**Project Title**

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

**Yamuna Sri. M, yamunasri1810@gmail.com**

Under the Guidance of

**Jay Rathod**

**ACKNOWLEDGEMENT**

We are deeply grateful for the support and guidance we received from numerous individuals throughout the course of this thesis work.

We would like to express our sincere appreciation to our supervisor, Pavan Kumar U, Jay Rathod, P Raja, for their exceptional mentorship and invaluable guidance. Their expertise, encouragement, provided a constant source of inspiration and innovation, driving us towards the successful completion of this project. We are incredibly grateful for the unwavering confidence they placed in our abilities, which served as a significant motivating force throughout our learning journey. The past 1 month spent working under their supervision has been an invaluable experience. Their support extended beyond the scope of our project, encompassing various aspects of our academic program. Their insightful discussions and invaluable lessons have not only enriched our analysis endeavors and academic pursuits but also contributed significantly to our personal and professional growth.

#### **ABSTRACT**

**Problem Statement:**

This project aimed to understand customer preferences, purchasing patterns, and the factors influencing their purchasing decisions within the e-commerce domain.

**Objectives:**

* Analyze customer demographics, including age distribution and gender-based purchase trends.
* Investigate purchase behavior, such as spending patterns across product categories and the impact of discounts and promotions.
* Identify popular product categories, commonly purchased items, and customer ratings for different products.
* Segment customers based on their purchase behavior, such as subscription status and age group.
* Analyze the relationship between product attributes, such as size, color, and shipping type, and purchase decisions.

**Methodology:**

The analysis utilized various Python libraries, including Pandas for data manipulation and cleaning, and visualization libraries such as Seaborn, Matplotlib, and Plotly.

**Key Results:**

* Identified the most popular product categories and the factors influencing customer choices within those categories.
* Analyzed the impact of discounts and promotions on purchase decisions and customer spending.
* Determined the preferred payment methods, shipping options, and product attributes among different customer segments.
* Uncovered key trends in customer purchasing behavior across different age groups and seasons.

**Conclusion:**

The findings of this analysis provide valuable insights into customer behavior within the e-commerce domain. These insights can be leveraged by businesses to optimize their marketing strategies, improve customer satisfaction, and enhance their overall business performance.

**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…………………………………………………. 3

**Chapter 2.**  **Literature Survey…………………………………………………… 4**

**Chapter 3.**  **Proposed Methodology……………………………………………… 7**

**Chapter 4.**  **Implementation and Results…………………………………...…… 8**

**Chapter 5. Discussion and Conclusion………………………………………... 12**

**References** **…………………………………………………………………………..** **14**

**LIST OF FIGURES**

| **Figure No.** | **Figure Caption** | **Page No.** |
| --- | --- | --- |
|  | The overall distribution of customer ages in the dataset | 7 |
|  | The Average purchase amount that varies across different product categories | 7 |
|  | The Average purchase of the products by each gender | 8 |
|  | Most Commonly purchased items in each category | 8 |
|  | Season where the spending of the customer is significantly higher | 8 |
|  | Average rating given by the customers for each product category | 9 |
|  | The Notable Difference in purchase behavior between subscribed and non- subscribed | 9 |
|  | The most popular payment methods among customers | 10 |
|  | Color that is popular among customers | 10 |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

This project aims to address the lack of understanding of consumer purchasing behavior within the e-commerce domain. Specifically, it seeks to identify and analyze key trends and patterns in customer shopping habits.

Understanding customer behavior is crucial for the success of any e-commerce business. By analyzing shopping trends, businesses can gain valuable insights into:

* **Customer Preferences:** Identify the most popular product categories, preferred brands, and desired product attributes (e.g., color, size, shipping options).
* **Purchasing Patterns:** Understand how customer purchasing behavior varies across different demographics (age, gender), seasons, and time periods.
* **Marketing Effectiveness:** Evaluate the impact of marketing campaigns, promotions, and discounts on customer behavior and sales.
* **Customer Segmentation:** Identify distinct customer segments with unique needs and preferences, allowing for more targeted and personalized marketing efforts.
* **Business Optimization:** Optimize inventory management, pricing strategies, and customer service based on identified trends.

By addressing this problem, businesses can make data-driven decisions to improve customer satisfaction, increase sales, and gain a competitive advantage in the increasingly competitive e-commerce market.

