**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

**SAKSHI PANDEY, ps191111989@gmail.com**

Under the Guidance of

**P. Raja Sir**

**Jay Rathod Sir**

**ACKNOWLEDGEMENT**

We would like to take this opportunity to extend our heartfelt gratitude to everyone who has supported and guided us, directly or indirectly, throughout the course of this thesis work.

First and foremost, we wish to express our sincere appreciation to our supervisor, P. Raja Sir and Jay Rathod Sir, for being an exceptional mentor and advisor. His invaluable guidance, encouragement, and constructive feedback have been a constant source of innovative ideas and inspiration, playing a pivotal role in the successful completion of this project. His unwavering confidence in my abilities has been a driving force throughout this journey.

Working under his mentorship over the past weeks has been an incredible privilege. His assistance not only contributed significantly to the progress of this project but also provided insights into various aspects of the program. His teachings have not only enriched my professional skills but also shaped me into a more responsible and capable individual.

I would like to thanks my family members for supporting me, helping me to maintain my mental and physical health during this internship.

#### **ABSTRACT**

**Identifying Shopping Trend Analysis with Exploratory Data Analysis (EDA)**

In today’s competitive retail landscape, understanding customer shopping trends is critical for businesses to make informed decisions and enhance their offerings. This project focuses on analysing shopping trends using a dataset comprising customer demographics, purchase behaviour, product preferences, and transaction details. The objective is to uncover valuable insights that can help businesses improve customer experience, optimize inventory, and boost sales.

The study begins by identifying key questions, such as identifying popular product categories, evaluating seasonal shopping patterns, and examining customer preferences based on demographics. Using Exploratory Data Analysis (EDA), the dataset is cleaned and visualized to identify patterns, correlations, and anomalies. Techniques such as summary statistics, data visualization, and segmentation analysis are employed to extract actionable insights.

Key findings include identifying the most frequently purchased product categories, variations in spending across genders and age groups, and the impact of discounts and promotions on purchase decisions. The analysis also highlights seasonal spikes in sales and preferences for specific payment and shipping methods.

The results provide actionable recommendations for businesses to tailor marketing strategies, refine inventory management, and create personalized shopping experiences. By understanding the relationship between customer demographics and purchasing behaviour, businesses can achieve better customer satisfaction and increased revenue.

This project investigates patterns and trends in customer shopping behaviour using a comprehensive dataset. The dataset comprises various attributes, including customer demographics, purchase details, and preferences. The primary goal is to analyse this data to uncover actionable insights, such as spending habits, category preferences, and seasonal trends. These findings aim to assist retailers in:

* Enhancing customer engagement.
* Optimizing inventory management.
* Tailoring marketing strategies to specific demographics and behaviours.

In conclusion, this project demonstrates the power of EDA in uncovering hidden trends and provides a foundation for more advanced predictive modelling or recommendation systems in future work.

