SDG – 13: CLIMATE ACTION

Project Title

Impact of Climate change on Population and Forest area

ES1101: COMPUTATIONAL DATA ANALYSIS

FACULTY GUIDE

Mr. Santosh Verma

Dr. Rabia Kamra

Mr. Arpan Gupta

PREPARED BY: GROUP 7 (Section A)

Amit Saraswat (2022Btech010

Keshav Pareek(2022Btech050)

Amit Talmale(2022Btech011)

Abhimanyu Sharma(202Btech003)

Institute of Engineering & Technology (IET)

JK LAKSHMIPAT UNIVERSITY, Jaipur

January 2022

INDEX

ABSTRACT

SDG Goal 13, also known as the Climate Action Goal, aims to take urgent action to combat climate change and its impacts. The goal includes targets such as strengthening resilience and adaptive capacity to climate-related hazards, integrating climate change measures into national policies, and increasing the ability to deal with the impacts of climate change. Additionally, it calls for enhancing education, awareness-raising, and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning. The goal also aims to mobilize increased financial resources from a variety of sources to address climate change, and to support developing countries in their efforts to combat climate change. Climate action refers to efforts taken by individuals, organizations, and governments to mitigate and adapt to the effects of climate change. This can include reducing greenhouse gas emissions using renewable energy sources, implementing policies to encourage sustainable practices, and investing in research and development of new technologies. Additionally, climate action can include measures to increase resilience and prepare for the impacts of climate change that are already occurring, such as sea level rise and increased frequency of extreme weather events. Overall, climate action is essential for addressing the urgent global challenge of climate change and protecting the planet for future generations.

INTRODUCTION

What do we mean by the "Sustainable"?

- Sustainable refers to the ability to meet the needs of the present without compromising the ability of future generations to meet their own needs. It encompasses environmental, social, and economic sustainability.

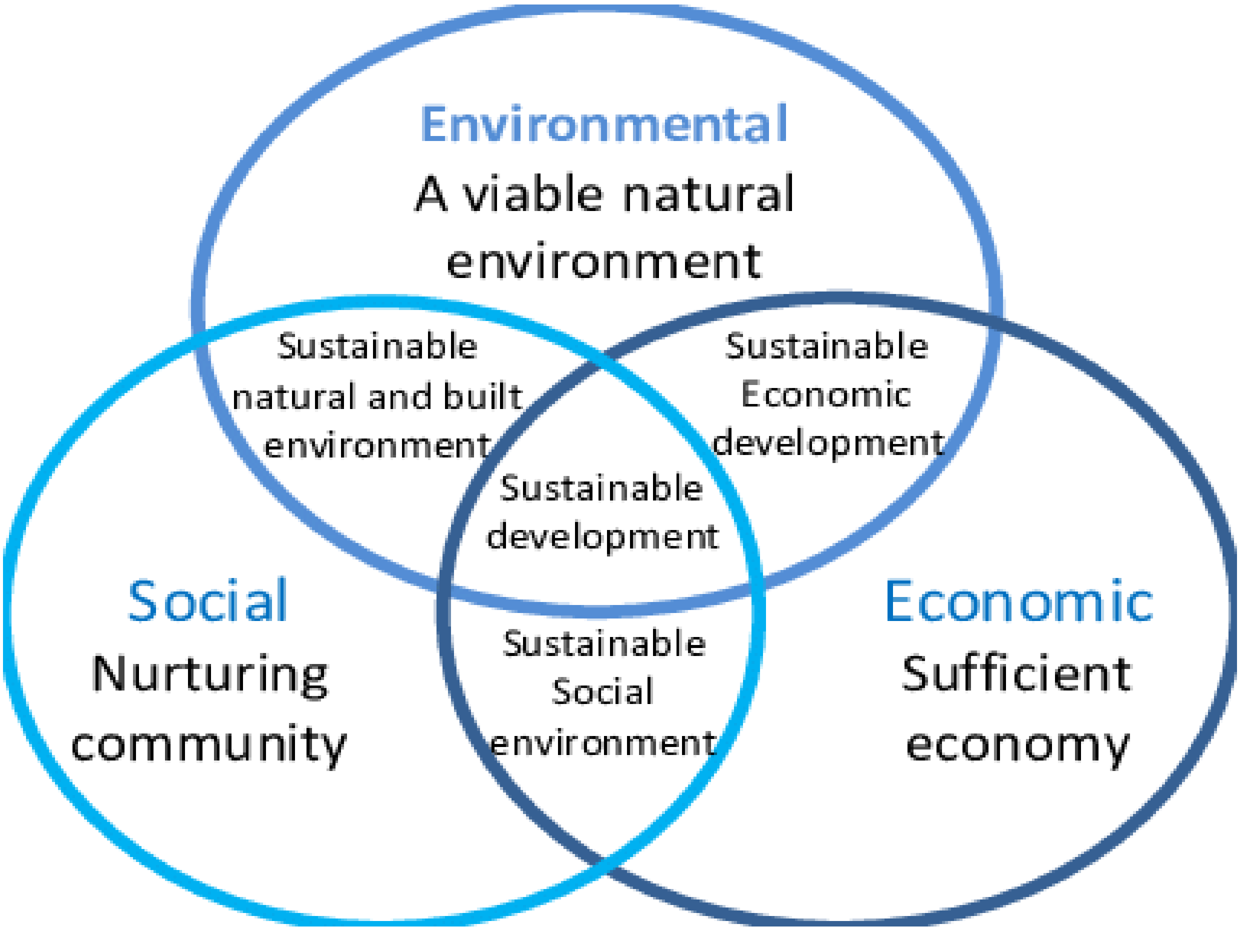


Figure 1: environment, economy, and equity

This includes responsible use of natural resources, reducing pollution and waste, and promoting economic and social equality. Adopting sustainable practices can also lead to cost savings and improved efficiency in the long run. It is important for individuals, businesses, and governments to actively work towards a more sustainable future.

“(UN) Sustainable Development goal”

The Sustainable Development Goals (SDGs) is made to end poverty, protect the planet, and ensure that all people enjoy peace and prosperity.

Figure 2: Sustainable Development Goal

- The main purpose of SDG goals is to promote sustainable development d improve the well-being of all people globally.
- Sustainability is a complex concept that has been derived from various fields of study such as ecology, economics, and social science. The concept of sustainability has evolved over time to encompass a wide range of issues including environmental, social, and economic concerns.
- The term "sustainability" was first used in a report by the United Nations World Commission on Environment and Development in 1987, where it was defined as "meeting the needs of the present without compromising the ability of future generations to meet their own needs."
- Sustainable development is a concept that aims to meet the needs of the present without compromising the ability of future generations to meet their own needs.
- It is often framed in terms of three pillars: economic, social, and environmental sustainability. Economic sustainability focuses on creating and maintaining a strong and stable economy that supports economic growth and job creation.

- Social sustainability focuses on creating and maintaining a just and inclusive society that supports the well-being and quality of life of all members.
- Environmental sustainability focuses on protecting and preserving the natural environment and natural resources for future generations.

There are 17 SDG Goals.

1) **No-poverty**: The goal aims to reduce the number of people living in extreme poverty, which is defined as living on less than $1.90 per day and ensure that all people have access to basic needs such as food, shelter, and healthcare. Additionally, the goal also aims to reduce inequalities and empower people living in poverty to improve their own livelihoods.
2) Zero-Hunger**:** The goal specifically aims to end hunger, achieve food security and improved nutrition, and promote sustainable agriculture. This includes addressing issues such as malnutrition, hunger, food waste, and the need for sustainable farming practices that protect the environment and biodiversity. The goal targets to end hunger, achieve food security and improved nutrition and promote sustainable agriculture by 2030**.**

3) Good health and wellbeing: established by the United Nations in 2015 to end poverty, protect the planet and ensure that all people enjoy peace and prosperity by 2030. SDG-3 specifically aims to "ensure healthy lives and promote well-being for all at all ages." This includes reducing maternal and child mortality rates, ending epidemics of communicable diseases, and improving mental health and well-being. It also includes increasing access to affordable and quality healthcare services, as well as reducing the number of deaths and illnesses from non-communicable diseases such as cancer and diabetes.

4) Quality Education: It Ensure that all people enjoy peace and prosperity by 2030. SDG-4 specifically focuses on ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all. This includes increasing access to quality education and vocational training, reducing gender and income disparities in education, and increasing the number of qualified teachers and education professionals.

5) Gender Equality: SDG-5 specifically aims to achieve gender equality and empower all women and girls. This includes eliminating all forms of discrimination and violence against women, promoting equal participation and leadership for women in decision-making, and ensuring universal access to sexual and reproductive health and rights.

6) Clean Water and Sanitation**:** To ensure availability and sustainable management of water and sanitation for all. This goal aims to address the global water crisis by increasing access to clean and safe water, improving sanitation and hygiene, and protecting and restoring water-related ecosystems. The target of SDG-6 is to achieve universal and equitable access to safe and affordable drinking water, sanitation, and hygiene for all by 2030**.**

7) Affordable And Clean Energy: SDG 7 specifically aims to ensure access to affordable, reliable, sustainable, and modern energy for all by 2030. This includes increasing the share of renewable energy in the global energy mix and providing access to energy in rural and remote areas. It also aims to improve energy efficiency and increase the use of clean energy technologies.

8) Decent Work and Economic Growth: Goal 8, is a goal set by the United Nations to promote sustained and inclusive economic growth, full and productive employment, and decent work for all. This goal aims to improve living standards and reduce poverty, while also promoting gender equality

and creating opportunities for all people to participate in the economy. Specific targets under this goal include reducing the proportion of workers living in extreme poverty, increasing the number of formal and decent jobs, and ensuring universal social protection.

9) Industry Innovation and Infrastructure: The aims to "Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation." This goal includes targets to develop quality, reliable, sustainable, and resilient infrastructure, in areas such as transportation, information and communications technology, energy, and water and sanitation. Additionally, it aims to promote industrialization that is inclusive and sustainable and to support innovation and research and development. The goal is to achieve a sustainable future for all by 2030.

10) Reduced Inequality: The United Nations in 2015 to guide global development efforts until 2030. It aims to reduce inequality within and among countries, by addressing issues such as poverty, discrimination, and social exclusion. Specific targets under SDG-10 include increasing the income of the bottom 40% of the population, reducing the number of people living in extreme poverty, and ensuring equal access to economic opportunities and basic services.

11) Sustainable Cities and Communities: The goal aims to reduce the number of deaths and number of people affected by disasters, improve access to safe and affordable housing, and improve the quality of urban life. It also aims to provide access to safe, affordable, accessible, and sustainable transport systems for all, and enhance inclusive and sustainable urbanization and capacity for participatory, integrated, and sustainable human settlement planning and management.

12) Responsible Consumption and Production: SDG 12 focuses on ensuring responsible consumption and production patterns. This includes promoting sustainable lifestyles and reducing waste through measures such as recycling and reusing resources, as well as encouraging businesses to adopt sustainable practices. The goal also aims to increase the use of renewable energy and improve energy efficiency, as well as reduce the use of harmful chemicals and pesticides.

13) Climate Action: Its Aims to "Take urgent action to combat climate change and its impacts. The goal of SDG-13 is to mitigate the negative impacts of climate change on the planet and its inhabitants, and to promote sustainable development practices that will help ensure a sustainable future for all. This is essential for the development of the countries and for the well-being of the people.

14) Life Below Water: The goal is to "conserve and sustainably use the oceans, seas and marine resources for sustainable development." The specific targets of SDG 14 include increasing the economic benefits to Small Island developing States and least developed countries from the sustainable use of marine resources, protecting marine biodiversity, and reducing marine pollution.

15) Life On Land: It aims to protect, restore, and promote the sustainable use of terrestrial ecosystems, forests, wetlands, and other habitats, and to combat desertification, land degradation, and loss of biodiversity. This goal includes targets such as promoting conservation and restoration of biodiversity, reducing deforestation, and improving land use practices.

16) **Peace Justice And Strong Institutions:** Goal 16 is a goal set by the United Nations to promote peaceful and inclusive societies, provide access to justice for all, and build effective, accountable, and inclusive institutions at all levels. This goal aims to reduce violence and crime, ensure equal access to justice for all, and strengthen the rule of law and good governance. It also aims to promote inclusive and participatory decision-making and ensure that all individuals and groups can participate in the political process and hold public officials accountable.

17) Partnership For Goals: Goal 17, is a goal set by the United Nations as part of their 2030 Agenda for Sustainable Development. It aims to strengthen the means of implementation and revitalize the global partnership for sustainable development by addressing key issues such as financing for development, trade, technology transfer, and capacity building. The goal includes 17 specific targets, including increasing the availability of financial resources for developing countries, improving access to affordable and reliable energy, and promoting fair trade. The goal is intended to ensure that all countries have the necessary resources and support to achieve the other 16 sustainable development goals.

GOAL 13

CLIMATE ACTION

Figure 3: Climate Action

- Reducing greenhouse gas emissions: This is one of the most important aspects of climate action, as greenhouse gases are the main drivers of climate change.
- Increasing renewable energy: This includes the use of solar, wind, geothermal, and hydro energy, which are less harmful to the environment than fossil fuels.
- Improving energy efficiency: This involves using less energy overall and using it more efficiently.
- Adapting to the effects of climate change: This includes taking steps to protect communities and infrastructure from the impacts of extreme weather, rising sea levels, and other effects of climate change.
- Promoting sustainable transportation: This includes increasing the use of electric cars, public transportation, and biking and walking to reduce carbon emissions from cars.
- Investing in research and development: This includes funding research on new technologies and ways to reduce greenhouse gas emissions.

- Encouraging international cooperation: Climate change is a global problem and requires cooperation between countries to address it effectively.

- Raising public awareness: It is important to educate the public about the causes and effects of climate change and the actions that can be taken to mitigate it.

INDIA's role in SDG goal 13:

- India has set a target to achieve 175 GW of renewable energy capacity by 2022, which includes 100 GW of solar, 60 GW of wind, 10 GW of biomass, and 5 GW of small hydro power.

- India has been actively participating in international negotiations on climate change, and it's playing a key role in the formation of International Solar Alliance.

