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

Sathwika Adepu, sathwika1227@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to our supervisor, Jay Rathod for his invaluable guidance and support throughout this project. His expert advice, constructive feedback, and encouragement were instrumental in shaping the direction of my work and overcoming the challenges I faced.

His insights and innovative ideas played a significant role in the successful completion of the project. I deeply appreciate the time and effort he dedicated to ensuring that my work met the highest standards.

It has been a privilege to work under his mentorship, and I am truly thankful for his unwavering support and confidence in our abilities.

ABSTRACT

Retail businesses generate large volumes of data through multiple channels, including in-store and online platforms. However, effectively analyzing this data to identify shopping trends, customer preferences, and seasonal patterns remains a challenge. Failure to act on such trends can lead to lost revenue, overstock or stockouts, ineffective marketing strategies, and a diminished competitive advantage.

This project, titled "Identifying Shopping Trends Using Data Analysis", aims to address these challenges by developing a data-driven framework to uncover actionable insights. The objectives include identifying emerging trends, predicting customer preferences, and enhancing inventory management and marketing strategies to optimize retail operations. The methodology involves collecting shopping data from diverse channels, preprocessing it, and applying statistical and machine learning techniques to uncover patterns. Advanced visualization tools are used to communicate insights effectively, aiding decision-making for retail managers.

Preliminary results demonstrate the framework's potential to predict trends and inform strategies that align inventory with demand and tailor marketing to customer behavior. These insights empower retail businesses to make accurate, timely decisions, improving overall efficiency and profitability.

Preliminary results demonstrate the framework's potential to predict trends and inform strategies that align inventory with demand, personalize marketing campaigns, and improve customer satisfaction. These insights empower retail businesses to make accurate, timely decisions, improving overall efficiency, reducing operational costs, and maximizing profitability.

In conclusion, this project highlights the importance of leveraging data analytics in retail, offering a practical solution for addressing operational and strategic challenges. It ensures businesses remain competitive in an evolving market by transforming raw data into valuable insights.

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 **4**

2.1 Review of Relevent Literature 4

2.2 Existing Models, Techniques, and Methodologies 5

2.3 Addressing the Limitations 6

Chapter 3. Proposed Methodology **7**

3.1 System Design 7

3.2 Requirement Specification 8

3.2.1 Hardware Requirements 8

3.2.2 Software Requirements 9

Chapter 4. Implementation and Results **10**

4.1 Snap Shots of Result 10

4.2 Github Link of code 13

Chapter 5. Discussion and Conclusion **14**

5.1 Future work 14

5.2 Conclusion 14

References 16

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	Proposed Solution	**7**
Figure 2	Snapshot 1 of result	**10**
Figure 3	Snapshot 2 of result	**11**
Figure 4	Snapshot 3 of result	**12**
Figure 5		
Figure 6		
Figure 7		
Figure 8		
Figure 9		

LIST OF TABLES

Table. No.	Table Caption	Page No.
1	Most purchased item	12

CHAPTER 1

Introduction

1.1Problem Statement:

Retail businesses generate large volumes of data through multiple channels, such as in-store and online transactions. However, many struggle to analyze this data effectively to identify trends, customer preferences, and seasonal patterns. This lack of actionable insights leads to significant challenges, including lost revenue, poor inventory management, and ineffective marketing strategies.

For instance, failing to predict demand accurately can result in stock shortages or overstock, causing customer dissatisfaction or increased operational costs.

This issue is critical as it impacts key areas of retail operations, such as inventory, marketing, and overall customer satisfaction. In today's competitive market, timely and accurate data-driven decisions are essential for success.

This project aims to address these challenges by developing a framework to analyze shopping trends, providing actionable insights to optimize retail operations and enhance profitability.

1.2Motivation:

The project on shopping trend analysis is chosen to gain insights into consumer behavior, preferences, and market dynamics. By analyzing shopping trends, businesses can better understand consumer patterns, optimize inventory management, and personalize marketing strategies to target specific customer segments.

