Major Project Report on

Exploratory data analysis and forecasting the Indian weather using machine learning.

Submitted By

Shubham Kaushik

Roll no: 2K19/DMBA/094

Under the Guidance of

Mr. Mohit Beniwal

Assistant Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

CERTIFICATE

This is to certify that the work titled 'Exploratory data analysis and forecasting the Indian weather using machine learning' as part of the final year Major Research Project submitted by Shubham Kaushik in the 4th Semester of MBA, Delhi School of Management, Delhi Technological University during January-May 2021 is his original work and has not been submitted anywhere else for the award of any credits/ degree whatsoever.

The project is submitted to Delhi School of Management, Delhi Technological University in partial fulfillment of the requirement for the award of the degree of Master of Business Administration.

ACKNOWLEDGMENT

It is a great pleasure for me to acknowledge the kind of help and guidance received during the research work. I would like to thank my faculty advisor Assistant Prof. Mohit Beniwal, who helped me to take up the topic ‘Exploratory data analysis and forecasting the Indian weather using machine learning’ and guided me to complete this project properly.

I am highly indebted to Delhi School of Management, Delhi Technological University for giving me an opportunity to work on this project. Lastly, I would like to express my gratitude to all the honorable faculty members for sharing their experience and expertise on this project. I have put all my efforts to ensure that the project is completed in the best possible manner and also ensured that the project is error-free.

Contents

ABSTRACT

India has a typical weather conditions consisting of various seasons and geographical conditions. Country has extreme high temperatures at Rajasthan desert, cold climate at Himalayas and heavy rainfall at Chirapunji. These extreme variations in temperatures make us to feel difficult in inferring / predictions of weather effectively. It requires higher scientific techniques / methods like machine learning algorithms applications for effective study and predictions of weather conditions.

Purpose: The purpose of this study is to do an exploratory data analysis and prediction of mean temperature and rainfall in India.

Methodology/Approach: The machine learning algorithms like Decision Tree Regression and Linear regression are used in this project to predict the temperature and rainfall. The exploratory data analysis is done using libraries of Python like matplotlib and plotly.

Research Limitations: The dataset is limited only to the year between 1901 to 2017.

Value: The insights will be used to analyze the weather and forecast it for optimizing the agricultural activities of our country.

1. Introduction

The foundation of Indian economy is the agriculture industry; each individual entail nourishment for sustaining. The agriculture industry ought to be aided, which will be helpful for them to acknowledge what all respect creates under numerous conditions. Agrobusiness depends upon work just as on several perspectives for example water, sort of soil, composts, weather, etc

Temperature and deluge assessing has been maybe the fundamental segments considered in climate influence focuses on spaces of agribusiness, vegetation, water resources and the movement business. It is essential for a country like India which is altogether dependent after farming. According to the new ecological change influence assessment thinks about structure, agribusiness, vegetation, water, resources and the movement business are the regions affected clearly by environment changes.

The economic development of every year relies upon the measure of term of monsoon rain, hence if the monsoon is bad it can lead to destruction of agriculture, annihilation of certain harvests, which may bring about shortage of some farming items which thus can cause food swelling, shortage, famine and public turmoil. In our examination we are attempting to comprehend the conduct of rainfall in India throughout the long term.

The analysis in this project will be done by using mean temperature and rainfall datasets provided by the Indian government website.

2. Literature Review

In India a few investigations have been completed to decide the progressions on relationship of temperature with environmental modification. Pramanik and Jagannathan (1954) concentrated that common patterns can be observed on the yearly average greatest & least temperatures throughout our country remained. They inferred that it is obsereved no broad inclination can be seen increment as well as lessening in these readings. The breaking down of the arrangement of time of average yearly readings done a bunch of 8 weather stations of India by Jagannathan and Parathasarathy (1973). It was a detailed and expanding pattern in the average yearly weather readings of Allahabad, Bangalore, Mumbai, Kolkata and a diminishing pattern at Chennai. Sarker and Thapliyal (1988) looked into environmental alteration of the past 70 years and demonstrated a minor heating pattern, dry-bulb temperature has an estimation of patterns of temperature; although precipitation displayed no pattern.

Srivastava et al. (1992) investigated the patterns of decadal in environment all across India provided the primary signs of the daytime deviation of weather patterns across India is very not moderately the same as which saw across numerous different pieces around the world. Srivastava et al. (1992) tracked down that greatest reading display a lot bigger expanding patterns as compared to the base temperatures, across a significant piece of the nation. The diminishing pattern in reading of temperature across the vast majority on the northerly pieces of the nation, combined through an expanding pattern across southerly parts, and an in general marginally expanding pattern of the request for 0.30°C remained during the most recent century.

It was shown by Rupa Kumar et al. (1994) that the nationwide average greatest temperature has ascended by 0.5°C, and the average least temperature was diminished by 0.2°C. Notwithstanding, as per outcome from average least temperature reading isn't genuinely critical, they reasoned that a large portion in expansions in average temperature of air and surface above India are because of the increment in daylight reading. Kumar and Gasp (1998) dissected the information for 1880–1996 and exhibited that a huge heating pattern of 0.60°C was observed each century. The extent of heating is increasing in the post-monsoon & wintertime. It can be seen temperature in rainy season doesn't demonstrate a critical pattern at all in significant pieces in our nation with the exception of a huge negative pattern across northwestern parts of our country. Not one critical pattern stood in the precipitation on any yearly or the occasional premise. The current data on environmental variation as well as on patterns are event of the

outrageous occasions across the nation was summed up by Sinha Ray and De (2003). The situation is accounted for that precipitation and external pressing factor found that middle value of over the country all in all show no huge pattern. An expanding pattern of the request for 0.35°C in the course of the most recent was observed in archives. Outrageous most stimulating and least readings display a diminishing pattern in the northside and expanding in southern parts. Forthcoming heating situations were created for the nation's mainland utilizing universal sequence models.

