

RAINFALL DATA ANALYSIS

A Project Report
submitted in partial fulfillment of the requirement for the degree
of

**BACHELOR OF TECHNOLOGY
(Computer Science and Engineering)**



INDIAN INSTITUTE OF
INFORMATION
TECHNOLOGY

By

B RAGAVAN	- 18BCS016
G K BHARATH BHUSHAN	- 18BCS026
K V S CHAITANYA	- 18BCS045
MEGHANA N	- 18BCS053

Under the guidance of
Dr. Uma Seshadri, HOD and Professor,
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, INDIAN INSTITUTE OF INFORMATION
TECHNOLOGY, DHARWAD.

Certificate

This is to certify that the work contained in the project report titled Rainfall Data Analysis by G K Bharath Bhushan, Meghana N, KVS Chaitanya, B Ragavan during the VIII semester - IV Year as a Major Project under the guidance of Dr. Uma Seshadri, IIIT Dharwad.

Signature of Supervisor
Dr. Uma Seshadri,
HOD and Professor, IIIT Dharwad

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Meghana N(18BCS053)

G K Bharath Bhushan(18BCS026)

B Ragavan(18BCS016)

K V S Chaitanya(18BCS45)

Approval Sheet

This project report entitled Rainfall Data Analysis by G K. Bharath Bhushan(18BCS026), Meghana N(18BCS053), KVS Chaitanya(18BCS045), B.Ragavan(18BCS016) of Indian Institute of Information Technology, Dharwad is approved for the degree of Bachelor of Technology in Computer Science and Engineering.

Supervisor,

Dr. Uma Seshadri,
HOD and Professor,
Computer Science and Engineering,
Indian Institute of Information Technology, Dharwad.

Table of Contents

Title	Page No
CHAPTER 1 INTRODUCTION	7-9
1.1 Overview.....	8
1.2 Problem Definition.....	9
CHAPTER 2 TECHNOLOGY STACK	10-11
CHAPTER 3 REVIEW OF LITERATURE	12-17
CHAPTER 4 IMPLEMENTATION	18-31
4.1 Module 1.....	18-25
4.2 Module 2.....	25-31
CHAPTER 5 RESULTS AND DISCUSSIONS	32-56
CHAPTER 6 SUMMARY AND CONCLUSION	57-58

Abstract

In the project we have analysis on rainfall data. To begin with, the world has faced a lot of water problems that are an effect of different rainfall patterns. We can understand the need for water and what made it necessary for us to choose this topic for our major project. The project also gives an introduction saying what are the existing difficulties related to rainfall(water management) which makes the rainfall analysis necessary. We discuss what are the traditional methods that have been used for ages and how rainfall analysis can be of help in order to understand the rainfall pattern. We can observe the disadvantages of the traditional methods used for rainfall analysis and how computer-based calculating methods are better and more useful. Then explain what is the problem statement and what is our view about dealing with it.

A description on technology stack where we will discuss the various technologies used in building the project and emphasizes the importance of every particular technology giving a brief about the same. A review of literature to know all the previous works on the similar or the same topic. What are the outcomes, what are the key derivatives, what are the disadvantages, what can be improvised etc. We have discussed the implementation to observe how we went about doing every particular step in the project having detailed description of all the procedures in the project. Finally conclude by analyzing the obtained graphs.

1. Introduction

In the modern scenario, rainfall is a sizable thing for maximum important matters going on all around the world. The farming quarter is seen as one of the most vital elements determining the nation's economy, and farming depends completely on rainfall. This project makes use of machine learning strategies for rainfall prediction in depicting an efficient rainfall prediction method.

Rainfall prediction enables water sources management, flood alerts, flight operations management, restricting transportation, production activities, and different elements which can be of maximum importance to humankind. Rainfall data for forecasting is collected using climate satellites, wired and wireless instruments and high-velocity computer systems are used. Rainfall prediction has been a charming and captivating sector since the sunrise of civilization, and it stays one of the most complicated domains. Scientists use various techniques and strategies to predict rainfall, a number of which can be more precise than others. Rainfall forecasting gathers atmospheric situations including humidity, temperature, pressure, rainfall, wind direction & velocity, evaporation, etc.

Presently, Rainfall prediction is the most critical thing for most water storage schemes worldwide. The uncertainty of rainfall data is one of the most complex problems. Today, most rainfall forecasting techniques are incapable of detecting hidden patterns or non-linear traits in rainfall data. Analyzing the quantity of rainfall offers alertness to farmers in advance in order to defend their plants and residences from rain. There are extra strategies for expecting rainfall. The ML algorithms are fine desirable for the prediction of rainfall. Here are a number of the fundamental ML algorithms used which might be Autoregressive Integrated Moving Average (ARIMA), Artificial neural network (ANN), Support Vector Machine, and Logistic regression.

Rainfall has a considerable effect on the availability of water in lots of places. Some areas can also additionally acquire little or no rainfall, others can also have mild rainfall at the same time as it but others can also acquire immoderate rainfall. Most areas include water reservoirs. However without rain, usage of such reservoirs can also get less resulting in water shortage. When it rains, rainwater harvesting is a technique of storing the rain water above the floor or recharging the groundwater supply. This is a herbal prevalence in rural places. We want to expand rainwater series technology in metropolitan areas. If precise techniques for storing water aren't implemented, it is able to have an effect on future generations

1.1 Overview :

Rainfall Analysis is the usage of technological knowledge and generation to anticipate the atmospheric situations for the possibility of rainfall in a given place and time. People were predicting the climate informally and inaccurately for lots of years in the nineteenth century and beyond. Rainfall Analysis is made through gathering statistics about the modern-day country of the ecosystem in a selected place after which the use of the data to analyze how much rainfall to expect. Individual opinion is required to set up the prediction. When it involves human interest that is primarily based on adjustments in modern-day weather, climate or cloud cover. Rainfall analysis is now mainly counting on technology.

1.2 Problem Definition:

Rainfall follows various patterns in different years. Since the last decade global warming has received the attention of many scientists ([11]). It is believed that global warming has an influence on rainfall and causes change in the rainfall pattern and thereby the climate. Analyzing the rainfall data is necessary to know the pattern followed. The rainfall is inaccurate or partially accurate due to the stagnant weather conditions, complex calculation requirements to analyze rainfall patterns, the error that came up while estimating the initial conditions, and a low or incomplete understanding of atmospheric conditions. The accuracy rate is still low as there is an increase in the difference between the current period and the period when the analysis is done. The usage of a more accurate model will make it certain to minimize the errors. So, we have taken 3 different cities namely Bangalore , Hubli & Vizag for our Rainfall data analysis using ARIMA model.

1.4 Key takeaways of the technology used:

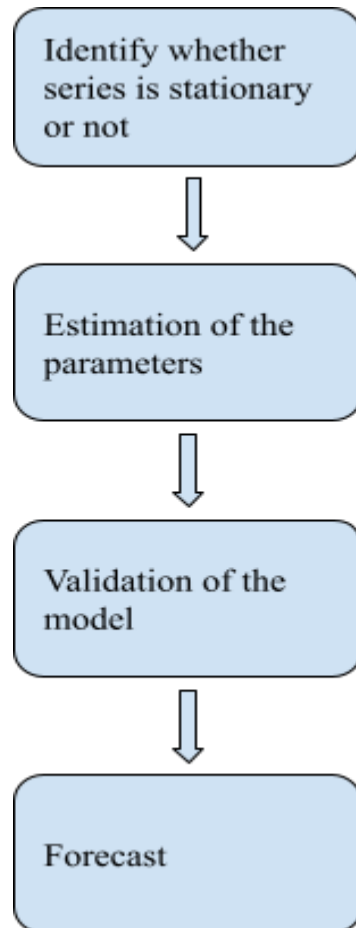
- The Autoregressive Integrated Moving Average (ARIMA) model forecasts future values using data from the past.
- ARIMA makes use of lagged moving averages in order to simplify the time series data.
- Autoregressive models believe the future will be similar to the past.
- It gauges the strength of one dependent variable to another changing variable.
- It has a significantly higher precision and accuracy.

2. Technology Stack

- **Google Colab, Spyder** - It is an open-supply cross-platform including improvement surroundings for clinical programming withinside the Python language.
- **netCDF4** - It is a Python interface to the netCDF C library.
- **Numpy** - NumPy is a library for the Python programming language, including aid for big, multi-dimensional arrays and matrices, at the side of a big series of high-stage mathematical features to function on those arrays.
- **Pandas** - Pandas is a software program library written for the Python programming language for statistics manipulation and evaluation.
- **Matplotlib** - Matplotlib is a plotting library for the Python programming language and its numerical arithmetic extension NumPy. It gives an object-oriented API for embedding plots into programs the usage of general-motive GUI toolkits like Tkinter, wxPython, Qt, or GTK.
- **Statsmodels** - It is a Python module that gives lessons and features for the estimation of many unique statistical fashions, in addition to for carrying out statistical tests, and statistical statistics exploration.
- **Scikit-learn** - It is an unfastened software program device gaining knowledge of libraries for the Python programming language. It capabilities diverse classification, regression and clustering algorithms inclusive of aid-vector machines .
- **Itertools** - Python's Itertool is a module that gives diverse features that paint on iterators to supply complicated iterators. This module works as a fast, memory-green device that is used both with the aid of using themselves or in aggregate to shape iterator algebra.

- **ARIMA** - It is an acronym that stands for Auto-Regressive Integrated Moving Average. It is a category of fashions that captures a set of various trendy temporal systems in time collection statistics.

ARIMA flow:



3. Review and Literature

As in keeping with the United Nations Council on Energy, Environment, and Water (CEEW) Council. Water control demanding situations are inevitable in a situation in which the population increases and weather alternates give unparalleled complications. Finding answers to those demanding situations calls for higher information and a more in-depth inspection of the problem at hand.

Now let's have a short study of the water-associated troubles in India.

1. Drought, floods, and absence of ingesting water
2. Surface and groundwater storage
3. Water exceptional troubles
4. Water Logging etc - A soil is stated to be waterlogged while it is saturated with water, that is due to water stagnation on flat land and low-lying areas. It takes place because of extra rainfall, floods, seepage excessive water table, obstruction to natural drainage, over-irrigation, etc

India's common temperature has risen by means of around 0.7°C in the course of 1901–2018. This upward push in temperature is essentially because of GHG-induced warming, partially offset by forcing because of anthropogenic aerosols and adjustments in LULC. By the end of the twenty-first century, the common temperature over India is projected to increase by about 4.4°C relative to the recent past (1976 - 2005 common).

Predicting rainfall is an extra vital aspect as heavy rainfall ends in many failures. The predicting of rainfall facilitates humans to take preventive measures on their plants and homes and moreover, the predictions have to be extra correct and green. There are sorts of predictions, short-time period rainfall prediction, and long-time period rainfall prediction. Predicting generally short-time period predictions can deliver us correct outcomes. The essential project for us is to construct a version for long-time period rainfall prediction. Heavy precipitation prediction might be a chief disadvantage for the climate forecasting branch due to the fact it's far more carefully related to the financial system and the life of human beings. Natural failures like floods and drought are a motive that is encountered via the means of people internationally each year. Predicting rainfall with extra accuracy may be very vital for international locations like India due to the fact their essential financial system is absolutely dependent on agriculture in India. With the dynamic nature of the atmosphere, implemented arithmetic strategies

fail to offer realistic correct outcomes for precipitation statements.

There are essential approaches to forecast rainfall: conceptual modeling and gadget theoretical modeling. Because it's far appeared to estimate the bodily mechanism that oversees the hydrologic process, conceptual modeling is regularly utilized in hydrological forecasting. It is normally primarily based totally at the capabilities and information of a watershed. Unfortunately, this method won't be possible for rainfall forecasting due to the fact critical calibration information for reasoning rainfall is tough to obtain, and rainfall calculations necessitate the usage of a complicated numerical instrument. System theoretical strategies are used to outline the hyperlink among inputs and outputs, while bodily shape strategies are ignored. In time collection forecasting, the maximum standard method is ARMAX, despite the fact that this version is not able to be expected.

Recent advances in sample popularity technique have ended in a number of gears that could without problems forecast rainfall instead of the formerly used traditional method of linear mathematical relationships supported via way of means of operator experience, mathematical curves, and guidelines. The use of a device gaining knowledge of devices to clear up hydrological problems, which include rainfall forecasting, is turning into an increasing number of popular. The importance of this modeling is that it lets the software program to devise input-output styles without the need for the aforementioned understanding of the elements affecting the forecast parameters. Researchers have considered all that has started to fine-track this ML to be expecting the shape of the modeling strategies and parameters if you want to enhance precision and consistency in depicting the predicting version. Furthermore, researchers have used devices gaining knowledge of and are expecting every day streamflow the usage of multi-layer perceptrons.

Some researchers used ANN to forecast rainfall the usage of numerous strategies, which include constructing a rainfall simulation version and imparting correct rainfall forecasting information, temporal and spatial distribution, the usage of short-time period rainfall for city catchment, and locating that the ANN version with decrease lag outflanked in phrases of forecasting specific index, and forecasting every day rainfall the usage of resilient propagation gaining knowledge of set of rules and locating that the ANN version with decrease lag outflanked in phrases of forecasting specific index. used seasonal ARIMA to forecast rainfall the usage of month-to-month rainfall information, and

observed that ARIMA (0,0,1)(1,1,1) turned into the simplest at predicting destiny precipitation with a 95% self assurance interval. examine the random forest (RF) and assist vector fashions.

Open Access Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining. It is vital to wager the correct rainfall for the powerful use of water sources and surest making plans of water sources and availability. For this purpose, numerous strategies and fashions are advanced to estimate rainfall in numerous researches the usage of information mining strategies. The correct and specific estimation of rainfall prediction and estimation of precipitation is viable through many to be had strategies. The utilization of information mining strategies to are expecting rainfall and its answers might also additionally show huge withinside the prediction of green rainfall so as to assist the farmers to assist with their choices on their plants and it additionally improves the increase of the agriculture sector. We observed that the PCA version and KMeans strategies are used some other place to research and are expecting the approaching rainfall and climate styles.

The studies led us to 4 one-of-a-kind rainfall-affected locations in India.

- Rainfall Problems in Different components of India:

India has confronted foremost issues concerning rainfall main to floods withinside the beyond years. We have accrued data approximately about some locations in India that confronted the worst rainfall.

1. Kurnool district(Andhra Pradesh):

The metropolis of Kurnool has been stricken by an intense loss of ingesting water. Colony citizens were up all night time looking to get a pot of water. The best desire becomes the summer time season garage tank, with the Sunkesula and Gajuladinne reservoirs absolutely dry. The tank's reserve is rarely sufficient to satisfy the thirst of the locals. Municipal government did not expect the ingesting water desires of the metropolis and are accused of plunging citizens right into a critical water crisis. Municipal governments declared that they had been presenting water on trade days, however the colonies neglected the promise and had been stored looking forward to as a minimum 3-five days. The revel in resorts is one-of-a-kind. Annapurna Hotel, in the front of Kurnool General Hospital, has pulled

down its shutters due to the fact there's no opportunity of having water on an everyday basis. According to the hotel's management, the meals enterprise can't function without water. As a result, we've quickly closed down.

In Andhra Pradesh, the Kurnool vicinity (rain-shadow zone) is a warm and arid vicinity with everyday climate droughts. Only around 30% of the farmed place in this vicinity is irrigated, with the relaxation counting on rain-fed farming. Tanks are a critical supply of water for the agricultural financial system withinside the vicinity. But the explosion of precise irrigation withinside the catchment and instructions for crop cultivation, which lessen the floor runoff into those tanks, has pushed tanks into disuse. The largest sufferers of this tragedy are the poor, small, and marginal farmers, who rely upon tanks for supplying supplementary irrigation. The officers concluded that 39 out of the fifty four mandals withinside the district had been suffering from low rainfall, mainly to crop failures.

2. Kerala's heavy rainfall in 2018 :

Kerala experienced the heaviest rainfall withinside the year 2018. Kerala has wet seasons, the primary begins off in June and the second one in mid-October, and subsequently ends around mid-November. For the primary 3 months of the monsoon i.e. from June 1st to August 31st, Kerala acquired extra than 35% of rainfall. Actual rainfall of 2428.9mm withinside the area of an everyday 1795.4mm. But withinside the year 2018, Idukki acquired 3580.6mm of rainfall that's 82% greater than everyday. thirteen districts out of 14 acquired extra than everyday rainfall, Kasaragod in north maximum of Kerala is an exception with 13% from everyday. Palakad district additionally had 65% greater than everyday with 2305.6mm falling below the class of Large Excess(LE) together with Idukki.

