SDG-15 Life on Land

ES1101: COMPUTATIONAL DATA ANALYSIS

FACULTY GUIDE-

Santosh Kumar Verma

Dr. Rabia Kamra

PREPARED BY:

(Group-1)

Disha Kumar Arora(2022Btech031)

Kanishk Tiwari(2022Btech045)

Aaditya Saxena(2022Btech001)

Neelesh Rana(2022Btech062)

Institute of Engineering and Technology (IET)

JK Lakshmipat University, Jaipur

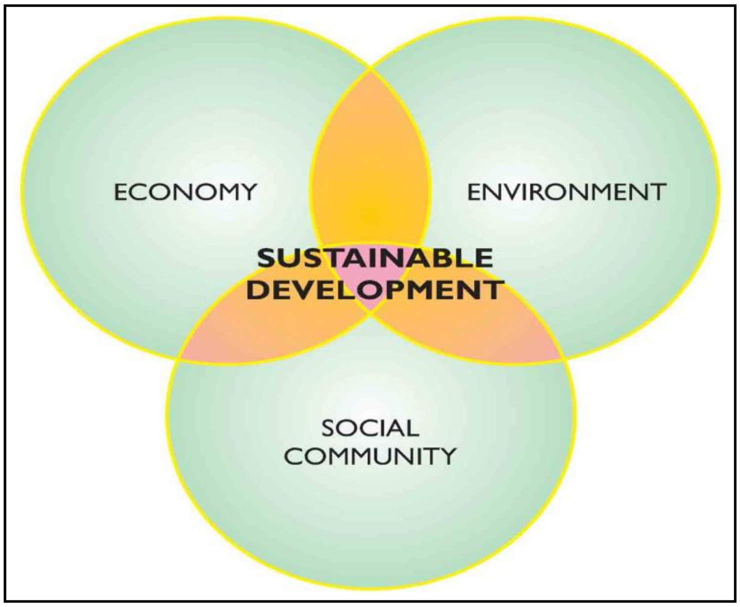
INDEX

Introduction

Meaning of Sustainable:

Sustainable refers to the ability to maintain or continue over a period of time. In the context of development, sustainable means meeting the needs of the present without compromising the ability of future generations to meet their own needs. It is often used to describe economic, social, and environmental practices that are intended to preserve resources and improve quality of life for current and future generations.

Sustainable development is a holistic approach to economic growth, social inclusion, and environmental protection. It involves balancing the three elements of economic growth, social progress, and environmental protection in order to create a more equitable and sustainable society. This can be achieved by implementing policies and practices that promote responsible resource use, social inclusion, and economic growth, while at the same time reducing environmental impacts and building resilience to climate change.



UN Sustainable Development Goals:

The Sustainable Development Goals (SDGs) are a set of 17 goals adopted by the United Nations in 2015 to end poverty, protect the planet, and ensure prosperity for all. They are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity by 2030. The SDGs cover a wide range of issues, including ending poverty and hunger, improving health and education, making cities more sustainable, combating climate change, and protecting oceans and forests.

List of all the 17 goals are:

- No Poverty: To eliminate extreme poverty and reduce inequality for all people.
- Zero Hunger: To end hunger and improve access to food for all people.
- Good Health and Well-being: To ensure that all people have access to good health care and are able to lead healthy lives.
- Quality Education: To provide all children and adults with access to quality education and lifelong learning opportunities.
- Gender Equality: To achieve gender equality and empower all women and girls.
- Clean Water and Sanitation: To ensure that all people have access to safe and clean water and sanitation.
- Affordable and Clean Energy: To increase access to affordable and clean energy for all people.
- Decent Work and Economic Growth: To promote sustained, inclusive and sustainable economic growth, full andproductive employment and decent work for all.
- Industry, Innovation and Infrastructure: To build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation.

- Reduced Inequalities: To reduce income and wealth inequalities within and among countries.
- Sustainable Cities and Communities: To make cities and human settlements inclusive, safe, resilient and sustainable.
- Responsible Consumption and Production: To ensure sustainable consumption and production patterns.
- Climate Action: To take urgent and significant action to combat climate change and its impacts.
- Life Below Water: To conserve and sustainably use the oceans, seas and marine resources for sustainable development.
- Life On Land: To protect, restore and promote the sustainable use of terrestrial ecosystems, forests, forests, wetlands, deserts and other ecosystems.
- Peace, Justice and Strong Institutions: To promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.
- Partnerships for the Goals: To strengthen the means of implementation and revitalize the global partnership for sustainable development.

GOAL-15 LIFE ON LAND

Sustainable Development Goal (SDG) 15 is about protecting and preserving the biodiversity of life on land, including forests, wetlands, deserts, and other ecosystems. The goal aims to halt and reverse the loss of natural habitats and biodiversity, and to restore degraded lands. The goal also focuses on sustainable management of all types of forests, which is crucial for the survival of many plant and animal species and for the livelihoods and well-being of millions of people worldwide. Additionally, the goal aims to combat desertification, which is the process of land becoming increasingly dry and unable to support life, and to strive towards a land-degradation-neutral world. The goal also focuses on ensuring the conservation of at least 17% of terrestrial and inland water areas and protecting and restoring at least 10% of coastal and marine areas.

India and SDG-15:

India has made significant progress in achieving some of the targets of SDG 15. For example, the country has made significant efforts to increase forest cover and has set a target of creating an additional carbon sink of 2.5 to 3 billion tonnes of CO2 equivalent through additional forest and tree cover by 2030. India has also implemented several policies and programs to combat desertification and land degradation, such as the National Action Plan on Desertification, the National Afforestation and Eco-development Board, and the National Watershed Development Project for Rainfed Areas.

However, India still faces many challenges in achieving the targets of SDG 15. Deforestation and illegal logging continue to be major issues, and the country still has a relatively low forest cover compared to other countries. Additionally, many of India's wetlands and other ecosystems are under threat due to urbanization, industrialization, and agriculture expansion.

India is also working with international organizations such as the United Nations Development Programme (UNDP) and the United Nations Framework Convention on Climate Change (UNFCCC) to implement SDG 15. India has also signed various international agreements, such as the United Nations Convention to Combat Desertification (UNCCD) and the United Nations Convention on Biological Diversity (CBD) to protect and preserve biodiversity

LITERATURE SURVEY

Literature survey for SDG 15, "Life on Land," involves reviewing and analyzing existing research and publications related to the protection, restoration, and sustainable use of terrestrial ecosystems, forests, wetlands, deserts, and other ecosystems. The survey could cover a wide range of topics, including:

Biodiversity conservation: This includes research on the conservation of different species and ecosystems, as well as the impacts of human activities on biodiversity.

Deforestation and forest degradation: This includes research on the causes and consequences of deforestation, as well as efforts to combat it and restore degraded forests.

Desertification and land degradation: This includes research on the causes and consequences of desertification and land degradation, as well as efforts to combat it and restore degraded lands.

Sustainable forest management: This includes research on the best practices for managing forests sustainably, including the use of certification schemes and community-based approaches.

Climate change and ecosystems: This includes research on the impacts of climate change on ecosystems and biodiversity, as well as efforts to adapt to and mitigate these impacts.

Policy and governance: This includes research on the effectiveness of different policies and governance mechanisms for protecting and preserving biodiversity and ecosystems.

Literature Survey Summary of Conservation of Endangered Species:

Conservation efforts often focus on large and charismatic animals such as the tiger and elephant that are undoubtedly in urgent need of protection. however, there are a host of species that do not rank very high on the conservation totem pole, although they are also under great threat and are classified as critically endangered by the IUCN(international union for conservation of nature). Critically endangered is the highest risk category assigned by the IUCN red list to wild species. There are five quantitative criteria to determine whether a taxon is threatened. Critically endangered means that the natural population of a species has decreased, or will decrease, by 80% within three generations, and all the available evidence indicates an extremely high risk of its extinction in the wild.

Literature Survey Summary of Poaching and Illegal Trafficking of Animals:

For several Critically Endangered taxa, such as rhinos, pangolins and wedgefish, trade-driven over-exploitation represents the greatest threat to their survival. Due to the impact of COVID-19 pandemic, a wide range of policy responses have been proposed. Extreme proposals include calls for a complete ban on the use and trade of wildlife or blanket global measures for entire Classes of wildlife. They aim to protect public health, while also improving animal welfare and meeting conservation goals. Others have called for a more balanced or targeted approach, targeting critical control points in supply chains or specific species that are more likely to harbor zoonotic viruses. In China, Top legislature passed a decision to thoroughly ban the illegal trading of wildlife and eliminate the consumption of wild animals to safeguard people's lives and health.This decision covers all terrestrial wild animals, fish, wild plants, amphibians and reptiles. Vietnam has temporarily banned the import of wildlife and wildlife products (exemptions for some non-edible products) and called for enforcement of existing laws to eliminate advertising, trade and consumption of illegal wildlife products. Similarly, Bolivia re-starting bans on wildlife trade and consumption as a matter of public health. Gabon has taken a more targeted approach by banning the consumption of bats and pangolins. In general, wildlife trade policies that incentivize sustainable use typically have more immediate positive effects on wildlife populations than outright trade bans.

Literature Survey Summary for Land Degradation and Desertification:

"Land degradation" means "reduction or loss of biological or economic productivity and complexity of rainfed cropland, irrigated cropland, or range, pasture, forest and woodlands resulting from land uses or from a process or combination of processes, including processes arising from human activities and habitation patterns, such as:

Soil erosion caused by wind and/or water.

Deterioration of the physical, chemical and biological or economic properties of soil.

India is a signatory to the UNCCD, which was adopted on June 17, 1994. Under UNCCD, a Thematic Programme Network - 1 on "Desertification Monitoring and Assessment", was identified as part of Regional Action Programme in Asia. Ahmedabad was identified as national nodal organisation to coordinate TPN-1 activities in the country by Ministry of Environment and Forest (MoEF), Government of India and subsequently SAC along with 17 participating agencies developed a three level national classification system (SAC 2007a).

Literature Survey Summary of forest cover area:

The total forest and tree cover of the country is 80.9 million hectare which is 24.62 percent of the geographical area of the country. As compared to the assessment of 2019, there is an increase of 2,261 sq km in the total forest and tree cover of the country. Out of this, the increase in the forest cover has been observed as 1,540 sq km and that in tree cover is 721 sq km.

Increase in forest cover has been observed in open forest followed by very dense forest. Top three states showing increase in forest cover are Andhra Pradesh (647 sq km) followed by Telangana (632 sq km) and Odisha (537 sq km). Area-wise Madhya Pradesh has the largest forest cover in the country followed by Arunachal Pradesh, Chhattisgarh, Odisha and Maharashtra. In terms of forest cover as percentage of total geographical area, the top five States are Mizoram (84.53%), Arunachal Pradesh (79.33%), Meghalaya (76.00%), Manipur (74.34%) and Nagaland (73.90%).17 states/ UT's have above 33 percent of the geographical area under forest cover. Out of these states and UT's, five states/UTs namely Lakshadweep, Mizoram, Andaman & Nicobar Islands, Arunachal Pradesh and Meghalaya have more than 75 percent forest cover while 12 states/UTs namely Manipur, Nagaland, Tripura, Goa, Kerala, Sikkim, Uttarakhand, Chhattisgarh, Dadra & Nagar Haveli and Daman & Diu, Assam, Odisha, have forest cover between 33 percent to 75 percent.Total mangrove cover in the country is 4,992 sq km. An increase of 17 sq Km in mangrove cover has been observed as compared to the previous assessment of 2019. Top three states showing mangrove cover increase are Odisha (8 sq km) followed by Maharashtra (4 sq km) and Karnataka (3 sq km).Total carbon stock in country's forest is estimated to be 7,204 million tonnes and there an increase of 79.4 million tonnes in the carbon stock of country as compared to the last assessment of 2019. The annual increase in the carbon stock is 39.7 million tonnes.

Literature Survey Summary of bio-diversity forest reserves:

Biodiversity refers to the variety of life on Earth, including the diversity within and among species, as well as the diversity of ecosystems. Biodiversity is essential for the functioning of ecosystems and the provision of a wide range of services that are important for human well-being, such as air and water purification, pollination of crops, and pest control.

National parks protect a wide range of plant and animal species, as well as their habitats, by providing a safe haven from human activities such as development, logging, and hunting.

Wildlife sanctuaries provide safe havens for a wide range of animal species, protecting them from hunting, poaching, and habitat loss.

Community reserves protect a wide range of plant and animal species, as well as their habitats, by providing a safe haven from human activities such as development, logging, and hunting.

Conservation of biodiversity: Conservation reserves protect a wide range of plant and animal species, as well as their habitats.

OBJECTIVES

Objective 1:

Take urgent action to end poaching and trafficking of protected species of flora and fauna and address both demand and supply of illegal wildlife products Enlist all the endangered and critically endangered species that are affected by poaching

Take urgent and significant action to reduce the degradation of natural habitats, halt the loss of biodiversity and, by 2020, protect and prevent the extinction of threatened species.

Problem Statement 1:

Which animal species have gone extinct in the last 150 years ranking of the species.

Problem Statement 2:

How much animals have been poached in 2020-30 prediction.

Problem Statement 3:

To see the correlation between extinct animals and extinct in wild animals.

Objective 2:

To take immediate action to reduce these nature harassing practices of Land degradation and Desertification, enlisting the factors effecting vegetation and biodiversity.

Problem Statement 1:

To designate the states based on their "Total area under desertification (ha)" over a period of 2011-13 (Power method).

Problem Statement 2:

Study, analyze, and correlate between the activities causing land degradation during 2003-05 and 2011-13 (Correlation).

Problem Statement 3:

To Find if there is any statistically significant difference in means of the "Vegetation Degradation 2011-13" and "Vegetation Degradation 2003-05", as to observe the extension of Vegetation degradation during those years. (Test Hypothesis)

Objective 3:

To integrate the ecosystem and biodiversity into national and local planning of the country, development processes.

Problem Statement 1:

To Predict the significant growth of the biodiversity in "National Parks", "Wild life sanctuaries" and "Forest Reserves" of India during 2022-2035 by using data of duration of 2006-2021.(Linear regression)

Problem Statement 2:

To designate all the states of India according to the area covered by national parks(km sqr) (power method).

Problem Statement 3:

To find if there is any statistical significant difference in the means of the "Area of Wild life sanctuaries (2022-2030)" and "Area of Wild life sanctuaries (2012-2021)",as to observe if there will be any extension of Area covered by Wild life sanctuaries. (Test Hypothesis)

Objective 4:

By 2030, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements.

Problem Statement 1:

To find if there is any statistical significance diffrence in mean of the "total forest area 2017" and "total forest area 2021"

Problem Statement 2:

To correlate forest area and desertification

Problem Statement 3: To predict the total forest area from 2025 to 2030.

Problem Statement 4: To compare the state on the basis of total forest area of 2021 and 2025.

Methodology

Correlation:

In statistics, correlation refers to the relationship between two variables. Correlation is measured using a correlation coefficient, which can range from -1 to 1. A correlation coefficient of 1 indicates a perfect positive correlation, meaning that as one variable increases, the other variable also increases. A correlation coefficient of -1 indicates a perfect negative correlation, meaning that as one variable increases, the other variable decreases. A correlation coefficient of 0 indicates no correlation.

It's important to remember that correlation does not imply causation. Just because two variables are correlated does not mean that one variable causes the other.

Formula:

$$r = \frac{(S_{xy})}{\sqrt{(S_{xx})(S_{yy})}}$$

$$S_{xy} = \Sigma xy - \Sigma x \Sigma y/n \qquad S_{xx} = \Sigma x^2 - (\Sigma x)^2/n \qquad S_{yy} = \Sigma y^2 - (\Sigma y)^2/n$$

Linear regression:

Linear regression is a statistical method that is used to model the relationship between a dependent variable (Y) and one or more independent variables (X). It is a linear approach to model the relationship between variables, and it is based on the assumption that there is a linear relationship between the independent and dependent variables.