Water assets assume an imperative part of the economic development of the sub-continent. The rural areas were probably going the most noticeably awful influenced by hotter environment. The significant streams of India originating from the Himalayas, in particular, the Ganga, the Indus, as well as north eastern Brahmaputra, and get the generous measure of water obtained by melting of snow as well as glacial masses. The Himalayan water produce from ice-took care of streams is generally double as compare to the peninsular waterways. A few investigations to appraise the commitment of snow and glacial masses in the yearly progressions of a couple of Himalayan streams were done by Singh et al. (1995, 1998) and Singh and Jain (2001).

3. Research Methodology

3.1 Dataset

The data is retrieved from the official database website of Indian government. The dataset has been downloaded from the government website "data.gov.in". This website is a program for subsidiary freely available information scheme of Indian government. The easy access gateway enables user to gain an one point entree to archives, administrations, apparatuses and applications distributed by services, divisions and associations of the Indian government.

a) Mean Temperature of India:

The dataset is in .csv provides information about monthly and seasonal temperature our country.

Source: "https://data.gov.in/resources/monthly-seasonal-and-annual-mean-temp-series-1901-2017"

b) Sub-divisional Rainfall data:

The dataset includes average rainfall in different sub-divisions of India from 1901-2017.

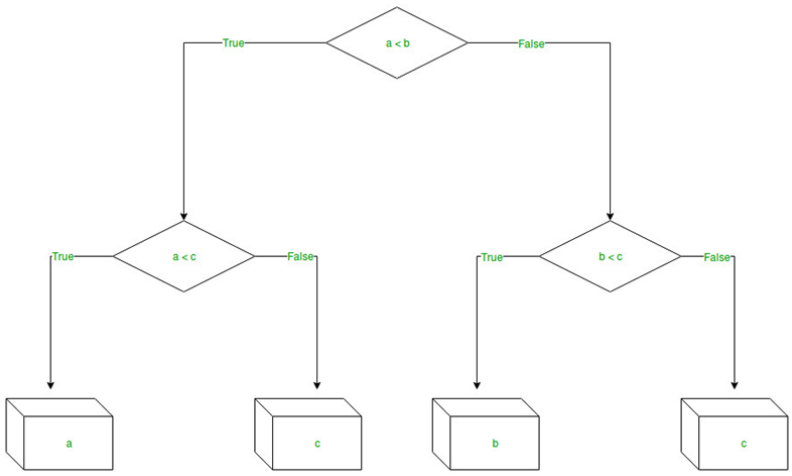
Source: "https://data.gov.in/resources/sub-divisional-monthly-rainfall-1901-2017"

3.2 Methods

3.2.1 Algorithm: Decision Tree Regressor using sklearn

Decision Tree can be defined as a conclusion creation instrument which practices a flowchart similar tree structure, it is a typical of decisions and the aggregate of their expected results, including results, input costs and utility. The branches/edges address the eventual outcome of the center point. Decision-tree computation falls under the grouping of oversaw learning estimations. It works for both steady similarly as full-scale yield factors.

The edges or branches address double or reality dealing of the attestation as well as receipts types a choice ward on those on the model underneath that displays a conclusion diagram which surveys the most diminutive of 3 figures:



Decision tree relapse notices highlights of an item and trains a prototype in the construction of a diagram so that can foresee data advanced in to distribute significant ceaseless output. Continuous output suggests that output obtained is not distinct, i.e., it isn't addressed just by a distinct, identified plan of statistics or assets.

- Discrete output example: A weather prediction prototype which forecasts that will there'll be rainstorm in a exact daytime.
- Continuous output example: Any advantage expectation prototype which couriers the likely advantage which can be shaped from that proposal of an entry.

3.3 Programming Tool Used

The programming tool used is Python 3.0 using Jupyter Notebooks.

Jupyter Notebooks

Jupyter Notebook can be explained as an application available on interet which is open source and that grants you to make and share files which cover aware cypher, circumstances, portrayals and record transcript. Usages include data scrubbing and modification, numerical re-institution, quantifiable showing, data portrayal, computer-based intelligence, and considerably further. It utilizes a scratch pad which incorporates cryptogram and its output into a lone record that unites portrayals, story text, mathematical conditions, and other rich media. All things considered it is a singular report in which one is able to access program, display the output, and moreover enhance explanations, conditions, graphs, and style our effort further clear, sensible & repeatable.

3.4 Libraries used

- Pandas

  It is a Python package that gives fast, flexible, and sensitive data assemblies envisioned to style working with systematized (plain, possibly assorted) and time plan data both straightforward and common. This library will help to utilize to import any source of data and make frame of data for examination.

- Sklearn

  It is the utmost valuable as well as durable collection for ML in Python. It stretches a purpose of effective strategies for ML and assessable demonstrating together with relapse, bunching and decrease through a consistence boundary in Python. Sklearn will be utilized to apply arbitrary timberland regressor for forecast and figure the precision of the model.

- PlotLy

  The plotly is a cooperative library and is an smart, open-source visualization package which upholds around 50 special outline kinds casing a extensive possibility of measurable, financial, terrestrial, logical, and 3-D cases. It empowers Python clients to make wonderful intuitive electronic representations which is able to be publicized in Jupyter scratch pad, saved to autonomous HTML credentials, or filled in as a feature of unadulterated Python-fabricated web applications utilizing Dash.

4. Objectives

- To execute an EDA on the data of mean temperature and mean rainfall experienced in India to analyze the weather of India.
- To predict the mean temperature for the next year on the basis of prior data with help of ML algorithm.
- To predict the mean rainfall every month for the next year on the basis of prior data with help of ML algorithm.
- To examine the trendline of rainfall and temperature over the years for analysis.
- To evaluate the model accuracy through R square value.

