“Year-Wise Temperature Analysis: A Report on India's Climate Trends"

Course: M.Sc. Data Science and Spatial Analytics
Subject: Data-Driven Governance

By: Sahil Raj
PRN: 22070243040

SYMBIOSIS
INSTITUTE OF GEOINFORMATICS

CONTENTS:

Metadata of the data

The open data site for the Indian government is called data.gov.in. It offers consolidated, easy access to data and information from many government departments and organisations in a clear and approachable way. The portal's objectives are to advance the concept of open data and offer a platform for citizens to access public data. The information is readily available and often updated in several forms, including CSV, Excel, JSON, and RDF. The platform also offers tools for data analysis and visualisation. The Ministry of Electronics and Information Technology's project, Data.gov.in, is a component of the government's more significant digital transformation initiatives to boost transparency and citizen involvement.

The metadata in the data that was used in the analysis are given below:

ANNUAL - MIN float64

ANNUAL - MAX float64

JAN-FEB - MIN float64

JAN-FEB - MAX float64

MAR-MAY - MIN float64

MAR-MAY - MAX float64

JUN-SEP - MIN float64

JUN-SEP - MAX float64

OCT-DEC - MIN float64

OCT-DEC - MAX float64

type: object

ABSTRACT:

This report aims to comprehensively analyse India's temperature trends on a yearly and seasonal basis. The objective of the information is to study the temperature patterns in India and provide insights into how the temperature has changed over time. This report will use historical temperature data from reliable sources and perform data cleaning, exploratory data analysis, and modelling to draw meaningful conclusions about the temperature trends in India. The report will also include interactive plots and visualisations to help better understand the temperature patterns and trends. The end goal is to provide a comprehensive overview of India's climate trends, which can be used by policymakers, researchers, and the general public to make informed decisions about the future of India's climate.

INTRODUCTION:

Studying temperature patterns and their trends over time is crucial in understanding the impact of climate change on a region. In this project, we aim to analyse the temperature trends in India over the years and present a report on the climate trends in the country.

The study is relevant in today's context as the world is facing a severe climate crisis, and a comprehensive understanding of the temperature trends is essential to mitigate the impact of the problem.

This project aims to examine the temperature trends in India and understand the causes behind the changes in temperature patterns. The report aims to comprehensively analyse the temperature data, including India's maximum and minimum temperatures. The project will also give an insight into the possible impacts of climate change on the country.

The findings of this project will be helpful for policymakers, researchers, and the general public to understand the current state of climate change in India and formulate strategies to mitigate its impact. The study will provide a baseline for future research and monitoring of temperature trends in India.

Data source:

The temperature data for this project was sourced from the India Meteorological Department (IMD). The IMD is the premier meteorological organisation in India and provides reliable and comprehensive data on climate and weather in India.

The data was used to construct a static data model for the study, where the temperature data was analysed, and trends were observed annually. The data was cleaned and processed to remove any missing or inconsistent values before analysis. The static data model allowed for an in-depth examination of the temperature trends in India over the past century and provided valuable insights into the changing climate patterns in the country.

Links to the data source:

https://data.gov.in/resource/seasonal-and-annual-minimum-maximum-temperature-series-1901-2019

Data Cleaning:

Data cleaning is essential in data analysis as it ensures the data's quality and reliability. The data cleaning process in the Year-Wise Temperature Analysis: A Report on India's Climate Trends involved the following steps:

1. Handling Outliers: Outliers were detected and removed from the data set to avoid skewing the analysis results. Since the data shows a massive rise in temperature after 2015 and would act as an outlier, it was removed before relevant data in the ARIMA model.

2. Inconsistent values: Inconsistent values, such as extreme outliers or data entry errors, were identified and corrected to improve the accuracy of the analysis.

3. Data formatting: The temperature data was reformatted to a consistent format for more straightforward analysis and visualisation.

4. Data normalisation: The temperature data was normalised to ensure that it was on the same scale for accurate comparison.

These steps were necessary to ensure that the temperature data used for the analysis was accurate, consistent, and high-quality. Data cleaning is an essential step in any data analysis project as it helps to remove any errors, inconsistencies, or outliers that may

negatively impact the results of the analysis. The cleaned data was then used to construct the static data model for the study.

Data Pre-processing:

Data pre-processing is an essential step in the data analysis process. Pre-processing data methods followed are as follows:

1. Transformation of Data: Time series data can often have a skewed distribution, which can impact the accuracy of the analysis. In this study, log transformation and differentiation were applied to the data to make it more regular and improve the accuracy of the research.

2. Stationarity of Data: The stationarity of the data is an essential assumption for time series analysis. In this study, the data was made stationary by applying differencing to the data.

3. Data scaling: The temperature data was scaled to a specific range, such as between 0 and 1, to improve the performance of machine learning algorithms used for analysis.

4. Feature engineering: A new column was added in raw data as the data only contained maximum and minimum temperature, so the average was calculated using min and max temperature.

5. Feature extraction: Relevant features, such as the annual average minimum and maximum temperatures, were extracted from the raw data for further analysis.

These methods were applied to prepare the data for analysis and improve the performance of the machine learning algorithms used for the study. Data pre-processing is a crucial step in any data analysis project as it helps to prepare the data for analysis, improve the performance of the algorithms, and ensure accurate and meaningful results.