Meteorologists say that the rainfall facts during the last six years now no longer advocate a seen pattern. After a top in 2013, rainfall withinside the nation dipped in 2014 and hit a low in 2015 earlier than slowly growing in 2016, 2017, and 2018. In 2013, all of the nation's districts, besides Thiruvananthapuram, acquired extra than 1400 mm of rainfall. Kannur, with 3119.4mm, received the maximum rain. "It is because of the inter-annual variability and modifications in climatic conditions," stated D Sivananda Pai, who's the top of the Climatic Prediction Group on the India

Meteorological Department in Pune.

Idukki has the best rainfall with 4060mm in 2013, and Thiruvananthapuram with 1154mm in 2011 is the lowest. The common rainfall for Kerala from 2010 to 2017 is 2721mm, it varies from 1885mm to 3222mm.

3. Floods in North Eastern regions :

In 2020, the ambiguity withinside the maximum, minimal and imply temperature over many components of the united states become usually withinside the variety of $\pm 1.0^{\circ}\text{C}$. Mean temperature over components of Jammu & Kashmir and Ladakh, Himachal Pradesh, and Andaman & Nicobar Islands have been above everyday via way of means of approximately 1°C . Maximum temperature over components of Jammu & Kashmir and Ladakh, Himachal Pradesh, Assam and Meghalaya, Tripura, Kerala, and Mahe become above everyday via way of means of approximately 1°C . . However, the Maximum temperature over components of Punjab, Haryana Chandigarh & Delhi, Uttar Pradesh, Jharkhand, Bihar, and Chhattisgarh have been beneath each other everyday by means of approximately 1°C . The monsoon season this year (with an anomaly of $+0.43^{\circ}\text{C}$ above common) became the 6th warmest considering that 1901. The Post-monsoon season this year (with an anomaly $+0.51^{\circ}\text{C}$ above common) became the 5th warmest considering that 1901. The annual rainfall over the united states as an entire become 110% of its Long Period Average (LPA) cost for the length 1961-2010 at some stage in the yr. The monsoon season rainfall over the United states as an entire becomes 109% of its LPA.

The seasonal rainfall at some stage in the Northeast monsoon season (October - December) over the NE Monsoon middle vicinity of the southern peninsula becomes 110% of its LPA. All the 5 subdivisions of the middle vicinity besides Kerala acquired excess/everyday rainfall. North-East India, which generally gets heavy rainfall at some stage in the monsoon months (June-September), has modified its man or woman for the worse. The flood-drought cycle now has all started to manifest inside a yr, specially at some stage in the monsoon. The rains are available in short bursts and flood the vicinity, observed via means of elongated dry durations that border on drought.

4. Southwest Monsoon Rainfall in Karnataka :

Karnataka remains seeing wet days and the October averages can also additionally have soared beyond ordinary for lots of parts. But as a way because the southwest monsoon (SWM) goes, the State has witnessed a comparable sample over the past 5 decades, barring minor variations. According to statistics from the Karnataka State Natural Disaster Monitoring Center (KSNDMC), the State has recorded an ordinary SWM from 2019, with +14% in 2019 and +17% in 2020, and -8 % in 2021 (ordinary rainfall). This, officers stated, turned into because of favorable elements which have been impacting the monsoonal winds, in flip bringing wet clouds over Karnataka because of large-scale phenomena along with the superiority of fine Indian Ocean Dipole in the course of 2019, La Nina/impartial circumstance in the course of 2020 and 2021.

KSNDMC officials[ref.10] admitted that even though the State as an entire acquired ordinary rainfall, distribution of rainfall performs a chief role, which can have severe occasions along with floods in a single element and dry spells in some others. “For the SWM 2021, we've 25 districts displaying ordinary rainfall, 3 districts which have acquired extra rainfall, and 3 are deficient,” stated an official. While the extended monsoon enables in lowering moisture stress, contributes to the second one crop for a specific season, and is right for reservoir and water conservation structures, above ordinary and under ordinary rainfall for a non-stop duration ought to bring about herbal calamities along with floods and droughts, main to monetary loss, they stated. Officials admitted that during current years, severe occasions were located more frequently, and green making plans and preparedness have been had to limit their effects.

4. Implementation

We have divided our undertaking into modules(components).

- Gathering facts and changing it into readable form.
- Using facts to construct a prediction model

4.1 Module -1 :

In the primary part of the undertaking we began accumulating facts from numerous assets just like the net and books. The essential part of our studies paintings become accumulating facts about rainfall in one-of-a-kind components of India. We desired to manually apprehend the rainfall pattern.

As an additional step in our project, we commenced searching out rainfall statistics withinside the preceding years. After special studies on one of a kind structures like Kaggle. With the assistance of our professor, we accumulated bulk statistics for the year 2018 in .nc form. In this format, we can now no longer be capable of studying the document for this reason we want to transform it into CSV format. Below are the stairs to transform the .nc document to CSV format.

Procedure to convert NC file to. CSV file

Step 1- Downloading netCDF4 Python library.

```
pip install netCDF4
```

Step 2 - In the same folder as your .nc file create a python file.

Step 3 - Reading the netCDF4 file

Create a variable and store the .nc file's path as arguments for Dataset().

```
data= Dataset(r'C:\ncdf_files\RF25_IMD0p252018.nc','r')
```

Step 4 - Displaying the names of the variables and Accessing the variables.

Print down the key variable and create a separate variable to hold the values in the keys.

```
print(data.variables.keys()) time = data.variables['TIME']
```

Step 5- Accessing the data from the variables. Store the data in the variables and print it.

```
lat_data = data.variables['LATITUDE'][:] print(lat_data)
```

Step 6 - Finding out the coordinates of latitude and longitude of a particular region.

We can find out the latitude and longitude of a place from Google Earth Pro. We have to open the properties of a particular region for the same.

Step 7- Storing the values of latitude and longitude in the python file.

Create a new entry for latitude and longitude and add the respective coordinates.

```
lat_hubli = 15.371598 lon_hubli = 75.101166
```

Step 8 - Squared difference of latitude and longitude.

Finding the closest difference between the coordinates and value of data in the variables.

```
sq_lat_dist = (lat - lat_hubli)**2 sq_lon_dist = (lon - lon_hubli)**2
```

Step 9 - Identifying the index of minimum value of latitude and longitude This can be done using the `argmin()` function of the NumPy library.

```
min_index_lat= sq_lat_dist.argmin() min_index_lon= sq_lon_dist.argmin()
```

Step 10 - Storing rainfall data in a temp variable.

We can use this to print the features of the data. The rainfall data in the particular .nc file takes three arguments i.e, (index of time, latitude, longitude).

Step 11- Defining the data range.

Here we store the starting and ending dates in variables either by indexing or by entering particular dates. We can use this to define the range of the data and scrap the information in a data frame using pandas.

```
starting_date = '2018-01-01' ending_date = '2018-12-31' date_range = pd.date_range(start  
= starting_date , end = ending_date)
```

Step 12 - Creating an empty data frame using pandas.

```
df = pd.DataFrame(0, columns = ['Precipitation'], index = date_range)
```

Step 13 - Creating a variable and giving it a size same as time variable in the .nc file We are going to simply store values from 0 to the size of the time variable.

```
dt = np.arange(0, data.variables['TIME'].size)
```

Step 14 - Loop over the time variable and store the precipitation values.

We have used a for loop to loop over that and store the temp variable for particular places.

```
for time_index in dt: df.iloc[time_index] = temp[time_index, min_index_lat, min_index_lon]
```

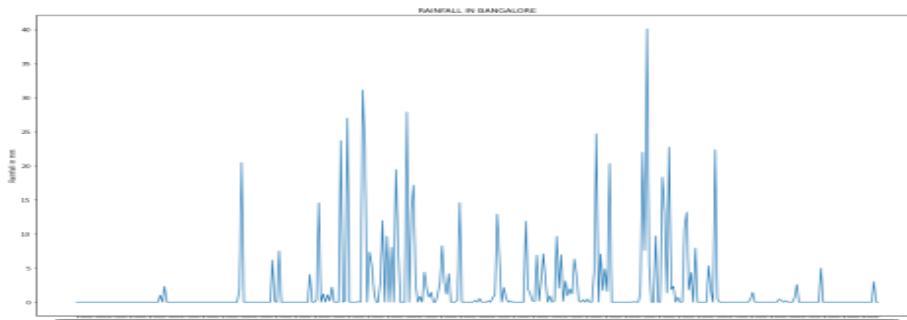
Step 15- Now save this to a CSV file.

```
df.to_csv('Rainfall_In_hubli.csv')
```

After converting the NC file to the CSV file (readable format) , the next step is to analyze and compare the rainfall in the cities considered (Bangalore, Vizag, Hubli).

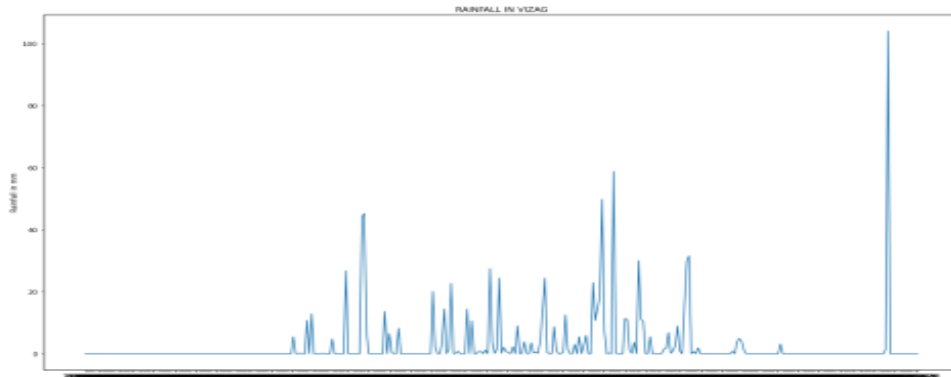
After successfully converting the .nc file to CSV format we could read the rainfall pattern of every particular day during 2018. The precipitation is shown in mm. Now we could derive graphs on all the different cities comparing the rainfall pattern.

Bangalore:



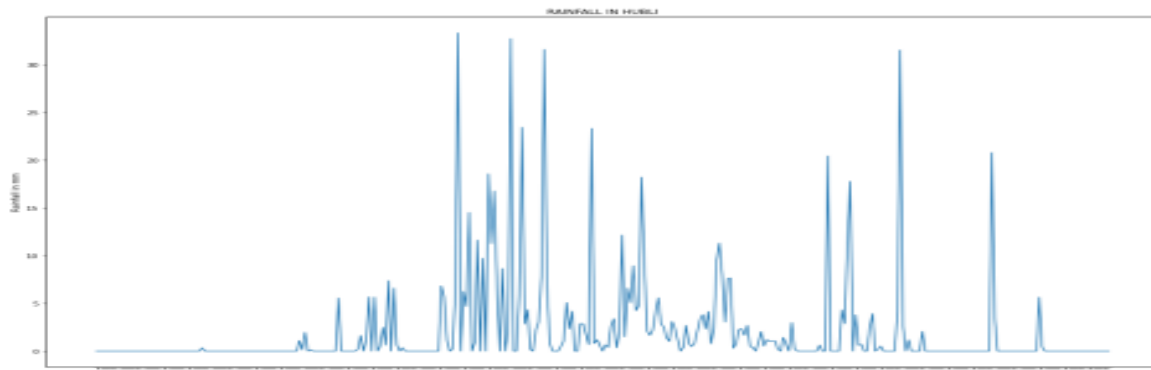
The highest recorded rainfall in Bangalore is around forty mm(40mm).

Vizag:



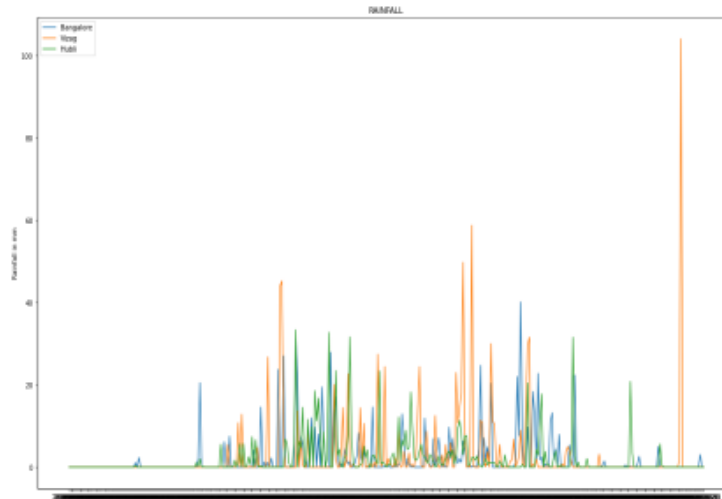
The maximum rainfall recorded in Vizag changed into more than a hundred mm. This befell because of the superiority of a cyclone in December.

Hubli:



The maximum rainfall recorded in Hubli is 30-40 mm.

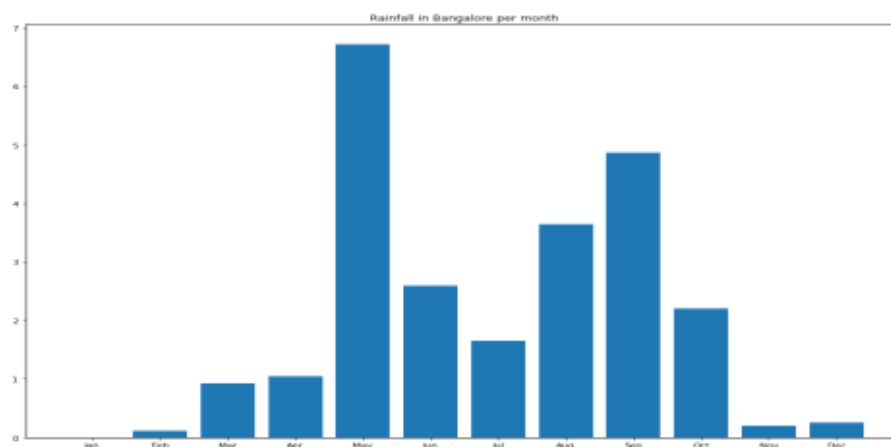
Rainfall in three cities:



This offers us a comparative evaluation of rainfall in 3 cities. It may be stated that the rains start in degrees all through the month of April and grow till the end of August. Vizag shows a unique practice session as December rains factor to a marvel peak.

