

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

Tanuj kumar, tanujkumar99995@gmail.com

Under the Guidance of

Jay Rathod and P Raja

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to everyone who supported me throughout the development of this project, "Identifying Shopping Trends Using Data Analysis."

First and foremost, I am deeply thankful to Jay Rathod , P Raja , Pavan kumar U, for their invaluable guidance, constructive feedback, and unwavering support during this project. Their expertise and mentorship played a crucial role in shaping the direction and success of this project.

I am also grateful to TechSaksham, Edunet for providing the necessary resources, tools, and an environment conducive to learning and exploration. The availability of data and technical infrastructure has been instrumental in conducting meaningful analysis.

This project has been an incredible learning experience, and I am sincerely appreciative of everyone who contributed to its completion.

Thank you all!

Tanuj kumar

27/01/2025

ABSTRACT

In the modern retail landscape, understanding customer behavior is pivotal for optimizing sales strategies and enhancing customer satisfaction. This project, *"Identifying Shopping Trends Using Data Analysis,"* aims to uncover actionable insights from customer purchase data.

Problem Statement: Retailers face challenges in identifying key factors that drive customer purchases, such as demographics, product preferences, and seasonal trends. Without a comprehensive understanding, it is difficult to make data-driven decisions to maximize profitability and customer engagement.

Objectives:

1. Analyze customer demographics, such as age and gender, and their impact on purchasing behavior.
2. Examine purchase trends across product categories, payment methods, seasons, and locations.
3. Identify factors like discounts, promo codes, and product features (e.g., size, color) that influence purchase decisions.
4. Compare behaviors between subscribed and non-subscribed customers.

Methodology:

The analysis was conducted using a dataset containing transactional records, including customer demographics, purchase details, and product attributes. Techniques such as statistical analysis, data visualization, and machine learning were employed to identify patterns and trends. Key tools used include Python libraries like Pandas, Matplotlib, Seaborn, and Plotly.

Key Results:

- The overall age distribution revealed a significant concentration of purchases from customers aged 20-70.
- Promo codes and discounts significantly increased purchase amounts.
- Males exhibited higher purchase frequency, while females had a higher average purchase amount.
- Seasonal spikes in spending were observed during holiday months.
- Subscription customers showed higher engagement and purchase amounts.
- Product attributes, such as specific colors and sizes, influenced preferences, while certain payment methods, such as credit cards, emerged as the most popular.

Conclusion:

The analysis highlights the importance of leveraging data-driven insights to optimize product offerings, marketing strategies, and customer experience. By tailoring approaches to demographic and seasonal preferences, businesses can enhance profitability while fostering customer loyalty.

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	3
Chapter 3. Proposed Methodology	4
Chapter 4. Implementation and Results	6
Chapter 5. Discussion and Conclusion	15
References	17

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	It show that the older and middle-aged groups dominate dataset or older and middle-aged groups purchase more items	6
Figure 2	It show that the male purchase more item as compare to the female.	6
Figure 3	It can be concluded that all genders have made an equal number of purchases. This suggests that there is no significant difference in the purchase frequency among different genders.	7
Figure 4	The graph shows that pants, sandales, jackets, and jewelry are the most purchased items in their respective categories.	7
Figure 5	It show that constumer purchase more item in spring season as compare to the other season.	7
Figure 6	It can be concluded that customer spending is significantly higher during the spring season	8
Figure 7	The graph shows that the average review ratings for all product categories—Accessories, Clothing, Footwear, and Outerwear—are consistently around 3.5. This indicates a uniform level of customer satisfaction across all categories, with no significant variation in ratings.	8
Figure 8	The findings indicate that, at a high level, there are no notable differences in purchase behavior between subscribed and non-subscribed customers, suggesting a consistent consumer landscape across both groups	9
Figure 9	It can be concluded that there is no significant difference in popularity among the listed payment methods among customers. The data indicates that Debit Card, Credit Card, Bank Transfer, Cash, PayPal, and Venmo all have similar popularity rates, with only slight variations observed.	9
Figure 10	The data suggests that businesses should consider offering a variety of payment options to cater to the diverse preferences of their customers.	10
Figure 11	It can be concluded that there is a notable difference in the usage of promo codes between male and female customers.	10
Figure 12	It can be concluded that the frequency of purchases varies across different age groups, with the "old" category exhibiting the highest frequency of purchases	11
Figure 13	It can be concluded that there is a correlation between the size of the product and the purchase amount.	11

