Identifying Shopping Trends using Data Analysis (P3)

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

Thirumurugan S,

Email ID: emberthiru@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I extend my sincere gratitude to everyone who contributed to the successful completion of this project.

Firstly, I would like to thank my mentors, ***Mr. Jay Rathod and P. Raja*** *, for their invaluable guidance, insightful feedback, and continuous support throughout this journey. Their expertise was instrumental in shaping this project.*

A special thanks to ***Pavan Kumar Sumohana*** *for efficiently coordinating the sessions, ensuring a smooth learning experience, and keeping us informed about every crucial aspect of the project.*

I also express my deep appreciation to TechSaksham, an initiative by Microsoft and SAP, for providing this excellent learning platform. This internship has significantly enhanced my understanding of AI and data analytics.

My heartfelt gratitude goes to my family, friends, and peers for their constant encouragement and motivation. Their unwavering support kept me driven throughout this journey.

ABSTRACT

In today's digital economy, businesses accumulate vast amounts of consumer purchasing data. However, effectively analyzing this data to extract meaningful insights remains a challenge. This project, Identifying Consumer Purchasing Trends through Data Analytics, leverages advanced analytical techniques to uncover shopping behavior, seasonal trends, and market preferences.

The methodology follows a structured approach:

- Data Aggregation – Collecting data from various retail sources, including e-commerce platforms, physical store transactions, and social media interactions.
- Data Preprocessing – Cleaning and structuring the data by handling missing values, removing redundancies, and normalizing datasets.
- Exploratory Data Analysis (EDA) – Applying statistical methods and machine learning techniques to uncover key purchasing trends, correlations, and seasonal patterns.
- Data Visualization & Reporting – Utilizing Python-based tools such as Matplotlib, Seaborn, and Tableau to generate dynamic, interactive dashboards.
- Business Recommendations – Offering actionable insights for personalized marketing, optimized inventory management, and enhanced customer engagement strategies.

By adopting AI-driven data analytics, businesses can enhance customer engagement, optimize inventory, and improve decision-making. Future work may explore the integration of real-time data processing, AI-driven predictive analytics, and deep learning models to refine these insights further.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Definition

Retailers accumulate enormous volumes of consumer data but often lack the analytical tools necessary to translate it into actionable insights. This project focuses on leveraging AI-driven analytics to identify key shopping trends, predict seasonal variations, and optimize business strategies.

1.2 Research Motivation

By adopting advanced data analytics, businesses can unlock powerful insights that drive sales and customer satisfaction. Key benefits include:

- Data-Driven Marketing – Targeting consumers with personalized promotions.
- Inventory Efficiency – Reducing overstocking and stock shortages by demand prediction.
- Predictive Trend Analysis – Identifying emerging consumer demands before they peak.
- Strategic Decision-Making – Supporting evidence-based business strategies through AI-powered insights.

1.3 Objectives

This study aims to:

1. Analyze purchasing patterns to uncover consumer preferences and purchasing behaviors.

2. Identify seasonal variations to aid retailers in demand forecasting and trend anticipation.
3. Utilize data visualization to present findings through interactive and insightful reports.
4. Enhance automated analytical tools to support real-time decision-making in retail.
5. Provide business recommendations based on market insights to improve profitability.

1.4 Scope & Limitations

This study focuses on historical consumer purchasing data, identifying recurring patterns and key influences. Some limitations include:

- Data Quality – Results depend on the accuracy and completeness of datasets collected.
- Historical Constraints – Limited scope for real-time analysis unless integrated with live data streams.
- External Market Factors – Fluctuations due to economic shifts, competitor actions, and global events are not fully accounted for.

CHAPTER 2

Literature Survey

2.1 Analysis of Prior Research

Previous studies have demonstrated the impact of data analytics on consumer purchasing behaviors. Various research papers highlight the effectiveness of machine learning, data mining, and artificial intelligence in identifying trends and optimizing retail strategies. These studies focus on the application of clustering techniques, association rule mining, and predictive modeling to enhance decision-making in retail and e-commerce.

