# Time Series Analysis of Global Temperature Data

**Debkamal Sarkar**
**M.Sc. Statistics and Computing(2025-27)**
**DST-CIMS(Banaras Hindu University)**

**Period of Internship: 21ST January 2026 – 17TH February 2026**

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

# 1. Abstract:

This project examines global temperature patterns using a monthly global temperature dataset collected from various scientific sources, mainly GCAG and GISTEMP. The aim is to investigate the temporal patterns of temperature variability and gain insights into long-term climate system behavior through data preprocessing, manipulation, and statistical analysis using Python and Pandas.

The dataset was initially processed to remove any inconsistencies by translating date data into a valid datetime format, followed by the extraction of year and month components for the purpose of temporal analysis. Exploratory data analysis was performed to verify the completeness of the dataset, which had no missing values. A 12-month moving average was calculated to filter out short-term variations and emphasize long-term temperature trends. The dataset was then filtered to analyze the last twenty years of temperature patterns.

Statistical aggregation methods were employed to calculate average temperatures based on source, year, and month. Yearly aggregation showed long-term warming trends, while monthly aggregation showed temperature patterns and the average warmest months. A pivot table was created to facilitate the comparison of temperature variability between months and years, offering a systematic approach to temporal analysis.

This project illustrates the application of data manipulation, grouping, and time series analysis methods to derive valuable information from climate datasets. The findings offer a comprehensive approach to understanding global temperature variability and facilitate further climate trend analysis and visualization.

## 2. Introduction:

Climate change and global warming have emerged as some of the most pressing scientific and environmental issues of the contemporary world. The study of long-term temperature trends is critical to the study of climate trends, the prediction of environmental changes, and informing policy decisions on sustainability. This project aims to analyze global temperature data using statistical and data analysis methods to determine trends, seasonal patterns, and long-term changes in global average temperature.

The significance of this project is based on the increasing need for data-informed climate analysis. Vast amounts of environmental data are being continually gathered by international bodies such as NASA and NOAA. The appropriate analysis of such data helps climate scientists make sense of warming trends, anomalies, and climate variability over time. By analyzing real-world temperature data from sources such as GCAG and GISTEMP, this project illustrates the real-world application of data science in climate studies.

The technologies that are used in this project include Python programming and its data analysis framework. The main tools used for this project are Pandas for data analysis, NumPy for numerical computations, and Google Colab as the cloud-based development platform.

A background survey of the topic shows that time series analysis and moving averages are popular techniques used in climate analysis to remove short-term variations and analyze long-term trends. Climate data is often analyzed by aggregating temperature data on a yearly or monthly basis to analyze seasonal patterns and long-term warming trends. Pivot tables and grouped statistical analysis are common analytical techniques used in environmental data analysis.

The procedure adopted in this project involves data acquisition, preprocessing, feature extraction, statistical aggregation, and structured analysis. First, the data was imported and preprocessed by handling date fields by converting them to datetime format. New variables such as year and month were derived to enable temporal analysis. Moving averages were calculated to analyze long-term trends, followed by grouping calculations to derive yearly and monthly averages. Finally, pivot tables were generated to compare temperature changes over years and months.

The aim of this project is to enhance skills in data preprocessing and analysis, exploratory data analysis, and time series analysis, applying statistical concepts to a real-world climate data set.

## 3. Project Objective:

The primary aim of this project is to examine the global temperature data through statistical and data analysis methods in order to understand the long-term climate trends and seasonal changes. The project will show how real-world environmental data can be handled and interpreted with the help of contemporary data science tools.

Objectives of the Project:

- To preprocess and handle the global temperature data by transforming the raw date information into a proper time-series format for analysis.
- To examine the long-term temperature trends by calculating the yearly average temperatures and determining the patterns of global warming.
- To examine the seasonal temperature changes by comparing the average temperatures for different months of the year.
- To use statistical smoothing methods, like the 12-month moving average, to suppress the short-term variations and emphasize the underlying trends.
- To compare the temperature data measured from different scientific sources (GCAG and GISTEMP) by using grouped statistical summaries.
- To create pivot tables and aggregated datasets for better visualization and interpretation of temperature variations for different years and months.
- To gain hands-on experience in data handling, exploratory data analysis, and time-series analysis with the help of Python, Pandas, and NumPy.