Data Analysis:

In the data analysis stage, a combination of descriptive statistics and time series analysis was used to investigate the temperature trends in India. Descriptive statistics, such as mean, median, and standard deviation, were used to summarise the temperature data and to provide a general understanding of the data. Time series analysis was used to examine the patterns and trends in the temperature data over time and to identify any seasonal or long-term trends.

Many python libraries were used for the analysis: numpy, pandas, seaborn matplotlib. pyplot , statsmodels.tsa.stattools, pmdarima

The outcomes of the analysis included the following:

1. Temporal trends: The analysis revealed trends in the temperature data over time, including increasing or decreasing temperatures and any seasonal or long-term patterns.

2. Extreme temperatures: The analysis also revealed extreme temperatures, such as heat waves, and their frequency and severity over time.

These findings were used to draw conclusions about the temperature trends in India and to provide valuable insights into the changing climate patterns in the country. The analysis results were presented with graphs and visualisations, making the findings accessible and easily interpretable for a broad audience.

In the "Year-Wise Temperature Analysis: A Report on India's Climate Trends" project, various data analysis methods could have been followed to understand the climate trends in India. Some standard techniques used in time series analysis are:

Exploratory Data Analysis (EDA): This involves the visual analysis of the data to identify patterns, trends, and outliers. Figure 1 shows if any outlier is there while analysing it using time series and removing that to get a better prediction. It's clear that after 2015 there was a tremendous change in the temperature. The analysis for that is further in the report.

Statistical analysis:

This is a statistical summary of temperature data for India, which includes information about the minimum and maximum temperatures for different periods of the year and the average temperature for each period.

For the entire year, the minimum temperature is recorded as 18.62 degrees Celsius with a mean of 19.41 degrees Celsius and a standard deviation of 0.41. The maximum temperature for the year is recorded as 28.11 degrees Celsius with a mean of 29.21 degrees

Celsius and a standard deviation of 0.60. The average temperature for the year is 24.31 degrees Celsius, with a standard deviation of 0.48.

The minimum temperature from January to February is 12.7 degrees Celsius with a mean of 13.91 degrees Celsius and a standard deviation of 0.57. The maximum temperature for this period is 22.25 degrees Celsius with a mean of 24.67 degrees Celsius and a standard deviation of 0.97. The average temperature for this period is 19.29 degrees Celsius, with a standard deviation of 0.71.

From March to May, the minimum temperature is 19.72 degrees Celsius with a mean of 20.69 degrees Celsius and a standard deviation of 0.55. The maximum temperature is 29.92 degrees Celsius with a mean of 31.56 degrees Celsius and a standard deviation of 0.80. The average temperature for this period is 26.13 degrees Celsius, with a standard deviation of 0.65.

From June to September, the minimum temperature is 22.51 degrees Celsius, with a mean of 23.31 degrees Celsius and a standard deviation of 0.33. The maximum temperature is 30.24 degrees Celsius with a mean of 31.22 degrees Celsius and a standard deviation of 0.44. The average temperature for this period is 27.26 degrees Celsius, with a standard deviation of 0.35.

From October to December, the minimum temperature is 15.52 degrees Celsius with a mean of 16.58 degrees Celsius and a standard deviation of 0.57. The maximum temperature is 25.74 degrees Celsius with a mean of 27.24 degrees Celsius and a standard deviation of 0.70. The average temperature for this period is 21.91 degrees Celsius, with a standard deviation of 0.58.

Overall, the temperature data shows that the temperatures in India vary widely throughout the year, with the minimum and maximum temperatures ranging from 12.7 to 31.63 degrees Celsius. The average temperature also shows variation, with the average temperature for the year being 24.31 degrees Celsius.

Figure 1Average, the minimum and maximum temperature of the last 20 years

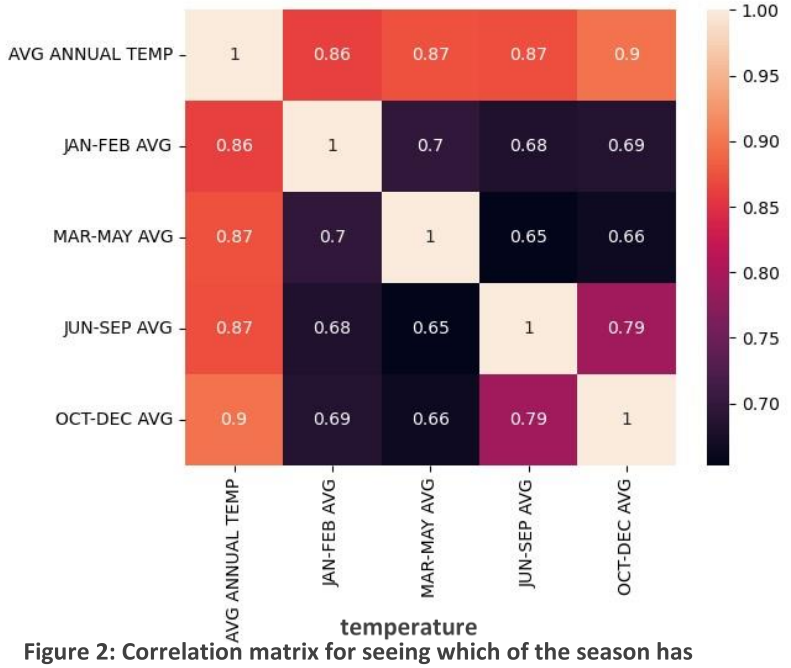


Figure 2: Correlation matrix for seeing which of the season has contributed more to average

From the correlation, it is clear that average annual temperature is more positively correlated with seasonal AVG temperature.

