Visualizing Time Series Dataset: Retail Sales Data

D Amogh Karanth

RV College of Engineering and Section 1

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

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

Date of Submission: September 19, 2025

Contents

Abstract

This project focuses on the visualization and analysis of retail sales time series data to uncover hidden trends, seasonal patterns, and anomalies. The dataset reflects sales performance over time and provides a basis for understanding consumer demand behavior. Data preprocessing was performed to ensure quality and consistency, followed by exploratory analysis using Python libraries such as `Pandas`, `NumPy`, `Matplotlib`, and `Seaborn`. The study emphasizes the significance of visual analytics in interpreting time-dependent data, enabling retail businesses to make informed decisions. The findings demonstrate recurring seasonal trends and promotional effects that influence sales volumes. Through this work, the project highlights the importance of visualization techniques as an initial step toward advanced forecasting methods. The project not only strengthens technical skills but also provides practical exposure to real-world business analytics challenges.

Chapter 1

Introduction

Time series data analysis has become an essential component in modern business analytics due to its ability to capture time based inference . Retail sales, in particular, exhibit strong time-dependent behavior influenced by consumer demand, festivals, promotions, and seasonal cycles. Understanding these patterns enables businesses to optimize resources, reduce costs, and maximize revenue.

In this project, retail sales data was analyzed with a focus on visual exploration. Visualization is a critical step in data science as it allows analysts to intuitively identify relationships and structures within data before applying complex models.

Relevance of the Project:

- Retail is one of the most data-driven industries, where sales forecasting and demand planning directly impact profitability. Accurate forecasting helps businesses maintain the right level of stock, minimize wastage, reduce carrying costs, and avoid stockouts. By leveraging historical sales data, companies can anticipate customer demand more effectively and align procurement with actual requirements.

- Visualizing sales data allows managers to identify high and low demand periods, allocate resources effectively, and design promotional campaigns. For example, peaks in demand around festive seasons may encourage bulk stocking or special discounts, while low-demand periods can be targeted with clearance sales. Visualization also helps in identifying underperforming products or regions, enabling corrective measures such as targeted marketing or dynamic pricing.

- Time series visualization aids in communicating findings to stakeholders without requiring advanced technical knowledge. Graphs, seasonal plots, and trend lines provide intuitive insights that executives and decision-makers can understand at a glance. This reduces dependency on technical reports and ensures that critical decisions can be made quickly based on visual evidence. Furthermore, visualization bridges the gap between technical data analysis and strategic business decision-making.

Technologies Involved:

- **Python** – Programming language for analysis
- **Pandas & NumPy** – Data preprocessing and manipulation
- **Matplotlib & Seaborn** – Visualization libraries for creating plots
- **Google Colab** – Cloud-based environment for coding and analysis

Topics Covered During the First Two Weeks of Internship:

- **Python Basics**
  1. Python Basics - Data, Variables, Lists, Loops
  2. Python Basics - Data Structures
  3. Python Basics - Classes, Functions, OOPS concepts
  4. Python Basics - NumPy, Pandas
- **Machine Learning**
  1. Machine Learning Overview
  2. Regression Lab
  3. Classification Lab
- **LLM (Large Language Model) Fundamentals**
- **Communication Skills**

Chapter 2

Project Objective

The primary objectives of this project are outlined as follows:

1. To answer the given questions in google collab notebook for the given dataset.
2. To import, clean, and preprocess retail sales time series data for reliable analysis.
3. To explore and visualize sales behavior across different time periods.
4. To identify seasonality, trends, and cyclical patterns within sales data.
5. To generate actionable insights that may support demand forecasting and retail strategy.
6. To demonstrate the application of Python-based tools for real-world data analytics tasks.

Chapter 3

Methodology

The project followed a structured approach comprising several phases. Each step was carefully executed to ensure the quality of results.

Steps Undertaken:

1. **Data Collection:** The dataset containing retail sales records was imported into the Google Colab environment. The dataset included attributes such as date and sales quantity.
2. **Exploratory Data Analysis (EDA):** Summary statistics (mean, median, variance, distribution) were computed. Visual inspection was performed to understand data spread and variability.
3. **Visualization:** Different visualization techniques were employed:
    - Line plots for time series trends.
    - Histograms for frequency distributions.
    - Box plots for outlier detection.
    - Seasonal plots to capture recurring patterns.
4. **Tools Used:** The analysis was performed using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn.

