MATHS REPORT

Exploratory Data Analysis on Indian Rainfall Data using R

DEPARTMENT OF COMPUTER SCIENCE

MSC COMPUTER SCIENCE SEM - 1

Submitted by:
Vallabh Tiwari
Sakshi
Srishty Sonal

CONTENTS

S.NO.	TITLE	Page
1.	Motivation	3
2.	About the Dataset	4
3.	Introduction	5
4.	Exploratory Data Analysis and Data Visualization	7
5.	Conclusion	28
6.	References	29

MOTIVATION

India's monsoon is considered as one of the most prominent monsoon system and its rainfall pattern may help us in visualizing various patterns of information such as agricultural produces /seasonal cropping, seasonal rainfall, intensity of rainfall in various regions, local precipitation expectation in various subdivisions ,occurrence of flood and drought and many more . We can derive such various inferences from the monsoon dataset which could be further used to extract hidden patterns or predict the future outcomes related to the occurrences of various events associated with it.

ABOUT THE DATASET

Dataset	Sub_Division_IMD_2017.csv
Size of Dataset	Number of rows: 4188 Number of columns: 19

Attribute	Description
Subdivision	Various subdivisions in India
Year	Year of rainfall(1901 to 2017)
Months(JAN to DEC)	Rainfall in 12 months for each year in a subdivision
Annual	Annual rainfall in each year for a subdivision
JF	Rainfall in January February
MAM	Rainfall in March, April and May
JJAS	Rainfall in June, July, August and September
OND	Rainfall in October, November and December

INTRODUCTION

Indian Monsoon is one of the most prominent monsoon systems around the world. It is depicted from the variation and amount of India's annual rainfall. Its effects are felt by India, the Indian subcontinent and the neighboring water bodies of the Arabian Sea, the Bay of Bengal and the Indian Ocean. During the cold months, the direction of the monsoon is from the north-east, known as the north east monsoon in India, while during the time of warmest months of the year the monsoon wind blows from the southwest hence known as the south west monsoon in India.

The Indian monsoon consists of four significant seasons, which are:

1. Winter Monsoon (**JF**)
2. Summer Monsoon (**MAM**)
3. South-West Monsoon (**JJAS**)
4. The North-East Monsoon (**OND**)

Out of the four seasons, most of the Indian rainfall is received by **SW Monsoon** and the **Winter Monsoon** (also known as retreating monsoon).
Based on the trends and patterns in Indian rainfall the Indian subcontinent is divided into 36 subdivisions.

The dataset used in the study has been taken from the Open Government Data Portal of India. It contains the rainfall data of India for every subdivision from the year 1901 to 2017. This study

is primarily focused on the study of rainfall trends of the last 50 years.
The data has been explored using various tools of language R.
The two main libraries used in the project are:

1. **readr** - To load the dataset and perform various operations on it.
2. **ggplot** - It is a very useful library to draw interactive plots. Interactive plots are more interesting and self explanatory than static plots. All the plots have been made using it.

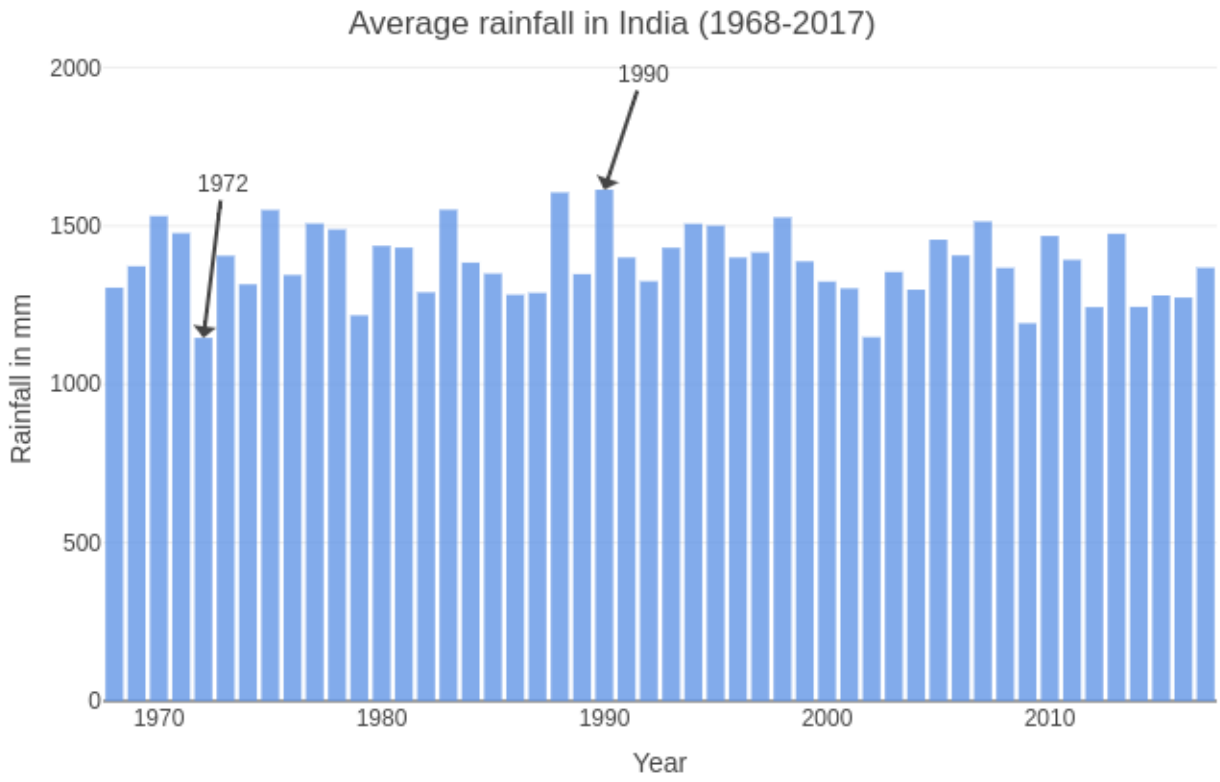
This report consists of various observations made by us in studying the dataset. Various questions can be formulated from the dataset. We have written some of them and answered them after observing the dataset through plots.

OBSERVATIONS

The observations made during study have been summarized as answers to questions as follows:

1. What are the trends in Indian rainfall for the past 50 years? Which years received the maximum and minimum rainfall respectively in the past 50 years?

We plotted a bar plot of year vs average rainfall to answer these questions.