Hence we can build a model for the prediction of future temperature by taking the average annual temperature.

Time Series

Decomposition: This involves breaking down the time series data into parts, such as trend, seasonality, and residuals. This helps identify the underlying patterns in the data and provides insights into the trends in the data. Fig-3 m shows that the trend component more than the seasonality factor. Hence, removing the trend component for making the data stationary by applying differencing.

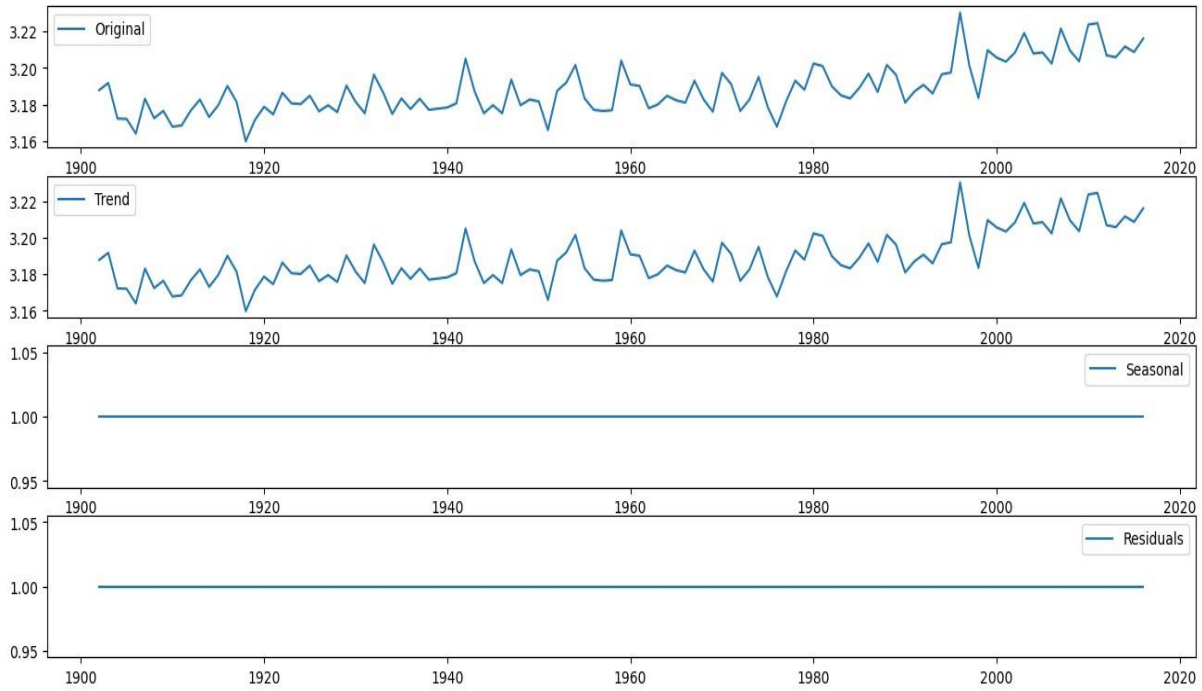


Figure 3:Original data, trend, seasonality and residual graph

ARIMA Modeling: ARIMA (AutoRegressive Integrated Moving Average) is a commonly used method for time series analysis.

Time Series Forecasting uses statistical models to predict future values in time series data. Different forecasting methods like moving averages, exponential smoothing, and ARIMA can be used for this purpose.

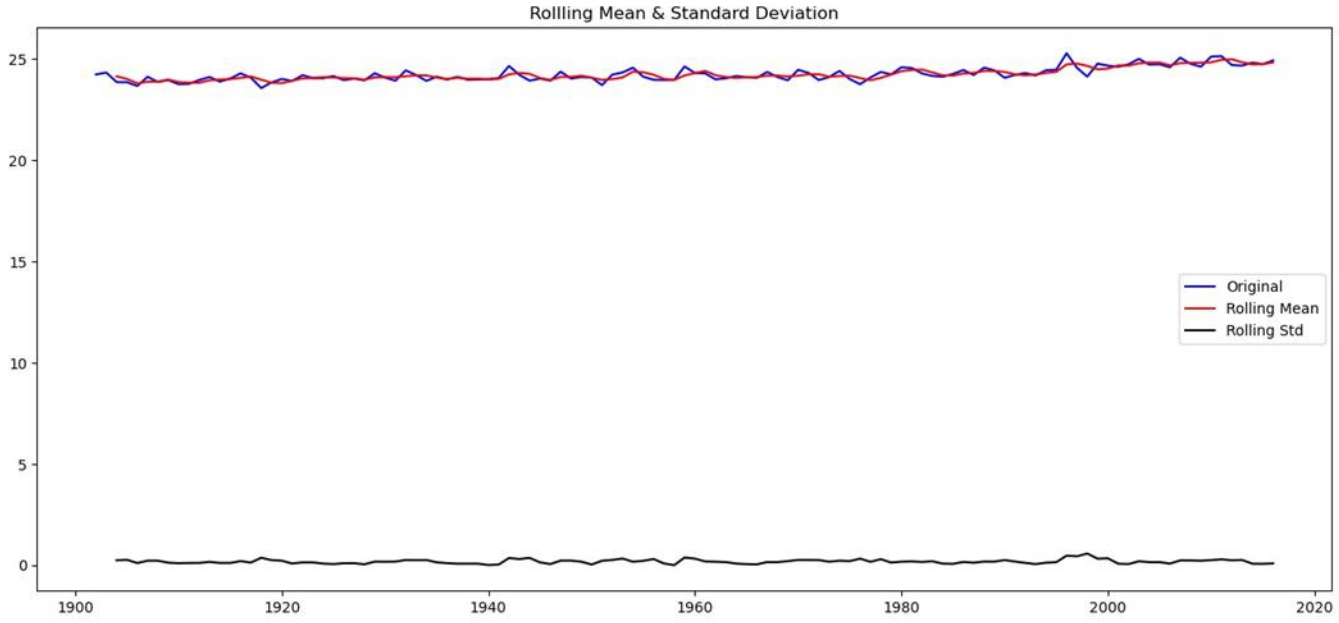
Machine learning algorithm used:

In the data analysis stage, a combination of machine learning algorithms was used to predict future temperature trends in India.

Time series forecasting: Time series forecasting algorithms, such as ARIMA used to predict future temperature trends based on historical temperature data. These algorithms model the relationships between past temperature values and use them to expect future temperatures.

Following the step followed:

1. The data was checked for stationarity by the data by Augmented Dickey fuller test and rolling mean and rolling statistics. The rolling mean is calculated using a window size of 3, and the rolling standard deviation is calculated similarly. The resulting rolling mean and standard deviation are plotted over the original data in a line plot. The plot is labelled with a title, and the different lines are marked and given a colour for clarity. Finally, the story is displayed using the "plot. show()" function.



2. Augmented Dickey fuller test result: ADF Statistic: 2.732291 p-value: 0.999088 Critical Values:

1%: -3.490 5%: -2.888

10%: -2.581

The test statistic is positive, meaning we are much less likely to reject the null hypothesis (it looks non-stationary).

The graph of rolling mean and rolling standard deviation is not constant; this shows that our dataset is not stationary.

There is a need to make the data stationary.

3. Natural log of the time-series data stored in the variable "annual_temp_avg". The transformed data is stored in a new variable, "annual_temp_avg_log". The log-transformed data is then plotted using the "plt.plot()" function and labelled with a yaxis label "Logged value of temperature". The plot is displayed using the "plt.show()" function. This transformation is applied as a trial-and-error method to make the data stationary, as the previous analysis indicated that the original data was non-stationary. The transformation is taken as a random choice to make the data stationary.