The basic **formula** for a simple linear regression model is:

$$Y = b0 + b1*X$$

Where:

Y is the dependent variable (the variable that we are trying to predict)

X is the independent variable (the variable that we are using to make predictions)

b0 is the y-intercept of the line (the point where the line crosses the y-axis b1 is the slope of the line (the rate of change in Y as X changes)

To determine the values of b0 and b1, we use the method of least squares. This method minimizes the sum of the squared differences between the predicted values of Y (based on the values of X) and the actual values of Y.

The formula for the method of least squares is:

$$b1 = (n*(\sum(xiyi)) - (\sum xi)(\sum yi)) / (n*(\sum(xi^2)) - (\sum xi)^2)$$

$$b0 = (\sum yi - b1*(\sum xi)) / n$$

Where:

n is the number of observations

xi is the ith value of X

yi is the ith value of Y

Once we have determined the values of b0 and b1, we can use the linear regression equation (Y = b0 + b1*X) to make predictions about the value of Y for a given value of X.

It's important to note that linear regression assumes that there is a linear relationship between the independent and dependent variables. If the relationship is non-linear, a different type of regression model may be more appropriate. Additionally, linear regression assumes that there is no multi collinearity and no autocorrelation among the independent variables.

Polynomial regression:

Polynomial regression is a type of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial. Polynomial regression can be used to fit a nonlinear relationship between the data-points.

The basic **formula** for a polynomial regression model is:

$$Y = b0 + b1X + b2X^2 + ... +bn*X^n$$

Where:

Y is the dependent variable (the variable that we are trying to predict)

X is the independent variable (the variable that we are using to make predictions) $b_0, b_1, \ldots, b_n$

b0, b1, b2, ..., bn are the coefficients of the polynomial.

To determine the values of the coefficients, we use a method similar to the method of least squares used in linear regression. We minimize the sum of the squared differences between the predicted values of Y (based on the values of X) and the actual values of Y.

The polynomial regression model can be fit by using a non-linear least squares method to fit a polynomial function to the data. It can also be fit by using linear regression on the polynomial feature of the independent variable.

It's important to note that polynomial regression can suffer from over fitting, meaning it can fit the noise in the data instead of the underlying trend. To prevent over fitting, techniques such as cross-validation and regularization can be used.

Additionally, as the degree of polynomial increases, the polynomial regression model becomes more complex and may not generalize well to new data. It is important to choose the right degree of polynomial that fits the data well and also generalizes well to new data.

Power Method:

The power method is an algorithm used to find the largest eigenvalue and corresponding eigenvector of a square matrix. It is particularly useful for large and sparse matrices, and it is a simple and efficient method for approximating the largest eigenvalue and eigenvector of a matrix.

The basic **formula** for the power method is:

x0 = an initial approximation of the eigenvector

for i = 1 to k (k is the number of iterations)

xi = A * xi-1

λ = (xi-1) * xi / (xi-2 * xi-1)

Where:

A is the square matrix

λ is the largest eigen value

xi is the eigenvector corresponding to the largest eigenvalue

The power method starts with an initial approximation of the eigenvector, and then it iteratively multiplies the matrix A by the approximation to find the eigenvector corresponding to the largest eigen value. The eigen value is then calculated as the dot product of the current approximation and the previous approximation divided by the dot product of the previous two approximations.

T-Test:

Methodology:

T-Test: It is a parametric statistical test which is used to compare and infer whether the two population's means differ from one another or not, when the standard deviation is not known. It is used when sample size is not large (n

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where ,

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.,$$

Here, $S_1{}^2$ and $S_2{}^2$ are the sample variance of sample 1 and sample 2. N1 and n_2 are the population size of two samples.

For one population:

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

Where,

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$

Here, n is the population size.

$\bar{X}$=mean of the sample

μ= population mean

s=Standard deviation of the sample

To see whether we've to apply independent t-test or relative t-test we've to see the correlation between the columns.

Objective 1

Take urgent action to end poaching and trafficking of protected species of flora and fauna and address both demand and supply of illegal wildlife products Enlist all the endangered and critically endangered species that are affected by poaching.

Take urgent and significant action to reduce the degradation of natural habitats, halt the loss of biodiversity and, by 2020, protect and prevent the extinction of threatened species

Datasets:

Tiger Population

Year	Bihar	Uttrakhand	Uttar Pradesh	Andhra Pradesh	Telangana	Chhattisgarh	Jharkhand	Madhya Pradesh	Maharashtra	...	Karnataka	Kerala
1986	0	152	72	56	28	13	25	278	81	...	277	21
1990	0	120	58	59	28	9	17	243	69	...	265	18
1994	0	98	53	57	20	9	13	231	71	...	261	24
1998	7	132	76	66	12	16	7	199	56	...	250	36
2002	6	152	100	79	14	19	5	291	82	...	264	38
2006	10	178	109	95	10	26	6	270	103	...	290	46
2010	8	227	118	72	8	26	10	257	168	...	300	71
2014	28	340	117	68	12	46	3	308	190	...	406	136
2018	31	442	173	65	26	49	5	526	312	...	524	190

Extinct and Extinct in Wild Species

	Name	EX(1900-1960)	EW(1900-1960)	Subtotal(EX+EW)(1900-1960)	EX(1960-2020)	EW(1960-2020)	Subtotal(EX+EW)(1900-1960).1
0	ACTINOPTERYGII	79	10	89	29	38	67
1	AMPHIBIA	35	2	37	12	8	20
2	ANTHOZOA	0	0	0	1	0	1
3	ARACHNIDA	9	0	9	9	0	9
4	ASTEROIDEA	0	0	0	0	1	1
...	...	...	...	...	...	...	...
63	POLYTRICHOPSIDA	0	0	0	0	0	0
64	SPHAGNOPSIDA	0	0	0	0	0	0
65	TAKAKIOPSIDA	0	0	0	0	0	0
66	ULVOPHYCEAE	0	0	0	0	0	0
67	Total	902	80	982	533	114	647

Poaching in India

	Species	Mean 1992-96	SD	Mean 1997-01	SD.1	t (Period1-Period2)	P	Mean 2002-2006	SD.2	t (Period1- Period3)	P.1	DF
0	Leopard skin	14.37	34.67	20.80	49.28	-1.710	0.04	22.60	46.50	-2.47	0.01	34
1	Leopard bone	1.48	8.66	7.25	29	-1541.000	0.06	1.91	6.88	-1.08	0.141	34
2	Leopard claw	0.34	1.99	51.91	300.14	-1.020	0.15	10.88	39.33	-1.99	0.02	34
3	Leopard equivalent	15.4	35.12	55.08	213.27	-2.870	<0.01	29.25	50.84	-3.01	<0.01	34
4	Tiger skin	3.97	8.39	7.91	16.39	-2.010	0.02	3.91	5.84	0.57	0.28	34
5	Tiger bone	19:48	84.84	29.65	166.11	-0.330	0.37	1.00	4.00	-2.05	0.02	34
6	Tiger equivalent	7.71	16.96	13.42	29.19	-1.480	0.07	7.28	10.66	-0.76	0.22	34
7	Elephant body	8.65	17.67	8.54	16.04	5.391	-0.01	7.77	16.67	3.98	<0.01	34
8	Elephant ivory (kg)	36.55	93.98	66.75	193.79	-0.990	0.16	24.44	68.99	0.73	0.23	34
9	Elephant Ivory (no)	0.62	265.00	5.20	21:36	-1.350	0.09	3.25	12.74	-1.16	0.12	34
10	Elephant tusk (no)	3.25	11.82	7.14	14.7	-2.950	<0.01	8.80	15.32	-2.94	<0.01	34
11	Elephant equivalent	13.17	22.21	19.34	32.7	-159.000	0.05	15.40	28.00	-1.09	0.14	34
12	Deer body (no)	3.6	9.63	8.82	41.2	-2.370	0.011	3.97	10.09	-3.37	<0.01	34
13	Deer skin (no)	1.6	4.87	4.85	11.91	0.790	0.21	9.82	22.18	-1.07	0.141	34
14	Deer antler (no)	2.97	10.05	13.71	54.52	-1.120	0.13	15.00	47.97	-2.02	0.02	34
15	Deer antler (kg)	67.54	193.08	977.83	3284.67	-1.650	0.05	218.45	1166.26	0.80	0.21	34
16	Deer equivalent (no)	25.45	71.93	206.17	668.77	-1.630	0.05	63.25	233.44	-2.11	0.02	34
17	Snake skin	2439.4	10818.63	184.68	640.86	1.720	0.04	2692.22	10558.26	0.19	0.42	34
18	Snake equivalent	2643.45	11337.13	207.48	697.22	-1.730	0.04	2703.40	10602.23	-0.02	0.49	34
19	Bird	827.62	2886.97	1054.62	3155.7	-1.011	0.16	348.17	1285.03	-0.41	0.33	34
20	Star tortoise	260.88	783.02	809.00	3541.95	-0.900	0.36	823.08	3214.30	-1.69	0.04	34
21	Red sander	41605	242567.00	3944.25	21633	-1.590	0.06	31538.97	107487.91	-1.97	0.02	34

Problem Statement 1:

To find how many species of each class of animals and trees have gone Extinct in last 150 years

Methodology

Linear algebra ranking:

Power method:

This method is used to find the dominant eigen value and a corresponding eigenvector. For square matrix S, we first guess the initial eigen vector of dominant eigen value. Then we multiply the RANK USING POWER METHOD • First we convert data into matrix form and assign number according to the condition. To be predict the data using the linear regression first we have to.

Find the value of intercept and coefficient value.

Equation of the Linear regression is Y= a + bx

Output:

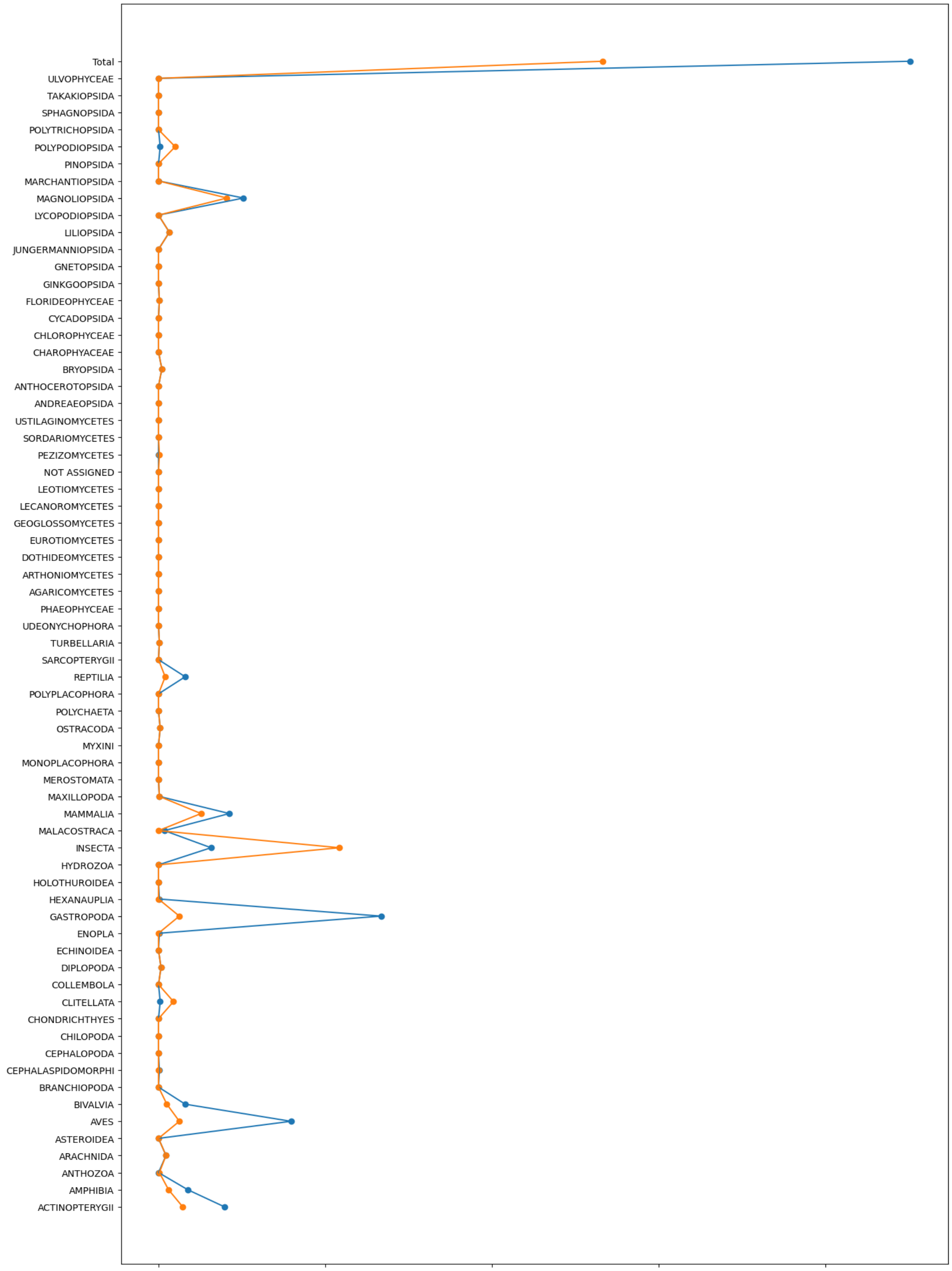
```
Total
GASTROPODA
AVES
MAGNOLIOPSIDA
MAMMALIA
ACTINOPTERYGII
INSECTA
AMPHIBIA
BIVALVIA
BIVALVIA
LILIOPSIDA
ARACHNIDA
MALACOSTRACA
BRYOPSIDA
DIPLOPODA
CLITELLATA
CLITELLATA
CLITELLATA
CEPHALASPIDOMORPHI
```

Conclusion:

From this graph we can see that more animals have gone extinct between 1900-1960 than 1960-2020, Small animal species like Insecta, Gastropoda(Snails), magnoliopsida(Flowering plants of Himalaya) are more prone to extinction as they do not get any consrvation programme unlike Tigers, Elephants and Rhinos. Their extinction can cause destructions of ecosystems. More animals must have got extinct in 20th century because there were no conservation efforts in most pars of world and poaching was done flawlessly especially of exotic birds species(Like Pink Duck found in India, got extinct in 1920s.