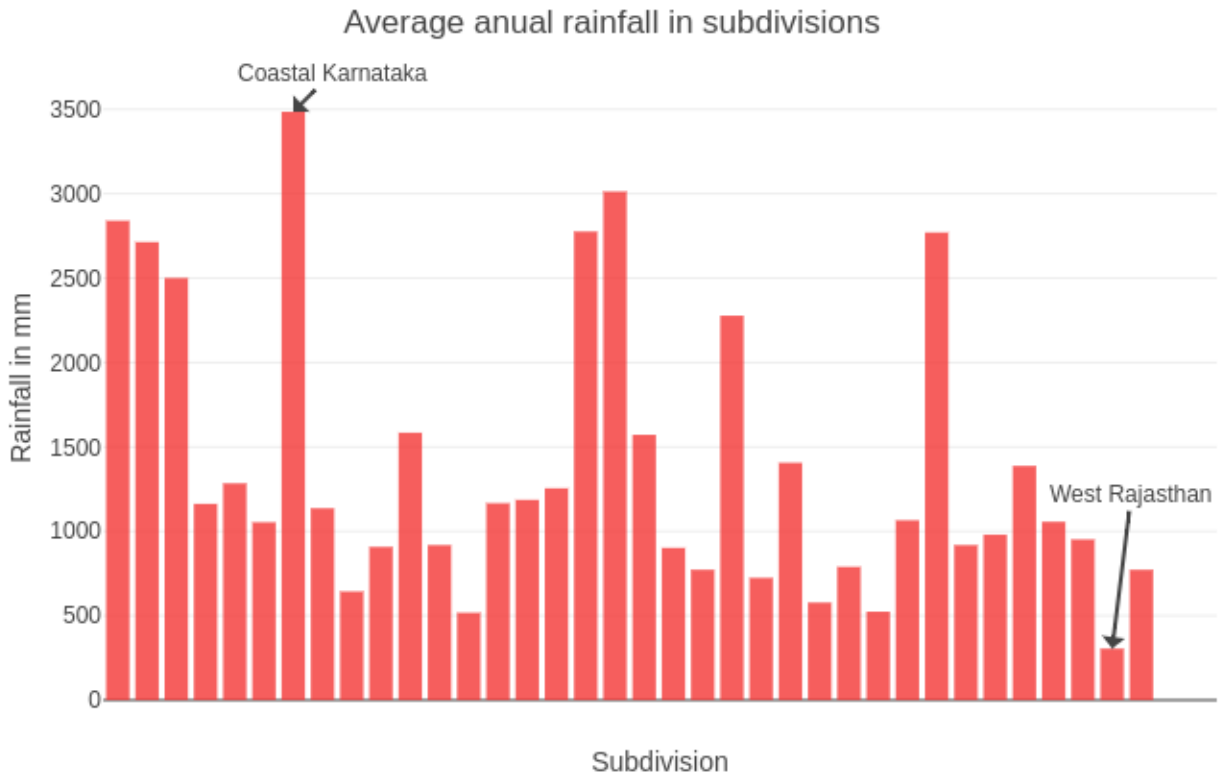
The average rainfall experienced in the Indian subcontinent for the past 50 years is 1385.07mm. The year **1990** shows the

maximum average rainfall of 1614.29 mm. **Minimum** rainfall was experienced in the year **1971** of 1146.49mm.

Also, before 1990 the average rainfall received in India was 1392.31mm which reduced to 1370.67mm after reaching the peak. This could be a result of increasing global warming and climate changes and factors like rapid industrialization, urbanization and deforestation. This would be worse in the present as our dataset has information up to 2017 only.

2. What are the rainfall trends in various subdivisions of India? Which subdivisions received the maximum and minimum rainfall in the past 50 years?

A bar plot of subdivision vs average rainfall can be studied to answer these questions.



Most of the rainfall received in India is due to SW Monsoon winds that originate from the Southern Hemisphere. They blow towards the NE direction and collect moisture from the Arabian sea and the Bay of Bengal.

The winds from the Arabian sea are stopped by the Western Ghats, a chain of mountains running parallel to India's Western Coast, approximately 30-50 km inland, the Ghats traverse the States of Kerala, Tamil Nadu, Karnataka, Goa, Maharashtra and Gujarat. Due to

this subdivisions like **Coastal Karnataka, Konkan and Goa, Kerala and Laksdweep** receive heavy rainfall. These subdivisions lie in the windward side(western) of the Western Ghats which is the main cause of heavy rainfall here.

Similarly the winds from the Bay of Bengal are stopped by Garo-Khasi-Jaintia hills in Assam and Meghalaya giving heavy rainfall to subdivisions of **Arunachal Pradesh, Assam and Meghalaya, Himalayan WB and Sikkim and Nagaland, Manipur, Mizoram and Tripura**. Mawsynram, situated on the southern slopes of the Khasi Hills in Meghalaya, India, is one of the wettest places on Earth with an average annual rainfall of 11,872mm.

Also, Coastal Karnataka received the **maximum** rainfall of **3485.58mm** while West Rajasthan received the **minimum** of **358.58mm**. The difference in the maximum and minimum indicates the spatial variation in rainfall distribution in various regions of India.
The different proportion of rainfall received in different divisions of India makes it unique in its own form and, thus making it interesting to look and examine our country's pattern of rainfall.
Subdivisions like **Marathwada, Vidarbha,Coastal Andhra Pradesh, Madhya Maharashtra** etc which lie on the leeward side(eastern) of the Western Ghats experience much less rainfall as compared to the windward side despite being close to the southern coast. **East Interior Karnataka** also experiences very less rainfall compared to Coastal Karnataka due to this only. Isn't is interesting that there is so much variation in the rainfall in the same state of Karnataka. All because of the Western Ghats.