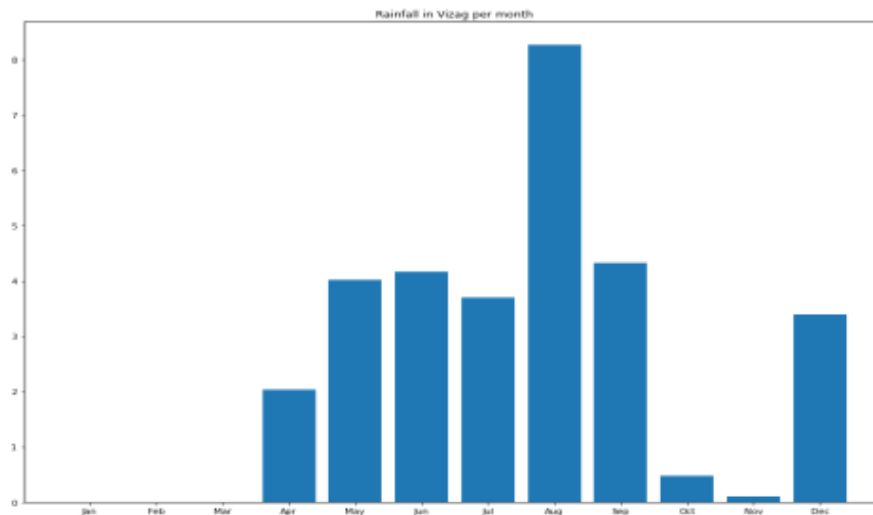
Analyzing the average monthly rainfall in each city considered:

1. Bangalore:



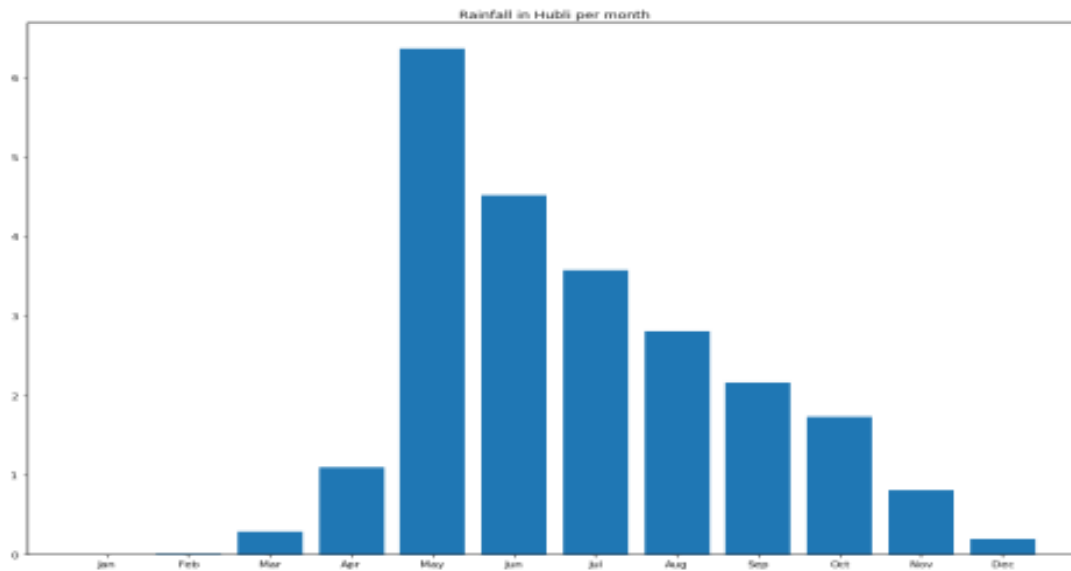
According to the data, It is observed that Bangalore experienced its highest rainfall during May. It can be observed that there was a peak in May and gradually went down during June and July which is supposed to be the monsoon period with more rainfall. Then there is a rise in August and more rainfall in September. Later it is seen to gradually decrease.

2. Vizag:



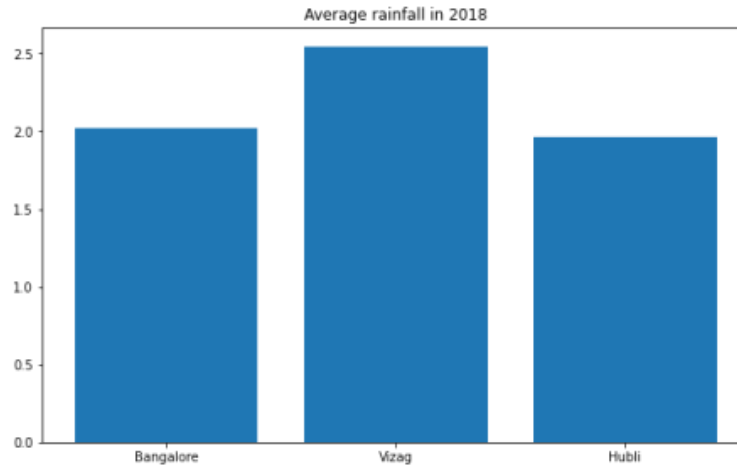
In 2018 we can conclude from the graph above that the rains became almost zero until April, in which the unexpected rains in April increased in May and stayed the same in June. It eased off somewhat in July, but peaked in August with the heaviest rainfall. It was partially reduced in September and significantly reduced in October and November. But for a realistic amount of rain in December.

3. Hubli:



In 2018, Hubli confronted minimum rainfall throughout March which barely elevated in April. There turned into an unexpected upward thrust throughout May which had the maximum rainfall of that year. It decreased in June and really regularly went on decreasing. It is determined that there's an indistinct sample of rainfall in Hubli this year. The rainfall turned in nonetheless there although the monsoon duration turned into over.

Comparative bar graphs w.r.t Bangalore, Vizag, Hubli



4.2 Module - 2:

This is the part that we carried out later wherein we have used the derived data to predict the rainfall pattern in the upcoming period.

Firstly we feed all the CSV files containing rainfall data of different cities.

TASK 1: Sum of annual rainfall

Step 1: Feeding the data of every city in different lists.

Creating separate lists containing annual rainfall of different cities.

```
b_r = list(blore_data['Rainfall in mm'])
h_r = list(hubli_data['Rainfall in mm'])
v_r = list(vizag_data['Rainfall in mm'])
```

Step 2: Finding the sum of rainfall data

Creating a list having all the cities. Then calculate the sum of the particular lists above and put all of them into a new list.

```
cities = ['Bangalore', 'Hubli', 'Vizag']
sum_of_annual_rainfall = [sum(b_r), sum(h_r), sum(v_r)]
```

Step 3: Plotting a bar graph considering the above two parameters

Now we plot a bar graph for the above lists. Using a loop to add the lists containing the sum of annual rainfall.

```
plt.figure(figsize=(10,7))
plt.bar(cities,sum_of_annual_rainfall)
for i in range(3):
    plt.text(i,sum_of_annual_rainfall[i],"{:.2f}".format(sum_of_annual_rainfall[i]))
```

TASK 2: Finding means of SouthWest and NorthEast Monsoon for the three cities.

Step 1- Creating a duplicate Dataframe.

Firstly we are creating a duplicate dataframe that consists of rainfall data of all the three cities.

Here we are trying to add a date column as well aligned with the precipitation data of different cities.

```
dup_df = blore_data
dup_df.columns = ['Date', 'rf_blore']
dup_df['rf_hubli'] = hubli_data['Rainfall in mm']
dup_df['rf_vizag'] = vizag_data['Rainfall in mm']
```

Step 2- Slicing the Month from the date.

Further, we are replacing the Date column with the Month it belongs to. Here, we create a list containing the items from the Date column. Then run a loop over it where we slice down the number representing a month and assign the month name replacing the date in the data frame.

```
date = list(dup_df['Date'])
for i in range(len(date)):
    if(date[i][5:7]=='01'):
        date[i] = 'Jan'
    elif(date[i][5:7]=='02'):
        date[i] = 'Feb'
    .....
dup_df['Date'] = date
```

Step 3- Grouping the southwest monsoon months.

We create a list and enter what all months come under SouthWest Monsoon seasons.

Then by running a loop over the date of the rainfall data we break out the rainfall data of the three cities based on the SouthWest Monsoon.

```
south_west = ['May', 'Jun', 'Jul', 'Aug', 'Sep']
blore_south_west=[]
hubli_south_west = []
vizag_south_west = []
for i in south_west:
    k=0
    b=0
    h=0
    v=0
    for j in range(len(date)):
        if(i==date[j]):
            k+=1
            b+=blore_rf[j]
            h+=hubli_rf[j]
            v+=vizag_rf[j]
            if(date[j+1]!=i):
                break
    blore_south_west.append(b/k)
    hubli_south_west.append(h/k)
    vizag_south_west.append(v/k)
```

Step 4- Plotting graphs for Southwest Monsoon season.

Giving the graph a certain size and name. Assigning the x and y coordinates then using a loop plotting a graph for all months in the southwest monsoon season.

```
plt.figure(figsize=(10,7))
plt.title('Bangalore rainfall in south west monsoon')
plt.xlabel('South west monsoon')
plt.ylabel('Average rainfall in mm')
plt.bar(south_west,blore_south_west)
for i in range(len(south_west)):
    plt.text(i,blore_south_west[i],"{:.2f}".format(blore_south_west[i]))
```

Step 5- Plotting a graph for the average southwest Monsoon season for three cities.

Collecting the mean of all the cities in a list and plotting a graph for the same.

```
plt.figure(figsize=(10,7))
```

```
plt.title('Mean rainfall in South West Monsoon of three cities')
plt.xlabel('CITIES')
plt.ylabel('Mean rainfall in mm')
plt.bar(cities,south_west_mean)
for i in range(3):
    plt.text(i,south_west_mean[i],"{:.2f}".format(south_west_mean[i]))
```

Task 3: The Same for northwest monsoon season by changing the months to October, November and December.

Task 4: Calculating the variance for overall rainfall throughout the year.

Step1- Finding the variance of three cities.

Using the statistics function of the python library we can calculate the variance. Finding the variance of all the cities and storing it in a list.

```
blore_variance = statistics.variance(b_r)
hubli_variance = statistics.variance(h_r)
vizag_variance = statistics.variance(v_r)
variance = [blore_variance,hubli_variance,vizag_variance]
```

Step 2 - Plotting a graph for variances of all the cities.

```
plt.figure(figsize=(10,7))
plt.title('Variance of rainfall throughout the year')
plt.xlabel('CITIES')
plt.ylabel('Variance values')
plt.bar(cities,variance)
for i in range(3):
    plt.text(i,variance[i],"{:.2f}".format(variance[i]))
```

TASK 5: Calculate the moving average and find the extremum.

Step 1 - Creating lists to store moving averages.

We have to create three different lists to store the moving average of the particular cities. Also, create an empty list to store the indexes by running a loop for a length of 363.

```

blore_mavg = []
hubli_mavg=[]
vizag_mavg = []
x=[]
for i in range(363):

    x.append(i)

```

Step 2 - Finding the moving average considering three consecutive days.

As we are considering a 3 days moving average there will be around 363 outcome data points. For every iteration of the for loop we are storing the three values in the data and taking the mean of that data and storing that data into the list.

```

for i in range(len(blore_rf)-2):
    ll = blore_rf[i:i+3]
    blore_mavg.append(statistics.mean(ll))

```

Step 3 - Plotting a graph for the moving average of particular cities.

```

plt.plot(x,blore_mavg)

```

ARIMA (Auto-Regressive Integrated Moving Average) is a class of models that 'explains' a time series based on its own previous values, that is, its own lags and lagged prediction errors, so that the equation can be used to anticipate future values.

```

def findParameters(train,test):
    p=range(0,8)
    q = range(0,8)
    d = (0,2)
    pdq_combinations = list(itertools.product(p,d,q))
    order=[]
    aic_list=[]
    for pdq in pdq_combinations:
        try:
            model = ARIMA(train,order=pdq).fit()
            aic = model.aic
            aic_list.append(aic)
            order.append(pdq)
        except:
            continue
    min_aic = min(aic_list)
    index = aic_list.index(min_aic)
    return(order[index])

```

An ARIMA version is characterized by way of means of three terms: p , d , q wherein

p is the order of the AR term

q is the order of the MA term

d is the wide variety of variations required to make the time collection desk bound

' p ' is the order of the 'Auto Regressive' (AR) term. It refers back to the wide variety of lags of Y for use as predictors. And ' q ' is the order of the 'Moving Average' (MA) term. It refers back to the wide variety of lagged forecast mistakes that ought to cross into the ARIMA Model.

AIC is a unmarried wide variety rating that may be used to decide which of a couple of fashions is maximum probable to be the high-quality version for a given dataset. It estimates fashions relatively, which means that AIC rankings are most effectively beneficial in contrast with different AIC rankings for the equal dataset. A decreased AIC rating is better.

Firstly the statistics ought to be cut up into education and checking out statistics. As the statistics taken into consideration is most effective south west monsoon (can also additionally - sept) the education statistics length taken into consideration is 122 and the checking out statistics length is 31.

In the ARIMA Model first of all we've to test whether or not the statistics is desk bound or no longer. To take a look at the statistics whether or not desk bound statistics or now no longer a statistical technique is finished, that is an adfuller check. If the p -price withinside the adfuller check is under zero.05 then the statistics is taken into consideration desk bound. So a feature is created which plays the adfuller check and returns the p -price of the adfuller check.

In our case, the 3 cities' rainfall statistics are long gone via adfuller check and all 3 of them have a price much less than zero.05 which states that the datasets are desk bound. So we are able to continue to the following step that is locating the high-quality parameter values (p, d, q) of the ARIMA version.

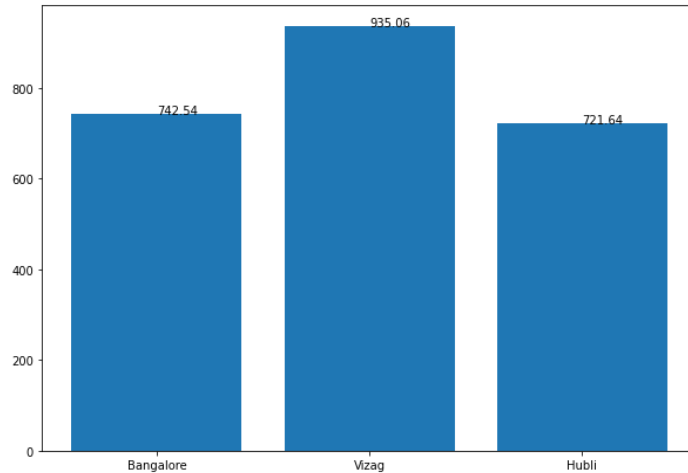
So to locate the parameters of the ARIMA version a feature is created wherein it'll teach statistics and check statistics as parameters. The AR (p), MA (q) price variety is taken into consideration as zero to 7 and the I (d) price degrees from zero to 2. Every viable mixture of p, d, and q withinside the sure limits already given is saved in a listing. There are 128 viable combos p, d, q had to carry out modeling. For each new release withinside the listing wherein the combos are saved, we're feeding the version with new parameters of p, d, q, and becoming the version. After becoming the version, take a look at for the aic rating of the version and keep the order of parameters in a listing and additionally keep the aic rating in some other listing. After finishing all of the iterations take a look at for the least aic price withinside the listing wherein all aic values are saved. The index of the least aic rating ought to be saved as this is the index of the order of parameters for which the version is giving the much less aic rating. At last, this feature will go back to the listing which includes the p, d, q values.

So the parameter listing back from the above feature is used to teach the ARIMA version and that version is used to expect the check statistics.

5. Result and Discussion

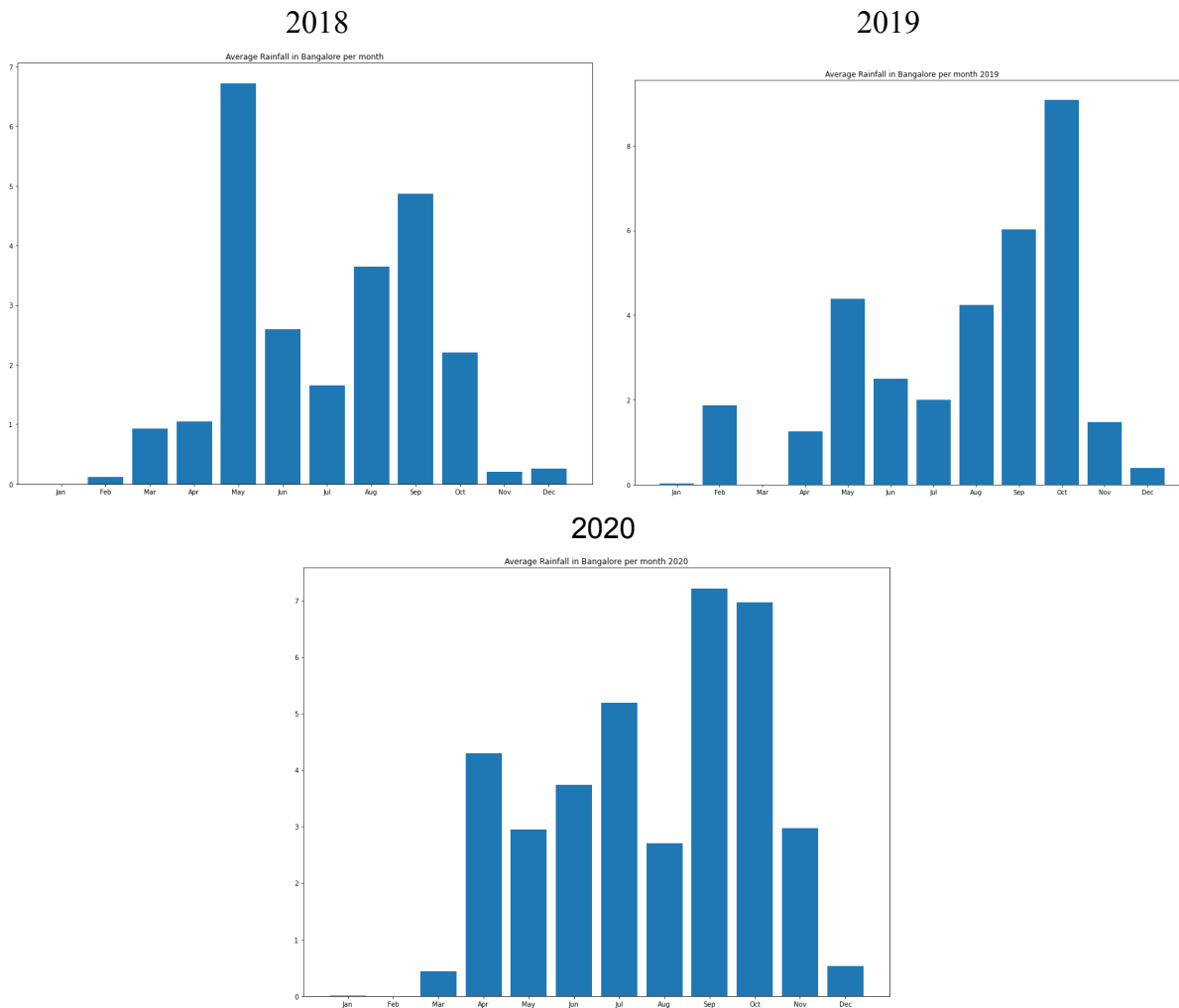
By using the rainfall data, we could successfully complete and conclude the above tasks.