Figure 14	It show that the discount applied by the constumer at the time of purchase.	12
Figure 15	It can be concluded that the presence of a discount has an impact on the purchase decisions of customers, with distinct patterns observed between genders	12
Figure 16	It can be concluded that certain colors are more popular among customers like yellow silver.	12
Figure 17	It can be concluded that there are no noticeable differences in purchase behavior between different locations (states).	13
Figure 18	It can be concluded that there is no apparent relationship between customer age and the category of products they purchase.	13
Figure 19	It show that male spend more amount than the female.	13
Figure 20	It can be concluded that there is no apparent relationship between customer age and the category of products they purchase. This conclusion is based on the observation that the average age of customers is 44 years, as determined from the bar plot.	14

CHAPTER 1

Introduction

1.1 Problem Statement:

The retail industry generates vast amounts of data, but many businesses struggle to derive actionable insights to understand customer purchasing behaviors. Key challenges include identifying how demographic factors, product preferences, and promotional strategies influence purchases. This lack of insight hinders businesses from optimizing their marketing strategies, improving customer satisfaction, and increasing profitability. Addressing these issues is crucial for maintaining a competitive edge in an ever-evolving market.

1.2 Motivation:

This project was chosen to bridge the gap between raw data and meaningful insights, enabling businesses to make informed decisions. By leveraging data analysis, companies can identify shopping trends, enhance marketing strategies, and tailor their offerings to better meet customer needs. The potential applications include personalized marketing, inventory optimization, and improved customer retention strategies. The impact extends beyond increased revenue, fostering customer loyalty and a more satisfying shopping experience.

1.3 Objective:

The primary objectives of this project are:

1. To analyze the distribution of customer ages and its impact on purchasing behavior.
2. To identify variations in purchase amounts across product categories.
3. To examine gender-based differences in purchase frequency and amount.
4. To determine the most commonly purchased items within each category.
5. To evaluate seasonal and monthly trends in customer spending.
6. To analyze the average rating given by customers for each product category.
7. To compare purchase behaviors of subscribed and non-subscribed customers.
8. To identify the most popular payment methods among customers.
9. To evaluate the impact of promo codes, discounts, and product features (e.g., size, color) on purchase decisions.
10. To explore correlations between product attributes and purchase amounts.
11. To assess differences in purchase behavior based on location.
12. To study relationships between customer age, product categories, and review ratings.

1.4 Scope of the Project:

This project focuses on analyzing customer transactional data to identify trends and patterns in purchasing behavior. Key aspects include examining demographics, product

preferences, and seasonal effects. The analysis is limited to the available dataset and may not account for external factors such as economic conditions or competitor strategies. While providing actionable insights, this project does not include the implementation of predictive models or recommendation systems

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

- Previous research on consumer behavior analysis, especially in e-commerce, has highlighted the impact of customer demographics (such as age, gender) and product attributes (such as category, color, size) on purchasing decisions.
- Several studies have focused on the role of promotional offers and discounts in influencing purchase behavior.
- Existing models like collaborative filtering and association rule mining have been used for product recommendations, and techniques like clustering are often employed for market segmentation.

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

- **Clustering Algorithms (e.g., K-means, DBSCAN):** Used to group customers based on purchasing behavior and demographics.
- **Regression Analysis:** Applied to understand the impact of various factors (like age, gender, discounts) on purchase amounts.
- **Classification Models (e.g., Decision Trees, Random Forest):** Used to predict customer behavior such as the likelihood of making a purchase or opting for discounts.
- **Time Series Analysis:** Employed to study seasonal trends and predict future purchasing behaviors.
- **Association Rule Mining:** Applied for identifying frequently bought items together and understanding customer preferences.

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

- **Lack of Granular Data Analysis:** Many existing studies focus on general demographic information without detailed analysis of purchase amounts across different product categories or specific customer behaviors.
- **Behavioral Insights on Subscription Status:** Few studies explicitly explore differences in purchase behaviors between subscribed and non-subscribed customers.
- **Limited Focus on Payment Methods and Promo Codes:** While promo codes are often discussed in isolated contexts, integrated analysis alongside payment methods and their influence on spending is less explored.
- **Regional Variations in Purchasing Behavior:** Most studies do not focus on location-based variations in purchasing behavior.