Potential Applications and impact: Predicting product performance to optimize inventory and marketing campaigns. Personalizing recommendations, offers, and ads based on consumer behavior. Identifying distinct customer groups for targeted marketing and promotions. Gaining a competitive edge by adapting quickly to changing market trends. Personalizing shopping experiences based on consumer preferences, leading to higher engagement and sales.

1.3Objective:

The primary objective of this project is to develop a robust framework for analyzing shopping trends that can help retail businesses make data-driven decisions, optimize their operations, and enhance profitability. Specifically, the objectives are:

To analyze large volumes of retail transaction data across multiple channels (in-store and online) to identify emerging shopping trends, customer preferences, and seasonal patterns.

To develop a model that accurately predicts product demand, helping businesses avoid stock shortages or overstocking, leading to better inventory management.

To provide insights for dynamic pricing models based on demand fluctuations, competition, and seasonal trends, ensuring competitive yet profitable pricing strategies.

To enhance operational decision-making by providing tools for better inventory forecasting, supply chain management, and campaign planning based on data-driven insights.

1.4Scope of the Project:

The key areas of focus will include:

Data Analysis:

Collect and analyze transactional data from both in-store and online retail channels.

Identify patterns related to customer preferences, seasonal trends, and purchasing behaviors.

Trend Identification and Forecasting:

Use historical data to identify emerging shopping trends and predict future demand for products.

Customer Segmentation:

Segment customers based on purchasing behaviors and preferences to enable targeted marketing and personalized recommendations.

Reporting and Visualization:

Provide dashboards and reports that offer actionable insights to businesses for improved decision-making in inventory management, marketing, and operations.

Limitations:

The quality and completeness of data may vary across different businesses. Inaccurate, missing, or inconsistent data could affect the reliability of the analysis.

The framework may not fully account for external factors (e.g., economic changes, global crises) that can influence shopping behavior in ways not captured by historical data.

The project will focus on transactional data but may not account for other sources of data, such as social media interactions or external market conditions, that could influence shopping trends.

CHAPTER 2

Literature Survey

2.1Review of Relevant Literature.

The field of shopping trend analysis and consumer behavior in retail has been studied extensively across various domains, including data analytics, machine learning, marketing, and supply chain management.

Consumer Behavior and Shopping Trends

Understanding consumer preferences and behavior is crucial for predicting shopping trends. Studies have shown that factors like price sensitivity, product recommendations, and seasonal trends significantly influence consumer purchasing decisions.

Data Analytics and Trend Identification

Machine learning algorithms, such as decision trees, random forests, and neural networks, have been applied to predict shopping trends by analyzing historical sales data. These models can help businesses forecast demand and optimize inventory.

Pricing Strategy Optimization

Studies have also focused on analyzing price elasticity to determine how sensitive customers are to price changes, which helps businesses set the optimal price for their products.

2.2 Existing Models, Techniques, and Methodologies

Several models, techniques, and methodologies have been developed and employed in the domain of shopping trend analysis, demand forecasting, customer segmentation, and inventory management.

Machine Learning Models for Trend Identification

Decision Trees and Random Forests: Decision trees are widely used for identifying patterns in retail data, such as identifying which products are likely to be popular at a given time. Random forests, an ensemble technique, improve prediction accuracy by averaging over multiple decision trees, making them more robust and less prone to overfitting.

XGBoost (Extreme Gradient Boosting): This popular algorithm for regression and classification tasks has been used to predict demand trends by evaluating numerous input features and their interactions. It is highly effective for structured retail data and known for its accuracy and efficiency.

Long Short-Term Memory (LSTM): LSTM, a type of recurrent neural network (RNN), is well-suited for sequential data like sales and transaction data, making it useful for demand forecasting and trend prediction. LSTM models capture long-term dependencies in time-series data, which are critical for retail trend analysis.

2.3 Gaps or Limitations in Existing Solutions and How This Project Addresses Them

While existing solutions have significantly advanced shopping trend analysis, several gaps and limitations remain, which this project aims to address. These include challenges related to data integration, scalability, personalization, and actionable insights.

The project will implement robust preprocessing techniques to clean and normalize data across multiple channels.

The project will focus on creating modular and customizable models that businesses can adapt to their specific needs.

The project will explore approaches like deep learning and graph-based models to overcome the cold start problem and enhance personalization.

