IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CONSUMER BEHAVIOUR ANALYSIS AND PREDICTION IN E-COMMERCE USING MACHINE LEARNING

Jiya Patel
Student
Department of DS
Drs. Kiran & Pallavi Patel Global University,
Vadodara, Gujarat, India

Vruti Parikh
Assistant Professor
Department of CSE
Drs. Kiran & Pallavi Patel Global University,
Vadodara, Gujarat, India

***Abstract* - The rapid growth of e-commerce platforms has resulted in the generation of large volumes of transactional data. Analyzing this data is essential for understanding business performance and improving decision-making processes. This study focuses on applying exploratory data analysis (EDA) techniques to evaluate e-commerce sales data using Python.**

The analysis includes monthly sales trends, category-wise and sub-category-wise performance, profit distribution, and customer segment behavior. Special emphasis is placed on the sales-to-profit ratio to evaluate the efficiency of converting sales into profit. Python libraries such as Pandas, NumPy, and visualization tools are used to process and represent the data.

The results highlight that sales performance varies across different months and categories, and higher sales do not always correspond to higher profit. The study demonstrates how simple data analysis techniques can provide meaningful insights for improving business strategies.

***Keywords* - Consumer Behaviour, Purchase Intention, E-commerce, Machine Learning, E-commerce, Data Analysis, Sales Analysis, Profit Analysis, Data Visualization, Python, EDA.**

I.INTRODUCTION

E-commerce platforms generate a large amount of data related to sales, products, and customers. Analyzing this data is important for understanding business trends and improving decision-making [1]. However, raw data is often complex and difficult to interpret.

Data analysis techniques help in identifying patterns and trends in the data. By analyzing sales and profit data, businesses can understand which products perform well and which areas need improvement [5]. Visualization techniques further help in presenting data in an easy-to-understand format [4].

This study focuses on analyzing e-commerce data using Python to extract meaningful insights related to sales, profit, and customer behavior.

II.LITERATURE REVIEW

The literature on e-commerce data analysis highlights the growing importance of extracting meaningful insights from large-scale transactional datasets [9]. Early studies primarily focused on descriptive statistical methods to analyze sales trends and customer behavior. These approaches were useful in summarizing historical data but lacked the ability to deeply explore relationships between multiple variables such as time, product categories, and customer segments [13].

Recent research emphasizes the use of Exploratory Data Analysis (EDA) techniques combined with visualization tools to better understand e-commerce datasets. Studies have shown that analyzing sales patterns over time helps in identifying seasonal trends and demand fluctuations, which are critical for inventory planning and marketing strategies. Similarly, category-wise analysis has been widely used to identify high-performing and low-performing product segments [1].

Customer segmentation has also been a major focus in existing literature. Researchers have used grouping techniques to classify customers based on their purchasing behavior, which helps businesses target specific groups more effectively. Visualization techniques such as bar charts, line graphs, and heatmaps are commonly used to represent these insights in a simple and interpretable format.

However, many existing studies focus either on advanced machine learning models or limited descriptive analysis. There is still a need for a balanced approach that provides both simplicity and meaningful insights. This study addresses this gap by applying structured EDA techniques to extract actionable insights from e-commerce data without relying on complex predictive models.

III.COMPARATIVE REVIEWS

The study by Zhou and Hudin (2024) focuses on analyzing time-based consumer behavior in e-commerce platforms. Their research highlights how time-series patterns in user interactions can help understand customer engagement trends. The study emphasizes the importance of analyzing behavioral data over time to identify purchasing patterns and improve marketing strategies.

Sharma et al. (2024) proposed an e-commerce sales prediction model using artificial neural networks combined with optimization techniques. Their work demonstrates that neural networks can effectively handle complex sales data and improve prediction accuracy. This study highlights the potential of AI-based methods in forecasting demand and improving business planning [2].

Wang and Chen (2024) explored fuzzy logic-based approaches for optimizing supply chain operations in e-commerce systems. Their findings show that fuzzy logic can handle uncertainty in supply chain decisions and improve operational efficiency. The study contributes to better inventory and logistics management in online retail systems.

Singh and Verma (2025) conducted a study on exploratory data analysis (EDA) of e-commerce datasets along with recommendation techniques. Their research shows that EDA plays a key role in identifying meaningful patterns in customer data. The study highlights how simple analytical techniques can support decision-making in recommendation systems.

Gupta and Jain (2024) performed a comparative study between traditional forecasting methods and machine learning approaches for e-commerce sales prediction. Their findings indicate that machine learning models generally outperform traditional statistical methods. However, they also note that model performance depends heavily on data quality.

Li et al. (2025) proposed a hybrid ARIMA and LSTM model for e-commerce sales forecasting. Their study demonstrates that combining statistical and deep learning models improves prediction accuracy. The research highlights the importance of hybrid approaches in time-series forecasting problems [6].

