

Visualizing Time Series Dataset

Retail Sales Data

Ritam Deb Barman

Four Year B.Sc (Physics Major), 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 is about analysing and visualizing sales data from a pharmacy dataset. The dataset includes details such as dates, medicine names, and sales amounts. At first, the raw data looks like a table filled with numbers, names, and dates, but when we apply data analysis techniques, hidden patterns start to appear. By cleaning and processing the data, I was able to see how monthly revenue changes over time and how the total revenue grows steadily when viewed cumulatively. Using Python libraries like Pandas and Matplotlib, I created different charts such as line graphs, bar charts, and donut charts to make the findings clear and easy to understand. The results showed that a few medicines contribute to most of the revenue while others contribute less. Seasonal changes also affect sales, as some months had higher revenue than others. The cumulative revenue graph confirmed that overall growth is positive. This kind of analysis can help a pharmacy to identify its best-selling products, manage inventory better, and make smarter business decisions. For me as a student, the project was also a good practice in Python programming, data handling, and visualization. It gave me confidence in working with real-world datasets and showed me how data can tell a meaningful story.

2. Introduction

Every time we walk into a pharmacy or even shop online, we leave behind a trail of data — what we bought, when we bought it, how much we paid, and sometimes even how often we return. At first, this information looks like nothing more than a set of numbers in a spreadsheet: dates, medicine names, and amounts. But when we start to analyse and visualize this data, hidden patterns begin to appear.

We can see which medicines are sold the most during certain months, how customer buying behaviour changes over time, and which products bring in the most revenue. Suddenly, pharmacy sales data isn't just a list of transactions — it becomes a story about customers, business growth, and the choices people make every day.

That's the true value of visualization. It doesn't just create colourful graphs, it helps to reveal trends, highlight problems, and show opportunities that might otherwise go unnoticed. Managers can use these insights to plan stock and marketing, and students like me can use them to practice data analysis and understand business patterns better.

In this project, I worked with a pharmacy dataset and used Python to explore and visualize it. Some of the tasks included cleaning the dataset, grouping sales by month, and calculating revenue growth. I also created visualizations to show monthly trends, cumulative growth, and top-selling medicines. By the end, this project helped me understand how simple sales records can be transformed into meaningful insights through analysis and visualization.

3. Project Objective

- To clean and prepare the pharmacy dataset so that it can be used for proper analysis.
- To study monthly revenue trends and understand how sales change over time.
- To calculate and visualize cumulative revenue growth of the pharmacy.
- To identify the top-selling medicines and their contribution to total revenue.
- To present the findings through different types of charts for easy understanding.

4. Methodology

- **Data Collection**

- The dataset (pharmacy_dataset.csv) was prepared to represent real-world pharmacy transactions.
- It contains information such as Patient ID, Date of purchase, Medicine Name, Quantity purchased, Price per unit, Prescribing Doctor, and Total Amount.
- Data was stored in CSV format for ease of processing.

- **Data Pre-processing**

- The dataset was imported into Python using the pandas library.
- The **Date column** was converted to datetime format (%d-%m-%Y) to enable time-series grouping and analysis.
- Checked for missing or duplicate values and ensured numerical fields (Quantity, Price_per_unit, and Amount) were properly formatted.
- Cleaned and structured data was used for further analysis.

- **Data Analysis**

- Revenue was aggregated **monthly** by grouping sales data using the month-year of the Date column.
- Product-wise revenue was calculated by grouping data by **Medicine_Name**.
- Cumulative revenue was computed using the cumulative sum (cumsum) of monthly revenue to track long-term growth.

- **Data Visualization**

- **Line charts** and **bar charts** were used to show monthly revenue trends.
- **Bar chart (medicine-wise)** identified the top-selling medicines by revenue.
- **Donut chart** represented the contribution of each medicine category to overall sales.

