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

Sanya Gupta,

Sanyaguptaa.1010@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

To start with, it's important to express my gratitude to my supervisor **Mr. Jay Rathod** for being an outstanding supervisor and the best adviser one can ever imagine. The guidance, motivation and the critics are source of fresh ideas, motivation and reasons behind the successful completion of this project. I must say that the confidence he had in me was the main great motivating factor towards me. I have enjoyed working with him for the last one year. He always assisted me during my project and many other facets that touched on the program. His talks and lessons if enable for the project work or any other conventions of the program are fully beneficial and enable me to become a good and responsible professional.

I also wish to acknowledge the **TechSaksham** initiative by **Microsoft** & **SAP** for providing this transformative learning opportunity. Lastly, I thank my family, peers, and friends for their unwavering support and encouragement.

Sanya Gupta.

ABSTRACT

The definitive feature of this work is the detailed exploration of the trends of the shoppers based on an exploratory data analysis (EDA) procedure. This report looks at the customer profile, purchases, and attitudes that underpin these areas as a way of understanding how spending happens. The objectives entail identifying relationships between various variables including; age, gender, the category, and the payment methods. Therefore, use of data visualization methods to extract intelligence measures was adopted. Essential insights indicate robust fluctuations throughout the year, customer specialty related to product categories, and the effect of promotional codes on spending. This paper also aims at benefiting the science of consumer examination by offering useful knowledge in the manner in which consumers can be managed for enhanced profitability by producers.

This project conducts an extensive exploratory data analysis (EDA) of shopping trends using a dataset containing 3,900 customer transactions. Using Python-based data analysis tools and visualization techniques, the study reveals important insights about customer segments, seasonal trends, and purchasing behaviors. The findings can help retailers optimize their inventory, marketing strategies, and customer experience.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

Knowing the shopping pattern is crucial for companies that wish to have a good stand in the constantly evolving marketplace. Consumer buying behavior depends on many factors some of which includes; age, gender, geographical locations and promotional appeals. However, the tools that are available in most businesses are not compatible to analyze all these elements at the same time, which hamper their marketing campaigns and inventory systems. This project fulfills the requirement of a holistic approach using exploratory data analysis (EDA) to optimize model interpretability and discover patterns within the data. Such reasoning allows businesses to increase customer interaction, refine products, and sharpen organizational performance. The influence of demographics, seasonality and promotional characteristics on consumers' decision making forms the basis of the study by developing extended recommendations for decision makers.

Yet, most companies find themselves devoid of these tools that would help them comprehend these elements as a single whole, which results in ineffective marketing strategies and inventory. The significance of these patterns is that it allows different businesses to better understand their customers, market, and improve the service offerings and company operations. The research is thus designed specifically to analyze the correlation between consumer characteristics, seasonal preferences and the impact of promotional deals to deliver a practical value for key decision makers.

1.2 Motivation:

The relevance of this project lies in the need to make business decisions based on analyzed data used in various sectors of the economy. It is no longer sufficient for businesses to present products in the current economy but it is necessary to be proactive in term of meeting customers' tastes and demands. This project is inspired by the need to gain insights of the data and be able to use it in areas such as marketing, inventory management and even relation to customers. Secondly, due to the increased availability of numerous and various types of data, analyzing changes in consumer behavior presents an interesting dilemma and prospect of realizing the potential of analytical solutions.

1.3 Objective:

The main concerns and goals of this project are concentrated around aspects of consumer behavior and their identification within the dataset. It covers the analysis of the frequency of the exhibited products and services and individual items that clients like to purchase, helping companies efficiently adapt to the market needs. The analysis also continues to explore how various demographic characteristics including age and gender affect the purchase behavior to understand the preferences of consumers in those segments. Moreover, the project considers the fluctuations related to the seasonal trends to identify main shopping periods and people's season preferences. Other important goals are understanding the efficiency of the use of promotional offers and payment options to ask about the potential model giving the most significant profit. In the end, this project is designed to find out where more optimal business strategies could be found through more comprehensive trend analysis which will reveal ways to enhance marketing, inventory control and in general customer satisfaction.