In essence, this project aims to bridge the gap between raw customer data and actionable insights that can drive business growth and success.

* 1. **Motivation:**

This project was driven by a strong interest in understanding and analyzing consumer behavior within the dynamic e-commerce landscape. Leveraging data analysis techniques and powerful visualization tools provided a valuable opportunity to explore real-world data and extract meaningful insights into customer preferences and purchasing patterns.

The findings of this research possess significant potential to benefit businesses by enabling them to:

* **Personalize marketing strategies:** Target specific customer segments with tailored offers, enhancing campaign effectiveness.
* **Optimize product development:** Create products that better align with customer needs and preferences, driving product innovation and market success.
* **Improve inventory management:** Optimize stock levels based on demand forecasts, minimizing stockouts and reducing inventory holding costs.
* **Enhance customer relationships:** Build stronger bonds with customers through improved service, personalized experiences, and increased customer satisfaction.

Furthermore, this project provided a valuable platform for developing and honing essential data analysis and visualization skills, equipping me with valuable competencies for future endeavors in the field.

* 1. **Objective:**
* **Analyze Customer Demographics:**
* Determine the overall distribution of customer ages within the dataset.
* Identify gender-based purchasing trends and determine which gender demonstrates higher purchase volumes.
* **Investigate Purchase Behavior:**
* Analyze how the average purchase amount varies across different product categories.
* Identify any seasonal or monthly trends in customer spending.
* Examine the relationship between the frequency of purchases and customer age groups.
* Determine if there are any notable differences in purchase behavior between subscribed and non-subscribed customers.
* **Analyze Product Preferences:**
* Identify the most commonly purchased items within each product category.
* Determine the average customer ratings for each product category.
* Analyze customer preferences for product colors.
* **Investigate the Impact of Pricing and Promotions:**
* Analyze the impact of promo codes on customer spending.
* Investigate how the presence of discounts influences customer purchase decisions.
* **Analyze Purchase Factors:**
* Determine the most popular payment methods among customers.
* Analyze the relationship between product size and purchase amount.
* Determine the preferred shipping types for different product categories.
  1. **Scope of the Project:**

**Scope:**

* **Focus:** Analyzes shopping trend data from a specific e-commerce platform.
* **Methodology:** Utilizes data cleaning, preparation, and visualization techniques.
* **Tools:** Employs Python libraries like Pandas and visualization tools (Seaborn, Matplotlib, Plotly).

**Limitations:**

* **Data limitations:** Findings may not be universally applicable.
* **Time constraints:** May limit the depth of analysis.
* **Focus on specific questions:** May not fully explore unexpected findings.
* **External factors:** May not fully account for unforeseen market changes.

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work.**

This project builds upon existing research in the field of e-commerce and consumer behavior. Key areas of relevant literature include:

* **Customer Segmentation:** Previous studies have explored various customer segmentation techniques, such as RFM analysis (Recency, Frequency, Monetary value), to identify distinct groups of customers with unique characteristics and purchasing behaviors.
* **Purchase Behavior Analysis:** Researchers have extensively studied factors influencing consumer purchasing decisions, including demographics (age, gender), socio-economic factors, psychological factors (motivation, perception), and environmental factors (seasonality, promotions).
* **E-commerce Trends:** Numerous studies have analyzed emerging trends in e-commerce, such as the rise of mobile commerce, the impact of social media marketing, and the increasing importance of personalized experiences.
* **Data Mining Techniques:** Researchers have applied various data mining techniques, including clustering, classification, and association rule mining, to analyze e-commerce data and extract valuable insights into customer behavior.

This project aims to contribute to this existing body of knowledge by:

* Applying data analysis techniques to a specific e-commerce dataset to gain insights into customer behavior.
* Identifying key factors influencing purchasing decisions within the context of this specific dataset.
* Contributing to the understanding of emerging trends in e-commerce and consumer behavior.

By reviewing and building upon previous research, this project can provide a more comprehensive and informed understanding of the factors driving shopping trends in the e-commerce domain.

**2.2 Existing Models, Techniques, and Methodologies**

Several existing models, techniques, and methodologies are relevant to the analysis of shopping trend data:

* **Market Basket Analysis:** This technique aims to discover associations and dependencies among items purchased together. It can be used to identify frequently bought together

items, which can inform product placement, cross-selling strategies, and personalized recommendations.