# 4. Methodology:

This project employs a systematic data analysis methodology to analyze global temperature patterns based on a real-world climate dataset. The project mainly involves data collection, cleaning, transformation, statistical analysis, and interpretation using Python-based analytical software. The methodology employed in this project ensures systematic processing of time-series data and precise extraction of valuable insights.

### 4.1 Data Collection

The dataset employed in this project is a monthly global temperature dataset collected from publicly available scientific literature. The dataset contains temperature data collected from two prominent climate observation centers:

- GCAG (Global Climate at a Glance)
- GISTEMP (NASA Goddard Institute for Space Studies Temperature Analysis)

The dataset was retrieved from a common Google Drive link and imported into the Google Colab platform for analysis.

Data Features:

- Monthly global mean temperature observations
- Date-specific temperature data
- Multi-source data
- Continuous time-series data format
- No survey or questionnaire was administered because the project is based on observational climate data.
- 4.2 Tools and Technologies Used
- Python – Analysis programming language
- Google Colab – Cloud-based development platform
- Pandas – Data processing and analysis
- NumPy – Numerical calculations
- CSV Data Format – Data storage and processing

### 4.3 Data Import and Initialization

The dataset was directly downloaded from Google Drive using the gdown function and imported into a Pandas DataFrame.

The following steps were executed:

- Required libraries (pandas, numpy) were imported
- Dataset was downloaded from Google Drive
- CSV file was imported into DataFrame (df_temp)
- Initial data was displayed for validation

4.4 Data Cleaning and Pre-processing

Data preprocessing was done to ensure that the data was in the correct format for time series analysis.

Steps included:

- Datetime Conversion
- The 'Date' column was changed from string to datetime format using pd.to_datetime().
- Sorting
- The data was sorted in chronological order according to date values.
- Missing Value Check
- The completeness of the data was checked using null value checks.
- No missing values were present.
- Feature Extraction

The Year and Month features were extracted from the Date column to facilitate group analysis.

## 4.5 Data Transformation and Analysis

Various analysis tasks were carried out to analyze temperature trends.

1. <u>Moving Average Calculation</u>

A 12-month moving average was calculated for each data source.

Purpose: To smooth out short-term variations and highlight long-term trends.

Method used:

GroupBy + Rolling Window Analysis

2. <u>Filtering Recent Data</u>

1. A new data subset was created with the last 20 years of data.
2. Facilitated the analysis of contemporary climate trends.

3. <u>Aggregation Analysis</u>

1. Various aggregation methods were used:
2. Mean temperature by Source
3. Mean temperature by Year
4. Mean temperature by Month
5. These steps facilitated the identification of:
6. Long-term warming trends
7. Seasonal changes
8. Variations between data sources

4. <u>Pivot Table Creation</u>

A pivot table was created with the following structure:

1. Rows → Year
2. Columns → Month
3. Values → Average Mean Temperature

This facilitated the comparison of data across years and months.

**4.6 Analytical Approach**

The project adopts an Exploratory Data Analysis (EDA) and statistical analysis of time series data approach, as opposed to predictive modeling.

- Techniques used include:
- Descriptive statistics
- Grouping and aggregation
- Rolling averages
- Time series filtering
- Data reshaping using pivot tables

No machine learning or predictive model was built since the aim was to explore trends and interpret results statistically.

### 4.7 Flow Chart of Project Activities

Data Collection

↓

Data Import (Google Colab)

↓

Data Cleaning & Validation

↓

Datetime Conversion

↓

Feature Extraction (Year, Month)

↓

Sorting & Preprocessing

↓

Moving Average Calculation

↓

Grouping & Aggregation

↓

Pivot Table Creation

↓

Trend Analysis & Interpretation

### 4.8 Code Repository

All analyses were conducted using Python code developed in Google Colab.
(You may upload your notebook to GitHub and provide the repository link below.)

GitHub Link: [Add your GitHub repository link here]

### 4.9 Summary of Methodology

The proposed methodology integrates structured data preprocessing with statistical aggregation methods for effective climate data analysis. By leveraging time-series processing and grouped analysis, the study has successfully identified significant temperature trends and demonstrated the applicability of data analysis to climate research.

## 5. Data Analysis and Results:

This section presents the findings obtained from analyzing the global temperature dataset. The analysis mainly focuses on descriptive statistical evaluation and time-series exploration of temperature trends. Since the project does not involve survey data or predictive machine learning models, the results are based on exploratory and descriptive analysis.

### 5.1. **Descriptive Analysis**

| | 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 |

Descriptive analysis was conducted to understand the structure, completeness, and general behavior of the dataset before performing deeper analysis.