CHAPTER 3

Proposed Methodology

3.1 System Design

[Data Sources] → [Data Ingestion] → [Data Storage] → [Preprocessing] → [Analysis & Modeling] → [Visualization & Reporting]

Data Source: Collect the consumer data from E-commerce platform or other website.

Data Ingestion: Scheduled imports of customer and transaction data from databases (SQL/NoSQL) or flat files (CSV, Excel).

Data Storage: Store the data using Structured Data: Relational databases (MySQL, PostgreSQL).

Data Processing: Extract, transform, and load data for cleaning and preprocessing.

- Tools: Python (Pandas, NumPy), SQL.

Analysis and modelling: Exploratory Data Analysis, Statistical Analysis of the data.

Visualization & Reporting: visualize the analysis data using matplotlib, seaborn and prepare a report on the analysis data for the further action.

Requirement Specification

The tools and technologies required to implement the solution.

3.1.1 Hardware Requirements:

Processor

RAM

Storage

Graphics Card

Network Connection

3.1.2 Software Requirements:

Pandas: For data manipulation and analysis.

NumPy: For numerical operations.

Matplotlib, Seaborn: For data visualization and plotting.

Plotly: For interactive visualizations and dashboards (optional).

SciPy: For statistical analysis.

Python: Programming Language

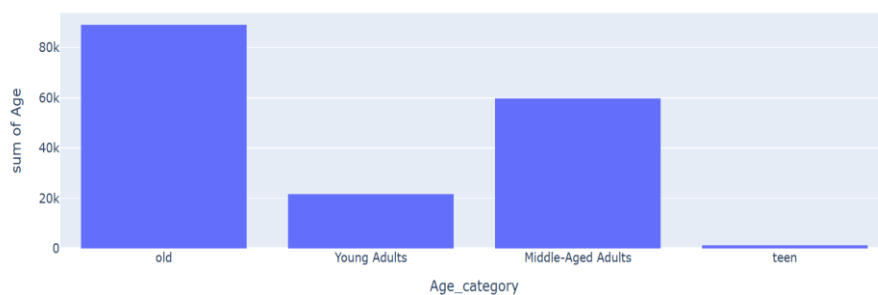
Jupyter Notebook: For interactive development and data analysis.

CHAPTER 4

Implementation and Result

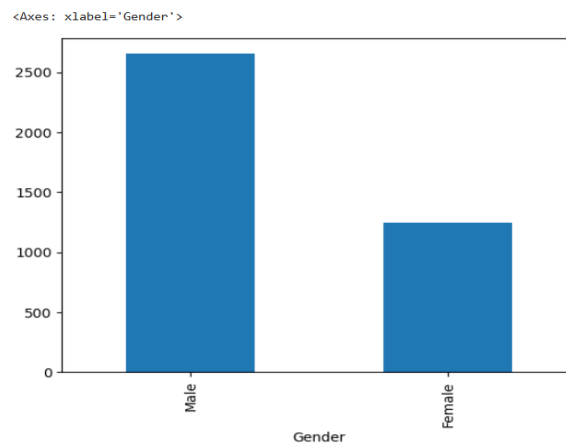
4.1 Snap Shots of Result:

Fig: 1



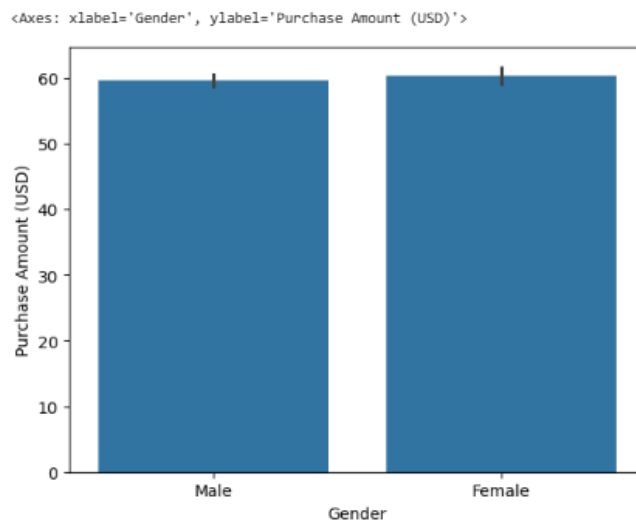
we are cutting the age into some category and storing in the different column. It show that the older and middle-aged groups dominate dataset or older and middle-aged groups purchase more items.