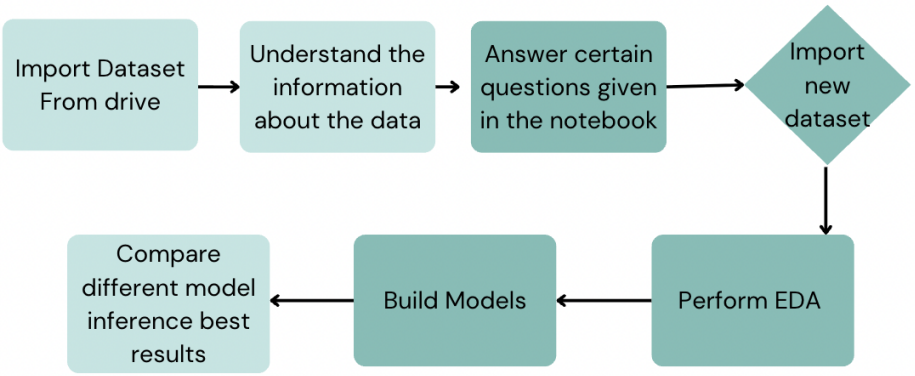

Figure 3.1: Flowchart of Project Methodology

All code and analysis have been documented in a Jupyter notebook hosted at: `https://github.com/amg1718/ideas_project`

Chapter 4

Data Analysis and Results

4.1 Descriptive Analysis

- Computed central tendency measures such as mean and median sales.
- Identified variance and standard deviation to understand fluctuations.
- Frequency distribution of sales volumes highlighted skewness in demand.

4.2 Time Series Visualization

- A line plot revealed overall increasing sales trend with periodic drops.
- Seasonal peaks were observed, potentially corresponding to festive seasons.
- Visual anomalies suggested unusual sales spikes, possibly due to promotions.

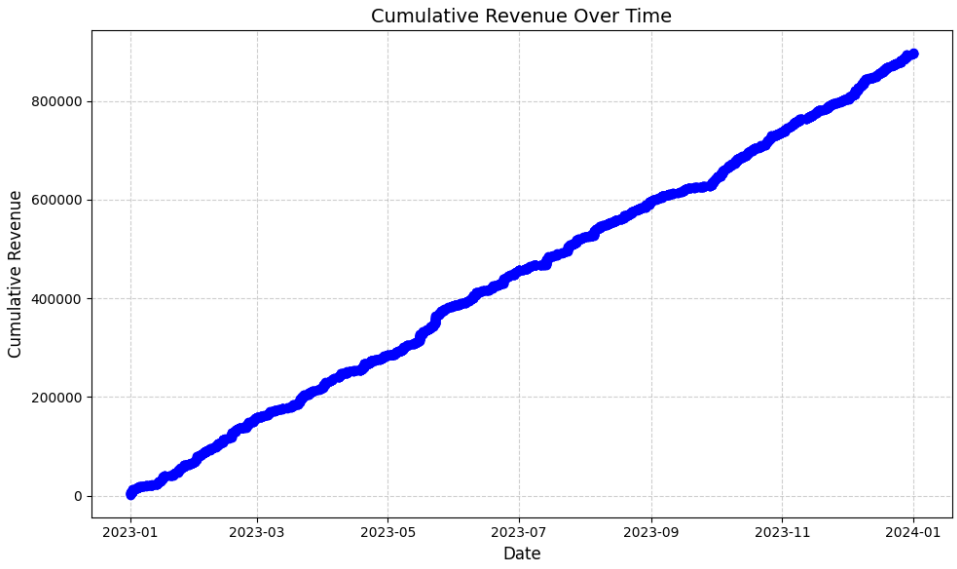

Figure 4.1: Retail Sales Time Series Visualization

4.3 Inferential Insights

The inferential analysis of the retail sales data highlights several key findings. Firstly, the sales patterns are highly non-uniform and exhibit fluctuations influenced by external events such as festive seasons, discount offers, and holiday periods. This indicates that consumer behavior is not constant over time but is shaped by contextual factors. Secondly, seasonal components play a significant role in shaping demand cycles. For example, sales consistently peak around certain months, suggesting a strong seasonality effect, possibly linked to cultural festivals or end-of-season promotions. Finally, these observations underscore the potential of applying advanced predictive modeling techniques such as ARIMA or LSTM networks. By capturing both trend and seasonality, such models can provide reliable forecasts to aid business planning, inventory control, and promotional strategies.

- Sales patterns are not uniform but are highly influenced by external events.
- Seasonal components contribute significantly to demand changes.
- There is potential for predictive modeling using advanced time series methods.