#### 5.1.1 Dataset Overview

| Description | Result |
|---|---|
| Dataset Type | Monthly Time-Series Dataset |
| Sources | GCAG, GISTEMP |
| Variables | Date, Source, Mean(Year, Month and Moving Average were later Calculated) |
| Missing Values | 0 |
| Time Coverage | Decades of Global Temperature records |

**Observation:**
The dataset contains complete observations with no missing values, making it suitable for direct statistical analysis.

#### 5.1.2 Data Preprocessing Results

The following transformations were successfully applied:

- Date column converted to datetime format
- Dataset sorted chronologically
- Year and Month extracted
- Dataset filtered for the last 20 years
- 12-month moving average computed

|   | Source | Date | Mean | Year | Month | Moving_Avg |
|---|---|---|---|---|---|---|
| 4 | GCAG | 2016-10-06 | 0.7292 | 2016 | 10 | 0.981917 |
| 3 | GISTEMP | 2016-11-06 | 0.9300 | 2016 | 11 | 1.017500 |
| 2 | GCAG | 2016-11-06 | 0.7504 | 2016 | 11 | 0.963992 |
| 1 | GISTEMP | 2016-12-06 | 0.8100 | 2016 | 12 | 0.992500 |
| 0 | GCAG | 2016-12-06 | 0.7895 | 2016 | 12 | 0.936292 |

These steps enabled structured time-series analysis.

### 5.2 Average Temperature by Source

| Source | Average Mean Temperature |
|---|---|
| GCAG | 0.048797 |
| GISTEMP | 0.024380 |

Both scientific sources show closely aligned average temperature trends, indicating consistency between independent climate monitoring systems. Although the numerical averages differ slightly due to differences in measurement methodology and baseline calculations, both datasets reflect similar global temperature behavior over time.

### 5.3 Yearly Temperature Trend

Temperatures were aggregated annually using group-by operations.

|   | Mean |
|---|---|
| Year | |
| 2012 | 0.629517 |
| 2013 | 0.661846 |
| 2014 | 0.742075 |
| 2015 | 0.882408 |
| 2016 | 0.964396 |

**Result:**

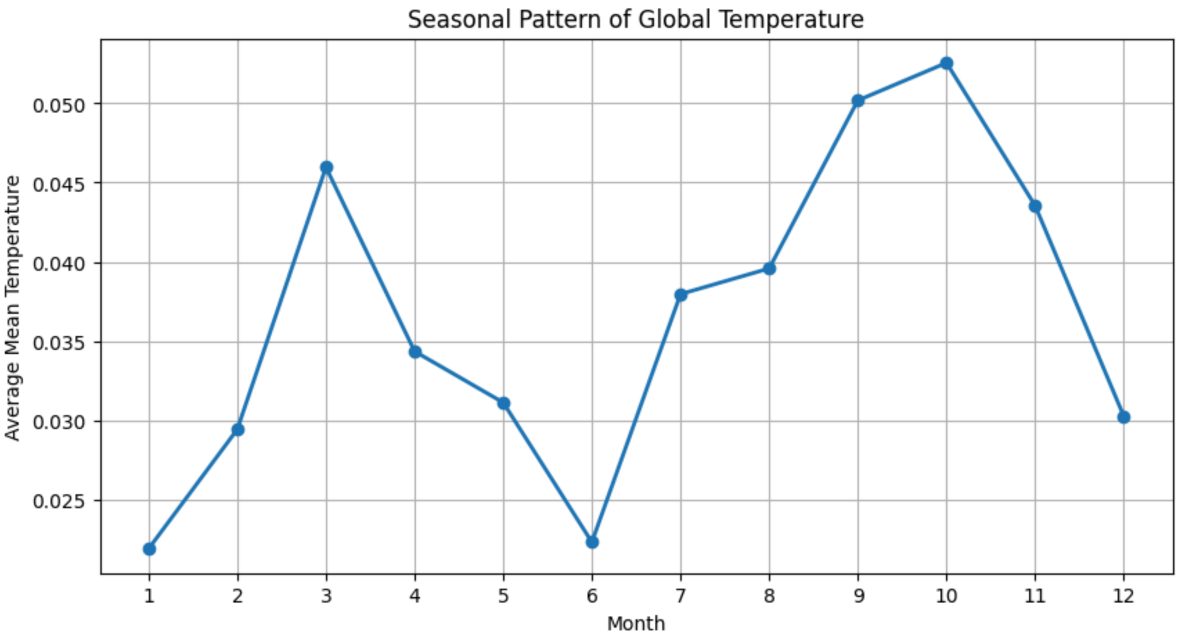
- A gradual upward trend is visible.
- Recent years show comparatively higher global temperatures.