- India has taken steps to promote sustainable transportation, including the launch of the National Electric Mobility Mission Plan, which aims to put 5-6 million electric vehicles on the road by 2020.

- India has been running campaigns to raise public awareness about climate change and the actions that can be taken to mitigate it.

OBJECTIVES

OBJECTIVE 1: To analyse and predict the impact on the temperature change in India.

Problem 1.1: To predict the mean temperature change from 2021 to 2025 using the data available years from 2000-2020(regression).

Problem 1.2: Ranking Indian states based on mean temperature change from 2000 to 2020.

Problem 1.3: To check the claim that there is significant difference between mean temperature change from 2020 to 2025(using hypothesis)

OBJECTIVE 2: To Analyse the change in seasonal rainfall [like rain in Winter (January-February), Summer (May-June), Monsoon (April-September)] in India due to climate change.

Problem 2.1: To predict the amount of seasonal rainfall in 2016-2031 using data from 2000-2015 of India. (Regression)

Problem 2.2: To rank the India based on seasonal rainfall in years 2000-2015 and 2016-2031

Problem 2.3: To study the change in the seasonal rainfall from 2000-2015 and predict the seasonal rainfall from 2016-2031.

OBJECTIVE 3: To analyse and study on impact of CO_2 emission on climate in INDIA's States.

Problem 3.1: To predict the amount of Green House Gas emission from 2022 to 2025, by historical data from 1990 to 2019 (By using regression analyses).

Problem 3.2: To determine the rank of all the Sectors, based on the amount of Green House Gases emission from 1990 to 2019 and for the year 2025. (Power method).

Problem 3.3: To check the claim that contribution of all the Sectors in releasing Greenhouse gas is not same in INDIA.

OBJECTIVE 4: Analysis on deforestation in India from 2001 to 2020 and ranking of state.

Problem statements:

Problem 4.1:To predict the deforestation in year 2021-2025 on basis from 2001-2020(regression).

Problem 4.2: To Implementing hypothesis on observed value of 2020 and predicted value of 2040.

Problem 4.3: To rank the states according to deforestation in 2020. (By ranking)

SURVEY OF LITERATURE

SR.NO	AUTHOR	LITERATURE
1	Intergovernmental Panel on Climate Change	• Climate action refers to the various measures and actions taken to address and combat climate change.

		This includes reducing greenhouse gas emissions, increasing resilience to the impacts of climate change, and integrating considerations of climate change into policy and decision-making. According to the WG3 of the Intergovernmental Panel on Climate Change (IPCC), global greenhouse gas emissions have grown significantly since pre-industrial times, particularly in the transport and buildings sectors. However, there is potential for reducing these emissions through various mitigation strategies. These strategies include increasing the use of renewable energy, implementing sustainable practices in urban areas, promoting energy-efficient buildings, and investing in sustainable transportation infrastructure. Climate action is a crucial step in addressing the pressing issue of climate change and requires the participation of governments, businesses, and individuals.
2	Intergovernmental Panel on Climate Change	• According to the report of IPCC (Intergovernmental Panel on climate change) Impact of climate change on India's agriculture. Climate change is having a significant impact on Indian agriculture. Rising temperatures are causing changes in weather patterns and weather extremes, which are affecting crop yields. For example, in the Indo-Gangetic Plain, an increase of 0.5 degrees Celsius in temperature during the pre-monsoon season is expected to negatively impact the wheat crop. In certain states such as Jharkhand, Odisha, and Chhattisgarh, rice production losses during droughts are estimated to be around 40% of total production, with an estimated value of $800 million. Additionally, an increase in CO_2 levels to 550 parts per million is expected to increase yields of crops such as rice, wheat, legumes, and oilseeds by 10-20%. However, an increase in temperature by 1 degree Celsius is expected to reduce yields of crops such as wheat, soybean, mustard, groundnut, and potato by 3-7%. Productivity of most crops is projected to decrease only

		slightly by 2020, but by as much as 10-40% by 2100 due to increasing temperatures
3	Intergovernmental Panel on Climate Change	• According to the report of IPCC (Intergovernmental Panel on climate change) Impact of climate change on India's rainfall. The majority of India's annual rainfall occurs during a specific period, the southwest monsoon months, which runs from June to September, accounting for 87% of the total annual rainfall. In contrast, the post-monsoon months (October to February) and pre-monsoon months (March to May) account for only 9.4% and 3.6% of annual rainfall, respectively. The study also revealed a decline in both monthly and annual normal rainfall during the period 1986-2015 in comparison to 1956-1985, with an annual rainfall deficiency of 205.3 mm recorded. The study also found that there is a decline in seasonal normal rainfall during June-September (kharif), which is a significant agricultural period. Additionally, the analysis revealed that the maximum temperature (T max) decreased by 0.4 °C, while the minimum temperature (T min) increased by 0.21 °C
4	Intergovernmental Panel on Climate Change	• The impact of changes in rainfall distribution on agricultural production. The research team finds that changes in rainfall distribution can have a significant negative impact on agricultural production. The study also highlights the need for drought protection and resilience plans in the face of changing climate conditions. The study cites the Intergovernmental Panel on Climate Change (IPCC) which reports that changes in climate are likely to cause distress to agriculture and increase the risk of hunger and water scarcity. Additionally, the study references other research which shows that the amount of freshwater in rivers in India is likely to decrease due to changing climate, which in combination with a growing population, could negatively affect a large population

		in India by 2050.
1.	Vijay Kumar, Sharad K. Jain and Yatveer Singh	The present study has examined trends in the monthly, seasonal and annual rainfall on the meteorological sub-division scale, the regional scale, and for the whole of India. A large data set was used, consisting of 306 stations with the length of data series of 135 years. As expected, the sub-divisional rainfall trends show a large variability – nearly half of the sub-divisions have shown an increasing trend in annual rainfall.
2.	R. H. Kripalani, Ashwini Kulkarni, S. S. Sabade & M. L Khandekar	They examined that In the light of the IPCC global warming projections on the Asian monsoon, the interannual and decadal variability in summer monsoon rainfall over India and its teleconnections have been examined by using observed data for the 131-year (1871–2001) period. While the interannual variations show year-to-year random fluctuations, the decadal variations reveal distinct alternate epochs of above and below normal rainfall. The epochs tend to last for about three decades. There is no clear evidence to suggest that the strength and variability of the Indian Monsoon Rainfall (IMR) nor the epochal changes are affected by the global warming. Though the 1990s have been the warmest decade of the millennium (IPCC, 2001), the IMR variability has decreased drastically
3	V. Sathiyamoorthy and P.C. Joshi	conducted they analysed rainfall simulated by 15 models that have participated in the IPCC-AR4 to study the likely changes in rainfall over India during the summer monsoon season due to doubling of CO_2. We selected 6 out of 15 models that have simulated the annual cycle of rainfall reasonably well over the Indian region in the 20th century experiments for rainfall projection study. The multi-model ensemble of rainfall simulated by 6-selected models in the doubling of CO_2 experiment suggest an increase in rainfall rate over the Indian summer monsoon region during entire year expect spring. Also 5 out of the 6 models suggest an increase in rainfall over the head Bay of Bengal and adjoining regions
4.	Bushra Praveen, , Swapan Talukdar, Shahfahad, Susanta Mahato, Jayanta Mondal, Pritee Sharma, Abu Reza Md. Towfiqul Islam & Atiqur Rahman	The temporary change in rainfall distribution significantly distresses the agricultural production. It will increase the drought protection and resilience plans under the changing climate conditions.

		Intergovernmental Panel on Climate change (IPCC) reported that upcoming changes in climate is to be likely to distress agriculture that will amplify the chance of hunger and water paucity, as well as the instructions on quicker thawing of glaciers115. Gosain *et al.*116 pointed that the amount of freshwater in the river of India is probable to be diminish because of changing climate. This reduction, along with growing population might unfavourably affect a large population in India by 2050.
5.	Yen YiLoo, LawalBilla , AjitSingh	This study has given some insights on the connections between global warming and monsoon rainfall. It is evident that the distribution of monsoon rainfall is greatly influenced by a number of weather systems, such as the Arctic Oscillation, Siberian High and Western Pacific Subtropical High, as well as the complex Asian topography, i.e. the Tibetan Plateau. The EAWM is regulated by the Arctic ice which influences the SH weather system. The EASM is affected by the westward shift of the WPSH and consequently impacting on the distribution and variability of monsoon precipitation. Excessive monsoon flooding which has become frequent in recent years in parts of Southeast Asia remain an issue to be overcome. Understanding the shift and predicting changing trends of monsoon may be central to managing the floods that impact on millions of people, damage to lives and property, destruction of ecology and farmlands and the long term effect on food security.

METHODOLOGY

STATISTCS

Statistics is used for collecting, analysing, and interpreting data to make informed decisions. It involves the use of mathematical and computational techniques to describe, understand, and make inferences about a population or sample of data.

Statistics divided into two types:

Statistics

DESCRIPTIVE	INFERENTIAL
Descriptive statistic is used to deal with the summarization and description of a set of data.	Inferential statistic is used to deal with making inferences and predictions about a population based on a sample of data.
It also includes graphical methods such as histograms, box plots, and scatter plots, to visually represent and understand the data.	Inferential statistics helps in estimating population parameters and testing hypotheses about relationships between variables.
Mean, mode, and median are used to calculate the central tendency. Range, variance, and standard deviation is calculated for measure the variability of data.	The goal of inferential statistics is to use sample data to make predictions or inferences about a larger population.

❖ Methods of Descriptive Statistics:

1. **Mean (‘$\bar{X}$’):**

Mean is the sum of observations divided by the total number of observations.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Here,

n= total number of observations

Σ= sum of total number of observations

Xi= i^{th} observations

2. **Mode:**

Mode defined as the number which occur the highest time.

3. **Median:**

The median is a statistical measure that represents the middle value of a dataset when it is arranged in ascending or descending order.

If n is odd:

$$\left(\frac{n+1}{2}\right)^{th} \text{ observation}$$

If n is even:

$$\frac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th} \text{ observation}}{2}$$

Here, n=total number of observations

4. **Standard Deviation:**

Standard deviation is a measure of the spread of a dataset, calculated as the square root of the variance of the dataset.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

Here, n=total number of observations

μ=sample mean

5. **Variance (σ^2):**

The variance is a measure of the spread of a dataset, calculated as the average of the squared differences of the data points from the mean.

$$\sigma^2 = \sum_{i=1}^{N} \frac{(X - \mu)^2}{N}$$

N= total number of Observations

Inferential Statistics Methods

Hypothesis Testing:

Hypothesis is an assumption according to the population parameter. It involves two hypotheses: the null hypothesis (H0), which represents the status quo or default assumption, and the alternative hypothesis (H1), which represents the opposite of the null hypothesis.

Null Hypothesis (H_0):

Null Hypothesis is a statistical hypothesis that there state is no difference between the parameter and specific value.

Alternative Hypothesis (H_1):

Alternative Hypothesis is a statistical hypothesis that there state is difference between the parameter and specific value.

Significance level (α):

Significance level (α) is a probability threshold (usually set at 0.05) that is used in hypothesis testing to determine the level of statistical significance and the likelihood of rejecting a null hypothesis when it is true.

Critical Region:

A critical region is a specific range of values defined by the significance level (α) in a hypothesis test, where if the test statistic falls within this range, the null hypothesis will be rejected in favour of the alternative hypothesis.

Non-Critical Region:

A non-critical region, also known as a "acceptance region", is the set of values for the test statistic that would not lead to the rejection of the null hypothesis in a hypothesis test. It is the complement of the critical region. If the test statistic falls within the non-critical region, then the null hypothesis is not rejected, and the sample data is considered consistent with the null hypothesis.

Critical Value:

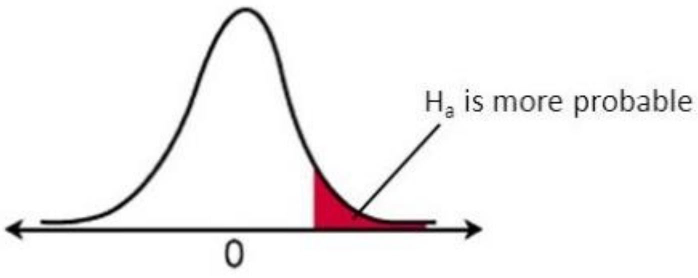
A critical value is a specific value of the test statistic that separates the critical region from the non-critical region in a hypothesis test and is used to determine whether to reject or fail to reject the null hypothesis.

One Tailed tests:

1. **Right Tailed Test:**

H_0: μ ≤ k

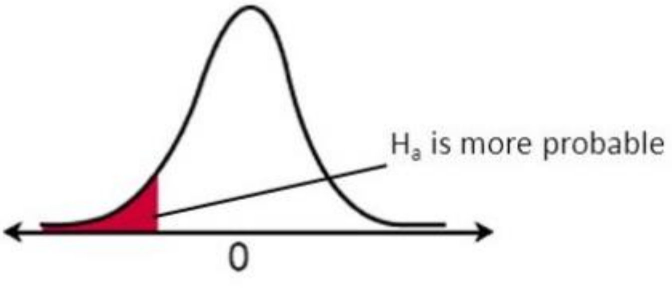
H_a: μ > k



2. **Left Tailed Test:**

H_0= μ ≥k

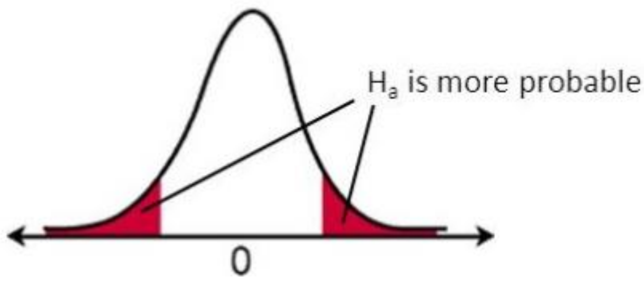
H_a= μ < k



Two Tailed test:

H_1: μ = k

H_a: μ ≠ k



T-Test :

In a t-test, the null hypothesis typically states that there is no difference between the means of two groups, while the alternative hypothesis states that there is a difference. The t-test is used to determine if the difference in means between the two groups is statistically significant.