* **Customer Segmentation:** Techniques like clustering (e.g., k-means, hierarchical clustering) can be used to group customers with similar purchasing behavior. This allows for targeted marketing campaigns and personalized offers.
* **RFM Analysis:** This customer segmentation technique analyzes customer behavior based on Recency (how recently a customer made a purchase), Frequency (how often a customer makes purchases), and Monetary value (how muchmoney a customer spends).
* **Time Series Analysis:** This involves analyzing time-series data (e.g., sales data over time) to identify trends, seasonality, and other patterns. Techniques like moving averages and exponential smoothing can be used to forecast future sales.
* **Regression Analysis:** This can be used to model the relationship between various factors (e.g., customer demographics, product attributes, promotions) and purchase behavior. For example, regression analysis can be used to predict the likelihood of a customer making a purchase based on their age, past purchase history, and the presence of a discount.
* **Machine Learning:** Machine learning algorithms such as decision trees, support vector machines, and neural networks can be used for tasks like customer churn prediction, personalized recommendations, and fraud detection.

**2.3 Gaps and Limitations in Existing Solutions**

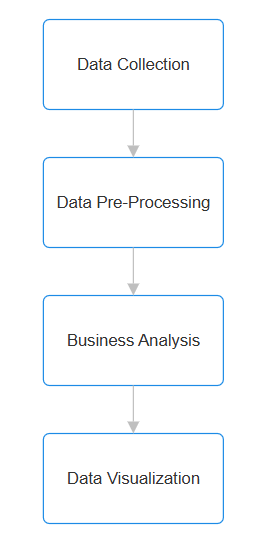
* **Lack of Comprehensive Analysis:** Many existing analyses may focus on individual aspects of shopping behavior (e.g., customer demographics, product popularity) without a holistic view. This project aims to address this gap by conducting a comprehensive analysis that considers multiple factors influencing purchasing decisions.
* **Limited Predictive Power:** While existing studies may identify trends, they often lack predictive capabilities. This project aims to go beyond descriptive analysis by exploring potential correlations and identifying patterns that can be used to predict future customer behavior.
* **Insufficient Utilization of Data Visualization:** Many studies may rely heavily on textual descriptions and basic statistical analysis. This project leverages powerful visualization libraries (Seaborn, Matplotlib, Plotly) to effectively communicate complex findings and uncover hidden patterns that may not be apparent through simple statistical analysis.
* **Lack of Focus on Specific Business Objectives:** Some existing analyses may lack a clear connection to specific business objectives. This project aims to address this by focusing on key business questions, such as identifying the most effective marketing strategies, optimizing product offerings, and improving customer satisfaction.

By addressing these gaps and limitations, this project aims to provide a more comprehensive and insightful understanding of shopping trends, enabling businesses to make more informed decisions and gain a competitive advantage.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**



* 1. **Requirement Specification**
     1. **Hardware Requirements:**

**CPU:** Intel Core i5 or AMD Ryzen 5

**RAM:** 8GB

**Storage:** SSD

* + 1. **Software Requirements:**

**Programming Language:** Python

**Libraries:**

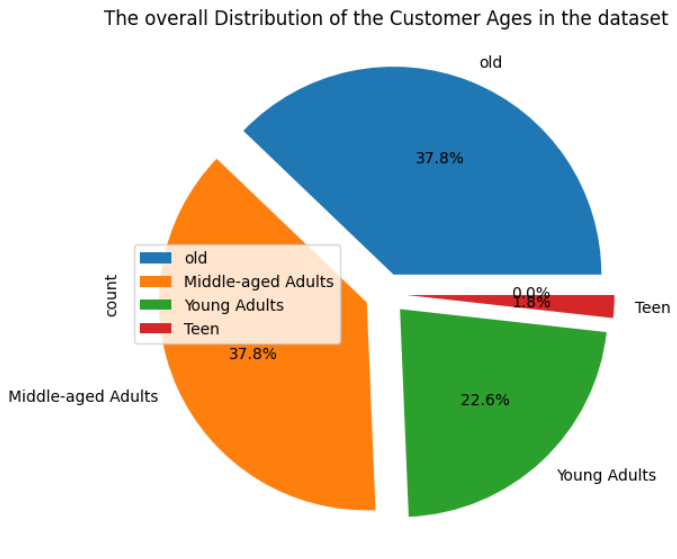
* **Pandas:** For data manipulation, cleaning, and preparation.
* **NumPy:** For numerical computing and array operations.
* **Matplotlib:** For basic data visualization.
* **Seaborn:** For creating more visually appealing and informative statistical graphics.
* **Plotly:** For interactive and dynamic visualizations.