This supports evidence of long-term warming patterns.

## 5.4 Monthly (Seasonal) Analysis

Average temperature across all years was computed for each month.

| | Mean |
|---|---|
| Month | |
| 1 | 0.021871 |
| 2 | 0.029409 |
| 3 | 0.045972 |
| 4 | 0.034341 |
| 5 | 0.031106 |
| 6 | 0.022287 |
| 7 | 0.037955 |
| 8 | 0.039577 |
| 9 | 0.050192 |
| 10 | 0.052566 |
| 11 | 0.043543 |
| 12 | 0.030244 |




## Key Observations

1. **Low temperatures at the beginning of the year (Jan–Feb)**
   Average temperatures are lowest during months 1 and 2. This corresponds to the Northern Hemisphere winter, which strongly influences global averages because most of the Earth's landmass lies in the Northern Hemisphere.
2. **Rise toward early-year warming (March peak)**
   Temperatures increase sharply around March as seasonal transition toward spring occurs in the Northern Hemisphere, increasing solar radiation and warming land surfaces.
3. **Mid-year dip (May–June)**
   A noticeable decrease occurs around June. This happens because:
   - The Southern Hemisphere enters winter.
   - Oceans, which dominate the global area, respond slowly to heating (thermal lag effect).

4. **Strong warming phase (July–October)**
   Temperatures steadily increase from July and reach the **maximum around September–October**.
   This period reflects:
   - Accumulated solar heating during Northern Hemisphere summer.

- Ocean heat storage releasing warmth gradually even after peak sunlight.

5. **Cooling phase (November–December)**
After October, temperatures decline as solar radiation decreases and winter conditions begin again in the Northern Hemisphere.

## Overall Conclusion

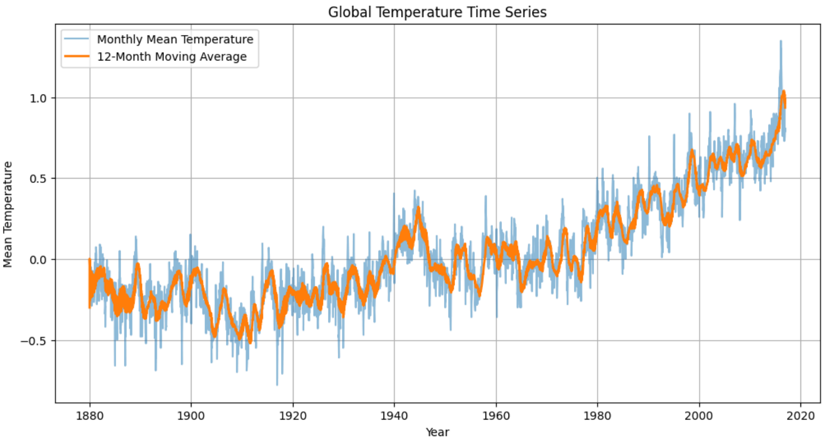
The graph confirms that global temperature follows a predictable seasonal cycle driven by Earth's orbital and atmospheric dynamics, while still maintaining the broader long-term warming trend observed in yearly analyses.

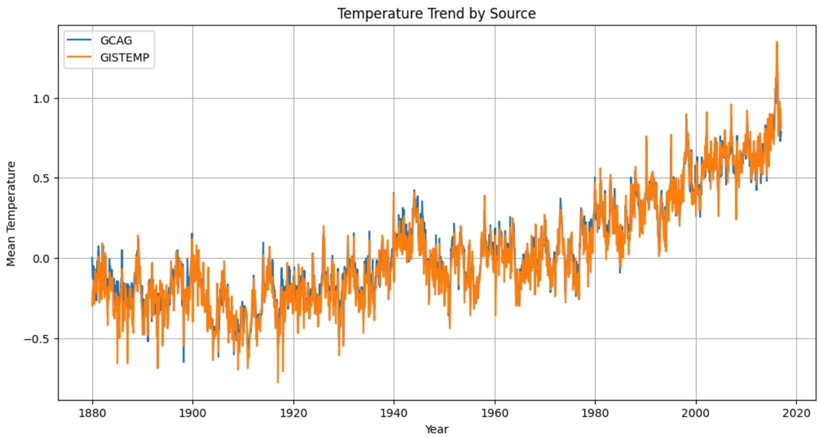
### 5.5 Moving Average Analysis

A **12-month rolling mean** was calculated for each source.

