TIME-SERIES ANALYSIS OF GLOBAL TEMPERATURE DATA

Souparna Adhikari

Section A, IDEAS TIH

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 study is based on Time series analysis of global temperature data using Pandas and Numpy. The data sheet uses GCAG and GISTEMP data sources, earliest data ranging from 1880-01-06 to as recent as 2016. We calculated the mean temperature by grouping the source (GCAG and GISTEMP respectively). We calculated the average mean temperature for each year. We also found the hottest month on average.

GCAG- Global Component of Climate at a Glance temperature data. It is provided by NOAA's National Centres for Environmental Information (NCEI), tracks global land and ocean temperature.

GISTEMP- GISS Surface Temperature Analysis. It is provided by NASA, tracks global surface temperature.

2. Introduction

Global Warming is a burning issue for 21st century Earth. Governments and Organisations throughout the world are trying to arrest the rising levels of global warming and striving to minimise the effects of global warming on human life and ecology. The aim is to use the Time series data in studying the progression of Global warming. The Data sources used include GCAG (source-NCEI) and GISTEMP (source-NASA). The data source is stored in csv file format. We are using Python (specifically Pandas and Numpy) in analysing the temperature changes and their peak over 1880 to 2016.

During the first two weeks of internship, I was trained in:

1)PYTHON PROGRAMMING (including PANDAS, NUMPY)
2)MACHINE LEARNING concepts like regression and classification using python functions and libraries.
3)LLM fundamentals
4)Communication Skills

3. Project Objective

The objective was to:

1)Sorting the data with respect to date.
2)Extracting Mean and Moving Average for each month.
4)Finding the mean temperature for both GCAG and GISTEMP sources.
5)Finding the hottest month on average (written as 1-12)

4. Methodology

1) I imported pandas and numpy and inserted ~~an~~a data frame from the csv file and displayed the first five rows of it.
2) Then sorted the data frame by date column using 'pd.to_datetime()' for conversion and '.sort_values(by='Date')' for sorting.
3) Printed the shape and then checked for the total number of missing (null) values. using 'df_temp.shape' to get the dimensions and 'df_temp.isnull().sum().sum()' to count all missing values.
4) Created two new columns in df_temp: 'Year' and 'Month'. Displayed the head of the DataFrame to show these new columns.
After converting to datetime, we can access the year with '.dt.year' and the month with '.dt.month' on the 'Date' column.
5) Calculated the 12-month moving average of the 'Mean' temperature for each 'Source'. Stored this in a new column named Moving_Avg. Display the last 5 rows of the DataFrame.

6) Created a new DataFrame called df_last20 by filtering df_temp to include only data from the last 20 years (inclusive of the latest year in the data). Printed the shape of this new DataFrame.
Used ‘df_temp['Year'].max()’ to find the latest year. Then, filterd the DataFrame where 'Year' is greater than or equal to latest_year - 19.
7) Grouped the original df_temp DataFrame by the 'Source' column and calculated the mean of the 'Mean' temperature for each source. Displayed the resulting Series.
Used the ‘.groupby('Source')['Mean'].mean()’ method to undertake this.
8) Grouped the df_temp DataFrame by 'Year' and calculated the average 'Mean' temperature for each year. Displayed the last 5 entries of this yearly aggregation.
Used ‘.groupby('Year')['Mean'].mean()’ and then ‘.tail()’ on the result to undertake this.
9) Across all years in the df_temp dataset, found which month has the highest average 'Mean' temperature. Printed the resulting Series which shows the average temperature for each month using ‘.groupby('Month')['Mean'].mean()’.
10) Using ‘df_last20.pivot_table(index='Year', columns='Month', values='Mean', fill_value=0)’ for creating a pivot table to show the average 'Mean' temperature for each 'Month' across each 'Year'.

5. Data Analysis and Results

1) Import Libraries and Load Data.

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

2) Convert to Datetime and Sort

DataFrame after converting 'Date' to datetime and sorting:

	Source	Date	Mean
0	GISTEMP	1880-01-06	-0.3000
1	GCAG	1880-01-06	0.0009
2	GCAG	1880-02-06	-0.1229
3	GISTEMP	1880-02-06	-0.2100
4	GCAG	1880-03-06	-0.1357

3) Data Inspection

```
Shape of the DataFrame: (3288, 6)
Total null values in the DataFrame: 0
```

4)Extract year and month

DataFrame with new 'Year' and 'Month' columns:

	Source	Date	Mean	Year	Month	Moving_Avg
0	GISTEMP	1880-01-06	-0.3000	1880	1	-0.3000
1	GCAG	1880-01-06	0.0009	1880	1	0.0009
2	GCAG	1880-02-06	-0.1229	1880	2	-0.0610
3	GISTEMP	1880-02-06	-0.2100	1880	2	-0.2550
4	GCAG	1880-03-06	-0.1357	1880	3	-0.0859

5)Calculating 12-month moving average.

DataFrame with 12-month Moving Average:

	Source	Date	Mean	Year	Month	Moving_Avg
3283	GCAG	2016-06-10	0.7292	2016	6	0.981917
3284	GISTEMP	2016-06-11	0.9300	2016	6	1.017500
3285	GCAG	2016-06-11	0.7504	2016	6	0.963992
3286	GISTEMP	2016-06-12	0.8100	2016	6	0.992500
3287	GCAG	2016-06-12	0.7895	2016	6	0.936292

6) Filter for the Last 20 Years

```
...  Shape of df_last20: (480, 6)
```

7) Group by Source

```
...  Mean 'Mean' temperature by Source:
                Mean
Source
GCAG      0.048797
GISTEMP   0.024380

dtype: float64
```

8) Aggregate by Year

```
Last 5 years of average mean temperature:
            Mean
Year
2012  0.629517
2013  0.661846
2014  0.742075
2015  0.882408
2016  0.964396

dtype: float64
```

9) Find the Hottest Month on Average

```
... Average 'Mean' temperature by Month:
            Mean
Month
1     -0.288450
2     -0.265682
3     -0.226420
4     -0.223945
5     -0.207092
6      0.074481
7     -0.177195
8     -0.178862
9     -0.153875
10    -0.175812
11    -0.219795
12    -0.178333
dtype: float64
```

10) Create a Pivot Table

Pivot table of average 'Mean' temperature by Year and Month:

Year	Month 6
1997	0.500171
1998	0.635542
1999	0.430671
2000	0.424779
2001	0.547404
2002	0.617800
2003	0.616287
2004	0.562087
2005	0.675517
2006	0.622479
2007	0.634167
2008	0.538054
2009	0.640442
2010	0.707358
2011	0.590650
2012	0.629517
2013	0.661846
2014	0.742075
2015	0.882408
2016	0.964396

6. Conclusion

It can be concluded that mean temperature is increasing with succeeding years. June(month-6) stays to be the hottest month throughout the dataset. The mean temperature for GCAG is twice than that of GISTEMP.

7. APPENDICES

1. data.giss.nasa.gov/gistemp/
2. DataHub.io - Global Land-Sea Average Temperatures. Monthly mean average temperatures from 1880–2016.
3.https://github.com/souparnoadhikari-commits/07-Time-Series_Analysis_of_Global_Temperature_Data_Spring_2026-Souparna-Adhikari.git