5. Code and Analysis

5.1 Analysis on mean temperature of India

5.1.1 Importing the required libraries

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
from datetime import datetime
df = pd.read_csv('new file.csv')
df.head()
```

Output:

	Unnamed: 0	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0	0	1901	17.99	19.43	23.49	26.41	28.28	28.60	27.49	26.98	26.26	25.08	21.73	18.95
1	1	1902	19.00	20.39	24.10	26.54	28.68	28.44	27.29	27.05	25.95	24.37	21.33	18.78
2	2	1903	18.32	19.79	22.46	26.03	27.93	28.41	28.04	26.63	26.34	24.57	20.96	18.29
3	3	1904	17.77	19.39	22.95	26.73	27.83	27.85	26.84	26.73	25.84	24.36	21.07	18.84
4	4	1905	17.40	17.79	21.78	24.84	28.32	28.69	27.67	27.47	26.29	26.16	22.07	18.71

```
df = pd.read_csv("new file.csv", index_col = 0)
df.head()
```

Output:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0	1901	17.99	19.43	23.49	26.41	28.28	28.60	27.49	26.98	26.26	25.08	21.73	18.95
1	1902	19.00	20.39	24.10	26.54	28.68	28.44	27.29	27.05	25.95	24.37	21.33	18.78
2	1903	18.32	19.79	22.46	26.03	27.93	28.41	28.04	26.63	26.34	24.57	20.96	18.29
3	1904	17.77	19.39	22.95	26.73	27.83	27.85	26.84	26.73	25.84	24.36	21.07	18.84
4	1905	17.40	17.79	21.78	24.84	28.32	28.69	27.67	27.47	26.29	26.16	22.07	18.71

Now, we'll make an attribute that would contain date (month, year). So that we could get temperature values with the timeline.

```
df1 = pd.melt(df, id_vars='YEAR', value_vars=df.columns[1:])
df1['Date'] = df1['variable'] + ' ' + df1['YEAR'].astype(str)
df1.loc[:,'Date'] = df1['Date'].apply(lambda x : datetime.strpti
me(x, '%b %Y'))
df1.head()
```

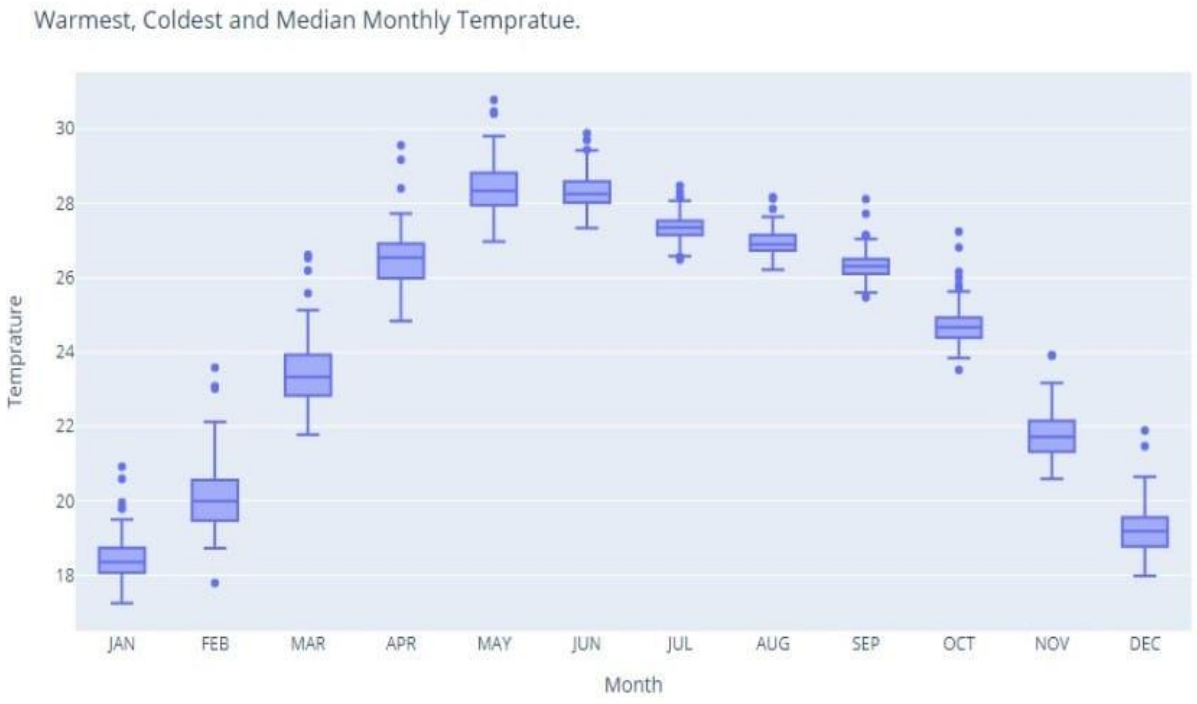
Output:

	YEAR	variable	value	Date
0	1901	JAN	17.99	1901-01-01
1	1902	JAN	19.00	1902-01-01
2	1903	JAN	18.32	1903-01-01
3	1904	JAN	17.77	1904-01-01
4	1905	JAN	17.40	1905-01-01

5.1.2 Exploratory Data Analysis.

```
df1.columns=['Year', 'Month', 'Temprature', 'Date']
df1.sort_values(by='Date', inplace=True) ## To get the time seri
es right.
fig = px.box(df1, 'Month', 'Temprature')