1.4 Scope of the Project:

This project is majorly centered on data analysis of a sample data base of 3900 customers where various details regarding the rights of customers and other important factors in business can be deduced accurately. The available variables include purchase category, purchase amount, and demographics of the buyers, which play a crucial role to deciphering customer behavior. The project therefore targets to focus solely at descriptive analytics where patterns and trends within the data are sought in a bid to provide an insightful snapshot beneficial for retailers, marketers and business decision makers alike.

Thus, the primary purpose of this study is to recognize trends and relationships that would be beneficial in increasing customer satisfaction and revenue as well as in refining business promotions and processes. Although the focus of this project is only to cover descriptive analytics, what has been done in this project laid down the groundwork to move to the next level of analytics namely predictive analytics and machine learning. They could be used to identify customer behavior trends, individualize promotions and make speculation about purchasing patterns that would prove beneficial to businesses.

In the given project, there is no attempt made to analyze data in real time or to feed data from outside the Bank into the structure, however what is shown is the power of structured data analysis.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

Scholarly studies in the consumer behavior field demonstrate that shopping behavior is a product of rational and complex processes. Demographic variables have been investigated in many research papers, demonstrating how age, gender and income influence a purchasing decision. The economic analyses also stress out the importance of the price policies and scenes, as well as appeals to the targeted consumer. For example, the studies about seasonal characteristics demonstrate that during festivals and holidays, people spend more money on items such as clothing or electronics. Nevertheless, most of the work done within this field uses fixed data putting a constraint on scenario based on dynamic markets. Prior studies of consumer behavior are characterized as segmented with most of them addressing the partial issue of consumer behavior e.g. consumer loyalty, consumer responses to pricing, or consumer attitude and perceptions rather than regarding it as an entire concept.

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

A number of theoretical frameworks and methods have been the subject of examination in relation to shopping trends. The customers have been segmented using cluster analysis and segmentation models whereby the customers are grouped depending on their buying frequency and average reorder. Regression analysis is useful in predicting variables that have a large impact on expenses, these include discount promotions. Recommendation systems make use of collaborative and content based filtering to suggest products to customers. Nevertheless, these models can be inadequate in addressing temporal and seasonality aspects of the shops. Second, traditional methods are history-oriented and cannot be adjusted quickly.

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

However, all documented methodologies have their own lacunas to that end. Demographic factors and promotional strategies are the two key areas of many models, however, their interaction is often left unaddressed. The same can be said about the static approaches they do not take into account fluctuations in the consumers' preferences due to such reasons as economic or others. Realtime flexibility and individualized recommendations are still things that can be enhanced. Further, very little literature focuses on the implementation of combining qualitative data like customer feedback into the analysis..

CHAPTER 3

Proposed Methodology

3.1 System Design

The methodology adopted for this project involves the following key steps:

1. **Data Cleaning:**

   The possessive technique of missing values must be managed properly for making the dataset complete. Any entry with the blank value was treated by either imputing it or excluding it depending on its necessity in the data set.

   Managing data formats by actually keeping the format of data consistent to allow for optimization when consuming and analyzing data.

2. **Exploratory Data Analysis:**

   To introduce fundamental aspects of the dataset and compile the basic comparison parameters for further analysis.

   Using value counts and grouping to gather insight and trends, and to examining outliers and variations.

3. **Data Visualization:**

   Using Seaborn and Plotly in order to create interactive and static visual aids where stakeholders can hover to know more about the specifics of the findings.

   Providing distinct and visible aspects including the seasonal spending and gender preferences as well as the popularity of the category.

4. **Analysis and Interpretation:**

   Determining factors controlling expenditure including the use of promotions or the relationship between age and spending.

   Approximating variables to make conclusions about naturally occurring relationships, for example, payment methods and average purchase size.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

An ordinary laptop or personal computer with 8GB or more RAM and 512 GB in disc space. It also suggested that higher configurations may be useful for accelerating the rate for analysing larger datasets.

