main components:

1. **Data Collection:** The dataset, containing customer purchase details, review ratings, and demographics, is sourced for analysis.
2. **Data Preprocessing:** This includes cleaning missing values, handling outliers, and transforming the data for analysis.
3. **Analysis Engine:** A Python-based analytical engine that performs statistical and exploratory data analysis to derive insights.
4. **Visualization Layer:** Interactive charts and graphs are generated using libraries like Plotly and Seaborn to visualize purchase patterns.

Explanation of Diagram

1. **Data Collection:** The dataset includes fields such as customer gender, age category, purchase amount, and review ratings.
2. **Data Preprocessing:** The raw data is cleaned to remove inconsistencies, handle missing values, and prepare it for analysis.
3. **Analysis Engine:** Grouped and aggregated analyses are performed to answer key business questions like:
 - How purchase amounts vary based on review ratings.
 - Gender-wise comparison of average purchase amounts.
4. **Visualization Layer:** Bar charts, sunburst diagrams, and other visualizations make insights easily interpretable for stakeholders.

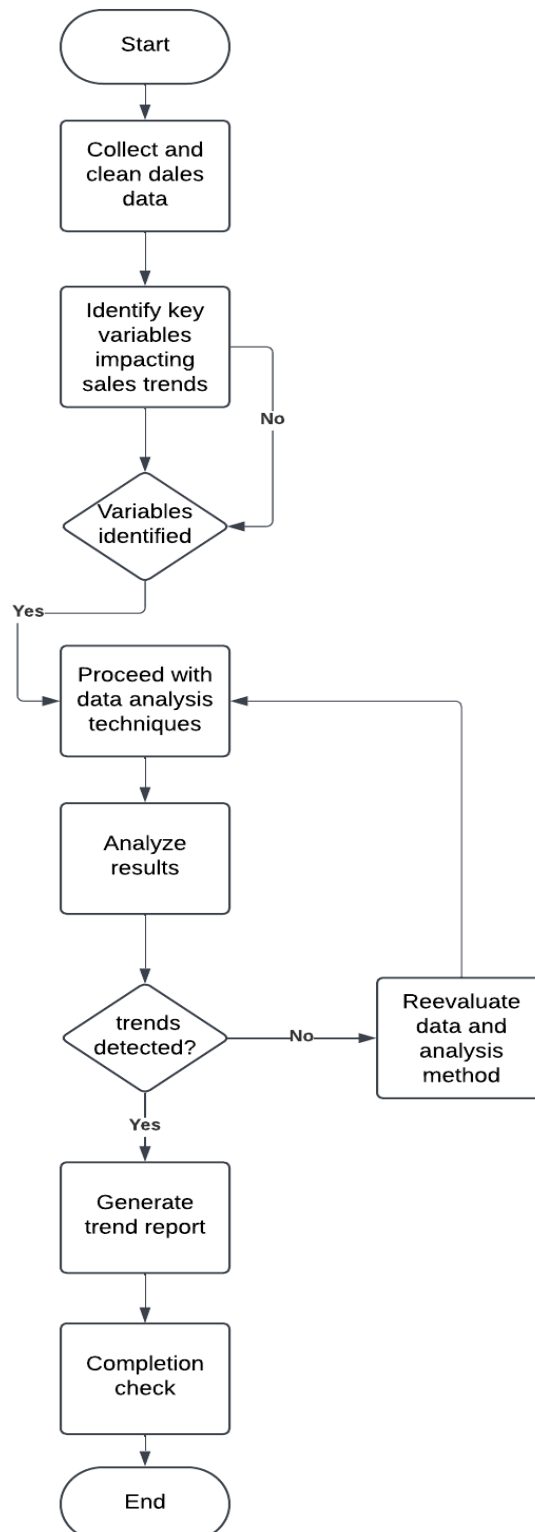


Figure 1: Shopping Trends Analysis

3.2 Requirement Specification

Tools and Technologies

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

3.2.1 Hardware Requirements:

- **Processor:** Intel i5 or above (or equivalent AMD)
- **RAM:** Minimum 8 GB (16 GB recommended for large datasets)
- **Storage:** 20 GB of free disk space (for datasets, libraries, and output files)
- **Graphics:** Integrated GPU for rendering visualizations