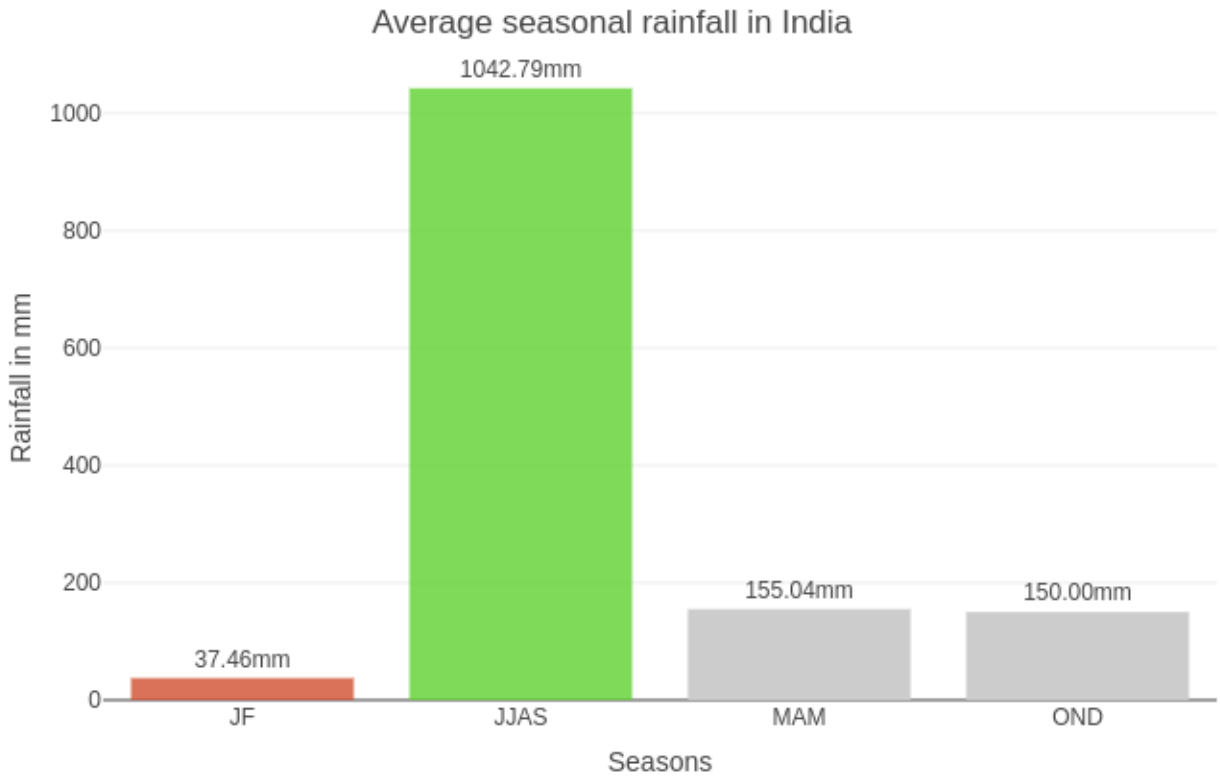
East Rajasthan and West Rajasthan experience less rainfall. Also rainfall in West Rajasthan is significantly less than that of East Rajasthan. This is due to two reasons. Firstly, the hot climate of the

Thar Desert due to which the SW monsoon winds do not condense. Secondly, lack of structure like plateaus or mountains. The Aravalli ranges running from Haryana to Rajasthan and Gujrat are not tall enough to act as a barrier for the monsoon winds. Also they are parallel to the path of monsoon winds causing no orographic rainfall.

A very peculiar and interesting observation made in this plot is the relatively less amount of rainfall in Northern Plains of India i.e. the subdivisions of **West Uttar Pradesh, East Uttar Pradesh, Bihar and also Jharkhand, Chhattisgarh, Orissa, Himachal Pradesh and Uttarakhand.** Although, ahead of these parts lies the great, tall and wide **Himalaya** which could be a potential barrier to monsoon bearing winds, these regions do not experience very heavy rainfall. This is because the strength of the SW monsoon winds is already exhausted by the time they cross the Western Ghats. Hence, rainfall in these regions is less.

3. How is rainfall in India distributed according to different seasons? Which season brings the maximum rainfall in India? Which season brings the minimum rainfall to the Indian subcontinent? How much average rainfall is received in different seasons in India?

We plotted a bar plot of seasons vs average rainfall to answer these questions.



The Indian rainfall is divided into four seasons based on the amount of rain received in different months:

1. Winter Monsoon **(JF)**
2. Summer Monsoon (**MAM**)
3. South-West Monsoon (**JJAS**)

4. The North-East Monsoon (**OND**)

Out of the Total average rainfall (about 1385mm) more than 75% i.e about **1042.79mm** is received in the **SW Monsoon Season**, in the months of June, July, August and September. This season is also called the **Rainy Season** of India.

The Thar Desert and adjoining areas of the northern and central Indian subcontinent heat up considerably during the hot summers(March, April, May and June). This causes a low pressure area over the northern and central Indian subcontinent. To fill this void, the moisture-laden winds from the Indian Ocean rush into the subcontinent. These winds, rich in moisture, are drawn towards the Himalayas. The Western Ghats followed by the Himalayas act like a high wall, blocking the winds from passing into Central Asia, and forcing them to rise. As the clouds rise, their temperature drops, and precipitation occurs.