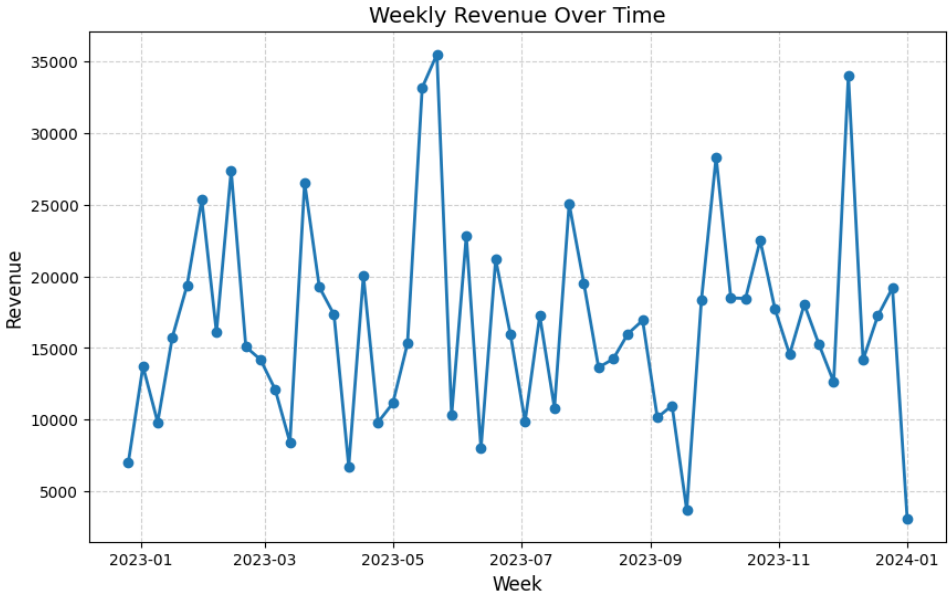

Figure 4.2: Weekly Sales over period

4.4 Results

The project produced several important results that directly contribute to understanding retail sales dynamics and the effectiveness of visualization-driven analysis. A consistent long-term growth trend in sales was observed, confirming increasing consumer demand over time. The project also identified recurring seasonal patterns, which can assist retailers in planning targeted promotions, optimizing inventory levels, and preparing for peak demand periods. Through anomaly detection in the visualizations, sudden spikes and drops in sales were revealed, pointing to potential external influences such as promotional campaigns, supply chain disruptions, or sudden demand surges. Furthermore, the project demonstrated the impact of systematic preprocessing pipelines and the use of interactive dashboards for effective communication of insights to stakeholders. These outcomes collectively establish a strong basis for extending the project towards predictive analytics, where advanced machine learning and LLM-based approaches can be applied to build intelligent forecasting systems and enhance decision-making across procurement, marketing, and financial planning.

