Visualizing Retail Sales Patterns

AVIK DAS,

2nd Year B.Sc. Computer Science, Bangabasi College

Period of Internship: 25th August 2025 - 19th September 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

1. Abstract

This project focused on analyzing an e-commerce dataset to gain insights into customer demographics, purchasing patterns, and product demand distribution. Using Python libraries such as Pandas, Matplotlib, Seaborn, and Plotly, the dataset was cleaned, pre-processed, and explored through descriptive statistics and visualizations. The analysis highlighted customer segmentation based on gender, age, and country, revealed sales trends, and identified popular products. The findings can be applied in real-world business decision-making for targeted marketing, inventory optimization, and customer engagement strategies.

2. Introduction

The growth of e-commerce has resulted in vast amounts of transactional data. Analyzing this data provides businesses with valuable insights into customer behavior and helps optimize decision-making.

In this project, we conducted an exploratory data analysis (EDA) on an e-commerce dataset. The dataset contained information on customers, transactions, countries, genders, and age groups. Python's data science stack (Pandas, NumPy, Matplotlib, Seaborn, Plotly) was used to process and visualize the data.

Topics covered in training during the first two weeks:

- Data collection and preprocessing techniques
- Exploratory data analysis (EDA)
- Visualization with Matplotlib, Seaborn, and Plotly
- Descriptive and inferential statistics
- Basics of hypothesis testing
- Hands-on practice with Python for data analysis

3. Project Objective

The objectives of this project are:

1. To perform exploratory analysis of customer demographics (gender, age, country).
2. To identify purchasing behavior patterns across users.
3. To analyze product-wise demand and sales distribution.
4. To create meaningful visualizations for business decision-making.
5. To recommend strategies for improving marketing and inventory management.

4. Methodology

- Data Collection: The dataset was sourced from Google Drive and imported into Pandas for processing.
- Data Cleaning: Duplicate values were handled, missing data was checked, and consistent formatting of features was ensured.
- Exploratory Data Analysis:
    - Customer demographics analysis (unique customers, repeat customers, gender distribution, age group breakdown).
    - Geographical distribution (top countries by customer count).
    - Sales and revenue trends (over time, across products).
- Visualization Tools: Matplotlib, Seaborn, and Plotly were used for interactive and static visualizations.
- Steps:
    1. Import libraries and dataset.
    2. Data exploration (head, shape, types).
    3. Cleaning & preprocessing.
    4. Customer-level analysis.
    5. Product-level analysis.
    6. Visualizations and interpretation of results.

Since no machine learning model was applied, the project focused on descriptive and inferential analysis rather than predictive modeling.

5. Data Analysis and Results

Descriptive Analysis:

- The dataset contained multiple unique customers with repeat purchases.

- Gender distribution showed a fairly even spread across male, female, and other categories.

- Age group distribution was balanced across young adults, middle-aged, and older demographics.

Geographical Insights:

- The highest number of unique customers came from China, India, and Japan.

Visualizations:

- Bar charts for gender and age distribution.

- Country-level distribution of unique customers.

- Additional charts for revenue trends and product performance.

Findings:

- Balanced demographic representation indicates wide user appeal.

- Key markets are concentrated in Asia.

- Potential to target repeat customers with loyalty programs.

6. Conclusion

Write the conclusions made after doing the project work, justify your conclusions by quoting your findings. May include recomendation for future work.

7. APPENDICES

References: Tutorials and documentation on Pandas, Seaborn, Matplotlib, Plotly.

GitHub Link: https://github.com/dasavik09/TIH-Internship-Section-1-.git

Dataset Source:

https://drive.google.com/file/d/1dx9sdncdQ3dY4Fqsjo77rL2oaC3zw-fU/view?usp=sharing