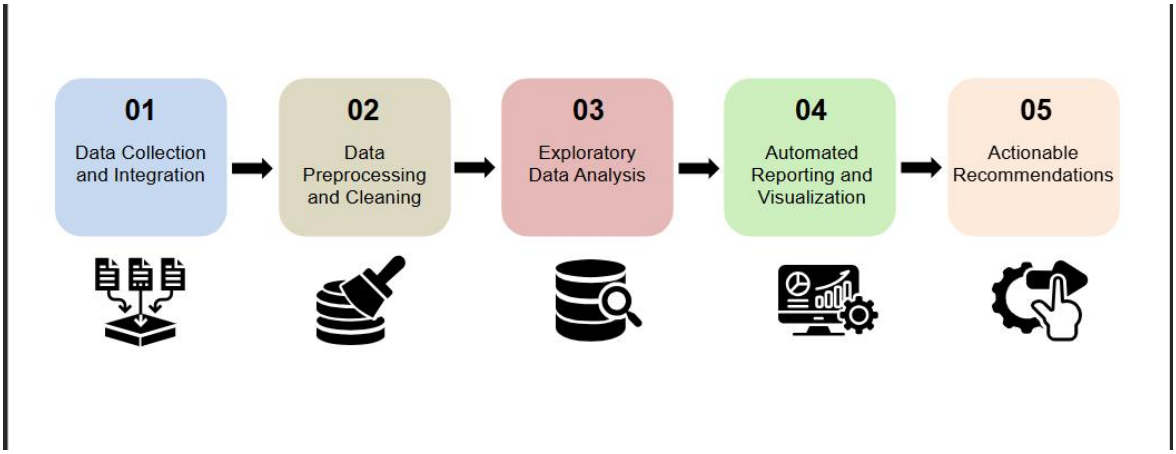
External data sources, such as market trends, competitor pricing, and macroeconomic indicators, will be integrated into the analysis.

CHAPTER 3

Proposed Methodology

3.1 System Design

Figure 1: Proposed Solution



Data Collection and Integration

This step involves gathering data from various sources such as retail stores, e-commerce platforms, customer transactions, and external datasets (e.g., market trends or competitor information).

Data Preprocessing and Cleaning

The collected data is often raw and unstructured. This stage focuses on cleaning the data to remove duplicates, handle missing values, and correct inconsistencies.

It also involves normalizing and formatting data to prepare it for analysis.

Exploratory Data Analysis (EDA)

This stage involves examining the dataset to uncover patterns, trends, and anomalies.

Automated Reporting and Visualization

After analyzing the data, the insights are visualized using tools like dashboards, graphs, and charts. These visualizations help stakeholders quickly interpret the results.

Actionable Recommendations

Based on the analysis and visualizations, specific recommendations are generated to optimize retail operations, such as inventory management, pricing strategies, and marketing campaigns.

3.2 Requirement Specification

This section outlines the tools, technologies, and platforms required to implement the shopping trend analysis solution effectively.

Python: For data analysis, machine learning, and visualization.

Libraries:

Pandas and NumPy: For data manipulation and preprocessing.

Matplotlib and Seaborn: For data visualization.

Scikit-learn: For machine learning models.

TensorFlow or PyTorch: For deep learning models (e.g., LSTM).

Jupyter Notebook: For building, testing, and documenting machine learning workflows.

Plotly or Dash: For building custom web-based data visualization tools.

3.2.1 Hardware Requirements:

Processor: Intel Core i5 or AMD Ryzen 5 (or higher).

RAM: At least 16 GB (32 GB recommended for larger datasets).

Storage:

500 GB SSD for faster data access.

Additional HDD for long-term data storage (optional).

Graphics Processing Unit (GPU):

NVIDIA GPU with at least 4 GB VRAM (e.g., NVIDIA GTX 1660 or better) for deep learning models.

For larger workloads, NVIDIA RTX 3060 or higher is recommended.

Operating System: Windows 10/11, Linux (Ubuntu 20.04+), or macOS.