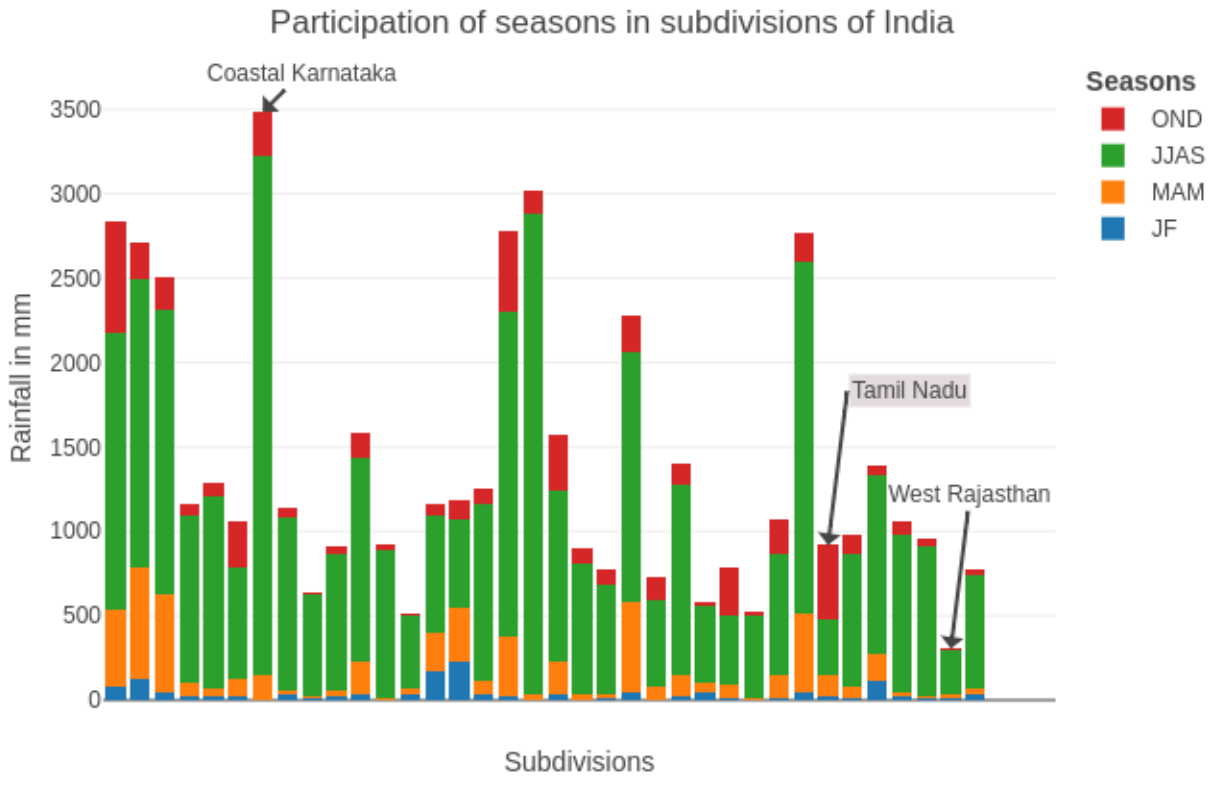
Minimum rainfall is received in the months of **January and February** of about 37mm(less than 0.03%).
The remaining 24% of the total rainfall is received in Winter (**October, November and December**) and Summer (**March, April and May**) monsoons respectively.

The monsoon of OND is also called the **North East Monsoon**. By the end of September, with the sun retreating south, the northern landmass of the Indian subcontinent begins to cool off rapidly, and air pressure begins to build over northern India. The Indian Ocean and its surrounding atmosphere still hold their heat, causing cold wind to sweep down from the Himalayas and Indo-Gangetic Plain towards the Indian Ocean and the Deccan plateaus. They pickup

moisture from the Bay of Bengal and pour it on parts of South India.

4. How much contribution are the different seasons making in the rainfall for different subdivisions of India? Are all the subdivisions experiencing the same proportion of seasonal rainfall?

We answered these questions by observing a bar plot of subdivisions vs annual average rainfall with every bar proportioned by four seasons.



This plot gave us various interesting observations.

While it has already become obvious that most of the subdivisions receive most of their rainfall by the SW Monsoon i.e. in the months of JJAS, there are some strange observations.
Coastal Karnataka that receives maximum rainfall among all the subdivisions, hardly receives any rain in the months of JF. This is because the NE Monsoon of previous year has already made the climate saturated and pressure is equalized to that of surrounding regions.

Himachal Pradesh, J&K and Uttarakhand located in the North West receive a relatively good amount of rainfall in these months. This is because of Western Disturbances, an extratropical storm originating in the Mediterranean region, which brings sudden and moderate to heavy rainfall to these parts. These rains are very beneficial for Rabi crops like Wheat in Punjab.

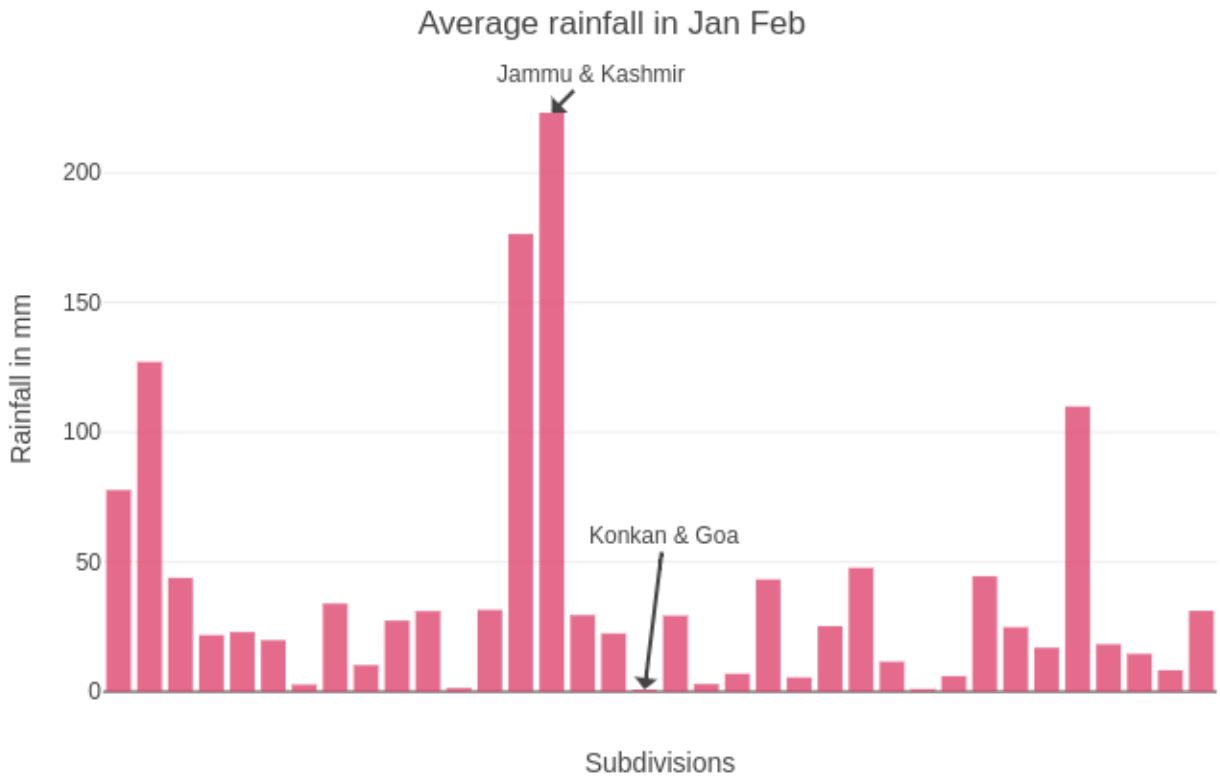
A very strange observation made was the seasonal rainfall distribution in Tamil Nadu. While most of the subdivisions received most of their rain in JJAS, Tamil Nadu received maximum rainfall in the seasons followed i.e. OND. The geographic location of Tamil Nadu and the tapering topography of Southern India are responsible for it. Situated on the Eastern Side of Western Ghats Tamil Nadu gets very less rainfall from the Arabian Sea branch of SW Monsoon. Also because of the tapering topography of Southern India, they don't come in the way of monsoon winds from the Bay of Bengal branch of SW Monsoon which blow towards NE. In the months of OND when the SW Monsoon retreats, i.e. NE Monsoon the moisture bearing winds are stopped

by the Eastern Ghats on the SE coasts of India, giving considerable amount of rainfall in Tamil Nadu. This is the reason why only Tamil Nadu received comparatively more rainfall in the months of **OND(436.63mm)** than in **JJAS(331.96)**
Other areas which receive good rain in these months are Andaman & Nicobar Islands, Arunachal Pradesh, Rayalaseema etc.