Graph:

Comparison of Number of Species gone Extinct between 1900-1960 & 1960-2020



Total
ULVOPHYCEAE
TAKAKIOPSIDA
SPHAGNOPSIDA
POLYTRICHOPSIDA
POLYPODIOPSIDA
PINOPSIDA
MARCHANTIOPSIDA
MAGNOLIOPSIDA
LYCOPODIOPSIDA
LILIOPSIDA
JUNGERMANNIOPSIDA
GNETOPSIDA
GINKGOOPSIDA
FLORIDEOPHYCEAE
CYCADOPSIDA
CHLOROPHYCEAE
CHAROPHYACEAE
BRYOPSIDA
ANTHOCEROTOPSIDA
ANDREAEOPSIDA
USTILAGINOMYCETES
SORDARIOMYCETES
PEZIZOMYCETES
NOT ASSIGNED
LEOTIOMYCETES
LECANOROMYCETES
GEOGLOSSOMYCETES
EUROTIOMYCETES
DOTHIDEOMYCETES
ARTHONIOMYCETES
AGARICOMYCETES
PHAEOPHYCEAE
UDEONYCHOPHORA
TURBELLARIA
SARCOPTERYGII
REPTILIA
POLYPLACOPHORA
POLYCHAETA
OSTRACODA
MYXINI
MONOPLACOPHORA
MEROSTOMATA
MAXILLOPODA
MAMMALIA
MALACOSTRACA
INSECTA
HYDROZOA
HOLOTHUROIDEA
HEXANAUPLIA
GASTROPODA
ENOPLA
ECHINOIDEA
DIPLOPODA
COLLEMBOLA
CLITELLATA
CHONDRICHTHYES
CHILOPODA
CEPHALOPODA
CEPHALASPIDOMORPHI
BRANCHIOPODA
BIVALVIA
AVES
ASTEROIDEA
ARACHNIDA
ANTHOZOA
AMPHIBIA
ACTINOPTERYGII
0
200
400
600
800

Problem Statement 2:

To predict the Tiger population for each Indian state for next 5 years using Linear Regression

Methodology

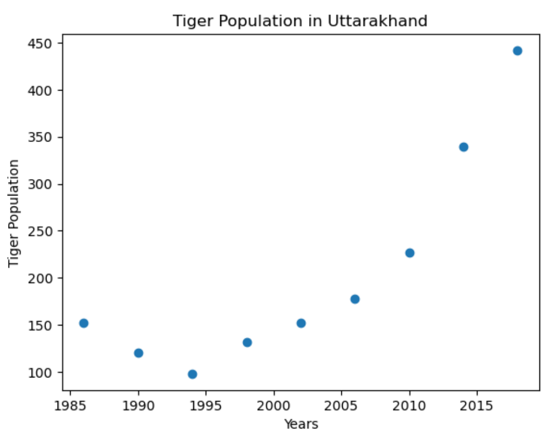
Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

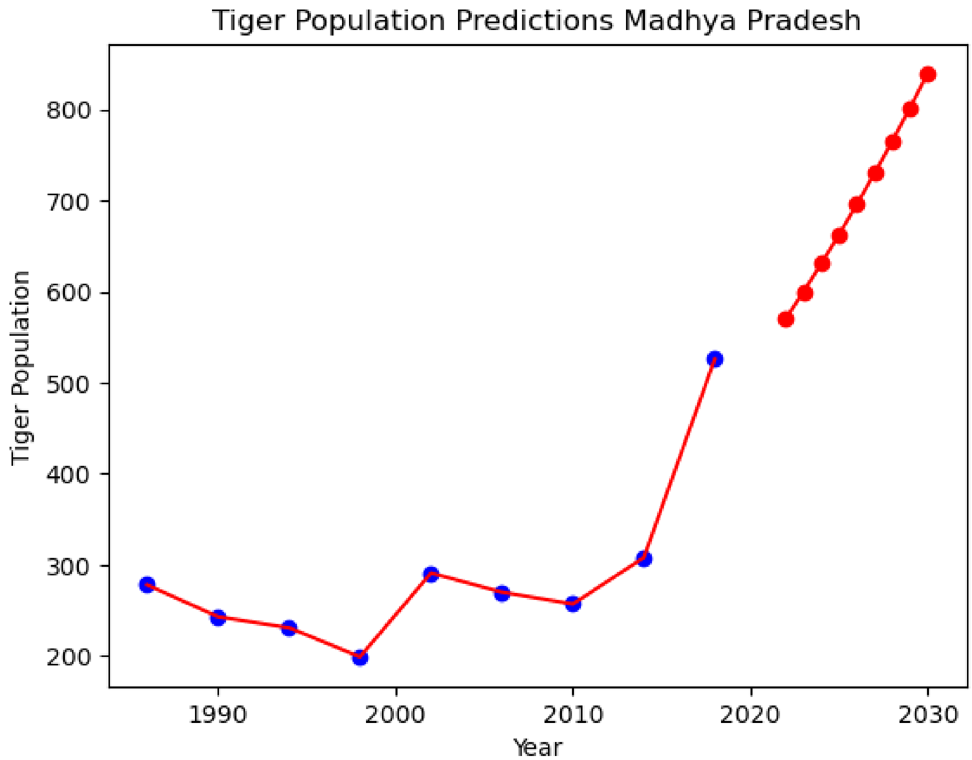
Output:

```
[[372.70555556]
 [381.55555556]
 [390.40555556]
 [399.25555556]
 [408.10555556]
 [416.95555556]
 [425.80555556]
 [434.65555556]
 [443.50555556]]
```

Graph:


```
[[372.70555556]
 [381.55555556]
 [390.40555556]
 [399.25555556]
 [408.10555556]
 [416.95555556]
 [425.80555556]
 [434.65555556]
 [443.50555556]]
```

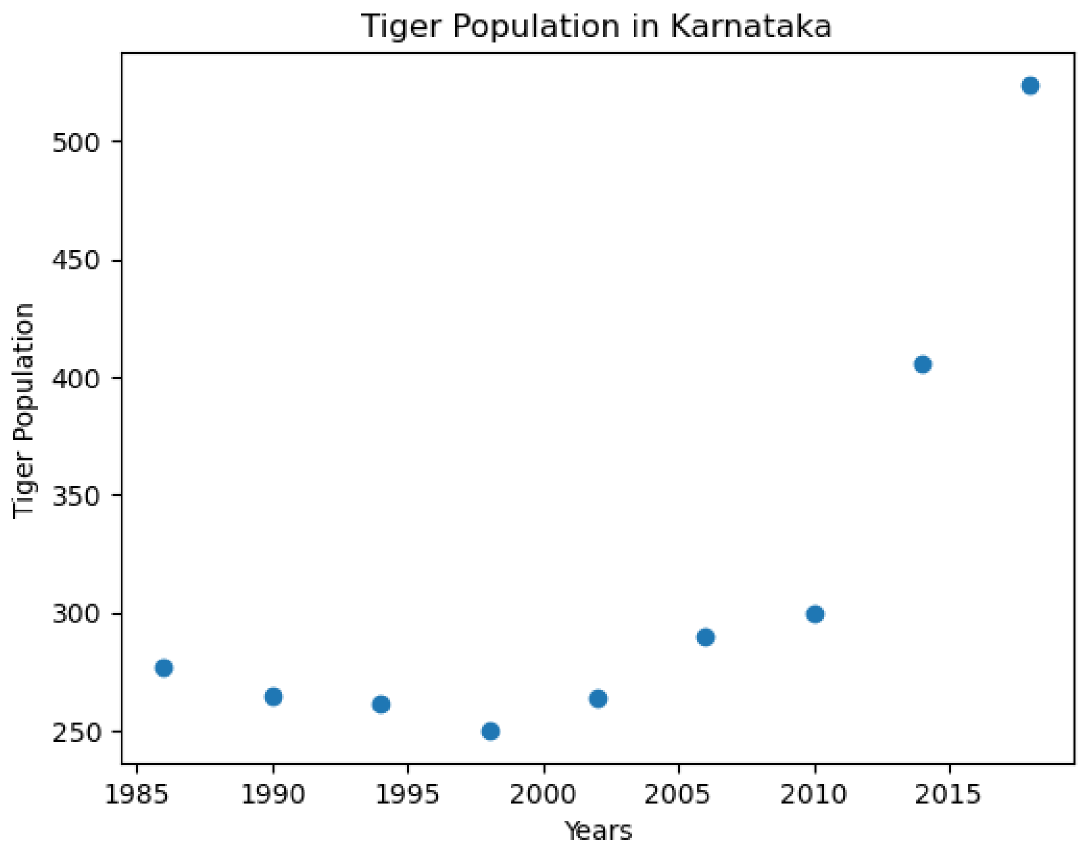





Output for Tiger Population in TN:

```
[[232.72916667]
 [238.75      ]
 [244.77083333]
 [250.79166667]
 [256.8125    ]
 [262.83333333]
 [268.85416667]
 [274.875     ]
 [280.89583333]]
```

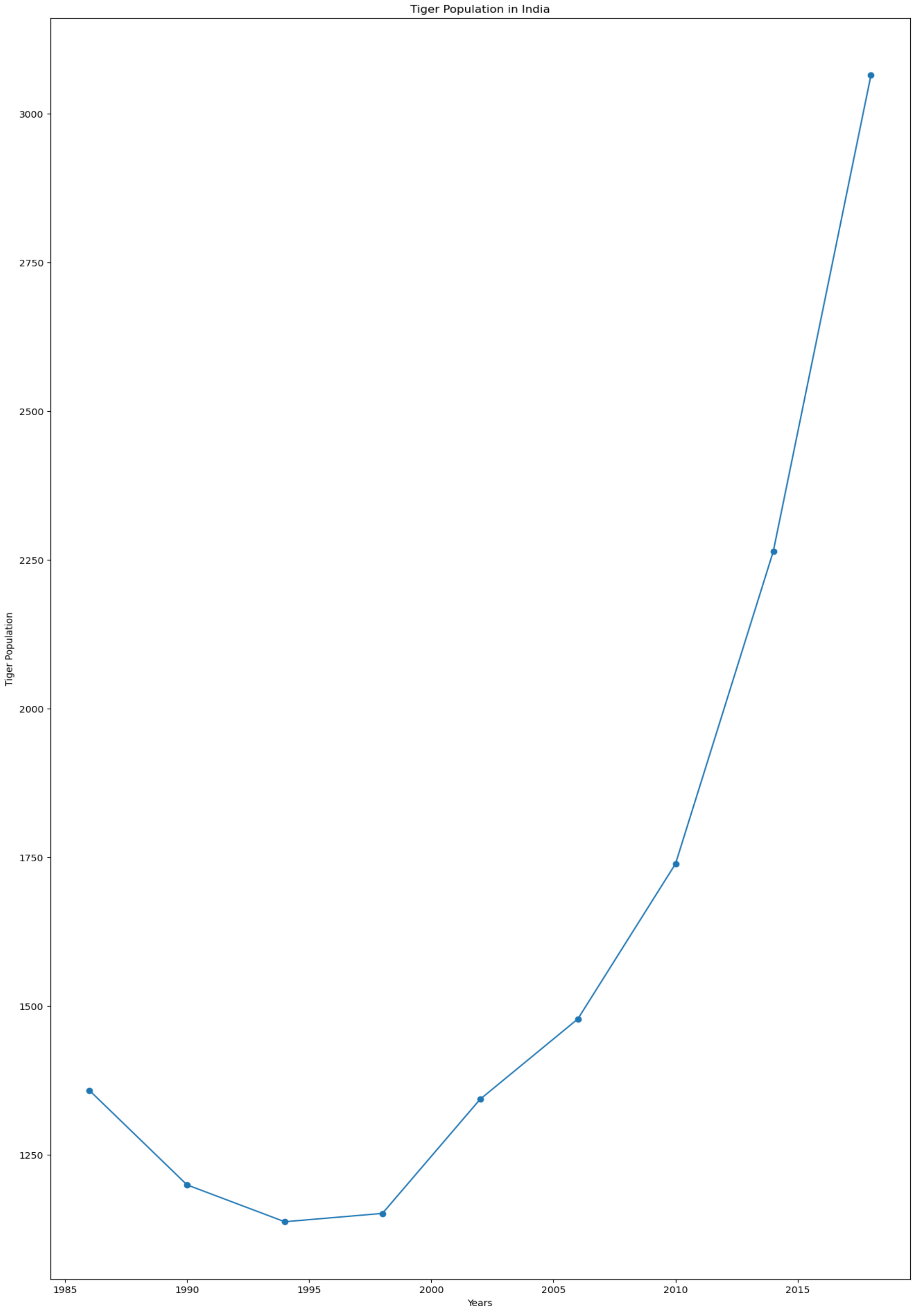
Graph:



Conclusion:

From this we can conclude that in future Tiger Population will continosly rise in most states of India. Especially in States of MP, Maharashtra, Uttrakhand etc Graph of Tiger population (last 30 years). We know that during late 1990's a large no of national parks were made, due to which tiger population almost doubled in last 15 years.

Tiger Population in INDIA



Statement 3:

Correlation between Extinct animals and extinct in wild animals

Methodology:

$$r = \frac{(S_{xy})}{\sqrt{(S_{xx})(S_{yy})}}$$

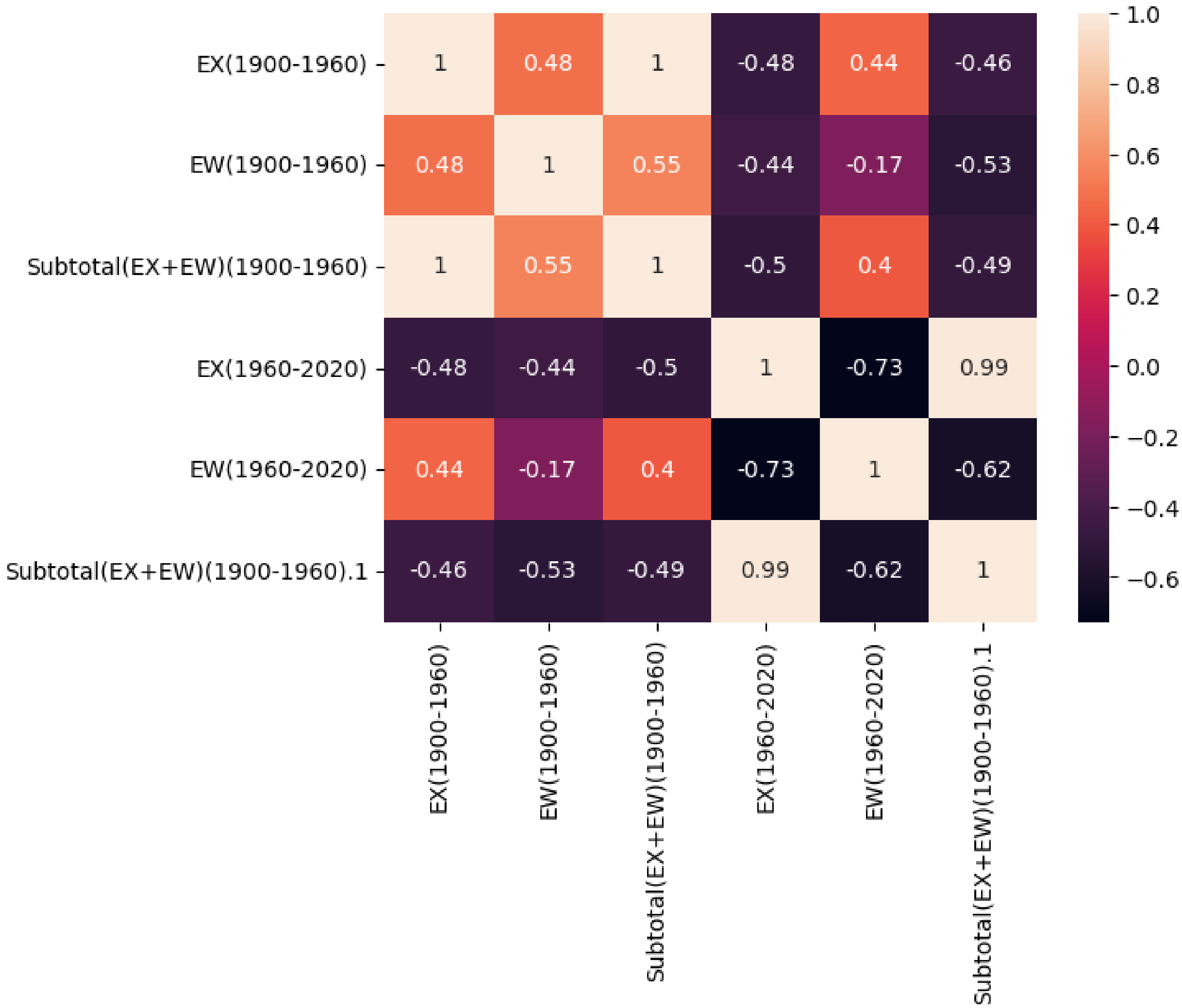
$S_{xy} = \Sigma xy - \Sigma x \Sigma y/n$ $S_{xx} = \Sigma x^2 - (\Sigma x)^2/n$ $S_{yy} = \Sigma y^2 - (\Sigma y)^2/n$

Here I use correlation to find what instant they are interrelated r (coefficient of correlation)

Output:

```
7385.529411764706
315.29411764705884
3554.9411764705883
76867.29411764706
72.60515379779524
0.946102408273731
0.868484821207163 9
```

Graph:



Conclusion:

From this we can conclude that there was a large amount of poaching conducted in the years 1900-1960 as the correlation between extinct animals and Extinct in wild is larger compared to 1960-2020.

This shows that the policies applied by the governments worldwide have contributed significantly in decreasing the amount of species going extinct

Objective 2

To take immediate action to reduce these 2 nature harassing practices of Land degradation and Desertification.