3.2.2 Software Requirements:

Python

Libraries and Frameworks

- **Data Processing**:
  - Pandas, NumPy .
- **Visualization**:
  - Matplotlib, Seaborn, Plotly.
- **Machine Learning**:
  - Scikit-learn.
- **Deep Learning**:
  - TensorFlow, PyTorch, or Keras.

CHAPTER 4

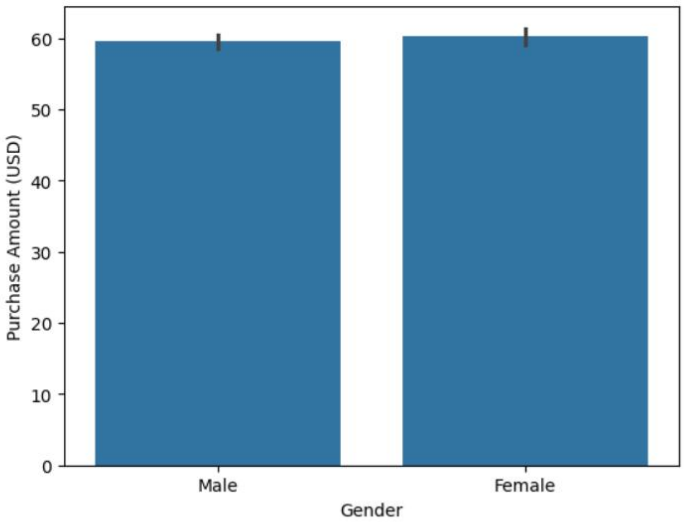
Implementation and Result

4.1 Snap Shots of Result:

Figure 2: Snapshot1 of result

3.Which gender has the highest number of purchases?

```
[25]: sns.barplot(shop , x = 'Gender' , y = 'Purchase Amount (USD)')
[25]: <Axes: xlabel='Gender', ylabel='Purchase Amount (USD)'>
```



This snapshot represents a bar plot that answers the question: "Which gender has the highest number of purchases?"

· Purpose: The chart compares the purchase amounts (in USD) for males and females.

· X-axis: Represents the gender categories ("Male" and "Female").

· Y-axis: Displays the average purchase amount in USD.

· Observation: From the bar plot, it seems both genders have a very similar purchase amount, with no significant difference.

This visualization helps businesses understand whether purchasing behavior differs between genders, which can influence marketing strategies and product targeting. If a specific gender had a higher average, campaigns could be tailored to engage that group more effectively.

Figure 3: Snapshot 2 of Result

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

```
[26]: shop.groupby('Category')['Item Purchased'].value_counts()
```

```
[26]: Category     Item Purchased
      Accessories  Jewelry           171
                   Sunglasses        161
                   Belt              161
                   Scarf             157
                   Hat               154
                   Handbag           153
                   Backpack          143
                   Gloves            140
      Clothing     Pants             171
                   Blouse            171
                   Shirt             169
                   Dress             166
                   Sweater           164
                   Socks             159
                   Skirt             158
                   Shorts            157
                   Hoodie            151
                   T-shirt           147
                   Jeans             124
      Footwear     Sandals           160
                   Shoes             150
                   Sneakers          145
                   Boots             144
      Outerwear    Jacket            163
                   Coat              161
      Name: count, dtype: int64
```

```
[27]: fig = px.histogram(shop , x = 'Item Purchased' , color = 'Category')
      fig.show()
```

This snapshot provides insights into the most commonly purchased items in each category.

Explanation:

- Purpose: It identifies the most popular items within different product categories such as Accessories, Clothing, Footwear, and Outerwear.
- Data: The table lists the counts of each item purchased, grouped by category, sorted by the frequency of purchases.
    - For instance:
        - In Accessories, Jewelry is the most purchased item with 171 purchases, followed by Sunglasses and Belt with 161 each.
        - In Clothing, Pants and Blouse are tied as the most purchased items with 171 purchases.
        - Other categories, such as Footwear and Outerwear, also showcase their most purchased items.
- Code: The method groupby('Category')['Item Purchased'].value_counts() is used to count item purchases by category.
- Significance: Retailers can focus on maintaining optimal stock levels of the most popular items in each category to meet customer demands.

This snapshot allows businesses to:

1. Tailor their inventory management based on demand.
2. Identify potential marketing opportunities for less popular items.