**Purpose:**

- Reduce short-term variability
- Highlight long-term climate trends




**Result:**

- Smooth temperature curve observed.
- The long-term warming trend becomes clearer compared to raw data.

### 5.7 Pivot Table Analysis

A pivot table was created using the filtered dataset containing the last 20 years of observations. The table presents the **average mean temperature** for each **month (columns)** across each **year (rows)**. This structure allows simultaneous comparison of seasonal variation and long-term temperature changes.

Structure of the Pivot Table

- **Rows:** Year
- **Columns:** Month (1–12)
- **Values:** Average Mean Temperature

| Month | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | | | | | | | | | | | | |
| 1997 | 0.35230 | 0.39940 | 0.49085 | 0.41755 | 0.40675 | 0.54040 | 0.43005 | 0.48955 | 0.59250 | 0.64435 | 0.62325 | 0.61510 |
| 1998 | 0.60710 | 0.88300 | 0.63480 | 0.68790 | 0.68895 | 0.72285 | 0.72290 | 0.68080 | 0.48855 | 0.48325 | 0.45660 | 0.56980 |
| 1999 | 0.49035 | 0.66665 | 0.35510 | 0.39950 | 0.36930 | 0.39515 | 0.40645 | 0.35475 | 0.41380 | 0.40365 | 0.39440 | 0.51895 |
| 2000 | 0.30335 | 0.56195 | 0.57805 | 0.61055 | 0.43285 | 0.42380 | 0.40595 | 0.44285 | 0.42920 | 0.30135 | 0.29785 | 0.30960 |
| 2001 | 0.45600 | 0.42730 | 0.60465 | 0.54330 | 0.59485 | 0.54365 | 0.58430 | 0.54660 | 0.53250 | 0.51030 | 0.69400 | 0.53140 |
| 2002 | 0.72310 | 0.77230 | 0.84935 | 0.57625 | 0.60780 | 0.56450 | 0.62145 | 0.54295 | 0.60805 | 0.52365 | 0.59090 | 0.43330 |
| 2003 | 0.71035 | 0.55630 | 0.57010 | 0.55505 | 0.60770 | 0.51405 | 0.54540 | 0.64010 | 0.65060 | 0.73855 | 0.56420 | 0.74305 |
| 2004 | 0.59395 | 0.71660 | 0.66820 | 0.60700 | 0.43210 | 0.45035 | 0.37030 | 0.48205 | 0.53155 | 0.63885 | 0.74005 | 0.51405 |
| 2005 | 0.67115 | 0.54715 | 0.69090 | 0.71070 | 0.65365 | 0.67110 | 0.65860 | 0.62735 | 0.73335 | 0.74650 | 0.74400 | 0.65175 |
| 2006 | 0.51335 | 0.66105 | 0.61155 | 0.50885 | 0.50995 | 0.63995 | 0.58090 | 0.67155 | 0.63685 | 0.68370 | 0.68245 | 0.76960 |
| 2007 | 0.92195 | 0.68045 | 0.68340 | 0.74680 | 0.63950 | 0.55870 | 0.57430 | 0.57545 | 0.60875 | 0.58350 | 0.54360 | 0.49360 |
| 2008 | 0.25730 | 0.36870 | 0.74625 | 0.50635 | 0.50555 | 0.50570 | 0.58980 | 0.49970 | 0.59090 | 0.66950 | 0.66515 | 0.55175 |
| 2009 | 0.60775 | 0.54895 | 0.55470 | 0.63195 | 0.61840 | 0.66205 | 0.68195 | 0.67990 | 0.70415 | 0.64090 | 0.72080 | 0.63380 |
| 2010 | 0.71290 | 0.74400 | 0.88350 | 0.84935 | 0.75410 | 0.68695 | 0.67435 | 0.65475 | 0.59120 | 0.66840 | 0.77910 | 0.48970 |
| 2011 | 0.49030 | 0.51380 | 0.61575 | 0.65110 | 0.55205 | 0.61825 | 0.67910 | 0.67300 | 0.57990 | 0.64745 | 0.52285 | 0.54425 |
| 2012 | 0.44140 | 0.45675 | 0.55090 | 0.71175 | 0.74105 | 0.66195 | 0.62920 | 0.64970 | 0.74330 | 0.74110 | 0.72935 | 0.49775 |
| 2013 | 0.63365 | 0.59285 | 0.63695 | 0.54050 | 0.66205 | 0.66690 | 0.62810 | 0.66025 | 0.73285 | 0.68435 | 0.81965 | 0.68405 |
| 2014 | 0.71180 | 0.49500 | 0.76975 | 0.79030 | 0.83190 | 0.71115 | 0.63440 | 0.80930 | 0.84360 | 0.81755 | 0.67975 | 0.81040 |
| 2015 | 0.81205 | 0.87215 | 0.89845 | 0.75705 | 0.81920 | 0.83175 | 0.75860 | 0.82730 | 0.86800 | 1.02565 | 1.00275 | 1.11595 |
| 2016 | 1.11345 | 1.27105 | 1.26225 | 1.08165 | 0.90190 | 0.83030 | 0.84935 | 0.93990 | 0.87335 | 0.80960 | 0.84020 | 0.79975 |

