

Sales Data Analysis For Retail & E-Commerce

Aalesh Ganpat Jadhao

Student, Department of CSE (Data Science)

Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India

aaleshjadhao2k3@gmail.com

Abstract: *The project titled “Sales Data Analysis for Retail and E-commerce” focuses on analysing Amazon sales data for the year 2025 to understand key business trends, customer behaviour, and sales performance. The main objective of this study is to identify patterns and insights that can help retail and e-commerce businesses improve their decision-making processes, product strategies, and marketing effectiveness. The analysis was conducted using Google Colab, employing Python-based libraries such as NumPy, Pandas, Matplotlib, and Seaborn for data cleaning, processing, and visualization. The dataset consisted of various sales-related attributes such as product category, quantity, price, region, order status, and profit.*

Keywords: Sales Data Analysis, Amazon Sales, Data Visualization, Google Colab, Python, Pandas, Matplotlib, Seaborn, Exploratory Data Analysis (EDA)

I. INTRODUCTION

Data analytics plays a vital role in this digital transformation. Through effective analysis of sales data, businesses can understand consumer purchasing patterns, identify high-performing products, evaluate seasonal trends, and forecast future demand. The insights generated through data analytics empower companies to make data-driven decisions, optimize marketing strategies, and enhance inventory management. Therefore, the application of Exploratory Data Analysis (EDA) on real-world sales datasets, such as Amazon’s sales data, provides a strong foundation for learning and implementing modern analytical techniques.

II. METHODOLOGY

Data Collection:

The dataset used is “Amazon Sales Data 2025,” which contains records of various sales transactions including order ID, product category, region, customer type, sales value, profit margin, and shipping.

Data Preprocessing:

This includes handling missing values, removing duplicates, correcting data types, and cleaning inconsistent data for accurate analysis.

Exploratory Data Analysis (EDA):

Descriptive statistics and visualizations are used to explore trends, correlations, and patterns in the data. Tools such as Python, Pandas, Matplotlib, and Seaborn are utilized for analysis and visualization.

Insights and Interpretation:

Key findings are derived from the EDA results to provide conclusions about sales performance and customer behavior.

Reporting and Visualization:

The outcomes are presented through graphical representations, summary tables, and interactive charts to make insights easy to interpret.

III. RESULTS AND DISCUSSION

The implementation of the project highlighted the value of Python programming and its extensive libraries, including Pandas, NumPy, Matplotlib, Seaborn, and Plotly, for conducting efficient data analysis and visualization. Using platforms such as Google Colab allowed for flexible, cloud-based execution, enabling handling of large datasets



without extensive local computational resources. The step-by-step application of code demonstrated the practical workflow of a data analyst, from initial data exploration to generating insights through graphs, dashboards, and statistical summaries.

A. Figures

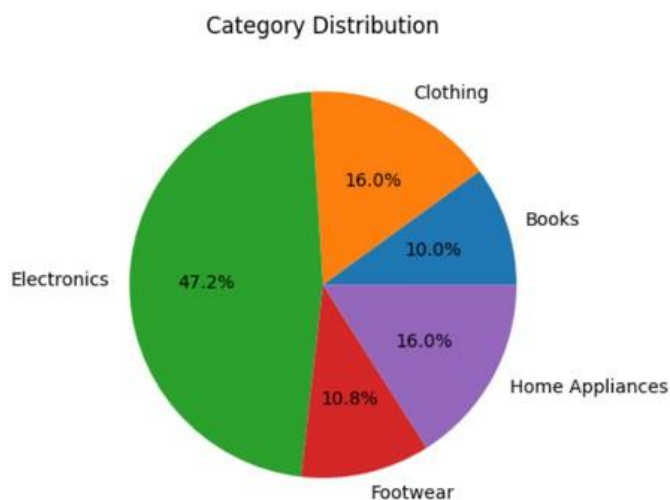


Fig. 1. Pie Chart Represented Category Distribution

Used for showing the proportional contribution of different categories or customer segments to total sales. They provide a quick snapshot of the distribution of products, regions, or customer types.

