Visualising Time-series Dataset

Rupsa Saha

4 Year BSc 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

In this report we looked at global temperature data which is a record of how our climate has been transforming over time. We used line graphs which we tracked how temperatures went year by year and applied rolling averages to put the tall waves and dips into perspective which in turn brought the bigger picture into focus. Also we created box plots which display how temperatures play out from month to month over the last 50 years in which we were able to identify what different seasons do to the data. All of which together makes it easy to see the short-term changes as well as the large scale warming trend in our world's climate.

2. Introduction

By itself, data is quite overwhelming, if not complicated, especially for long durations of time, and with thousands of entries. As already mentioned, visualization is a data pattern recognition technique and is very helpful in drawing out relevant and useful interpretations from its raw data. Python has excellent tools for processing and visualization of data, in particular, Pandas, Matplotlib, and Seaborn libraries.

This report is based on data on global temperatures, which states the changes in our climate over time. The larger picture became clearer when we employed rolling averages to put the tall waves and dips into perspective and line graphs to show how temperatures changed year over year. In order to determine how different seasons affect the data, we also made box plots that show how temperatures change month over month over the previous 50 years. When taken as a whole, these factors make it simple to observe both the short-term shifts and the global warming trend.

During the first two weeks of classes, the following topics were covered :

• Basics Of Python

• OOPs

• Machine Learning (Supervised and unsupervised learning)

• Neural Networks

• Pytorch

• Large Language Models

• Ollama

3. Project Objective

- To convert raw temperature records into a usable format by analysing dates and extracting important features such as Year and Month.
- Use line plots with rolling averages (60-month window) to smooth fluctuations and highlight long-term changes in global temperatures.
- To visually identify patterns and anomalies in recent climate behaviour by transforming raw temperature data into an interpretable heatmap, making long-term and seasonal changes easier to observe.
- To compare raw and smoothed datasets in order to highlight underlying climate patterns and provide a clearer view of gradual global warming over time.
- Provide meaningful insights into global temperature rise and seasonal variations, supporting awareness of climate change patterns.