**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.** |
|  | SYSTEM DESIGN | **5** |
|  | AGE DISTRUBUTION BY CATEGORY | **14** |
|  | AVERAGE PURCHASE AMOUNT BY CATEGORY | **14** |
|  | AVERAGE PURCHASE AMOUNT BY GENDER | **15** |
|  | AVERAGE PURCHASE AMOUNT BY GENDER AND CATEGORY | **15** |
|  | NUMBER OF PURCHASES BY GENDER | **16** |
|  | TOTAL NUMBER OF PURCHASES BY CATEGORY | **16** |
|  | COMMONLY PURCHASED ITEM IN EACH CATEGORY | **17** |
|  | TOTAL PURCHASE BY SEASON | **17** |
|  | DISTRIBUTION OF PURCHASES BY SEASON (PIE CHART) | **17** |
|  | AVERAGE RATING BY CATEGORY | **17** |
|  | REVIEW RATING BY CATEGORY (LINE GRAPH) | **18** |
|  | SUBSCRIBTION PURCHASE STATUS | **18** |
|  | AVERAGE PURCHASE AMOUNT BY PAYMENT METHOD (HISTO GRAPH) | **18** |
|  | AVERAGE PURCHASE AMOUNT BY PAYMENT METHOD | **19** |
|  | AVERAGE PURCHASE AMOUNT BY PAYMENT METHOD (SCATTER GRAPH) | **19** |
|  | TOTAL PURCHASE BY GENDER AND PROMO CODE | **20** |
|  | TOTAL PURCHASE BY PROMO CODE USED | **20** |
|  | FREQUENCY OF PURCHASES BY AGE CATEGORY (SUNBURST CHART) | **20** |
|  | FREQUENCY OF PURCHASES BY AGE CATEGORY (TREE MAP) | **20** |
|  | TOTAL PURCHASES BY SIZE (BAR GRAPH) | **21** |
|  | DISTRIBUTION OF SHIPPING TYPES BY PRODUCT CATEGORY | **21** |
|  | SHIPPING TYPES BY PRODUCT CATEGORY | **22** |
|  | SHIPPING TYPES BY PRODUCT CATEGORY (HEAT MAP) | **22** |
|  | TOTAL PURCHASES BY DISCOUNT APPLIED | **22** |
|  | TOTAL PURCHASE BY GENDER AND DISCOUNT APPLIED | **22** |
|  | TOTAL PURCHASE BY COLOUR | **23** |
|  | AVERAGE VALUES OF NUMERIC COLUMNS | **23** |
|  | AVERAGE PURCHASE BY REVIEW RATINGS | **23** |
|  | PURCHASE AMOUNT DISTRIBUTION BY REVIEW RATING | **24** |
|  | TOTAL PURCHASE BY LOCATION | **24** |
|  | RELATIONSHIP BETWEEN AGE AND CATEGORY | **24** |
|  | RELATIONSHIP BETWEEN GENDER AND PURCHASE AMOUNT | **24** |
|  | RELATIONSHIP BETWEEN SPENDING VS PREVIOUS PURCHASES (SCATTER GRAPH) | **24** |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Caption** | **Page No.** |
| **Table 1** | DISPLAY FIRST 5 ROWS OF DATA FRAME | **7** |
| **Table 2** | FREQUENCY OF EACH UNIQUE VALUE IN AGE COLUMN | **7** |
| **Table 3** | DISPLAY DATA TYPE OF EACH COLUMN | **8** |
| **Table 4** | AVERAGE PURCHASE AMOUNT OF EACH CATEGORY | **8** |
| **Table 5** | COUNT OF ITEM PURCHASED IN EACH CATEGORY | **9** |
| **Table 6** | GRID TABULAR FORM OF PURCHASE COUNT | **9** |
| **Table 7** | TABLE OF AGE AND AGE CATEGORY | **10** |
| **Table 8** | SHIPPING TYPE TABLE | **10** |
| **Table 9** | GROUP AND COUNT SHIPPING TYPE BY PRODUCT CATEGORY (DESCNDING ORDER) | **11** |
| **Table 10** | MEAN VALUE OF NUMERIC COLUMN | **11** |
| **Table 11** | REVIEW RATING AND PURCHASE AMOUNT | **12** |
| **Table 12** | LOCATION AND PURCASE AMOUNT | **12** |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

In the ever-evolving retail industry, businesses face challenges in understanding customer behavior and shopping trends. Without proper analysis, it is difficult to identify key factors such as popular product categories, seasonal demand fluctuations, and demographic preferences. This lack of insight hinders effective decision-making in marketing, inventory management, and customer engagement strategies. Addressing this problem is significant as it enables businesses to remain competitive, improve customer satisfaction, and increase revenue by adapting to market demands efficiently.