- 1) The sum of annual rainfall of all cities in 2018.



The above graph shows the annual rainfall of the three cities where Bangalore receives about 742.54mm of rainfall, Vizag 935.06mm and Hubli experiences a rainfall of 721.64 annually according to the data from 2018. We can see that, vizag experiences the most average annual rainfall in that year when compared to the other two cities.

2) Monthly Average rainfall is different in Bangalore in three years.



From the above graphs we can see Bangalore experiencing rainfall in the three years.

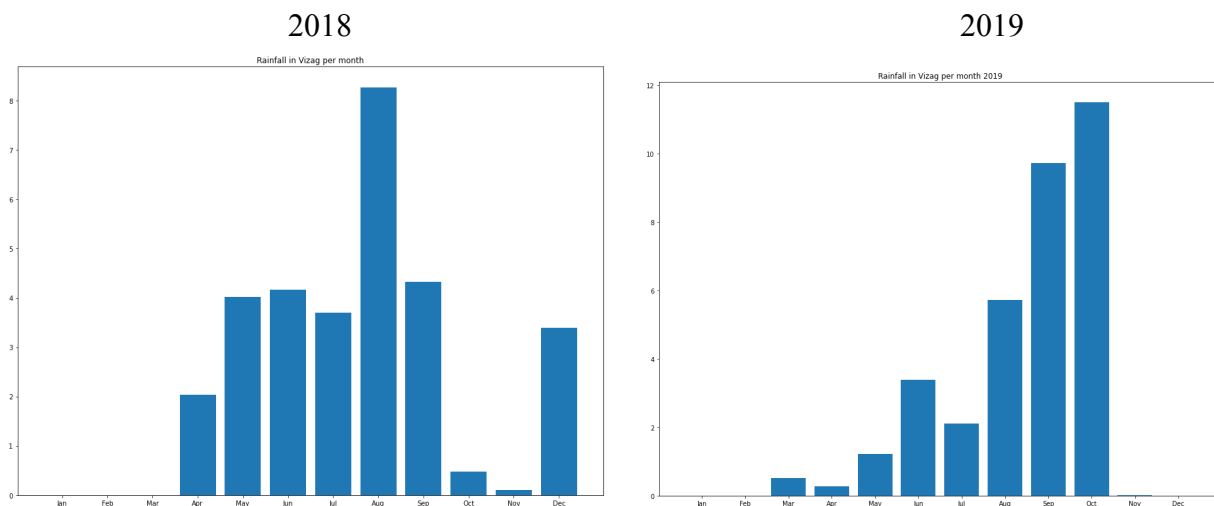
It is evident that during Jan there is no rainfall or there is very little rainfall in all the three years.

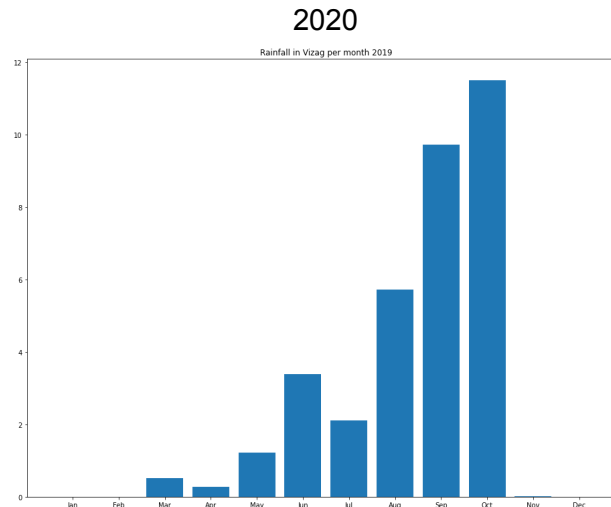
Whereas in the month of february there is a negligible amount of rainfall during 2018 and 2020, but in the year 2019 the rainfall is more compared to the other two years which is about 2mm. It is seen that one particular day there 58mm which was the highest rainfall in february from the past 10years. It was predicted due to the formation of a trough extending from TamilNadu to Telangana including southern parts of Karnataka. This particular difference is caused due to heavy rainfall on a particular day. Hence we can ignore it for prediction.

For the month of March, all of the 3 years are comparable and comply with the identical sample. The common rainfall is 1-2mm. We can finish with the aid of using pronouncing that March follows the identical sample and now no longer plenty of versions are observed. In the month of April, at some stage in 2018 and 2019 there's a minimum common rainfall that is about the same. But withinside the 12 months 2020, April studies a unexpected more common while in comparison to the rainfall at some stage in March withinside the preceding years. The common is extra than 4mm. On research, it's far said to be the “wettest” April in seven years and claims to be due to cyclonic flow off the coast of Tamil Nadu.

After this all of the month belongs to either the southwest Monsoon season or northeast monsoon season. Below is the distinctive cause of the same.

3) Monthly Average rainfall is different in Vizag in three years.





The above graph shows the annual rainfall in 3 years 2018, 2019, 2020 for the city of Vizag city. In the year of 2018 , the rainfall was very fluctuating throughout the year . In the month of Jan , Feb & Mar there is no rainfall at all. And in the year of 2019, the months Jan, Feb had no rainfall and March did see a little rainfall of about 1.26 mm of rainfall. But the year of 2020 had a good start with the average monthly rainfall starting from the month Jan with 1.12 mm, for Feb with the average monthly rainfall of 2.00 mm and for March 1.8 mm.

There is a variation across all the months for all the years and it is seen that in the year 2018, the average is too high which is caused due to rainfall throughout the month. In the month of April, the year 2018 the average monthly rainfall is 2.1 , for 2019 it is about 1.3 mm and for 2020 it is around 3.67 mm. It shows the idea that the course of time is not exactly the same and easily predictable.

In the year 2018, the average monthly rainfall in May is about 3.90 mm. In the year 2019, the average monthly rainfall is 1.45mm. In the year 2020, the average rainfall is 2.0mm. And in the month of May 2018 it had the maximum average monthly rainfall compared to all the 3 years. In the month of June , the average monthly rainfall is zig zag from 2018 to 2020. With the highest rainfall in 2020 with 4.0 mm , the average monthly rainfall being 3.5 mm for 2019 and 4.4mm for the year 2020.

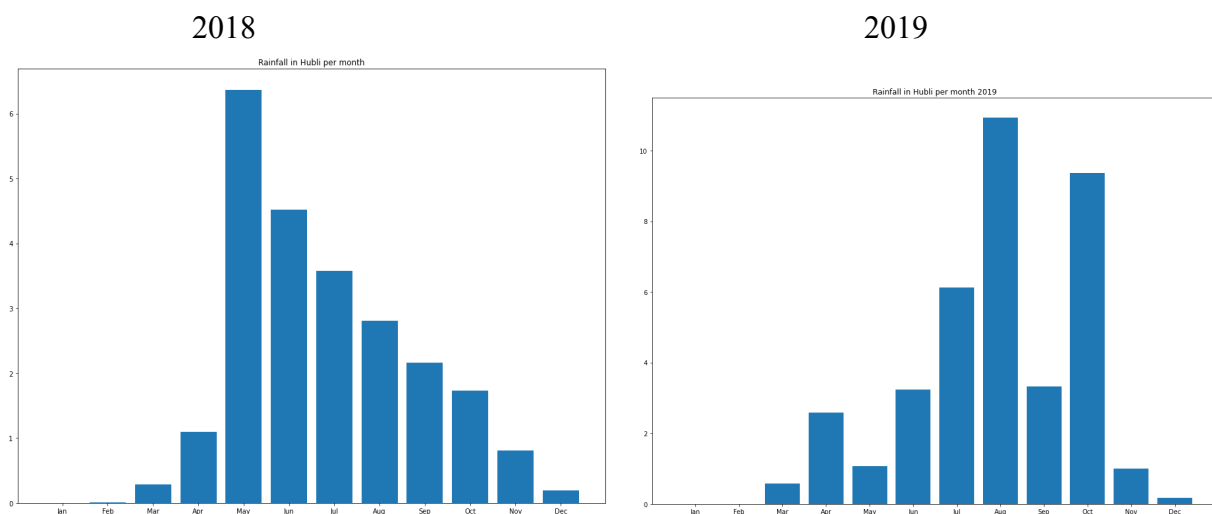
In the month of July , as we will see that the common month-to-month rainfall is fluctuating from 2018 to 2020 with the very best rainfall in 2018 of three.sixty seven mm and sizeable lower withinside the yr of 2019 of approximately 176 mm common month-to-month rainfall and and once more with a

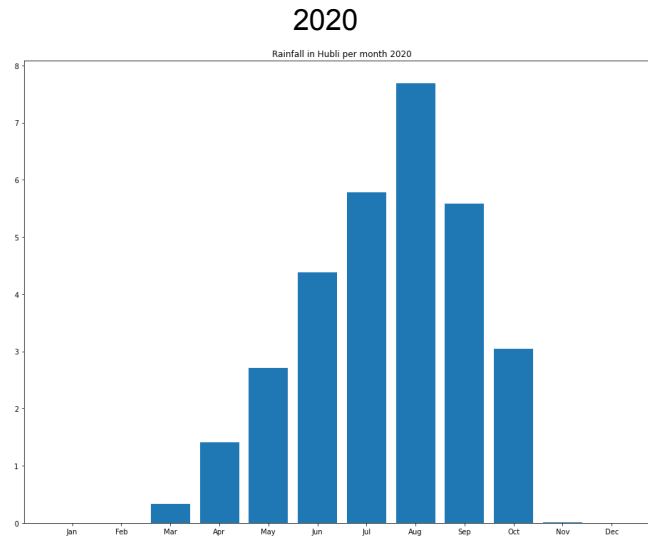
moderate boom which debts to 2.four mm for the year 2020. In the month of August, the common month-to-month rainfall is progressively reducing from 2018 to 2020. With the very best rainfall in 2018 with 8.zero mm now no longer best this being the very best of three however the maximum of all of the month-to-month rainfall on this yr 2018 , the common month-to-month rainfall being 5.zero mm for 2019 and 2.three mm for the year 2020.

In the month of September, as we can see that the average monthly rainfall is fluctuating from 2018 to 2020 with the highest rainfall in 2019 of 9.0 mm and significant decrease in the year of 2020 of about 5.0 mm average monthly rainfall and again with the least which accounts to 4.0 mm for the year 2019. In the month of Oct , there is a drastic increase in the monthly average rainfall from the year 2018 to 2020. In 2018 it showed the least rainfall of 0.75 mm , In 2019, it showed a drastic increase in the amount of rainfall of 9 mm. In 2020 it had further increased to 12.3 mm.

In the month of Nov, the years 2018, 2019 had a very meager amount of monthly average rainfall, but in the year of 2020 it showed a significant increase. But it is completely the other way round as in the case of Dec, the year 2019, 2020 doesn't have rainfall at all. In 2018 there was a significant amount of rainfall.

4) Monthly Average rainfall is different in Hubli in three years.





The above graph indicates the month-to-month annual rainfall in three years 2018, 2019, 2020 for the Hubli metropolis. In the Febmonth of the 12 months 2018, the rainfall turned into very meager, nearly negligible. And it's far from an unusual place for all of the three years in Jan month and Feb month for 2018 and 2022 that there's no rainfall at all. In the month of March the month-to-month rainfall of all of the years i.e. from 2018, 2019, 2020 all are nearly equal that is 0.45 mm.

But in the month of April we have a dim and dip again with the monthly rainfall of 2018 as 1.0 mm and for 2020 as 1.35 mm which is almost similar for both. But there is a slight increase in the monthly rainfall of 2019 as it has recorded a value of 2.9 mm. In the month of May the Hubli city saw the highest rainfall of the year 2018 with 6.4 mm. But it ultimately decreased very rapidly in 2019 to 1.67 mm and it had a significant rise in the year 2020 with 2.87 mm as monthly rainfall. This shows that we can't anticipate things on our own without proper research and analysis.

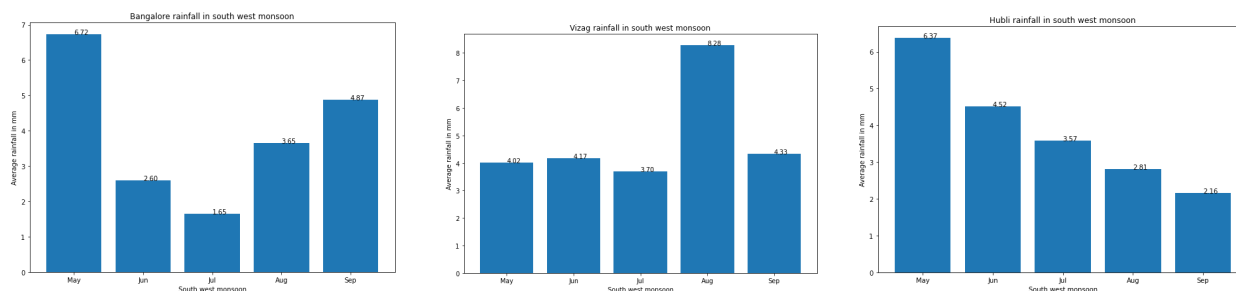
In the month of June, we can see from the above graph that the years 2018 and 2020 had almost the same amount of monthly rainfall with 4.5 mm and 4.45 mm respectively. In 2019, the monthly rainfall amounts to 3.0 mm. The month July has a different trend, as it had the highest record in the year of 2018 with 3.5 mm of monthly rainfall. In 2019, the monthly rainfall amounts to 6.1 mm and in 2020 the monthly rainfall amounts to 5.5 mm. The August monthly rainfall data is the highest ever in 2019 with

10.5 mm. But in the year of 2018 the monthly rainfall is 2.87 mm and in 2020 the monthly rainfall accounts to 7.5 mm.

The month September had an increasing trend, as it had the highest record in the year of 2020 with 5.5 mm of monthly rainfall. In 2018, the monthly rainfall had the least and it amounted to 2.34 mm and it had a significant increase. In 2019 the monthly rainfall amounts to 3.5 mm .

In the month of October ,the annual monthly rainfall in 2018 is the least of all, which is 1.85 mm. And in the year 2019, it had a significant increase in this year compared to last year which is 9.0 mm. The monthly rainfall in 2020 is 3.0 mm. In the month of November ,the monthly rainfall started decreasing eventually from 2018 to 2020 which is 0.87 mm in the year 2018 , in 2019 it is 1 mm and in 2020 it accounts to 0.35mm. In the December month, the year 2018 had some amount of rainfall like 0.25 mm , in the month of Dec of 2019 there was a small decrease compared to last year, the year 2020 had no rainfall at all.

5) South West Monsoon season in the three cities.



