Visualizing Time Series Dataset - 07

Pratik Shyam Ijantkar,
RV College of Engineering, 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

1. Abstract

The internship project “Visualizing Global Temperature Trends” focuses on analyzing and visualizing global temperature data to identify trends like warming, cooling, etc. using Python. Using the data file provided, the project first imports and installs the necessary Python packages, plots the required graph, smoothens the data to visualize the peaks and falls in temperature trends, plots heatmaps and then draws the conclusion that there is an upward rise in global temperature trends. A second dataset titled “Microsoft Stock – Time Series Analysis” is used to plot the raw graph, smoothened graph, color scatter plot and come to the conclusion that there is a clear upward rise in stock prices from 2015 to early 2021. The analysis is conducted using Python in the Google Colab environment, utilizing libraries NumPy, Pandas, Matplotlib, Plotly and Seaborn.

The "Visualizing Global Temperature Trends" internship project uses Python to analyze and visualize global temperature data, identifying trends such as warming or cooling. Using a provided dataset, the project imports necessary Python packages, plots a graph, smooths the data to highlight temperature peaks and falls, creates heatmaps, and concludes an upward trend in global temperatures.

A second dataset, "Microsoft Stock – Time Series Analysis," is used to plot raw and smoothed graphs, and a color scatter plot, leading to the conclusion of a clear upward rise in stock prices from 2015 to early 2021. The analysis is performed in Google Colab using NumPy, Pandas, Matplotlib, Plotly, and Seaborn.

2. Introduction

This project, “Visualizing Global Temperature Trends” addresses the critical issue of climate change by providing a comprehensive analysis and visualization of historical global temperature data from 1880 to 2016. Increasing global temperature has many devastating consequences like rise in sea levels, rapid melting of glaciers, habitat loss, breakdown of the ecosystem, change in migratory patterns, vulnerability of a larger population to heatstroke, etc. to name a few. These changes highlight the need for a simple plot to visually analyze the trends in global temperature. This project does the same by converting the large dataset to simplified time series 12 month moving average and 60 month moving average plots to better read and analyze the global temperature trends.

The project uses Python as its language of choice, in Google Colab, a cloud based online integrated development environment (IDE) for Python that requires no setup. Python is known for its extensive collection of libraries, which are well utilized in the project. The libraries used are as follows:

The "Visualizing Global Temperature Trends" project addresses climate change by analyzing and visualizing historical global temperature data from 1880 to 2016. Rising global temperatures have severe consequences, such as increased sea levels, melting glaciers, habitat loss, ecosystem disruption, altered migratory patterns, and a

higher risk of heatstroke for populations. To address these concerns, this project aims to provide a simple visual representation of global temperature trends. It achieves this by converting a large dataset into simplified 12-month and 60-month moving average time series plots, making the trends easier to read and analyze. The project is developed in Python, utilizing Google Colab as a cloud-based IDE, and leverages Python's extensive library collection for its implementation.

- Pandas – To analyze, clean, explore and manipulate the dataset provided.
- NumPy – To convert the large data stored in Python list into array for faster computation.
- Matplotlib – To integrate with Seaborn to plot graphs. It is a low-level graph plotting library on its own.
- Plotly – To create the graphs, line charts, heatmaps and scatter plot.
- Seaborn – To work with Matplotlib to visualize the random distribution plot graphs.

A standard data science procedure was followed to complete this project – data collection from a reliable source, raw data plotting, cleaning (or smoothening) the data to handle missing values and inconsistencies, exploratory data analysis using different types of graphs and charts, and finally, drawing a credible conclusion from the different plots present. The primary purpose of this project was to map the change in global temperature over the last 200 years and to visualize the upward rise of Microsoft Stock prices from 2015 – 2021.

This project required extensive reading and understanding of time-series datasets [3][4], moving averages – its types and functions, especially in meteorology and finance [1], global temperature trends, stock price fluctuations, basic data science processes – data collection, data cleaning, data visualization, and knowledge of Python – the language and its libraries [4][5][6][7][8][9]. The first two weeks of the internship were crucial for understanding and working with the Python language in an IDE (Google Colab).