3.2.2 Software Requirements:

- Python software environment.
- Libraries: Pyspark or pandas for data manipulation, seaborn or matplot lib for static data visualization, plotly.
- Markdown, and in particular, Jupyter Notebook for development for the distinct steps in the process as well as the iteration.

CHAPTER 4

Implementation and Result

4.1 Implementation:

1. **Age Distribution:**

   On the population, an age range was divided into some bins like child, teen, young adult, middle-aged, and elderly. This classification was able to show that middle aged adult were the most active shoppers. To cater for this distribution, a histogram was used thus exposing essential demography that business should focus on.

2. **Product Popularity:**

   Clothing was found to be the most commonly bought item in the selected category, and some of the products most in demand for this segment are blouses and pants. This was accompanied by bar graphs displaying an analysis of the product rating grouping in the categories.

3. **Seasonal Trends:**

   On spending activity, the highest mean was obtained in Spring season second to which was Fall showing a pattern of the season the spending activity is likely to occur. To highlight peak shopping months, line graphs and histograms were used while selling trends were displayed in bar graphs.

4.2 Result:

1. **Insights on Gender:**

   Gender segmentation showed that male and female shoppers participated in the event to almost similar extent, although certain differences existed with respect to the event categories and types of payments made by them. These differences were illustrated by bar plots and were ideal for defining the scope of targeted campaigns.

2. **Promo Code Effectiveness:**

   Those who had promo codes had a slightly higher average spending as was expected hence encouraging the promo code strategy. A comparison chart was used to give emphasis on this tendency.

3. **Payment Methods:**

Debit card was the most preferred method of payment; this shows that consumers prefer demonstrations of swift and secure payment. Pie chart showing the manner in which different payments were made.

4.3 Snap Shots of Result:

1. **What is the overall distribution of customer ages in the dataset?**

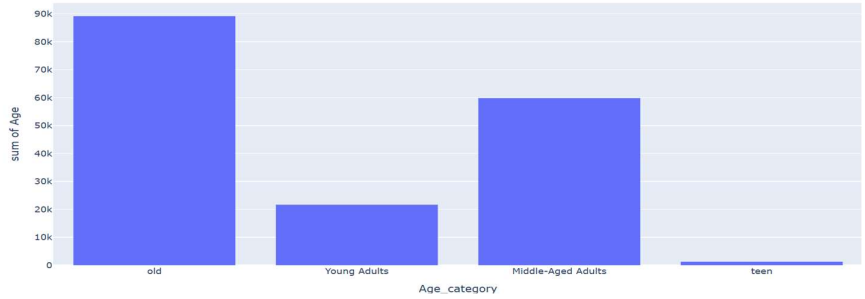

Fig.1

2. **How does the average purchase amount vary across different product categories?**

Category	Purchase Amount (USD)
Accessories	59.838710
Clothing	60.025331
Footwear	60.255426
Outerwear	57.172840