fig.update_layout(title='Warmest, Coldest and Median Monthly Tem
pratue.')
fig.show()
```

Output:



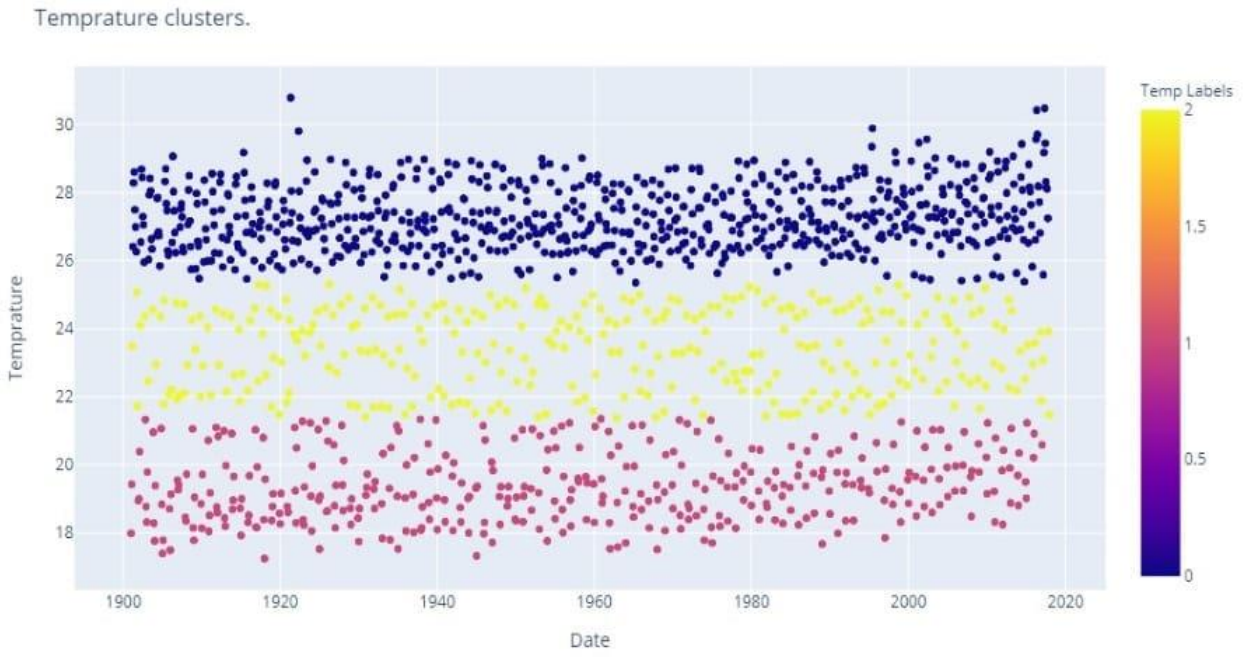
The above boxplot explains the following information:

- The coldest days are in the month of January in a year.
- The hottest days of month in a Year.
- The least Standard Deviation is in the month of July which means, in July the temperature differ slightest.

5.1.3 Clustering

```
from sklearn.cluster import KMeans
sse = []
target = df1['Temprature'].to_numpy().reshape(-1,1)
num_clusters = list(range(1, 10))

for k in num_clusters:
    km = KMeans(n_clusters=k)
    km.fit(target)
    sse.append(km.inertia_)

km = KMeans(3)
km.fit(df1['Temprature'].to_numpy().reshape(-1,1))
df1.loc[:,'Temp Labels'] = km.labels_
fig = px.scatter(df1, 'Date', 'Temprature', color='Temp Labels')
fig.update_layout(title = "Temprature clusters.",
                  xaxis_title="Date", yaxis_title="Temprature")
fig.show()
```

Output:



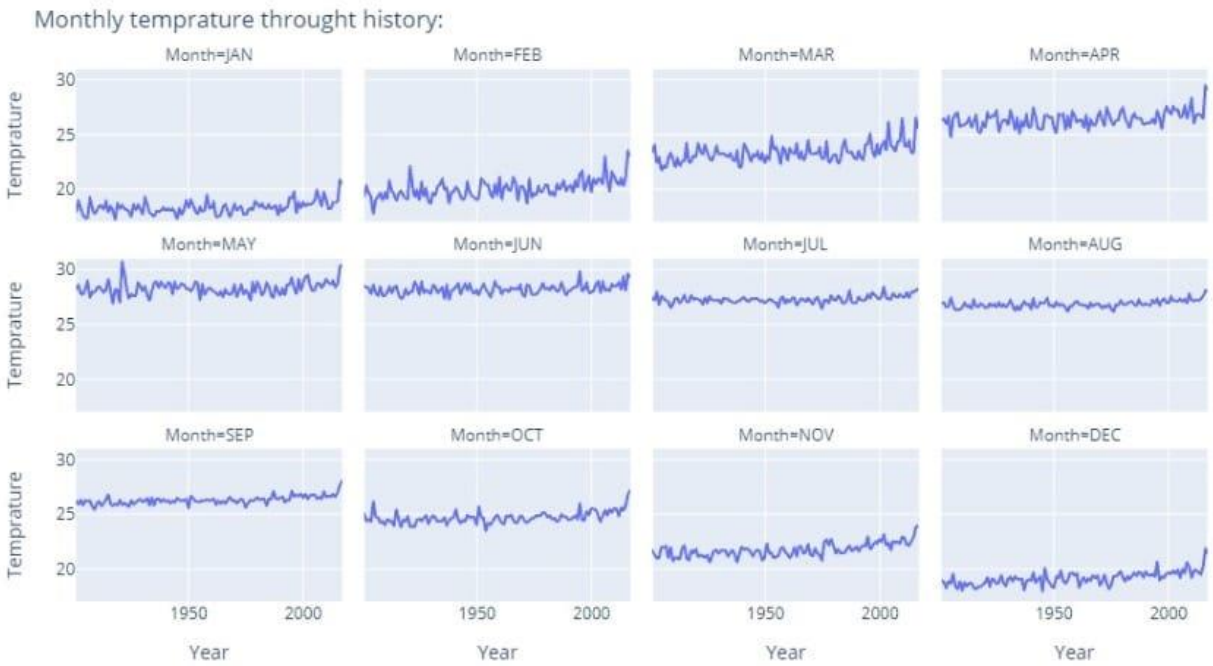
The scatter plot explains that:

- There can be seen three key clusters on the basis of temperature in spite of having 4 seasons.
- The coldest months are January, February and December.
- The months with higher degree of temperature May, April, July, August, June and September.
- March, November and October are the months with temperature not too high and not too low.

5.1.4 Monthly temprature throughout history:

