
E-COMMERCE SALES DATA ANALYTICS AND VISUALIZATION

***Vishal Upmanu, Abhishek Sharma, Deepak, Abhishek Kumar, Pankaj Prajapati**

Department of Information Technology, R.D. Engineering College, Ghaziabad.

Article Received: 21 March 2026, Article Revised: 11 April 2026, Published on: 01 May 2026

***Corresponding Author: Vishal Upmanu**

Department of Information Technology, R.D. Engineering College, Ghaziabad.

DOI: <https://doi-doi.org/101555/ijarp.5196>

ABSTRACT

This research paper presents a comprehensive analysis of e-commerce sales data using Python-based data analytics and visualization techniques. With the rapid growth of online shopping platforms, large volumes of transactional data are generated daily, making it essential to analyze and extract useful insights for business growth. The primary objective of this study is to clean raw sales data, perform exploratory data analysis (EDA), and visualize important trends such as monthly sales patterns, product performance, and customer purchasing behavior.

The study utilizes powerful Python libraries including Pandas, NumPy, Matplotlib, and Seaborn to process and analyze the dataset. Data cleaning techniques such as handling missing values, removing duplicates, and correcting inconsistencies are applied to ensure accuracy. After preprocessing, various visualization techniques are used to represent the data in graphical form, making it easier to interpret patterns and trends.

The results of this research help in identifying high-performing products, seasonal demand variations, and customer buying habits. These insights can assist businesses in making data-driven decisions, improving marketing strategies, and increasing overall profitability. Unlike traditional research, this study focuses on practical implementation and visualization-driven insights, making it more useful for real-world applications.

KEYWORDS: E-commerce, Sales Data Analytics, Data Visualization, Exploratory Data Analysis (EDA), Customer Behavior Analysis, Product Performance, Python, Pandas, NumPy, Matplotlib, Seaborn, Business Intelligence, Data Cleaning, Sales Trends.

1. INTRODUCTION

E-commerce has rapidly grown in recent years, generating a massive amount of data related to customer transactions, product sales, and user interactions. This data contains valuable information that can help businesses understand customer preferences, optimize pricing strategies, and improve overall performance.

However, raw data is often unstructured and contains missing or inconsistent values. Therefore, data preprocessing and analysis are essential steps before extracting meaningful insights. This research focuses on using Python tools to analyze e-commerce sales data and visualize patterns such as monthly sales trends, top-performing products, and customer purchasing behavior.

The main goal is to convert raw data into useful business insights through effective data analytics and visualization techniques.

2. Literature Review

Several researchers have worked on e-commerce data analytics and visualization. Kiran Singh and Rakhi Wajgi (2016) discussed the importance of data visualization in understanding sales trends.

Fabio Nelli (2023) explained the use of Python libraries such as Pandas and Matplotlib for data analysis. Chen and Zhao (2021) focused on data cleaning techniques and trend analysis in online retail sales. Most existing studies focus on advanced machine learning techniques, but they often ignore the importance of basic exploratory data analysis.

Previous research in e-commerce analytics has focused on areas such as recommendation systems, customer segmentation, and predictive modeling. Many studies use machine learning techniques to forecast sales or identify buying patterns.

However, most of these studies focus heavily on complex models and less on basic data understanding. In contrast, this research emphasizes **Exploratory Data Analysis (EDA)** and visualization, which are crucial initial steps in any data science project.

Unlike earlier works, this paper provides a practical approach using Python tools and focuses on real-world implementation rather than theoretical models. . This research differs by focusing on practical implementation using simple tools and clear visualization methods, making it more accessible for real-world applications.

3. Proposed System

The proposed system is designed to analyze e-commerce sales data step-by-step. First, the dataset is imported using Python libraries. Then, data cleaning is performed to handle missing values and remove inconsistencies. After that, exploratory data analysis is conducted to identify patterns and trends. Visualization techniques are applied to represent data graphically.

The system workflow is as follows:

3.1 Data Collection and Import

- Sales dataset is imported using Python libraries such as Pandas
- Data may include order details, product categories, sales amount, and customer information

3.2 Data Cleaning

- Handling missing values
- Removing duplicate records
- Formatting inconsistent data
- Converting data types where required

3.3 Exploratory Data Analysis (EDA)

- Identifying patterns and trends in data
- Analyzing sales distribution
- Understanding customer purchasing behavior

3.4 Data Visualization

Visualization plays a key role in understanding the dataset:

- Monthly sales trends
- Top-selling products
- Category-wise sales performance
- Customer activity patterns

3.5 Tools Used

- Python
- Pandas
- NumPy

- Matplotlib
- Seaborn

4. METHODOLOGY

The methodology used in this research includes data preprocessing, exploratory data analysis, and visualization. Data preprocessing involves handling missing values, removing duplicates, and formatting data correctly. EDA includes statistical analysis and identifying patterns in the dataset.