dtype: float64

Fig.2

3. **Which gender has the highest number of purchases?**

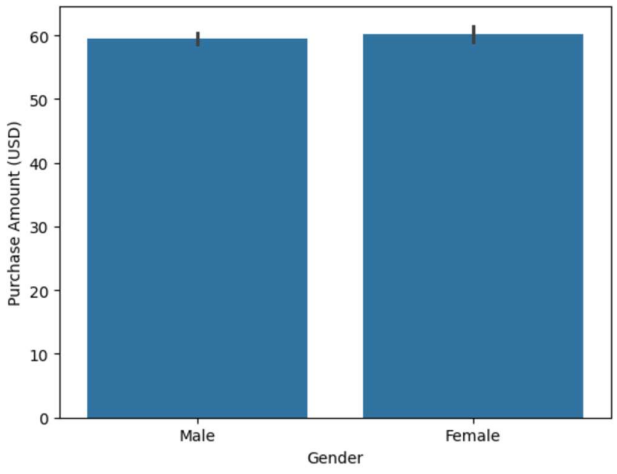

Fig.3

4. **What are the most commonly purchased items in each category?**

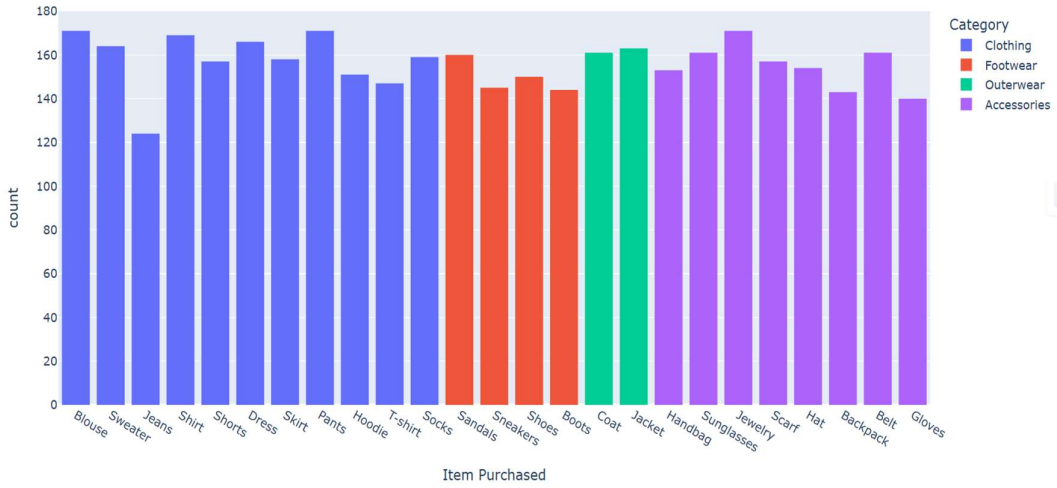

Fig.4

5. **Are there any specific seasons or months where customer spending is significantly higher?**

Fig.5

6. **What is the average rating given by customers for each product category?**

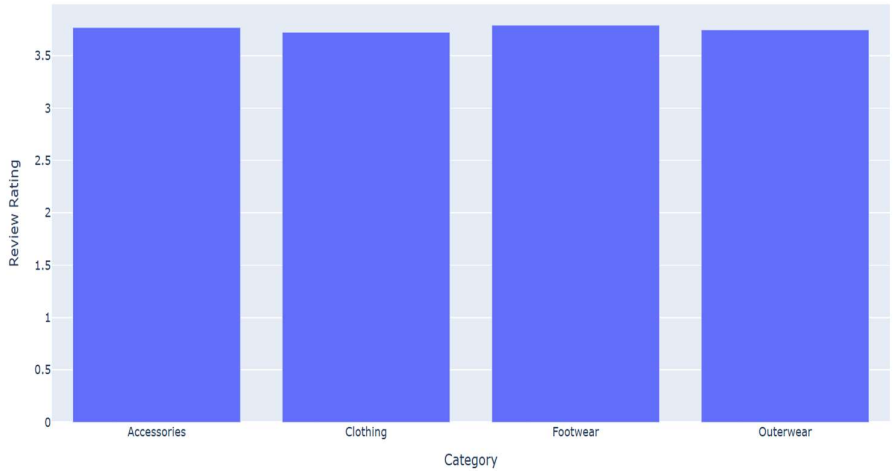

Fig.6

7. **Are there any notable differences in purchase behavior between subscribed and non-subscribed customers?**

