**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

**Atharva Sakpal , atharvasakpal2715@gmail.com**

Under the Guidance of

**Pavan Sumohana**

**ACKNOWLEDGEMENT**

My special thanks go to Mr. Pavan Sumohana for his unstinting support, perfect guidance and encouragement throughout this project. We shaped up this work under his leadership as he is knowledgeable, insightful, adroit and committed to excellence.

His mentorship gave me the confidence and clarity to overcome challenges and developed my technical and analytical skills. His positive remarks helped me improve this project by enhancing my knowledge of data analysis and developing the concept. His explanation of a concept helped me understand more about it and also influence the quality of this project.

With gratitude to his encouragement mind, tackle a problem in a creative way, find different ways of solving a problem and to inspire me to keep higher standards in my work. This particular project has definitely taught me a lot and his guidance has left a mark on me academically and professionally.

I am truly thankful to Edunet Foundation for providing the necessary resources and help for the successful completion of the project.

The knowledge and learning infrastructure they helped establish is very useful to overcome problems and accomplish the goals of the project.

This project was possible because of the cooperation and resources given by Mr Pavan Sumohana and Edunet Foundation. I am motivated by their dedication to work and I hope to continue the actions I have taken while on this journey.

Appreciate you all for making this journey worthwhile and enriching.

#### **ABSTRACT**

This project is mainly based on analyzing the shopping trends of the customers in different aspects and situations. The primary objective of this project was to take raw data and identify the key trends which will help in enhancing the customer satisfaction and improve the revenue.

In this project I had to identify shopping trends using data analysis for identification of unique and repetitive patterns of shopping of customers. This will in fact help in increasing the revenue and improving the customer experience.

The methodology involves the development of a Python script which uses different libraries such as Pandas, NumPy, Matplotlib and Seaborn. This script offers several functionalities which include sales tracking, customer segmentation and product insights. And this script was divided into several functions which had unique operations to be performed. This is a menu driven script which was done to make it more user-friendly to access these analyses.

The primary analysis from this project were the purchasing behavior of the customer, popular products, average spending of the consumers, and customer preferences. This insight helps to develop targeted marketing and increase the revenue.

In conclusion, this project demonstrates how Python based tools can efficiently analyze the given data and can give the results we desire and which can help us in decision making and operating efficiently.

**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 2

**Chapter 2.**  **Literature Survey 3**

**Chapter 3.**  **Proposed Methodology**

**Chapter 4.**  **Implementation and Results**

**Chapter 5. Discussion and Conclusion**

**References**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Caption** | **Page No.** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

In retail industry, it is important to understand customer pattern and to develop unique marketing strategy. To accomplish all of this, businesses face a huge challenge to derive some insights and patterns. The problem lies with analyzing the data such as purchasing trends and preferences effectively and efficiently.

This problem is highly significant because customer expectations are evolving constantly. By addressing this issue, we can enhance the customer experience, improve the inventory as per the customer’s needs, increase revenue by understanding the impact of discounts and promo codes and by marketing strategically.

By solving this problem, the gap between raw data and useful insights can be filled.

* 1. **Motivation:**

This project was chosen because of the presence of large amount of data generated by daily customer interactions, which can be used to enhance the customer experience as well as increase the profit of businesses. This project provides an opportunity to understand data analysis and how it can be used to solve real world problems.

This project can be used to develop personalized marketing which will be based on customer preferences and their purchasing habits, allows businesses for better inventory planning so that there is less waste generated, and understanding the effect of discounts and promotions to develop a more effective pricing model.

It will impact the revenue of the businesses, help businesses align their operations with customer needs and provide businesses with tools to make data based decisions.