The t-test is a type of inferential statistical test that is used when the sample size is small and the population standard deviation is unknown. The t-test uses a t-statistic to determine the probability that the difference in means is due to chance. If the probability is low (typically less than 0.05), the null hypothesis is rejected, and it is concluded that there is a statistically significant difference between the means of the two groups.

The formula for hypothesis testing using a t-test is:

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

Where:

t = the calculated t-value

$\bar{x}$ = the sample mean

μ = the population mean (the value being tested in the hypothesis)

s = the sample standard deviation

n = the sample size

ANOVA Test:

i. The ANOVA (Analysis of Variance) test is used to determine whether there are significant differences between the means of two or more groups. Here are the steps to perform an ANOVA test:
ii. State the null and alternative hypotheses. The null hypothesis is that there is no significant difference between the means of the groups, while the alternative hypothesis is that there is at least one significant difference between the means of the groups.

iii. Select the level of significance, typically denoted as alpha (α). Common values for alpha are 0.05 or 0.01.
iv. Determine the degrees of freedom (df) for the test. The degrees of freedom for the numerator are the number of groups - 1, and the degrees of freedom for the denominator is the total sample size - the number of groups.
v. Collect data for the groups being compared and calculate the mean, variance, and sample size for each group.
vi. Calculate the overall mean for the entire sample.
vii. Calculate the sum of squares between groups (SSB) by summing the squared differences between each group mean and the overall mean, multiplied by the size of each group.
viii. Calculate the sum of squares within groups (SSW) by summing the squared differences between each individual value and the group mean, for all groups.
ix. Calculate the mean square between groups (MSB) by dividing SSB by the degrees of freedom for the numerator.
x. Calculate the mean square within groups (MSW) by dividing SSW by the degrees of freedom for the denominator.
xi. Compare the calculated F-value (MSB/MSW) to the critical F-value from a table of F-distributions, using the appropriate degrees of freedom and level

Regression:

Regression is a statistical method used to model the relationship between a dependent variable (y) and one or more independent variables (x). The goal of regression is to find the best-fitting line or equation that describes the relationship between the variables.

Linear regression: It is the most common type of regression analysis, which assumes a linear relationship between the independent and dependent variables. The formula for a simple linear regression is:

$y = b_0 + b_1*x$

where:

y = dependent variable

x = independent variable

b_0 = y-intercept, the point where the line crosses the y-axis

b_1 = slope of the line, representing the change in y for a unit change in x

e = error term

POWER method:

The power method is a simple and efficient algorithm used to find the dominant eigenvalue and corresponding eigenvector of a matrix. The eigenvalue is the scalar value that, when multiplied by the eigenvector, results in the same vector as the original matrix multiplied by the eigenvector. The dominant eigenvalue is the eigenvalue with the largest absolute value. Here are the steps to perform the power method:

- Select an initial approximation for the eigenvector, typically denoted as x. This can be any non-zero vector.
- Multiply the initial approximation of the eigenvector by the matrix to get a new approximation of the eigenvector.
- Normalize the new approximation of the eigenvector by dividing it by its magnitude.
- Repeat steps 2 and 3 multiple times until the approximation of the eigenvector converges or reaches a desired level of accuracy.
- The dominant eigenvalue is the ratio of the magnitudes of the new and old approximations of the eigenvector. The dominant eigenvector is the final approximation of the eigenvector.

Then ranking based on eigenvector.

RESULT AND DISCUSSION

OBJECTIVE:

To analyse and predict the impact on the temperature change in India.

Problem 1: To predict the mean temperature change from 2021 to 2025 using the data available from keggle(regression)

Problem 2: Ranking Indian states on the basis of mean temperature change from 2020 to 2025(by rank())

Problem 3: To check the claim that there is significant difference between mean temperature change from 2020 to 2025(using hypothesis)

DATASET:

States	2000	2001	2002	2003	2004	2005	..	2016	2017	2018	2019	2020
Andaman and Nicobar	26.98	27.08	27.18	27.26	27.29	27.32	..	27.69	27.58	27.32	27.76	27.8
Andhra Pradesh	27.69	27.97	28.05	28.06	27.68	27.82	..	28.23	28.29	28.03	28.37	28.06
Arunachal Pradesh	18.21	18.71	18.63	18.73	18.43	18.97	..	18.99	19.09	18.76	19.01	18.89
Assam	22.85	23.56	23.33	23.33	23.27	23.6	..	23.58	23.57	23.21	23.44	23.33
Bihar	25.38	25.73	25.79	25.69	25.79	25.82	..	25.81	25.57	25.2	25.37	25.18
Chandigarh	23.89	24.01	24.04	23.39	24.16	23.4	..	24.83	24.45	24.42	23.56	23.71
Chhattisgarh	26.08	26.21	26.5	26.32	26.13	26.21	..		26.4	26.07	26.29	25.98
Dadra and Nagar Haveli	26.51	26.55	26.8	26.7	26.79	26.26	..	26.85	26.89	26.97	26.79	26.8
Daman and Diu	26.7	26.7	26.92	26.89	27.01	26.4	..	27.03	27.04	27.19	26.96	27.04
Delhi	25.46	25.47	25.84	25.09	25.8	25.06	.	26.45	26.07	26.03	25.28	25.31
Goa	26.66	26.99	27.18	27.18	26.88	26.74	..	27.25	27.39	27.27	27.39	27.28
Gujarat	27.23	27.15	27.48	27.29	27.58	26.85	..	27.61	27.56	27.74	27.41	27.44
Haryana	25.29	25.28	25.63	24.83	25.59	24.8	..	26.17	25.84	25.84	25.06	25.09
Himachal Pradesh	10.26	10.42	10.33	9.79	10.49	9.69	..	11.18	10.78	10.74	9.95	10.07
Jharkhand	24.96	25.35	25.47	25.39	25.34	25.43	..	25.31	25.08	24.65	24.96	24.69
Karnataka	25.89	26.18	26.33	26.4	26.02	26.02	..	26.51	26.59	26.44	26.67	26.51
Kerala	26.38	26.57	26.74	26.78	26.43	26.56	..		26.9	26.73	27.05	27.01
Lakshadweep	27.58	27.74	27.96	27.99	27.69	27.84	…	28.41	28.13	27.92	28.26	28.3

Madhya Pradesh	25.81	25.78	26.24	25.91	26.04	25.69	..	26.29	26.17	26.04	25.94	25.75
Maharashtra	26.75	26.9	27.14	27.06	26.93	26.67	..	27.2	27.27	27.14	27.2	27.04
Manipur	19.28	20.03	19.79	19.74	19.66	20.05	..	19.94	19.81	19.5	19.85	19.7
Meghalaya	21.36	22.16	21.87	21.88	21.9	22.12	..	22.05	21.89	21.51	21.82	21.66
Mizoram	21.15	21.9	21.67	21.55	21.54	21.89	..	21.73	21.51	21.1	21.54	21.33
Nagaland	18.8	19.49	19.26	19.29	19.13	19.58	..	19.49	19.47	19.13	19.41	19.32
Orissa	25.95	26.28	26.47	26.35	26.13	26.3	..	26.35	26.25	25.86	26.16	25.86
Puducherry	28.38	28.57	28.66	28.72	28.36	28.52	..		28.87	28.72	29.08	28.93
Punjab	24.85	24.85	25.01	24.24	25.01	24.2	..	25.59	25.28	25.29	24.45	24.54
Rajasthan	26.18	26.03	26.51	25.86	26.51	25.67	..	26.8	26.55	26.68	26.1	26.08
Sikkim	4.54	4.97	4.69	4.75	4.83	5.03	..	5.08	4.97	4.49	4.79	4.57
Tamil Nadu	27.5	27.69	27.8	27.85	27.49	27.64	..	28.15	27.99	27.82	28.18	28.07
Tripura	24.35	25.14	24.86	24.8	24.84	25.11	..	24.94	24.7	24.3	24.72	24.51
Uttar Pradesh	25.87	25.93	26.24	25.87	26.21	25.81	..	26.56	26.26	26.12	25.74	25.62
Uttarakhand	14.28	14.4	14.49	13.99	14.57	13.95	..	15.26	14.87	14.81	14.05	14.14
West Bengal	25.47	26.16	26.05	26.01	26.08	26.23	..	26.07	25.82	25.42	25.73	25.51

Problem 1.1: To predict the mean temperature change from 2021 to 2025 using the data available from keggle(regression)

- **Prediction using Linear Regression**

```
In [58]: from sklearn.linear_model import LinearRegression

In [59]: regr=LinearRegression()

In [60]: y = df['2020']
         x = df['2019']
         Sxx = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
         Sxx

Out[60]: 931.2333529411771

In [61]: Syy = np.sum(y**2) - ((np.sum(y)**2)/y.shape[0])
         Syy

Out[61]: 927.2875764705896

In [62]: Sxy = np.sum(y*x)-((np.sum(x)*np.sum(y))/y.shape[0])
         Sxy

Out[62]: 928.9755411764745
```

- **Find slope and intercept**

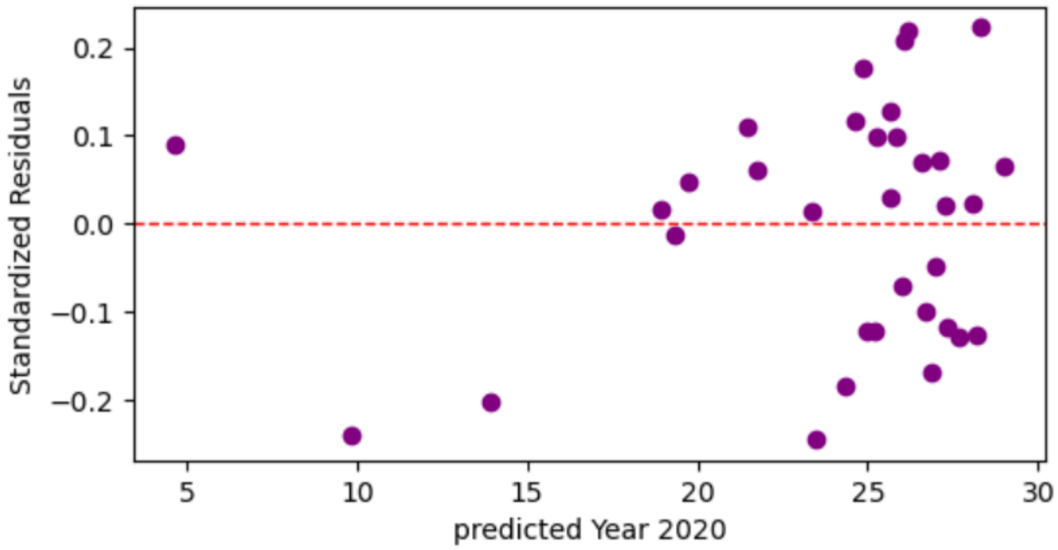
```
In [64]: b_xy = Sxy/Syy
         b_xy

Out[64]: 1.001820324944188

In [65]: b_yx=Sxy/Sxx
         b_yx

Out[65]: 0.9975754608041351

In [66]: a=np.mean(y)-b_xy*np.mean(x)
         a

Out[66]: -0.13830480632499587
```

```
In [67]: y=a+b_xy*x
y
```

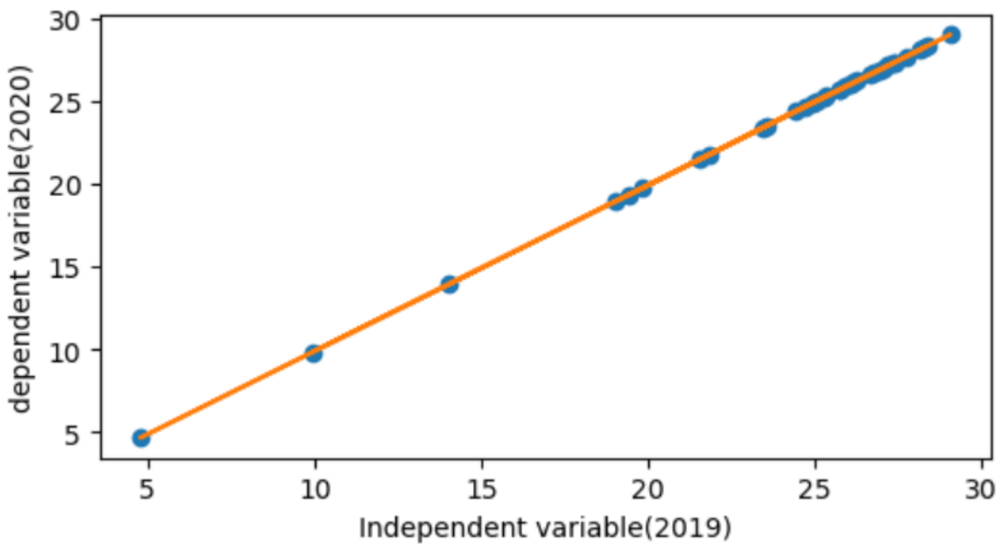
- **Residual**

```
In [68]: residual=y-df['2020']
residual
```