```
fig = px.line(df1, 'Year', 'Temprature', facet_col='Month', face
t_col_wrap=4)
fig.update_layout(title='Monthly temprature throught history:')
fig.show()
```

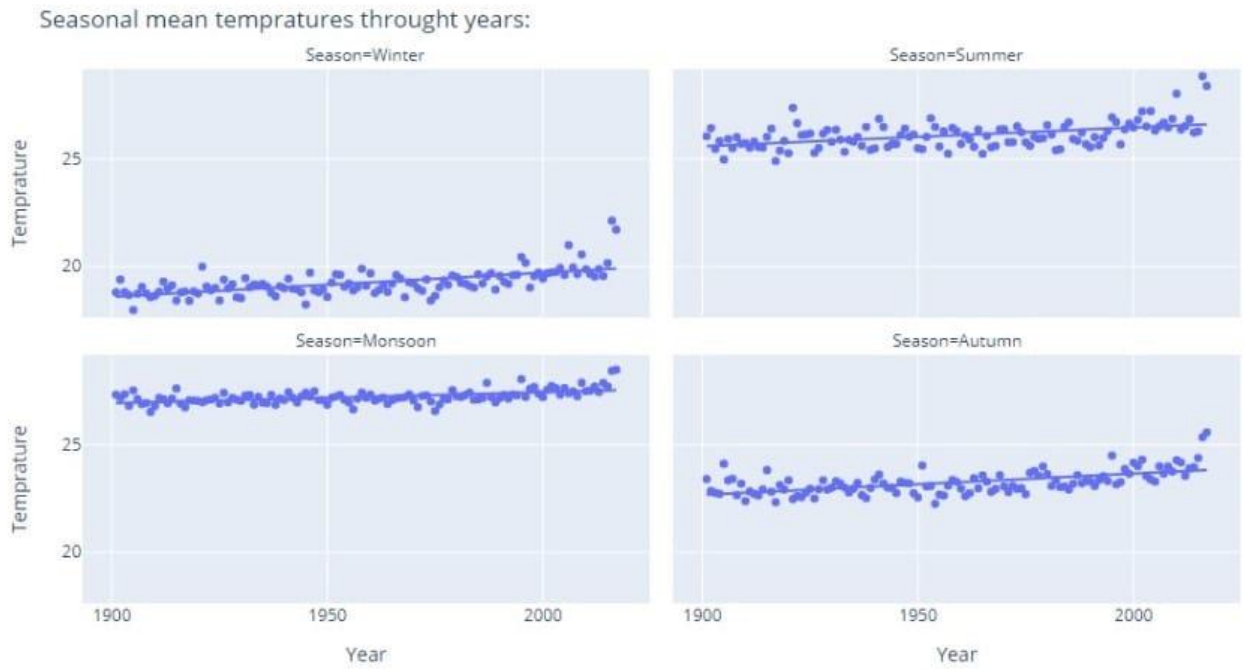
Output:



5.1.5 Seasonal Temperature Analysis

```
df['Winter'] = df[['DEC', 'JAN', 'FEB']].mean(axis=1)
df['Summer'] = df[['MAR', 'APR', 'MAY']].mean(axis=1)
df['Monsoon'] = df[['JUN', 'JUL', 'AUG', 'SEP']].mean(axis=1)
df['Autumn'] = df[['OCT', 'NOV']].mean(axis=1)
seasonal_df = df[['YEAR', 'Winter', 'Summer', 'Monsoon', 'Autumn
']]
seasonal_df = pd.melt(seasonal_df, id_vars='YEAR', value_vars=se
asonal_df.columns[1:])
seasonal_df.columns=['Year', 'Season', 'Temprature']

fig = px.scatter(seasonal_df, 'Year', 'Temprature', facet_col='S
eason', facet_col_wrap=2, trendline='ols')
fig.update_layout(title='Seasonal mean tempratures throught year
s:')
fig.show()
```

Output:



A positive trendline can be seen amongst time and temperature. Even though there is not a very high positive correlation of trendline with time but still it is not insignificant.

5.1.6 Trendline over the years

```
df['Yearly Mean'] = df.iloc[:,1:].mean(axis=1)
fig = px.scatter(df,x = 'YEAR', y = 'Yearly Mean', trendline = '
lowess')
fig.update_layout(title='Trendline Over The Years :',
                  xaxis_title='Time', yaxis_title='Temprature in
Degrees')
fig.show()
```

Output:

From the above trendline analysis we can conclude the following points:

- The yearly mean temperature was not increasing till 1980. It was only after 1979 that we can see the gradual increase in yearly mean temperature.
- After 2015, yearly temperature has increased drastically.
- We are seeing a monthly like up-down pattern in yearly temperatures as well.

5.1.7 Forecasting and Accuracy of the model using Decision Tree Regressor

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
df2 = df1[['Year', 'Month', 'Temprature']].copy()
df2 = pd.get_dummies(df2)
y = df2[['Temprature']]
x = df2.drop(columns='Temprature')
dtr = DecisionTreeRegressor()
train_x, test_x, train_y, test_y = train_test_split(x,y,test_siz
e=0.3)
dtr.fit(train_x, train_y)
pred = dtr.predict(test_x)
r2_score(test_y, pred)
```

Output:

```
0.966190110865615
```

- **Prediction for the year 2018:**

```
next_Year = df1[df1['Year']==2017][['Year', 'Month']]
next_Year.Year.replace(2017,2018, inplace=True)
next_Year= pd.get_dummies(next_Year)
temp_2018 = dtr.predict(next_Year)
temp_2018 = {'Month':df1['Month'].unique(), 'Temprature':temp_20
18}
temp_2018=pd.DataFrame(temp_2018)
temp_2018['Year'] = 2018
temp_2018
```

Output:

	Month	Temprature	Year
0	JAN	20.92	2018
1	FEB	23.08	2018
2	MAR	25.58	2018
3	APR	29.17	2018
4	MAY	30.41	2018
5	JUN	29.44	2018
6	JUL	28.31	2018
7	AUG	28.12	2018
8	SEP	28.12	2018
9	OCT	27.24	2018
10	NOV	23.92	2018
11	DEC	21.47	2018