* 1. **Objective:**
* Optimize Inventory: to provide actionable insights for inventory planning so that there is less amount of waste generated.
* Understanding Purchasing Patterns: to identify product preferences such as categories, sizes and colors, and their correlation with seasonal factors.
* Asses Discounts: to observe how the presence of discounts and promo codes influence the purchase amount and frequency of the customer.
* Identifying Popular Payment Methods and Shipping Preference: to determine which mode of payment is most favorable and the type of shipping they prefer to enhance convenience.
* Leverage Data Visualization: to use graphs like histograms, bar graph and pie charts to make the findings more accessible.
* Develop Data Analysis Skills: To deeply understand data analysis and visualization techniques in Python and tools like Pandas, NumPy, Seaborn and Matplotlib.
  1. **Scope of the Project:**
* Visualization of trends: the use of graphical tools such as pie chart, bar graphs and histogram makes it easier to understand the complex data.
* Data Based Decision Making: by analyzing the spending patterns, preferred payment methods helps in developing key strategies which could be beneficial for the businesses.
* Evaluating Customer Loyalty: it displays the focus of project on understanding the loyalty of the customer.
* Skill Development: This project gives hands-on experience on important tools like Python and it’s libraries such as Pandas, Matplotlib, Seaborn and NumPy.

But there are certain limitations to this project such as:

* Resource Availability: This project assumes access to majority of the necessary tools and resources which might not be the case in real world scenarios.
* Applicability Challenges: the insights developed in this project might not be applicable to all types of businesses, because the type and preferences of the customers may vary.
* Dependency on Dataset: the analysis developed in this are solely dependent on the dataset provided so the absence of any data will affect the results generated.
* Limited Features: As there can be endless number of factors which can be taken into consideration to analyze the data but taking all factors in account is not possible.

.

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work in this domain.**

For decades now, scholars and corporate practitioners have conducted a lot of research to try and get a clear understanding of customer shopping behaviour, explore what determines a purchase decision, and seek ways to create a better shopping experience. This has given them valuable insights that informed this project. Some key areas examined in earlier work are as follows:

* Getting to Know Customers More: Historically, companies have used demographic or purchase-based segmentation to group customers in a way that adapts their businesses accordingly. For example, aggregating similar customer profiles enables the creation of personalized recommendations and targeted advertising.
* How Customers Pay: As trend shows, there is a significant increase in the use of digital wallets and contactless payments. Being aware of these changes makes it easier and more convenient for businesses to conduct transactions.
* The Role of Reviews: Customer buying decisions can be greatly influenced by product reviews and ratings. Positive feedback can lead to increased sales, while negative reviews point out room for improvement for businesses.
* Seasonal Shopping Trends: businesses have spent a lot of time in understanding how a particular season or a festival can affect the buying behaviour. They use this analysis to stock right products onto their shelves.
* Discounts and Promo Codes: Research has continuously shown how discounts and promo codes have made the customers to purchase certain products they might not have otherwise.
* Building Loyalty: these programs often show that rewarding the customer for their business keeps them coming back, they increase the number of purchase and spend more.
  1. **Mention any existing models, techniques, or methodologies related to the problem.**

In shopping trend analysis there are several existing models, techniques and methodologies which have been developed. These are crucial for getting important insights from large datasets and enable businesses to make more data driven decisions. Some models and methodologies related to this problem are:

* Predictive Analysis Model: Decision Trees and Forests are mainly used in predicting customer behavior based on purchasing patterns. They help businesses understand the customer pattern and plan accordingly.
* Customer Segmentation Model: RFM analysis (Recency, Frequency, Monetary), this model helps in understanding the customer loyalty, their purchase amount and how often they purchase.
* Market Basket Analysis: it is often used to understand which products are purchased together and helps in optimizing product placements.
* Recommendation Systems: This method suggests customers the products according to their previous purchases and preferences of similar users.
* Time Series Analysis: this method helps in analyzing the trends in customer purchases as per the seasons and festivals.
* A/B testing: This technique is widely used to find out a more effective marketing strategy such as pricing models and promotional offers. By comparing two strategies, the businesses can adopt the more efficient strategy.
* Sentiment Analysis: is used to understand the product reviews by using Natural Language Processing (NLP) techniques which also understands the product issues.

