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

Ritik Yadav, ry9812262@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude to all individuals who supported and guided me throughout this project. I am immensely thankful to my supervisor, **Jay Rathod**, for their invaluable advice, encouragement, and constructive feedback. Their guidance has been a source of inspiration and innovation.

This internship opportunity has provided a platform to enhance my technical skills and broaden my perspective in data analytics and AI-driven insights. I extend my gratitude to TechSaksham and AICTE for organizing this transformative learning experience, which has been instrumental in honing my abilities. Additionally, I thank my peers and colleagues for their cooperation and collaboration throughout this journey. Their insights and perspectives enriched my learning experience and inspired me to achieve the project's objectives effectively.

I deeply appreciate the unwavering support of my family, whose encouragement and belief in my abilities have been the cornerstone of my success. This project has been a remarkable learning curve, and the collective efforts of everyone involved made it a rewarding endeavor.

ABSTRACT

This project focuses on analyzing customer shopping trends through exploratory data analysis (EDA) of a well-structured dataset. The dataset encompasses customer demographics, purchase details, and behavioral preferences, offering a comprehensive view of shopping patterns. The primary objectives are to identify significant trends, such as age demographics, seasonal variations, and the impact of promotional strategies like discounts.

Using Python and its powerful libraries like Pandas, Matplotlib, and Seaborn, the project employs various EDA techniques to extract actionable insights. Key findings include the dominance of a 25-34 age group in purchases, higher shopping activity during holiday seasons, and a noticeable preference for discounted products. These insights serve as a valuable foundation for businesses to tailor their strategies, enhance customer satisfaction, and optimize marketing campaigns.

In conclusion, the project underscores the importance of data-driven decision-making in retail and provides a robust framework for analyzing consumer behavior effectively. By leveraging such insights, businesses can align their offerings with customer preferences, ultimately driving growth and profitability.

TABLE OF CONTENT

CHAPTER 1

Introduction

1.1 Problem Statement:

Understanding shopping trends is crucial for businesses to make data-driven decisions. The lack of actionable insights from customer purchase data often results in missed opportunities for improving customer satisfaction and optimizing marketing strategies.

1.2 Motivation:

This project was chosen to bridge the gap between raw data and actionable insights in retail. The analysis aims to unlock patterns and trends that can guide business decisions, enhance customer engagement, and improve profitability. The potential applications include personalized marketing, inventory optimization, and enhanced customer retention.

1.3 Objective:

- To analyze customer demographics and their purchasing behavior.
- To identify trends based on seasons, categories, and other factors.
- To evaluate the impact of discounts, promo codes, and customer subscriptions.
- To provide actionable insights for strategic decision-making.

1.4 Scope of the Project:

The project focuses on EDA of shopping data using Python. While it emphasizes trend identification, it does not delve into predictive analytics or machine learning, making it an exploratory rather than predictive study.

CHAPTER 2

Literature Survey

Review of Relevant Literature

Numerous studies in the domain of retail analytics and customer behavior have explored the factors influencing shopping trends. These include demographic segmentation, the impact of seasonal variations, and promotional strategies. Research highlights the importance of data-driven approaches to understanding consumer preferences. For example, studies have demonstrated the effectiveness of personalized promotions in increasing customer loyalty and sales.

Existing Models, Techniques, and Methodologies

Current methodologies employed in shopping trend analysis include:

- Market Basket Analysis: A technique to identify items frequently bought together, aiding in product placement and promotion strategies.
- Cluster Analysis: Grouping customers based on similar purchasing behaviors, enabling targeted marketing.
- Seasonal Trend Analysis: Studying variations in sales patterns across different times of the year.
- Sentiment Analysis: Evaluating customer feedback to understand satisfaction levels and preferences.

Gaps and Limitations in Existing Solutions

While the methodologies mentioned above provide valuable insights, several limitations persist:

1. Limited Integration of Real-Time Data: Existing systems often fail to incorporate real-time behavioral data, limiting their responsiveness to market dynamics.
2. Over-Reliance on Historical Data: Many models focus on past data without accounting for emerging trends or external factors.
3. Lack of Comprehensive EDA: Most studies provide high-level insights but lack an in-depth exploratory approach to identify hidden patterns.

Addressing the Gaps

This project aims to overcome these limitations by:

- Conducting a thorough EDA to uncover intricate patterns in customer behavior.
- Integrating multiple variables such as age, gender, seasonal trends, and promotional strategies to provide a holistic analysis.
- Offering actionable insights that businesses can implement immediately to enhance customer satisfaction and optimize marketing efforts.

CHAPTER 3

Proposed Methodology

3.1 System Design

The methodology involves:

- **Data Collection**: Extracting a comprehensive dataset with variables like age, gender, purchase details, and discounts.
- **Data Cleaning**: Handling missing values and inconsistencies.
- **Exploratory Data Analysis**: Utilizing Python libraries (Pandas, Matplotlib, Seaborn) to visualize trends and derive insights.