To study these trends more precisely, we plotted two more bar plots, first of subdivisions vs average rainfall in the months of January and February, second subdivisions vs average rainfall in the months ofJune July August and September.

5. What are the trends in rainfall of India in the months of January and February? Do subdivisions receiving more annual rainfall receive good rainfall in these months and vice versa?

One might expect that subdivisions which experience higher total rainfall, would be receiving more rainfall than other subdivisions in all the months. But the bar plot portrayed something quite different.

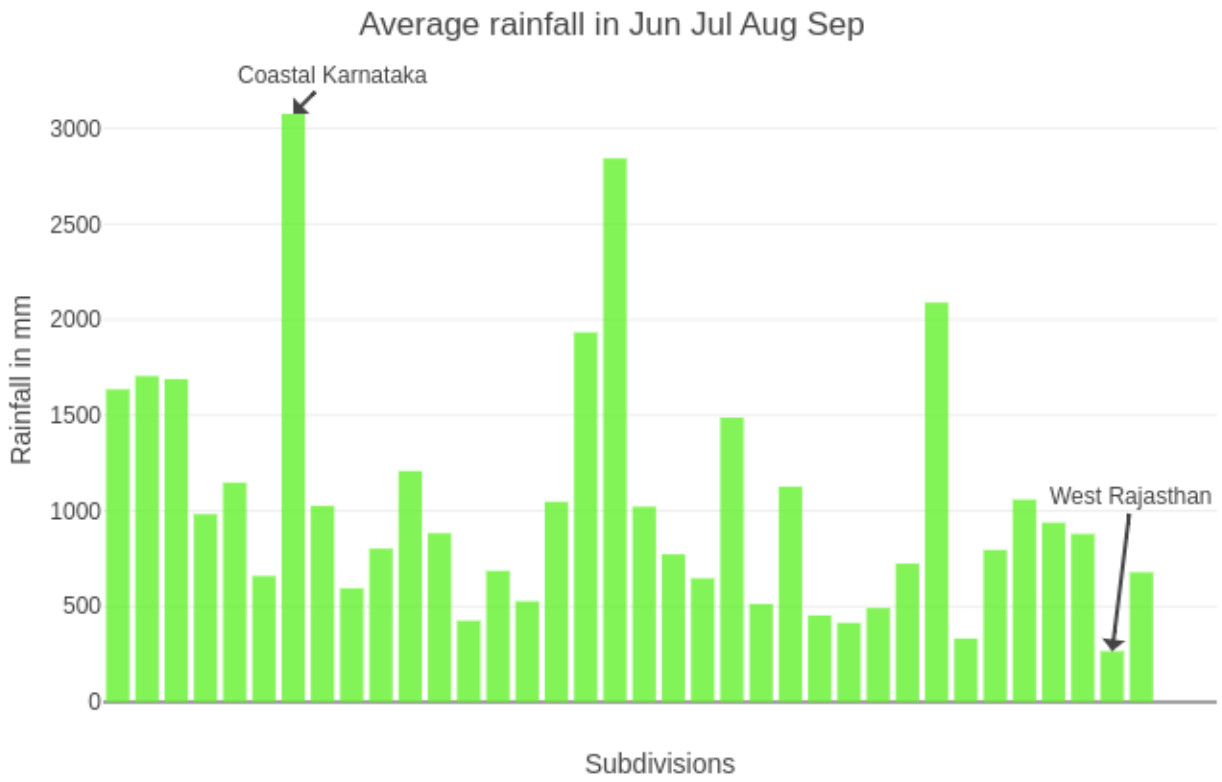


It is observed that **Jammu and Kashmir** that on average receives considerably less annual rainfall, experiences maximum rainfall in the months of January and February, higher than that of coastal subdivisions like . This is because J&K are the first place encountered by the Western Disturbances coming from the

Mediterranean Sea hence get the most rain brought by them, followed by Himachal Pradesh, Arunachal Pradesh and Uttarakhand.
Another peculiar observation is seen that, **West Rajasthan and East Rajasthan** which are presumed to receive less rainfall, actually receive more rainfall than Konkan and Goa and even Coastal Karnataka which are coastal regions with heaviest rainfall.
Thus, this gives a different pattern of rainfall that can be seen and observed in India subdivisions over the following years .
Andaman and Nicobar receive moderate rainfall in January and February. A&N receive good rainfall in all the seasons because of their location in the South Eastern edge of Bay of Bengal, the place of origin of Indian monsoons.

All the subdivisions which receive highest rainfall experience low rainfall in the months of January and February.

6. What are the trends in rainfall of India in the months of JJAS? Do subdivisions receiving more annual rainfall receive good rainfall in these months and vice versa?