Datasets:

	State Name	Vegetation Degradation 2011-13	Vegetation Degradation 2003-05	Water Erosion 1	Water Erosion 2	...	Total Area under Desertification (%) 1	Total Area under Desertification (%) 2	NO Apparent adation 1	NO Apparent adation 2
0	Andhra Pradesh	0.236822	0.237352	0.097281	0.102604	...	0.186327	0.188875	13447078	13476591
1	Arunachal Pradesh	0.022675	0.020051	0.017695	0.025781	...	0.000000	0.000000	8144850	8162237
2	Assam	0.094784	0.064039	0.003176	0.003874	...	0.108728	0.085469	6591013	6735134
3	Bihar	0.047711	0.050216	0.039148	0.039689	...	0.082514	0.080947	8511828	8527091
4	Chhattisgarh	0.274539	0.274165	0.096562	0.100840	...	0.216265	0.218119	11130592	11166012
	....									
25	Telangana	0.108979	0.108292	0.353628	0.387090	...	0.439380	0.455683	7689491	7631019
26	Tripura	0.046449	0.023578	0.022478	0.024621	...	0.593536	0.445885	608776	716717
27	Uttar Pradesh	0.082785	0.082814	0.072145	0.079924	...	0.067173	0.090292	22115961	21831845
28	Uttarakhand	0.122411	0.109742	0.000757	0.001318	...	0.153113	0.139282	4667750	4738936
29	West Bengal	0.052379	0.052111	0.164334	0.170274	...	0.263628	0.261079	6884910	6926022

<u>Problem Statement 1:</u>

To designate the state ranks on the basis of Total area under desertification (ha) over a period of 2011-13 (Power method).

Methodology:

LINEAR ALGEBRA

POWER METHOD: This method is used to find the dominant eigen value and a corresponding eigenvector. For square matrix S, we first guess the initial eigen vector of dominant eigen value.

Then we multiply the RANK USING POWER METHOD

- First we convert data into matrix form and assign number according to the condition.

In order to rank these sectors, we created the matrix containing 1, 2 and 3.

Where: - (i) 1 is marked where the first value is greater than second value.

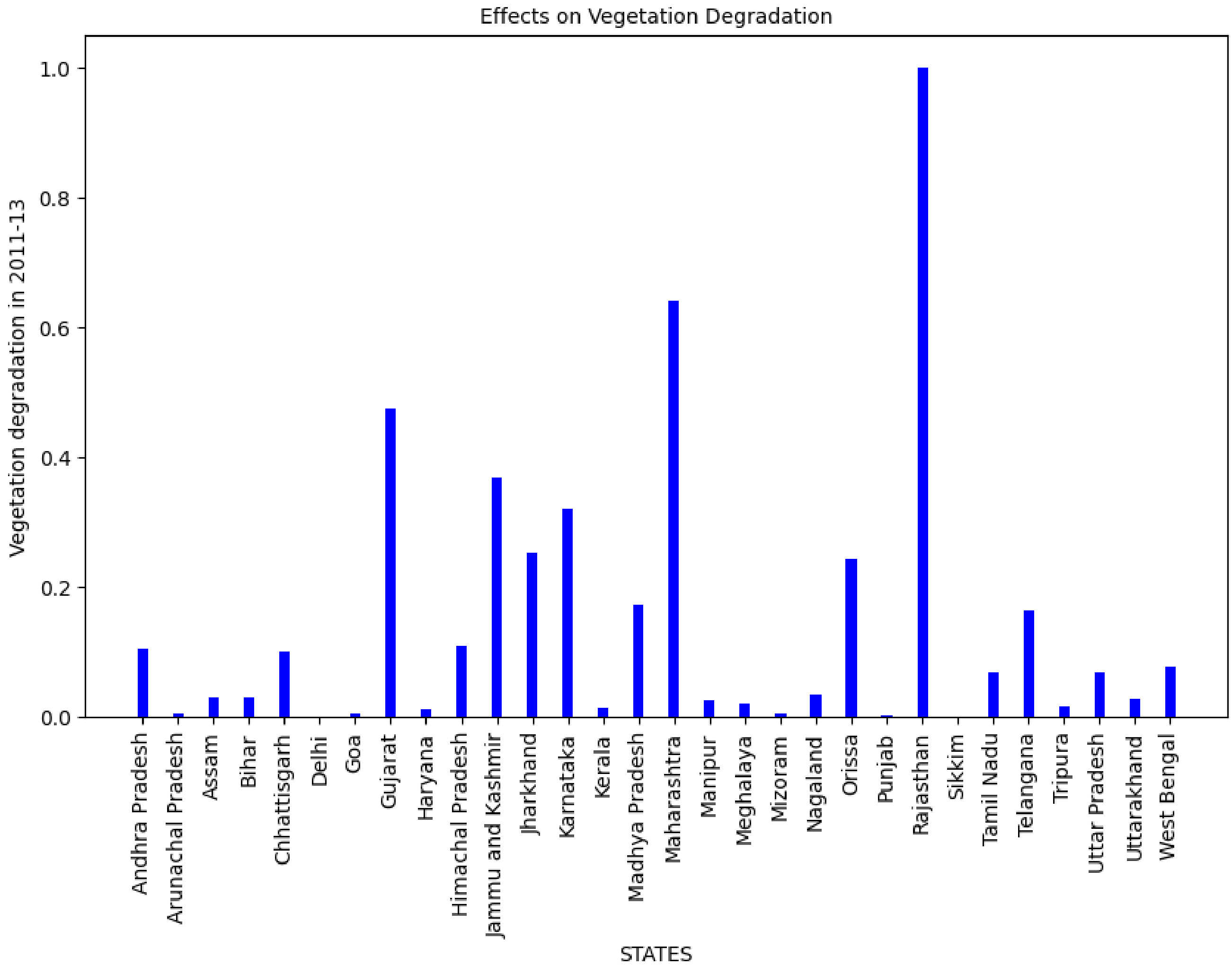
(ii) 2 is marked where both values are same.

(iii) 3 is marked where the first value is greater than second value.

Output:

```
Sikkim
Delhi
Punjab
Arunachal Pradesh
Mizoram
Goa
Haryana
Kerala
Tripura
Meghalaya
Manipur
Uttarakhand
Bihar
Assam
Nagaland
Uttar Pradesh
Tamil Nadu
West Bengal
Chhattisgarh
Andhra Pradesh
Himachal Pradesh
Telangana
Madhya Pradesh
Orissa
Jharkhand
Karnataka
Jammu and Kashmir
Gujarat
Maharashtra
Rajasthan
```

Graph:



Conclusion:

As we can notice in the output above 'Rajasthan' got last position in ranlking states according to the ascending order of having 'Total area under desertification during 2011-13' because of having largest desert like 'THAR', whereas 'Sikkim' is at first because of having smallest land area with negligible desert area.

Problem Statement 2:

Study, analyze, and correlate between the activities causing land degradation during 2003-05 and 2011-13 (Correlation).

Methodology:

Here I use correlation to find what instant they are interrelated r (coefficient of correlation)

$$r = \frac{(S_{xy})}{\sqrt{(S_{xx})(S_{yy})}}$$

$S_{xy} = \Sigma xy - \Sigma x \Sigma y/n$ $\quad$ $S_{xx} = \Sigma x^2 - (\Sigma x)^2/n$ $\quad$ $S_{yy} = \Sigma y^2 - (\Sigma y)^2/n$

Correlation between Vegetation Degradation and Water Erosion in 2011-13

Output:

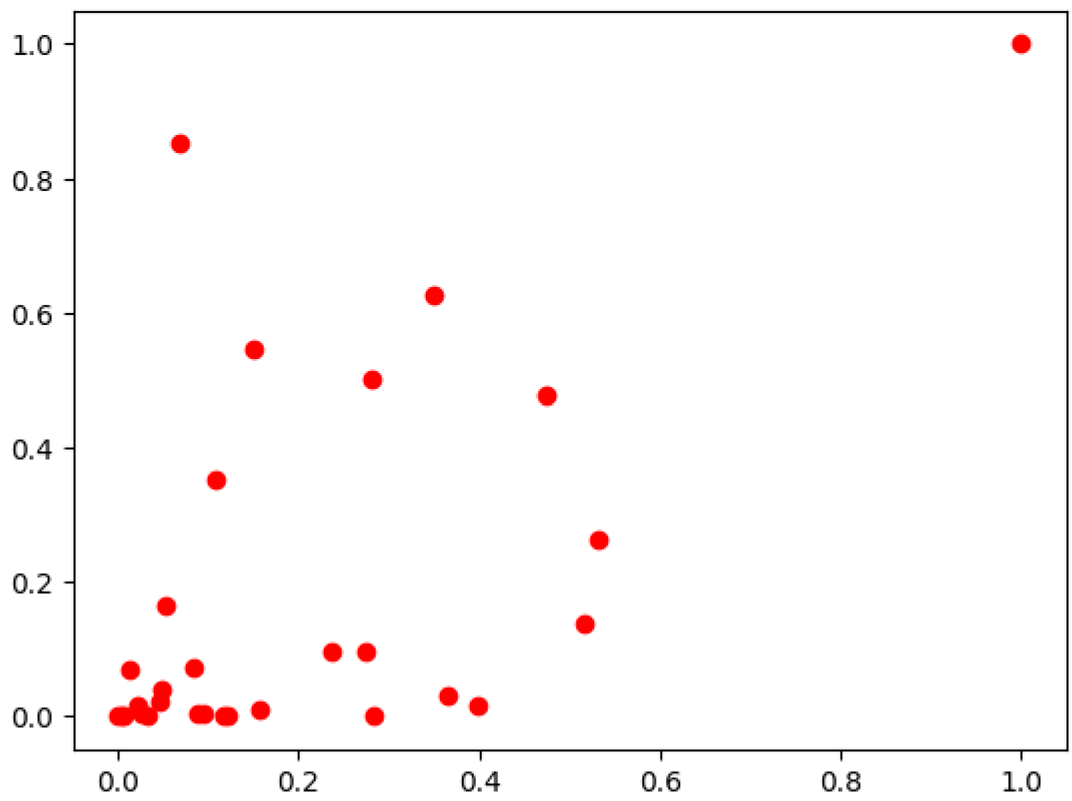
0.561387334684744

Correlation between Vegetation Degradation and Water Erosion in 2003-05

Output:

0.5580721759026515

Graph for the above two outputs:



Correlation between Vegetation Degradation and Wind Erosion in 2011-13

Output:

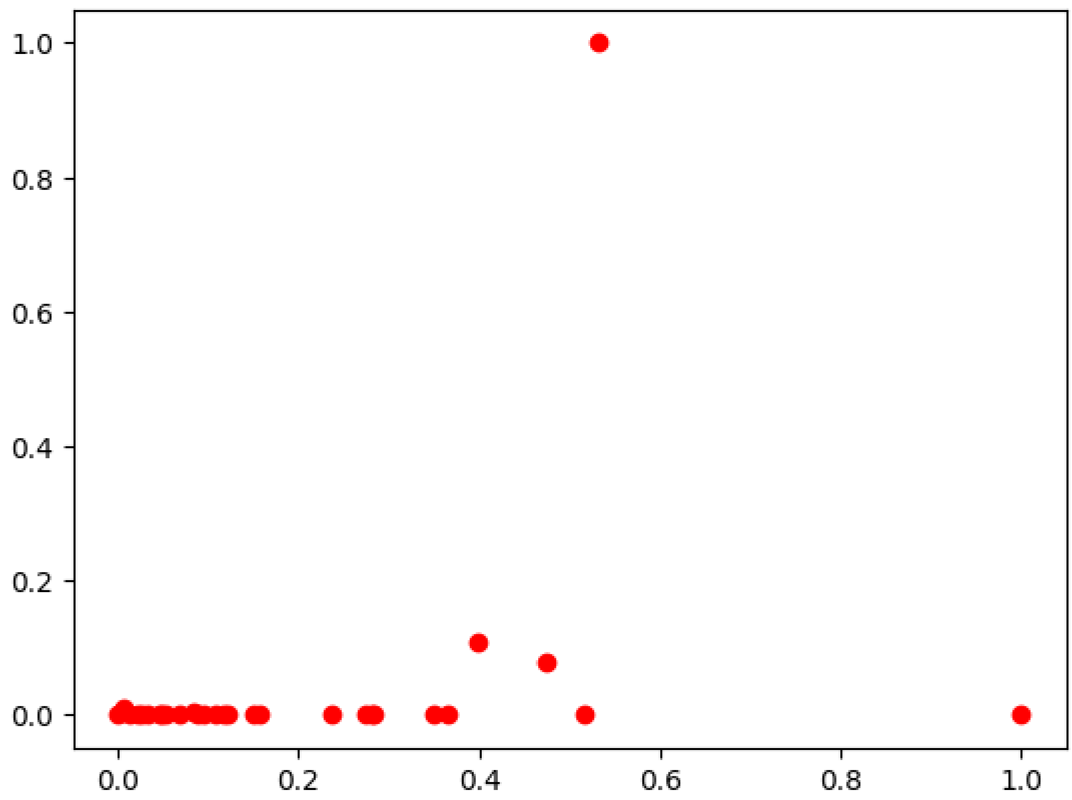
0.319944189722402335

Correlation between Vegetation Degradation and Wind Erosion in 2003-05

Output:

0.3205017378482241

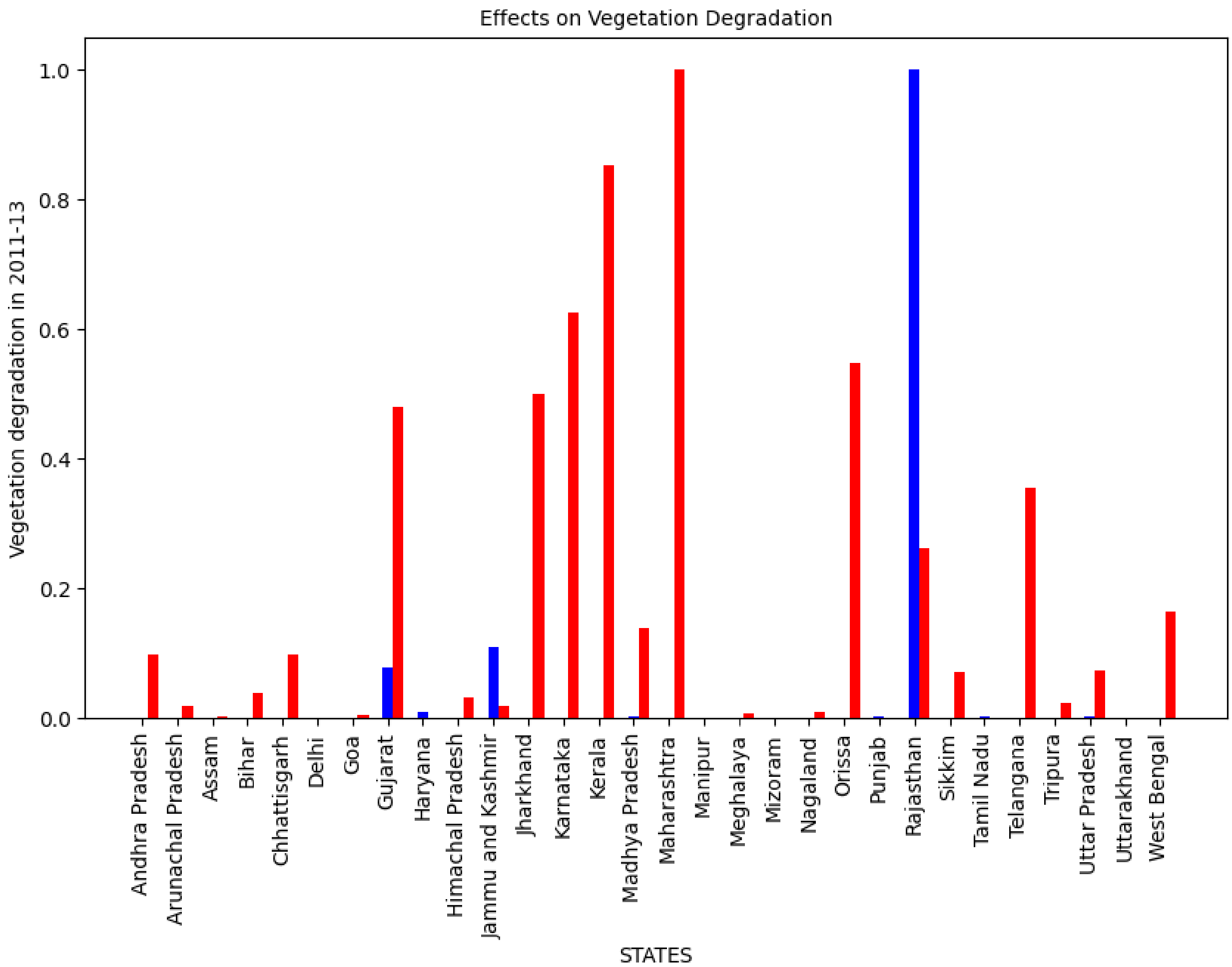
Graph for the above two outputs:



Conclusion:

We can draw a conclusion from this correlation that water erosion is a greater cause of vegetation degradation than wind erosion, hence top soil in farm lands is at a greater risk from flash floods in India than wind storms. Water erosion is a continuous process which can lead to the destruction of vegetation over time, unlike wind erosion which is occasional and in most cases not continuous. The force of water erosion can uproot plants, wash away soil, and erode riverbanks.