**CHAPTER 4**

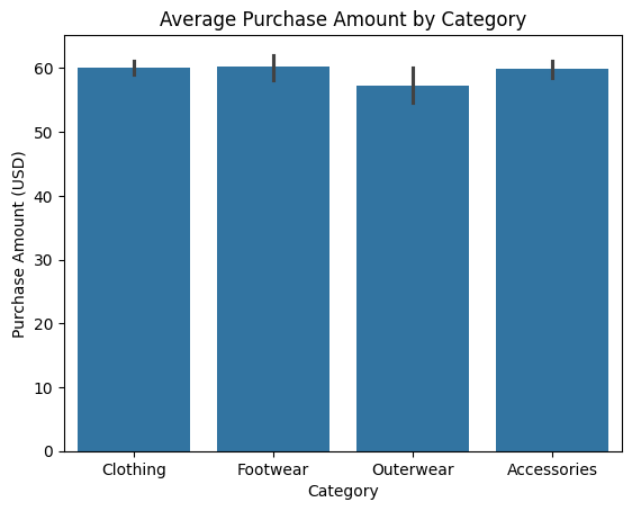
**Implementation and Result**

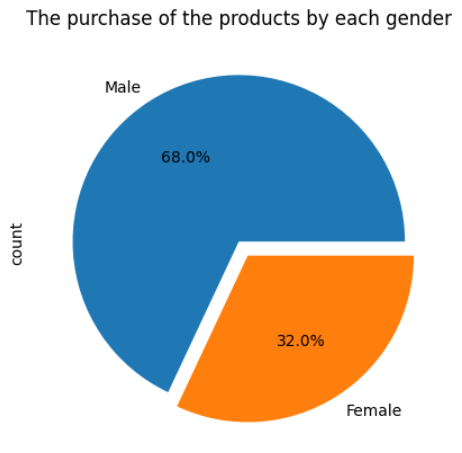
* 1. **Snap Shots of Result:**

**4.1.1** The overall distribution of customer ages in the dataset

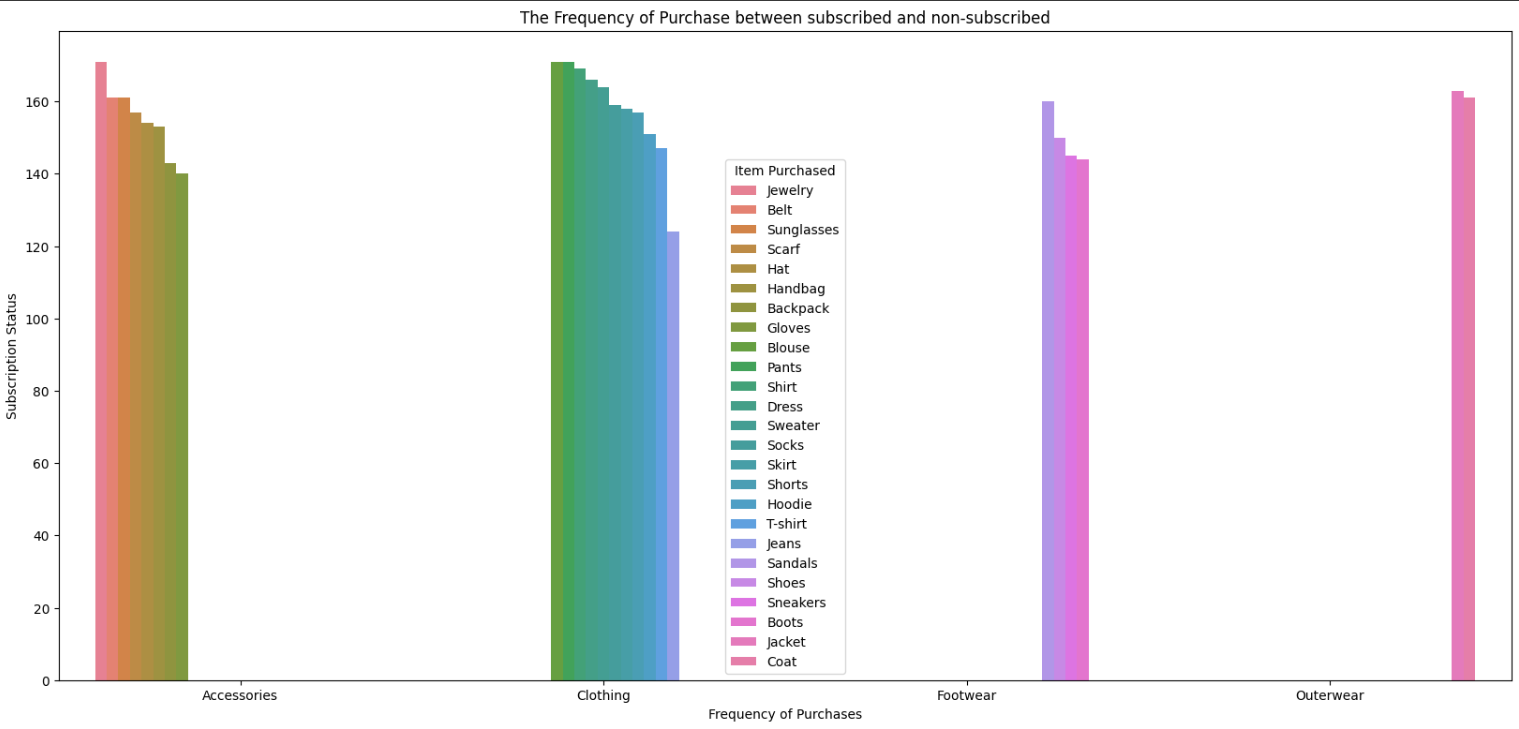