- **Residual Graph**



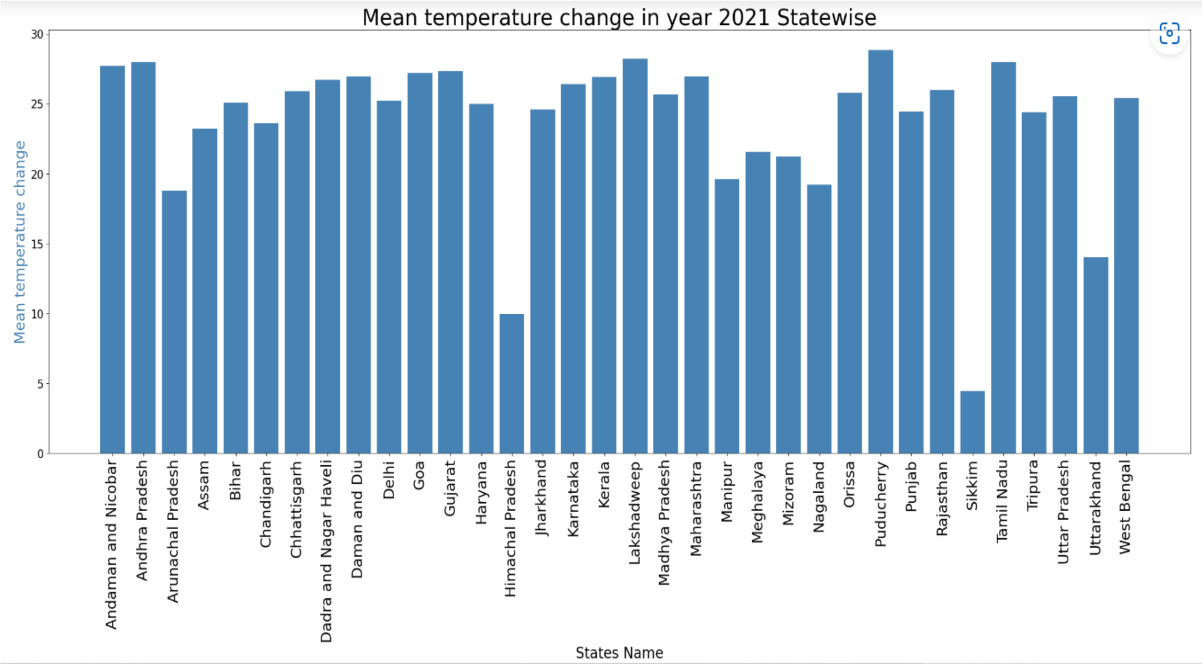
- **Regression line graph**



- **Prediction**

states	2021	2022	2023	2024	2025
Andaman and Nicobar	27.712300	27.624441	27.536421	27.448242	27.359902
Andhra Pradesh	27.972774	27.885388	27.797844	27.710140	27.622277
Arunachal Pradesh	18.786081	18.681973	18.577676	18.473188	18.368511
Assam	23.234163	23.138152	23.041966	22.945606	22.849069
Bihar	25.087531	24.994894	24.902088	24.809113	24.715969
Chandigarh	23.614855	23.519537	23.424045	23.328380	23.232540
Chhattisgarh	25.888987	25.797809	25.706464	25.614954	25.523276
Dadra and Nagar Haveli	26.710480	26.620797	26.530951	26.440941	26.350767
Daman and Diu	26.950917	26.861671	26.772264	26.682693	26.592959
Delhi	25.217768	25.125367	25.032799	24.940062	24.847156
Goa	27.191354	27.102546	27.013577	26.924445	26.835152
Gujarat	27.351645	27.263129	27.174452	27.085613	26.996613
Haryana	24.997367	24.904566	24.811595	24.718456	24.625146
Himachal Pradesh	9.950026	9.829833	9.709422	9.588792	9.467941
Jharkhand	24.596639	24.503108	24.409407	24.315535	24.221493
Karnataka	26.419952	26.329740	26.239364	26.148823	26.058118
Kerala	26.920862	26.831562	26.742099	26.652474	26.562685
Lakshadweep	28.213210	28.126263	28.039157	27.951892	27.864469
Madhya Pradesh	25.658569	25.566971	25.475206	25.383274	25.291175
Maharashtra	26.950917	26.861671	26.772264	26.682693	26.592959
Manipur	19.597556	19.494925	19.392107	19.289102	19.185910
Meghalaya	21.561123	21.462067	21.362830	21.263412	21.163814
Mizoram	21.230523	21.130864	21.031025	20.931003	20.830800
Nagaland	19.216864	19.113540	19.010028	18.906328	18.802439
Orissa	25.768769	25.677372	25.585808	25.494078	25.402180
Puducherry	28.844357	28.758558	28.672604	28.586492	28.500224
Punjab	24.446366	24.352561	24.258586	24.164440	24.070122
Rajasthan	25.989169	25.898173	25.807011	25.715684	25.624190
Sikkim	4.440014	4.309792	4.179332	4.048635	3.917700
Tamil Nadu	27.982792	27.895425	27.807899	27.720213	27.632368
Tripura	24.416311	24.322452	24.228422	24.134221	24.039848
Uttar Pradesh	25.528332	25.436497	25.344495	25.252325	25.159988
Uttarakhand	14.027435	13.914664	13.801689	13.688507	13.575120
West Bengal	25.418132	25.326096	25.233893	25.141522	25.048983

- **Graph of mean temperature change in 2021**



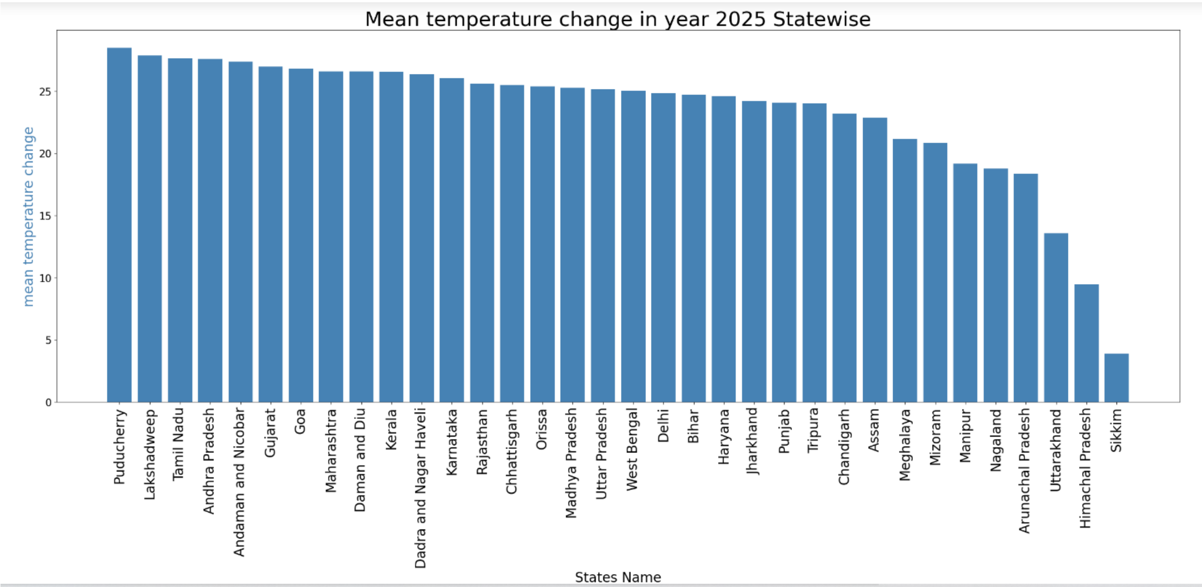
Problem 2: Ranking Indian states on the basis of mean temperature change from 2020 to 2025(by rank())

```
In [114]: df['rank']=df['2025'].rank(method="min",ascending=False)
          df.sort_values('rank',inplace=True)
          df
```

Rank	States
1	Puducherry
2	Lakshadweep
3	Tamil Nadu
4	Andhra Pradesh
5	Andaman and Nicobar
6	Gujarat
7	Goa
8	Maharashtra
9	Daman and Diu
10	Kerala
11	Dadra and Nagar Haveli
12	Karnataka
13	Rajasthan
14	Chhattisgarh
15	Orissa
16	Madhya Pradesh
17	Uttar Pradesh
18	West Bengal
19	Delhi
20	Bihar
21	Haryana

22	Jharkhand
23	Punjab
24	Tripura
25	Chandigarh
26	Assam
27	Meghalaya
28	Mizoram
29	Manipur
30	Nagaland
31	Arunachal Pradesh
32	Uttarakhand
33	Himachal Pradesh
34	Sikkim

- **Mean temperature change ranking graph in 2025**



Problem 3: To check the claim that there is significant difference between mean temperature change from 2020 to 2025(using hypothesis)

```
In [116]: from scipy import stats
#Null Hypothesis: Mean of prdicted mean temperature change is less than mean of Mean temperature change.
#Alternte Hypotheisi: Mean of predicted mean temperature change is greater than Mean of mean temperature change .

temperature1 =df["2020"]
temperature2 =df["2025"]


t, p = stats.ttest_ind(temperature1, temperature2)

# Print results
print("t-value: ", t)
print("p-value: ", p)

# Interpret results
alpha = 0.05
if p > alpha:
    print("Accept null hypothesis")
else:
    print("Reject null hypothesis")
```

```
t-value:  0.36864516896752636
p-value:  0.7135721143331109
Accept null hypothesis
```

Objective:

Analyse and prediction on deforestation in India from 2001 to 2020 and ranking of state.

Problem statements:

1) To predict the deforestation in year 2021-2025(regression).
2) Implementing hypothesis on observed value of 2020 and predicted value of 2040.
3) To rank the states according to deforestation in 2020. (By ranking)

Methodology:

1) To predict the deforestation in year 2021-2040(regression):

Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables. It is a linear approach to model the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The goal of linear regression is to find the best-fitting straight line through the data points. The line is represented by an equation of the form Y = a + bX, where X is the independent variable and Y is the dependent variable. The coefficient 'a' is the Y-intercept and 'b' is the slope of the line. Linear regression can be used for both simple and multiple regression analysis.

$$y = \theta_1 + \theta_2 . x$$

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

2) Implementing hypothesis on observed value of 2020 and predicted value of 2040:

Hypothesis:

A hypothesis is a statement or set of statements that describe an expected or predicted relationship between variables. It is an educated guess or prediction about the relationship between variables based on previous knowledge or observation. In scientific research, a hypothesis is a proposed explanation for an observable phenomenon. It is typically a testable statement that can be proven or disproven through experimentation or observation. Hypotheses are important in scientific research as they provide a starting point for further investigation, and they help to guide the design of experiments and the collection of data. Hypotheses can be either null (no relationship) or alternative (there is a relationship).

The general form of a null hypothesis is: $H0: \mu = \mu0$ Where H0 represents the null hypothesis, μ represents the population mean, and μ0 represents the hypothesized value of the population mean.

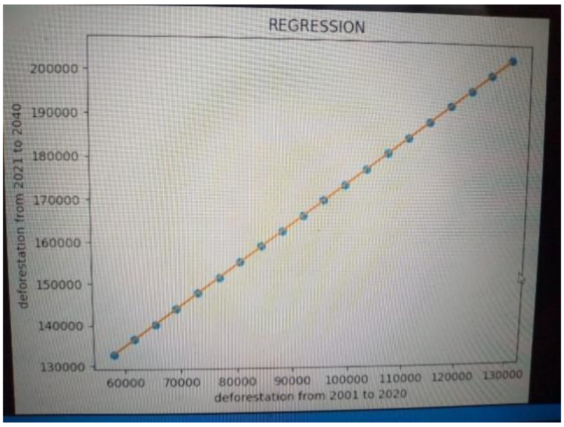
The general form of an alternative hypothesis is: $H1: \mu \neq \mu0$ or $H1: \mu < \mu0$ or $H1: \mu > \mu0$ Where H1 represents the alternative hypothesis, μ represents the population mean, and μ0 represents the hypothesized value of the population mean.

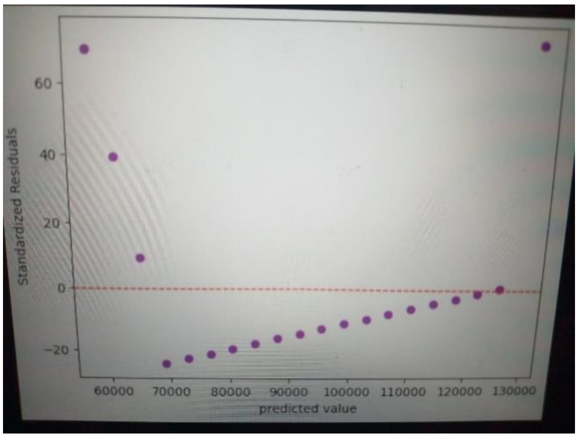
3) To rank the states according to deforestation in 2020. (By ranking)

Ranking in Python can be done using the ‘rankdata()’ function from the ‘scipy.stats’ module. This function assigns ranks to the elements of an array, with the highest element receiving rank 1, the second highest receiving rank 2, and so on.

Step(1): To predict the deforestation in year 2021-2040(regression):

Firstly, I implement the regression on dataset, then predict the value from year 2021 to 2040 and plot the graph on it.





Step(2): Implementing hypothesis on observed value of 2020 and predicted value of 2040:

Then, I implement the hypothesis on it and apply T-Test on it.

And claimed that the hypothesis is alternative.

Step(3): To rank the states according to deforestation in 2020. (By ranking)

Then have taken a new dataset and apply rank method on it. By python inbuilt function rank(), it sorted the dataset and I got the highest deforested state.

Result:

As a result, I conclude that the deforestation will increase from 2021 to 2040.