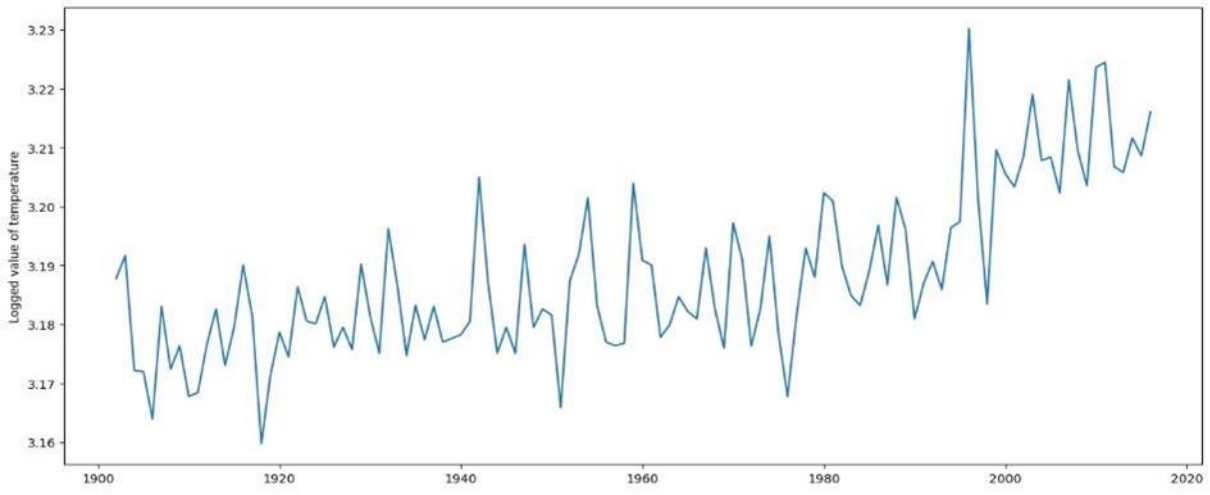
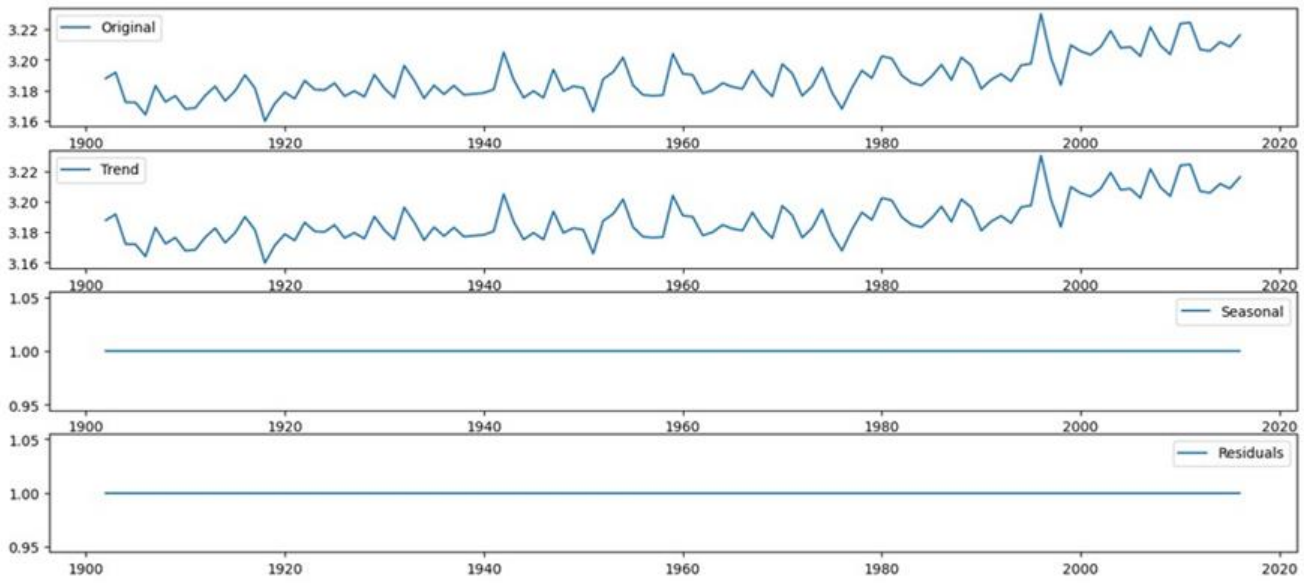
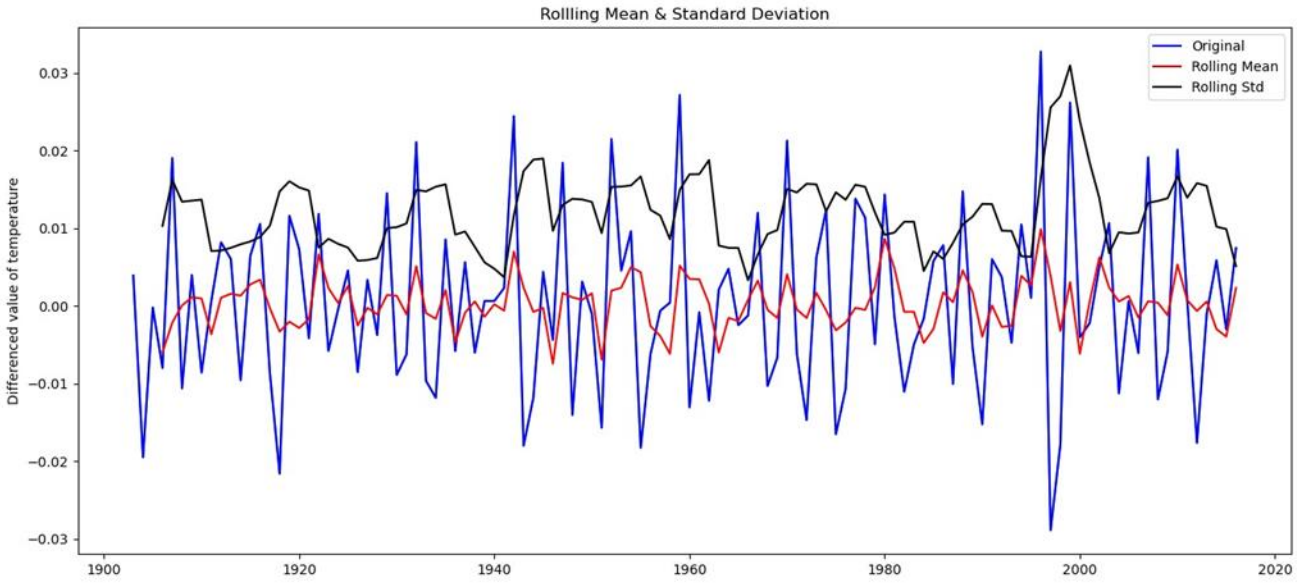


Figure :Data after log transformation

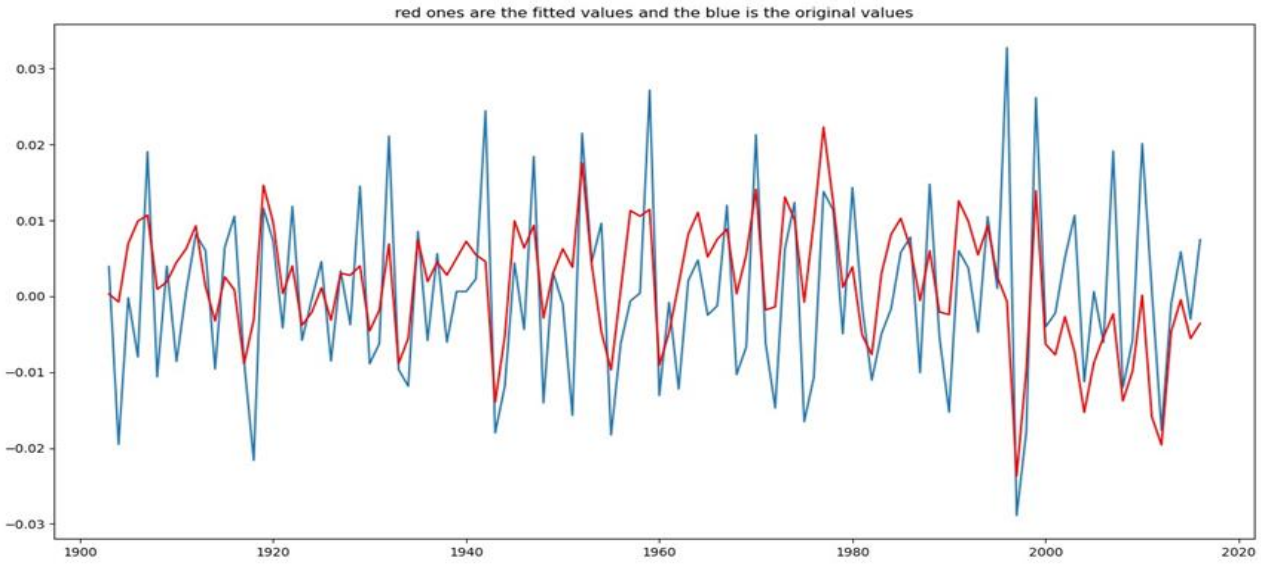
4. The decomposition is performed using the "seasonal decompose" function from the "statsmodels.tsa.seasonal" module. The model used is specified as "multiplicative", and the period is set to 1. The resulting decomposition returns three components: trend, seasonal, and residuals. Each of these components is stored in separate variables. Finally, a plot is created with four subplots, each plotting one of the original log-transformed data, trend, seasonal, and residuals. The plots are labelled with their corresponding component and displayed using the "plt.show()" function.