Kim and Park (2024) focused on deep learning-based sales prediction in live-streaming commerce. Their model effectively captures real-time consumer behavior during live sessions. The study shows that deep learning can significantly improve accuracy in dynamic e-commerce environments [7].

Kumar and Singh (2024) studied product-based sales forecasting using consumer browsing data. Their research emphasizes that browsing behavior provides valuable insights into customer intent. The study highlights the importance of behavioral data in improving forecasting accuracy.

Chen and Zhang (2024) conducted a comparative analysis of machine learning techniques for e-commerce analytics [9]. Their work shows that different algorithms perform differently based on dataset characteristics. The study helps in selecting appropriate models for specific business problems.

Brown and Wilson (2025) analyzed retail demand forecasting using tree-based machine learning models. Their findings show that models like Random Forest and Gradient Boosting are efficient for structured retail data. The study also highlights their ability to handle large datasets effectively [11].

Liu and Zhao (2024) proposed a transformer-based demand prediction model that incorporates external factors. Their research shows that attention-based models can capture long-term dependencies in data. The study improves forecasting accuracy by considering external influences.

Davis and Clark (2023) explored multivariate retail demand forecasting using regression and machine learning techniques [13]. Their study shows that combining multiple variables improves prediction performance. It emphasizes the importance of considering multiple factors in forecasting models.

Khan and Ali (2025) focused on improving e-commerce performance using optimized LSTM models [14]. Their research demonstrates that optimized deep learning models can enhance prediction accuracy. The study highlights the importance of tuning model parameters for better results.

IV.METHODOLOGY

The methodology adopted in this study follows a structured pipeline consisting of data collection, preprocessing, analysis, and visualization. The primary objective is to systematically analyze e-commerce data to extract meaningful business insights.

The dataset used in this study contains key attributes such as order date, sales, profit, product category, sub-category, and customer segment [8]. These attributes provide a comprehensive view of business operations and customer behavior. The dataset is suitable for performing time-based, category-based, and segment-based analysis.

Data preprocessing is an essential step in ensuring data quality. In this study, preprocessing includes handling missing values, correcting data types, and converting the order date into a DateTime format. From this, additional features such as month and year are extracted to enable time-series analysis. Duplicate records are also removed to maintain data consistency.

After preprocessing, exploratory data analysis (EDA) is performed. This includes grouping and aggregating data based on different attributes [16]. Monthly sales and profit trends are analyzed to identify performance variations over time. Category-wise and sub-category-wise analysis is performed to evaluate product performance. Customer segment analysis is conducted to understand how different groups contribute to sales and profit.

Finally, data visualization techniques are applied to represent the findings. Graphical tools such as line charts, bar charts, and pie charts are used to make the results more interpretable. Python libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly are used throughout the implementation process.

V.IMPLEMENTATION

The implementation of the project is carried out using Python programming in the Jupyter Notebook environment. The process begins with importing the dataset using the Pandas library, which allows efficient handling of structured data.

Once the dataset is loaded, initial exploration is performed to understand its structure, including column names, data types, and sample records. This helps in identifying the key attributes required for analysis.

Data cleaning is then performed to ensure consistency and accuracy. Missing values are handled appropriately, and unnecessary or duplicate records are removed. The “Order Date” column is converted into a DateTime format, which enables the extraction of time-based features such as month and year.

After preprocessing, the dataset is used for analysis. Grouping operations are applied using Pandas to calculate total sales and profit across different dimensions such as time, category, sub-category, and customer segment. Sorting functions are used to identify top-performing and low-performing segments [11].

Visualization is an important part of implementation. Line graphs are used to show monthly trends, bar charts are used for category comparisons, and pie charts are used to show proportional distributions [1]. These visualizations help in better understanding the patterns present in the dataset.

The implementation demonstrates how Python can be effectively used for data analysis in a structured and efficient manner, providing meaningful insights from raw e-commerce data.

VI.RESULTS AND DISCUSSION

The analysis of the e-commerce dataset provides several important insights into sales performance, profitability, and customer behavior. The monthly sales analysis shows significant variations across different months, indicating the presence of seasonal trends in customer purchasing behavior.

Category-wise analysis reveals that certain product categories contribute a major portion of total sales, while others have relatively lower performance [5]. This highlights the importance of focusing on high-performing categories for revenue generation.

Sub-category analysis provides a more detailed breakdown of product performance. It identifies specific products that generate high sales as well as those that underperform. This information is useful for inventory management and product optimization.

Profit analysis shows that high sales do not always correspond to high profit [13]. Some categories generate high revenue but lower profit margins due to operational costs or pricing strategies. This highlights the importance of analyzing profitability alongside sales.