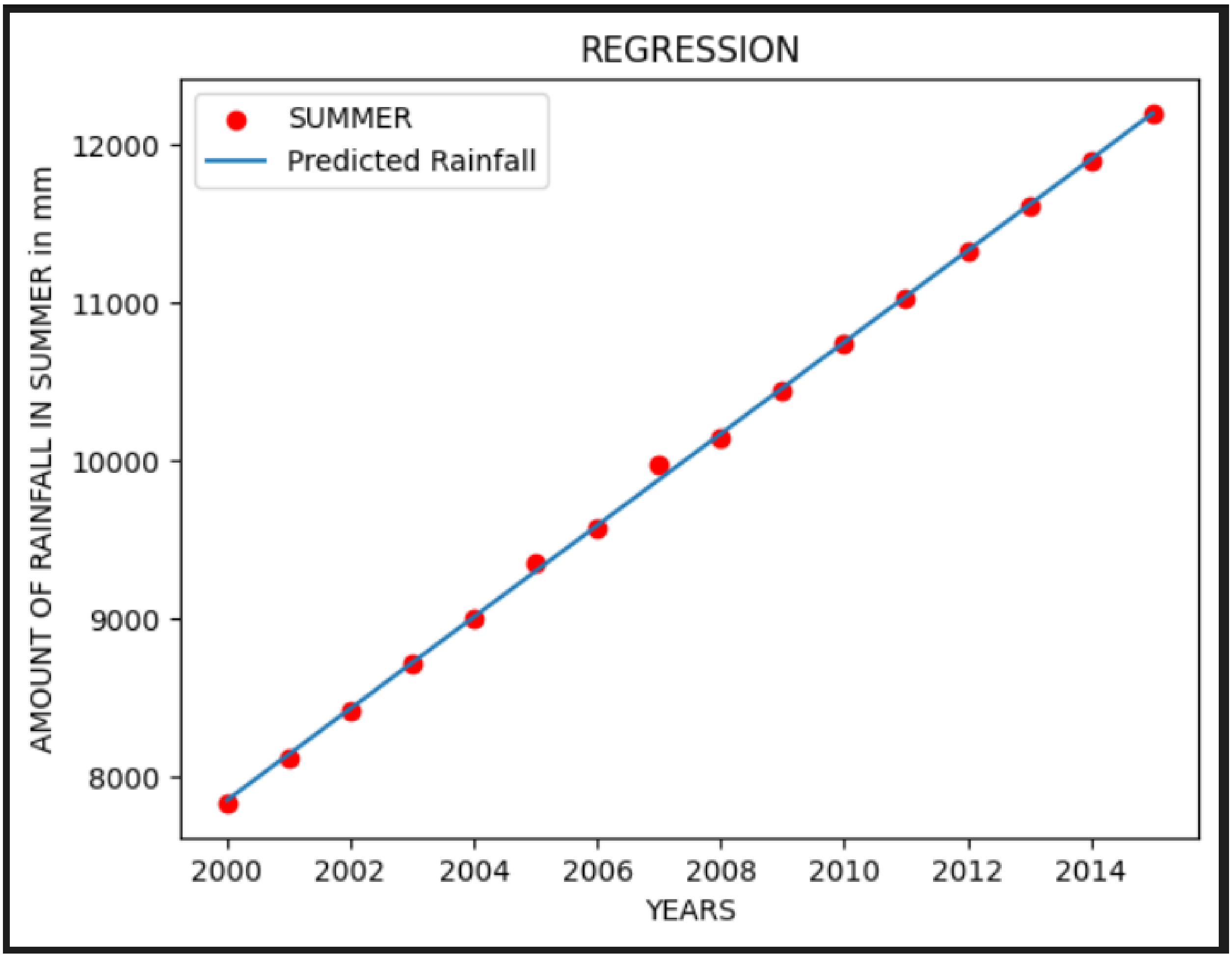
Objective 2:

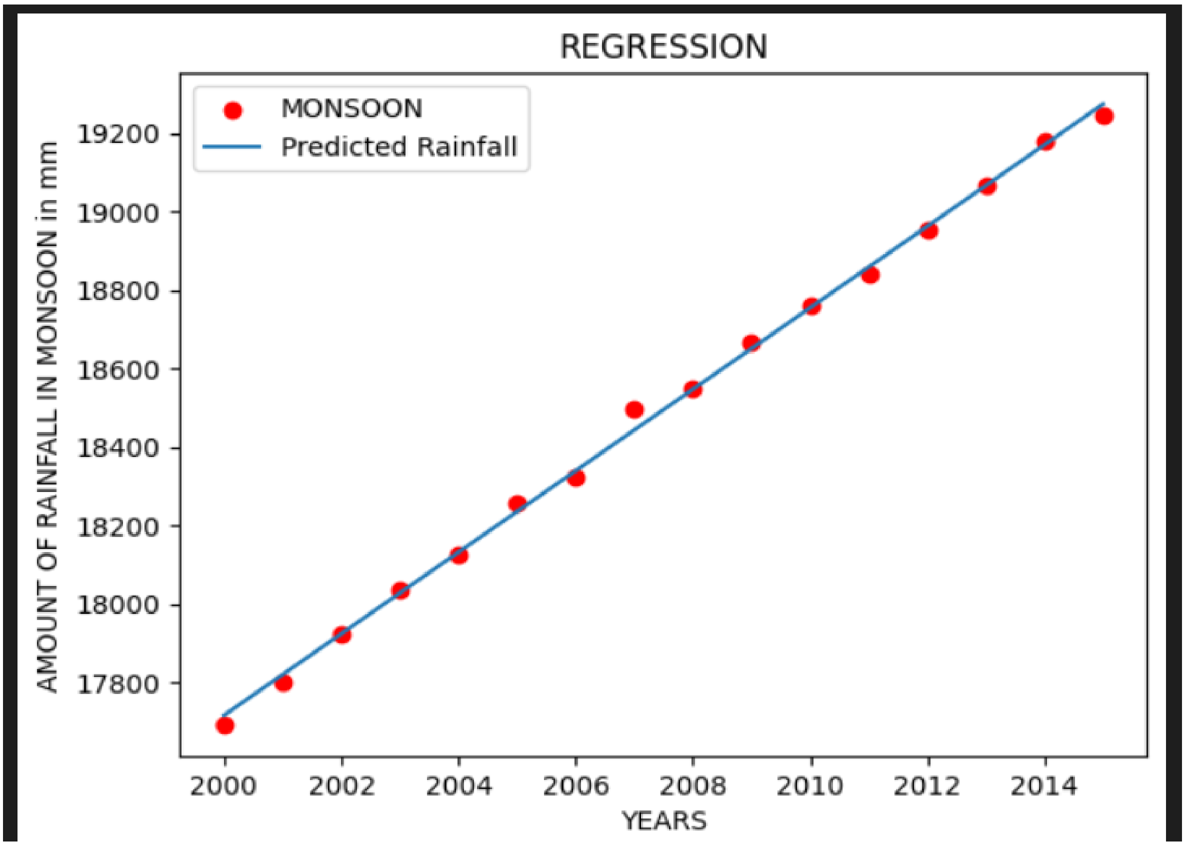
To Analyse the change in seasonal rainfall [like rain in Winter (January-February), Summer (May-June), Monsoon(April-September)] in India due to climate change.

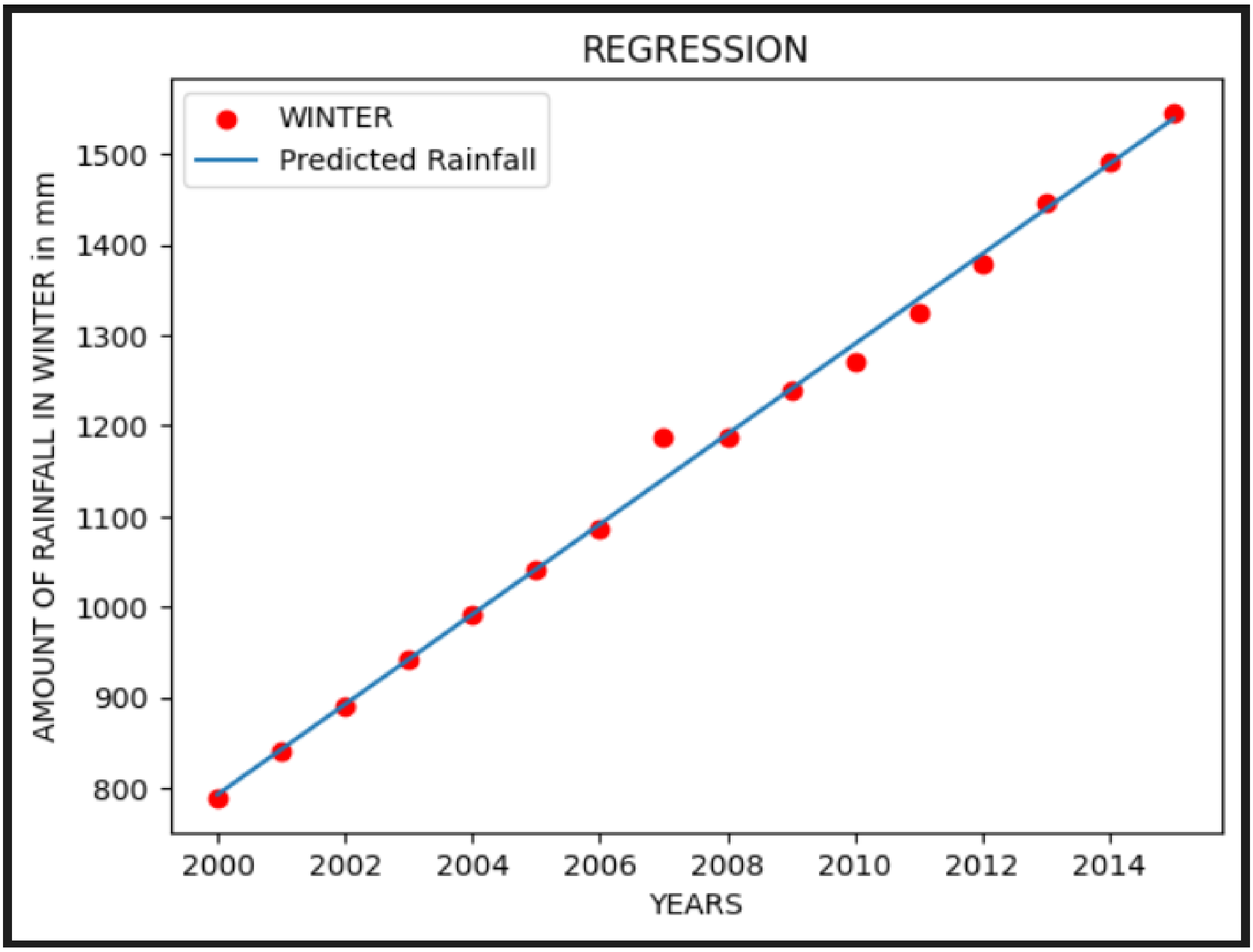
country	YEAR	SUMMER	MONSOON	WINTER
India	2000	7838.05	17693.7	789.2
India	2001	8128.3	17802.1	840
India	2002	8418.6	17925.7	891
India	2003	8708.8	18035.1	942.5
India	2004	8999.1	18127.8	990.9
India	2005	9359.83	18259.7	1042
India	2006	9579.7	18325.7	1085.4
India	2007			
India	2008	10146.01	18549.5	1186.9
India	2009	10450.5	18664.6	1239.3
India	2010	10740.8	18759.7	1271.6
India	2011	11031.1	18840.6	1325.8
India	2012	11321.3	18955.7	1379.1
India	2013	11611.6	19067.69	1446.5
India	2014	11901.9	19180.3	1492.19
India	2015	12192.2	19242.8	1545.6

Problem statement 2.1:

To predict the amount of seasonal rainfall in 2016-2031 using data from 2000-2015 of India. (Regression)


REGRESSION
SUMMER
Predicted Rainfall
AMOUNT OF RAINFALL IN SUMMER in mm
12000
11000
10000
9000
8000
2000
2002
2004
2006
2008
2010
2012
2014
YEARS


REGRESSION
MONSOON
Predicted Rainfall
AMOUNT OF RAINFALL IN MONSOON in mm
19200
19000
18800
18600
18400
18200
18000
17800
2000
2002
2004
2006
2008
2010
2012
2014
YEARS



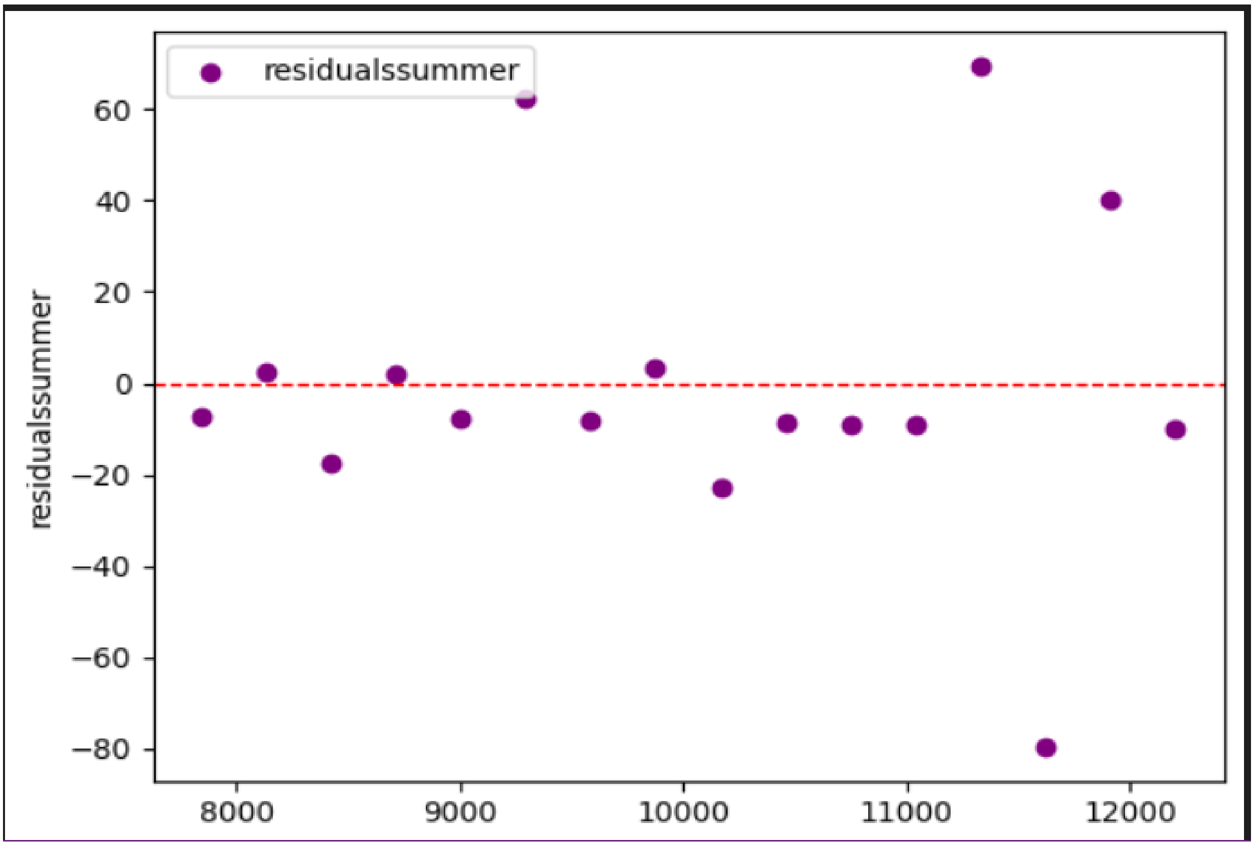
SUMMER	MONSOON	WINTER	predictsummer2000_2015	predictsummer2016_2031	predictmonsoon2000_2015	predictmonsoon2016_2031	predictwinter2000_2015	predictwinter2016_2031
2000	7838.05	17693.700000	789.20	7845.218382	12495.185000	17717.135098	19377.389467	792.933897
2001	8138.30	17802.100000	840.00	8135.676015	12784.517059	17820.900996	19481.155365	842.666794
2002	8408.60	17925.700000	891.00	8426.133647	13073.849118	17924.666894	19584.921263	892.399691
2003	8718.80	18035.100000	942.50	8716.591279	13363.181176	18028.432792	19688.687161	942.132588
2004	8999.10	18127.800000	990.90	9007.048912	13652.513235	18132.198690	19792.453059	991.865485
2005	9359.83	18259.700000	1042.00	9297.506544	13941.845294	18235.964588	19896.218957	1041.598382
2006	9579.70	18325.700000	1085.40	9587.964176	14231.177353	18339.730486	19999.984855	1091.331279
2007	9881.62	18495.379333	1186.90	9878.421809	14520.509412	18443.496384	20103.750753	1141.064176
2008	10146.01	18549.500000	1186.90	10168.879441	14809.841471	18547.262282	20207.516651	1190.797074
2009	10450.50	18664.600000	1239.30	10459.337074	15099.173529	18651.028180	20311.282549	1240.529971