The topics learnt during training are as follows:This project utilized a standard data science methodology, including data collection from reliable sources, initial plotting, cleaning for missing values and inconsistencies, exploratory data analysis with various visualizations, and finally, drawing conclusions from the plots. The main goals were to map global temperature changes over the past 200 years and visualize Microsoft stock price increases from 2015-2021.

This involved extensive research and understanding of time-series datasets, different types and functions of moving averages (especially in meteorology and finance), global temperature trends, stock price fluctuations, basic data science processes (collection, cleaning, visualization), and proficiency in Python and its libraries. The first two weeks of the internship were crucial for learning and working with Python in Google Colab, covering various topics.

Week 1

Sl. No.	Date	Day	Topic
1.	25.08.2025	Monday	Introduction - Welcome Note - What to expect from this internship
2.	26.08.2025	Tuesday	Python Basics - 1 (Data, Variable, Lists, Loop)
3.	27.08.2025	Wednesday	Python Basics - 2 (Data Structures)
4.	28.08.2025	Thursday	Python Basics - 3 (Class, Functions, OOPS)
5.	30.08.2025	Saturday	Python Basics - 4 (NumPy, Pandas)

Week 2

Sl. No.	Date	Day	Topic
1.	01.09.2025	Monday	Machine Learning Overview
2.	02.09.2025	Tuesday	Regression Lab
3.	03.09.2025	Wednesday	Classification Lab
4.	04.09.2025	Thursday	LLM Fundamentals
5.	06.09.2025	Saturday	Communication Skills

3. Project Objective

The project contains two datasets – “Visualizing Global Temperature Trends” and “Microsoft Stock Prices – Time Series Analysis”. The main objective of this is:

- **Clean:** Process large datasets into usable, clean data.
- **Graph:** Visualize datasets using graphs, plotting both raw and smoothed data for easier comparison and analysis.
- **Color Map:** Display datasets using color-coded maps, presenting them in a format other than traditional graphs to better illustrate trends.
- **Conclude:** Analyze trends within the datasets to draw conclusions.

It's important to note that this project solely uses pre-existing datasets, either provided by the institute or downloaded from public sources, eliminating the need for surveys. The focus is strictly on the in-depth analysis and visualization of existing scientific data.

4. Methodology

The project's methodology followed a structured data science pipeline, encompassing data collection, cleaning, preprocessing, smoothing, exploratory analysis, and visualization, leading to the final conclusions. This project relied solely on existing data, thus eliminating the need for on-site data collection, surveys, questionnaires, or other surveying methodologies.

For Dataset 1, visualizing global temperature trends:
a. Data was provided by the institute at the project's outset.

b. The dataset was plotted without alterations.
c. Data was smoothed using both 12-month and 60-month moving averages to address inconsistencies and clarify trends.
d. Graphs were plotted for the smoothed data.
e. Two heatmaps, one for the last 50 years and another for the last 20 years, were generated. The first heatmap was subsequently visualized as a line chart.
f. The graphs and heatmaps indicated a global temperature increase over the past 200 years.

For Dataset 2, Microsoft Stock – Time Series Analysis:
a. Data was obtained from Kaggle.
b. The dataset was plotted as is.
c. The dataset was smoothed using a 30-day moving average.
d. A graph of the smoothed data was plotted.
e. A scatter plot illustrating the closing trend of stock prices was created, then re-rendered as a line chart.
f. Despite market fluctuations, the graphs and scatter plot suggested a general rise in Microsoft Stock Prices between 2015 and 2021.

5. Data Analysis and Results

This section will present all the graphs, plots and maps used in the project. All analyses were done in Google Colab using Python and its libraries.