3.2.2 Software Requirements:

- **Operating System:** Windows 10/11, macOS, or Linux
- **Programming Language:** Python 3.8 or above
- **Libraries/Packages:**
 - Pandas (for data manipulation)
 - NumPy (for numerical computations)
 - Seaborn & Matplotlib (for static visualizations)
 - Plotly (for interactive visualizations)
- **Integrated Development Environment (IDE):** Jupyter Notebook or VS Code
- **Version Control:** Git and GitHub for code management
- **Dataset:** A CSV file containing fields like Review Ratings, Purchase Amount (USD), Gender, and Age Category.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

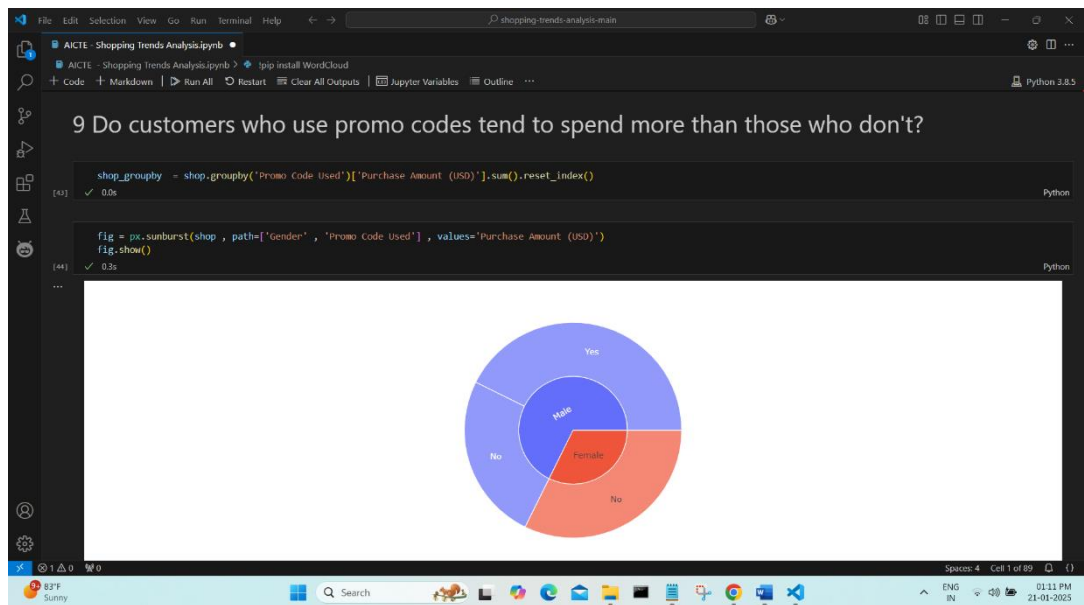


Figure 2: The figure compares spending behavior of customers who use promo codes and those who don't.