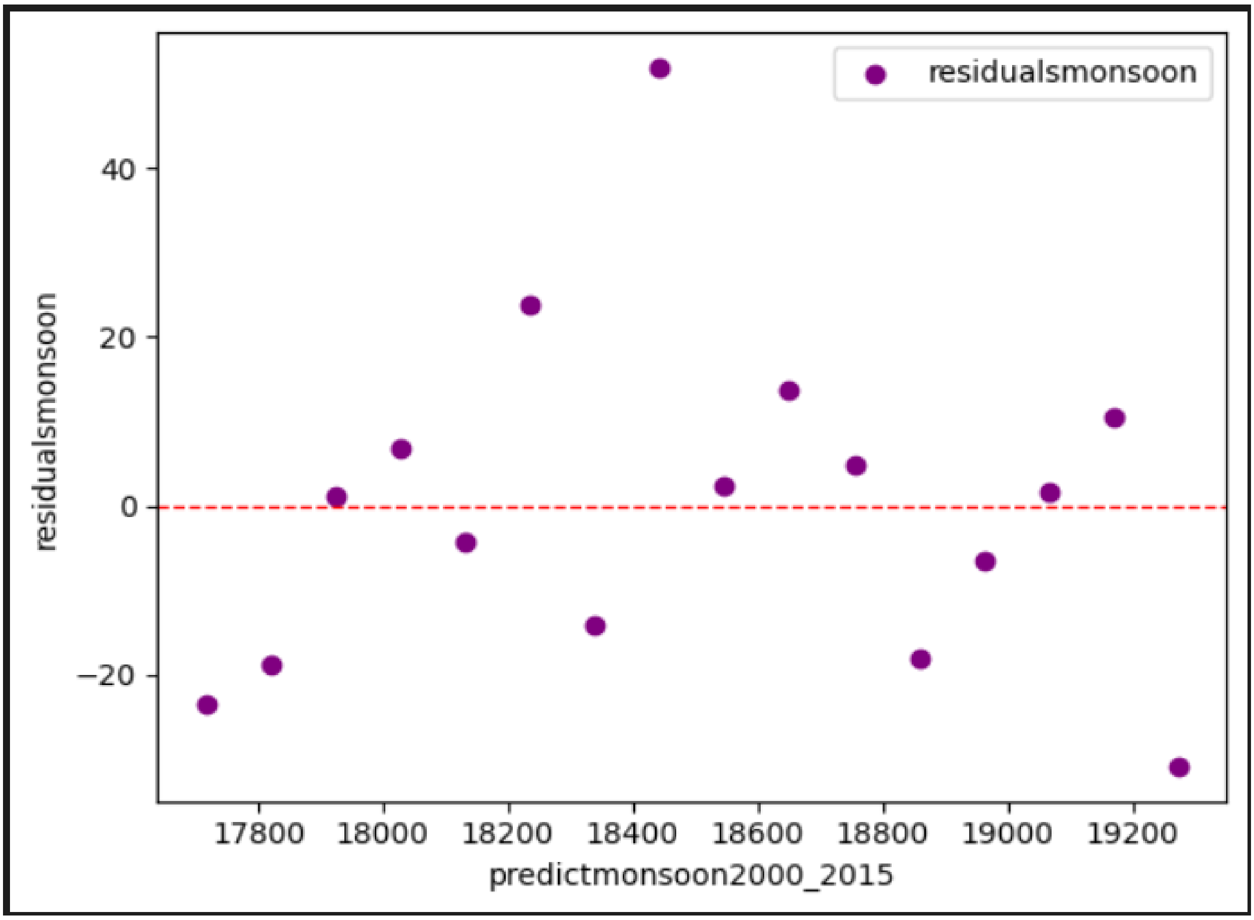
SUMMER	MONSOON	WINTER	predictsummer2000_2015	predictsummer2016_2031	predictmonsoon2000_2015	predictmonsoon2016_2031	predictwinter2000_2015	predictwinter2016_2031
2010	10740.80	18759.700000	1271.60	10749.794706	15388.505588	18754.794078	20415.048447	1290.262868
2011	11031.10	18840.600000	1325.80	11040.252338	15677.837647	18858.559976	20518.814345	1339.995765
2012	11400.30	18955.700000	1379.10	11330.709971	15967.169706	18962.325875	20622.580243	1389.728662
2013	11541.60	19067.690000	1446.50	11621.167603	16256.501765	19066.091773	20726.346141	1439.461559
2014	11951.90	19180.3000	1492.19	11911.625235	16545.833824	19169.857671	20830.112039	1489.194456
2015	12192.20	19242.800000	1545.60	12202.082868	16835.165882	19273.623569	20933.877937	1538.927353

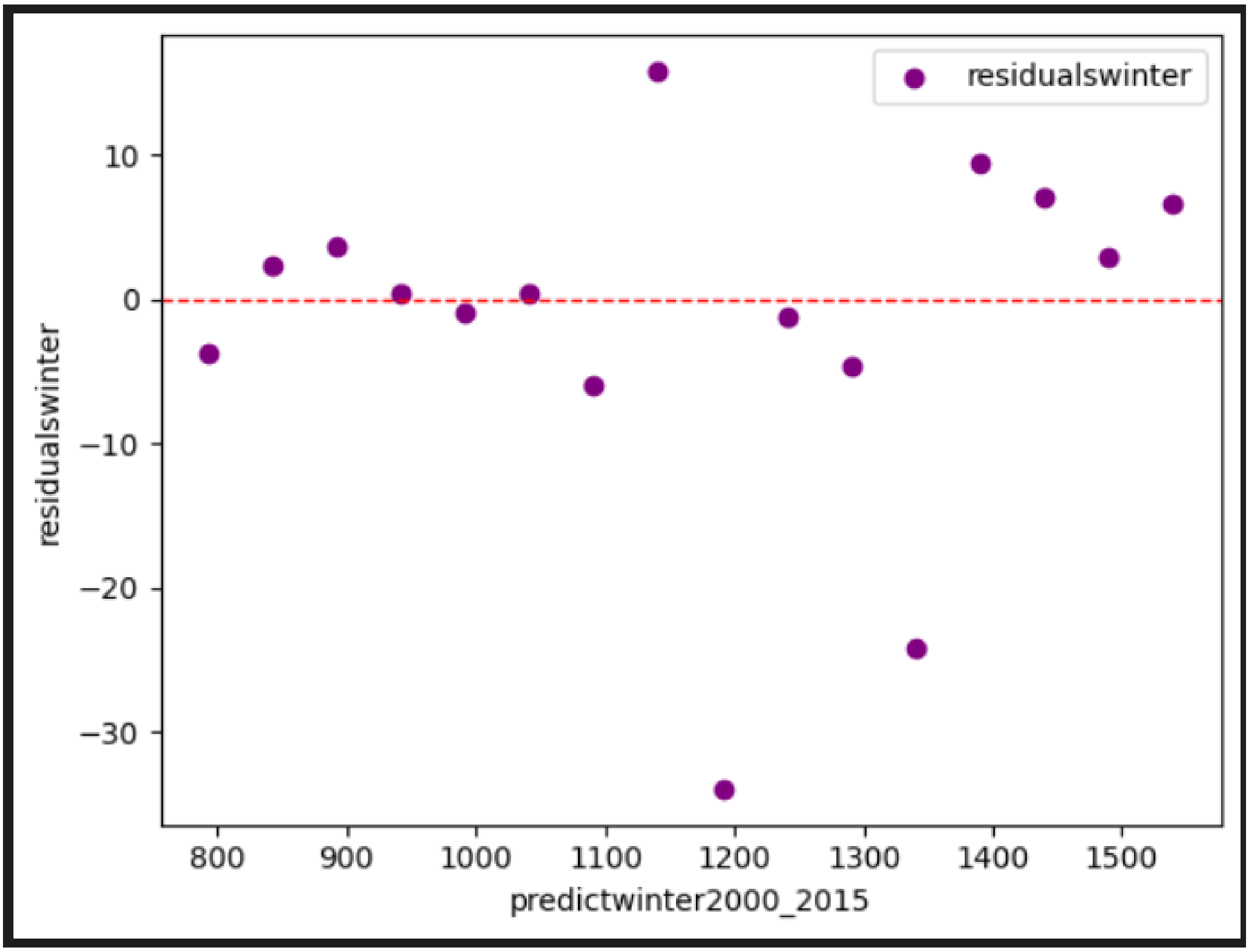
- In my first step i plot the graph of regression to check the fitness of my dataset.
- Then do the prediction on rainfall in india from 2016-2031 by using dataset of years 2000-2015 by regression.
- I find R_square is 0.9965127121037636 by method the ratio of ssres/sstot then 1-ssres.
- SSres which is sum square of residual.
- SStot which is total sum of squares.

residualssummer	residuals_squaredsummer	residualsmonsoon	residuals_squaredmonsoon	residuals squared	residualswinter	residuals_squaredwinter	residuals_squaredwinter
1588.660250	-7.168382	-376.730625	-23.435098	549.203820	642689.753494	-3.733897	141925.963813
1638.393147	2.623985	-325.930625	-18.800996	353.477454	480636.234027	-2.666794	106230.772313
1688.126044	-17.533647	-274.930625	1.033106	1.067308	324534.542827	-1.399691	75586.848563
1737.858941	2.208721	-223.430625	6.667208	44.451660	211857.064694	0.367412	49921.244188
1787.591838	-7.948912	-175.030625	-4.398690	19.348475	135114.566294	-0.965485	30635.719688
1837.324735	62.323456	-123.930625	23.735412	563.369772	55544.748160	0.401618	15358.799813
1887.057632	-8.264176	-80.530625	-14.030486	196.854545	28791.076160	-5.931279	6485.181563

residualssummer	residuals_squaredsummer	residualsmonsoon	residuals_squaredmonsoon	residuals squared	residualswinter	residuals_squaredwinter	residuals_squaredwinter
1936.790529	3.198191	20.969375	51.882949	2691.840399	0.000000	45.835824	439.714688
1986.523426	-22.869441	20.969375	2.237718	5.007380	2929.046560	-3.897074	439.714688
2036.256324	-8.837074	73.369375	13.571820	184.194287	28635.634027	-1.229971	5383.065188
2085.989221	-8.994706	105.669375	4.905922	24.068066	69865.414827	-18.662868	11166.016813
2135.722118	-9.152338	159.869375	-17.959976	322.560755	119177.308694	-14.195765	25558.217063
2185.455015	69.590029	213.169375	-6.625875	43.902213	211895.116160	-10.628662	45441.182438
2235.187912	-79.567603	280.569375	1.598227	2.554331	327539.499180	7.038441	78719.174188
2284.920809	40.274765	326.259375	10.442329	109.042244	469116.319627	2.995544	106445.179775
2334.653706	-9.882868	379.669375	-30.823569	950.092383	558637.652960	6.672647	144148.834313




residualsmonsoon
40
20
0
−20
residualsmonsoon
17800
18000
18200
18400
18600
18800
19000
19200
predictmonsoon2000_2015

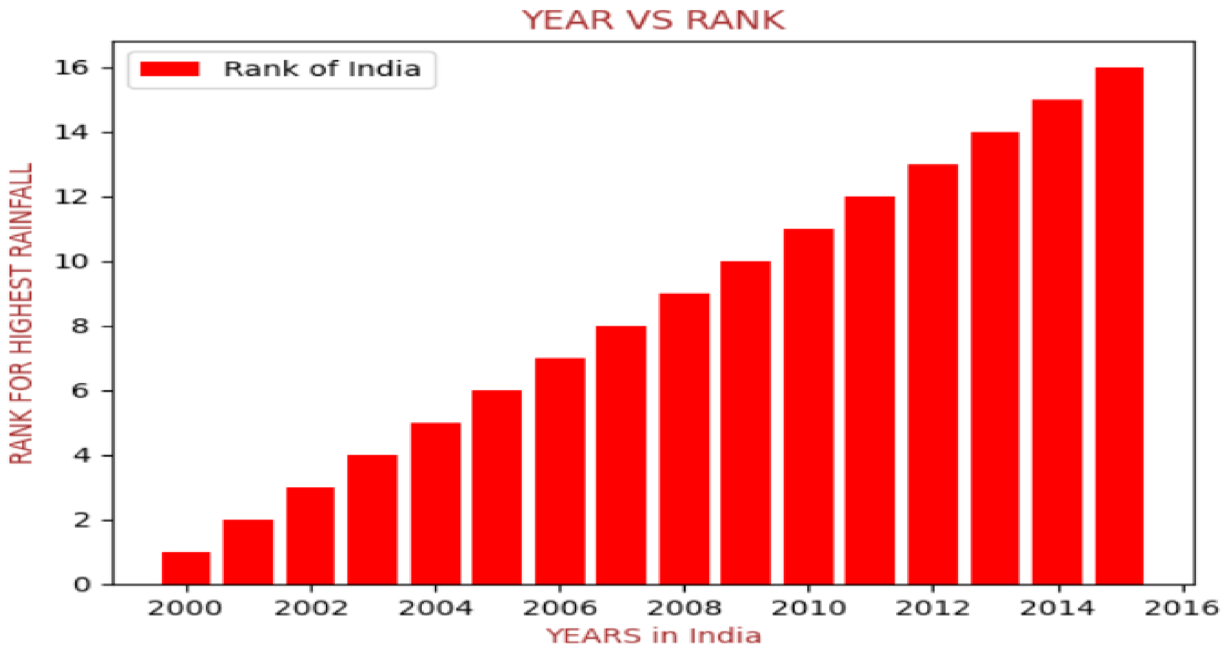

residualswinter
10
0
−10
−20
−30
residualswinter
800
900
1000
1100
1200
1300
1400
1500
predictwinter2000_2015

Problem statement 2.2:

To rank the India based on seasonal rainfall in years 2000-2015 and 2016-2031.