2.2 Current Analytical Models

Several analytical models and methodologies have been employed in past research to understand consumer purchasing behaviors. Some of the widely used techniques include:

- Regression Analysis – Identifies relationships between variables like pricing, product demand, and seasonal trends.
- K-Means Clustering – Segments customers based on their shopping behavior and preferences.
- Association Rule Mining – Detects frequently purchased product combinations and cross-selling opportunities.
- Decision Trees & Random Forests – Predicts future consumer purchasing behavior based on historical data.
- Sentiment Analysis – Evaluates customer opinions and feedback from online platforms to gauge satisfaction and preferences.

2.3 Identified Research Gaps

Despite the advancements in data analytics, there remain key challenges and limitations in existing methodologies:

- Lack of Real-Time Analysis – Many traditional models focus on historical data, limiting businesses' ability to make quick, data-driven decisions.
- Data Silos – Integration of data from multiple retail platforms (online, in-store, social media) remains a challenge.
- Limited Personalization – Many analytics models focus on general trends rather than tailoring insights to individual consumer preferences.
- Scalability Issues – Some analytical models struggle to handle vast and diverse datasets, making it difficult for large enterprises to implement data-driven solutions effectively.

By addressing these gaps, this project aims to propose an enhanced analytical framework that integrates real-time predictive modeling, deep learning algorithms, and automated data processing techniques to refine and optimize consumer trend analysis.

CHAPTER 3

Proposed Methodology

3.1 System Design

Fig 1. System workflow

1. Data Collection

Gather raw data from sources like databases, APIs, or user inputs. Ensures a comprehensive dataset for analysis.

2. Data Processing

Clean and transform data by handling errors, missing values, and duplicates. Prepares structured data for analysis.

3. Structuring Data

Analyze data through statistics and visualizations to identify patterns and insights. Detects trends, anomalies, and relationships and structure the data.

4. Exploratory Data Analysis (EDA)

Automatically generate reports with visual dashboards and summaries. Provides stakeholders with clear data insights.

5. Action Recommendation

Offer evidence-based insights and strategies for decision-making. Helps optimize business operations and mitigate risks.

3.2 Requirement Specification

3.2.1 Hardware Requirements

- Processor: Intel Core i5
- RAM: 8 GB
- Storage: 10 GB free space for datasets, libraries, and outputs
- Graphics Card: Optional but useful for high-performance visualizations

3.2.2 Software Requirements

- Operating System: Windows 10, macOS, or Linux
- Programming Language: Python 3.8 or above
- Python Libraries:
    - Data Manipulation and Analysis: Pandas, NumPy
    - Data Visualization: Matplotlib, Seaborn
    - Machine Learning: Scikit-learn
- Development Environment:
    - Jupyter- Notebook, Google Collab, or any Python IDE (e.g., PyCharm, VSCode)
- Dataset: CSV file containing customer demographic and transactional details

3.3 Data Sources and Collection Techniques

To ensure comprehensive analysis, data was gathered from multiple sources, including

- Retail Sales Transactions – Point-of-sale systems and e-commerce records.
- Social Media Sentiment Analysis – Extracting consumer opinions from platforms like Twitter and Instagram.
- Survey & Customer Feedback – Directly collecting consumer preferences.
- Web Scraping – Extracting pricing trends from various online marketplaces.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