Overview:



Conclusion:

As per the graph we can conclude that Wind Erosion in Dry states is more than compared to others, i.e. 'Rajasthan' and 'jammu and kashmir' have two of the largest deserts like 'Thar' and 'Ladakh'(Containing semi arid, Barren, vegetation restricted areas) respectivly.

Problem Statement 3:

To Find if there is any statistically significant difference in means of the "Vegetation Degradation 2011-13" and "Vegetation Degradation 2003-05", as to observe the extension of Vegetation degradation during those years. (Test Hypothesis)

Methodology

Test of hypothesis difference of mean (two population)

$\mu 1$= Mean of the Vegetation Degradation 2011-13.

$\mu 2$= Mean of the total Vegetation Degradation 2003-05.

Level of significance (α) = 0.05

Null Hypothesis

This Hypothesis presumes that there is no statistically significance difference between the means of "Vegetation Degradation 2011-13" and "Vegetation Degradation 2003-05".

$H0$: $\mu 1 - \mu 2 = 0$

Alternative Hypothesis

This Hypothesis presumes that there is an enough significant difference between the means of "Vegetation Degradation 2011-13" and "Vegetation Degradation 2003-05".

Ha: $\mu 1 - \mu 2 \neq 0$

Level of significance(α)=0.05

Output:

```
p-value: 0.9001477022491438
t-value 0.1260253579841153
Difference of means: 0.007207192633333304
Null Hypothesis is accepted
```

Conclusion:

As per the output our 'T-value' is in the appropriate range to be selected as in the accepetence region of 'Null hypothesis'. Hence we can conclude that no major change was scene in the vegetation degradation during those mid years, which is good, for our biodiversity and Land conservation. As per the output, we can draw a conclusion, due to 'P-value' being larger than Level of significance( ), that we are aware of accepting 'Null hypothesis', because of the difference α between the means of "Vegetation Degradation 2011-13" and "Vegetation Degradation 2003-05", being negligible(0.007...).

Objective 3

To designate all the states of India according to the area covered by national parks(km sqr). (Power method)

Dataset:

Area and number of different types of biodiversity reserves.

	Year	No. of National parks	Area Under National parks (km2)	No. of Wild Life sanctuaries	Area Under Wild Life sanctuaries (km2)	No. of Community reserves	Area Under Community reserves (km2)	No. of Conservation reserves	Area Under Conservation Reserves (km2)	No. of Protected areas	Total Area under areas (km2)
0	2000.0	89.0	37803.10	485.0	10862.50	0.0	0.00	2.0	28.33	574.0	146665.60
1	2006.0	96.0	38392.12	503.0	111229.48	1.0	0.31	4.0	42.87	604.0	149664.78
2	2007.0	95.0	35428.88	507.0	111529.04	5.0	21.00	7.0	94.82	617.0	150073.74
3	2008.0	99.0	39441.74	510.0	113123.35	5.0	21.00	45.0	1259.84	659.0	153845.93
4	2009.0	99.0	39441.74	512.0	113395.36	5.0	21.00	45.0	1259.84	661.0	154117.94
5	2010.0	102.0	40283.62	516.0	113842.87	5.0	21.00	47.0	1382.28	670.0	155529.77
6	2011.0	102.0	40283.62	518.0	113998.75	5.0	21.00	52.0	1801.29	677.0	156104.66
7	2012.0	103.0	40500.13	526.0	114933.44	5.0	21.00	59.0	2012.93	693.0	157467.50
8	2013.0	102.0	40500.13	532.0	117123.63	19.0	30.94	64.0	2232.61	717.0	159887.31
9	2014.0	103.0	40500.13	535.0	118290.66	43.0	58.22	64.0	2232.61	745.0	161081.62
10	2015.0	103.0	40500.13	541.0	118866.44	44.0	59.51	71.0	2548.82	759.0	161974.90
11	2016.0	103.0	40500.13	543.0	118917.71	45.0	59.66	72.0	2566.20	763.0	162043.70
12	2017.0	103.0	40500.13	544.0	118931.80	46.0	72.61	76.0	2587.95	769.0	162092.49
13	2018.0	104.0	40501.13	544.0	118931.80	46.0	72.61	77.0	2594.03	771.0	162099.47
14	2019.0	101.0	40564.43	553.0	119756.97	163.0	833.34	86.0	3858.25	903.0	165012.59
15	2020.0	104.0	43716.00	566.0	122420.00	214.0	1302.00	97.0	4483.00	981.0	171921.00
16	2021.0	106.0	44372.56	564.0	122509.33	219.0	1445.71	99.0	4726.24	987.0	173053.69

State-wise area of national parks

	state	State Area (km2)	No. of NP	Area National Parks (km2)
0	Andhra Pradesh	160229.0	3.0	1368.87
1	Arunachal Pradesh	83743.0	2.0	2290.82
2	Assam	78438.0	7.0	2664.58
3	Bihar	94163.0	1.0	335.65
4	Chhattisgarh	135191.0	3.0	2899.08
5	Goa	3702.0	1.0	107.00
6	Gujarat	196022.0	4.0	480.12
7	Haryana	44212.0	2.0	48.25
8	Himachal Pradesh	55673.0	5.0	2256.28
9	Jharkhand	79714.0	1.0	226.33
10	Karnataka	191791.0	5.0	2794.05
11	Kerala	38863.0	6.0	558.16
12	Madhya Pradesh	308245.0	11.0	4349.14
13	Maharashtra	307713.0	6.0	1273.60
14	Manipur	22327.0	2.0	140.00
15	Meghalaya	22429.0	2.0	267.48
16	Mizoram	21081.0	2.0	150.00
17	Nagaland	16579.0	1.0	202.02
18	Odisha	155707.0	2.0	99070.00
19	Punjab	50362.0	0.0	0.00
20	Rajasthan	342239.0	5.0	3947.07
21	Sikkim	7096.0	1.0	1784.00
22	Tamil Nadu	130058.0	5.0	827.51
23	Telangana	114840.0	3.0	19.62
24	Tripura	10486.0	2.0	36.71
25	Uttar Pradesh	240928.0	1.0	49530.00
26	Uttarakhand	53483.0	6.0	4915.02
27	West Bengal	88752.0	6.0	1981.48
28	Jammu and Kashmir	163090.0	4.0	2432.45

Predicted area of wild-life sanctuaries and existing area

	Area of Wild life sanctuaries (2022-2030)	Area of Wild life sanctuaries (2013-2021)
0	1.124686	117123.63
1	1.152678	118290.66
2	1.180669	118866.44
3	1.208661	118917.71
4	1.236652	118931.80
5	1.264644	118931.80
6	1.292635	119756.97
7	1.320627	122420.00
8	1.348619	122509.33

Problem Statement 1:

To Predict the significant growth of the biodiversity in "National Parks", "Wild life sanctuaries" and "Forest Reserves" of India during 2022-2035 by using data of duration of 2006-2021.(Linear regression)

Methodology:

The y value is predicted on the basis of x value whose line equation is as follows:

$$y-\bar{y}=\ddot{\beta}(x-\bar{x})$$

Where,

$$\hat{\beta}=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}=\frac{n\sum xy-\sum x\sum y}{n\sum x^2-(\sum x)^2}$$

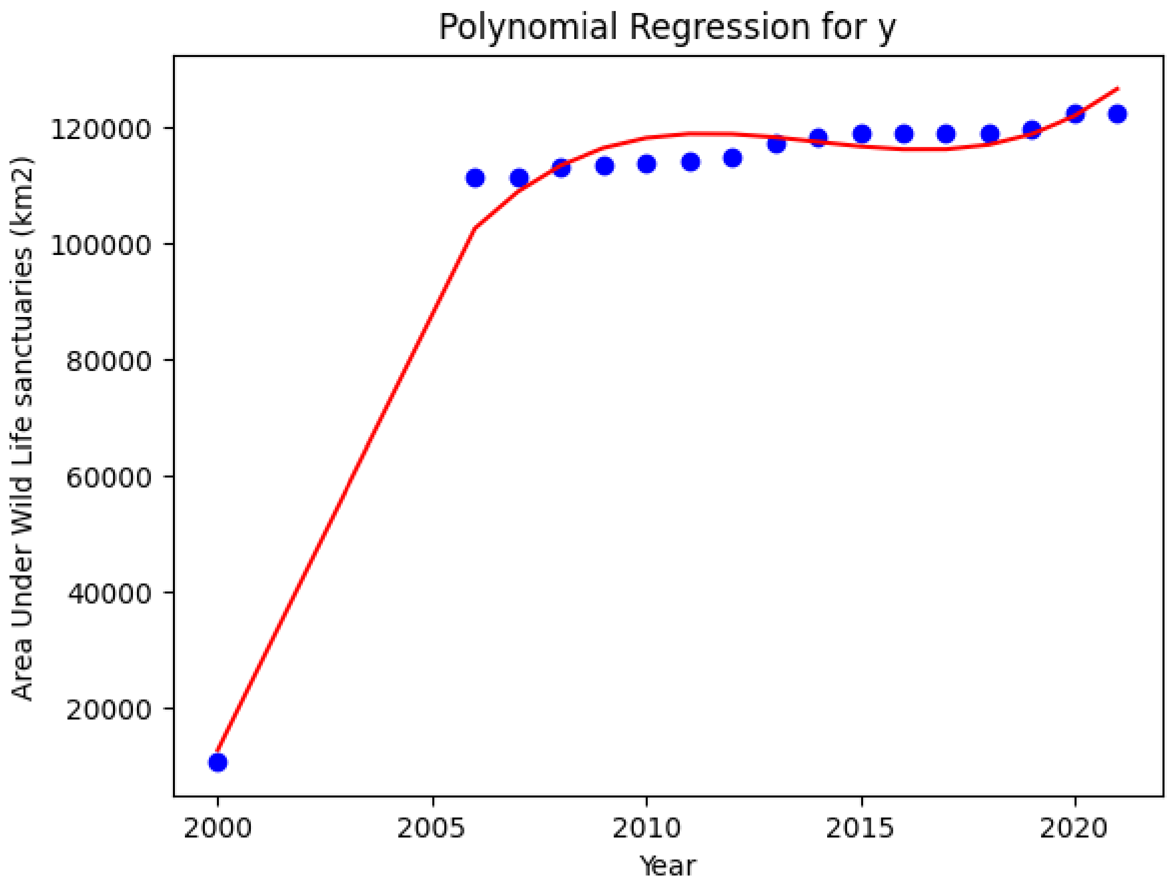
Area under wildlife sanctuaries

(2022-2035)

Output:

```
array([[136430.1251127 ],
       [139555.29530359],
       [142680.46549448],
       [145805.63568537],
       [148930.80587627],
       [152055.97606716],
       [155181.14625805],
       [158306.31644894],
       [161431.48663983]])
```

Graph:



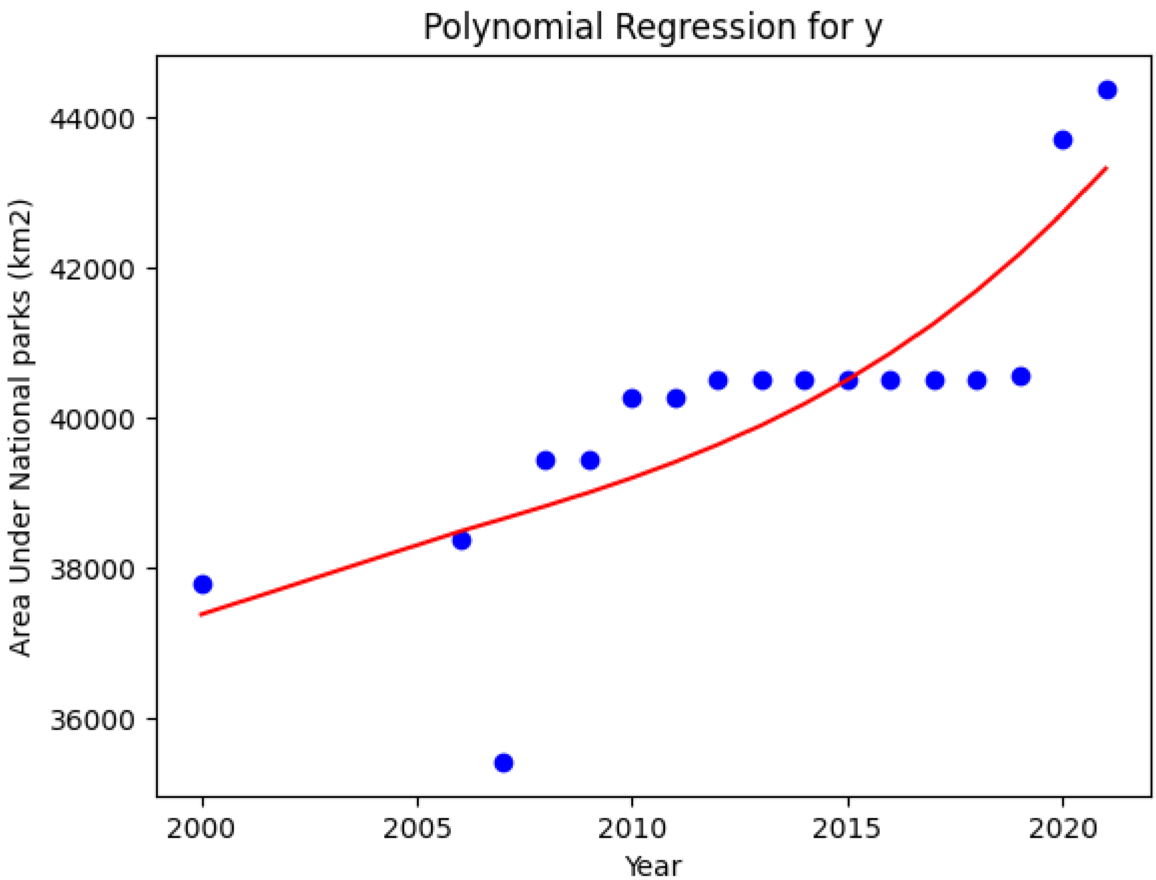
Area under National Parks

(2022-2035)

Output:

```
array([[42511.61302438],
       [42791.52580037],
       [43071.43857636],
       [43351.35135235],
       [43631.26412833],
       [43911.17690432],
       [44191.08968031],
       [44471.0024563 ],
       [44750.91523229]])
```

Graph:



Area under Conservation reserves

(2022-2035)

Output:

```
array([[4027.88185833],
       [4260.23704692],
       [4492.59223551],
       [4724.9474241 ],
       [4957.3026127 ],
       [5189.65780129],
       [5422.01298988],
       [5654.36817847],
       [5886.72336707]])
```