SUMMER	MONSOON	WINTER	predictsummer2000_2015	predictsummer2016_2031	predictmonsoon2000_2015	predictmonsoon2016_2031	predictwinter2000_2015	predictwinter2016_2031	rank
2000	7838.05	17693.700000	789.20	7845.218382	12495.185000	17717.135098	19377.389467	792.933897	1
2001	8138.30	17802.100000	840.00	8135.676015	12784.517059	17820.900996	19481.155365	842.666794	2
2002	8408.60	17925.700000	891.00	8426.133647	13073.849118	17924.666894	19584.921263	892.399691	3
2003	8718.80	18035.100000	942.50	8716.591279	13363.181176	18028.432792	19688.687161	942.132588	4
2004	8999.10	18127.800000	990.90	9007.048912	13652.513235	18132.198690	19792.453059	991.865485	5
2005	9359.83	18259.700000	1042.00	9297.506544	13941.845294	18235.964588	19896.218957	1041.598382	6
2006	9579.70	18325.700000	1085.40	9587.964176	14231.177353	18339.730486	19999.984855	1091.331279	7
2007	9881.62	18495.379333	1186.90	9878.421809	14520.509412	18443.496384	20103.750753	1141.064176	8
2008	10146.01	18549.500000	1186.90	10168.879441	14809.841471	18547.262282	20207.516651	1190.797074	9
2009	10450.50	18664.600000	1239.30	10459.337074	15099.173529	18651.028180	20311.282549	1240.529971	10
2010	10740.80	18759.700000	1271.60	10749.794706	15388.505588	18754.794078	20415.048447	1290.262868	11
2011	11031.10	18840.600000	1325.80	11040.252338	15677.837647	18858.559976	20518.814345	1339.995765	12
2012	11400.30	18955.700000	1379.10	11330.709971	15967.169706	18962.325875	20622.580243	1389.728662	13
2013	11541.60	19067.690000	1446.50	11621.167603	16256.501765	19066.091773	20726.346141	1439.461559	14
2014	11951.90	19180.3000	1492.19	11911.625235	16545.833824	19169.857671	20830.112039	1489.194456	15
2015	1219	19242.	1545.60	12202.08286	16835.165882	19273.62356	20933.8779	1538.92	16

SUMMER	MONSOON	WINTER	predictsummer2000_2015	predictsummer2016_2031	predictmonsoon2000_2015	predictmonsoon2016_2031	predictwinter2000_2015	predictwinter2016_2031	rank
	2.20	800000		8		9	37	7353	



- In my second problem statement I perform ranking of India according to seasonal rainfall so I use .rank() function.

<u>Problem statement 2.3:</u>

To check claim the increment in seasonal rainfall from 2016-2031 by given data 2016-2031.

Step1.

1.Find the mean of given data according to seasons from2000-2015 and predicted data from 2016-2031.

2. let u1 and u2 be the mean of data according to seasons from 2000-2015 and predicted data from 2016-2031.

3.let s1 and s2 be the standard deviation of mean of data according to seasons from 2000-2015 and predicted data from 2016-2031.

4.Null hypothesis H0=u1<=u2.

5.Alternate Hypothesis (Ha)=u1>u2.

Step 2

1.u1-u2>=0, therefore we going to left side tailed test

2.level of significance=0.01

3. •Degree of freedom(df)=n -1=4

4. •Now as our sample size i.e., n < 30 and the population variance is unknown, so we will use t test to find the difference of mean.•

5.On calculating through t distribution table for the given value of significance and

df, the value $t\alpha$ is found to be 3.25.

step3.

calculating through python

The t value is 529.1796871414085 for summer

The t value is 330.37475438677785 for monsoon

The t value is 261.82330778861495 for winter

step 4

criteria of rejection

Since, the null hypothesis is rejected for all three, through the criteria of left-tailed test, and the alternative hypothesis is accepted for all three.

Step5:

Conclusion

From the test of hypothesis, we can conclude that our null hypothesis is rejected, and the impact of climate change on seasonal rainfall is that the rainfall increases day by day.

Objective 4:

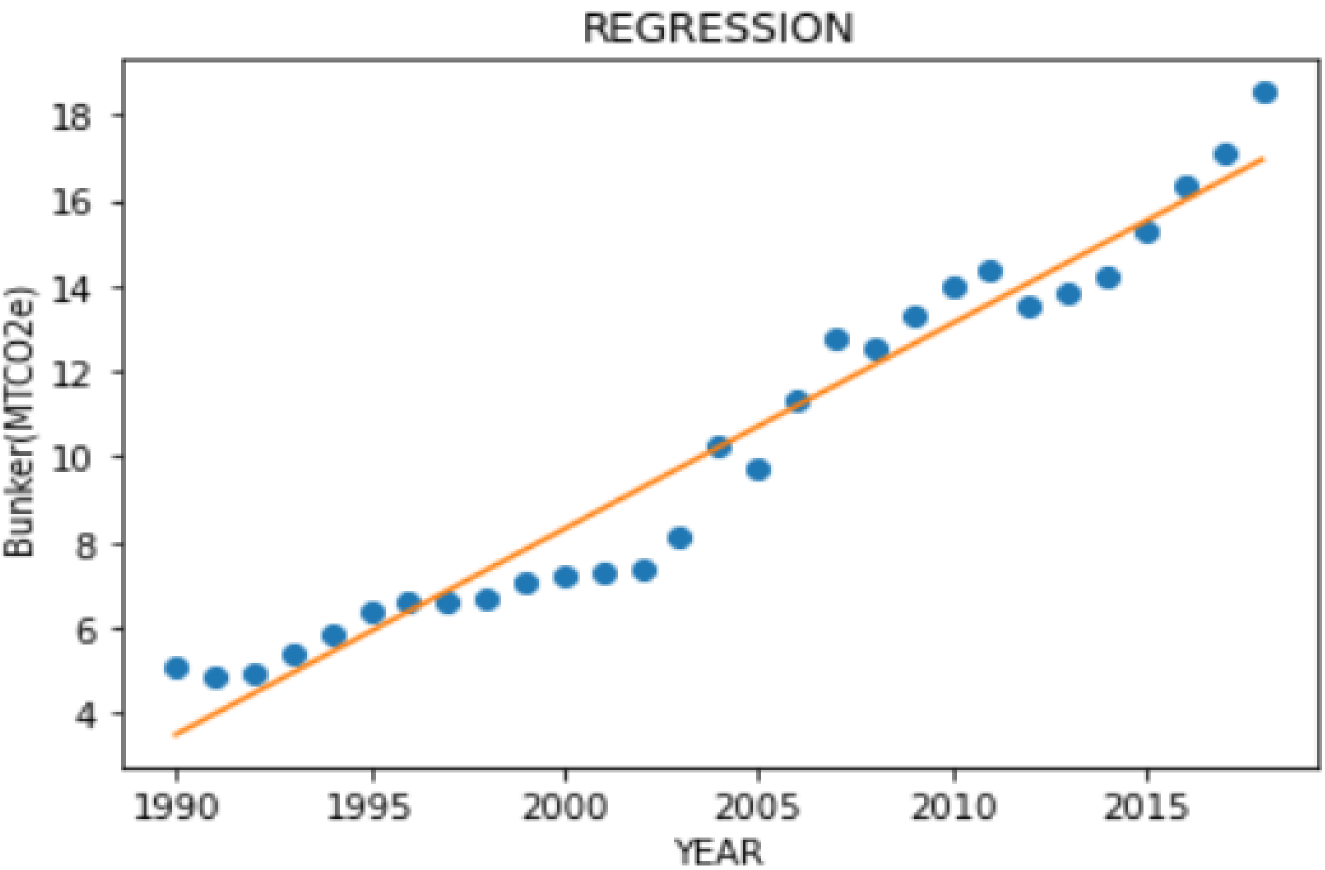
Problem 4.1:

To predict the amount of Green House Gas emission from 2022 to 2025, by historical data from 1990 to 2019 (By using regression analyses).

Step1: A graph of a linear regression model was plotted firstly which is between Bunker ($MtCO_2e$) and years.

Here, Bunker ($MtCO_2e$) is dependent Variable

Year is independent Variable



- *Figure 4: Linear Regression Graph b/w Bunker ($MTCO_2e$) and year*

Step 2: Predict the value from 2020-2025 by using linear regression equation.

Regression equation: $Y = a + bx$

Problem Statement 4.2:

To determine the rank of all the Sectors, based on the amount of Green House Gases emission from 1990 to 2019 (Power Method).

Rank by Power method:

sector	Agriculture	Industrial Process	Bunker fuel	waste	Bunker fuel.1	manufacture and construction	transportation	Fugative emission	Other
Agriculture	0	2	2	2	1	1	1	2	1
Industrial Process	1	0	2	2	1	1	1	2	1
Bunker fuel	1	1	0	1	1	1	1	1	1
waste	1	1	2	0	1	1	1	2	2
Bunker fuel	2	2	2	2	0	1	2	2	2
manufacture and construction	2	2	2	2	2	0	2	2	2
transportation	1	2	2	2	1	1	0	2	2
Fugative emission	1	1	2	1	1	1	1	0	1
Other	1	2	2	2	1	1	1	2	0

Do iteration 10 times:

```
Dominant eigenvector: [0.32586751 0.30158784 0.23739963 0.30339686 0.40668255 0.439423
 0.35210185 0.25651175 0.32586751]
Dominant eigenvalue: 11.421411796889661
```

Conclusion:

Rank 1: manufacture and construction

Rank 2: Bunker fuel

Rank 3: transportation

Rank 4: Other

Rank 5: Agriculture

Rank 6: waste

Rank 7: Industrial Process

Rank 8: Fugative emission

Rank 9 Bu

Problem 4.3:

To check the claim that contribution of all the Sectors in releasing Greenhouse gas is not same in INDIA.

ANNOVA Hypothesis:

Step1: Hypothesis Formulation

Let the region 1,2, 3...9 mean value is μ1, μ2, μ3…μ9. then the Null hypothesis and Alternative hypothesis are as follow.

Null Hypothesis:

μ1=μ2=μ3…μ8=μ9 = k

If there is no significant difference of all population mean.

Alternative hypothesis:

μ1≠μ2≠μ3…μ8≠μ9 =k

Not all μ1 are equal to k

Step 2: Calculate the squared sum difference between the sample means. (SSB)

SSB:

$$SSW = \sum_{j=1}^{k}\sum_{j=1}^{l}(X-\overline{X}_j)^2$$

Python calculation,

SSB= 20748208.896820296

Step 3: Calculate the sum of squares of difference between the sample. (SSW)

SSW:

$$SSB = \sum_{j=1}^{k} (\bar{X}_j - \bar{X})^2$$

Python calculation,
SSW= 6079933.586357133

Step 4: calculate the total square sum deviation. (SST)

SST= SSB+SSW

Python Calculation,
SST= 26828142.48317743

Step 5: calculate the Mean Squares of SSB:

MSB=SSB/df

Python Calculation,
MSB= 2593526.112102537

Step 6: calculate the Mean squares of SSW:

MSW= SSW/df

Python calculation,
MMW= 319996.5045451123

Step 7: Make ANOVA table:

Put the value in table which is obtained before steps (2 to 4), ANOVA table as shown below.

Variation Source	Sum of Squares	MEAN Square	F Value
SSB	20748208.896820296	2593526.112102537	8.104857632083629
SSW	6079933.586357133	319996.5045451123	
SST	26828142.48317743		

Step 8: Accept and Reject the Null Hypothesis by the analysis.

The critical value is 3.52

The F value is greater than the critical value and not lie in Critical region.