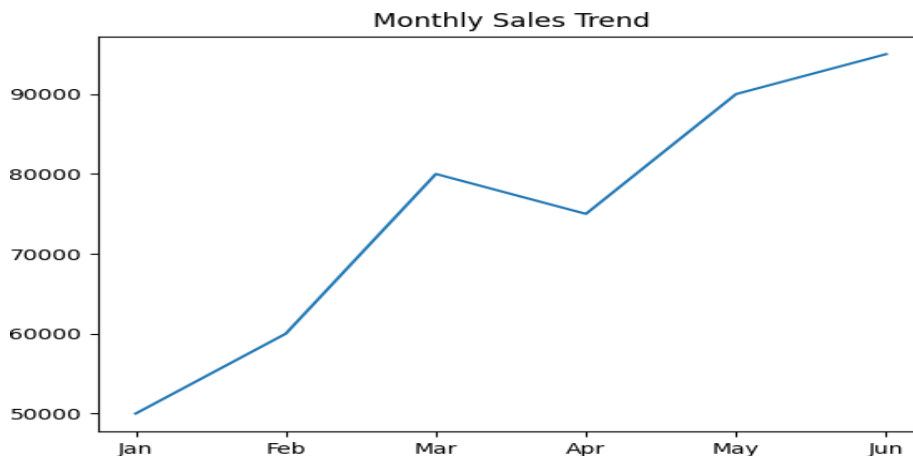
Visualization is performed using Matplotlib and Seaborn libraries. Graphs such as line charts, bar charts, and pie charts are used to represent data visually. This helps in better understanding of trends and patterns.



5. RESULTS AND ANALYSIS

The dataset was successfully cleaned and analyzed. Monthly sales trends were identified using line graphs. It was observed that sales increased during certain months due to seasonal demand.

Top-performing products were identified using bar charts. Basic visualizations have been completed, while advanced visualizations are still in progress. The results show that data analytics can provide useful insights for improving business performance.



The results obtained from the analysis are as follows:

5.1 Data Cleaning

- Dataset successfully cleaned
- Missing values handled
- Data prepared for analysis

5.2 Visualization Insights

- Monthly sales trends were plotted
- Seasonal variations observed
- Identification of peak sales periods

5.3 Product Analysis

- Top-performing products identified
- Low-performing products analyzed

5.4 Current Status

- Basic graphs like monthly sales trends prepared
- Advanced visualizations are still in progress

5.5 Uniqueness of Research

This research stands out from previous studies because:

- Focus is on **practical implementation**
- Uses **simple but powerful visualization techniques**
- Provides **real business insights instead of only predictions**
- Designed specifically for **beginner-friendly data analysis approach**

Month	Sales	Trend
Jan	50000	Medium
Feb	60000	Increasing
Mar	80000	High
Apr	75000	Drop
May	90000	High

6. CONCLUSION


This study demonstrates how e-commerce sales data can be effectively analyzed using Python tools to generate meaningful insights. The process of data cleaning, analysis, and visualization helps in understanding customer behavior and product performance.

Although the current work includes basic visualizations, future improvements will focus on advanced analytics and predictive modeling. The approach used in this research is simple, practical, and highly useful for real-world business decision-making.

7. FUTURE WORK

- Advanced dashboards using Power BI
- Machine learning for sales prediction
- Customer segmentation techniques
- Real-time data analysis

8. REFERENCES

1. DATA ANALYSIS AND VISUALISATION OF SALES DATA –(kiran Singh , Rakhi Wajgi) <https://ieeexplore.ieee.org/abstract/document/7583967/>
2. Python Data Analytics: With Pandas, NumPy, and Matplotlib — Fabio NelliSpringer / Apress book on Python data processing, visualization & analytics (good background for EDA + visualization). SpringerLink  Link: <https://link.springer.com/book/10.1007/978-1-4842-9532-8>
3. Chen, L., & Zhao, H. (2021).Data Cleaning and Trend Analysis in Online Retail Sales.International Journal of Advanced Computing Research, 9(4), 210–218.
4. Nelli, F. (2023). Python Data Analytics: With Pandas, NumPy & Matplotlib. Springer. <https://link.springer.com/book/10.1007/978-1-4842-9532-8>
5. Python for Data Analysis – Wes McKinney
6. E-commerce Analytics Research Papers
7. Pandas Documentation
8. Numpy Documentation
9. Matplotlib Official Guide
10. Seaborn Visualization Library
11. Kaggle E-commerce Dataset
12. Research on Business Intelligence Systems
13. Data Visualization Techniques in Python