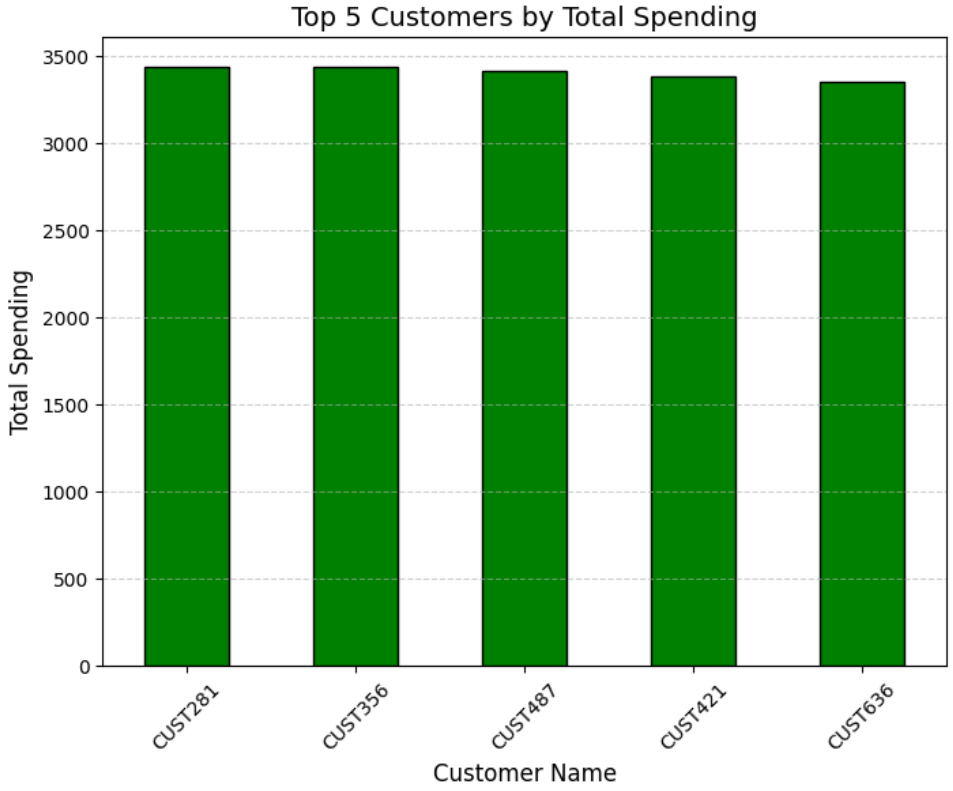

Figure 4.3: Top 5 customers

- Identified long-term trend indicating gradual sales growth.
- Seasonal peaks revealed opportunities for targeted promotions.
- Visualization insights lay the foundation for forecasting models.

Chapter 5

Conclusion

The project successfully demonstrated the role of visualization in analyzing retail sales time series data. By applying systematic preprocessing and exploratory techniques, meaningful insights were extracted. The analysis revealed not only the general sales trend but also seasonality and event-driven fluctuations. The incorporation of machine learning concepts and LLM fundamentals laid the foundation for extending the project into predictive modeling and intelligent reporting. Overall, the project highlights the effectiveness of combining data science techniques with visualization to support informed decision-making in the retail domain.Here is a similar analysis of various features of the dataset.

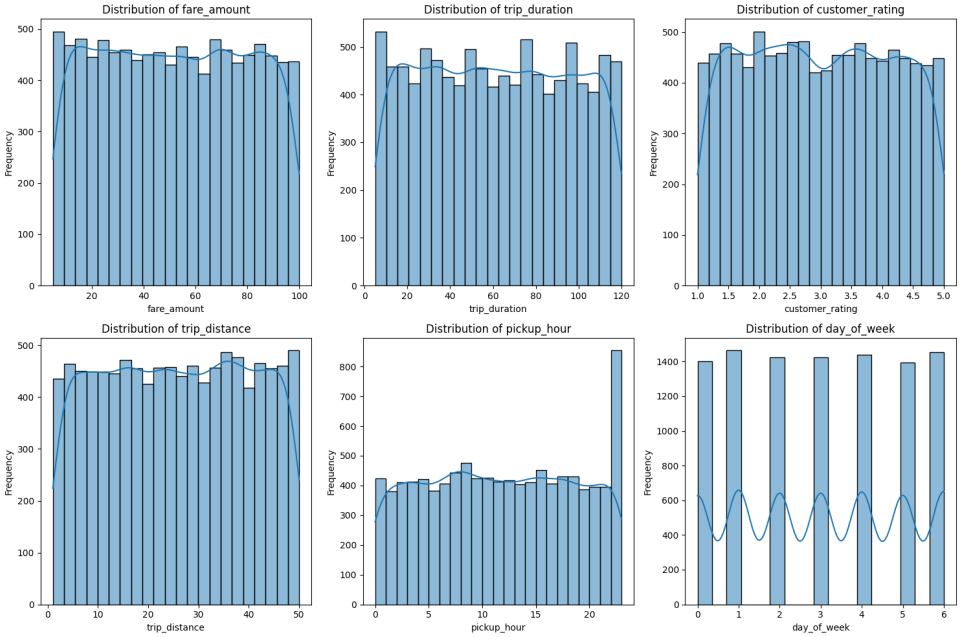

Figure 5.1: Distribution of features.

Future Work:

- Implement advanced forecasting models (ARIMA, Prophet, LSTM).
- Conduct correlation studies between sales and external variables (holidays, promotions, economic indicators).
- Develop an interactive dashboard for real-time retail analytics.
- Ingest real time data and get real time analysis

Appendix A

References

- Hyndman, R.J. and Athanasopoulos, G. (2018), *Forecasting: Principles and Practice*.
- Shumway, R.H. and Stoffer, D.S. (2017), *Time Series Analysis and Its Applications: With R Examples*.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M. (2015), *Time Series Analysis: Forecasting and Control*.
- Bishop, C.M. (2006), *Pattern Recognition and Machine Learning*.
- Chollet, F. (2018), *Deep Learning with Python*.
- Python official documentation: `https://docs.python.org/`
- Pandas documentation: `https://pandas.pydata.org/`
- Matplotlib documentation: `https://matplotlib.org/`
- Seaborn documentation: `https://seaborn.pydata.org/`

Appendix B

GitHub Link for Codes

https://github.com/amg1718/ideas_project

Appendix C

Other Documents

A copy of this report, the dataset and the presentation is stored in the GitHub/Google Drive link.