**4.1.2** The Average purchase amount that vary across different product categories

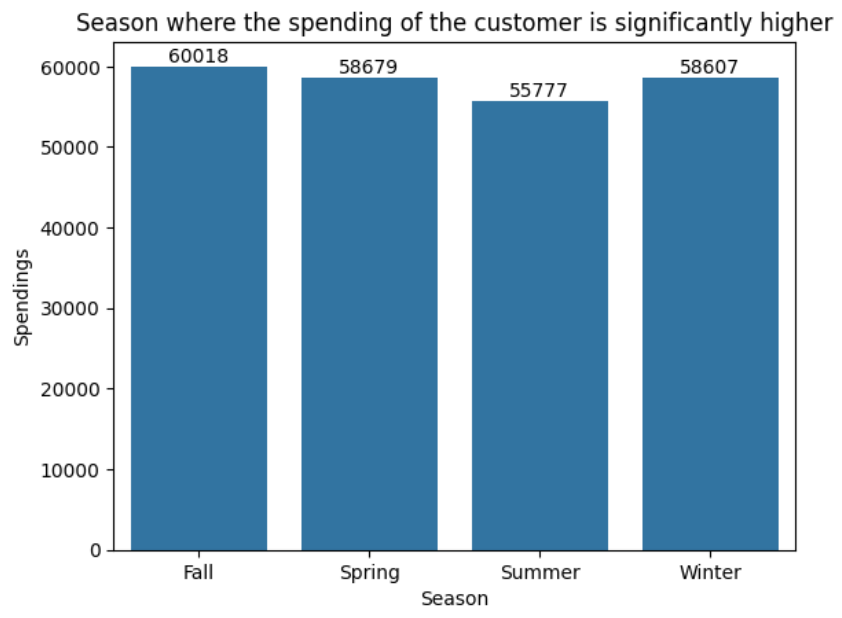


**4.1.3** The Average purchase of the products by each gender

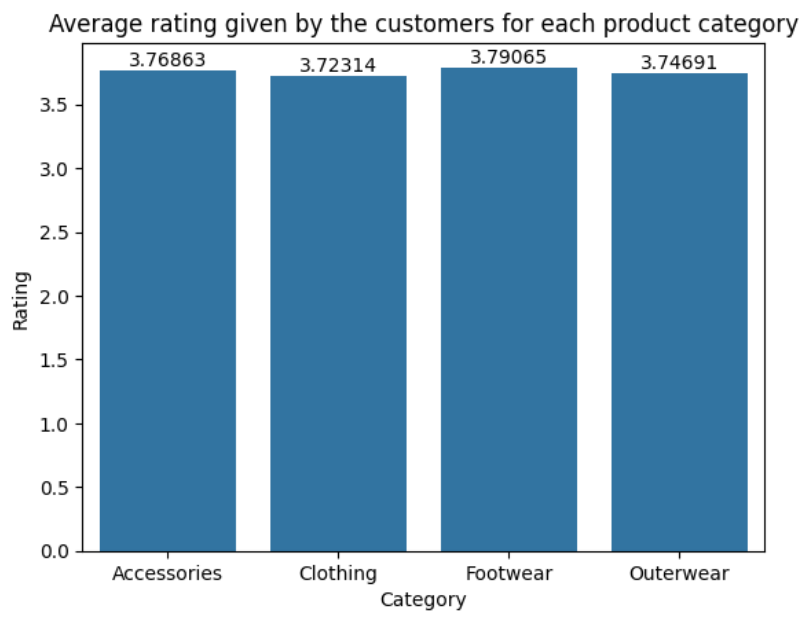