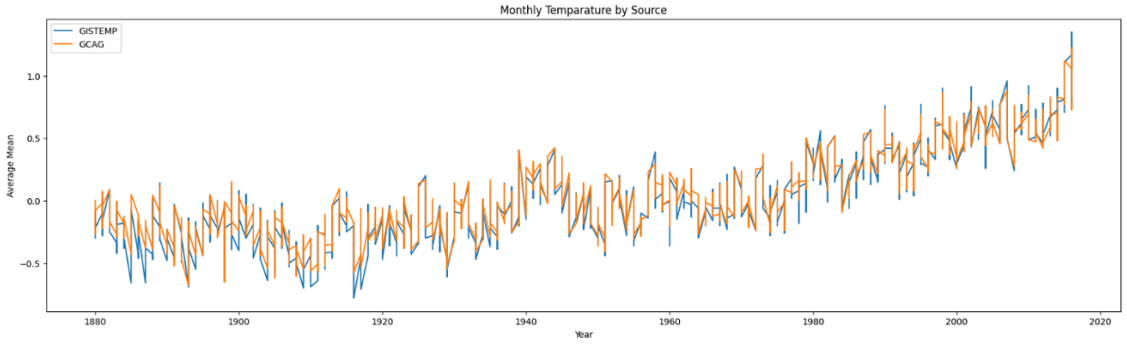

Figure 1. Raw graph of global temperature dataset

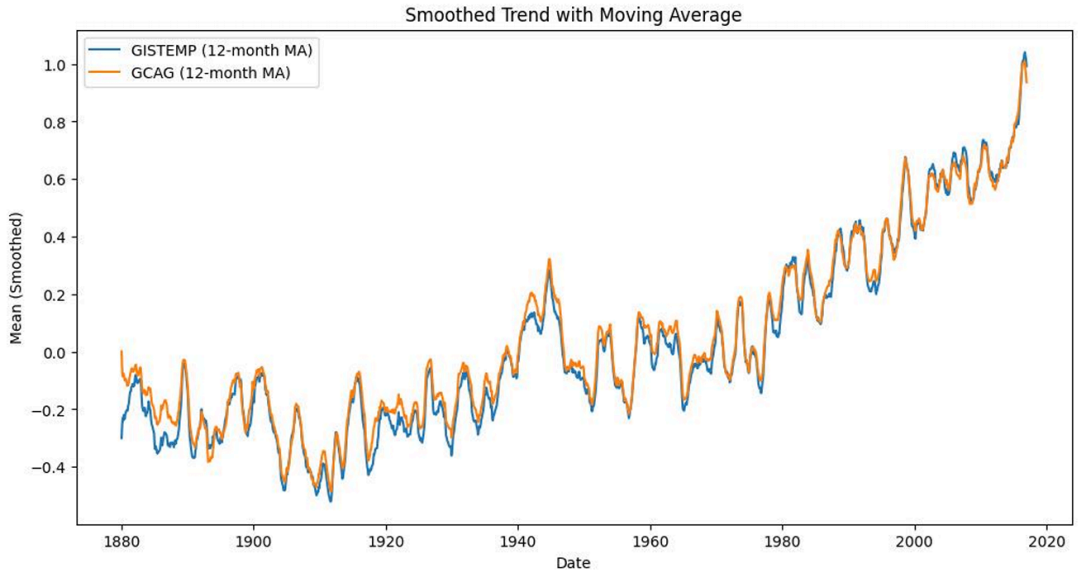

Figure 2. 30-day moving average graph

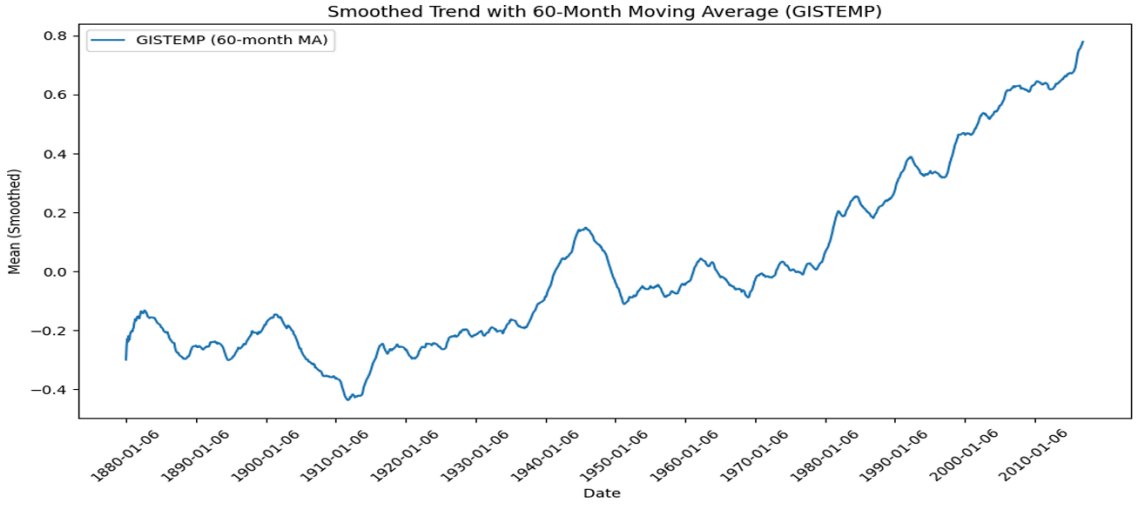

Figure 3. 60-day moving average graph

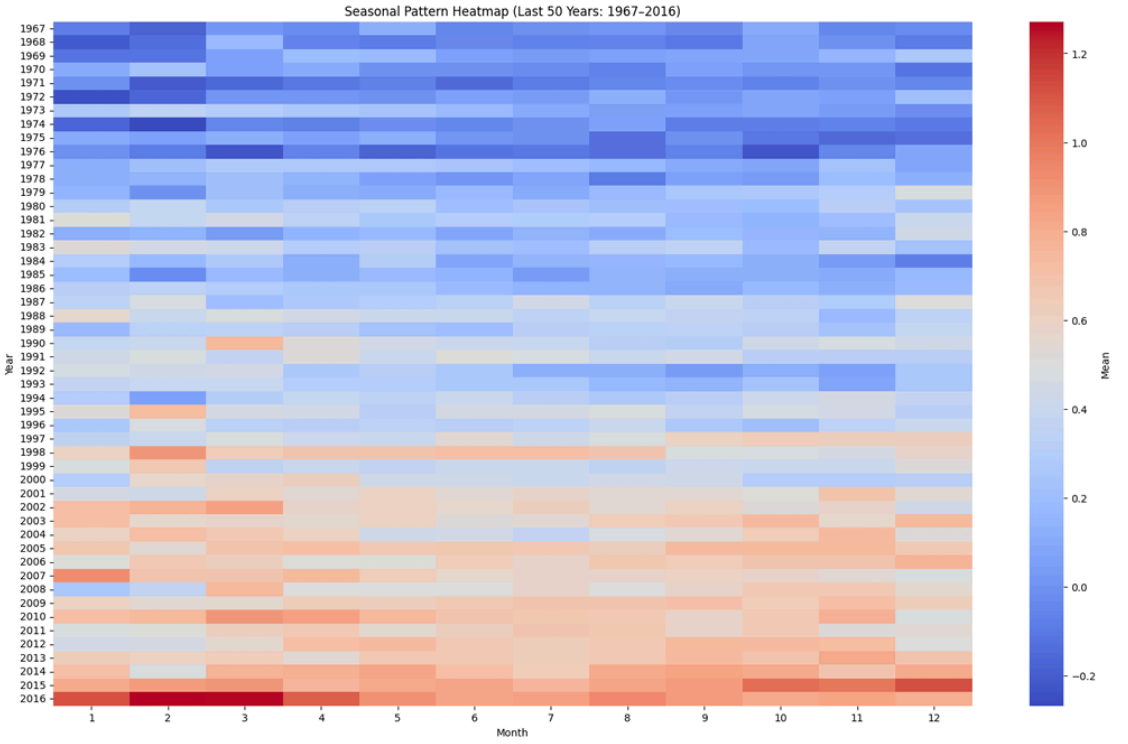

Figure 4. 1967-2016 Heatmap

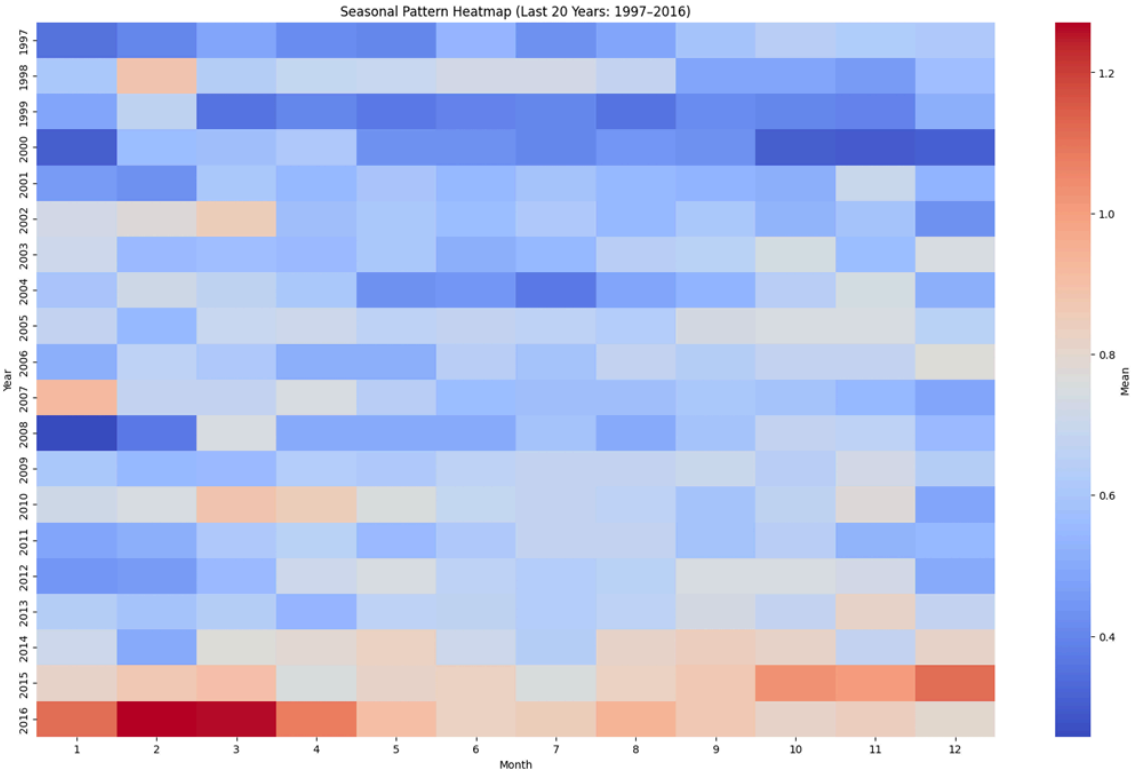

Figure 5. 1997-2016 Heatmap

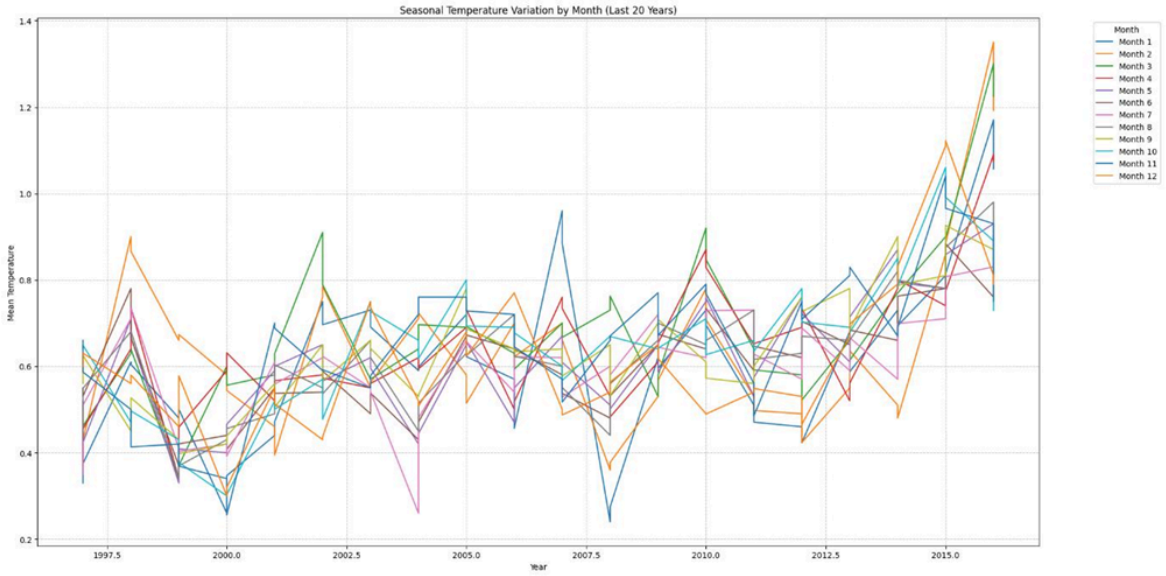

Figure 6. Line Plot of 20-year heatmap

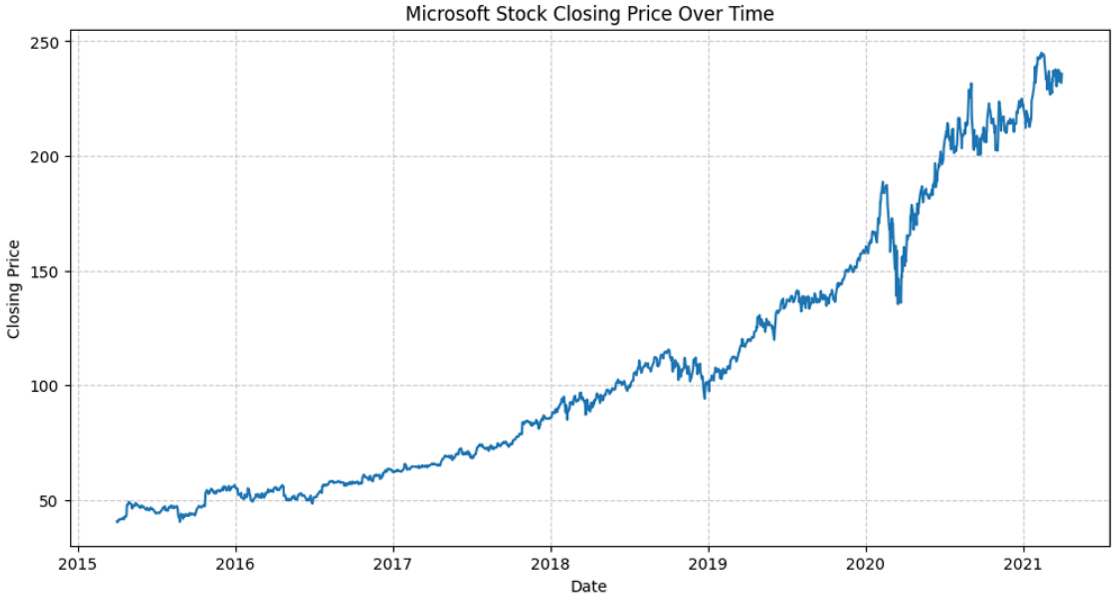

Figure 7. Raw graph of closing stock price dataset

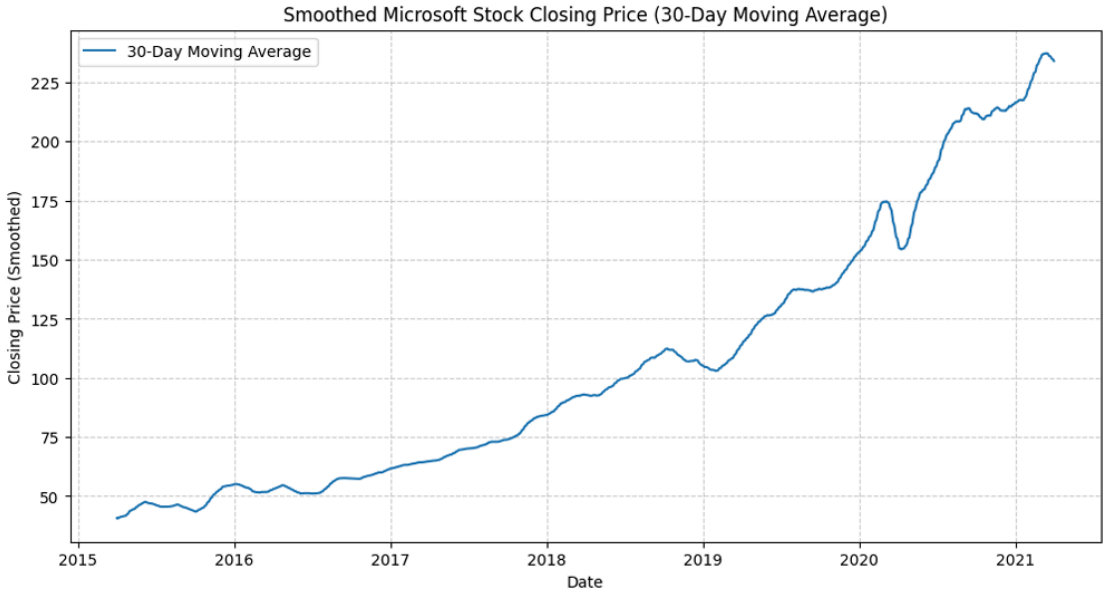