5.2 Analysis on mean rainfall from 1901-2017

- Importing the required libraries

```
import numpy as np ## For Linear Algebra
import pandas as pd ## To Work With Data
from matplotlib import pyplot as plt
## For visualizations I'll be using plotly package, this creates
 interesting and interective visualizations.
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
from datetime import datetime ## Time Series analysis.
df = pd.read_csv("Mean rainfall data major project.csv", index_c
ol = 0)
df.head()
```

Output:

	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
0	1901	35.22	40.51	21.10	46.39	70.43	194.61	291.21	301.33	134.15	73.70	64.76	10.81
1	1902	8.52	8.31	20.95	45.77	72.04	168.82	351.15	239.80	258.61	86.71	43.46	33.15
2	1903	16.85	14.34	26.58	16.97	80.24	171.99	362.11	299.48	216.62	125.35	48.52	22.31
3	1904	11.70	9.39	28.88	41.40	92.45	237.49	311.97	217.01	167.40	84.55	19.76	14.77
4	1905	23.20	20.48	40.94	34.75	76.44	164.81	288.31	246.78	179.26	78.17	16.03	17.01

```
df1 = pd.melt(df, id_vars='YEAR', value_vars=df.columns[1:])
df1.head()
```

Output:

	YEAR	variable	value
0	1901	JAN	35.22
1	1902	JAN	8.52
2	1903	JAN	16.85
3	1904	JAN	11.70
4	1905	JAN	23.20

```
df1['Date'] = df1['variable'] + ' ' + df1['YEAR'].astype(str)
df1.loc[:,'Date'] = df1['Date'].apply(lambda x : datetime.strpti
me(x, '%b %Y')) ## Converting String to datetime object
df1.head()
```

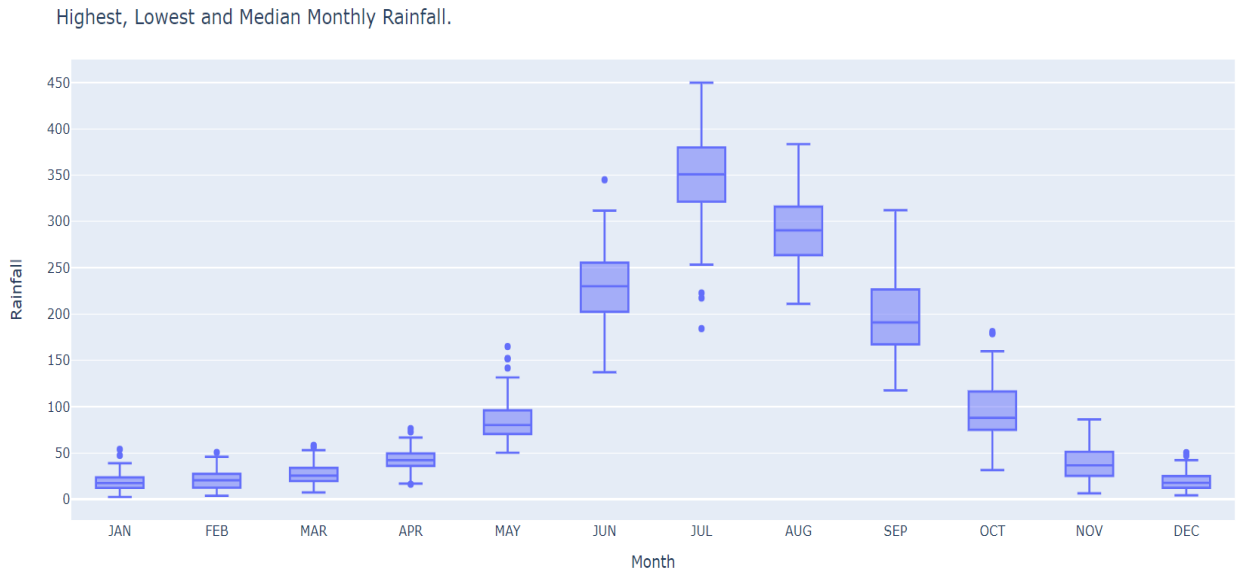
Output:

	YEAR	variable	value	Date
0	1901	JAN	35.22	1901-01-01
1	1902	JAN	8.52	1902-01-01
2	1903	JAN	16.85	1903-01-01
3	1904	JAN	11.70	1904-01-01
4	1905	JAN	23.20	1905-01-01

5.2.1 Exploratory Data Analysis.

```
df1.columns=['Year', 'Month', 'Rainfall', 'Date']
df1.sort_values(by='Date', inplace=True) ## To get the time seri
es right.
fig = px.box(df1, 'Month', 'Rainfall')

fig.update_layout(title='Highest, Lowest and Median Monthly Rain
fall.')
fig.show()
```

Output:



The above boxplot explains the following information:

• January and February have the least rainfall in a Year.

• July has the most rainfall in a Year.

5.2.2 Clustering

```
from sklearn.cluster import KMeans
sse = []
target = df1['Rainfall'].to_numpy().reshape(-1,1)
num_clusters = list(range(1, 10))

for k in num_clusters:
    km = KMeans(n_clusters=k)
    km.fit(target)
    sse.append(km.inertia_)

km = KMeans(3)
km.fit(df1['Rainfall'].to_numpy().reshape(-1,1))
df1.loc[:,'Temp Labels'] = km.labels_
fig = px.scatter(df1, 'Date', 'Rainfall', color='Temp Labels')
fig.update_layout(title = "Rainfall clusters.",
                  xaxis_title="Date", yaxis_title="Rainfall")
fig.show()
```

Output:

The scatter plot explains that:

• There 3 main clusters based on rainfall.

• August, June and July experience the most rainfall throughout the year.