Fig: 2



It show that the male purchase more item as compare to the female.

Fig: 3



It can be concluded that all genders have made an equal number of purchases. This suggests that there is no significant difference in the purchase frequency among different genders.

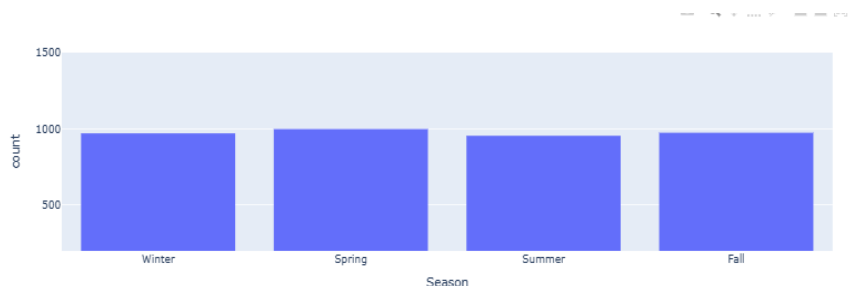
Fig: 4



This histogram show that which item most purchase in the each category.

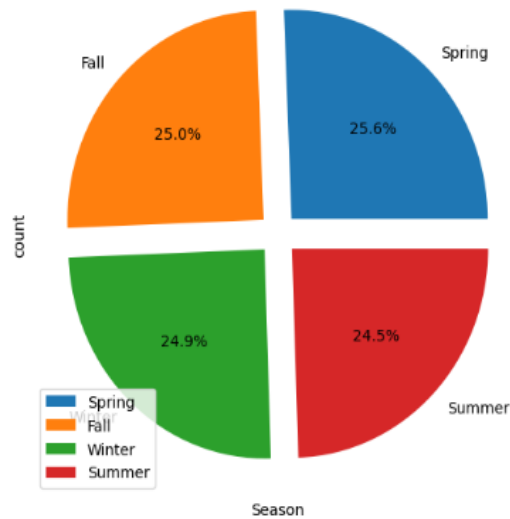
The graph shows that **pants, sandales, jackets, and jewelry** are the most purchased items in their respective categories. These items stand out as customer favorites, indicating areas to prioritize for inventory and marketing strategies.

Fig: 5



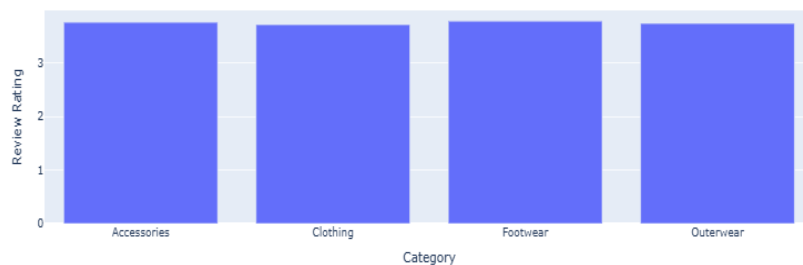
It show that constumer purchase more item in spring season as compare to the other season.

Fig: 6



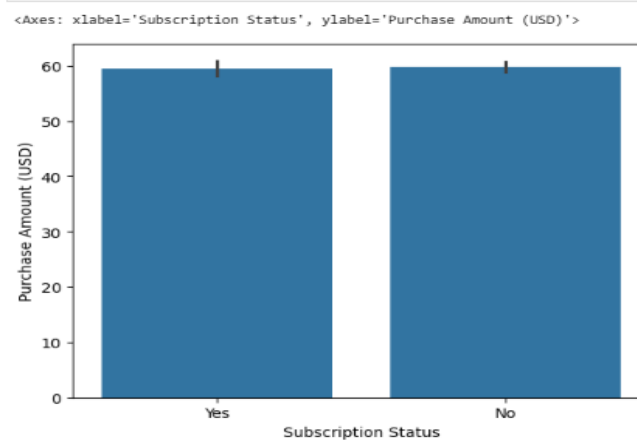
It can be concluded that customer spending is significantly higher during the spring season