Figure 8. 30-day moving average graph

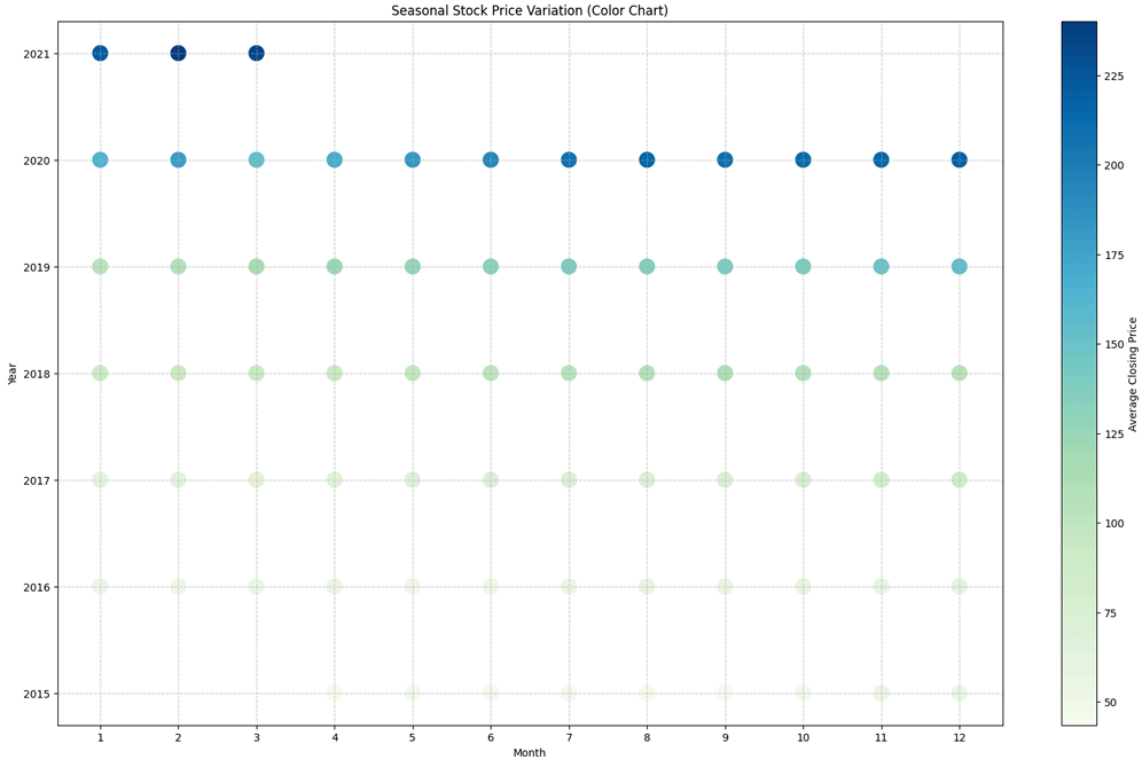

Figure 9. Closing Stock Price scatter plot

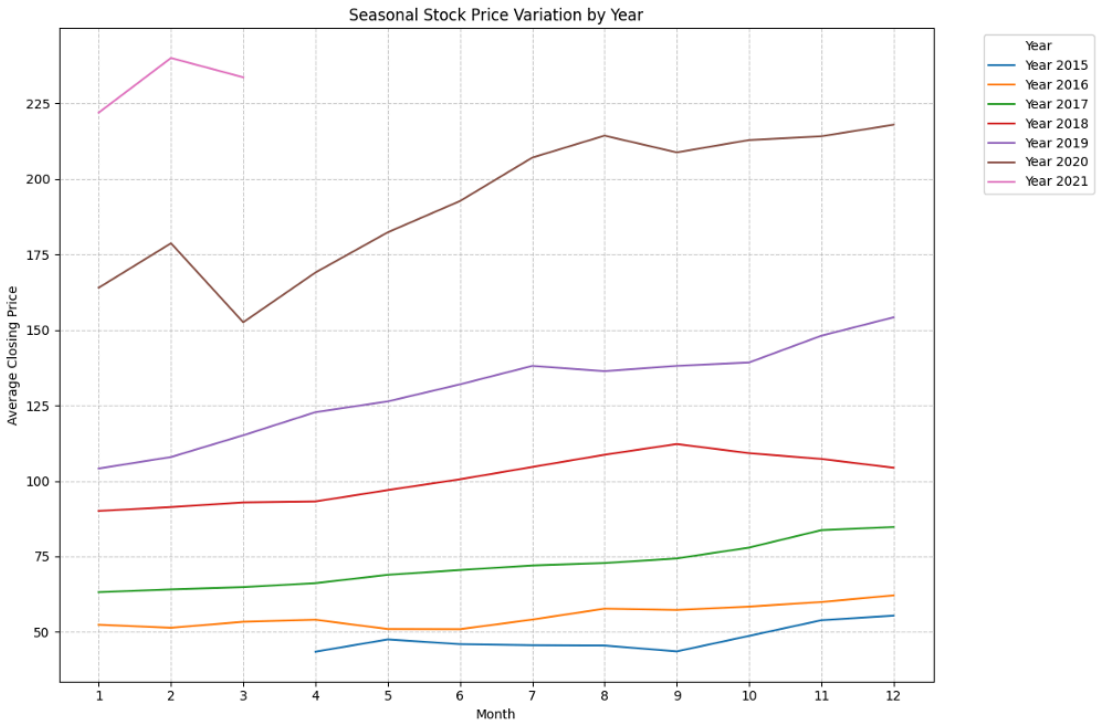

Figure 10. Line chart of scatter plot

6. Conclusion

The successful completion of this project marks a significant achievement in the application of Python for comprehensive data analysis and intuitive visualization. The primary objective, focused on understanding trends within historical global data and Microsoft Stock prices, was thoroughly met. Through rigorous analysis, the project unequivocally confirmed existing scientific and analytical consensus regarding