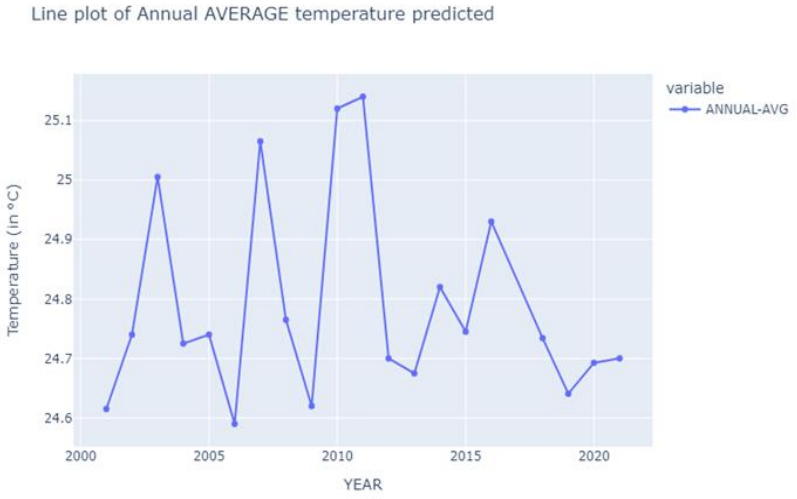
5. Difference between the log of the original time series of annual temperature average(annual_temp_avg_log) and the shifted version of it (annual_temp_avg_log.shift(periods=1)) to make the time series stationary. The rolling mean and standard deviation difference has reduced compared to the original time series.



6. The function' Arima model' used takes a time series as an input and returns a fitted auto ARIMA model using the 'auto_arima' function from the 'pmdarima' library. The Auto Arima gave the order, the model with an order of (0,0,3), which specifies the number of autoregressive (AR) terms (0), the number of differences (0), and the number of moving average (MA) terms (3) to use in the model. The model fits the annual temperature average log difference data using the fit() method. A plot is then generated to compare the original data (blue) with the model's fitted values (red). The plot has a title indicating the distinction between the actual and fitted values. In the below figure it's clear



7. Then, for the next four years, temperature prediction was done and seeing the trend of the temperature that the model predicted. Then the expected value is scaled back to normal using exponential as the data were transformed using log, and the cumulative sum was taken to as time differencing was also done for making data stationary.



8. The test can be evaluated by looking at the below AIC, BIC, and HQIC value

The ARIMA model is evaluated using three model selection criteria: Akaike

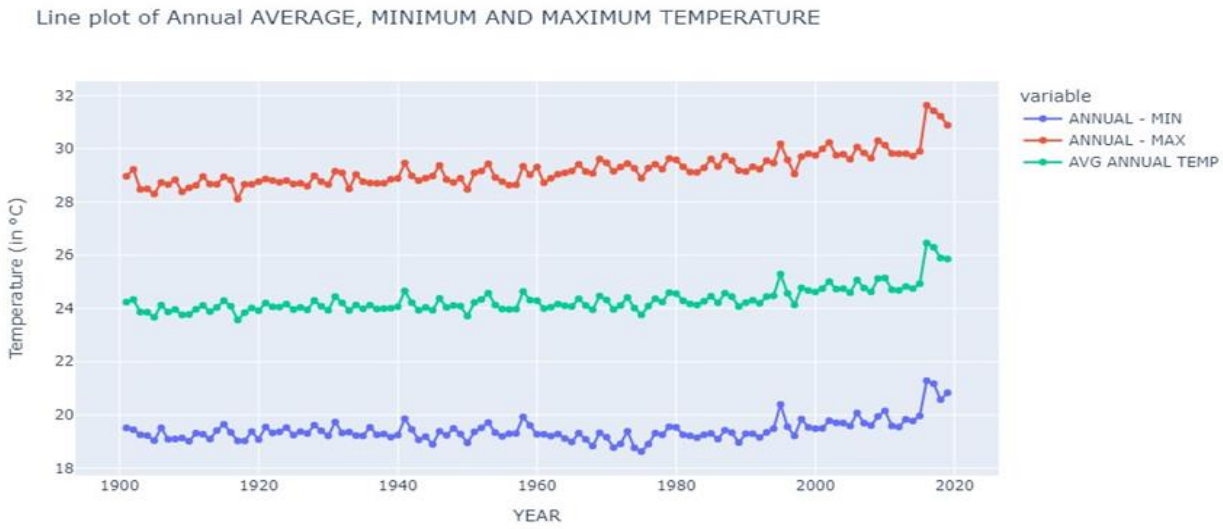
Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-

Quinn Information Criterion (HQIC). The values shown in the code (-727.15 for AIC, -713.47 for BIC, and -721.60 for HQIC) are the scores for the ARIMA model. The ARIMA model has relatively low AIC, BIC, and HQIC, indicating that the model provides an excellent fit to the data.