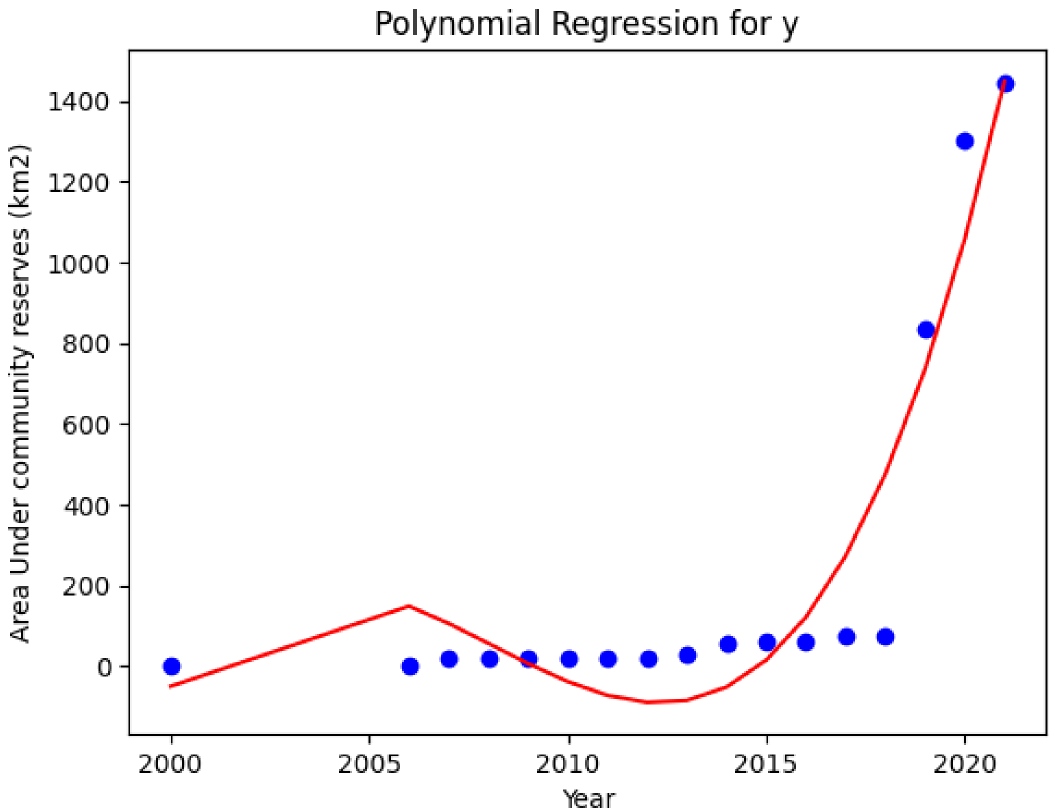
Area under reserves **Community**

(2022-2035)

Output:

```
array([[ 683.80932498],
       [ 737.45364075],
       [ 791.09795653],
       [ 844.74227231],
       [ 898.38658809],
       [ 952.03090386],
       [1005.67521964],
       [1059.31953542],
       [1112.9638512 ]])
```

Graph:



Conclusion:

As per the predicted results of significant growth of each section of 'Forests Reserves', we can precisely Observe that the growth of the 'Wild life sanctuaries' is way more than that of 'Community reserves' and 'Conservation Reserves'.

Problem Statement 2:

To designate all the states of India according to the area covered by national parks(km sqr) (power method).

Methodology:

Let the eigen values of a matrix A be 1, 2, 3,.... and n. If |1 | = |I |, for I = 2,..., n, 1 is considered the dominant eigen value of A. The eigenvectors that correlate to 1 are known as A's dominating eigenvectors. To use this method on a square matrix S, start with the first prediction for the dominating eigenvector. Multiply the resulting vector on the left by S, convert to conventional form, and continue until the desired eigenvector is not discovered or the results are no longer equivalent. If this happens, the dominating eigen value's absolute value will be taken as the norm of that eigen vectos.

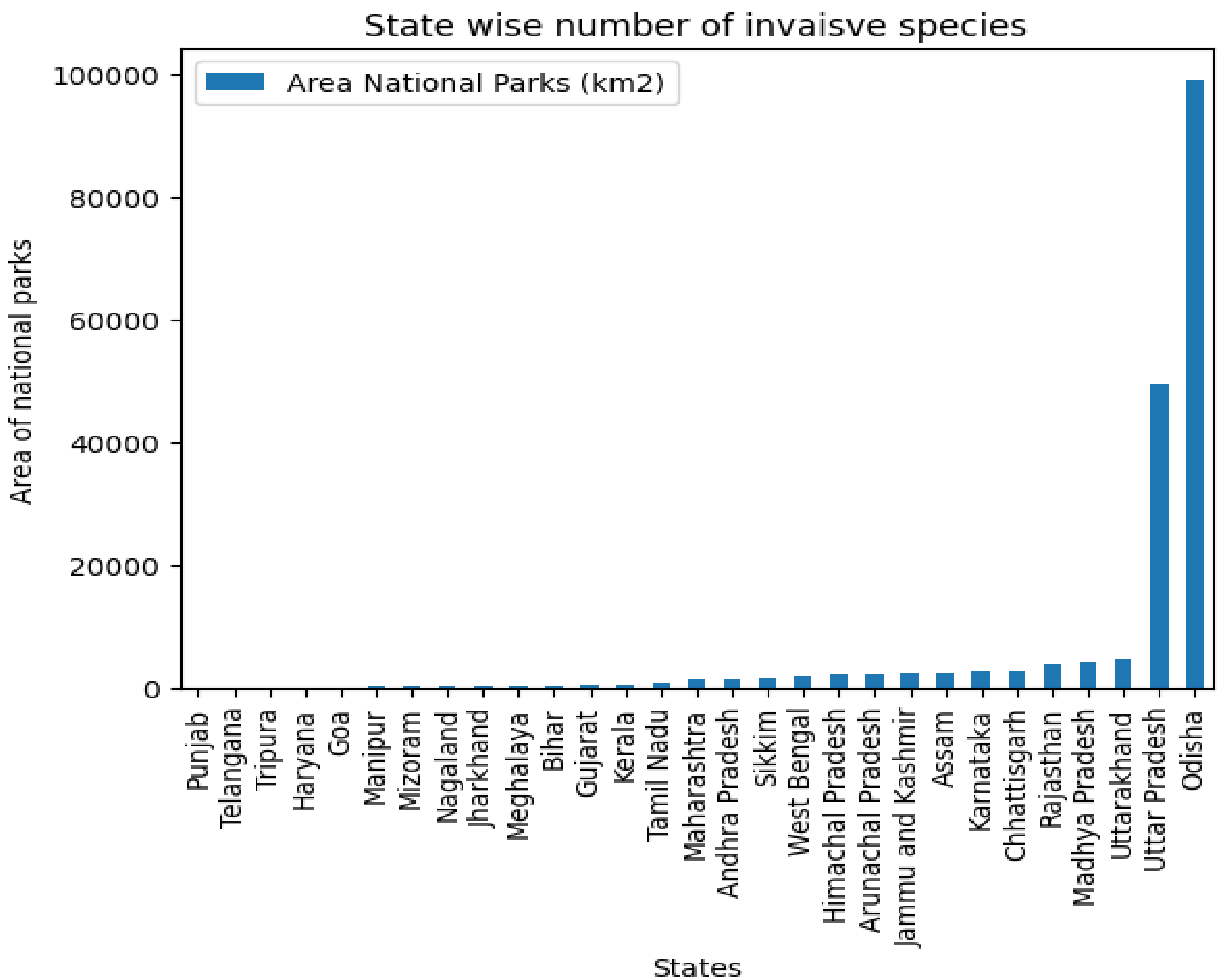
Steps:

- First, we create a 1, 2, and 3 matrix based on the data we've gathered.
- To determine a matrix's rank, we must first determine the matrix's dominant eigen value. The power approach can be used to accomplish this (described above).
- It will be difficult to calculate manually because the matrix is of higher order. As a result, we use Python to obtain the answer.
- We can utilize the dominant eigen value and corresponding eigenvector to find the rank based on our data after we have the dominant eigenvalue and corresponding eigenvector. In the dominating eigenvector, for example, the country with the greatest value will be ranked 1, the country with a lower value will be ranked 2, and so on.

Output:

```
Punjab
Telangana
Tripura
Haryana
Goa
Manipur
Mizoram
Nagaland
Jharkhand
Meghalaya
Bihar
Gujarat
Kerala
Tamil Nadu
Maharashtra
Andhra Pradesh
Sikkim
West Bengal
Himachal Pradesh
Arunachal Pradesh
Jammu and Kashmir
Assam
Karnataka
Chhattisgarh
Rajasthan
Madhya Pradesh
Uttarakhand
Uttar Pradesh
Odisha
```

Graph:



Conclusion:

Odisha has the highest area for national parks where as punjab do not have any area for national parks.

Problem Statement 3:

To find if there is any statistical significant difference in the means of the "Area of Wild life sanctuaries (2022-2030)" and "Area of Wild life sanctuaries (2012-2021)",as to observe if there will be any extension of Area covered by Wild life sanctuaries. (Test Hypothesis)

Methodology:

T-Test: It is a parametric statistical test which is used to compare and infer whether the two population's means differ from one another or not, when the standard deviation is not known. It is used when sample size is not large (n

Output:

```
t-statistic: 1.0720752915615275
p-value: 0.3149588721608356
Difference of means: 0.053639559042794216
Null Hypothesis is accepted
```

Conclusion:

As per the results(output of the code), we can draw a conclusion, due to 'P-value' being larger than Level of significance( ), that we are aware of accepting 'Null hypothesis', α because of the difference between the means of "Area of Wild life sanctuaries (2012-2021)" and "Area of Wild life sanctuaries (2022-2032)", being negligible(0.053...).

Objective 4

By 2030, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements**.**

Target 15.1.1
Forest area as a proportion of total land area

Dataset:

	State / UT	Geographical Area 2019	Very dense	Moderately dense	Open forest	no of workers in industry 19-20	Total forest area 2019	% of forested area	% change since 2019	Total forest area 2021	no of workers in 2020-21	Total forest area 2017	Total forest area 2013	x	Total Area under Desertification (ha) 1
0	Andhra Pradesh	162968	1994.00	13929.00	13861.00	663574	29784.00	0.18	0.02	29784	2139763	28147.00	46116.00	8.870867e+08	2298758
1	Arunachal Pradesh	83743	21058.00	30176.00	15197.00	2581	66431.00	0.79	0.00	66431	3687	66954.00	67321.00	4.413078e+09	153933
2	Assam	78438	3017.00	9991.00	15304.00	263106	28312.00	0.36	0.00	28312	429003	28105.00	27671.00	8.015693e+08	716596
3	Bihar	94163	333.00	3286.00	3762.00	128203	7381.00	0.08	0.01	7381	1082685	7299.00	7291.00	5.447916e+07	694809
4	Chhattisgarh	135192	7068.00	32279.00	16370.00	228927	55717.00	0.41	0.00	55717	531766	55547.00	55621.00	3.104384e+09	2211153
5	Delhi	1483	6.72	56.60	131.68	108232	195.00	0.13	0.00	195	626909	192.41	179.81	3.802500e+04	89868
6	Goa	3702	538.00	576.00	1130.00	74197	2244.00	0.61	0.00	2244	1638	2229.00	2219.00	5.035536e+06	192973
7	Gujarat	196244	378.00	5032.00	9516.00	2068743	14926.00	0.08	0.00	14926	1266676	14757.00	14653.00	2.227855e+08	10261641
8	Haryana	44212	28.00	445.00	1130.00	1023667	1603.00	0.04	0.00	1603	553399	1588.00	1586.00	2.569609e+06	338964
9	Himachal Pradesh	55673	3163.00	7100.00	5180.00	227975	15443.00	0.28	0.00	15443	130120	15100.00	14683.00	2.384862e+08	2394240
10	Jharkhand	79716	2601.00	9689.00	11431.00	205076	23721.00	0.30	0.00	23721	276070	23241.00	22538.00	5.626858e+08	7969607
11	Karnataka	191791	4533.00	20985.00	13212.00	1081116	38730.00	0.20	0.00	38730	1638703	23553.00	23473.00	1.500013e+09	5498726
12	Kerala	38852	1944.00	9472.00	9837.00	341411	21253.00	0.55	0.01	21253	1114661	37550.00	36132.00	4.516900e+08	6951000
13	Madhya Pradesh	308252	6665.00	34209.00	36619.00	413475	77493.00	0.25	0.00	38730	1344584	20321.00	17922.00	3.001304e+09	379587
14	Maharashtra	307713	8734.00	20589.00	21475.00	2038255	50798.00	0.17	0.00	21253	2051494	77414.00	77522.00	1.079610e+09	3804315
15	Manipur	22327	905.00	6228.00	9465.00	6898	16598.00	0.74	-0.01	77493	136811	50682.00	50632.00	1.286229e+09	13825935
16	Meghalaya	22429	560.00	9160.00	7326.00	12904	17046.00	0.76	0.00	50798	65586	17346.00	16990.00	8.659027e+08	601959
17	Mizoram	21081	157.00	5715.00	11948.00	1664	17820.00	0.85	-0.01	16598	24850	17146.00	17288.00	2.957764e+08	494880
18	Nagaland	16579	1272.00	4449.00	6530.00	5361	12251.00	0.74	-0.02	17046	56795	18186.00	19054.00	2.088305e+08	187453
19	Odisha	155707	7213.00	20995.00	23948.00	281911	52156.00	0.34	0.01	17820	923176	12489.00	13044.00	9.294199e+08	766678

Problem Statement 1:-

To find if there is any statistical significance diffrence in mean of the "total forest area 2017" and "total forest area 2021"

NULL HYPOTHESIS: This Hypothesis pressumes that there is no statistically significance difference between the means of "Total forest area 2017" and "Total forest area 2021".

ALTERNATE HYPOTHESIS: This Hypothesis presumes that there is an enough significant difference between the means of "Total forest area 2017" and "Total forest area 2021".

Output:-

Z-statistics= -0.09827759280363911

Accept Null Hypothesis

Conclusion:

As null hypothesis is accepted so the the difference between Total forest area 2017" and "Total forest area 2021 is very less.

Problem Statement 2:-

To correlate forest area and desertification

Methodology: Here I use correlation to justify that due to desertification forest area is decreasing

Output:-

0.25228107050528226

Conclusion:

There is only 30% area under desertification in India. Due to which value of correlation is low.

Problem Statement 3:-

To predict the total forest area from 2025 to 2030.

Methodology

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable

Output:

Output exceeds the size limit. Open the full output data in a text editor

```
0     24236.569269
1     34519.065211
2     23823.552269
3     17950.686207
4     31512.907504
5     15934.422506
6     16509.335437
7     20067.678918
8     16329.482246
9     20212.739916
10    22535.399381
11    26746.657645
12    21842.922507
13    26746.657645
14    21842.922507
```

15 37622.865526
16 30132.723654
17 20536.812359
18 20662.513185
19 20879.683809
20 19317.121529
21 30513.754284
22 16397.944303
23 20552.805544
24 16817.134112
...

Conclusion:

From this we can conclude that in future Total forest area will continuously fluctuate in most states of India..

Problem Statement 4:-

To compare the state on the basis of total forest area of 2021 and 2025

Methodology

Linear algebra ranking:

Power method:

This method is used to find the dominant eigen value and a corresponding eigenvector. For square matrix S, we first guess the initial eigen vector of dominant eigen value. Then we multiply the RANK USING POWER METHOD • First we convert data into matrix form and assign number according to the condition. To be predict the data using the linear regression first we have to.

Find the value of intercept and coefficient value.