4. Methodology

The topics were visualising time-series dataset and generate a synthetic dataset and perform data visualising from that dataset. Before proceeding we have to import the following modules first:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot s plt
```

- seaborn(sns): A statistical plotting library for making plots more

attractive and easier to build.

- Matplotlib(plt): main plotting library in python.

The goal of this visualization is to analyze long-term temperature trends using real-world climate data.

Process:

- Data collection: the two data that we used were synthetic data containing age and corresponding values and monthly dataset containing time series data.
- Data preparing: Importing dataset using pandas in Python.

Checked structure and missing values using .head() and .info().

Sorted the datasets (by date for monthly data, by age for synthetic data).

- Data visualization: Used matplotlib and seaborn libraries for visualization.

Plotted rolling average graph, heatmap, boxplots for both datasets. For the synthetic dataset we used Age vs Corresponding values and for monthly dataset we used Date vs values.

During the process of rolling average for the first dataset We filter the dataset so that we only focus on one source for consistency.

Apply smoothing using a rolling average instead of looking at raw monthly values, we calculated a 60 month rolling average which smooths out seasonal fluctuations.

Visualize both raw and smoothed data and Plot the original monthly temperature data (to see actual variations). Plot the rolling average line (to see the long-term trend).Use different colours and styles so the difference is clear. Interpret the visualization.

The jagged line shows short-term variability. The smooth red line shows the long-term warming or cooling trend.

Here's a rough explanation the code :-

- For rolling average graph

import matplotlib.pyplot as plt

Imports Matplotlib, (for plotting a graph)

Creates a new column Rolling_60M.

Uses a rolling window of 60 months (5 years) to compute the moving average of the Mean temperature.

min_periods=1 ensures the calculation starts immediately, even if fewer than 60 months of data are available at the beginning.

plt.figure(figsize=(15,7))

Sets up the plotting canvas with a size of 15x7 inches (wider for long time series).

Plots the original monthly mean temperatures. Plots the smoothed rolling average line in red. This shows the long-term climate trend.

plt.title(f"Long-Term Temperature ({source_name})")

plt.xlabel("Year")

plt.ylabel("Mean Temperature")

Adds a title with the data source name.

Labels x-axis as Year and y-axis as Mean Temperature.

plt.legend()

plt.grid(True, linestyle="--", alpha=0.6)

plt.legend() → shows labels for the two lines (Original vs Rolling Average).

plt.grid() → adds dashed, semi-transparent gridlines for better readability.

plt.show()

Displays the final plot.

- For the heatmap

Prepare the data – Converted the Date column into year and month, then filtered records from the last 20 years.