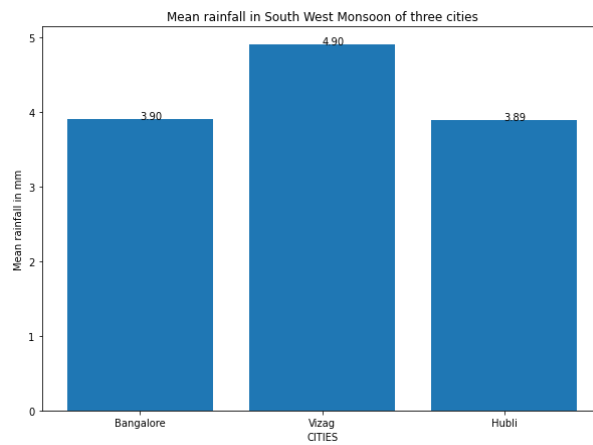
The above graph represents the southwest monsoon season(May- June) and the rainfall during that season. It can be observed that the average rainfall in Bangalore during may was 6.72, june it was 2.60, july it was 1.65, aug it was 3.65 and in sept it was 4.87. The month with the most rainfall was observed to be May. Also, June and July had the least rainfall.

It can be observed that the average rainfall in Vizag during May was 4.02, June it was 4.17, July it was 3.70, Aug it was 8.28 and in September it was 4.33. Vizag has the highest average rainfall during the month of Aug. But a very notable pattern is the lowest average rainfall of the southwest monsoon season is July and there is a sudden rise observed in Aug but it also comes down in the month of

September.

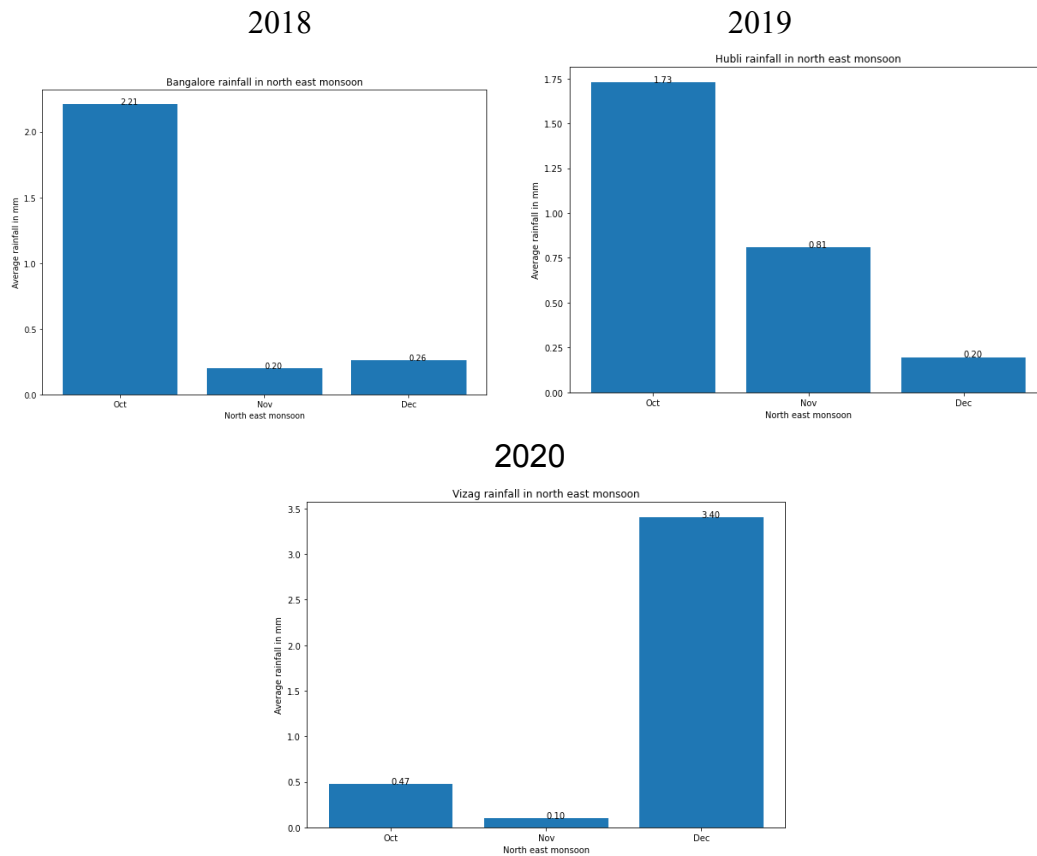
It can be observed that the average rainfall in Hubli during may was 6.37, june it was 4.52, july it was 3.52, aug it was 2.81 and in sept it was 2.16. Here Hubli experienced a peak in May which gradually decreased till September.

6) The mean rainfall of the southwest monsoon season of all the three cities.



In the mean rainfall experienced by the three cities during the southwest monsoon season graph, we can observe that Vizag experienced the highest rainfall of 4.90 and the other two cities are having nearly the same means.

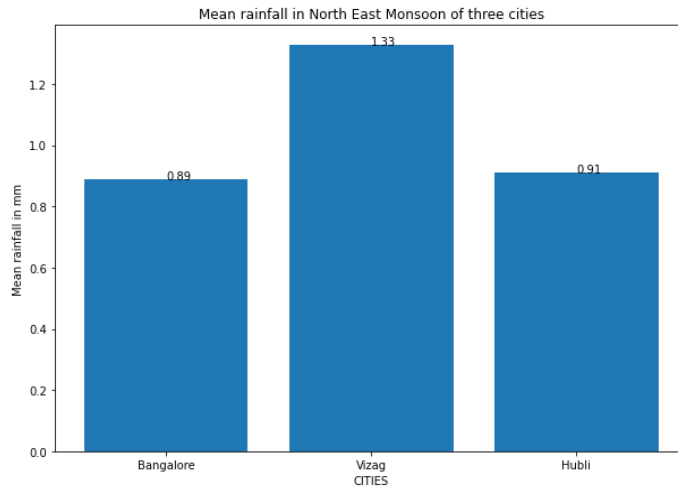
7) The rainfall experienced by the cities during the northeast(Oct-Dec) monsoon season.



The above graph shows us the average rainfall experienced by the three cities during the northwest monsoon. It can be observed that the average rainfall in Bangalore during Oct was 2.21, Nov it was 0.20 ,and in Dec it was 0.26.

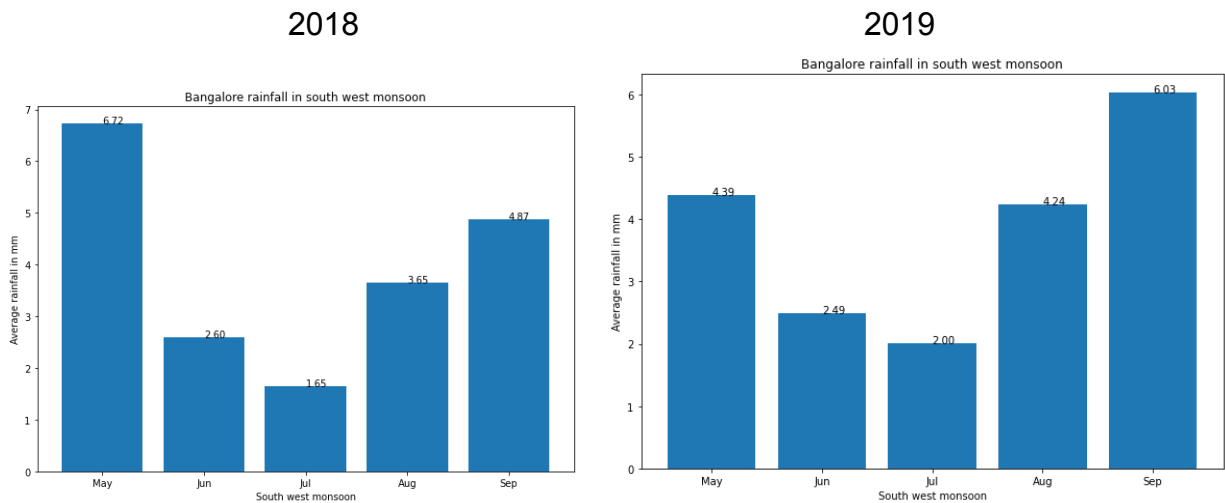
It can be observed that the average rainfall in Hubli during Oct was 1.73, Nov it was 0.81 ,and in Dec it was 0.20. It can be observed that the average rainfall in Vizag during Oct was 0.47, Nov it was 0.10 ,and in Dec it was 3.40.

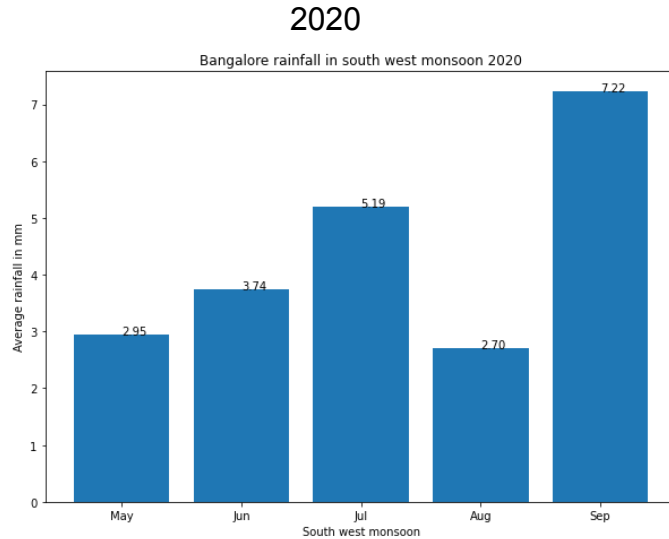
8) This means the average rainfall experienced by the three cities during the northeast monsoon season.



It can be observed that when we take the mean of the rainfall experienced during northwest monsoon season Vizag again experienced the highest rainfall and the other two cities have been facing considerably similar amounts of rainfall.

9) The southwest monsoon season in Bangalore with respect to the different years.





The graph above shows Bangalore's rainfall during the southwest monsoon season. In 2018, the average rainfall in May is about 6.72 mm. In 2019, the average rainfall is 4.39. In 2020, the average rainfall is 2.45. There is variation over the years and it can be seen that the average in 2018 is too high due to the rainfall during the month.

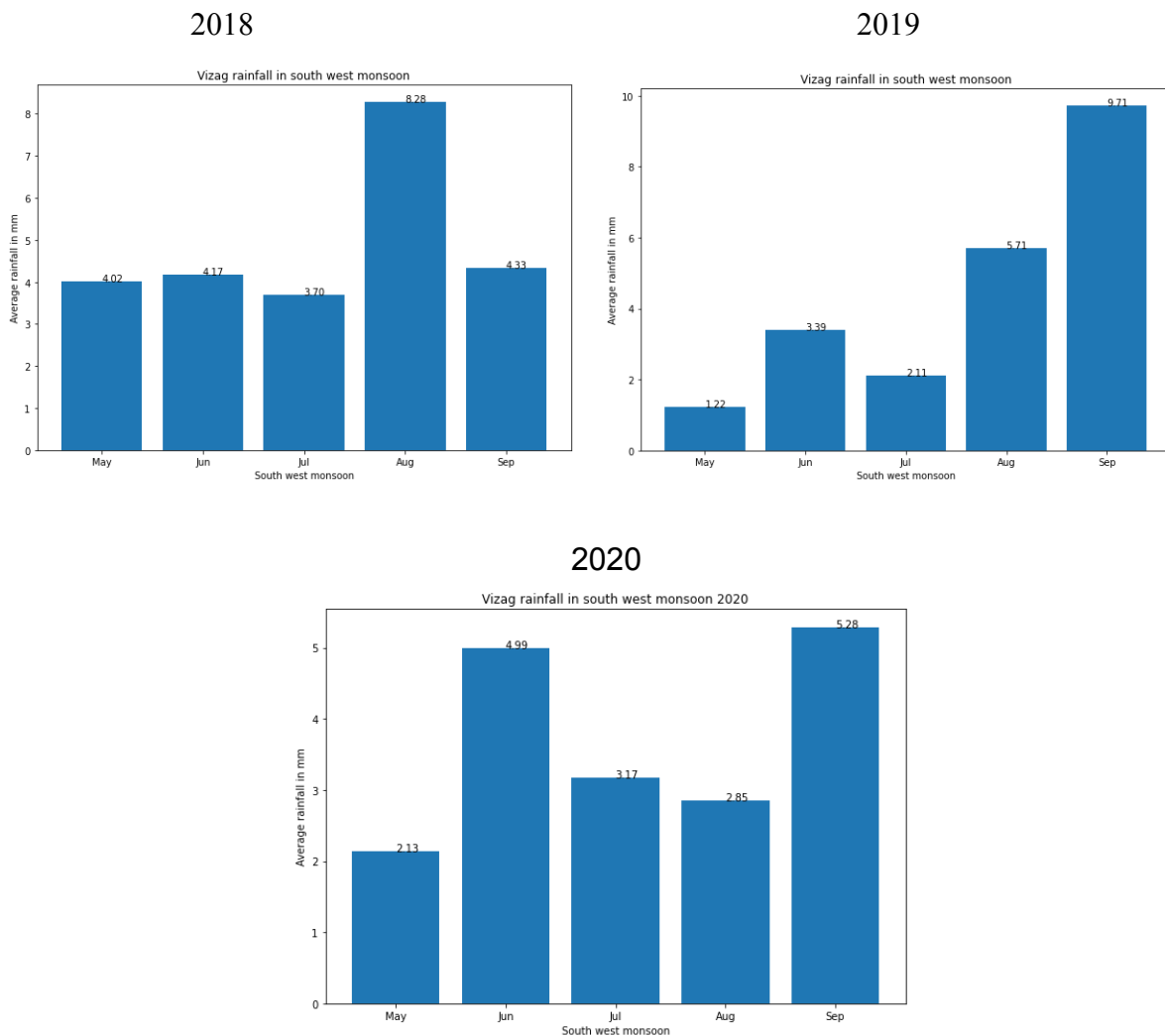
In the month of June, for the year 2018 the average rainfall is about 2.60mm, for 2019 it is about 2.49mm and for 2020 it is around 3.74mm. For the month of June the pattern is similar for all the three years. We can consider this for the analysis as the average is following almost the same pattern. In the month of July, the year 2018 has an average rainfall of about 1.65mm, in the year 2019 the rainfall average is about 2.00mm and in the year 2020, the average rainfall is 5.39mm. It can be observed that during 2020 there is a comparatively higher average rainfall.

In the month of August, the year 2018 has an average rainfall of about 3.65mm, in the year 2019 the rainfall average is about 4.24mm and in the year 2020, the average rainfall is 2.70mm. Here the average is again similar through the years. In the month of Sept, the year 2018 has an average rainfall of about 4.87mm, in the year 2019 the rainfall average is about 6.30mm and in the year 2020, the average rainfall is 7.22mm.

As an overall comparison between the average rainfall in the southwest monsoon season through the three years. We can observe that, in the year 2018 there was a greater amount of rainfall during May.

However, the rains reduced soon by June and July. The reason for the reduction in these rains can be attributed with the passage of the well-marked low-pressure area which was prevailing in Southeast Arabian Sea. But the average again rose during August and increased by September.

10) The southwest monsoon season in Vizag with respect to the different years.