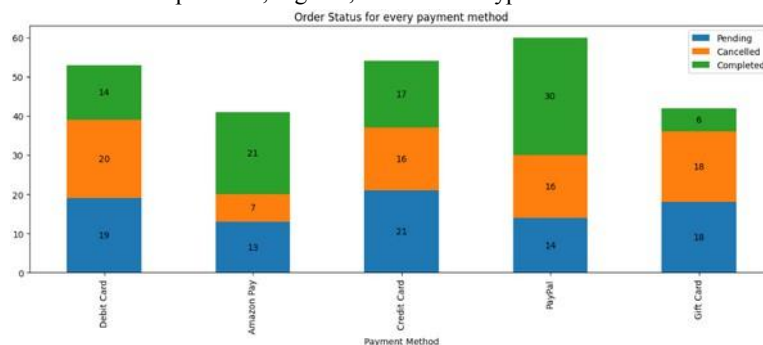


Fig. 2. Boxplot Represented Order Status

Display the distribution of sales or profit values, highlighting median, quartiles, and outliers. Boxplots help detect anomalies, identify data spread, and analyze variability in sales performance.



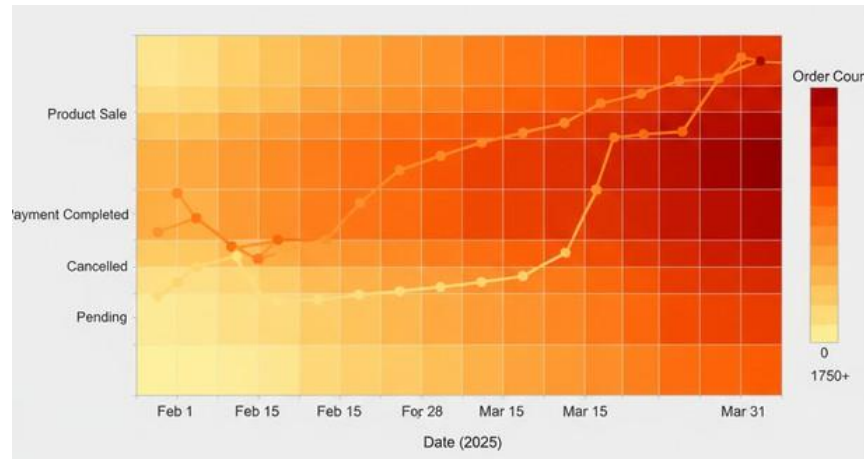


Fig. 3. Heatmap Represented Matrices

Represent correlations between variables, such as sales, profit, and discounts, using color intensity. Heatmaps make it easier to identify positive or negative relationships and dependencies among different metrics.

IV. CONCLUSION

The project “Sales Data Analysis for Retail and E-commerce” aimed to analyze the Amazon Sales Data 2025 using data analytics techniques to uncover trends, patterns, and actionable insights. The study focused on applying a structured methodology for data import, cleaning, preprocessing, exploratory data analysis (EDA), visualization, and interpretation, reflecting the practices commonly used in the retail and e-commerce domain.

V. ACKNOWLEDGMENT

With profound feeling of immense gratitude and affection, I would like to thank my guide Prof. Ujwala Khartad, Computer Science and Engineering Data Science for her continuous support, motivation, enthusiasm and guidance. Her encouragement, supervision with constructive criticism and confidence enabled me to complete this project.

REFERENCES

- [1]. GeeksforGeeks. (2025, July 23). Top 15 Python libraries for data analytics. Retrieved from <https://www.geeksforgeeks.org/blogs/python-libraries-for-data-analytics/>
- [2]. Coursera. (2025, August 13). Python libraries for data analysis: Essential tools. Retrieved from <https://www.coursera.org/articles/python-libraries-for-data-analysis>