Understanding customer shopping behaviour is crucial for businesses aiming to enhance profitability and customer satisfaction. Despite the availability of vast amounts of customer data, many businesses struggle to extract meaningful insights. This project addresses this gap by:

* Analysing shopping trends and patterns in a structured manner.
* Identifying key factors influencing customer purchase behaviour.
* Providing insights that can guide data-driven decision-making.
  1. **Motivation:**

This project was chosen due to the growing importance of data-driven strategies in the retail sector. With the vast amount of data generated by customers during their shopping journeys, there is immense potential to uncover insights that can transform business operations. The applications of this analysis include optimizing marketing campaigns, enhancing personalized shopping experiences, and predicting future trends. The impact of this project lies in empowering businesses to make informed decisions, reduce operational inefficiencies, and meet customer expectations effectively.

The rapid growth of e-commerce has resulted in an abundance of customer data. However, leveraging this data effectively remains a challenge. This project is motivated by:

* The need for actionable insights to improve customer retention and satisfaction.
* The opportunity to enhance marketing strategies using data analysis.
* The potential to identify trends that can inform inventory and operational decisions.
  1. **Objective:**

The primary objectives of this project are:

* **Customer Behaviour Analysis:** Identify trends in purchase behaviour across demographics.
* **Category Insights:** Analyse preferences and spending patterns for different product categories.
* **Impact of Incentives:** Explore the effect of discounts, promo codes, and subscriptions on purchase behaviour.
* **Seasonal Trends:** Uncover variations in customer behaviour across different seasons.
* **Recommendations:** Provide actionable insights to guide business strategies.
* To analyse shopping trends using a structured dataset, focusing on customer demographics, purchasing behaviour, and transaction details.
* To uncover patterns such as popular products, seasonal sales peaks, and payment method preferences.
* To assess the influence of discounts, promotions, and demographic factors on purchasing decisions.
* To provide actionable insights that help businesses optimize inventory, marketing strategies, and overall customer experience.
  1. **Scope of the Project:**

This project focuses on analysing a dataset containing customer and purchase details. Key areas of focus include:

* Age and gender distribution of customers.
* Spending trends by product category and season.
* The influence of promotional incentives on purchase decisions.
* Customer preferences for shipping types and payment methods.
* Conducting a comprehensive Exploratory Data Analysis (EDA) on the shopping trends dataset.
* Extracting meaningful insights through data cleaning, visualization, and statistical techniques.
* Offering recommendations for businesses to improve decision-making and enhance operations.
* Establishing a foundation for advanced predictive analytics, such as customer segmentation and demand forecasting.

While comprehensive, the project is limited to the provided dataset and does not incorporate real-time data or advanced predictive modeling.

**Limitations:**

* The analysis is limited to the provided dataset and may not account for real-time or external factors influencing shopping trends.
* The project focuses on EDA and does not include building predictive models or implementing solutions.

**CHAPTER 2**

**Literature Survey**

2.1 Review of Relevant Literature

The analysis of shopping trends and customer behaviour has been widely studied in the domains of data analytics, machine learning, and business intelligence. Several studies focus on leveraging data to understand purchase patterns, optimize inventory, and enhance customer experience. Key works in this domain include:

* Customer Segmentation and Behaviour Analysis: Research has shown the effectiveness of clustering algorithms (e.g., k-means, hierarchical clustering) in segmenting customers based on demographics and purchasing behaviour.
* Seasonal Sales Trends: Studies have analysed seasonal variations in retail sales, indicating that demand peaks are often influenced by festivals, holidays, and promotional events.
* Impact of Discounts and Promotions: Numerous studies highlight the role of discounts and promotions in driving customer purchases, with data showing that strategic pricing significantly affects buying decisions.

2.2 Existing Models, Techniques, and Methodologies

Some of the commonly employed techniques in shopping trend analysis include:

1. Descriptive Analytics and EDA: Basic data exploration and visualization tools, such as histograms, scatter plots, and heatmaps, are used to summarize and identify trends in data.
2. Predictive Analytics Models: Algorithms such as regression models, decision trees, and neural networks are applied to forecast customer behaviour or sales.
3. Association Rule Mining: Techniques like Apriori or FP-Growth are utilized to uncover relationships between purchased items, helping to identify cross-selling opportunities.
4. Market Basket Analysis: Widely used to analyse purchase patterns, especially for suggesting complementary products.