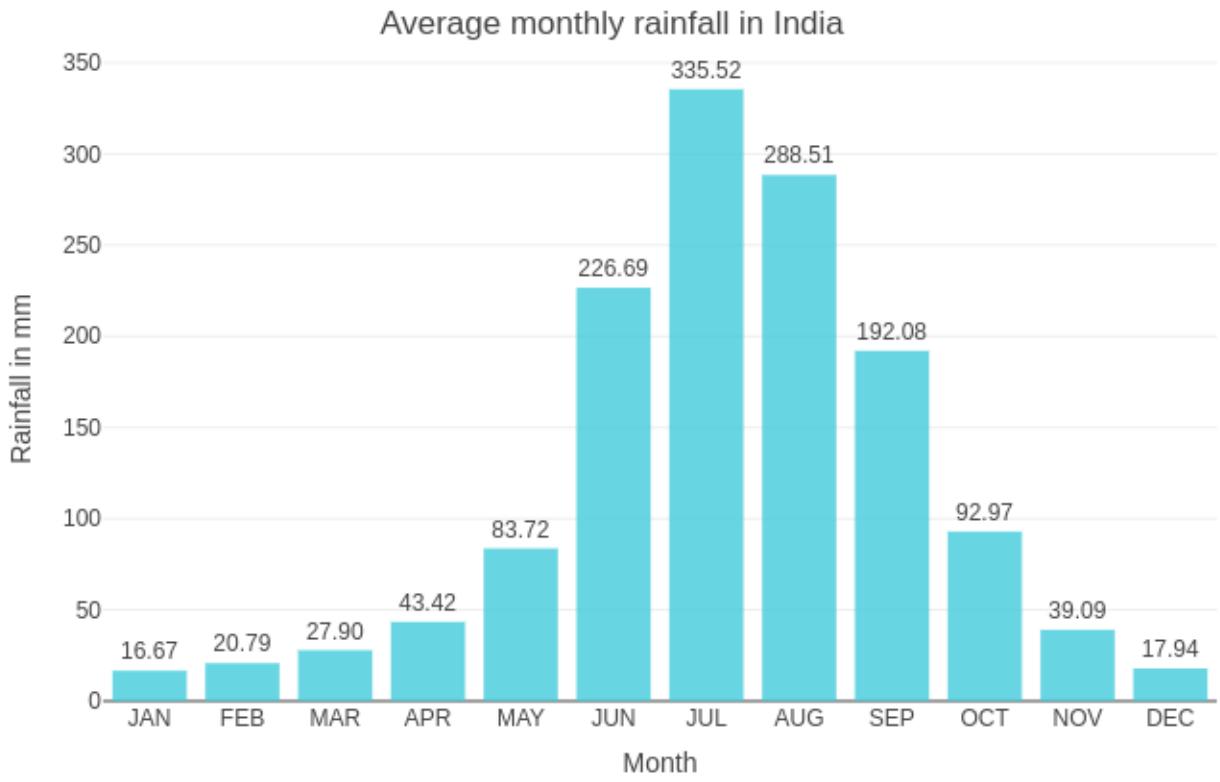


As we can see from the graph,
Coastal Karnataka receives high rainfall during the JJAS (June to September) months due to its geographical location and topography.
The Western Ghats mountain range runs parallel to the coastline of Karnataka and acts as a barrier to the moist winds coming from the Arabian Sea. These winds bring heavy rainfall to the region, especially during the monsoon season.

Additionally, the region has several rivers and streams that flow from the Western Ghats to the Arabian Sea, which also contributes to the high rainfall. The hilly terrain and lush vegetation of the region also aid in capturing and retaining moisture, leading to a higher average rainfall.

Here we can see West Rajasthan had the lowest average rainfall during the JJAS months .This is because West Rajasthan is located in the western part of India and lies in a semi-arid region that is surrounded by the Thar Desert. The region is far away from the Bay of Bengal and the Arabian Sea, which are the primary sources of moisture for the monsoon winds. As a result, the region receives very little rainfall during the monsoon season.

7. What was the minimum and maximum rainfall experienced in different months in India?



Maximum average rainfall was seen in the month of July, followed by **August and June**.

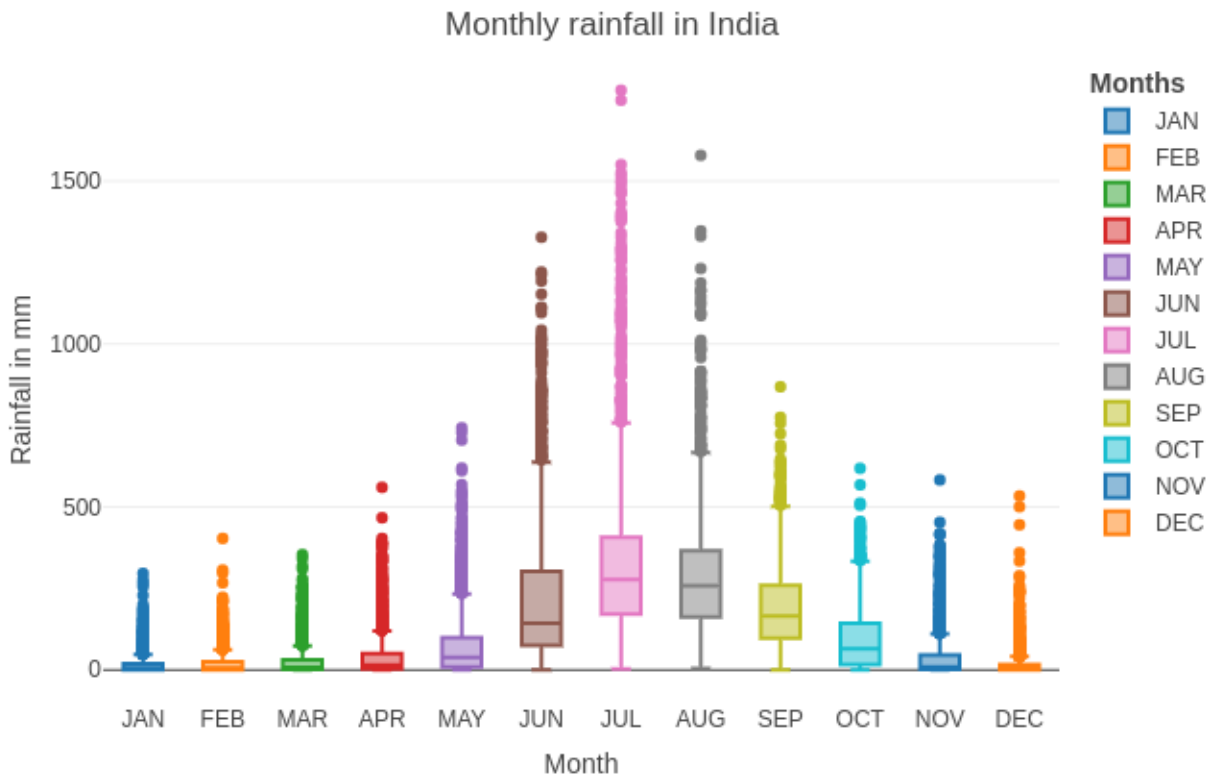
It can be seen from the graph that in **June,** the average rainfall is 226.69 mm as the monsoon season typically begins with the onset of rainfall in the southernmost part of India. This is followed by the gradual spread of the monsoon towards the northern parts of the country. The average monthly rainfall in June is relatively low because the monsoon has just begun, and the moisture content in the atmosphere is still building up.