- **Cumulative line chart** demonstrated overall revenue growth over time.
- **Tools and Techniques Used**
 - **Python** programming language was used for analysis.
 - **Pandas** for data manipulation and grouping.
 - **Matplotlib** for generating charts and visualizations.
 - **Google Drive/Local system** for dataset storage.
- **Optional Analytical Extension**
 - The methodology can be extended to include predictive models such as regression for sales forecasting.

To complete this project, I followed a step-by-step process.

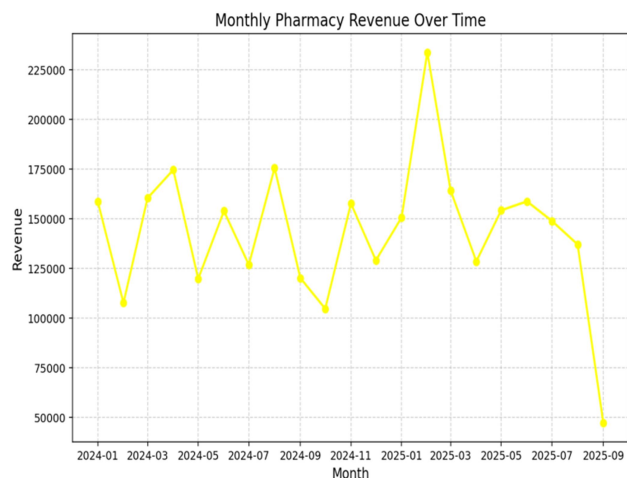
1. **Data Loading and Cleaning**
 - I loaded the dataset into Python using the Pandas library.
 - The **Date** column was converted into the proper date-time format so that grouping by months became easier.
 - I also checked for missing or duplicate values.
2. **Data Analysis**
 - I grouped the sales data by month and calculated the total revenue.
 - Then, I created a cumulative sum of the monthly revenue to study overall growth.
 - Next, I grouped the data by medicine names to identify which medicines generated the most revenue.
3. **Data Visualization**
 - A line chart was made to show the monthly revenue trend.
 - A bar chart was used to compare monthly revenues.
 - A cumulative revenue line chart showed overall growth across time.
 - Another bar chart displayed the top-selling medicines.
 - Finally, a donut chart was created to show the revenue contribution of each medicine category.

By following these steps, I was able to turn raw sales records into useful insights.

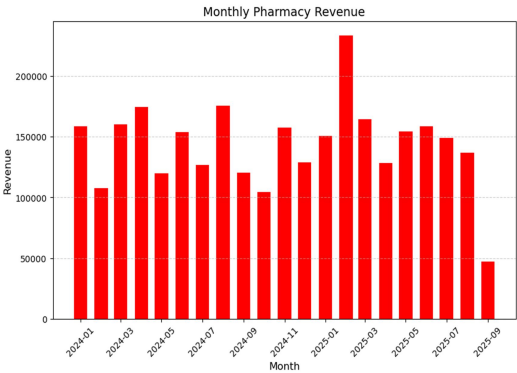
Github Link : https://github.com/RitamDebBarman/Visualizing_Time_Series_Dataset-

5. Data Analysis and Results

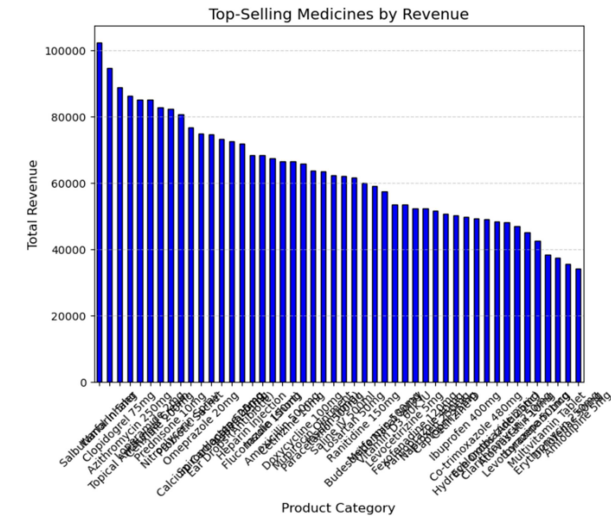
After performing the analysis, I observed several important findings:



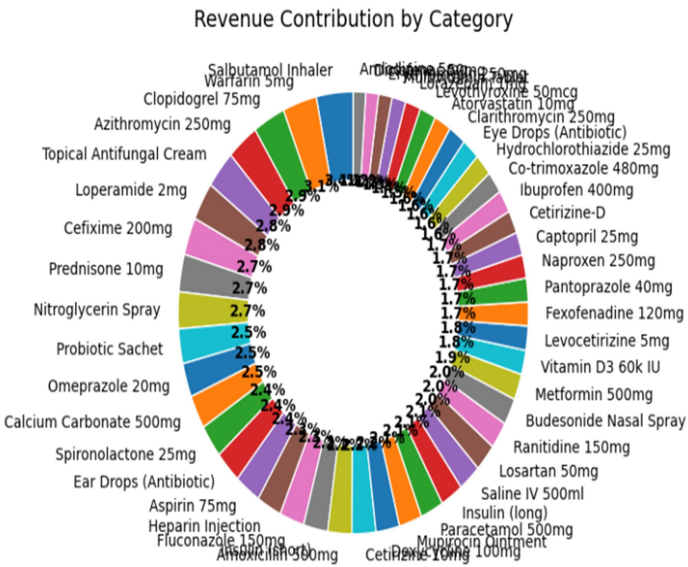
- Monthly Revenue:** The line and bar charts showed that revenue was not the same every month. Some months had higher sales, which could be because of seasonal demand or customer needs.



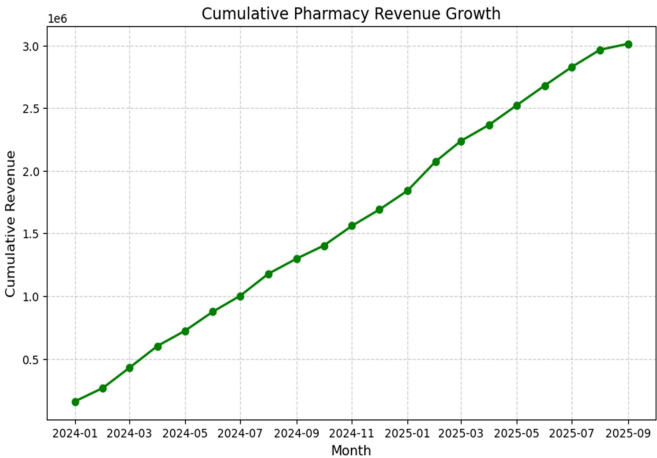
- Top-Selling Medicines:** The bar chart revealed that only a few medicines generated most of the revenue. These medicines are crucial for the business.



- Revenue Contribution:** The donut chart showed that the revenue distribution is not equal. A small number of medicines contribute the majority share, while many others contribute less.



These results prove that visualization makes it much easier to understand business performance and identify the most important factors.



- Cumulative Revenue:** The cumulative revenue graph went upward steadily, showing that overall sales are growing continuously.

6. Conclusion

From this project, I learned that sales data is much more than numbers in a file. With proper analysis, it can provide useful insights about customer behaviour, business growth, and product performance. The pharmacy dataset showed that only a few medicines contribute most of the revenue, while the rest play a smaller role. The cumulative revenue graph showed consistent growth, which means the business is moving in the right direction. Seasonal effects were also visible in monthly sales.

This type of analysis can help a pharmacy to focus on its best-selling products, plan stock in advance, and take better business decisions. On a personal level, the project gave me hands-on practice with Python, Pandas, and Matplotlib. It also showed me how data can tell a story when visualized properly. Overall, it was a very good learning experience.

Appendices

1. Scientific Computing in Python 3rd edition (Abhijit Kar Gupta)
2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems
3. Introduction To Machine Learning With Python: A Guide For Data Scientists
4. Chat-gpt