Customer segment analysis shows differences in purchasing behavior among various customer groups [16]. Some segments contribute more to sales, while others generate higher profit margins. This information is useful for targeted marketing strategies.

Segment	Sales to Profit Ratio
Consumer	8.66
Corporate	7.68
Home Office	7.13

The sales-to-profit ratio analysis further highlights inefficiencies in certain areas where high sales do not translate into proportional profit. This helps in identifying areas that require business optimization.

Overall, the results demonstrate that data analysis can provide deep insights into business performance and support better decision-making.

VII.CONCLUSION AND FUTURE WORK

A. Conclusion

This study demonstrates the effectiveness of exploratory data analysis techniques in analyzing e-commerce datasets. The analysis successfully identifies key trends in sales, profit, product performance, and customer behavior.

The results show that sales performance varies across different months and categories. It is also observed that higher sales do not always lead to higher profit, emphasizing the importance of profitability analysis. Customer segment analysis provides additional insights into purchasing behavior.

Overall, the study highlights the importance of data-driven decision-making in e-commerce businesses. By analyzing raw data, meaningful insights can be extracted to improve business strategies and performance.

B. Future Work

Although this study focuses on exploratory data analysis, there are several possible extensions for future improvement. Machine learning techniques can be applied to predict future sales and customer behavior based on historical data.

The project can also be extended by integrating real-time data streams from e-commerce platforms to enable dynamic analysis. Interactive dashboards can be developed using tools such as Power BI or Tableau for better visualization and decision support.

Additionally, advanced analytical techniques such as customer clustering and recommendation systems can be implemented to enhance customer targeting. Future work can also focus on combining structured and unstructured data to gain deeper insights into consumer behavior.

VIII.REFERENCES

[1] Z. Zhou and Y. Hudin, “Time analysis of online consumer behavior in e-commerce platforms,” *Journal of Marketing Analytics*, vol. 12, no. 2, pp. 145–160, 2024.

[2] A. Sharma, R. Patel, and S. Mehta, “E-commerce sales prediction using artificial neural networks and optimization algorithms,” *Scientific Reports*, vol. 14, no. 1, pp. 1–15, 2024.

[3] L. Wang and T. Chen, “Fuzzy logic-based supply chain optimization in e-commerce systems,” *Scientific Reports*, vol. 14, no. 2, pp. 1–12, 2024.

[4] M. Singh and P. Verma, “Exploratory data analysis of e-commerce datasets with recommendation techniques,” *International Journal of Business Intelligence*, vol. 20, no. 5, pp. 258–270, 2025.

[5] K. Gupta and R. Jain, “A comparative study of traditional and machine learning approaches for e-commerce sales forecasting,” *Journal of Data Science and Analytics*, vol. 9, no. 3, pp. 112–125, 2024.

[6] H. Li, X. Zhao, and Y. Sun, “Hybrid ARIMA and LSTM model for e-commerce sales forecasting,” *Mathematics*, vol. 13, no. 11, p. 1838, 2025.

[7] J. Kim and S. Park, “Deep learning-based sales prediction in live-streaming commerce,” *Scientific Reports*, vol. 14, no. 3, pp. 1–14, 2024.

[8] R. Kumar and A. Singh, “Product-based e-commerce sales forecasting using consumer

browsing data," *International Journal of Forecasting*, vol. 40, no. 2, pp. 210–225, 2024.

[9] Y. Chen and L. Zhang, "Comparative analysis of machine learning techniques for e-commerce analytics," *Applied Economics and Management Sciences*, vol. 5, no. 1, pp. 45–60, 2024.

[10] S. Patel and D. Shah, "Enhanced e-commerce forecasting using EEMD-integrated LSTM models," *Journal of Retail Analytics*, vol. 11, no. 2, pp. 78–92, 2025.

[11] P. Brown and T. Wilson, "Retail demand forecasting using tree-based machine learning models," *IEEE Access*, vol. 13, pp. 45678–45690, 2025.

[12] X. Liu and H. Zhao, "Transformer-based demand prediction incorporating external factors," *Expert Systems with Applications*, vol. 240, p. 120123, 2024.

[13] R. Davis and M. Clark, "Multivariate retail demand forecasting using regression and machine learning techniques," *Journal of Retailing and Consumer Services*, vol. 70, p. 103145, 2023.

[14] A. Khan and S. Ali, "Improving e-commerce performance using optimized LSTM models," *Scientific Reports*, vol. 15, no. 1, pp. 1–10, 2025.

[15] N. Gupta and P. Arora, "Sentiment-aware deep learning models for sales prediction in e-commerce," *Knowledge-Based Systems*, vol. 280, p. 110987, 2024.

[16] V. Reddy and K. Kumar, "Customer segmentation in e-commerce using K-means clustering," *Procedia Computer Science*, vol. 215, pp. 345–352, 2023.