All these models, techniques and methodologies help us in understanding the customer patterns and offer to enhance the customer experience, increase the profits and optimize the marketing strategies.

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

As there are numerous models, methodologies and techniques, there exists certain gaps or limitations. These gaps often arise due to the presence of large complex real world data and inability to handle these huge and diverse datasets. Some of the limitations and how this project will address them are as shown:

* Limited Customization for specific businesses:

Gap: Many models such as K-means clustering and RFM analysis may not fully capture the purchasing behavior of customers in different industries.

How this project addresses it: This project will analyze the product categories and customer demographics. By including multiple features such as subscription status, shipping type etc. ,it will provide a more personalized insights for businesses which will allow them to develop customized strategies.

* Inability to handle multi-dimensional analysis:

Gap: many models only analyze a single thing such as age of customers or purchase amounts. This limits the ability to develop useful insights.

How this project addresses it: This project uses a multi-dimensional analysis approach by considering various factors such as gender, location, product category and purchase amount and how these factors interact with each other.

* Difficulty in Handling Complex Customer Behavior Pattern:

Gap: Many models struggle with complex trends that can get generated from large data sets.

How this project addresses it: This projects uses combination of statistical analysis and machine learning to detect a more complex pattern in data, such as relationship between review rating and purchasing decisions or how promo codes affect customer loyalty.

* Limited Focus on Customer Loyalty:

Gap: Many existing solutions mainly focus on one-time transaction or short-term behavior without considering long term customer loyalty.

How this project addresses it: This project focuses on factors like Subscription Status, Previous purchases and frequency of purchase to understand loyalty of the customer. This allows businesses to develop strategies which can benefit the long term customers as well as maintain their revenue.

This project fills these gaps by giving a more detailed and flexible approach to analyzing the shopping trends. This provides businesses with tools to improve decision making and optimize their strategies for more customer engagement.

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

****

The diagram represents the workflow for the analysis of shopping trends. It is divided into 6 steps, which are:

1. Input Data: Initially we read the CSV file, which contains all the important information such as purchase history, product preferences, location details, etc.
2. Data Preprocessing: The raw data is cleaned thoroughly to remove any inconsistency and extracts unique values from columns such as Gender, Category and Payment Method. This ensures that dataset is ready for further analysis.
3. Exploring Data: We need to examine the distribution of data using different types of graphs and then find useful insights such as customer location, their purchasing behavior and popularity of products.
4. Analysis Model: Customer Behavior Analysis, based on understanding subscription patterns and gender based purchases. Product Analysis, helps in determining product categories and effect of discounts. Financial Analysis, finds out the purchasing patterns, frequency of purchases and impact of discounts and promo codes.
5. Visualization: Tools like histogram, pie charts and bar graphs are used to convert complex data into readable form which will make the insights more understandable.
6. Interactive User Input: This allows user to select any specific analysis according to their requirement from a variety of options.