Fig: 7



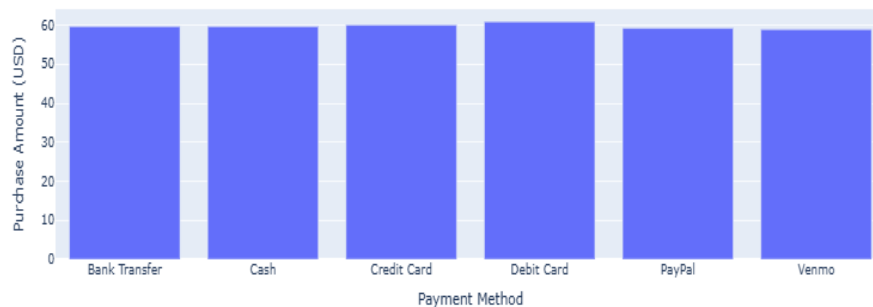
The graph shows that the average review ratings for all product categories—Accessories, Clothing, Footwear, and Outerwear—are consistently around 3.5. This indicates a uniform level of customer satisfaction across all categories, with no significant variation in ratings.

Fig: 8



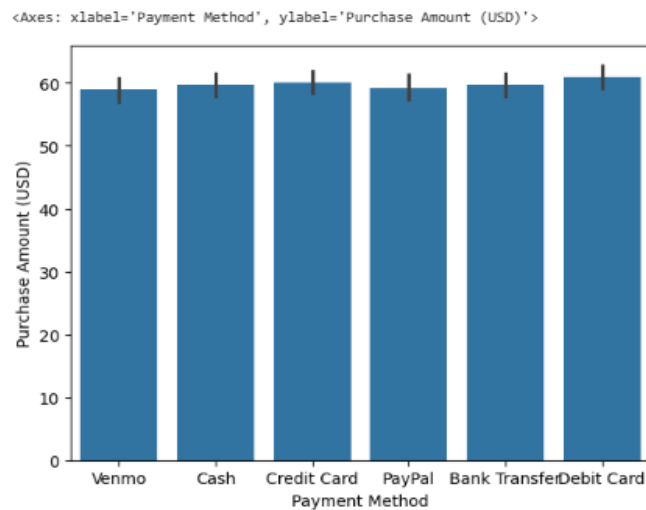
The findings indicate that, at a high level, there are no notable differences in purchase behavior between subscribed and non-subscribed customers, suggesting a consistent consumer landscape across both groups.

Fig: 9



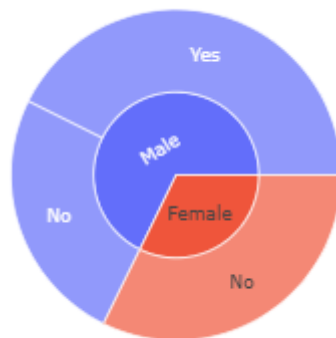
It can be concluded that there is no significant difference in popularity among the listed payment methods among customers. The data indicates that Debit Card, Credit Card, Bank Transfer, Cash, PayPal, and Venmo all have similar popularity rates, with only slight variations observed.

Fig: 10



The data suggests that businesses should consider offering a variety of payment options to cater to the diverse preferences of their customers. Providing a range of payment methods can enhance convenience, improve customer satisfaction, and accommodate various customer preferences and needs.

Fig: 11



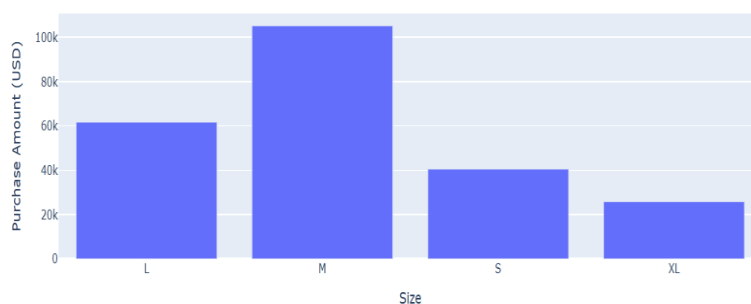
It can be concluded that there is a notable difference in the usage of promo codes between male and female customers. The data suggests that male customers tend to use promo codes more frequently compared to female customers, while female customers have a lower tendency to utilize promo codes.