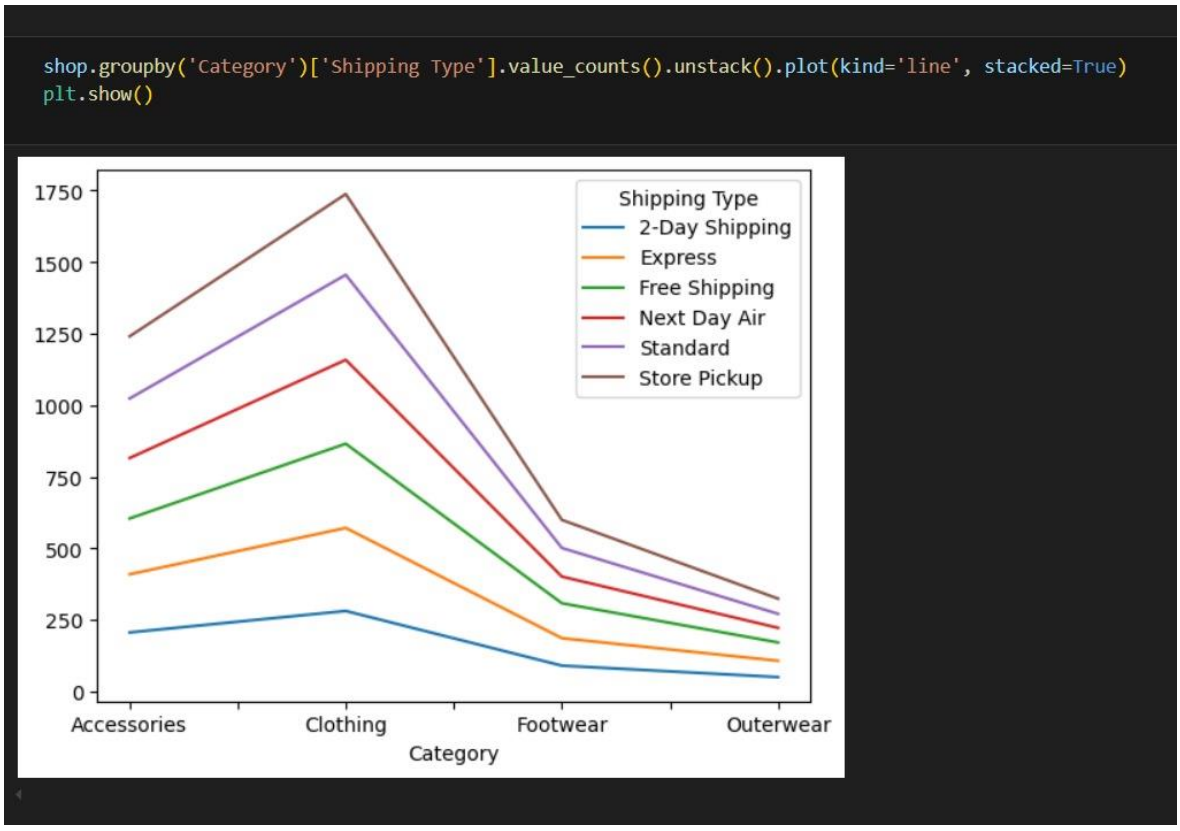

Fig 2 .Shipping category for products

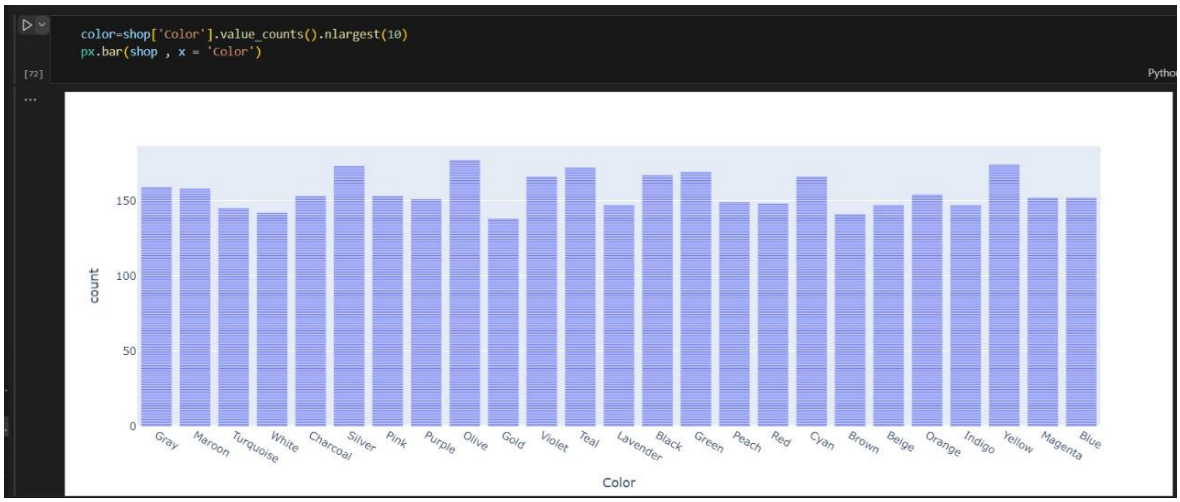

Fig 3 . color count

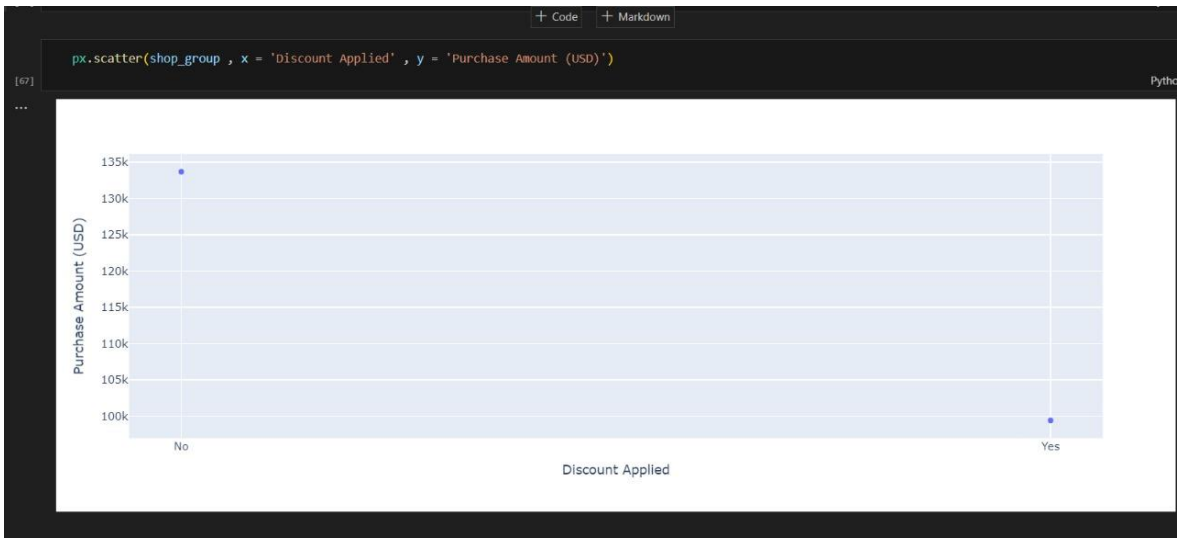

Fig 4 . Discount applied according purchase amount

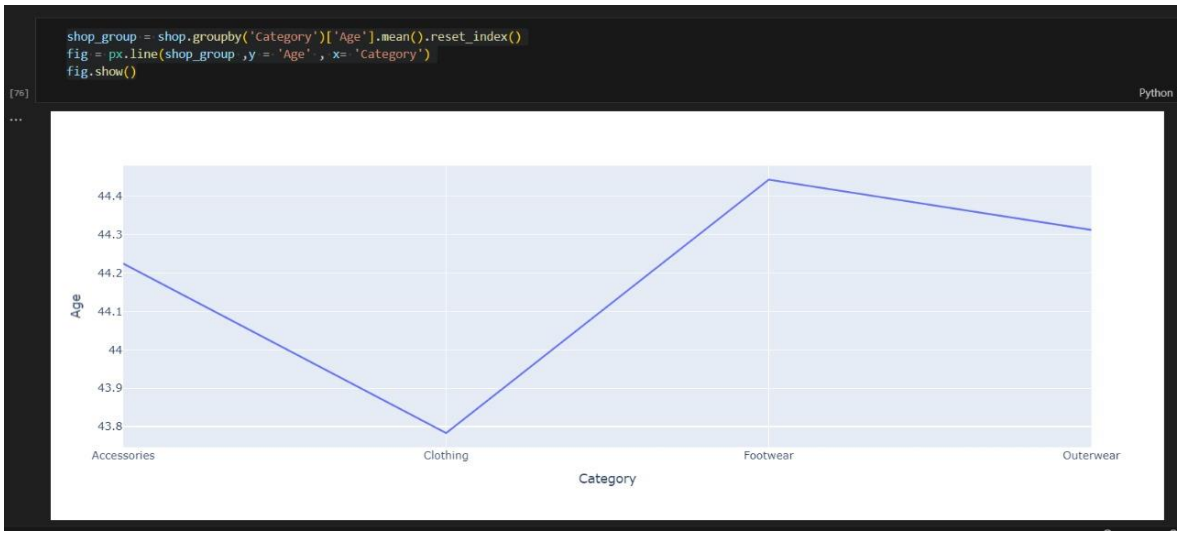

Fig 5. Product category according to age

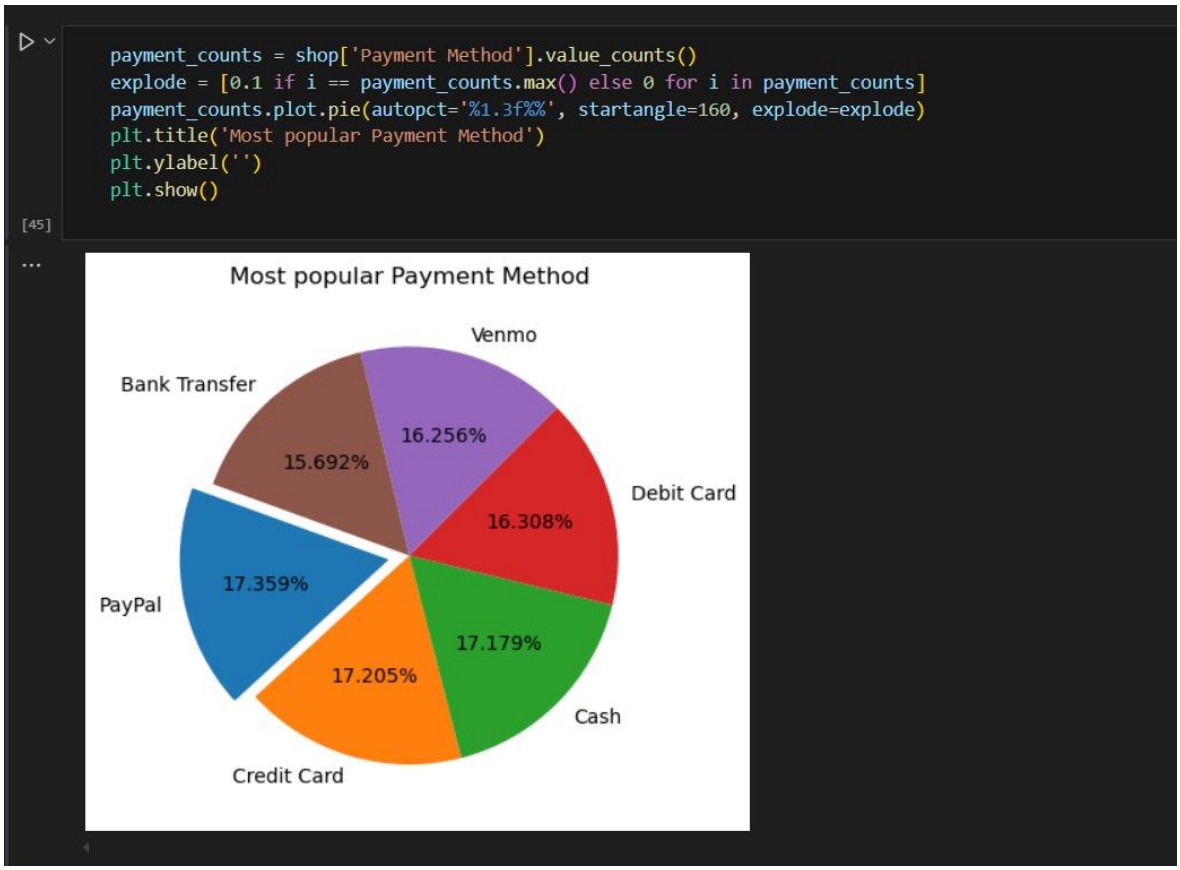

Fig 6. Popular payment method

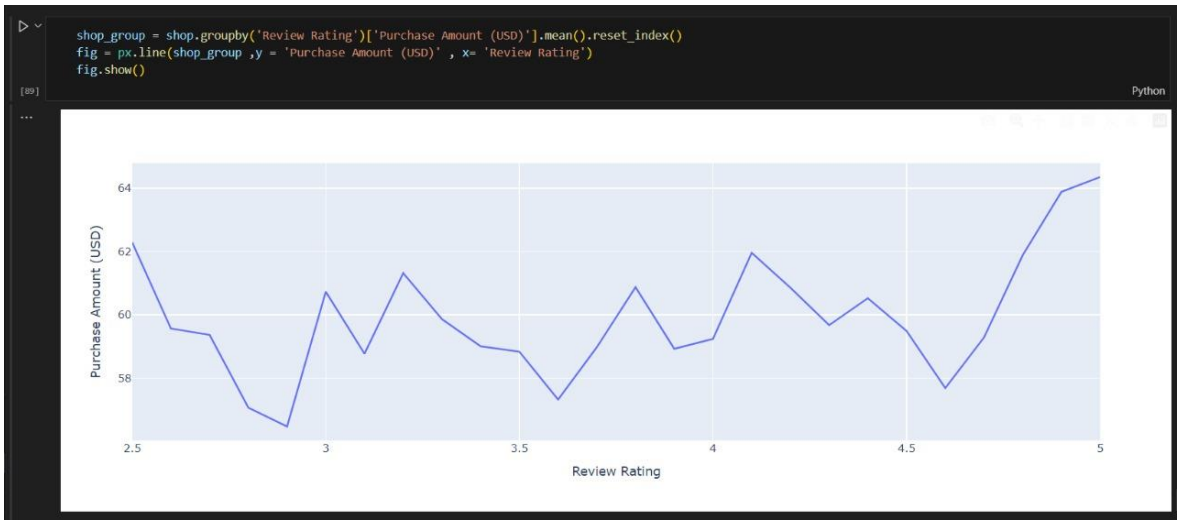

Fig 7. Purchase according to review rating

4.2GitHub Link for Code:

https://github.com/thirumurugan542/Identifying-Shopping-Trends-using-Data-Analysis