Since, we reject the null hypothesis and accept the alternative hypothesis

Reference

https://link.springer.com/article/10.1007/s10531-015-1033-2

https://www.tandfonline.com/doi/abs/10.1177/0971852412459428

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Analysing+the+gross+and+net+deforestation+rates+in+India%3A&btnG=

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=ipcc+climate+action&btnG==

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Earth+observations+for+estimating+greenhouse+gas+emissions+from+deforestation+in+developing+countries%3A&btnG=

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Analysing+deforestation+rates%2C+spatial+forest+cover+changes+and+identifying+critical+areas+of+forest+cover+changes+in+North-East+India+during+1972%E2%80%931999%3A&btnG=

https://www.sciencedirect.com/science/article/abs/pii/S2352938519301880

https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG027i001p00115

https://www.science.org/doi/abs/10.1126/science.248.4960.1217

https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/RG027i001p00115

https://www.jstor.org/stable/2233863

https://journals.sagepub.com/doi/abs/10.1177/030913339001400101

Appendix

Appendix 1:

- Appendices 1.1

```
from sklearn.linear_model import LinearRegression
regr=LinearRegression()
y = df['2020']
x = df['2019']
Sxx = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
Sxx
Syy = np.sum(y**2) - ((np.sum(y)**2)/y.shape[0])
Syy
Sxy = np.sum(y*x)-((np.sum(x)*np.sum(y))/y.shape[0])
Sxy
corr_xy = Sxy/np.sqrt(Sxx*Syy)
corr_xy
b_xy = Sxy/Syy
b_xy
b_yx=Sxy/Sxx
b_yx
a=np.mean(y)-b_xy*np.mean(x)
a
y=a+b_xy*x
y
residual=y-df['2020']
residual
residuals_squared = residual ** 2
ssres = residuals_squared.sum()
ssres
y_mean = df['2020'].mean()
```

```
# create a column for the residuals
residuals_of_Mean = df['2020'] - y_mean

# square the residuals
residuals_sq = residuals_of_Mean ** 2

# calculate the total sum of squares
sstot = residuals_sq.sum()
print(sstot)
R_square=1-ssres/sstot
R_square
plt.scatter(y,residual,color='purple')
plt.xlabel('predicted Year 2020')
plt.ylabel('Standardized Residuals')
plt.axhline(y=0, color='red', linestyle='--', linewidth=1)
plt.show()
r=(b_xy*b_yx)**0.5
r
plt.plot(x, y, 'o')

#obtain b_xy(slope) and a(intercept) of linear regression line
b_xy, a = np.polyfit(x, y, 1)

#add linear regression line to scatterplot
plt.plot(x,b_xy*x+a)
plt.xlabel('Independent variable(2019)',size=10)
plt.ylabel('dependent variable(2020)',size=10)
y2=a+b_xy*df['2020']
df['2021']=y2
y2
y3=a+b_xy*df['2021']
```

```
df['2022']=y3
y3
y4=a+b_xy*df['2022']
df['2023']=y4
y4
y5=a+b_xy*df['2023']
df['2024']=y5
y5
y6=a+b_xy*df['2024']
df['2025']=y6
y6
df
plt.rcParams['figure.figsize']=(30,10)
col1='steelblue'
col2='red'
#define the subplot here
fig,ax=plt.subplots()
ax.bar(df['States'],df['2021'],color=col1)
#add the x label
ax.set_xlabel('States Name',fontsize=20)
# rotate of x axis by Xticks
plt.xticks(rotation=90,size=20)
plt.yticks(size=15)
#add the y lable
ax.set_ylabel('Mean temperature change',color=col1,fontsize=20)
plt.title('Mean temperature change in year 2021 Statewise',fontsize=30)
plt.show()
plt.rcParams['figure.figsize']=(30,10)
col1='steelblue'
col2='red'
#define the subplot here
fig,ax=plt.subplots()
```

```
ax.bar(df['States'],df['2025'],color=col1)
#add the x label
ax.set_xlabel('States Name',fontsize=20)
# rotate of x axis by Xticks
plt.xticks(rotation=90,size=20)
plt.yticks(size=15)
#add the y lable
ax.set_ylabel('mean temperature change',color=col1,fontsize=20)
plt.title('Mean temperature change in year 2025 Statewise',fontsize=30)
plt.show()
```

- Appendices 1.2

```
df['rank']=df['2020'].rank(method="min",ascending=False)
df.sort_values('rank',inplace=True)
df
plt.rcParams['figure.figsize']=(30,10)
col1='steelblue'
col2='red'
#define the subplot here
fig,ax=plt.subplots()
ax.bar(df['States'],df['2020'],color=col1)
#add the x label
ax.set_xlabel('States Name',fontsize=20)
# rotate of x axis by Xticks
plt.xticks(rotation=90,size=20)
plt.yticks(size=15)
#add the y lable
ax.set_ylabel('mean temperature change',color=col1,fontsize=20)
plt.title('Mean temperature change in year 2020 Statewise',fontsize=30)
plt.show()
```

```
df['rank']=df['2025'].rank(method="min",ascending=False)

df.sort_values('rank',inplace=True)

df

plt.rcParams['figure.figsize']=(30,10)

col1='steelblue'

col2='red'

#define the subplot here

fig,ax=plt.subplots()

ax.bar(df['States'],df['2025'],color=col1)

#add the x label

ax.set_xlabel('States Name',fontsize=20)

# rotate of x axis by Xticks

plt.xticks(rotation=90,size=20)

plt.yticks(size=15)

#add the y lable

ax.set_ylabel('mean temperature change',color=col1,fontsize=20)


plt.title('Mean temperature change in year 2025 Statewise',fontsize=30)

plt.show()
```

- Appendices problem 1.3

```
from scipy import stats

#Null Hypothesis: Mean of prdicted mean temperature change is less than mean of Mean temperature change.

#Alternte Hypotheisi: Mean of predicted mean temperature change is greater than Mean of mean temperature change .


temperature1 =df["2020"]

temperature2 =df["2025"]
```

```
t, p = stats.ttest_ind(temperature1, temperature2)


# Print results
print("t-value: ", t)
print("p-value: ", p)


# Interpret results
alpha = 0.05
if p > alpha:
   print("Accept null hypothesis")
else:
   print("Reject null hypothesis")


print(df['2020'].mean())
print(df['2025'].mean())
```

Appendix 2:

Prediction

```
y=df["SUMMER"]
x=df["YEAR"]
Sxx=np.sum(x**2)-((np.sum(x)**2/x.shape[0]))
Syy=np.sum(y**2)-((np.sum(y)**2)/y.shape[0])
Sxy=np.sum(y*x)-(np.sum(x)*np.sum(y))/y.shape[0]
b_yx = Sxy/Sxx
a = np.mean(y)-b_yx*np.mean(x)
print(b_yx)
print(a)
y1=(a+b_yx*2000)
y2=(a+b_yx*2001)
y3=(a+b_yx*2002)
```

```
y4=(a+b_yx*2003)
y5=(a+b_yx*2004)
y6=(a+b_yx*2005)
y7=(a+b_yx*2006)
y8=(a+b_yx*2007)
y9=(a+b_yx*2008)
y10=(a+b_yx*2009)
y11=(a+b_yx*2010)
y12=(a+b_yx*2011)
y13=(a+b_yx*2012)
y14=(a+b_yx*2013)
y15=(a+b_yx*2014)
y16=(a+b_yx*2015)
y1,y2,y3,y4,y5,y6,y7,y8,y9,y10,y11,y12,y13,y14,y15,y16
import matplotlib.pyplot as plt
sum=[y1,y2,y3,y4,y5,y6,y7,y8,y9,y10,y11,y12,y13,y14,y15,y16]
ye=[2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015]
plt.scatter(df["YEAR"],df["SUMMER"],color="red")
plt.plot(ye,sum)
plt.legend(['SUMMER','Predicted Rainfall'])
plt.title('REGRESSION')
plt.xlabel('YEARS')
plt.ylabel("AMOUNT OF RAINFALL IN SUMMER in mm")
plt.show()
for i in range(2016,2032):
    predictsummer2016_2030=a+b_yx*i
    print(predictsummer2016_2030)
```

Residual

```
df['residualssummer']=df['SUMMER']-df["predictsummer2000_2015"]
df['residualssummer']
df['residuals_squaredsummer'] = df['residualssummer']**2
ssres = df['residuals_squaredsummer'].sum()
print("ssres=",ssres)
y_mean = df['SUMMER'].mean()
df['residuals B']=df['SUMMER']-y_mean
df['residuals_squaredsummer']=df['residuals B']**2
sstot = df['residuals_squaredsummer'].sum()
print("sstot=",sstot)
R_square=1-ssres/sstot
print("R_square=",R_square)
plt.scatter(df['predictsummer2000_2015'],df['residualssummer'],color='purple')
plt.xlabel('predictsummer2000_2015')
plt.ylabel('residualssummer')
plt.legend(['residualssummer'])
plt.axhline(y=0, color='red', linestyle='--', linewidth=1)
plt.show()
```

Ranking

```
df["rank"]=df["SUMMER"].rank()
df.sort_values('rank',inplace=True)
df
plt.bar(df["YEAR"],df["rank"],color="red")
plt.title('YEAR VS RANK',color="brown")
plt.xlabel('YEARS in India',color="brown")
plt.ylabel("RANK FOR HIGHEST RAINFALL",color="brown")
plt.legend(["Rank of India"])
plt.show()
```

Hypothesis

```
#Null Hypothesis: Mean of prdicted rainfall is less than Mean rainfall.
#Alternte Hypotheisi: Mean of predicted rainfall is greater than Mean of rainfall.
data_group1 = df['predictsummer2016_2031']
data_group2 = df['SUMMER']
t_stat,p_val= stats.ttest_rel(a=data_group1, b=data_group2)
print("t-statistics=",t_stat)
print("p-value",p_val)
#Level of significance = a
a=0.01
#Confidence Interval
(-3.25,3.25)
if t_stat >-3.35 and t_stat <3.25:
    print("Accept Null Hypothesis")
else:
    print("Accept Alternate Hypothesis")
```

Appendix 3.

Linear Regression code

```
x=df['loss per hactre']
y=df['loss_per_hectare2021-40']
plt.plot(x, y, 'o')
m, b = np.polyfit(x, y, 1)
plt.xlabel('deforestation from 2001 to 2020')
plt.ylabel('deforestation from 2021 to 2040')
plt.title('REGRESSION')
plt.plot(x, m*x+b)
plt.show()
```

Residual code:

```
df['residuals']=df['loss per hactre']-df['loss per hactare in 2001-2020']
df['residuals']
df['residuals_squared'] = df['residuals'] ** 2
ssres = df['residuals_squared'].sum()
```

```
print(ssres)
y_mean = df['loss per hactre'].mean()
df['residuals B'] = df['loss per hactre'] - y_mean
df['residuals_squared'] = df['residuals B'] ** 2
sstot = df['residuals_squared'].sum()
print(sstot)
R_square=1-ssres/sstot
R_square
```

Residual graph code

```
import matplotlib.pyplot as plt

plt.scatter(df['loss per hactare in 2001-2020'],df['residuals'],color='purple')
plt.xlabel('predicted value')
plt.ylabel('Standardized Residuals')
plt.axhline(y=0, color='red', linestyle='--', linewidth=1)
plt.show()
```

Hypothesis code

```
from scipy import stats
#Null Hypothesis: Mean of prdicted deforestation is less than Mean deforestation.
#Alternte Hypotheisi: Mean of predicted deforestation is greater than Mean of deforestation.
data_group1 = df['loss_per_hectare2021-40']
data_group2 = df['loss per hactre']
t_stat,p_val= stats.ttest_rel(a=data_group1, b=data_group2)
print("t-statistics=",t_stat)
#Level of significance = a
a=0.01
#Confidence Interval
(-3.25,3.25)
if t_stat >-3.35 and t_stat <3.25:
   print("Accept Null Hypothesis")
else:
```

```
    print("Accept Alternate Hypothesis")
```

Ranking code

```
d['rank']=d['tc_loss_ha_2020'].rank()
d.sort_values('rank',inplace=True)
d
plt.rcParams['figure.figsize']=(30,10)
col1='steelblue'
col2='red'
#define the subplot here
fig,ax=plt.subplots()
ax.bar(d['subnational1'],d['rank'],color=col1)
#add the x label
ax.set_xlabel('States Name',fontsize=20)
# rotate of x axis by Xticks
```

Ranking graph code:

```
plt.xticks(rotation=90,size=20)
plt.yticks(size=15)
#add the y lable
ax.set_ylabel('change in deforestation',color=col1,fontsize=20)
plt.title('Deforestation',fontsize=30)
plt.show()
```

Appendix 4:

4.1

1.

```
import numpy as np
y=df["Bunker (MTC2e)"]
 plt.plot(x, y, 'o')
#obtain m (slope) and b(intercept) of linear regression line
m, b = np.polyfit(x, y, 1)
plt.xlabel('YEAR')
plt.ylabel('Bunker(MTCO2e)')
```

```
plt.title('REGRESSION')
#add linear regression line to scatterplot
plt.plot(x, m*x+b)
plt.show()
```

2.

```
x=df["Year"]
Sxx=np.sum(x**2) -((np.sum(x)**2/x.shape [0]))
Syy=np.sum(y**2) -((np.sum(y)**2)/y.shape[0])
Sxy=np.sum(y*x) -(np.sum(x)*np.sum(y))/y.shape[0]
```

3.

```
b_yx = Sxy/Sxx
b_xy=Sxy/Syy
a = np.mean(y) - b_yx * np.mean(x)
print(b_yx)
```

4.

```
import matplotlib.pyplot as plt
plt.scatter(df['y_pre Bunker'],df['residuals'],color='purple')
plt.xlabel('predicted value')
plt.ylabel('Standardized Residuals')
plt.axhline(y=0, color='yellow', linestyle='--', linewidth=1)
plt.show()
```

4.2

1.

```
ranking_matrix=pd.read_csv('C:\\Users\\91820\\Documents\\green house matrix.csv')
ranking_matrix
```

2.

```
matrix=ranking_matrix.iloc[:,1:].to_numpy()
matrix
```

3.

```
import numpy as np


def power_method(A, x0, num_iterations):
```

```
    x = x0
    for i in range(num_iterations):
        x_new = np.dot(A, x)
        lambda_new = np.dot(x_new, x) / np.dot(x, x)
        x = x_new / np.linalg.norm(x_new)
    return x, lambda_new
A = np.array(matrix)
x0 = np.ones(A.shape[1])
sector=ranking_matrix['sector']
eigenvector, eigenvalue = power_method(A, x0, 10)
print("Dominant eigenvector:", eigenvector)
print("Dominant eigenvalue:", eigenvalue)
```

4.

```
show_eig=ranking_matrix[['sector','eigenvector']]
show_eig
```

4.3

1. µ1=df['Agriculture (MTCO2e)'].mean()

```
µ1
µ2=df['Industrial process(MTCO2e)'].mean()
µ2
```

2.

```
X=(µ1+µ2+µ3+µ4+µ5+µ6+µ7+µ8+µ9)/9
```

3.

```
SS_bet=(n*((µ1-X)**2)) + (n*((µ2-X)**2)) + (n*((µ3-X)**2)) + (n*((µ4-X)**2)) + (n*((µ5-X)**2)) + (n*((µ6-X)**2)) +(n*((µ7-X)**2)) + (n*((µ8-X)**2))
SS_bet
```

4. SS_w=a_sum+b_sum+c_sum+d_sum+e_sum+f_sum+g_sum+h_sum+i_sum

```
SS_w
```

5. SS_tot= SS_b+SS_w

```
SS_tot
```

6. # MS

```
Ms_SS_b=SS_b/x
```

Ms_SS_b

7. Ms_SS_w=SS_w/y

Ms_SS_w

8. F_ratio= Ms_SS_b/Ms_SS_w

F_ratio