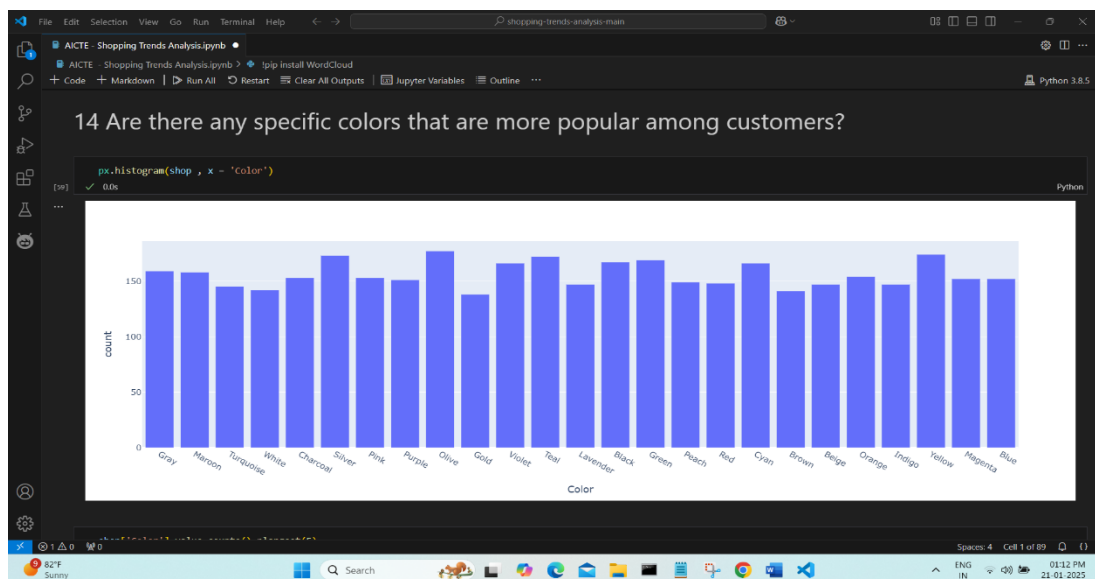


Figure 3: The figure highlights the popularity of different colors based on customer preferences.

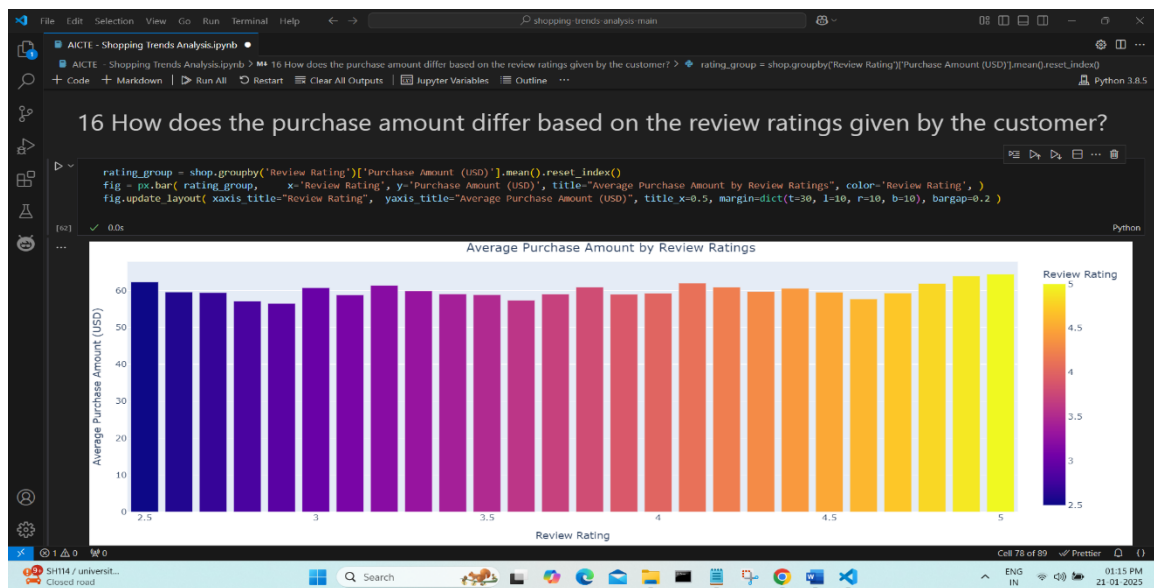


Figure 4: The figure shows how purchase amounts vary with different customer review ratings.