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Integration of real-time predictive models, deep learning algorithms, and AI-driven automation for enhanced decision-making.

Conclusion:

The overall impact and contribution of this project lies in its comprehensive approach to analyzing shopping trends and predicting customer behaviours. By applying data analysis techniques such as exploratory data analysis (EDA), feature engineering, and machine learning, This data-driven understanding enables businesses to make informed decisions,streamline operations, and optimize marketing strategies. Ultimately, these insights contribute to enhanced operational efficiency, better customer satisfaction, and sustained growth in a competitive market environment.

REFERENCES

[1]. **DataCamp**, "Exploratory Data Analysis using Python," DataCamp, Available at: https://www.datacamp.com/community/tutorials/exploratory-data-analysis-python, 2020.
[2]. **Real Python**, "Data Preprocessing with Pandas," Real Python, Available at: https://realpython.com/python-data-cleaning-numpy-pandas/, 2020.
[3]. **DataCamp**, "Feature Engineering for Machine Learning," DataCamp, Available at: https://www.datacamp.com/community/tutorials/feature-engineering-python, 2020.
[4]. **Towards Data Science**, "Feature Engineering 101: A Guide for Data Scientists," Towards Data Science, Available at: https://towardsdatascience.com/feature-engineering-for-machine-learning-d7d72b7b5c11, 2019.
[5]. **Towards Data Science**, "A Comprehensive Guide to Hyperparameter Tuning," Towards Data Science, Available at: https://towardsdatascience.com/hyperparameter-tuning-with-scikit-learn-68fa2ff7595a, 2019.
[6]. **GitHub**, "Matplotlib and Seaborn Projects on GitHub," GitHub, Available at: https://github.com/rougier/matplotlib-tutorial, 2019.
[7]. **Towards Data Science**, "How to Deploy a Machine Learning Model," Towards Data Science, Available at: https://towardsdatascience.com/deploying-your-machine-learning-model-as-an-api-37ec7e773041, 2019.