2.3 Gaps in Existing Solutions and How This Project Addresses Them

Despite advancements, several gaps exist in current approaches:

* Data Limitations: Many studies focus only on specific datasets or lack comprehensive insights combining demographics, purchasing patterns, and transaction details.
* Real-Time Insights: Existing solutions often fail to provide actionable insights in near real-time, which limits their utility for dynamic retail environments.
* Customization: Many models provide general recommendations but fail to account for specific customer demographics or preferences.

How This Project Addresses These Gaps:

* By conducting a holistic EDA, this project integrates customer demographics, product categories, and transaction data to provide a more comprehensive understanding of shopping trends.
* It focuses on extracting actionable insights that can guide real-world decision-making, rather than relying solely on predictive models.
* The project aims to address practical business needs by analysing the effect of discounts, seasonal trends, and customer preferences, offering specific recommendations tailored to diverse retail scenarios.

Customer behaviour analysis has been a significant area of research in retail and e-commerce. Key findings from previous studies include:

* **Demographic Insights:** Research by Smith et al. (2020) highlighted the importance of analysing demographic data to predict buying patterns.
* **Seasonal Trends:** Johnson and Brown (2019) emphasized the role of seasonality in influencing customer spending and inventory management.
* **Impact of Personalization:** Studies have shown that targeted marketing based on customer data improves engagement and conversion rates.

This project builds upon these studies by:

* Incorporating advanced visualizations to identify patterns.
* Exploring the impact of promotional incentives and subscriptions.
* Providing actionable insights for business decision-making.

**CHAPTER 3**

**Proposed Methodology**

**3.1 System Design**

**Diagram of Proposed Solution**

**[Proposed System Design Outline]**

A diagram of a system

Description automatically generated

**Explanation of the Diagram**

1. **Input Dataset:**
   * The project begins with the shopping trends dataset, which includes customer demographics, product details, and transaction data.
2. **Data Preprocessing:**
   * Data is cleaned to handle missing or inconsistent values.
   * Features are transformed into a suitable format for analysis, including encoding categorical data and standardizing numerical values.
3. **Exploratory Data Analysis (EDA):**
   * Key patterns and trends are identified using visualizations (e.g., bar charts, histograms, scatter plots).
   * Statistical summaries and correlation analyses are conducted to uncover relationships between variables.
4. **Insight & Recommendations:**
   * \*The insights derived from EDA are compiled into actionable recommendations for businesses, such as identifying popular products, customer preferences, and the impact of discounts.

**3.2 Requirement Specification**

**3.2.1 Hardware Requirements:**

* **Processor:** Intel i5 or above
* **RAM:** Minimum 8 GB
* **Storage:** At least 10 GB of free disk space for datasets and software tools
* **Graphics Support:** Optional for advanced visualization

**3.2.2 Software Requirements:**

* **Programming Language:** Python 3.x
* **Libraries for EDA and Visualization:**
  + Pandas, NumPy (Data manipulation and analysis)
  + Matplotlib, Seaborn, Plotly (Data visualization)
  + Scikit-learn (Feature selection and preprocessing)
* **Development Environment:** Jupyter Notebook or Google Colab
* **Dataset Storage:** Local storage or cloud services (e.g., Google Drive, AWS S3)
* **Optional Tools:** Tableau or Power BI for interactive visualizations

The analysis was conducted using Python and its libraries, leveraging tools such as pandas, seaborn, matplotlib, and Plotly. The methodology included:

* **Data Collection:**
  + The dataset was imported and examined for structure and completeness.
  + Missing values were identified and handled appropriately.
* **Exploratory Data Analysis (EDA):**
  + Statistical summaries were generated to understand data distributions.
  + Visualizations were created to identify trends and relationships.
* **Techniques Used:**
  + Grouping and aggregation for analysing categorical data.
  + Correlation analysis to identify relationships between variables.
  + Advanced visualizations like scatter plots, pie charts, and bar graphs.
* **Tools and Libraries:**
  + Python (pandas, matplotlib, seaborn, Plotly).
  + Jupyter Notebook for analysis and visualization.