As we can see from the graph,in **July,** the average rainfall value is maximum with the value 335.52mm, the monsoon is well-established across most of India, resulting in increased cloud formation and higher amounts of rainfall. This is also the time when the monsoon trough, a region of low-pressure in the atmosphere, moves across the country, bringing heavy rainfall to many parts of the country.

In **August,** the monsoon trough starts moving towards the Himalayas, resulting in a decrease in rainfall over the northern parts of India,hence the average rainfall value is 288.51 mm.However, many parts of the country continue to receive significant amounts of rainfall during this month.

PLOT 8:

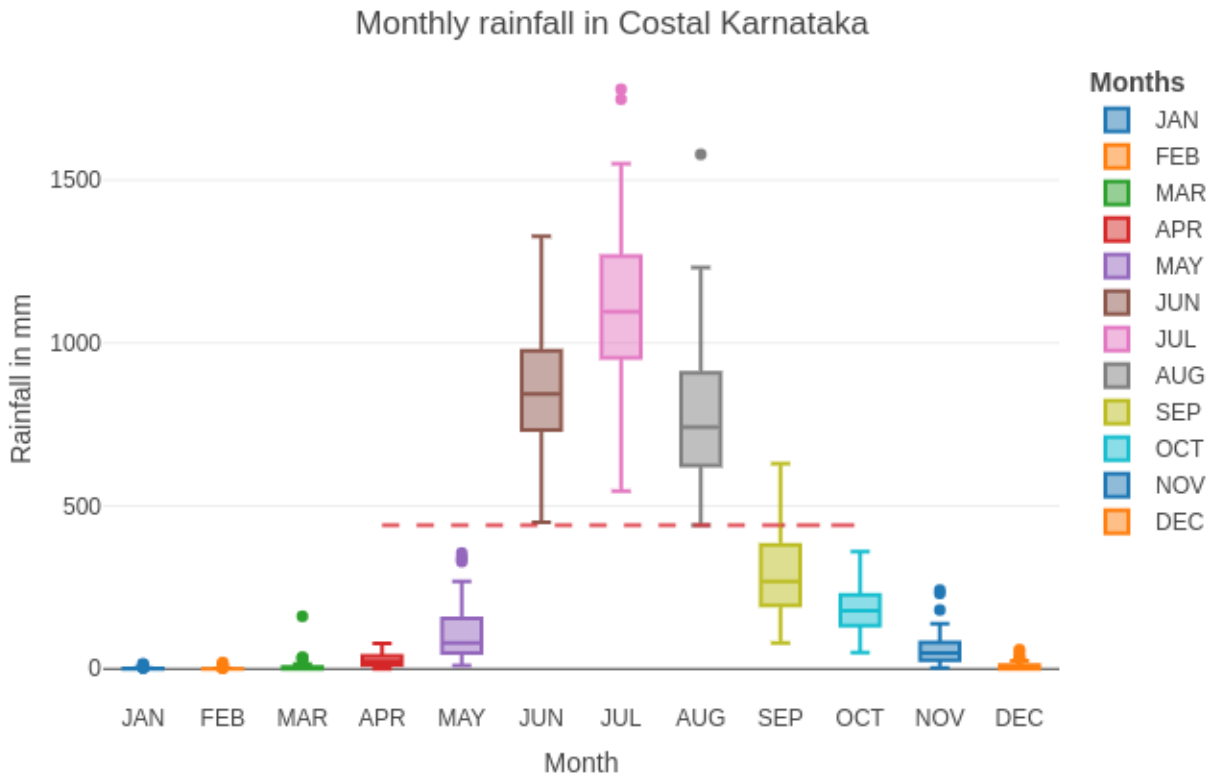
Monthly rainfall in different divisions:



Maximum rainfall was seen in July with median value 277.6 mm with maximum rainfall 778.9 mm and minimum 2.4 mm
Minimum rainfall was seen in January with maximum value 296 mm and minimum value 47.8 mm.
May's minimum rainfall was less than January's minimum value of 40.8 mm.

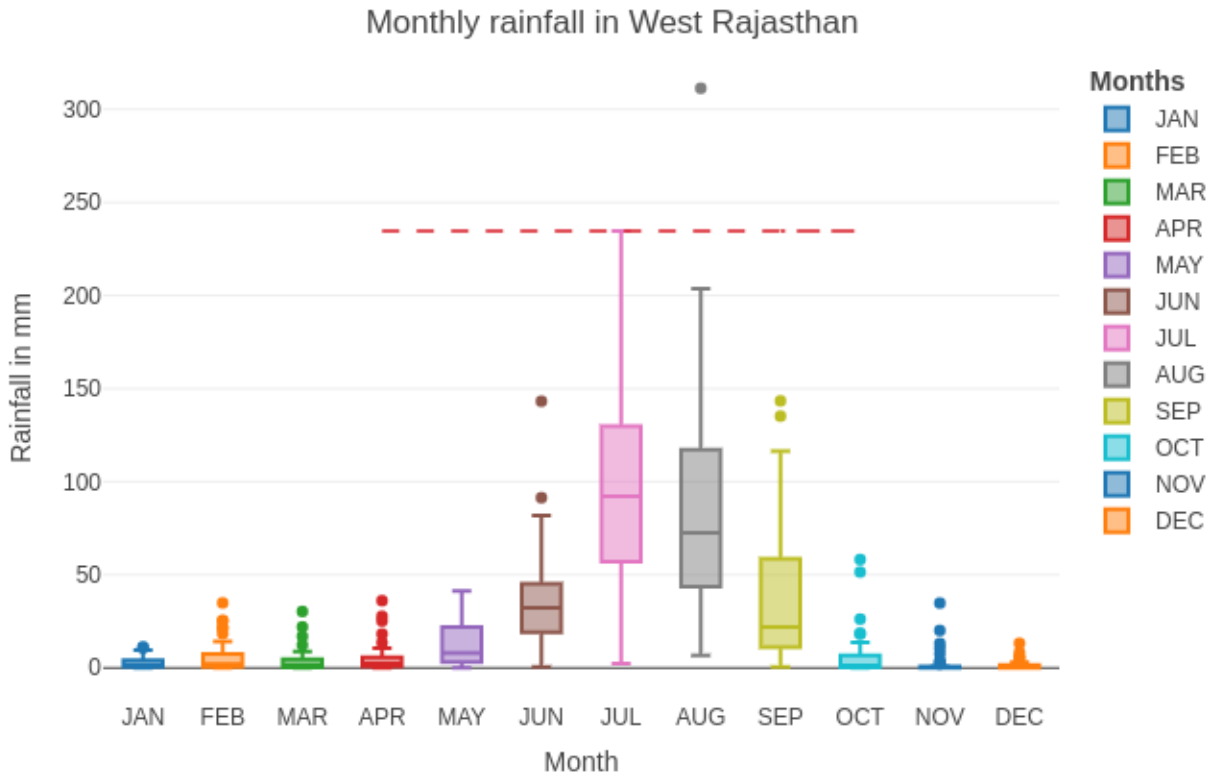
From the above observations we picked **three subdivisions of interest** to study further viz **Coastal Karnataka, West Rajasthan and Tamil Nadu**.

