DOI: 10.5281/zenodo.16021454

International Journal of Research and Applications
July – Sep © 2025 Transactions, 12(47): 3717-3720

eISSN : 2349 – 0020
pISSN : 2394 – 4544
www.ijraonline.com

COMPUTERS REVIEW REPORT

Identifying shopping trends using data analysis

[1] E. Narasimha Swamy, [2] K. Ranjith Kumar
[1] PG(Pursuing) , [2]Assistant Professor in CSE
[1] Master of Computer Applications,
[1,2] Vaagdevi Engineering College, Bollikunta, Warangal, India

ABSTRACT

This project, "Identifying Shopping Trends using Data Analysis," aims to uncover patterns in customer purchasing behaviors to provide actionable insights that inform business decision-making. The analysis explores key factors such as the influence of discounts on spending, regional variations in purchasing habits, and preferences for shipping methods across different product categories. Using Python, the project leverages data analysis libraries such as pandas, matplotlib, and seaborn for data preprocessing, exploratory analysis, and visualization. The dataset includes customer demographics, product categories, purchase amounts, and shipping preferences. The data preprocessing phase involved cleaning and organizing the dataset to ensure accuracy and consistency. Key findings reveal significant regional differences in spending, with certain areas exhibiting higher purchase volumes. Additionally, the analysis showed that discounts strongly influenced purchase amounts, while shipping preferences varied by product category. A further examination of product color preferences provided valuable insights into inventory management. Visualizations, including pie charts, bar charts, and histograms, effectively conveyed these findings, such as regional purchase patterns and distribution of shipping preferences. These insights have practical applications for businesses, enabling them to optimize operations, improve customer satisfaction, and devise targeted marketing strategies. The project highlights the importance of data-driven decision-making in understanding consumer behavior and suggests future integration of machine learning models to predict customer preferences and seasonal trends.

Keywords:- Pandas, Matplotlib, Seaborn for data preprocessing, Exploratory analysis, Visualization.

1. INTRODUCTION

In today's data-driven world, understanding customer behaviour has become a critical factor for businesses seeking to stay competitive in the retail industry. With vast amounts of shopping data being generated daily, businesses are presented with an enormous opportunity to uncover valuable insights that can help optimize operations, tailor marketing strategies, and enhance customer experiences. However, despite having access to such rich data, many organizations struggle to derive meaningful conclusions without effective data analysis techniques.

This project, titled "Identifying Shopping Trends using Data Analysis," is aimed at addressing this challenge by applying data science methods to explore patterns and trends in customer purchasing behavior. Conducted as part of the AICTE Internship on Artificial Intelligence under the TechSaksham initiative—a joint CSR effort by Microsoft and SAP—this project leverages real-world shopping data to uncover actionable insights.

Using Python as the primary programming language and powerful data analysis libraries such as pandas, matplotlib, and seaborn, the project involves a comprehensive process of data preprocessing, exploratory data analysis (EDA), and data visualization. The dataset contains information related to customer demographics, product categories, purchase amounts, discounts, and shipping preferences.

By answering these questions, the project aims to assist retail businesses in making informed decisions that are grounded in data. The insights can be used to develop more efficient inventory systems, target promotions effectively, and deliver personalized shopping experiences. Visual representations such as bar charts, pie charts, and histograms have been used extensively to simplify complex data and make the findings easily interpretable by business stakeholders.

2. LITERATURE SURVEY

The study of shopping trends through data analysis has increasingly become a critical component of modern business strategy. The availability of vast datasets capturing customer purchasing behaviors, demographics, preferences, and engagement metrics has presented businesses with unparalleled opportunities to gain insights and optimize operations. Over the last two decades, significant academic and industrial research has been dedicated to uncovering the underlying patterns within shopping behaviors. Through the advent of data science, machine learning, and artificial intelligence, businesses have transitioned from intuition-based strategies to data-driven decision-making, allowing them to react faster to market demands and to personalize customer experiences with unprecedented accuracy.