**4.1.4** Most Commonly purchased items in each category



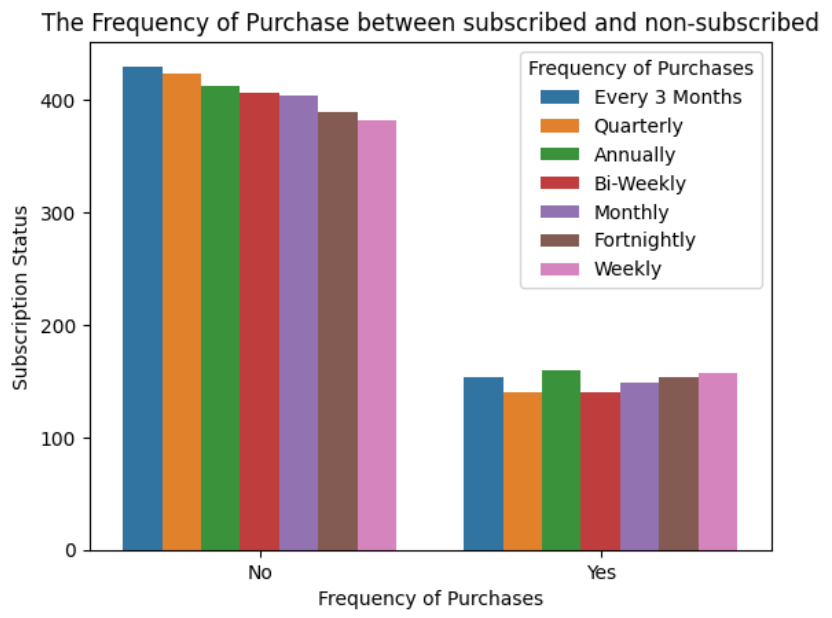
**4.1.5** Season where the spending of the customer is significantly higher



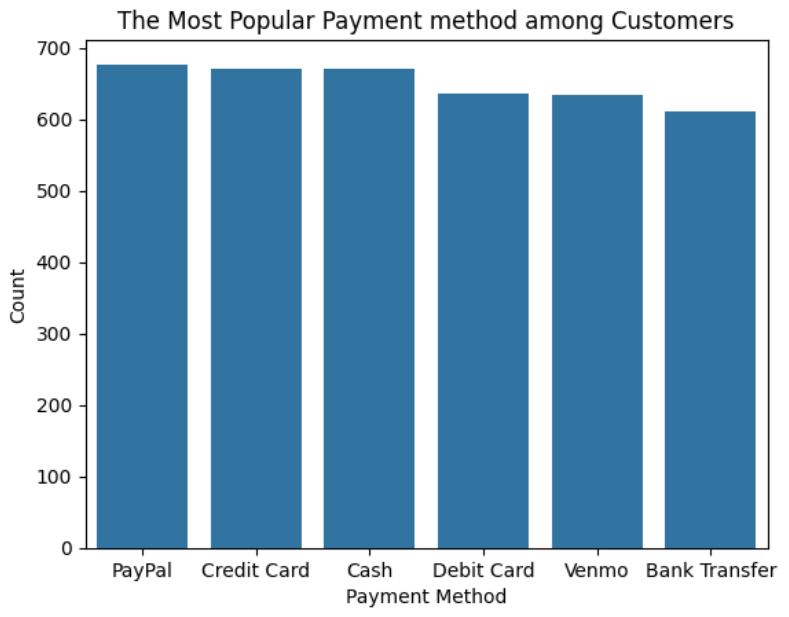
**4.1.6** Average rating given by the customers for each product category