Fig: 12



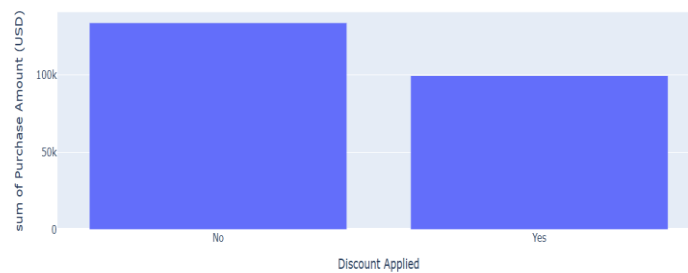
It can be concluded that the frequency of purchases varies across different age groups, with the "old" category exhibiting the highest frequency of purchases.

Fig: 13



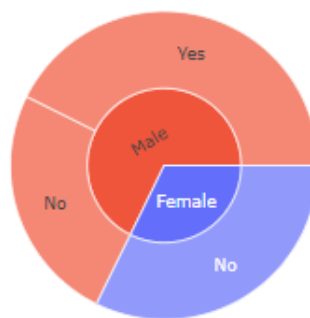
It can be concluded that there is a correlation between the size of the product and the purchase amount. the correlation between product size and purchase amount can assist businesses in optimizing their product offerings and pricing strategies, ensuring they align with customer preferences and maximize sales potential.

Fig: 14



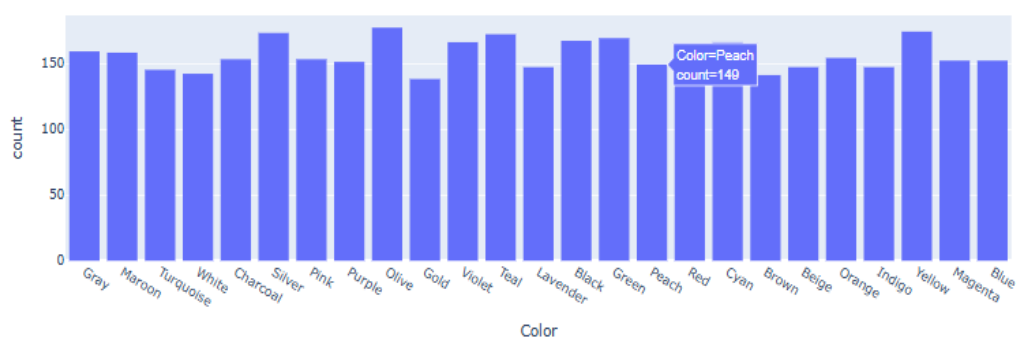
It show that the discount applied by the consumer at the time of purchase.

Fig: 15



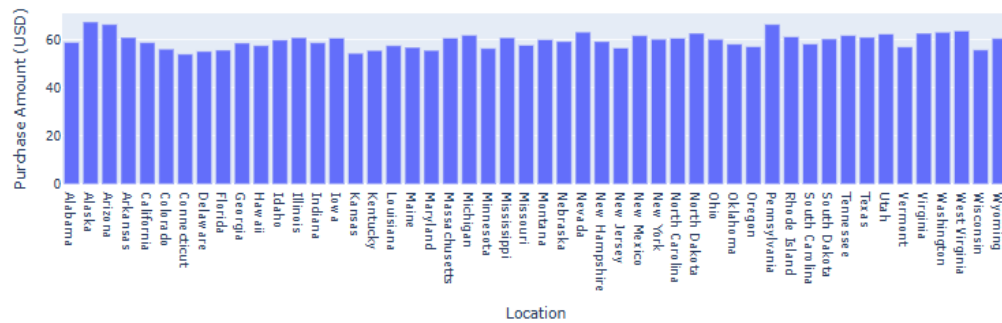
It can be concluded that the presence of a discount has an impact on the purchase decisions of customers, with distinct patterns observed between genders

Fig: 16



It can be concluded that certain colors are more popular among customers like yellow silver.