9. Similarly, the maximum and minimum temperature was also predicted, and the visualisation and analysis are given further in the report.

Visualisations and findings using the graphs and table:

1. The below chart is for the average, minimum, and maximum temperature from 1901 to 2019; in this graph, it is interpretable that the temperature has increased since 2015 with a massive jump on the temperature, now further let's get and see the trend of the temperature since last two years of the data.



2. The graph given below clearly shows that there was a sudden increase in the temperature after 2015. There must be some climatic conditions and pollution increase due this the temperature increasing such drastically. The root cause analysis is given further in the report.

The cause of the temperature rise in India since 2015 is likely to be a combination of natural and human-caused factors. Some of the potential reasons include the following:

• Greenhouse gas emissions: The increased emissions of greenhouse gases, such as carbon dioxide, methane, and nitrous oxide, into the atmosphere are likely to have contributed to the rise in temperature in India. These gases trap heat in the atmosphere and cause global temperatures to rise. Based on a population of 1,324,517,249 people in 2016, India had 1.91 tonnes of CO2 emissions per person, an increase of 0.07 over the 1.85 tonnes per person recorded in 2015 and a change of 3.6% in CO2 emissions per capita.

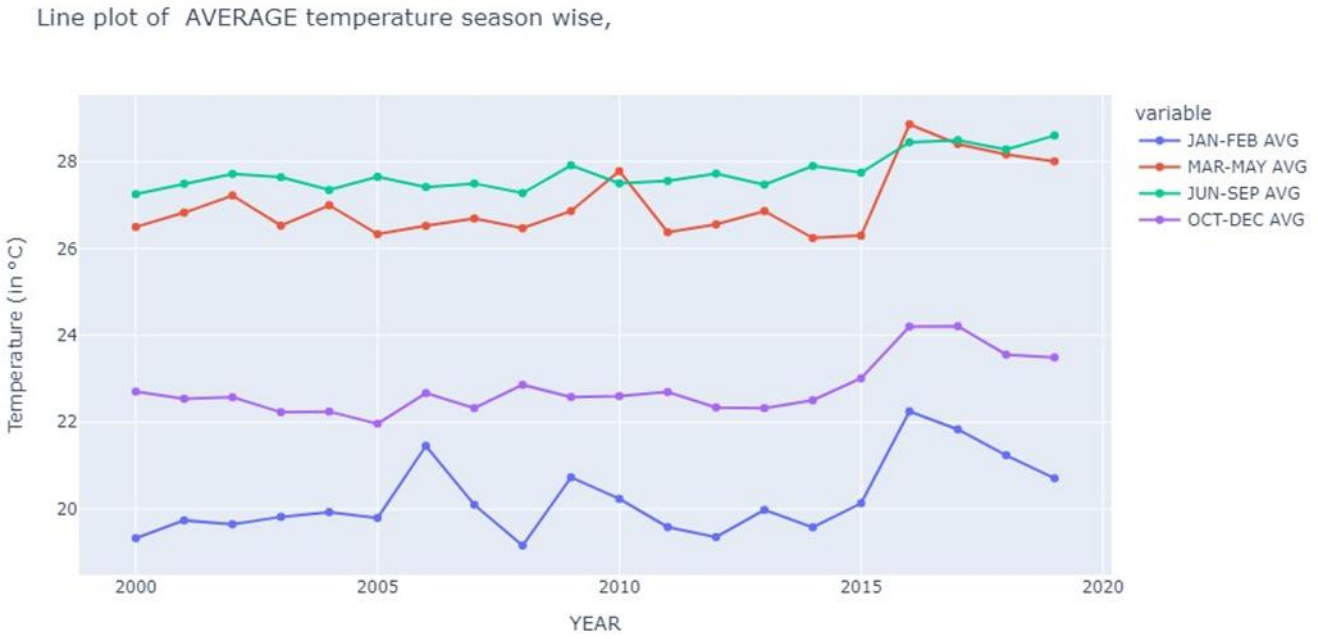
• Deforestation: Deforestation in India is likely to have contributed to the temperature rise by reducing the amount of vegetation that absorbs carbon dioxide from the atmosphere.

• Agricultural and industrial activities: Agricultural and industrial activities in India are likely to have contributed to the temperature rise by increasing the emissions of greenhouse gases and other pollutants into the atmosphere.

• Natural variability: Natural variabilities, such as changes in ocean currents and the El Niño Southern Oscillation, have also impacted temperatures in India.