**3.2 Requirement Specification**

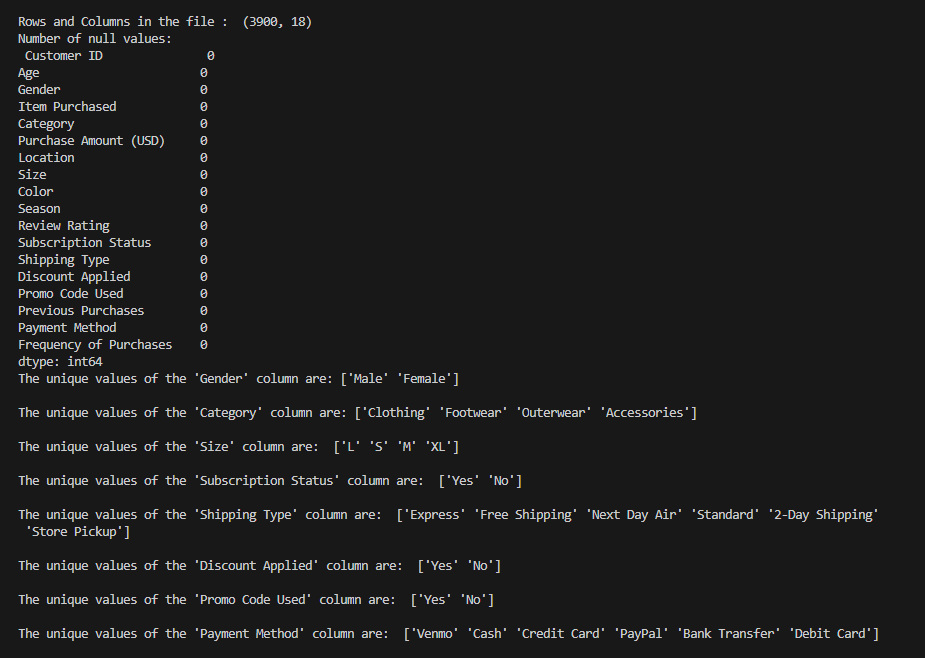
Tools required to make this project are as follows:

1. Python: Core implementation is written in Python as there are numerous libraries which are useful in this project.
2. Analysis Library: Libraries such as Pandas and NumPy are mainly used for data manipulation and for performing numerical operations.
3. Visualization Library: Libraries such as Matplotlib and Seaborn are used to provide graphical representation of data to make it easier to understand.
4. Data Storage: Input is taken in CSV file and later on this file is converted into Excel file.
5. IDE: Tools like VS Code, Jupyter Notebook can be used to write and debug the code.
6. Environment and Dependencies: Latest version of Python has to be installed and ‘pip’ is used to install all the necessary libraries required for this project.
   * 1. **Hardware Requirements:**
7. Processor: Quad-core processor such as Intel Core i5/i7 or AMD Ryzen is recommended.
8. RAM: minimum 4GB is required but 8GB or higher is recommended.
9. Storage: SSD with maximum 10GB free space is required to get better performance.
10. Operating System: Compatible with Windows, macOS and Linux.
11. Optional Hardware: External monitor for better visualization and multitasking. GPU for accelerated data processing.
    * 1. **Software Requirements:**
12. Programming Language: Python 3.7 or latest version.
13. Python Libraries: Pandas, NumPy, Matplotlib and Seaborn.
14. Development Environment: Visual Studio Code, Jupyter or PyCharm.
15. Dependency Management: pip to install all the libraries.
16. Operating System: Windows 10/11, macOS or Linux.
17. Data file requirements: ‘shopping\_trends\_updated1.csv’ file as input.