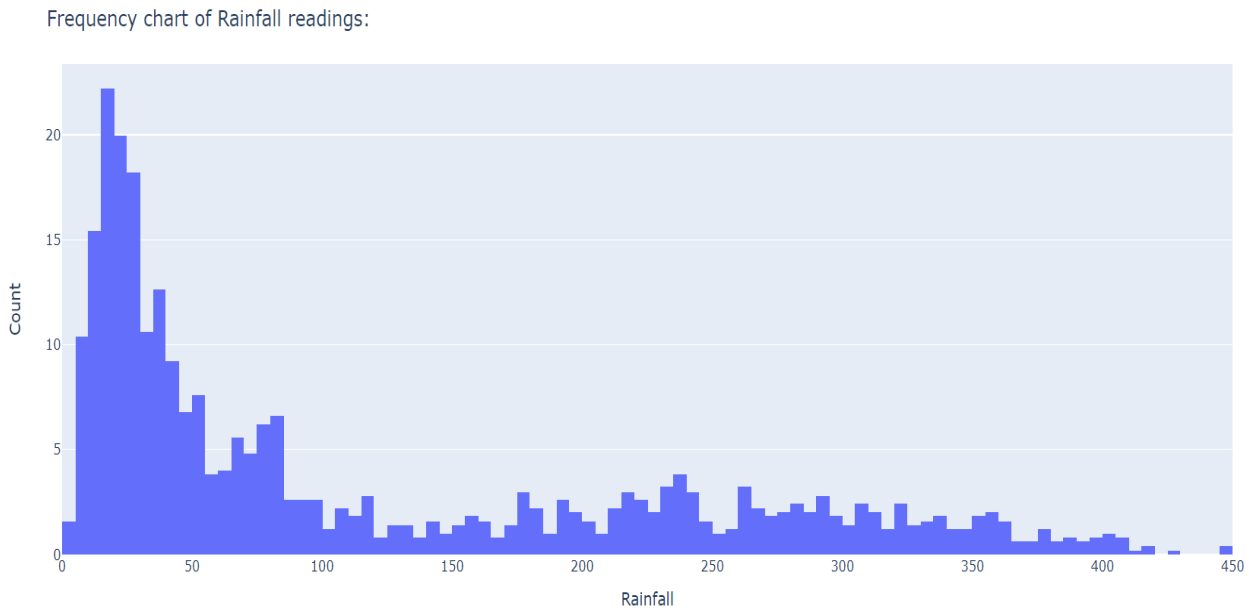
• all have hotter temperatures.

• Mar, Oct and Nov are the months that have temperatures neither too hot nor too cold.

5.2.3 Frequency chart of rainfall:

```
fig = px.histogram(x=df1['Rainfall'], nbins=200, histnorm='densi
ty')
fig.update_layout(title='Frequency chart of Rainfall readings:',
 xaxis_title='Rainfall', yaxis_title='Count')
```

Output:



5.2.4 Monthly rainfall throughout history:

```
fig = px.line(df1, 'Year', 'Rainfall', facet_col='Month', facet_
col_wrap=4)
fig.update_layout(title='Monthly Rainfall throught history:')
fig.show()
```

Output:

5.2.5 Seasonal Temperature Analysis

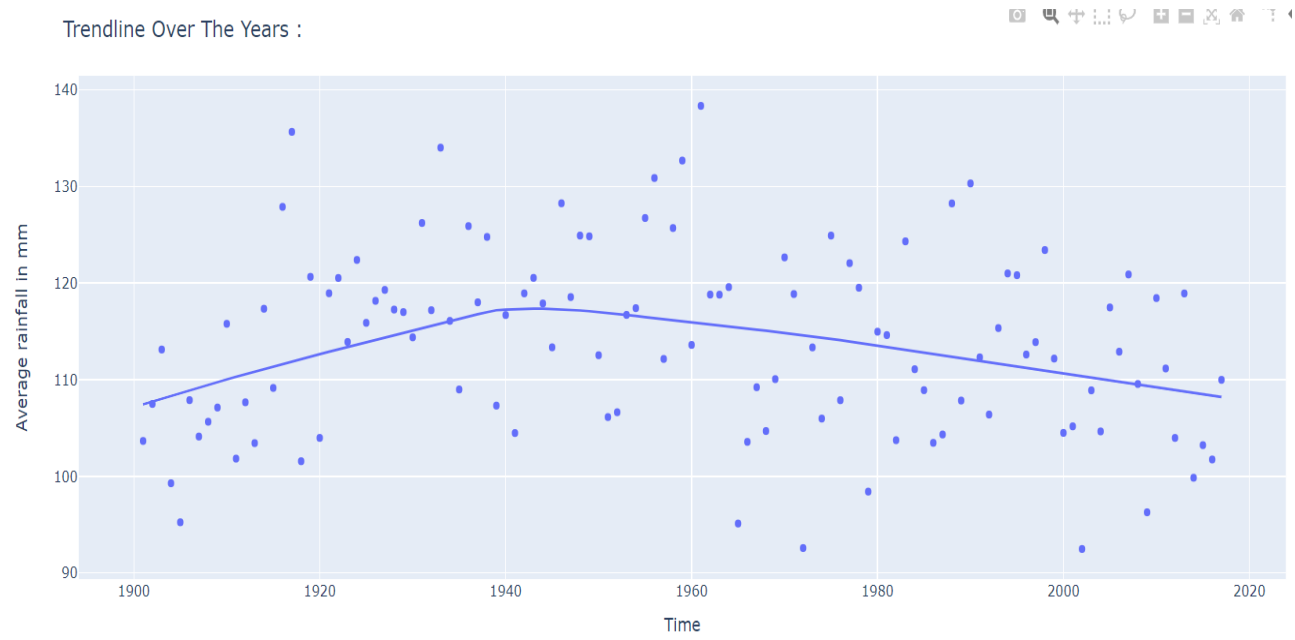
```
df['Winter'] = df[['DEC', 'JAN', 'FEB']].mean(axis=1)
df['Summer'] = df[['MAR', 'APR', 'MAY']].mean(axis=1)
df['Monsoon'] = df[['JUN', 'JUL', 'AUG', 'SEP']].mean(axis=1)
df['Autumn'] = df[['OCT', 'NOV']].mean(axis=1)
seasonal_df = df[['YEAR', 'Winter', 'Summer', 'Monsoon', 'Autumn
']]
seasonal_df = pd.melt(seasonal_df, id_vars='YEAR', value_vars=se
asonal_df.columns[1:])
seasonal_df.columns=['Year', 'Season', 'Rainfall']

fig = px.scatter(seasonal_df, 'Year', 'Rainfall', facet_col='Sea
son', facet_col_wrap=2, trendline='ols')
fig.update_layout(title='Seasonal mean Rainfall throught years:'
)
fig.show()
```

Output:

5.2.6Trendline over the years

```
df['Yearly Mean'] = df.iloc[:,1:].mean(axis=1)
fig = px.scatter(df,x = 'YEAR', y = 'Yearly Mean', trendline = 'lowess')
fig.update_layout(title='Trendline Over The Years :',
                  xaxis_title='Time', yaxis_title='Temprature in Degrees')
fig.show()
```

Output:



From the above trendline analysis we can conclude the following points:

• The yearly mean temperature was increasing till 1940. It was only after 1940 that we can see the gradual decrease in yearly mean temperature.