Table 1: Most common purchased item

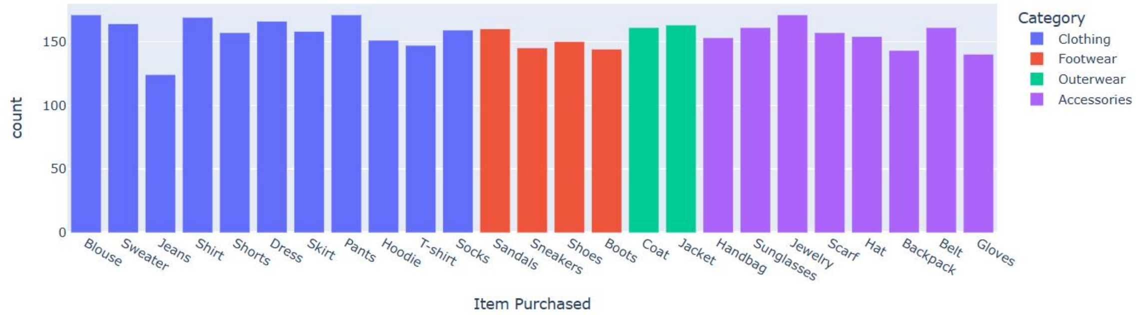


Figure 4: Snapshot 3 of Result

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

```
[38]: shop['Age_category'].unique()
```

```
[38]: ['old', 'Young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'Young Adults' < 'Middle-Aged Adults' < 'old']
```

```
[39]: shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
```

```
[40]: px.sunburst(shop , path=['Frequency of Purchases','Age_category'] , values='Age')
```

```
C:\Users\DELL\AppData\Local\Programs\Python\Python312\Lib\site-packages\plotly\express\_core.py:1727: FutureWarning:

The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.
```

This visualization represents a sunburst chart that illustrates the frequency of purchases across different age groups. Here's the explanation:

Explanation:

Outer Circle (Age Categories): The chart is divided into segments representing age categories (child, teen, Young Adults, Middle-Aged Adults, and old).

These categories show how purchase frequencies differ across age groups.

Inner Circle (Frequency of Purchases): The chart's inner layer represents the purchase frequency levels, such as low, medium, or high (assumed categories). Each frequency level acts as the root for its respective age group distribution.

Size of Segments: The size of each segment corresponds to the number of purchases or the sum of the ages in that frequency category.

4.2 GitHub Link for Code:

https://github.com/Sathwika-A/shopping-trends-analysis

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

To further enhance the current solution, the following improvements and extensions can be considered:

Integration of Real-Time Analytics:

- Implement real-time data collection and analysis to provide immediate insights into shopping trends, enabling dynamic decision-making for inventory management and marketing.

Multi-Channel Data Integration:

- Extend the model to integrate data from social media, customer reviews, and web traffic to gain a more comprehensive understanding of customer preferences and sentiment.

Address Data Limitations:

- Improve the data collection process by ensuring completeness and consistency across channels.

Ethical AI and Data Privacy:

- Implement mechanisms to ensure data privacy and compliance with GDPR or other regulations.

5.2 Conclusion:

This project successfully demonstrates the application of data analysis and visualization techniques to uncover actionable insights in the retail domain. By leveraging structured datasets and state-of-the-art tools, the project provides valuable findings regarding customer behavior, purchase patterns, and product preferences, which can significantly enhance decision-making processes.

The analysis revealed critical patterns such as gender-based purchasing trends, commonly purchased items, and the impact of age groups on shopping frequency. These insights help businesses tailor their strategies for targeted marketing and inventory management.

The visualizations and findings provide a clear understanding of customer behavior, enabling businesses to make informed decisions that align with consumer needs and preferences.

Automating the analysis and reporting process saves time and effort, allowing organizations to focus on strategy development and execution.

Overall, this project contributes to the digital transformation of retail businesses by emphasizing the importance of data-driven decision-making, offering a scalable and efficient framework for analyzing and interpreting customer data. It provides a robust foundation for future advancements and paves the way for achieving greater customer satisfaction and business success.

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.