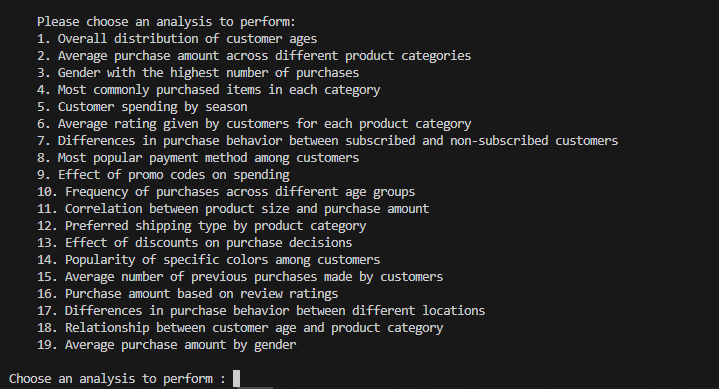
**CHAPTER 4**

**Implementation and Result**

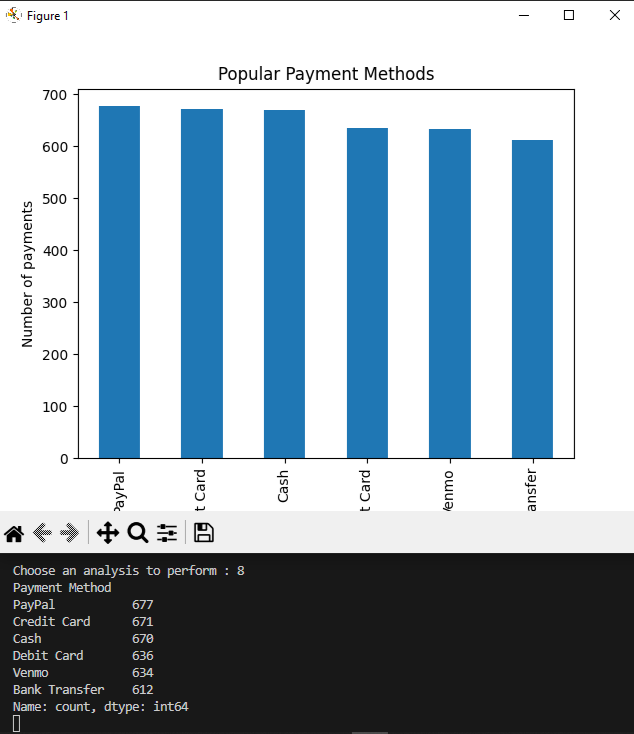
* 1. **Snap Shots of Result:**

. 

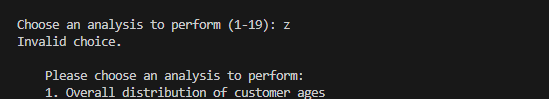
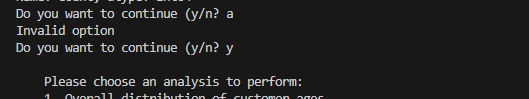
In the beginning of the program, it will give us all the initial information of the given dataset such as total number of rows and columns, number of null values and all the unique parameters in the dataset. Which will be useful in analyzing the dataset.



In the second image, number of operations is displayed or the type of analysis the user wants to perform. This gives a more user-friendly experience, which allows them to understand it clearly.



In the third image, an analysis was chosen to find out the most popular payment method among the customers. This analysis will show us the number of customers who use different payment methods as well as a graph to show the variation across all payment methods.



This program is also designed to handle invalid inputs given by the user, which displays a message saying “Invalid Option”.

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

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

To improve the given model, here are some suggestions:

1. Data Cleaning: Ensure that the dataset being used is cleaned thoroughly before processing. It can be done by creating automated data cleaning methods.
2. Data Visualizations: Implement features which allows user to interact with visualizations for a deeper analysis.
3. Scalability: Ensure that the model can handle large subsets by using optimization techniques such as parallel processing or using a database management system to sore data.
4. Advanced Analytical Features: Add more features such as predictive modeling for future purchases.
5. User Interface: Build a web or desktop application that allows user to interact with data and visualize results without writing a code.
6. Performance Optimization: Implement a more efficient algorithms to handle large datasets quickly when running more complex analysis.
   1. **Conclusion:**

The overall impact of this project is summarized as follows:

1. Data-driven insights: This project provides a thorough analysis of customer shopping data, allowing businesses to understand customer behavior and purchasing patterns. This can be used to develop targeted strategies.
2. Interactive tool: The menu-driven interface allows user to choose the analysis they want to perform based on their needs.
3. Improved Decision Making: The insights developed in this project will greatly affect the businesses by increasing their revenue and understanding how the presence of discounts and promo code affect the customer behavior.
4. Scalability: It is capable to handle big datasets and can be integrated with machine learning for predictive analysis and improve the visualization.
5. Contribution to Data Science Practices: This project demonstrates the application of data analysis, which can provide a solid foundation for future work in data science. It can be used as a stepping stone for creating a more advanced models.

**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.