The above graphs shows the rainfall in Vizag during the south west monsoon season in the year 2018, 2019 and 2020. The average rainfall in May is about 4.2mm. In the year 2019, the average rainfall is 1.2. In the year 2020, the average rainfall is 2.13mm. In the month of Jun, for the year 2018 the average rainfall is about 4.17mm, for 2019 it is about 3.39mm and for 2020 it is around 4.99mm. For

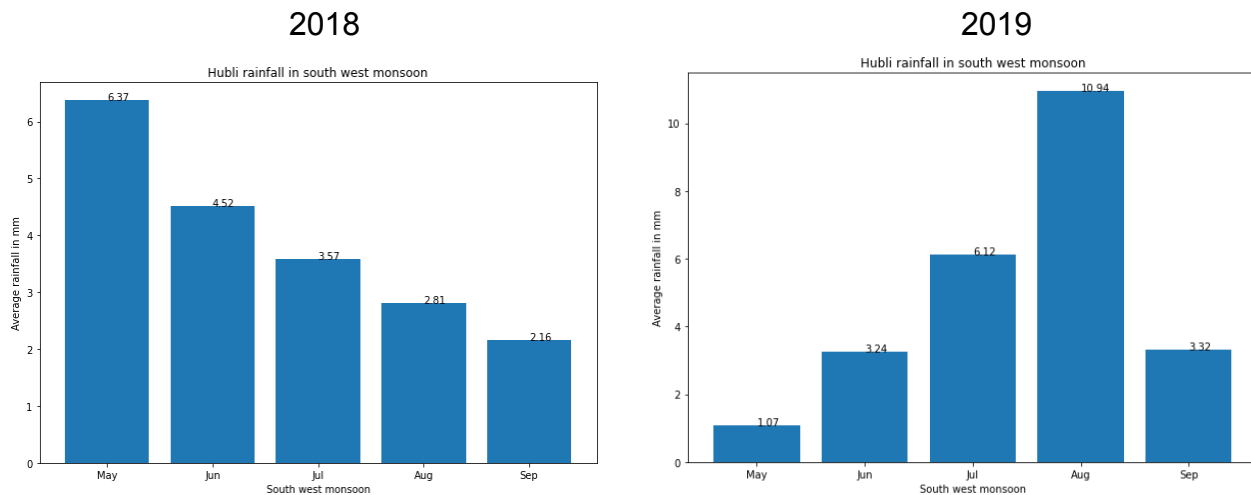
the month of June the pattern is similar for all the three years.

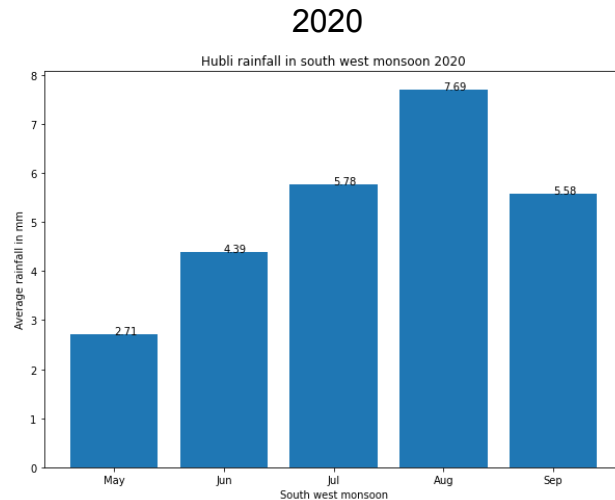
In the month of July, the year 2018 has an average rainfall of about 3.70mm, in the year 2019 the rainfall average is about 2.19mm and in the year 2020, the average rainfall is 3.17mm. The pattern is observed to be similar in all the three years.

In the month of August, the year 2018 has an average rainfall of about 8.28mm, in the year 2019 the rainfall average is about 5.71mm and in the year 2020, the average rainfall is 2.85mm.

In the month of Sept, the year 2018 has an average rainfall of about 4.33mm, in the year 2019 the rainfall average is about 9.71mm and in the year 2020, the average rainfall is 5.28mm.

11) The southwest monsoon season in Hubli with respect to the different years.

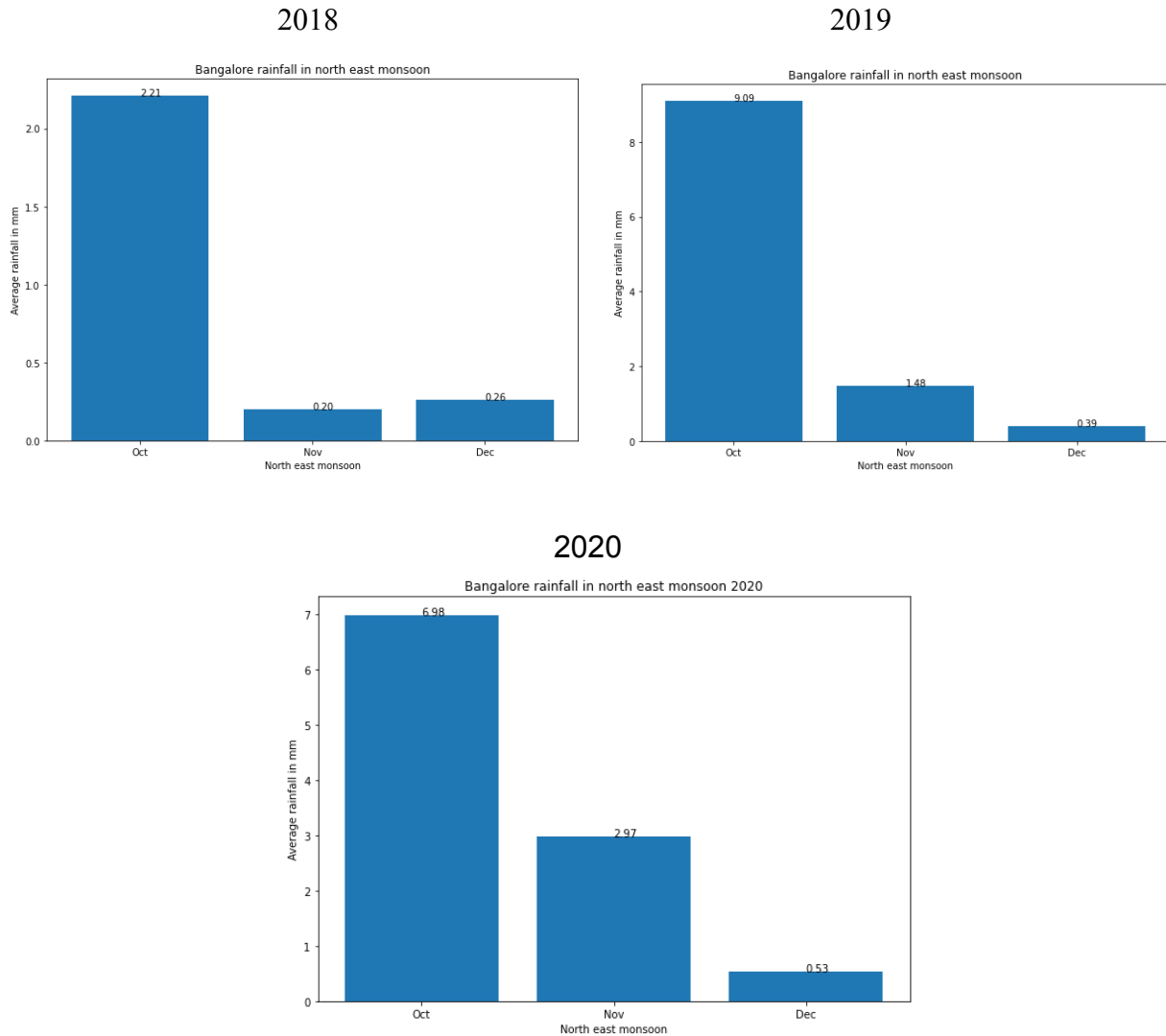




The graphs above depict Hubli's rainfall throughout the south west monsoon seasons of 2018, 2019, and 2020. In May, the average rainfall is 6.37mm. The average rainfall for the year 2019 is 1.10mm. The average rainfall in 2020 will be 2.71mm. For the month of June in 2018, the average rainfall was roughly 4.52mm, 3.24mm in 2019, and 4.39mm in 2020. The pattern for the month of June is consistent across all three years.

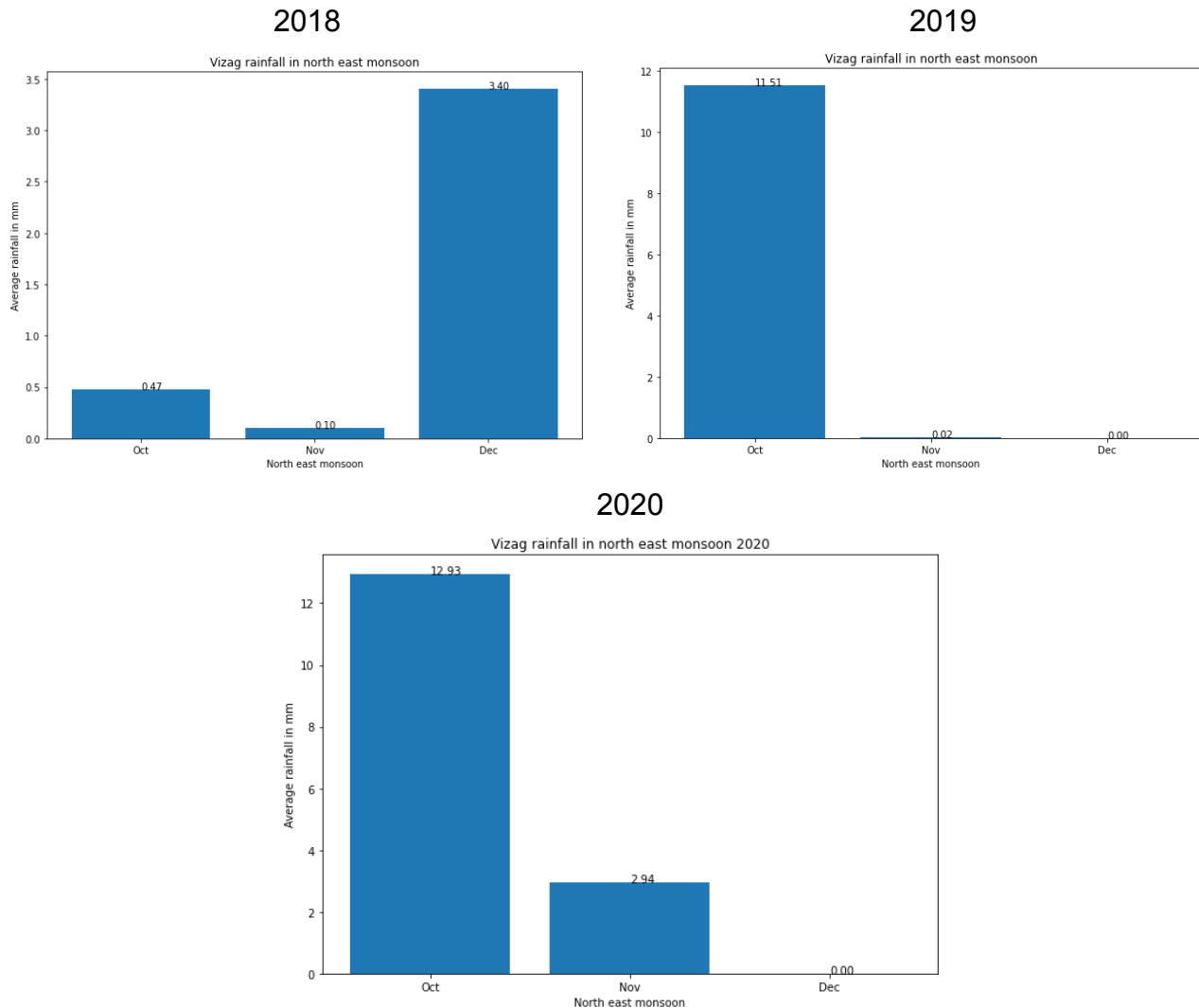
In the month of July, the year 2018 has an average rainfall of about 3.57mm, in the year 2019 the rainfall average is about 6.12mm and in the year 2020, the average rainfall is 5.78mm. The pattern is observed to be similar in all the three years. In the month of August, the year 2018 has an average rainfall of about 2.81mm, in the year 2019 the rainfall average is about 10.83mm and in the year 2020, the average rainfall is 7.69mm. In the month of Sept, the year 2018 has an average rainfall of about 2.16mm, in the year 2019 the rainfall average is about 0.00mm and in the year 2020, the average rainfall is 5.58mm.

12) The average rainfall during the northeast monsoon season in the Bangalore region.



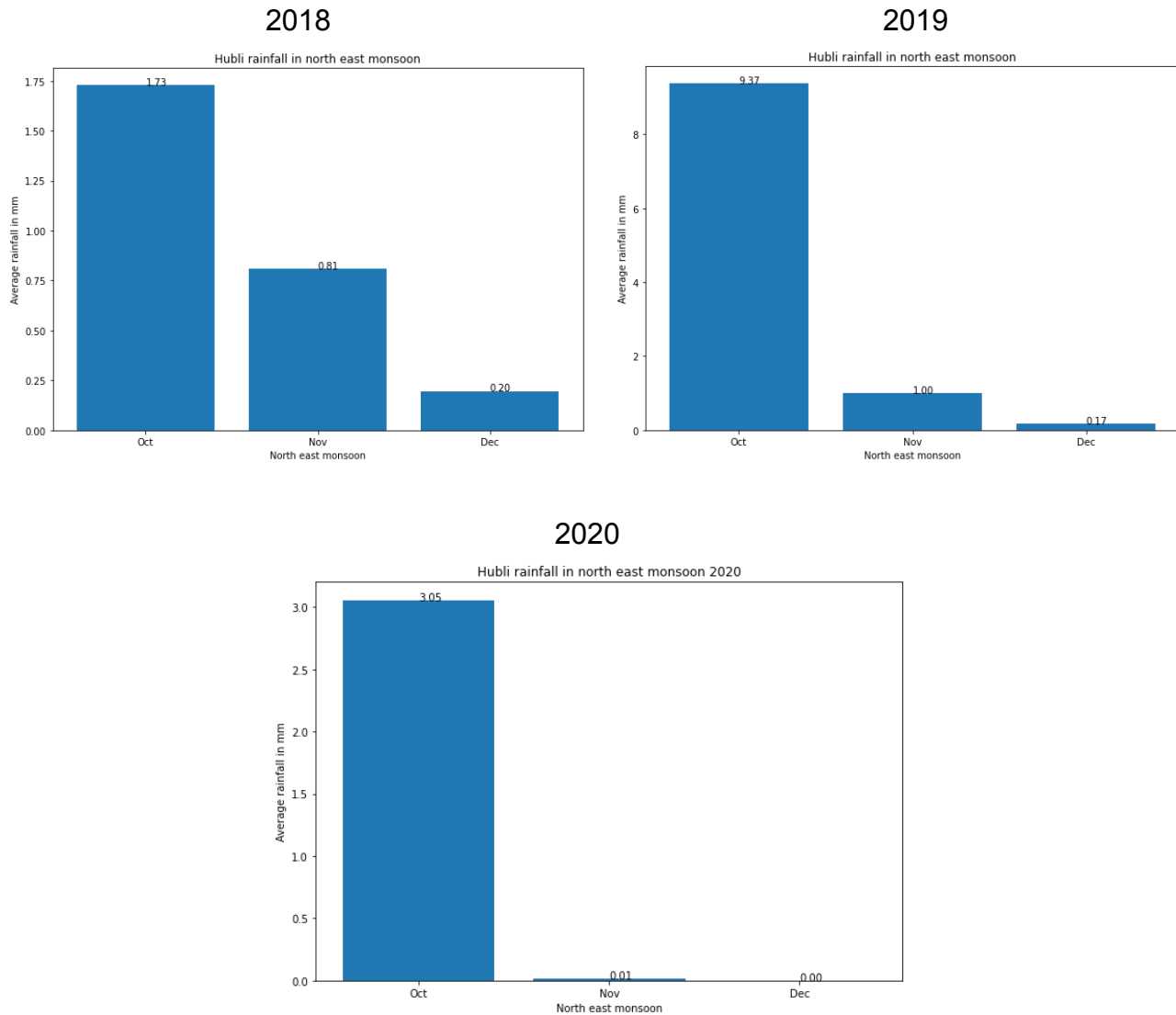
The above graph shows the Rainfall experienced by Bangalore during the NorthEast Monsoon season for the years 2018, 2019,2020. In the year 2018 October, there is a greater average rainfall of 2.21mm, during November the average rainfall is about 0.20mm and in December the average rainfall is about 0.26. In the year 2019, the average rainfall in October is 9.90mm, in November the average rainfall is about 1.48mm and in the month of December it is about 0.39mm. In the year 2020 October, the average rainfall is about 6.98mm, in November the average rainfall is 2.97mm and in December the average rainfall is about 0.53mm.

13) The average rainfall during the northeast monsoon season in the Vizag region.