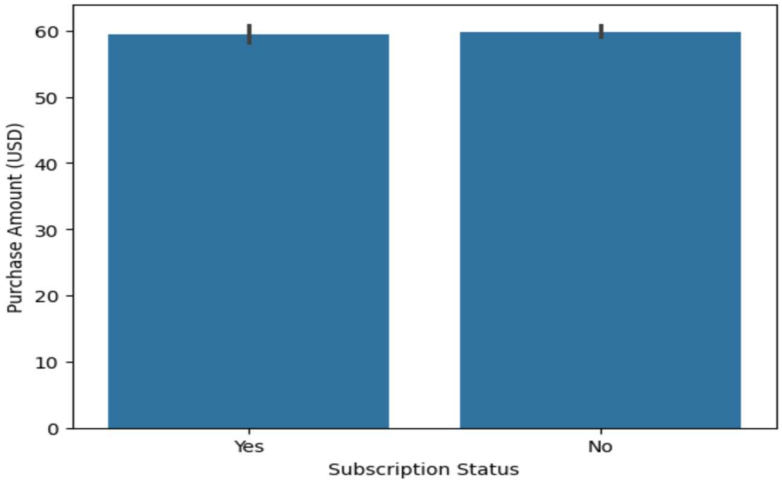

Fig.7

8. **Which payment method is the most popular among customers?**

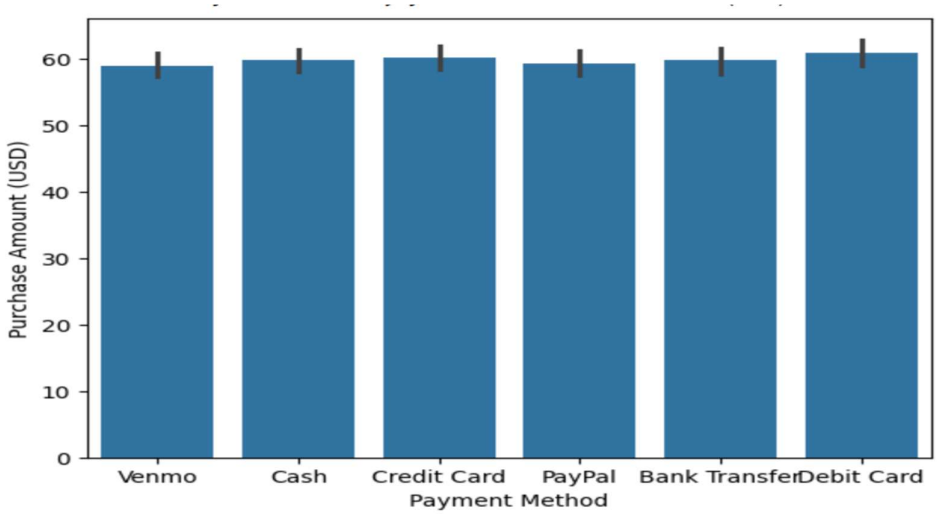

Fig.8

9. **Do customers who use promo codes tend to spend more than those who don't?**

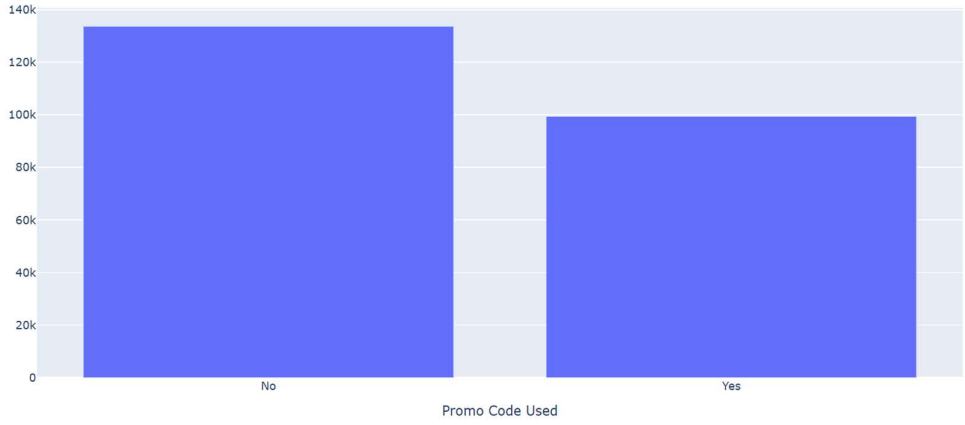

Fig.9

10. How does the frequency of purchases vary across different age groups?

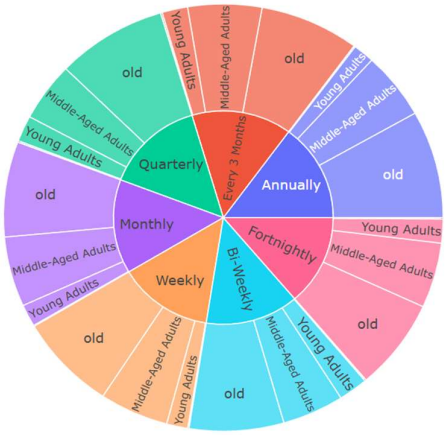

Fig.10

11. Are there any correlations between the size of the product and the purchase amount?

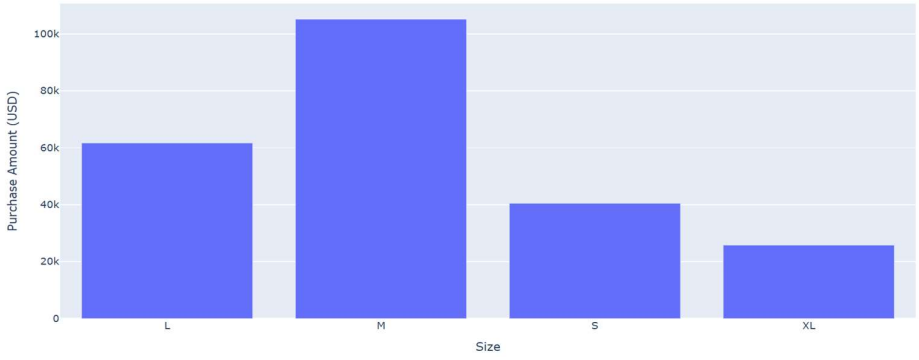

Fig.11

12. Which shipping type is preferred by customers for different product categories?

Category	Shipping Type	count
Clothing	Standard	297
	Free Shipping	294
	Next Day Air	293
	Express	290
	Store Pickup	282
	2-Day Shipping	281
Accessories	Store Pickup	217
	Next Day Air	211
	Standard	208
	2-Day Shipping	206
	Express	203
	Free Shipping	195
Footwear	Free Shipping	122
	Standard	100
	Store Pickup	98
	Express	96
	Next Day Air	93
	2-Day Shipping	90
Outerwear	Free Shipping	64
	Express	57
	Store Pickup	53
	Next Day Air	51
	2-Day Shipping	50
	Standard	49