• After 1953, yearly rainfall has decreased drastically.

5.2.7 Forecasting and Accuracy of the model using Decision Tree Regressor

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
df2 = df1[['Year', 'Month', 'Rainfall']].copy()
df2 = pd.get_dummies(df2)
y = df2[['Rainfall']]
x = df2.drop(columns='Rainfall')
dtr = DecisionTreeRegressor()
train_x, test_x, train_y, test_y = train_test_split(x,y,test_siz
e=0.3)
dtr.fit(train_x, train_y)
pred = dtr.predict(test_x)
r2_score(test_y, pred)
```

Output:

```
0.8985433738907581
```

```
next_Year = df1[df1['Year']==2017][['Year', 'Month']]
next_Year.Year.replace(2017,2018, inplace=True)
next_Year= pd.get_dummies(next_Year)
rain_2018 = dtr.predict(next_Year)
rain_2018 ={'Month':df1['Month'].unique(), 'Rainfall':rain_2018}
temp_2018=pd.DataFrame(rain_2018)
temp_2018['Year'] = 2018
temp_2018
```

Output:

	Month	Rainfall	Year
0	JAN	25.45	2018
1	FEB	9.14	2018
2	MAR	51.73	2018
3	APR	47.88	2018
4	MAY	84.33	2018
5	JUN	232.58	2018
6	JUL	322.09	2018
7	AUG	277.41	2018
8	SEP	191.01	2018
9	OCT	66.06	2018
10	NOV	22.36	2018
11	DEC	23.78	2018

6. Conclusion

The exploratory data analysis on the dataset of 117 years, provided by the government of India shows a significant change in the weather of our country.

6.1 Key Findings

i. The warmest year recorded was 2016
ii. In the years between 2002 & 2016 there were 13 years amongst the hottest 15 years. The most recent period of 2001to 2010 and the years between 2007 to 2016 were considered likewise the warmest ever recorded.
iii. Yearly average temperature of our country expanded by around 1.3℃ subsequently the start of the 1900s.
iv. Yearly average temperature of our country quickly expanded later 1996. In light of present conditions of increment, it can break the 2℃ mark inside the following twenty years. 2℃ is the optimistic objective established underneath the Paris Contract.
v. Throughout the hurricane period that the temperature increment is around 1.2℃.
vi. The rainstorm has been very dynamic in the long stretch of September in light of different elements. One significant factor is the low-pressure frameworks that dropped over Bay of Bengal moved towards Rajasthan and under its impact, an East-West

arranged low-pressure zone happened. This supported the rainstorm for a significant long time.

vii. Thereafter, a downturn framed over the upper east Arabia Sea and it crossed the Gujarat coast and it moved toward that path supporting the rainstorm highlights.

6.2 Implications

The recurrence and force of outrageous climate occasions is expanding through the increasing temperature. In 2017, during winter season the normal temperature was 3℃ more than the 1901-1930 pattern, It is noticeable that it was a terrible dry season in a century occurred in southerly parts of our country. States like Kerala, Karnataka, Tamil Nadu and Andhra Pradesh were most exceedingly awful hit, with 330 million individuals going under the grasp of dry spell.

Additionally, In 2010's summer period, the normal reading was 2.3℃ more than the benchmark, once more the most elevated in written past, extreme warmth tendency circumstances beat huge pieces of India. These conditions asserted in excess of 300 lives. Also, four cyclonic tempests hit India that year.

The temperature of our country is increasing at an alarming rate. The ramifications of this essential truth are not kidding for financial, communal and environmental prosperity of India. We are encountering incessant limit climate occasions, and our climate is getting unusual. Misfortunes because of outrageous climate occasions are mounting and it is our helpless who are enduring the torments of environmental change.

7. Biblography

T V, R. K., V V SSS, B., & N, R. (2014). Analysis of Indian weather data sets using data mining techniques. *Computer Science & Information Technology (CS & IT)*.

Dash, Y., Mishra, S. K., Sahany, S., & Panigrahi, B. K. (2018). Indian summer monsoon rainfall prediction: A comparison of iterative and non-iterative approaches. *Applied Soft Computing*, *70*, 1122-1134.

Yaduvanshi, A., Nkemelang, T., Bendapudi, R., & New, M. (2021). Temperature and rainfall extremes change under current and future global warming levels across Indian climate zones. *Weather and Climate Extremes*, *31*, 100291.

Arora, M., Goel, N. K., & Singh, P. (2005). Evaluation of temperature trends over India / Evaluation de tendances de temperature en Inde. *Hydrological Sciences Journal*, *50*(1).

Radhakrishnan, K., Sivaraman, I., Jena, S. K., Sarkar, S., & Adhikari, S. (2017). A climate trend analysis of temperature and rainfall in India. *Climate Change and Environmental Sustainability*, *5*(2), 146.

Rupa Kumar, K. ; Sahai, A. K. ; Krishna Kumar, K. ; Patwardhan, S. K. ; Mishra, P. K. ; Revadekar, J. V. ; Kamala, K. ; Pant, G. B. (2006) *High-resolution climate change scenarios for India for the 21st century* Current Science, 90 (3). pp. 334-345.