Initially, shopping behavior studies relied heavily on traditional statistical methods such as descriptive analysis, regression models, and time-series forecasting. Scholars like Kotler (2000) emphasized the importance of understanding consumer behavior through surveys and structured interviews. However, as e-commerce and online retailing emerged and expanded exponentially, researchers soon recognized the necessity of analyzing digital footprints, transactional logs, and behavioral data collected across multiple touch points. This transition to digital data analysis led to a surge of literature focused on electronic commerce analytics, predictive modeling, and customer segmentation.

Existing research outlines various techniques for studying shopping behaviors. Exploratory Data Analysis (EDA) has been a foundational method, serving to summarize the main characteristics of datasets and to uncover underlying structures. EDA methodologies, pioneered by John Tukey, emphasize visualization techniques like histograms, box plots, scatter plots, and pie charts, which remain highly relevant today. Numerous studies have demonstrated how EDA facilitates identifying purchase frequencies, average order values, and the correlation between product categories and customer demographics.

3. METHODOLOGY

Data Understanding

- Before even preprocessing, you should understand the data:
- What each column means.
- Identify categorical, numerical, and text data.
- Check how many missing values are there.
- This step ensures correct assumptions during analysis.

Feature Engineering (optional but powerful)

- Feature engineering means creating new columns from existing data to extract more insights.
- Example:
  - Create a "discounted price" column = original price - discount.
  - Create customer age groups like "18-25", "26-35", etc.
- Helps reveal hidden patterns not visible otherwise.
-
- **Correlation Analysis**
  - Beyond basic EDA, you can check correlations:
  - Does discount have a strong relation with purchase amount
  - Are ages and purchase behavior connected?
  - Correlation plots (heatmaps) can reveal hidden connections.

Hypothesis Testing

- Instead of just observing, you can prove statistically:
- Is the average spending really different across genders.
- Does region A spend more than region B?

Data Quality Reporting

- After preprocessing, create a data quality report:

- What assumptions were made during cleaning?

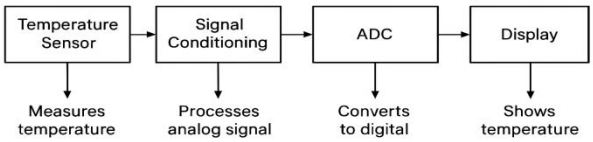

Fig: - 1 Block Diagram

TOOLS

Jupyter Notebook or Google Colab for coding and visualization.

4. RESULTS

The result figures reveal key shopping trends: young adults shop frequently; older groups spend more. Gender influences purchase amounts. Discounts significantly boost sales. Regional differences highlight varied buying behaviors. Product type affects shipping preferences. These insights, visualized through bar and pie charts, support data-driven strategies to enhance retail operations and marketing.

In Fig 2

The bar chart shows the total age sum across four age categories. Older adults contribute the highest total age, followed by middle-aged adults. Young adults have a significantly lower total, while teens contribute the least. This suggests a higher population or frequency of older individuals in the dataset.

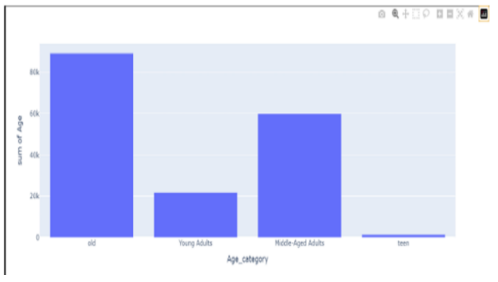

Fig:-2 Pie chart is representing Age categories.

In Fig 3

The bar chart displays the purchase counts of various items categorized by type. Clothing and Accessories dominate with higher counts, especially items like sweaters, jackets, and sunglasses. Footwear and outerwear have relatively fewer purchases. This indicates a higher consumer preference for clothing and accessory items in the dataset.