Equation of the Linear regression is Y= a + bx

Output:

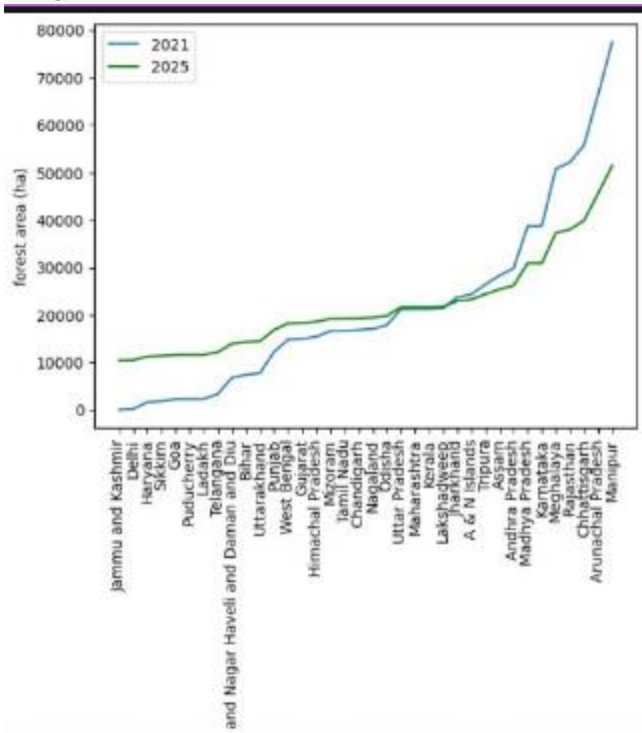
2021

State / UT	Geographical Area 2019	Very dense	Moderately dense	Open forest	no of workers in industry 19-20	Total forest area 2019	% of forested area	% change since 2019	Total forest area 2021	no of workers in 2020-21	Total forest area 2017	Total forest area 2013	x	Total Area under Desertification (ha) 1	2025	2029	rank
Jammu and Kashmir	54624	4155.00	8117.00	9115.00	66759	21387.00	0.39	0.00	23	152699	21358.00	23425.00	4.919010e+05	513428	10393.112620	15886.162367	1.0
Delhi	1483	6.72	56.60	131.68	108232	195.00	0.13	0.00	195	626909	192.41	179.81	3.802500e+04	89868	10484.221058	15934.422506	2.0
Haryana	44212	28.00	445.00	1130.00	1023667	1603.00	0.04	0.00	1603	553399	1588.00	1586.00	2.569609e+06	338964	11230.038786	16329.482246	3.0
Sikkim	7096	1102.00	1551.00	688.00	23039	3341.00	0.47	0.00	1847	24368	16572.00	16086.00	6.170827e+06	21526512	11359.285612	16397.944303	4.0
Goa	3702	538.00	576.00	1130.00	74197	2244.00	0.61	0.00	2244	1638	2229.00	2219.00	5.035536e+06	192973	11569.576554	16509.335437	5.0
Puducherry	490	0.00	17.53	35.77	50958	53.30	0.11	0.02	2272	35217	52.00	2486.00	1.210976e+05	242365	11584.408157	16517.191739	6.0
Ladakh	168055	2.00	512.00	1758.00	214	2272.00	0.01	0.01	2286	487	2252.00	22391.00	5.193792e+06	64286	11591.823958	16521.119890	7.0
Telangana	112077	1624.00	9119.00	10471.00	787105	21214.00	0.19	0.03	3341	943582	26281.00	23844.00	7.087597e+07	1543898	12150.657570	16817.134112	8.0
Dadra and Nagar Haveli and Daman and Diu	602	1.40	85.56	140.79	132695	227.75	0.38	0.00	6744	42012	21.56	17.26	1.535946e+06	654316	13953.227028	17771.955344	9.0
Bihar	94163	333.00	3286.00	3762.00	128203	7381.00	0.08	0.01	7381	1082685	7299.00	7291.00	5.447916e+07	694809	14290.645995	17950.686207	10.0
Uttarakhand	53483	5055.00	12768.00	6482.00	437816	24305.00	0.45	0.00	7722	529720	14679.00	14349.00	1.876832e+08	1528997	14471.273731	18046.364737	11.0
Punjab	50362	11.00	793.00	1043.00	661932	1847.00	0.04	0.00	12251	908576	51345.00	50347.00	2.262760e+07	5304114	16870.285508	19317.121529	12.0
West Bengal	88752	3037.00	4208.00	9587.00	736165	16832.00	0.19	0.00	14818	2169105	24295.00	24508.00	2.494166e+08	648253	18230.025679	20037.376040	13.0
Gujarat	196244	378.00	5032.00	9516.00	2068743	14926.00	0.08	0.00	14926	1266676	14757.00	14653.00	2.227855e+08	10261641	18287.233290	20067.678918	14.0
Himachal Pradesh	55673	3163.00	7100.00	5180.00	227975	15443.00	0.28	0.00	15443	130120	15100.00	14683.00	2.384862e+08	2394240	18561.088244	20212.739916	15.0
Mizoram	21081	157.00	5715.00	11948.00	1664	17820.00	0.85	-0.01	16598	24850	17146.00	17288.00	2.957764e+08	494890	19172.891866	20536.812359	16.0
Tamil Nadu	130060	3593.00	11034.00	11792.00	2663069	26419.00	0.20	0.00	16655	2018137	3344.00	3358.00	4.400084e+08	78749	19203.084772	20552.805544	17.0
Chandigarh	114	1.36	13.51	0.01	10167	22.88	0.20	0.04	16832	48252	6742.00	6711.00	3.851162e+05	1902456	19296.841691	20602.468594	18.0

2025

State / UT	Geographical Area 2019	Very dense	Moderately dense	Open forest	no of workers in industry 19-20	Total forest area 2019	% of forested area	% change since 2019	Total forest area 2021
Jammu and Kashmir	54624	4155.00	8117.00	9115.00	66759	21387.00	0.39	0.00	23
Delhi	1483	6.72	56.60	131.68	108232	195.00	0.13	0.00	195
Haryana	44212	28.00	445.00	1130.00	1023667	1603.00	0.04	0.00	1603
Sikkim	7096	1102.00	1551.00	688.00	23039	3341.00	0.47	0.00	1847
Goa	3702	538.00	576.00	1130.00	74197	2244.00	0.61	0.00	2244
Puducherry	490	0.00	17.53	35.77	50958	53.30	0.11	0.02	2272
Ladakh	168055	2.00	512.00	1758.00	214	2272.00	0.01	0.01	2286
Telangana	112077	1624.00	9119.00	10471.00	787105	21214.00	0.19	0.03	3341
Dadra and Nagar Haveli and Daman and Diu	602	1.40	85.56	140.79	132695	227.75	0.38	0.00	6744
Bihar	94163	333.00	3286.00	3762.00	128203	7381.00	0.08	0.01	7381
Uttarakhand	53483	5055.00	12768.00	6462.00	437816	24305.00	0.45	0.00	7722
Punjab	50362	11.00	793.00	1043.00	661932	1847.00	0.04	0.00	12251
West Bengal	88752	3037.00	4208.00	9587.00	736165	16832.00	0.19	0.00	14818
Gujarat	196244	378.00	5032.00	9516.00	2068743	14926.00	0.08	0.00	14926
Himachal Pradesh	55673	3163.00	7100.00	5180.00	227975	15443.00	0.28	0.00	15443
Mizoram	21081	157.00	5715.00	11948.00	1664	17820.00	0.85	-0.01	16598
Tamil Nadu	130060	3593.00	11034.00	11792.00	2663069	26419.00	0.20	0.00	16655
Chandigarh	114	1.36	13.51	0.01	10167	22.88	0.20	0.04	16832

State / UT	no of workers in 2020-21	Total forest area 2017	Total forest area 2013	x	Total Area under Desertification (ha) 1	2025	2029	ran
Jammu and Kashmir	152699	21358.00	23425.00	4.919010e+05	513428	10393.112620	15886.162367	1.
Delhi	626909	192.41	179.81	3.802500e+04	89868	10464.221038	15934.422506	2.
Haryana	553399	1588.00	1586.00	2.569609e+06	338964	11220.036786	16329.482246	3.
Sikkim	24368	16572.00	16086.00	6.170827e+06	21526512	11359.285612	16397.944303	4.
Goa	1638	2229.00	2219.00	5.035536e+06	192973	11569.576554	16509.335437	5.
Puducherry	35217	52.00	2486.00	1.210976e+05	242365	11584.408157	16517.191739	6.
Ladakh	487	2252.00	22391.00	5.193792e+06	64288	11591.823958	16521.119890	7.
Telangana	943582	26281.00	23844.00	7.087597e+07	1543898	12150.657570	16817.134112	8.
Dadra and Nagar Haveli and Daman and Diu	42012	21.56	17.26	1.535946e+06	654316	13953.227028	17771.955344	9.
Bihar	1082685	7299.00	7291.00	5.447916e+07	694809	14290.645995	17950.686207	10.
Uttarakhand	529720	14679.00	14349.00	1.876832e+08	1528997	14471.273731	18046.364737	11.
Punjab	908576	51345.00	50347.00	2.262760e+07	5304114	16970.285508	19317.121529	12.
West Bengal	2169105	24295.00	24508.00	2.494166e+08	648253	18230.025679	20037.376040	13.
Gujarat	1266676	14757.00	14653.00	2.227855e+08	10261641	18287.233290	20067.678918	14.
Himachal Pradesh	130120	15100.00	14683.00	2.384862e+08	2394240	18561.098244	20212.739916	15.
Mizoram	24850	17146.00	17288.00	2.957764e+08	494880	19172.891866	20536.812359	16.
Tamil Nadu	2018137	3344.00	3358.00	4.400084e+08	78749	19203.084772	20552.805544	17.
Chandigarh	48252	6742.00	6711.00	3.851162e+05	1902456	19296.841691	20602.468594	18.



Conclusion:

From this graph we can see that there is drastic change in total forest area 2025.

Appendix

Objective 1

Problem Statement 1

```
# Ranking of species that got extinct in 1900-1960

species=ex['Name']

def answer(x):          # Power method

    e_value = abs(x).max()       # Eigenvalue

    e_vector = x / x.max()      # Eigenvector

    return e_value, e_vector



def rank(p):

  lol=np.empty([68,68])

  for v in range(len(p)):

    for r in range(len(p)):

      if p[v]>p[r]:

        lol[v][r] = 3

      elif p[v] == p[r]:

        lol[v][r]=2

      elif p[v] < p[r]:

        lol[v][r]=1

  global matrix

  matrix = pd.DataFrame(lol)

  x = np.random.randint(1,2,68).reshape(68,1)

  for i in range(1):

    x = np.dot(matrix, x)

    lambda_1, x = answer(x)

  a = sorted(list(x),reverse=True)
```

```
    print(a)
    for i in list(a):
        if i in list(x):
            c = list(x).index(i)
            print(species[c])
rank(ex['EX(1900-1960)'])
```

Alternative method for Problem Statement 1

```
# Ranking of species that got extinct in 1960-2020 using sort we can check the results of power method
ex['rank']=ex['EW(1960-2020)'].rank(method="min",ascending=False)
ex.sort_values('rank',inplace=True)
ex
```

Problem Statement 2

```
#For Madhya Pradesh
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression


tigg= pd.read_csv('C:\\Users\\hp\\Desktop\\CDA Main Project\\Tiger.csv')


X = tigg['Year'].values.reshape(-1,1)
y = tigg['Madhya Pradesh'].values.reshape(-1,1)
```

```
poly = PolynomialFeatures(degree=3)

X_poly = poly.fit_transform(X)

model = LinearRegression()

model.fit(X_poly, y)

future_years = np.array([[i] for i in range(2022, 2031)])

future_years_poly = poly.fit_transform(future_years)

future_predictions = model.predict(future_years_poly)

plt.scatter(X, y, color='blue')

plt.plot(X, y, color='red')

plt.plot(future_years, future_predictions, color='red')

plt.scatter(future_years, future_predictions, color='red')

plt.xlabel('Year')

plt.ylabel('Tiger Population')

plt.title('Tiger Population Predictions Madhya Pradesh')

plt.show()
```

Problem Statement 3

```
y=poaching['EX(1900-1960)']

x=poaching['EW(1900-1960)']

Stx_4=np.sum(x**2)-((np.sum(x)**2)/x.shape[0])
```

```
print(Stx_4)

Stx_4 = np.sum(x*2) - ((np.sum(x)*2)/x.shape[0])

print(Stx_4)

Sty_4= np.sum(y*2) - ((np.sum(y)*2)/(x.shape[0]))

print(Sty_4)

Sr_4 = np.sum(y*x) - (np.sum(x) * np.sum(y)/y.shape[0])

print(Sr_4)

corr_4 = Sr_4/np.sqrt(Stx_4*Sty_4)

print(corr_4)
```

```
col1 = poaching['EX(1900-1960)']

col2 = poaching['EW(1900-1960)']

correlation = np.corrcoef(col1, col2)[0, 1]

print(correlation)


col1 = poaching['EX(1960-2020)']

col2 = poaching['EW(1960-2020)']

correlation = np.corrcoef(col1, col2)[0, 1]

print(correlation)
```

Heat Map

```
import matplotlib.pyplot as plt
import numpy as np

data =poaching.corr()
plt.imshow(data, cmap='hot')
plt.colorbar()
plt.show()

import seaborn as sns
import numpy as np

data = poaching.corr()
sns.heatmap(data.corr(), annot=True)


import plotly.express as px
import numpy as np

data = poaching.corr()
fig = px.imshow(data, color_continuous_scale='hot',)
fig.show()
```

Comparison of Number of Species gone Extinct between 1900-1960 & 1960-2020

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
plt.subplots(figsize=(16,25))
plt.scatter(x=ex['EX(1900-1960)'], y=ex['Name'])
plt.plot(ex['EX(1900-1960)'],ex['Name'])
plt.scatter(x=ex['EX(1960-2020)'], y=ex['Name'])
plt.plot(ex['EX(1960-2020)'],ex['Name'])
plt.show()
```

Tiger Population

```
tigg=pd.read_csv("C:\\Users\\hp\\Desktop\\CDA Main Project\\Tiger.csv")


plt.subplots(figsize=(16,25))
plt.scatter(x=tigg['Year'], y=tigg['Total'])
plt.plot(tigg['Year'],tigg['Total'])
plt.xlabel('Years')
plt.ylabel('Tiger Population')
plt.title('Tiger Population in India')

plt.show()
```

Poaching Data

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
```

```
ex=pd.read_csv("C:\\Users\\hp\\Desktop\\CDA Main Project\\Poaching-Data.csv")

desert = pd.concat([ex['Mean 2002-2006'], ex['Species']], axis=1)
f, ax = plt.subplots(figsize=(16, 8))
fig = sns.boxplot(x=ex['Mean 2002-2006'], y=ex['Species'])
```

Objective 3

#LAND DEGRADATION AND DESERTIFICATION IN INDIA

```
#Importing important Librarires
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
#Importing dataset
degrade=pd.read_csv('C:\\Users\\hp\\Downloads\\Partially Normalized Data(new).csv')
degrade
```

NORMALIZING THE DATA

```
import pandas as pd
degrade = pd.read_csv("C:\\Users\\hp\\Downloads\\desertification.csv")

#Select the columns that you want to normalize
columns_to_normalize = ["Vegetation Degradation 2011-13","Vegetation Degradation 2003-05","Water
Erosion 1","Water Erosion 2"
,"Wind Erosion 1","Wind Erosion 2","Water Logging 1","Water Logging 2",
"Total Area under Desertification (ha) 1","Total Area under Desertification (ha) 2",
"Total Area under Desertification (%) 1","Total Area under Desertification (%) 2","Manmade 1","Manmade
2"]

#Normalizing the selected columns using the min-max normalization method
degrade[columns_to_normalize] = (degrade[columns_to_normalize] - degrade[columns_to_normalize].min()) /
(degrade[columns_to_normalize].max() - degrade[columns_to_normalize].min())

#Saving the normalized data back to the CSV file
degrade.to_csv("Normalized Data(new)1.csv", index=False)
```

PROBLEM STATEMENT 1

```
state = degrade["State Name"]
def answer(x):          # Power method
        e_value = abs(x).max()       # Eigenvalue
        e_vector = x / x.max()      # Eigenvector
        return e_value , e_vector

def rank(p):
    lol=np.empty([30,30])
    for v in range(len(p)):
        for r in range(len(p)):
            if p[v]>p[r]:
                lol[v][r] = 3
            elif p[v] == p[r]:
                lol[v][r]=2
            elif p[v] < p[r]:
                lol[v][r]=1
    global matrix
    matrix = pd.DataFrame(lol)
    #print(matrix)
    x = np.random.randint(1,2,30).reshape(30,1)
    for i in range(200):
        x = np.dot(matrix, x)
        lambda_1, x = answer(x)
    #print(lambda_1)
    #print(x)
    a = sorted(list(x),reverse=False)
    #print(a)
    for i in list(a):
        if i in list(x):
            c = list(x).index(i)
            print(state[c])


rank(degrade['Total Area under Desertification (ha) 1'])
answer(x = degrade['Total Area under Desertification (ha) 1'])




#Graph
plt.figure(figsize=(10,6))
X = degrade["State Name"]
```

```
X_axis = np.arange(len(X))
plt.bar(X_axis,degrade['Total Area under Desertification (ha) 1'],0.3,color="blue")
plt.xticks(X_axis,X)
plt.xlabel("STATES",fontsize="10",)
plt.ylabel("Area under desertfication in 2011-13",fontsize='10')
plt.title("Total area covered under desertification",fontsize='10')
plt.xticks(rotation=90)
plt.show()
```