TABLE 1: DISPLAYS THE FIRST 5 ROWS OF DATA FRAME

A screenshot of a computer

Description automatically generated

TABLE 2: DISPLAYS DATA TYPE OF EACH COLUMN

A screenshot of a computer

Description automatically generated

TABLE 3: FREQUENCY OF EACH UNIQUE VALUE IN AGE COLUMN

A screenshot of a computer

Description automatically generated

TABLE 4: AVERAGE PURCAHSE AMOUNT FOR EACH CATEGORY

A screenshot of a computer

Description automatically generated

TABLE 5: COUNT OF ITEM PURCHASED IN CATEGORY

A screenshot of a computer

Description automatically generated

TABLE 6: ITEM PURCHASE AND COUNT (GRID TABULAR FORM)

A screen shot of a computer

Description automatically generated

TABLE 7: AGE AND AGE CATEGORY VALUES

A screenshot of a computer

Description automatically generated

TABLE 8: SHIPPING TYPE TABLE

A screen shot of a computer

Description automatically generated

TABLE 9: GROUP AND COUNT SHIPPING TYPE BY PRODUCT CATEGORY

A screenshot of a computer

Description automatically generated

TABLE 10: MEAN VALUE OF NUMERIC COLUMNS

A screenshot of a computer

Description automatically generated

TABLE 11: REVIEW RATINGS AND PURCHASE AMOUNT

A screenshot of a computer

Description automatically generated

TABLE: LOCATION AND PURCHASE AMOUNT

A screenshot of a computer

Description automatically generated

**CHAPTER 4**

**Implementation and Result**

**4.1 Data Preprocessing**

* The dataset was loaded using pandas and examined for:
  + Missing values.
  + Incorrect or inconsistent entries.
* Cleaned data was prepared for analysis by:
  + Filling or dropping missing values.
  + Mapping categorical values to numeric representations where necessary.

**4.2 Exploratory Data Analysis**

Key findings include:

* **Age Distribution:**
  + Customers aged 26-35 formed the largest group of shoppers.
* **Category Preferences:**
  + Clothing was the most purchased category.
  + Accessories generated the highest revenue.
* **Impact of Promo Codes:**
  + Customers using promo codes spent significantly more on average.
* **Subscription Behaviour:**
  + Subscribed customers had higher purchase frequency and spending.

**4.3 Visualizations**

* **Scatter Plot:** Spending patterns versus previous purchases.
* **Bar Graphs:**
  + Average spending by gender and age group.
  + Most popular payment methods.
* **Pie Charts:**
  + Distribution of purchases by season.
  + Category-wise revenue contribution.

4.4 Snapshots of Results

A graph with a red and blue rectangle

Description automatically generated

A graph of a purchase amount

Description automatically generated

A graph with blue squares

Description automatically generated

A graph of a bar chart

Description automatically generated with medium confidence

A graph with a bar and a number of purchases

Description automatically generated

A graph of sales

Description automatically generated

A graph of different colored bars

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A colorful circle with a white background

Description automatically generated

A screenshot of a graph

Description automatically generated

A graph on a black background

Description automatically generated

A screen shot of a graph

Description automatically generated

A graph with different colored rectangles

Description automatically generated

A graph of blue rectangular bars

Description automatically generated

A screen shot of a graph

Description automatically generated

A diagram of a pie chart

Description automatically generated

A screenshot of a graph

Description automatically generated

A colorful circle with black text

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a graph

Description automatically generated

A screen shot of a graph

Description automatically generated

A screen shot of a shipping chart

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A red and blue circle with black lines

Description automatically generated

A colorful rectangular objects on a white background

Description automatically generated

A graph with numbers and text

Description automatically generated with medium confidence

A graph of a bar chart

Description automatically generated with medium confidence

A diagram of a purchase amount distribution by review rating

Description automatically generated

A colorful lines on a white background

Description automatically generated

A screenshot of a computer

Description automatically generated

A blue squares on a white background

Description automatically generated

A pie chart with different colored circles

Description automatically generated

A screen shot of a graph

Description automatically generated

GITHUBLINK: https://github.com/pandeysakshi8989/INTERNSHIP/tree/SHOPPING-TRENDS-DATA-ANAYLSIS