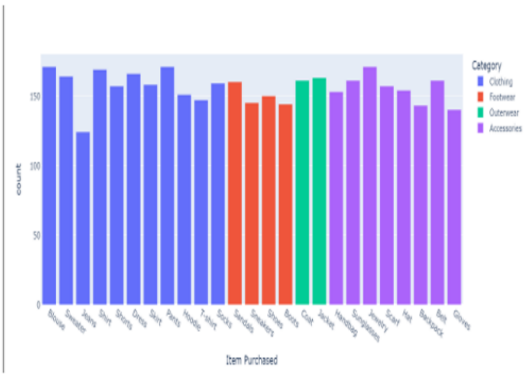

Fig:- 3 Bar chart showing count of Item Purchased category

In Fig 4

The bar chart compares total purchase amounts by gender. Males show significantly higher purchase amounts Than females, indicating that men either make more purchases or spend more per transaction. This suggests a gender- based difference in shopping behavior within the analyzed dataset.

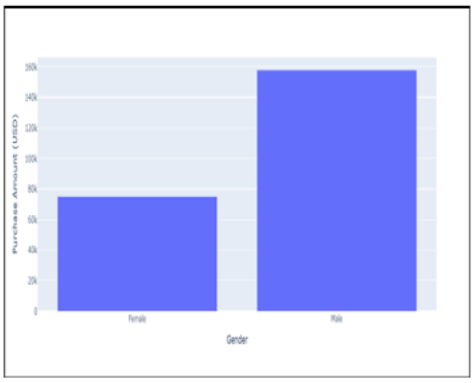

Fig:- 4 Bar chart showing average purchase amount with respect to gender.

In Fig 5

The bar chart displays average purchase amounts by location (U.S. states). While most states show similar

spending patterns around $55–$65, a few states like South Dakota and Alabama stand out with higher average purchases. This suggests regional

variations in consumer spending behavior across different U.S. locations in the dataset.

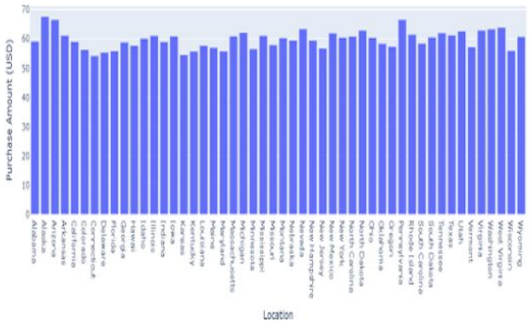

Fig:- 5 Bar chart showing of purchase Amount

5. CONCLUSION

This study has clearly demonstrated data analysis to be a process capable of distinguishing shopping trends and giving crucial insight into business development. From analyzing purchasing behaviors, shipping preferences and the impact of discounts on categories and geographical locations, we were able to unravel some valuable insights that firms could use in sharpening their processes for operations, market strategies, and customer engagement.

The created visualizations allow for a clear understanding of the data and complex trends understandable to the stakeholders. This initiative, therefore, highlights the importance of decision based on data and provides a foundation for further work in predictive analytics, real-time observation, and personalized customer interaction. The insights obtained from this study have the potential to positively impact organizational activities; they may be helpful in allowing businesses to better serve client's needs to improve efficiency, and to respond more quickly to the demands of the marketplace. As the science of data evolves, the use of advanced techniques like machine learning and real-time analytics should meaningfully enhance the accuracy and applicability of insights gained from studies of this sort.

REFERENCES

1) Kotler, P. (2000). *Marketing Management* (10th ed.). Pearson Education.
2) Neslin, S. A., et al. (1994). "A Model for Evaluating the Profitability of Customer Promotions." *Marketing Science*, 13(3), 204–220.
3) Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
4) Tsiptsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. Wiley.
5) Kumar, V., & Reinartz, W. (2016). *Creating Enduring Customer Value*. Journal of Marketing, 80(6), 36–68.
6) Redman, T. C. (1998). "The Impact of Poor Data Quality on the Typical Enterprise." *Communications of the ACM*, 41(2), 79–82.
7) Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830.
8) McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
9) Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering*, 9(3), 90–95.
10) Sharma, A., & Goyal, D. (2020). "Customer Segmentation and Behavior Analysis in E-Commerce Using Machine Learning." *IEEE Access*, 8, 144629–144649.