The above graph shows the Rainfall experienced by Vizag during the NorthEast Monsoon season for the years 2018, 2019, 2020. In the year 2018 October, there is a greater average rainfall of 0.47mm, during November the average rainfall is about 0.10mm and in December the average rainfall is about 3.40. In the year 2019, the average rainfall in October is 11.51mm, in November the average rainfall is about 0.02mm and in the month of December it is about 0.00mm. In the year 2020 October, the average rainfall is about 12.93mm, in November the average rainfall is 2.94mm and in December the average rainfall is about 0.00mm.

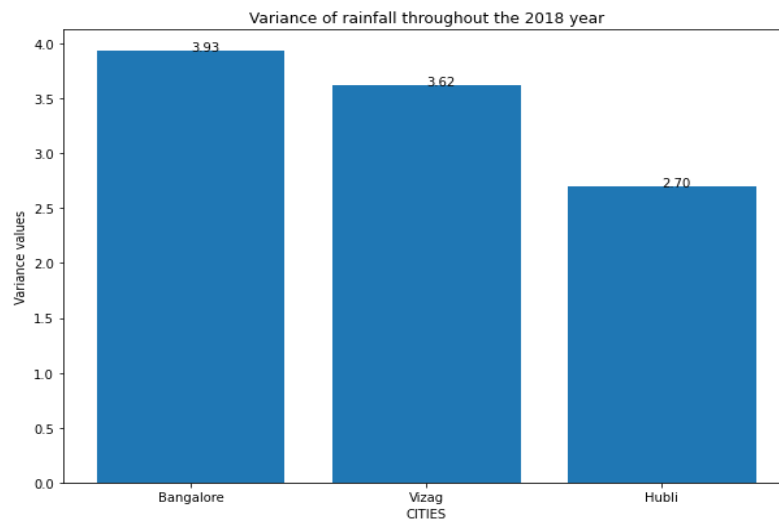
14) The average rainfall during the northeast monsoon season in the Hubli region



The graph above depicts Hubli's rainfall during the NorthEast Monsoon season for the years 2018, 2019, and 2020. The average rainfall in October is 1.73mm, 0.81mm in November, and 0.20mm in December. In the year 2019, the average rainfall in October is 9.37 millimeters, 1.00 millimeters in November, and 0.17 millimeters in December. In the year 2020, the average rainfall in October is approximately 3.05mm, 0.01mm in November, and 0.00mm in December.

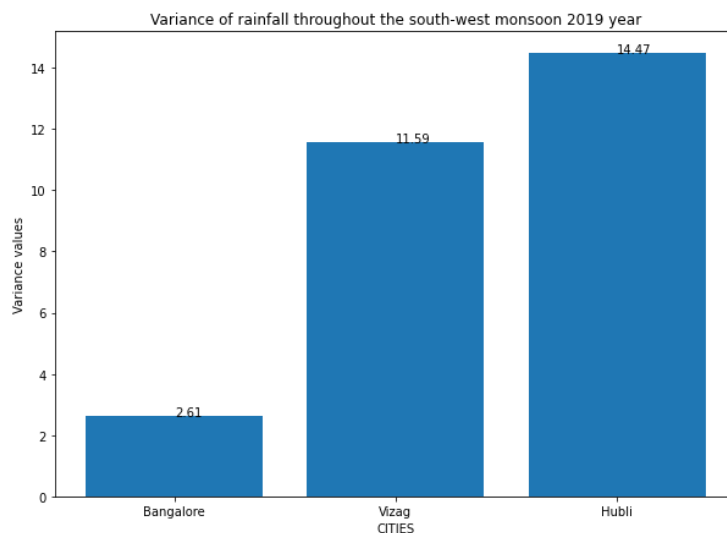
15) Variance of average rainfall throughout the southwest monsoon season.

☐ For the year 2018



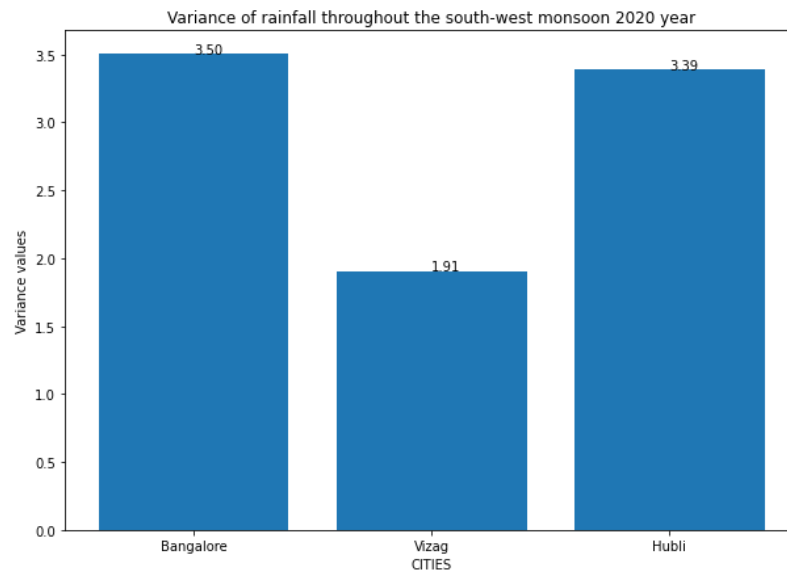
The above graph shows variance of average rainfall of the three cities throughout the southwest monsoon season in the year 2018. We can see that Bangalore has the highest variance of 3.93 , Vizag has a variance of 3.62 and Hubli has the least variance of 2.70.

☐ For the year 2019



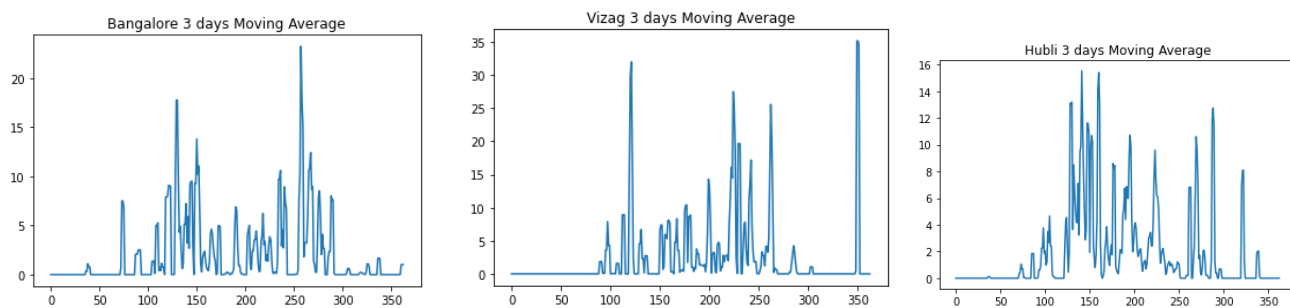
The above graph shows variance of average rainfall of the three cities throughout the southwest monsoon season in the year 2019. We can see that Bangalore has the least variance of 2.61 ,Vizag has a variance of 11.59 and Hubli has the highest variance of 14.47.

□ For the year 2020



The above graph shows variance of average rainfall of the three cities throughout the southwest monsoon season in the year 2020. We can see that Bangalore has a variance of 3.50 ,Vizag has a variance of 1.91 and Hubli has the highest variance of 3.39.

16) The moving averages by considering three days.



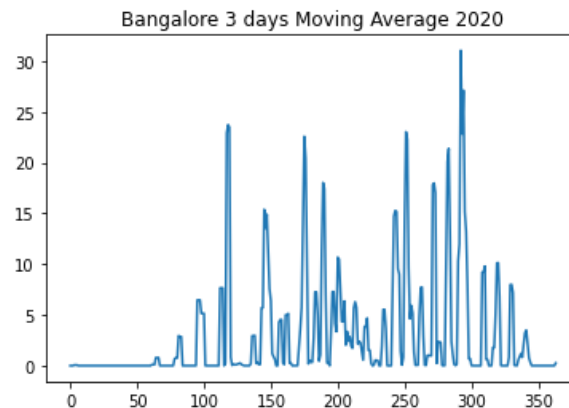
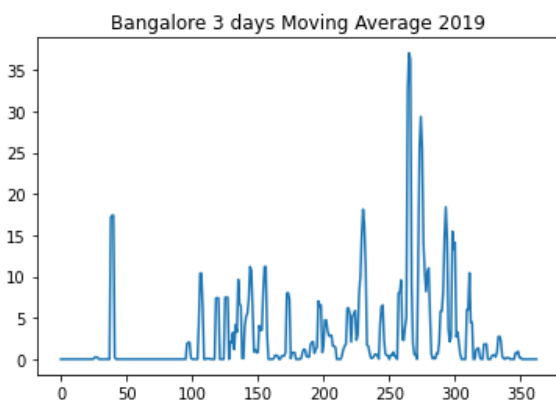
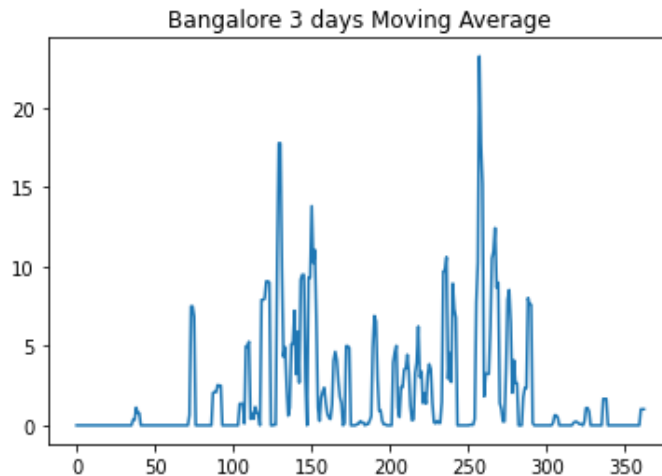
The above graphs show the moving averages of the three cities that can be observed by taking consecutive three days into account.

There are not many similar patterns between three different cities as Vizag has its peak around (100-150) and (350), Bangalore has the peak during (100-150) and (250-300).

Whereas Hubli faced the highest around (100-200) also during (250-300).

17) Moving average considering three consecutive days for cities in three years.

□ Bangalore



An IMD reliable reports that The deep despair over the southeast Bay of Bengal has moved in a west-Northwest course and intensified right into a cyclone Asani. At 5:30 hours on May 8. It will pass withinside the Northwest course and might accentuate right into an excessive cyclone. It will keep to transport withinside the Northwest course till the tenth May night and attain West Central and adjoin the northwest Bay of Bengal.

Extremum- [257, 129, 130]

Over the North Central Bay of Bengal, a well-marked low-pressure area is likely to intensify into a Depression. In fact, it would cause severe depression. It is more likely to move throughout the

west path towards Central India. Furthermore, a cyclonic circulation may be seen across the northeast Arabian Sea, with the trough extending as far as Coastal Karnataka.

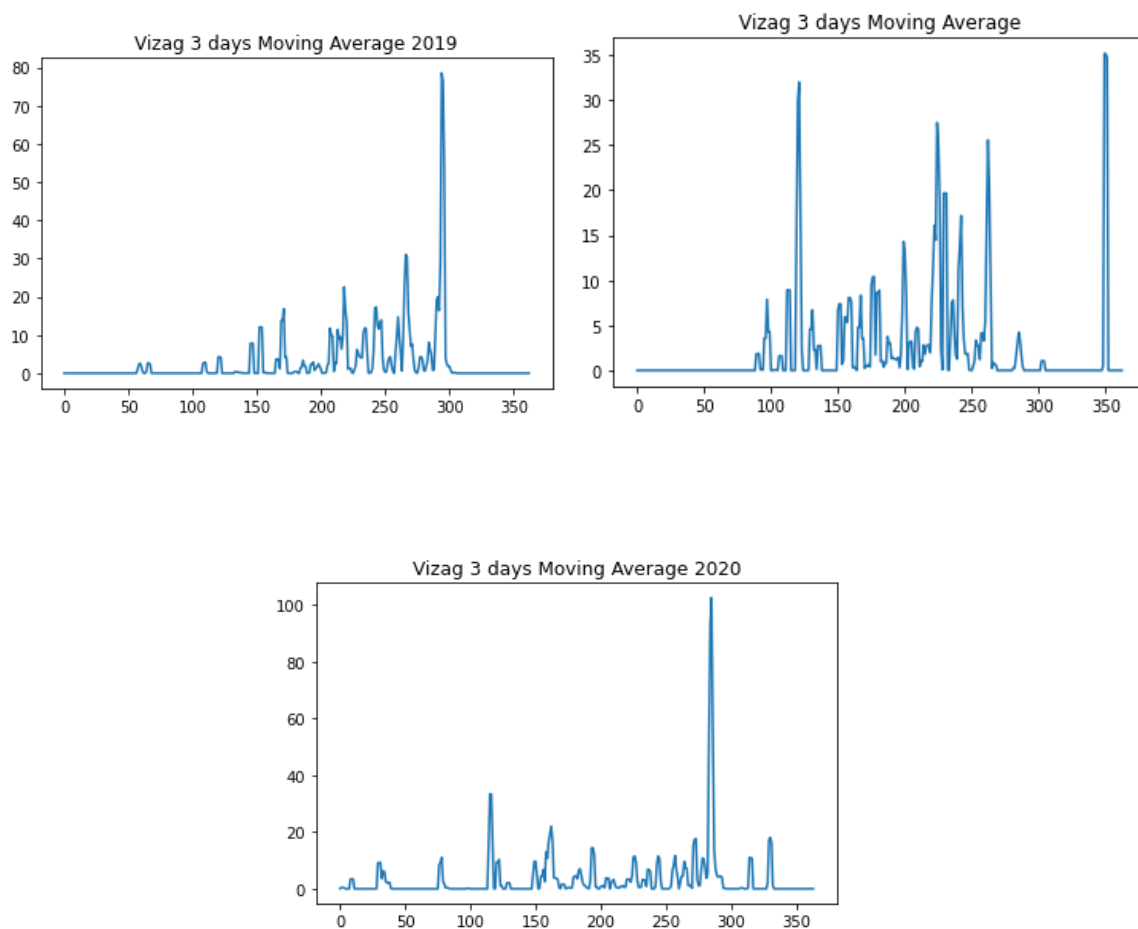
Extremum -[265, 266, 264]

These climate structures might beautify the rain sports over Karnataka and we assume some excellent spells of rain over many components of the country till September 21. Weathermen are predicting light to heavy rain and thundershower with sturdy winds over Bagalkot, Bangalore, etc

The Indian Meteorological Department reported on Monday that a deep sorrow forming over the northeast and nearby east-important Arabian sea traveled practically westwards at a speed of 17 kmph and strengthened into Cyclonic Storm 'Hikka.'

Extremum- [292, 294, 118]

□ Vizag



The above graph indicates the rainfall sample in Vizag while taken 3 consecutive days into attention.

In the 12 months 2018, the rainfall slowly rises across the one hundred and twentieth day this is April. A comparable sample is observed till days someplace among 2 hundred-250th days this is round August however the rainfall isn't regular till day 348. Here there's non-stop rainfall for more than 3 consecutive days. On similar studies it's far discovered to be induced because of a cyclonic hurricane named Titli, equal to a class IV and V hurricane. This is maximum in all likelihood because of depressions in Bay of Bengal induced throughout the transition from southwest monsoon season to northeast monsoon season as a prediction with the aid of Andhra University. We can finish with the aid of thinking about the extremum to be someplace round from the 348th day to 352th day.