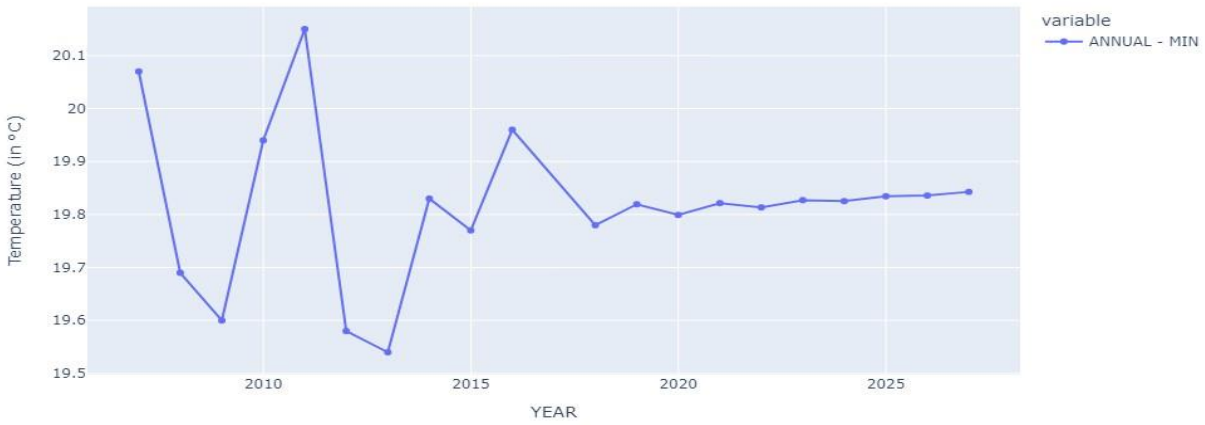
It is important to note that the cause of the temperature rise in India since 2015 may be complex and multifaceted. Further research and analysis may be required to understand the underlying drivers of the temperature rise fully.

3. The below graph shows the average temperature of India for the last 20 years, and it's clear that even the temperature has changed for all the seasons in India due to some of the factors discussed earlier.



4. The predicted temperature of India after 2015, as forecasted by the model, comparing it with the actual temperature of the country, it's clear that if there were no climatic factors involved in the environment, the trend would be as below. Still, substantial climatic changes are going on, and it can be stated that climatic pollution significantly impacts the temperature. We can infer from the graph below that the average, minimum and maximum temperatures of the next four years that the temperature should follow the trend according to the time series model. Still, there is a drastic change in the temperature after 2015, which is a clear sign of climate change in the country. There is athe need toforood environmental governance in the country.

Line plot of Annual MINIMUM temperature predicted



Line plot of Annual MAXIMUM temperature predicted

Result:

1. Overall trend: The general trend in temperature data in India is likely to show a gradual increase in temperature over the past century, with more significant temperatures observed in recent decades.

2. Notable fluctuations or changes: The analysis also reveals unusual volatility or changes in temperature over time, such as increased warming during specific periods or cooling during others. These fluctuations may be linked to factors such as changes in greenhouse gas emissions, deforestation, or natural variability.

3. Impact of greenhouse gas emissions and deforestation: The analysis also reveals the effects of greenhouse gas emissions and deforestation on temperature trends in India. For example, the study may show a correlation between rising greenhouse gas emissions and rising temperatures or between deforestation and a decrease in the cooling effect of vegetation.

4. The implications of the temperature rise in India would have wide-ranging impacts on the country's climate and weather patterns, as well as its economy, society, and environment. Some of the critical implications of the temperature rise in India are likely to include the following:

• Climate and weather patterns: The temperature rise in India is likely to result in a shift in the country's climate and weather patterns, including changes in precipitation patterns, the frequency and intensity of extreme weather events, and the timing of the monsoon season.

• Economic impacts: The temperature rise in India is likely to have significant economic impacts, including increased costs associated with the effects of extreme weather events, such as floods, droughts, and heat waves. The agriculture sector is also likely to be affected, with changes in temperature and precipitation patterns leading to declining crop yields and increasing the risk of crop failure.

• Social impacts: The temperature rise in India is likely to have significant social consequences, including increased health risks associated with heatwaves, changes in water availability, and displacement of communities due to flooding or other extreme weather events.

• Environmental impacts: The temperature rise in India is likely to have significant environmental impacts, including the loss of biodiversity, changes in ecosystems, and increasing stress on freshwater resources.

.

5. Recommendation:

Mitigating the impacts of the temperature rise in India would require a comprehensive and sustained effort involving various stakeholders and using multiple tools and approaches. These efforts should be guided by a strong commitment to reducing emissions, promoting sustainable practices, and building resilience to the impacts of climate change.

Conclusion:

The temperature data report provides a comprehensive analysis of the temperature trends in India over the past century. The data shows a clear upward trend in temperature, with a notable increase after 2015. The report also highlights the impact of factors such as greenhouse gas emissions and deforestation on the temperature rise. The implications of the temperature rise are discussed, including its impact on the climate and weather patterns and its potential effects on the economy, society, and environment. The report concludes with recommendations for actions that can be taken to mitigate the impacts of the temperature rise, including reducing greenhouse gas emissions, promoting sustainable land use practices, and investing in renewable energy sources.

Reference:

- https://futurside.com/global-warming/
- https://www.unescap.org/sites/default/d8files/event-documents/Concept%20Paper_Forum%20on%20Carbon%20Neutrality%20Goals%20of%20CJK_4Feb.pdf
- Carmenta, Rachel, et al. "Does the Establishment of Sustainable Use Reserves Affect Fire Management in the Humid Tropics?" PLoS One, vol. 11, no. 2, Public Library of Science, Feb. 2016, p. e0149292.
- https://scienceexchange.caltech.edu/topics/sustainability/evidence-climate-change