8. How is the rainfall distributed in Coastal Karnataka ? What is the least rainfall in the months which receive maximum rainfall?



In July,Maximum Rainfall in JJA is 1778.9 mm and minimum rainfall is 441.10 mm.This value of minimum is same for both June and August.This means at least 441.10 mm of rainfall took place between these months

9. How is the rainfall distributed in West Rajasthan ? What is the highest rainfall in the months which receive maximum rainfall?



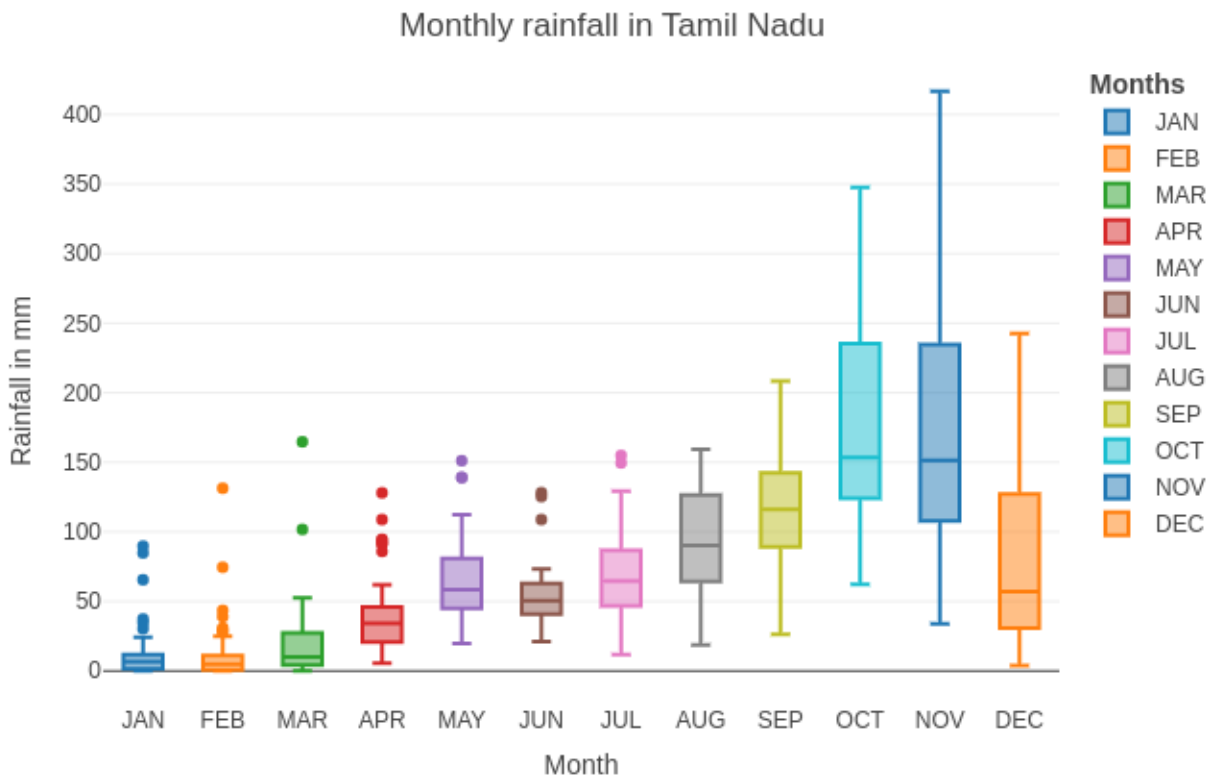
We plotted a box plot of months vs average rainfall for West Rajasthan to answer these questions.
Monthly Rainfall in West Rajasthan :
Maximum Rainfall was seen in JJA With a value 234.6 mm and in May with value 0 mm.

Another very interesting thing we showed in the box plots was that the maximum rainfall experienced in West Rajasthan in the dominant months was less than the minimum rainfall experienced in Coastal Karnataka in these months. From this one can imagine

the extent of diverse distribution of Indian rainfall. Really incredible is our India

10. Rainfall distribution in Tamil Nadu.

We plotted a box plot of months vs average rainfall for Tamil Nadu for this purpose.



Tamil Nadu was a special region of interest in our study because of its strangely different rainfall distribution. Most of the subdivisions are seen with a low rainfall in the beginning of the year which gradually increases in the months of June and July reaching its peak in mid July to August. Followed by that, again it starts decreasing in the later months. This was very clear from the

box plots and bar plots. Monthly rainfall in all the other subdivisions was **normally distributed**.
But for Tamil Nadu the monthly rainfall is **left skewed**. Maximum rainfall is experienced in the later months of **October, November and December**. This is observable in the box plot of months vs average rainfall for Tamil Nadu. The reasons for which have been discussed earlier.

Maximum value of Rainfall was seen at 416.7mm in November and minimum was seen in March 0 mm.

CONCLUSION

The dataset of rainfall can thus give a lot of information regarding the seasonal changes in pattern of rainfall that affected the agricultural produce, or may have resulted in flood or drought like situation in different subdivisions of india. This may or might help us to take precaution if situations like such occur in future. The hidden pattern while visualization has helped us in finding out various patterns within and across different sub-division of India. We can also see patterns of local precipitation in different years/months and during different seasons across different division.Thus, we have inferred such important information from our dataset by doing various exploratory visualization techniques as stated above.

References

- https://mausam.imd.gov.in
- https://www.wikipedia.org
- https://data.gov.in