Extremum- [349, 350, 351]

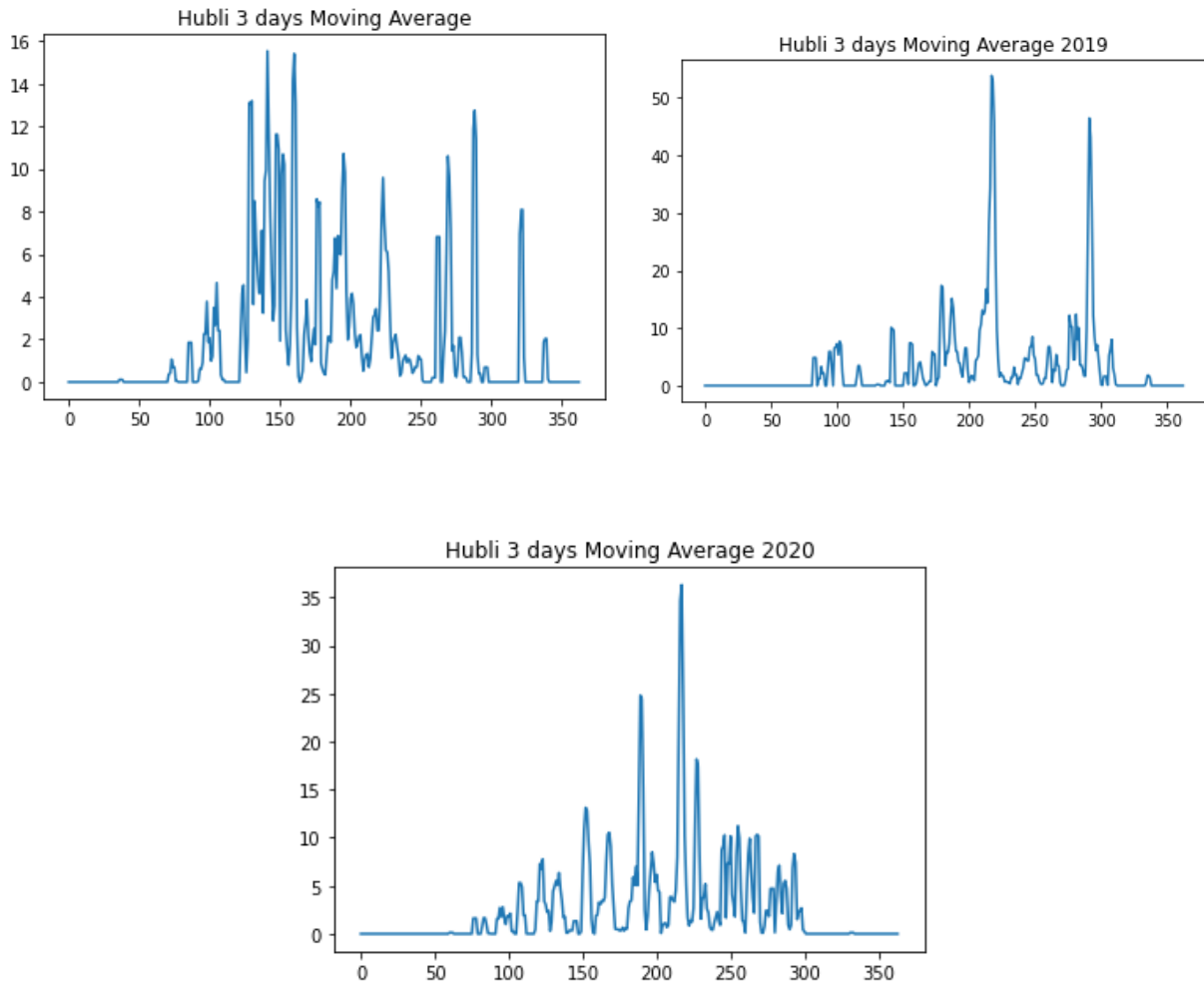
In the 12 months 2019, the rainfall has minute ups and downs till the times round 2 hundred wherein there's a mild upward push and the upward push is going on growing slowly. During the length from 293rd day there's a rainfall for consecutive days, throughout the 299th day additionally there's non-stop rainfall consequently this era may be taken into consideration as extremum. Understanding the motive for this non-stop rainfall, the sample is so due to a cyclone named Gulab which has shaped over the Bay of Bengal which had crossed Andhra Pradesh coast and weakened right into a deep despair. Vizag changed into near proximity to the hurricane, and had obtained extreme rainfall sports for twenty-four hours to the tune of 267mm. This changed into the best rainfall in 24 hours ever obtained with the aid of using Vizag throughout the month of September.

Extremum- [294, 295, 296]

In the 12 months 2020, the rainfall has minute ups and downs till the times round 2 hundred wherein there's a mild upward push and the upward push is going on growing slowly. The steep upward push is discovered throughout the 282th day to 286th day approximately. This length is in October starting (October 8th). The studies states that this rainfall is in all likelihood due to the fact the despair over West-Central Bay of Bengal intensified similarly right into a deep despair which induced heavy rainfall withinside the location of vizag and different components of the state. thirteenth October 2020, as a Deep Depression with the most sustained wind pace of 55-sixty five kmph gusting to seventy five kmph.

Extremum-[285, 284, 283]

□ Hubli



The rainfall pattern in 2018 is quite varied, with numerous peaks throughout the year. The rainfall is minimal for the first 50 days, but after the 100th day, the rainfall pattern begins to change. There is a peak around the 139th day, which is approximately May 20th. Rain fell for almost four days, possibly as a result of the cyclone that slammed the coastal areas. The Department of Meteorology has provided no specific proof.

Extremum- [141, 160, 159]

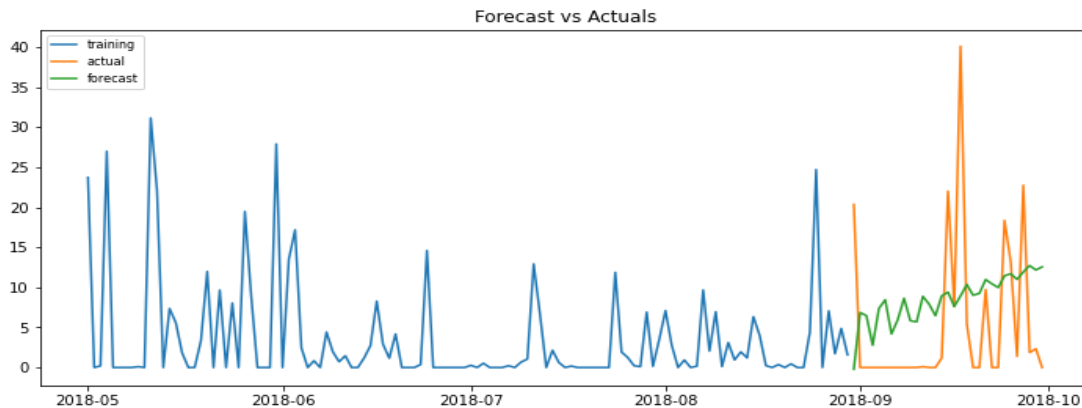
In the year 2019, the rainfall in Hubli is comparatively low while at the beginning of the year there are minimal peaks. Around the 170th day there are little peaks found and they gradually decrease. On the 213th day there is a peak and continuous rainfall during this period. At this period Hubli received the highest rainfall in 118 years for this exact period and it was predicted due to cyclonic waves.

Extremum-[217, 218, 291]

In the year 2020, it is observed that there is limited rainfall in the beginning of the year, with some peaks as the year progresses. There is a noticeable increase around the 150th day i.e, May ending to June starting. There is also a peak around 180-190th day that is in June. The major peak can be observed around 282nd day there is continuous rainfall in this period. This is predicted to be because of the southwest monsoon active conditions over west coast of karnataka. The flow has strengthened over the Arabian sea at both surface and lower troposphere. It was observed that there were strong squally winds with wind speed 50-60kmph prevailing over east central Arabian sea August 4th and July 7th is said to have experienced the highest rainfall in the city.

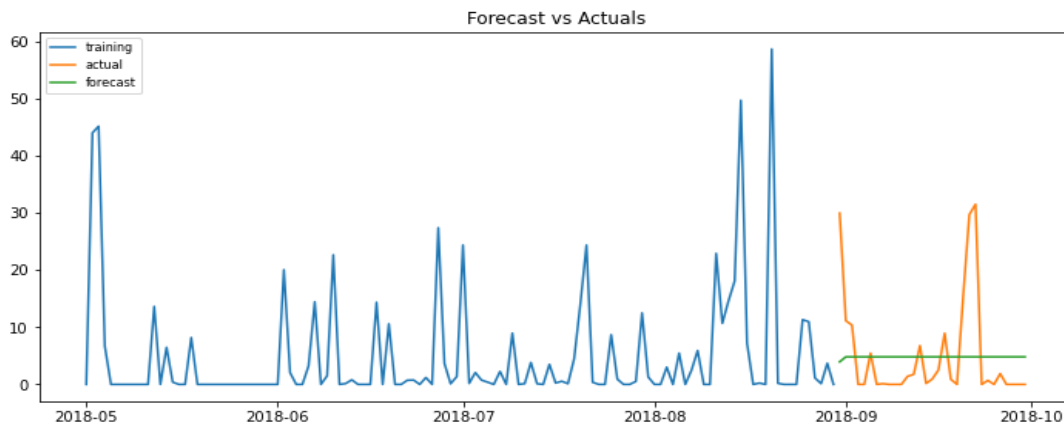
Extremum-[285, 284, 283]

Bangalore:



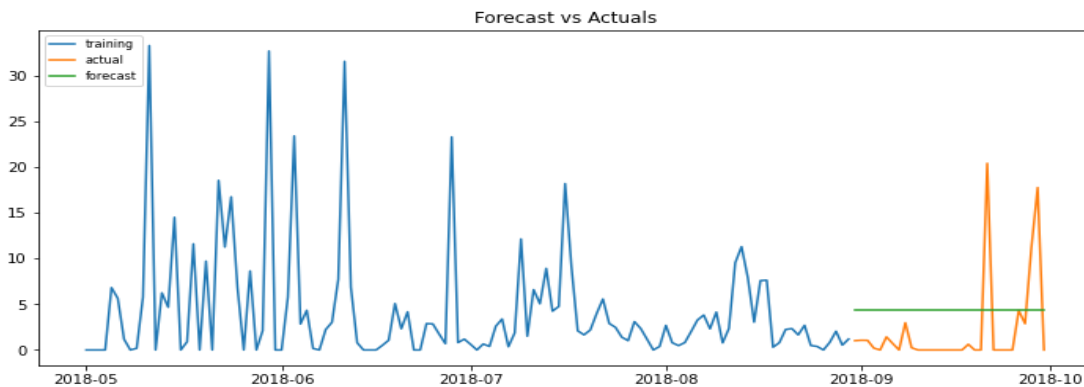
The above graph represents the rainfall data of the South west monsoon season of the Bangalore city in the year 2018. The blue line represents the training data considered which is of size 122 data points , the orange line represents the testing data considered with a size of 31 data points and the green line represents the forecasted data for the testing data considered.

Vizag:



The above graph represents the rainfall data of the South west monsoon season of the Vizag city in the year 2018. The blue line represents the training data considered which is of size 122 data points , the orange line represents the testing data considered with a size of 31 data points and the green line represents the forecasted data for the testing data considered.

Hubli:



The above graph represents the rainfall data of the South west monsoon season of the Hubli city in the year 2018. The blue line represents the training data considered which is of size 122 data points , the orange line represents the testing data considered with a size of 31 data points and the green line represents the forecasted data for the testing data considered.

6. Summary and Conclusion :

The rainfall forecasting is essential in urban planning, however, it is difficult and complicated to model because of its extremely irregular and turbulent behavior. While the smart models are used to predict the rainfall, the results are unreasonable. The main reason for this is that the hydrological series is a combination of stochastic and algebraic processes; whereas the smart models are algebraic and cannot predict the stochastic events of the hydrological series.

From the implementation of this project, we have been able to address the problem of lack of Accurate methods of weather forecasting by various Authorities. Here we have implemented the Rainfall prediction using ARIMA modeling that is used for forecasting the rainfall of Vizag , Hubli and Bangalore . The results show that the modeling accuracy increases with normalization and standardization of data.

While building and developing this complete project, we have had the privilege to get exposed to a wide range of tools and technologies. As we started off nothing but the idea of just doing the rainfall prediction. And later on we searched for the data set , learnt how to process a netcdf file to a csv file and retrieve data from it. And later we brought in python tools to manipulate the csv file and started inferring about the results which ultimately guided us in idea of calculating mean averages, variances and moving averages not only this but in the path of segregating data in the way of north east and south west and brought in brief idea of how our weather pattern works in the real world. This just does not stop us from doing it for one single year but we did try for three consecutive years and were able to compare the data with respect to each other. But we are not just convinced that the small parameter which we took into consideration with a data set of three years made us understand the complete weather and rainfall prediction pattern but we are in a notion of extending this project in a bigger scope if time would have permitted us.

This project has also helped us think and brainstorm extreme scenarios and come out with feasible solutions. Overall, we have gained a good experience and vast knowledge and the ability to work and collaborate in a team . We learned that the rainfall phenomenon is affected by various parameters such

as precipitation, temperature, humidity, evaporation etc. In this project, rainfall is used as the only prediction parameter. As a future research, other parameters can be used for rainfall forecasting.

References

- S. Zhang, L. Lu, J. Yu, and H. Zhou, “Short-term water level prediction using different artificial intelligent models,” in 2016 5th International Conference on AgroGeoinformatics, Agro-Geoinformatics 2016, 2016
- S. Zainudin, D. S. Jasim, and A. A. Bakar, “Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1148–1153, 2016
- R. Venkata Ramana, B. Krishna, S. R. Kumar, and N. G. Pandey, “Monthly Rainfall Prediction Using Wavelet Neural Network Analysis,” *Water Resour. Manag.*, vol. 27, no. 10, pp. 3697–3711, 2013.
- D. Nayak, A. Mahapatra, and P. Mishra, “A Survey on Rainfall Prediction using Artificial Neural Network,” *Int. J. Comput.*, vol. 72, no. 16, pp. 32–40, 2013.
- S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, “An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives,” *Expert Syst. Appl.*, vol. 85, pp. 169–181, 2017.
- Aftab, Shabib & Ahmad, Munir & Hameed, Noureen & Bashir, Salman & Ali, Iftikhar & Nawaz, Zahid. (2018). Rainfall Prediction using Data Mining Techniques: A Systematic Literature Review. *International Journal of Advanced Computer Science and Applications*. 9. 10.14569/IJACSA.2018.090518.
- Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water (Switzerland)*, 10(11). <https://doi.org/10.3390/w10111536> .

- Janani, B; Sebastian, P. (2014). Analysis on the weather forecasting and techniques. International Journal of Advanced Research in Computer Engineering & Technology, 3(1), 59–61. <http://ijarcet.org/wp-content/uploads/IJARCET-VOL3-ISSUE-1-59-61.pdf>
- Chaudhari, M. S., & Choudhari, N. K. (2017). Open Access Study of Various Rainfall Estimation & Prediction Techniques Using Data Mining. American Journal of Engineering Research (AJER), 7, 137– 139. [http://www.ajer.org/papers/v6\(07\)/Q0607137139.pdf](http://www.ajer.org/papers/v6(07)/Q0607137139.pdf)
- Aakash Parmar, Kinjal Mistree, M. S. (2017). Machine Learning Techniques for rainfall prediction: A Review. International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS). https://www.researchgate.net/profile/Aakash_Parmar/publication/319503839_Machine_Learning_Techniques_For_Rainfall_Prediction_A_Review/links/59afb922458515150e4cc2e4/MachineLearningTechniques-For-Rainfall-Prediction-AReview.pdf
- <http://www.ksndmc.org/Default.aspx>
- Global Warming- Wikipedia, the free encyclopedia

9. Acknowledgement

We would like to thank the Indian Institute of Information Technology for providing us the opportunity to put forth our idea and was there as a constant support. It is indeed a great pleasure to express our sincere gratitude to our supervisor Dr. Uma Seshadri, Department of Computer Science and Engineering, IIIT Dharwad for her continuous support in this project. To start with she was very supportive during the project topic selection. She was always there to listen, advise and share her expertise with us at every stage of the project development. She showed us different ways to approach a real world problem and the need to be persistent to accomplish any goal. She was always there to meet and talk about our ideas, share her expertise and views on developing a solid prototype, and to ask us good questions to help us think through our problems. Without her encouragement and constant guidance, we could not have completed this project on time. We are also indebted to our own team members for their rigorous efforts in questioning the most difficult edge cases and extracting the best out of it and also for their persistent coordination till the end. We would also like to thank Dr. Narasimha, Visiting faculty IIT Dharwad, retired scientist TIFR for giving us good guidance throughout the project. He divided the project into steps and because of this approach we were able to carry our project forward smoothly. We were able to understand everything about rainfall with his help. We also would like to thank Dr. R Krishna, Director, IITM(Indian Institute of Metrology) for providing us with real time data having rainfall precipitation. The data was of great help and it helped us throughout the project.

Meghana N

G K Bharath Bhushan

B Ragavan

K V S Chaitanya

Date: 30/04/2022