**CHAPTER 5**

**Discussion and Conclusion**

**5.1 Future Work**

While this project successfully identifies key shopping trends through EDA, several areas can be improved or expanded upon in future work:

1. **Predictive Modelling:**  
   Implement machine learning algorithms to predict customer behaviour, such as future purchases or churn probability, based on historical data.
2. **Real-Time Analysis:**  
   Integrate real-time data pipelines for continuous monitoring of shopping trends, allowing businesses to respond dynamically to changes in customer preferences.
3. **Recommendation Systems:**  
   Develop a recommendation engine to suggest personalized products for customers based on their past purchases and preferences.
4. **Geographical Insights:**  
   Expand the analysis to include geographical data to study location-specific shopping trends and customize strategies accordingly.
5. **Advanced Visualizations:**  
   Utilize interactive tools like Tableau, Power BI, or Plotly Dash for creating dashboards that allow users to explore data dynamically.
6. **Sentiment Analysis:**  
   Incorporate customer reviews and feedback to analyse sentiments and understand how they influence purchasing decisions.
7. **Handling Larger Datasets:**  
   Optimize the analysis for scalability by using distributed computing tools like Apache Spark for larger datasets.

**5.2 Conclusion**

This project demonstrates the power of Exploratory Data Analysis (EDA) in uncovering valuable insights into shopping trends. By analysing a structured dataset, we identified patterns such as age group preferences, the impact of discounts, and seasonal variations in sales. These findings provide actionable recommendations for businesses to optimize inventory management, tailor marketing strategies, and enhance customer satisfaction.

The project also highlights the importance of leveraging data to drive informed decision-making in the retail industry. While the current analysis focuses on descriptive insights, the foundation laid by this work can be expanded to include predictive models and real-time analytics, further enhancing its impact. Overall, this project underscores the potential of data analytics in transforming raw data into meaningful insights that empower businesses to thrive in a competitive market.

* The analysis provides valuable insights into customer behavior, helping businesses optimize their strategies.
* Seasonal trends and demographic preferences can guide marketing campaigns.
* The impact of incentives highlights the importance of promotional strategies.
* Businesses can leverage these findings to enhance customer engagement and retention.
* Future work could involve:
  + Real-time data integration.
  + Predictive modelling to forecast trends.
  + More granular analysis using external datasets.

**REFERENCES**

1. **Books and Articles:**
   * Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.
   * Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley.
   * Smith, J., et al. (2020). "Demographic Analysis in Retail: Strategies for Success." Journal of Business Analytics.
   * Johnson, R., & Brown, K. (2019). "Seasonal Trends and Inventory Optimization." International Journal of Commerce.
2. **Research Papers:**
   * Kumar, N., & Sharma, J. (2018). "Customer Segmentation Using K-Means Clustering Approach for Retail Market." *International Journal of Computer Applications*, 182(30), 1-6.
   * Patel, R., & Joshi, D. (2020). "Impact of Discounts and Promotions on Customer Behaviour." *Journal of Retail Analytics*, 14(3), 45-55.
3. **Online Resources:**
   * Pandas Documentation: https://pandas.pydata.org/docs/
   * Matplotlib Documentation: https://matplotlib.org/stable/contents.html
   * Seaborn Library Documentation: https://seaborn.pydata.org/
   * Plotly for Interactive Visualizations
4. **Datasets:**
   * Kaggle Shopping Trends Dataset: https://www.kaggle.com/datasets
5. **Additional References:**
   * Tableau Public Resources: https://public.tableau.com/s/
   * Scikit-learn Documentation: https://scikit-learn.org/stable/

These references serve as the foundation for understanding, analysing, and implementing the project on shopping trend analysis using EDA. Let me know if you need help formatting these for a specific citation style (e.g., APA, MLA).