Fig: 17



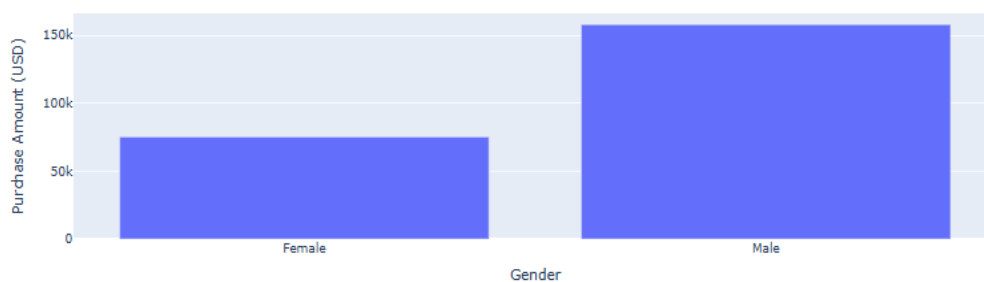
It can be concluded that there are no noticeable differences in purchase behavior between different locations (states). The data suggests that all states exhibit similar purchasing patterns, with no significant variations observed. This finding indicates that customers across all states demonstrate consistent purchasing behavior, regardless of their geographical location.

Fig: 18



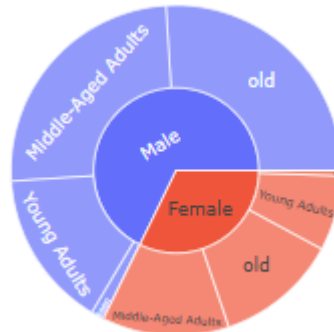
It can be concluded that there is no apparent relationship between customer age and the category of products they purchase. This conclusion is based on the observation that the average age of customers is 44 years, as determined from the bar plot.

Fig: 19



It show that male spend more amount than the female.

Fig: 20



It can be concluded that there is no apparent relationship between customer age and the category of products they purchase. This conclusion is based on the observation that the average age of customers is 44 years, as determined from the bar plot.

4.2 GitHub Link for Code:

<https://github.com/Tanujkumarsingh/identifying-shopping-trends-project/blob/main/identifying%20shopping%20%20.ipynb>

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

The project has achieved its objectives of analyzing customer purchasing trends and deriving actionable insights. However, there are areas where future work can build on the current findings to enhance the analysis and extend its applicability:

- ☐ **Real-Time Integration:** Incorporate real-time data to dynamically update trends and enable live decision-making through interactive dashboards.
- ☐ **Advanced Models:** Develop predictive models using advanced techniques like neural networks or ensemble methods to improve accuracy and uncover complex patterns.
- ☐ **Recommendation Systems:** Build personalized product recommendations based on purchase history and customer preferences using collaborative and content-based filtering.
- ☐ **Expanded Demographics:** Analyze additional factors such as income level, profession, and regional trends to refine customer segmentation and insights.
- ☐ **Improved Privacy Measures:** Strengthen data privacy mechanisms to comply with regulations like GDPR and ensure ethical handling of customer data.

Conclusion:

This project successfully analyzed customer purchasing trends, providing valuable insights into factors such as age distribution, gender-based purchasing patterns, seasonal spending behaviors, and the influence of discounts, promo codes, and payment methods. The findings contribute to a deeper understanding of customer preferences and behavior, enabling businesses to:

- ☐ Optimize product offerings based on customer preferences.
- ☐ Implement targeted marketing strategies for different customer segments.
- ☐ Improve decision-making regarding pricing, discounts, and promotional campaigns.

- Enhance customer satisfaction through personalized recommendations and tailored experiences

The analysis has highlighted significant patterns, such as the impact of discounts on purchasing decisions, popular payment methods, and category-specific customer preferences. These findings can help businesses improve customer engagement, boost sales, and foster long-term loyalty.

REFERENCES

<https://www.geeksforgeeks.org/python-introduction-matplotlib/> visit for Matplotlib

<https://www.geeksforgeeks.org/introduction-to-pandas-in-python/> visit for pandas

<https://seaborn.pydata.org/tutorial/introduction.html> Visist for seaborn in python

https://www.w3schools.com/python/numpy/numpy_intro.asp Visit for numpy in python

<https://www.geeksforgeeks.org/python-plotly-tutorial/> Visit for plotly

Python Software Foundation. (n.d.). *Python Programming Language*. Retrieved from <https://www.python.org>