the upward trajectory of global temperatures and the consistent increase in Microsoft's stock closing prices.

The methodology employed leveraged Python's robust libraries for data manipulation, statistical analysis, and advanced visualization. This allowed for the extraction of meaningful insights from large datasets, transforming raw information into readily understandable visual representations. For instance, the analysis of global temperature data involved scrutinizing historical records, identifying anomalies, and presenting long-term warming trends through clear and concise graphical formats. Similarly, the examination of Microsoft Stock prices involved analyzing historical closing prices, identifying patterns, and visualizing periods of growth and stability.

The project's findings serve as a compelling affirmation of previously established facts, lending further credence to the work of scientists and financial analysts. The observed increase in global temperatures, a critical environmental concern, was consistently evident across various data subsets and visualization techniques,

reinforcing the urgency of addressing climate change. Concurrently, the sustained upward shift in Microsoft Stock prices underscores the company's strong market position and consistent performance over time.

In conclusion, this project not only achieved its stated objectives but also validated the power of data-driven insights in confirming established knowledge. The successful execution, coupled with the clear and compelling visualization of complex data, solidifies the project's success and demonstrates the efficacy of Python as a tool for impactful data analysis.

7. APPENDICES

A. References

[1] Moving Average: What It Is, Types, and How It's Used. Investopedia.
[2] Global Temperature Report for 2023. (2024, January 12). NASA GISS.
[3] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.
[4] McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.
[5] VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.
[6] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95.
[7] Perktold, J., & Ross, G. (2009). SciPy: A Reference Guide for Scientific Computing in Python. *Computing in Science & Engineering*, *11*(3), 20–25.
[8] Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference*, 57–62.
[9] Plotly Technologies Inc. (2015). *Plotly Graphing Library*.
[10] Kaggle. (n.d.). *Microsoft Stock Price Data*.

B. Github Link

https://github.com/iSavage05/IDEAS-TIH