## Key Observations

### 1. Increasing Temperature Trend Over Years

From earlier years (late 1990s and early 2000s) to recent years (2014–2016), temperature values generally increase across most months.
For example:

- Earlier years show values mostly between **0.3 – 0.7**
- Recent years frequently exceed **0.8 – 1.2**

This indicates a clear **warming trend over time**.

**2. Consistent Seasonal Pattern**

Across nearly all years:

- Mid-to-late months (approximately **September–November**) tend to record higher temperatures.
- Early months often show comparatively lower averages.

This confirms the seasonal cycle previously observed in monthly aggregation analysis.

**3. Strong Warming in Recent Years**

Years such as **2015 and 2016** display noticeably higher temperatures across almost all months, suggesting intensified warming during the most recent period of the dataset.

**4. Month-to-Month Stability**

Although yearly averages increase, the relative ordering of warmer and cooler months remains similar across years. This suggests:

- Seasonal climate structure remains stable,
- But the overall temperature baseline is rising.

## Interpretation

The pivot table effectively combines **time-series** and **seasonal analysis**, revealing two simultaneous patterns:

1. **Long-term global warming trend** (increase across years)
2. **Recurring seasonal temperature cycle** (variation across months)

Thus, the data suggests that global temperatures are increasing while maintaining natural seasonal variability

## Conclusion from Pivot Analysis

The pivot table provides strong visual and numerical evidence that recent decades exhibit higher global temperatures compared to earlier years, supporting observations derived from yearly averages and moving-average analysis.

## 6. Conclusion:

This project applied time-series analysis to examine global temperature data and identify long-term trends and seasonal patterns. Using Python and Pandas for preprocessing, aggregation, and analysis, meaningful insights were derived from real-world datasets provided by GCAG and GISTEMP.

The dataset was complete, with no missing values, making it suitable for statistical evaluation. Converting date fields into datetime format and extracting year and month variables enabled structured temporal analysis. The 12-month moving average effectively reduced short-term fluctuations and highlighted long-term temperature behavior.

The findings show a steady rise in global average temperatures over the years. Recent years recorded noticeably higher values compared to earlier periods, reinforcing evidence of global warming. Both GCAG and GISTEMP datasets produced closely matching results, supporting the reliability of the analysis.

Seasonal patterns followed a consistent annual cycle, with certain months regularly showing higher averages. However, despite stable seasonal variation, the overall temperature baseline has increased over time. Pivot table results further confirmed that rising temperatures are observed across most months in recent years.

**Key Conclusions**

- Global temperature shows a clear long-term upward trend.
- Seasonal patterns remain consistent over time.
- GCAG and GISTEMP datasets demonstrate strong agreement.
- Moving-average analysis effectively reveals underlying climate trends.
- Recent decades exhibit higher temperature averages than earlier years.

## 7. Appendices:

I would like to express my sincere gratitude to **IDEAS-TIH, Indian Statistical Institute (ISI), Kolkata**, for providing me with the valuable opportunity to undertake this internship and work on the present project. The internship served as an excellent platform to enhance my practical understanding of data analysis and time-series methodologies while applying theoretical concepts learned during my academic coursework.

**Github Link:**
**07_Time_Series_Analysis_of_Global_Temperature_Data_Spring_2026.ipynb**