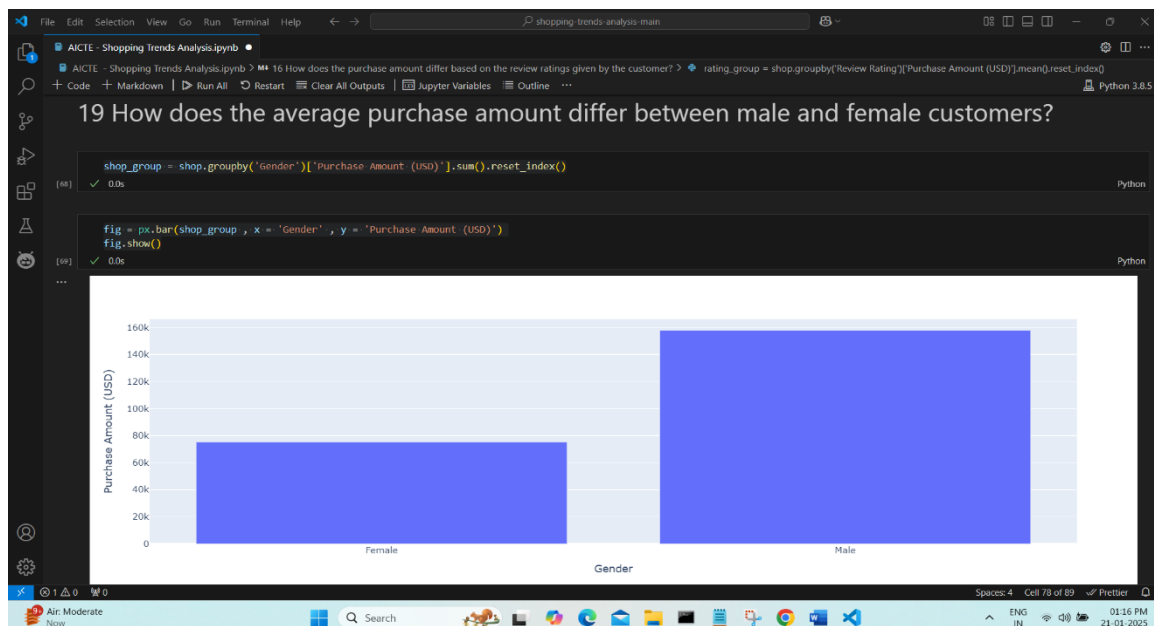


Figure 5: The figure compares the average purchase amounts between male and female customers.

4.2 GitHub Link for Code:

<https://github.com/nikitagote/Identifying-Shopping-Trends-Using-Data-Analysis.git>

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

The **Shopping Trends Analysis** project has successfully explored customer purchasing behavior based on various factors such as review ratings, gender, and age categories. However, there is room for further improvement and expansion:

1. **Integration with Machine Learning Models:** Future work could involve building predictive models to forecast purchase trends, customer churn, or lifetime value using machine learning techniques like regression or classification.
2. **Incorporating Additional Data Sources:** Integrating external data sources, such as social media reviews or competitor pricing, could enhance the analysis and provide deeper insights.
3. **Real-Time Data Analysis:** Implementing real-time data pipelines for live analysis of purchase trends and customer reviews would improve the system's responsiveness and usability.
4. **Advanced Visualizations:** Expanding the visualization layer to include dashboards with real-time updates and customizable metrics could make the findings more accessible to stakeholders.
5. **Addressing Limitations:** Addressing issues such as missing data, biases in the dataset, or inconsistencies in review ratings could further refine the analysis.

5.2 Conclusion:

The Shopping Trends Analysis project has provided valuable insights into customer purchasing behavior, offering a detailed understanding of how factors like review ratings, gender, and age categories influence purchase amounts. By leveraging Python-based analytical tools and interactive visualizations, this project demonstrates the potential of data-driven decision-making for businesses.

The findings can assist businesses in personalizing their marketing strategies, optimizing customer engagement, and increasing sales revenue. For instance, the correlation between positive reviews and higher purchase amounts emphasizes the importance of maintaining high product quality and customer satisfaction.

Overall, this project has laid a strong foundation for data-driven shopping trend analysis and opens new avenues for further advancements in the domain. By addressing the outlined future work, this project can evolve into a robust analytical tool for retail and e-commerce industries.

REFERENCES

- [1]. Chevalier, J. A., & Mayzlin, D., “The influence of online reviews on consumer behavior: An exploratory study,” *Journal of Marketing Research*, Vol. 43, No. 3, 2006, pp. 345-354.
- [2]. Gupta, S., & Mehta, R., “Impact of demographic factors on consumer buying behavior: A study of consumers in India,” *International Journal of Management Research*, Vol. 12, No. 4, 2019, pp. 45-58.
- [3]. Ahuja, N., & Yang, M. H., “Data visualization techniques in retail analytics,” *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 1, 2018, pp. 12-20.
- [4]. Mishra, P., & Choudhury, R., “Retail analytics and consumer behavior analysis,” *Springer Advances in Big Data*, Vol. 8, 2021, pp. 123-140.
- [5]. Pandas Documentation, Available at: <https://pandas.pydata.org/>
- [6]. Plotly User Guide, Available at: <https://plotly.com/python/>
- [7]. Seaborn Tutorials, Available at: <https://seaborn.pydata.org/>