Aggregate monthly means – Grouped data by Year and Month and calculated the average temperature for each. Reshape for heatmap – Pivoted the table so rows = Years and columns = Months, making it heatmap-ready. Visualize – Plotted a heatmap with colour intensity showing average temperatures, allowing seasonal and yearly patterns to be compared easily.

We also did a boxplot representation to quickly spot whether one group has consistently higher or lower values than another.

As for the second dataset we performed the same thig all over again.

Data source:

- GCAG: Global Climate
- www.syngendata.ai

All the codes uploaded in Github, it can be accessed from the link:

[https://github.com/Rupmin/Data-visualization]

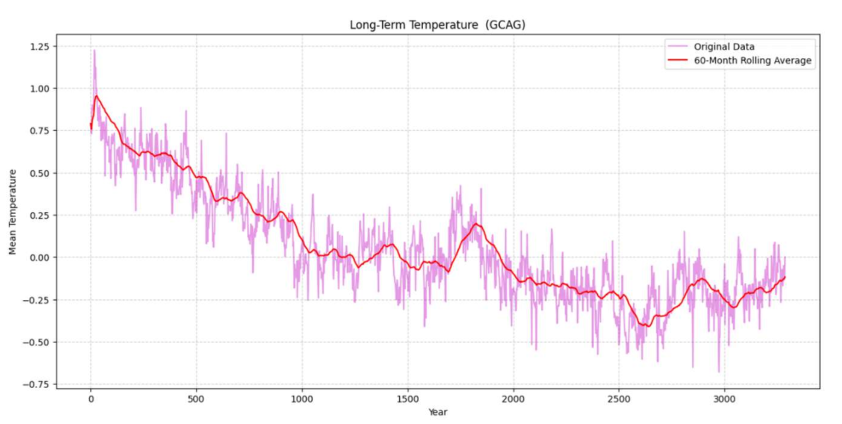
5.Data Analysis and Result

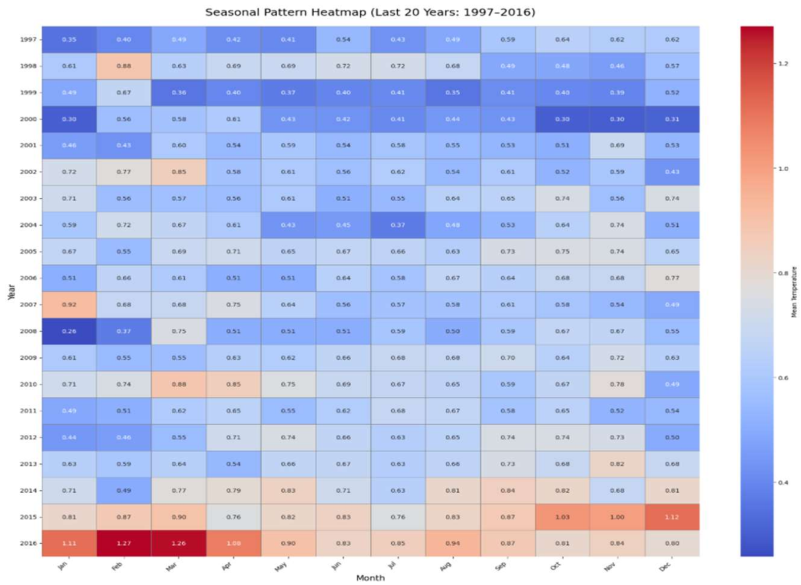
For the first dataset,

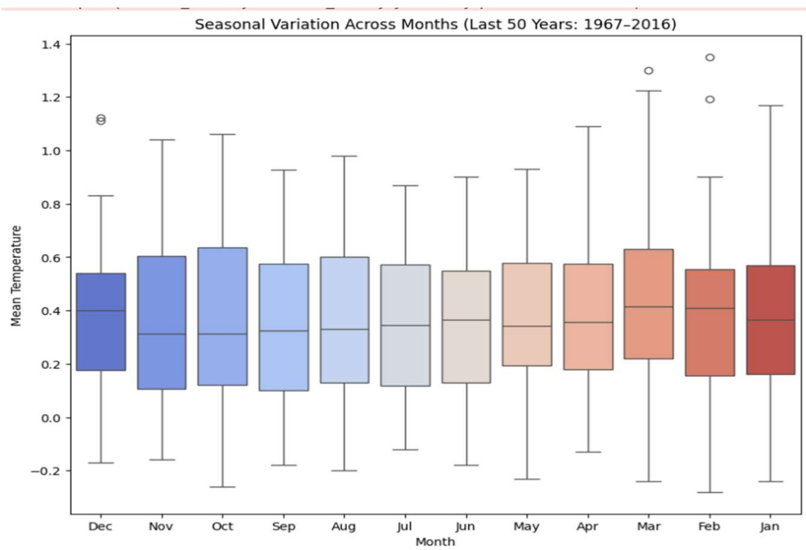
Saving monthly_csv.csv to monthly_csv.csv

	Source	Date	Mean
0	GCAG	2016-12-06	0.7895
1	GISTEMP	2016-12-06	0.8100
2	GCAG	2016-11-06	0.7504
3	GISTEMP	2016-11-06	0.9300
4	GCAG	2016-10-06	0.7292

- <u>Graphs</u>







For the second dataset,

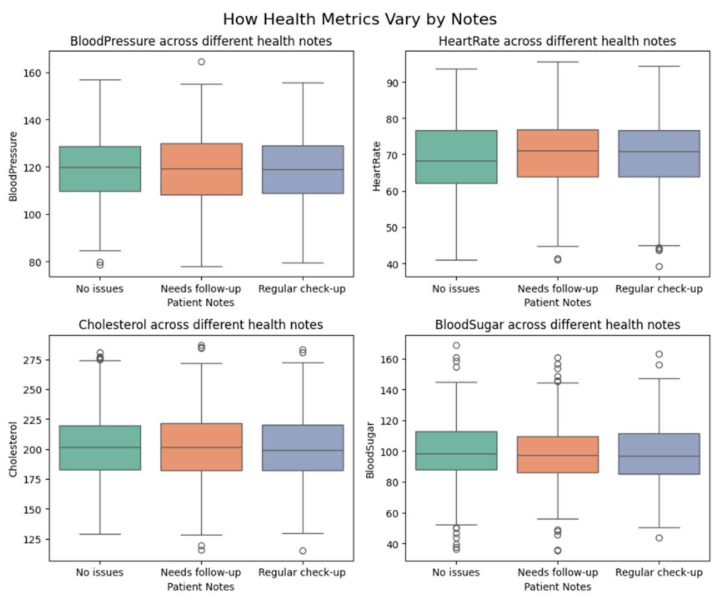
Saving Synthetic_Data_2025-09-18.csv to Synthetic_Data_2025-09-18 (2).csv

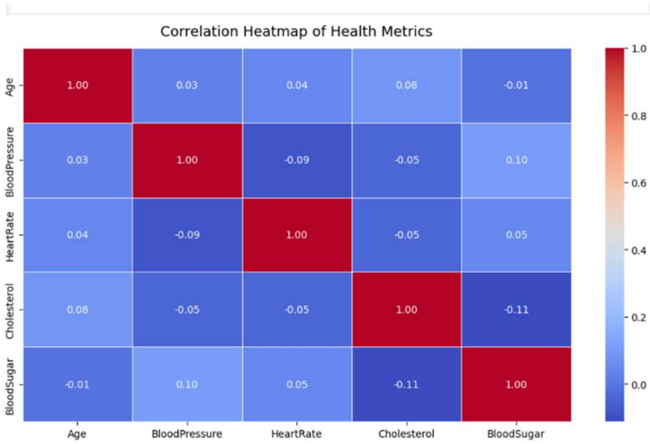
	Age	BloodPressure	HeartRate	Cholesterol	BloodSugar
0	48.9672	96.6822	75.3878	209.3334	132.1444
1	80.0036	141.6217	57.0018	169.5862	108.6848

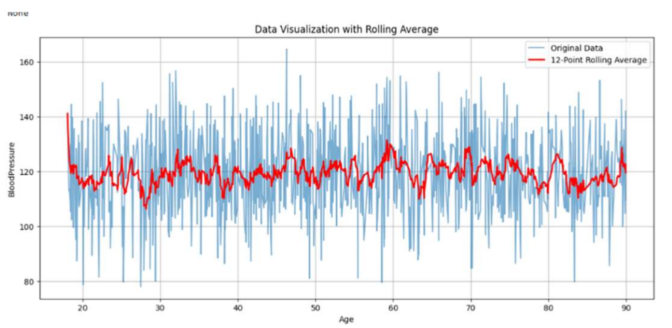
2	62.1090	90.9614	56.9324	174.1043	109.6714
3	27.0628	132.0603	85.2470	196.6521	71.7100
4	68.0029	122.7896	58.3116	236.8029	77.6854

	Notes
0	No issues
1	Needs follow-up
2	Regular check-up
3	Needs follow-up
4	Regular check-up

- Graphs







6.Conclusion

<u>Inference between the two dataset:</u>

Nature of the data:

Synthetic data → individual type data (age-based, discrete).

Monthly data → time centric data (continuous over years, seasonal)

How rolling average woks:

In synthetic data, rolling averages is mainly smooth and we see random variability across ages

In monthly data, rolling averages expose many kinds of seasonal data.

Usefulness:

Synthetic data visualization is useful for studying any kind of relationship of ages and measured values.

Monthly climate data visualization helps in climate trends.

The two datasets comparison shows how rolling averages improve data interpretation. Rolling averages smooth out random fluctuations across age groups in the synthetic dataset, making age-related variation patterns more obvious. Rolling averages remove short-term seasonal noise from the monthly dataset, making long-term climate or trend shifts more apparent. Therefore, even though the two datasets are different in that one is demographic and the other is temporal, rolling averages are equally useful in revealing significant insights that may be hidden by raw data alone.

7.Appendics

1.References:

- https://youtube.com/playlist?list=PLfP3JxW-T70EwL12zs3a3ZJAe8DLo6qlA&si=v-W8ejbb22ufe3Hk/
- https://youtu.be/FYji-5S8D2s?si=yN1HZrY2N_zCF1Iz/
- https://books.google.co.in/books?id=2vOoCwAAQBAJ&newbks=1&newbks_redir=0&lpg=PP1&pg=PR1#v=onepage&q&f=false/

2.Github link for the code:

https://github.com/Rupmin/Data-visualization

3. Google collab link:

https://colab.research.google.com/drive/1ROE_vvfxXIHKSsXvPUKi71mE_NQJegwa?usp=sharing