PROBLEM STATEMENT 2

```
#Correlating Vegetation Degradation to Water Erosion.
#Start of Rgression
y = degrade['Water Erosion 1']
x = degrade['Vegetation Degradation 2011-13']
Sxx_1 = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
Sxx_1

Syy_1= np.sum(y**2) - ((np.sum(y)**2)/x.shape[0])
Syy_1

Sxy_1 = np.sum(y*x) - (np.sum(x) * np.sum(y))/y.shape[0]
Sxy_1

#Correlation between Vegetation Degradation and Water Erosion in 2011-13
corr_1 = Sxy_1/np.sqrt(Sxx_1*Syy_1)
corr_1

#Start of Rgression
y = degrade['Water Erosion 2']
x = degrade['Vegetation Degradation 2003-05']
Sxx_2 = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
Sxx_2

Syy_2= np.sum(y**2) - ((np.sum(y)**2)/x.shape[0])
Syy_2

Sxy_2 = np.sum(y*x) - (np.sum(x) * np.sum(y))/y.shape[0]
Sxy_2
```

```
#Correlation between Vegetation Degradation and Water Erosion in 2003-05
corr_2 = Sxy_2/np.sqrt(Sxx_2*Syy_2)
corr_2

import matplotlib.pyplot as plt
plt.scatter(x=degrade['Vegetation Degradation 2011-13'], y=degrade['Water Erosion 1'],color = 'Red')
plt.show()

#Start of Rgression
y = degrade['Wind Erosion 1']
x = degrade['Vegetation Degradation 2011-13']
Sxx_3 = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
Sxx_3
Syy_3= np.sum(y**2) - ((np.sum(y)**2)/x.shape[0])
Syy_3
Sxy_3 = np.sum(y*x) - (np.sum(x) * np.sum(y))/y.shape[0]
Sxy_3
#Correlation between Vegetation Degradation and Wind Erosion in 2011-13
corr_3 = Sxy_3/np.sqrt(Sxx_3*Syy_3)
corr_3



#Start of Rgression
y = degrade['Wind Erosion 2']
x = degrade['Vegetation Degradation 2003-05']
Sxx_4 = np.sum(x**2) - ((np.sum(x)**2)/x.shape[0])
Sxx_4

Syy_4= np.sum(y**2) - ((np.sum(y)**2)/x.shape[0])
Syy_4

Sxy_4 = np.sum(y*x) - (np.sum(x) * np.sum(y))/y.shape[0]
Sxy_4

#Correlation between Vegetation Degradation and Wind Erosion in 2003-05
corr_4 = Sxy_4/np.sqrt(Sxx_4*Syy_4)
corr_4

import matplotlib.pyplot as plt
plt.scatter(x=degrade['Vegetation Degradation 2011-13'], y=degrade['Wind Erosion 1'],color = 'Red')
plt.show()
```

```
plt.figure(figsize=(10,6))
X = degrade["State Name"]
X_axis = np.arange(len(X))
plt.bar(X_axis,degrade['Wind Erosion 1'],0.3,color="blue")
plt.bar(X_axis+0.3,degrade['Water Erosion 1'],0.3,color="red")
plt.xticks(X_axis, X)
plt.xlabel("STATES",fontsize="10",)
plt.ylabel("Vegetation degradation in 2011-13",fontsize='10')
plt.title("Effects on Vegetation Degradation",fontsize='10')
plt.xticks(rotation=90)
plt.show()
```

PROBLEM STATEMENT 3

```
import pandas as pd
from scipy.stats import ttest_ind
from scipy import stats

'''In PS-3, we now can assume our level of Significance('α') equal to 5% of the complete dataset.
Hence its value would be [0.05].'''

# load degrade into a DataFrame
degrade = pd.read_csv("C:\\Users\\hp\\Downloads\\Partially Normalized Data(new).csv")

# perform t-test
t_value, p_value = stats.ttest_ind(degrade['Vegetation Degradation 2011-13'], degrade['Vegetation
Degradation 2003-05'])

# print results of t-test
print("p-value:",p_value)
print("t-value:",t_value)

# calculate difference mean of after t-test
column1_mean = degrade['Vegetation Degradation 2011-13'].mean()
column2_mean = degrade['Vegetation Degradation 2003-05'].mean()
print('Difference of means:', column1_mean - column2_mean)
if -3.25 < t_value < 3.25:
    print("Null Hypothesis is accepted")
else:
    print("We've to switch to alternate hypothesis")
```

Objective 3

Problem Statement 1

```
import sklearn as sklearn
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

x = nparks['Year'].values.reshape(-1, 1)
y = nparks['Area Under Wild Life sanctuaries (km2)'].values.reshape(-1, 1)

# Create and fit the linear regression model
reg = LinearRegression().fit(x,y)

# Make predictions for the years from 2022 to 2030
years_to_predict = np.array(range(2021,2030)).reshape(-1, 1)
degradation_predictions = reg.predict(years_to_predict,)
degradation_predictions


from sklearn.preprocessing import PolynomialFeatures
# Reshape the data
x= x.reshape(-1,1)
y = y.reshape(-1,1)

# Transform the data to include polynomial terms
poly = PolynomialFeatures(degree=3)
year_poly = poly.fit_transform(x)

# Create and fit the model
model = LinearRegression()
model.fit(year_poly, y)

# Make predictions
y_pred = model.predict(year_poly)

# Plot the results
plt.scatter(x, y, color='blue')
plt.plot(x, y_pred, color='red')
plt.xlabel('Year')
plt.ylabel('Area Under Wild Life sanctuaries (km2)')
```

```
plt.title('Polynomial Regression for y')
plt.show()



#Mean Squared Error

from sklearn.metrics import mean_squared_error

mse = mean_squared_error(y, y_pred)
print("Mean Squared Error: ", mse)


from sklearn.linear_model import LinearRegression

x = nparks['Year'].values.reshape(-1, 1)
y = nparks['Area Under National parks (km2)'].values.reshape(-1, 1)

# Create and fit the linear regression model
reg = LinearRegression().fit(x,y)

# Make predictions for the years from 2022 to 2030
years_to_predict = np.array(range(2021,2030)).reshape(-1, 1)
degradation_predictions = reg.predict(years_to_predict,)
degradation_predictions


from sklearn.preprocessing import PolynomialFeatures
# Reshape the data
x= x.reshape(-1,1)
y = y.reshape(-1,1)

# Transform the data to include polynomial terms
poly = PolynomialFeatures(degree=3)
year_poly = poly.fit_transform(x)

# Create and fit the model
model = LinearRegression()
model.fit(year_poly, y)

# Make predictions
y_pred = model.predict(year_poly)

# Plot the results
plt.scatter(x, y, color='blue')
plt.plot(x, y_pred, color='red')
```

```
plt.xlabel('Year')
plt.ylabel('Area Under National parks (km2)')
plt.title('Polynomial Regression for y')
plt.show()



from sklearn.linear_model import LinearRegression

x = nparks['Year'].values.reshape(-1, 1)
y = nparks['Area Under Conservation Reserves (km2)'].values.reshape(-1, 1)

# Create and fit the linear regression model
reg = LinearRegression().fit(x,y)

# Make predictions for the years from 2022 to 2030
years_to_predict = np.array(range(2021,2030)).reshape(-1, 1)
degradation_predictions = reg.predict(years_to_predict,)
degradation_predictions


from sklearn.preprocessing import PolynomialFeatures
# Reshape the data
x= x.reshape(-1,1)
y = y.reshape(-1,1)

# Transform the data to include polynomial terms
poly = PolynomialFeatures(degree=2)
year_poly = poly.fit_transform(x)

# Create and fit the model
model = LinearRegression()
model.fit(year_poly, y)

# Make predictions
y_pred = model.predict(year_poly)

# Plot the results
plt.scatter(x, y, color='blue')
plt.plot(x, y_pred, color='red')
plt.xlabel('Year')
plt.ylabel('Area Under Conservation Reserves (km2)')
plt.title('Polynomial Regression for y')
plt.show()
```

```
from sklearn.linear_model import LinearRegression

x = nparks['Year'].values.reshape(-1, 1)
y = nparks['Area Under Community reserves (km2)'].values.reshape(-1, 1)

# Create and fit the linear regression model
reg = LinearRegression().fit(x,y)

# Make predictions for the years from 2022 to 2030
years_to_predict = np.array(range(2021,2030)).reshape(-1, 1)
degradation_predictions = reg.predict(years_to_predict,)
degradation_predictions
```

```
from sklearn.preprocessing import PolynomialFeatures
# Reshape the data
x= x.reshape(-1,1)
y = y.reshape(-1,1)

# Transform the data to include polynomial terms
poly = PolynomialFeatures(degree=3)
year_poly = poly.fit_transform(x)

# Create and fit the model
model = LinearRegression()
model.fit(year_poly, y)

# Make predictions
y_pred = model.predict(year_poly)

# Plot the results
plt.scatter(x, y, color='blue')
plt.plot(x, y_pred, color='red')
plt.xlabel('Year')
```

```
plt.ylabel('Area Under community reserves (km2)')
plt.title('Polynomial Regression for y')
plt.show()
```

PROBLEM STATEMENT 2 Ranking

```
#Dataset 2
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
aparks = pd.read_csv("C:\\Users\\Rajesh\\Desktop\\Area Parks.csv")
aparks
```

```
state = aparks['state']
def answer(x):          # Power method
        e_value = abs(x).max()       # Eigenvalue
        e_vector = x / x.max()      # Eigenvector
        return e_value, e_vector

def rank(p):
    f=np.empty([29,29])
    for i in range(len(p)):
        for j in range(len(p)):
            if p[i]>p[j]:
                f[i][j] = 3
            elif p[i] == p[j]:
                f[i][j]=2
            elif p[i] < p[j]:
                f[i][j]=1
    global matrix
    matrix = pd.DataFrame(f)
    #print(matrix)
    x = np.random.randint(1,2,29).reshape(29,1)
    for i in range(200):
        x = np.dot(matrix, x)
```

```
        lambda_1, x = answer(x)
    #print(lambda_1)
    #print(x)
    a = sorted(list(x),reverse=False)
    #print(a)
    for i in list(a):
        if i in list(x):
            c = list(x).index(i)
            print(state[c])
rank(aparks['Area National Parks (km2)'])
```

Verification by Rank function

```
aparks['rank']=aparks['Area National Parks (km2)'].rank()
aparks.sort_values('rank', inplace=True)
aparks
```

PROBLEM STATEMENT 3

```
import pandas as pd
nparks = pd.read_csv("C:\\Users\\Rajesh\\Desktop\\15.9(appended).csv")
nparks
nparks.corr()

import pandas as pd
from scipy.stats import ttest_rel

'''In every peice of code in PS-3, we now can assume our
level of Significance('α') equal to 5% of the complete dataset.
Hence its value would be [0.05].'''



# load parks into a DataFrame
parks = pd.read_csv("C:\\Users\\Rajesh\\Desktop\\Normalized Data1.csv")

# perform t-test
t_stat, p_value =
ttest_rel([0,0.124999994,0.249999989,0.374999983,0.499999978,0.624999972,0.749999967,0.874999961,1],
```

```
[0,0.216690495,0.323599532,0.333119186,0.335735373,0.335735373,0.488950369,0.983413484,1])

# print results of t-test
print("t-statistic:", t_stat)
print("p-value:", p_value)

# calculate difference mean of after t-test
column1_mean = parks['Area of Wild life sanctuaries (2013-2021)'].mean()
column2_mean = parks['Area of Wild life sanctuaries (2022-2030)'].mean()
print('Difference of means:', column2_mean - column1_mean)
if p_value>0.05:
    print("Null Hypothesis is accepted")
else:
    print("We've to switch to alternate hypothesis")
```

Objective 4

PROBLEM STATEMENT 1

```
import pandas as pd

df = pd.read_csv("C:\\Users\\Amit\\OneDrive\\Desktop\\sdg 15.1.csv")


#Select the columns that you want to normalize

columns_to_normalize = ["Total forest area 2017","Total forest area 2021"]


#Normalizing the selected columns using the min-max normalization method

df[columns_to_normalize] = (df[columns_to_normalize] - df[columns_to_normalize].min()) /
(df[columns_to_normalize].max() - df[columns_to_normalize].min())


#Saving the normalized data back to the CSV file

df.to_csv("Normalized Data(new)1.csv",index=False)


from statsmodels.stats.weightstats import ztest as ztest



data_group1 = df['Total forest area 2017']

data_group2 = df['Total forest area 2021']


z_stat,p_val= ztest(data_group1,data_group2, value=0)


print("Z-statistics=",z_stat)


#Level of significance=a

a=0.05
```

```
if z_stat >-1.96 and z_stat <1.96:
    print("Accept Null Hypothesis")
else:
    print("Accept Alternate Hypothesis")
```

PROBLEM STATEMENT 2

```
y=df['Total forest area 2017']
x=df['Total Area under Desertification (ha) 1']
Stx_4=np.sum(x*2)-((np.sum(x)*2)/x.shape[0])
print(Stx_4)
Stx_4 = np.sum(x*2) - ((np.sum(x)*2)/x.shape[0])
print(Stx_4)
Sty_4= np.sum(y*2) - ((np.sum(y)*2)/(x.shape[0]))
print(Sty_4)
Sr_4 = np.sum(y*x) - (np.sum(x) * np.sum(y)/y.shape[0])
print(Sr_4)
corr_4 = Sr_4/np.sqrt(Stx_4*Sty_4)
print(corr_4)
```

```
col1 = df['Total forest area 2017']
col2 = df['Total Area under Desertification (ha) 1']
correlation = np.corrcoef(col1, col2)[0, 1]
```

```
print(correlation)
```

PROBLEM STATEMENT 3

```
from sklearn.linear_model import LinearRegression
regr=LinearRegression()
y = df['Total forest area 2021']
x = df['Total forest area 2017']
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
y = df['Total forest area 2021']
x = df['Total forest area 2017']
```

```
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
plt.plot(x, y, 'o')

#add linear regression line to scatterplot
plt.plot(x,b_xy*x+a)
plt.xlabel('Total forest area 2017')
plt.ylabel('Total forest area 2021')

y2=a+b_xy*df['Total forest area 2021']
df['2025']=y2
y2
y2=a+b_xy*df['2025']
df['2029']=y2
y2
```

POBLEM STATEMENT 4

```
df['rank']=df['Total forest area 2021'].rank()
df.sort_values('rank',inplace=True)
df

df['rank']=df['2029'].rank()
df.sort_values('rank',inplace=True)
df

plt.plot(df['State / UT'],df['Total forest area 2021'])
plt.plot(df['State / UT'],df['2025'],color='green')
plt.xlabel('states')
plt.ylabel('forest area (ha)')
plt.legend(['2021','2025'])
plt.xticks(rotation=90)
plt.show()
```

References:

Reid, A. J., Brooks, J. L., Dolgova, L., Laurich, B., Sullivan, B. G., Szekeres, P., ... & Cooke, S. J. (2017). Post-2015 Sustainable Development Goals still neglecting their environmental roots in the Anthropocene. *Environmental Science & Policy*, *77*, 179-184.

Karanth, K. U., & Nichols, J. D. (1998). Estimation of tiger densities in India using photographic captures and recaptures. *Ecology*, *79*(8), 2852-2862.

Eswaran, H., Lal, R., & Reich, P. F. (2019). Land degradation: an overview. *Response to land degradation*, 20-35.

Reddy, V. R. (2003). Land degradation in India: Extent, costs and determinants. *Economic and Political Weekly*, 4700-4713.

Keenan, R. J., Reams, G. A., Achard, F., de Freitas, J. V., Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, *352*, 9-20.