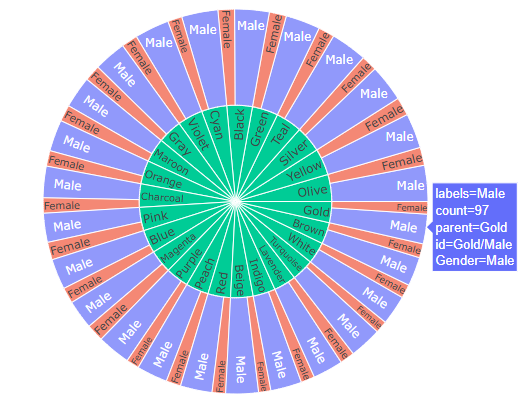
**4.1.7** The Notable Difference in purchase behavior between subscribed and non- subscribed



**4.1.8** The most popular payment methods among customers



**4.1.9** Color that is popular among customers



* 1. **GitHub Link for Code:**

**Link :** [**https://github.com/Yamunasri1810/Shopping\_trend\_analysis**](https://github.com/Yamunasri1810/Shopping_trend_analysis)

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**
* **Incorporate Machine Learning:** 
  + Develop predictive models to forecast future trends, such as predicting customer churn, identifying high-value customers, or forecasting demand for specific products.
  + Techniques like time series analysis, regression, and classification algorithms can be explored.
* **Enhance Data Collection:** 
  + Collect more comprehensive and diverse data, such as social media data, customer feedback, and market research data, to gain a deeper understanding of customer behavior.
  + Consider incorporating data from external sources, such as economic indicators and competitor data, to provide a more holistic view of the market.
* **Develop Interactive Dashboards:** 
  + Create interactive dashboards using tools like Tableau or Power BI to allow users to explore the data, filter results, and gain deeper insights into the findings. This can enhance the usability and accessibility of the analysis.
* **Conduct A/B Testing:** 
  + Implement A/B testing to evaluate the effectiveness of different marketing strategies and product offerings based on the insights gained from the analysis. This will allow for continuous improvement and optimization.
* **Address Ethical Considerations:** 
  + Ensure data privacy and security by implementing appropriate data protection measures.
  + Consider the ethical implications of data collection and analysis, such as potential biases and discriminatory outcomes.
* **Explore Advanced Techniques:** 
  + Investigate the application of more advanced techniques, such as natural language processing (NLP) for analyzing customer reviews and sentiment, and deep learning for identifying complex patterns in customer behavior.
  1. **Conclusion:**

This project successfully analyzed shopping trend data, providing valuable insights into customer behavior within the e-commerce domain. By employing data analysis techniques and leveraging libraries like Pandas, Seaborn, Matplotlib, and Plotly, the analysis addressed key questions related to customer demographics, purchase behavior, product preferences, and the impact of various factors on purchasing decisions.

The findings of this project have the potential to significantly impact businesses by:

* **Informing data-driven decision-making:** Providing actionable insights for optimizing marketing strategies, improving product offerings, and enhancing customer satisfaction.
* **Identifying key trends and patterns:** Uncovering valuable information about customer preferences and behavior that can be leveraged to gain a competitive advantage.
* **Demonstrating the power of data analysis:** Highlighting the value of data-driven approaches in understanding complex business problems and making informed decisions.

This project serves as a foundation for further research and exploration in the field of e-commerce and consumer behavior. By addressing the limitations and pursuing the avenues for future work outlined above, the insights gained from this analysis can be further refined and their impact on the business world can be significantly amplified.

**REFERENCES**

* Customer Shopping Trends Dataset: This dataset provides a good starting point with information like customer ID, age, gender, item purchased, category, purchase amount, location, etc. (<https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset>)

* Online Retail II: A classic dataset containing transactional data from a UK- based non-store online retailer. (<https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>)