Fig.12

13. How does the presence of a discount affect the purchase decision of customers?

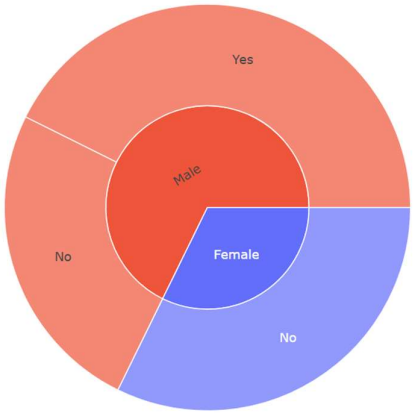

Fig.13

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

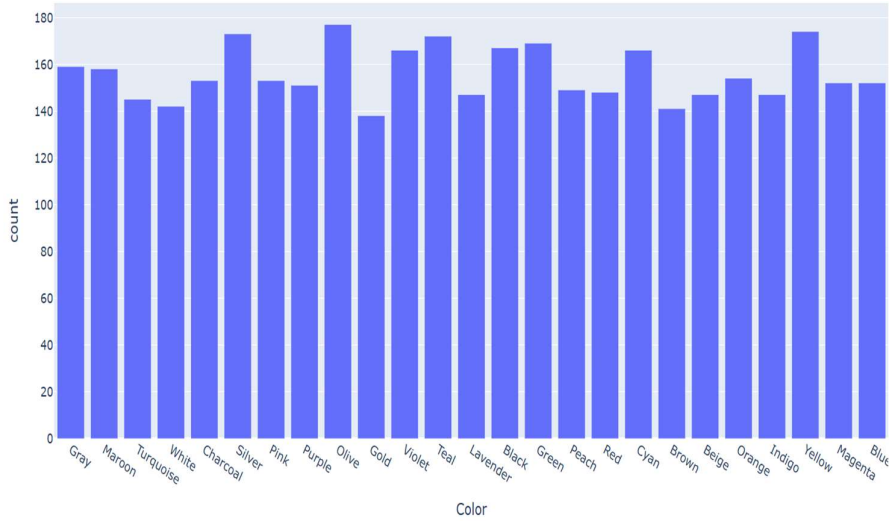

Fig.14

15. What is the average number of previous purchases made by customers?

```
25.35153846153846
```

Fig.15

4.4 GitHub Link for Code

https://github.com/SanyaGuptaa/Shopping-Trends-Analysis

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

1. The company should incorporate the appropriate form of predictive analytics to help inform future sales trends and locate avenues for expansion.
2. In order to get more granular in segmenting, extend the data set to include the geographic or cultural characteristics. This can be useful in giving understanding of shopping patterns of the particular region.
3. Examine how to apply the techniques of sentiment analysis on the customer's feedback to address the analysis of qualitative attributes of the purchase decision. This can assist the companies to respond to the complains from the customer properly.

5.2 Conclusion:

This work aims at explaining a shopping trend analysis through the use of EDA methods in this particular project. The observations highlighted in the paper help to identify key behavioral characteristics, like demographic factors, categories of consumer preference, and an effect of seasonality or promotional campaigns. Hence, these views provide critics and useful suggestions to organizations desiring to fine-tune their approaches in relation to customer desires. The strength of this analysis lies in showcasing how data analytics can help close the gap between quantitative data and solutions within the contextual focus of the retail sector.

In conclusion, the goals of this project are in three fold: This high level overview is not only indicative of the effectiveness of exploratory data analysis in identifying significant trends but also of the potential for further development by employing advanced analysis techniques such as predictive modeling and real-time data processing. They help present strategies of how business can change in a world where data is rapidly becoming a key resource and a force to reckon with in fulfilling the needs of the consumer.

Shopping trends are explored in this project using EDA as a way of analyzing shopping trends more comprehensively. The results obtained enable uncovering of important patterns, trends and associations, including the demographic characteristics of the customers, their preference by the product category, or the effects that the seasonal and promotion factors may impose on the buying behaviour of the customers. All these ideas present good sources of information for firms that seek to enhance their operations in order to meet the buyer needs. Thus, demonstrating how this study has identified the need of data analytics in the new model of retailing by closing the gap between big data and managerial use.

This particular project involves carrying out Exploratory Data Analysis (EDA) to study shopping behavior and paying special attention on demographic features, preference as well as seasonality and promotional information. The findings of this analysis help refine the strategies and their understanding of customers. Thus, through illustrating how data analytics helps in closing the gap between data and solutions this project shows the importance of data analytics in the retail industry.

REFERENCES

[1]. Kumar, A., & Gupta, R. "**Consumer Behavior Analysis in Retail**". Journal of Retail Studies, 2019.

[2]. Silverstein, M., Fiske, N., & Butman, J. "**Trading Up: The New American Luxury**". Portfolio, 2003.

[3]. Kotler, P., & Keller, K. "**Marketing Management**". Pearson Education, 15th Edition, 2016.

[4]. McKinsey & Company. "**The State of Fashion 2023**". Retrieved from McKinsey Insights, 2023.

[5]. GfK Research. "**Global Consumer Trends Report**". Retrieved from GfK Insights, 2022.

[6]. Kaggle. "**Retail Data Analysis: Trends and Insights Dataset**". Retrieved from Kaggle Datasets, 2021.

[7]. Harvard Business Review. "**Data-Driven Marketing: Lessons from the Leading Edge**". Harvard Business Publishing, 2020.

[8]. Kumar, A., & Gupta, R. "**Consumer Behavior Analysis in Retail**". Journal of Retail Studies, 2019