3.2 Requirement Specification

Mention the tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

- Minimum 8GB RAM.
- Core i5 Processor or equivalent.
- 256GB SSD storage.

3.2.2 Software Requirements:

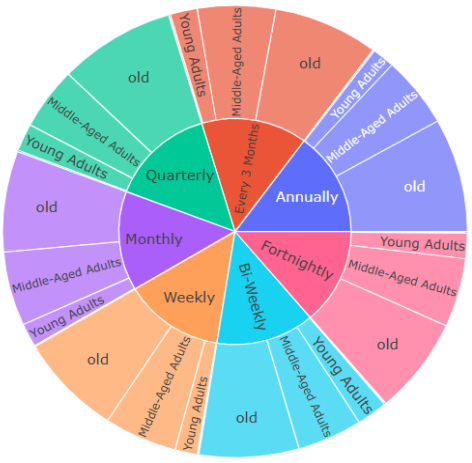
- Python 3.8 or above.
- Libraries: Pandas, Matplotlib, Seaborn.
- Jupyter Notebook.

CHAPTER 4

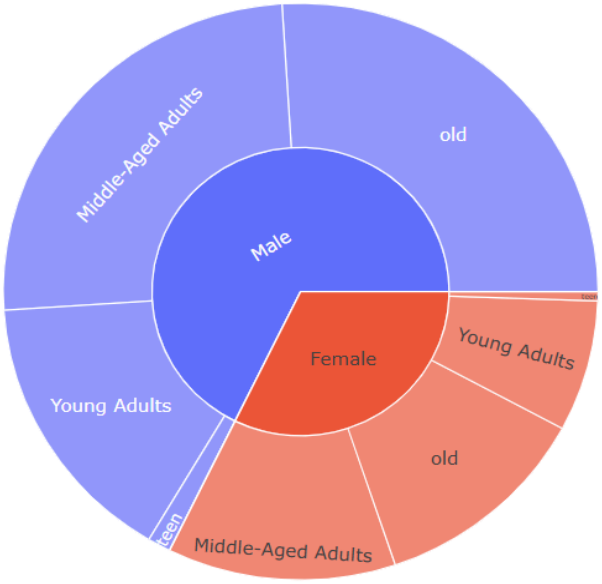
Implementation and Result

4.1 Snap Shots of Result:

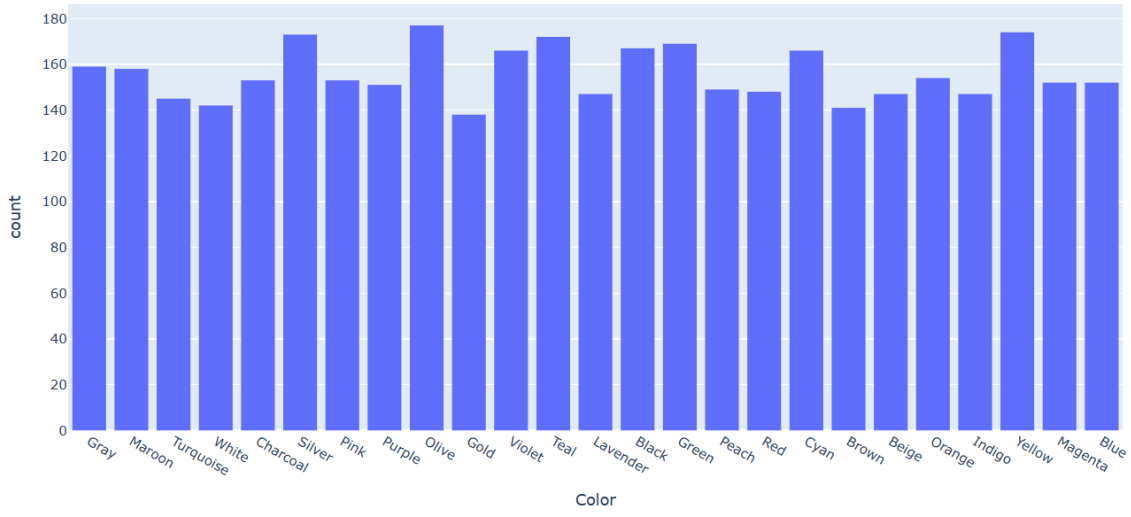
How does the frequency of purchases vary across different age:-



How does the average purchase amount differ between male and female customers?



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



4.2 GitHub Link for Code:

https://github.com/Ritik650/Shopping-Trend-Analysis

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- Integrating predictive analytics to forecast future trends.
- Expanding the dataset to include additional variables like customer feedback

5.2 Conclusion:

The project demonstrates the significance of EDA in retail analytics, offering insights into customer behavior that can inform business strategies and drive growth.

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.