IC 272: Lab1: Data visualization and statistics

Report

Name- Akanksha Sinha

Roll no.- B19125

Mobile no.- 9284606382, 9928960217 (Whatsapp)

Q1)

Mean, median, mode, minimum, maximum and standard deviation of all attributes are recorded in the following table using built in numpy functions [np.mean(), np.median(), np.std()] , scipy function [stats.mode()], min() and max() functions.

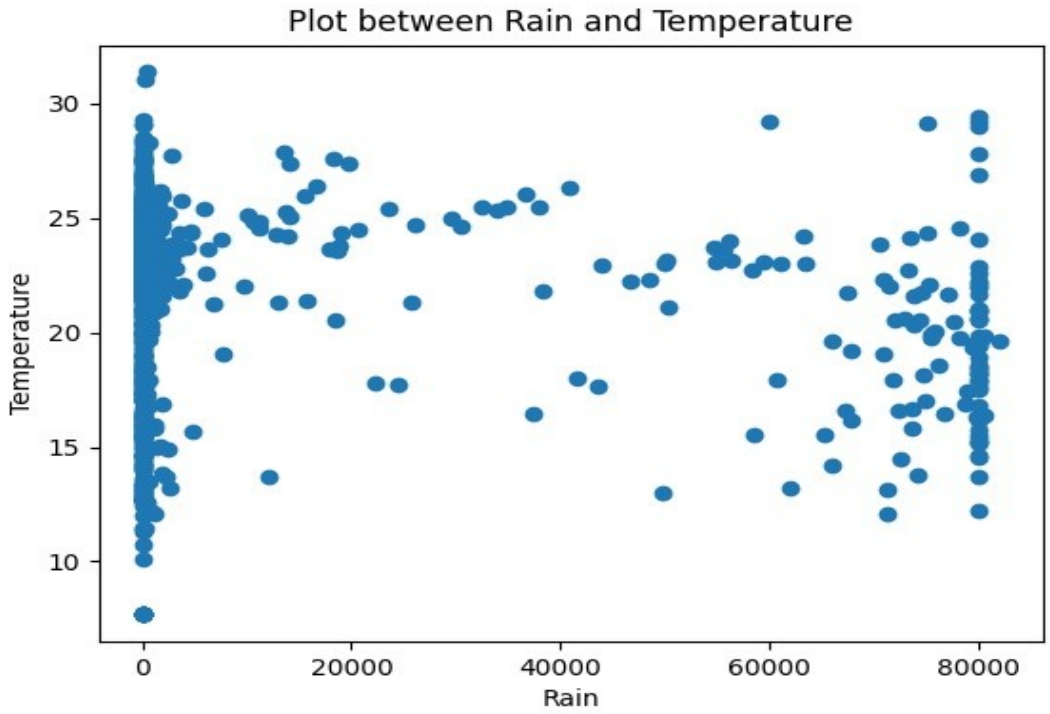
	Mean	Median	Mode	Minimum	Maximum	Standard Deviation
Temperature (in ºC)	21.214888105820148	22.27273	12.72727	7.6729	31.375	4.353512664832555
Humidity (in %)	83.47993155555568	91.38095	99.00	31.00	99.72	18.200427162850392
Pressure (in mb)	1009.0087737986486	1014.6778321678	789.39269231	452.0978873239	1079.162	46.955613275347304
Rain (in mm)	10701.53837037037	18.00	0.00	0.00	82037.25	24839.10246612766
Lightavgw/ o0 (in lux)	4438.428453333335	1656.88	4488.9103	0.00	54612.00	7569.154781086212
Lightmax (in lux)	21788.62328042328	6634.00	4000.00	2259.00	54612.00	22053.315399022737
Moisture (in %)	32.386052592592 58	16.7042	0.00	0.00	100.00	33.635433988151505

Table for observations

Q2 & Q3)

a)

1. Scatter Plot between Rain and Temperature :



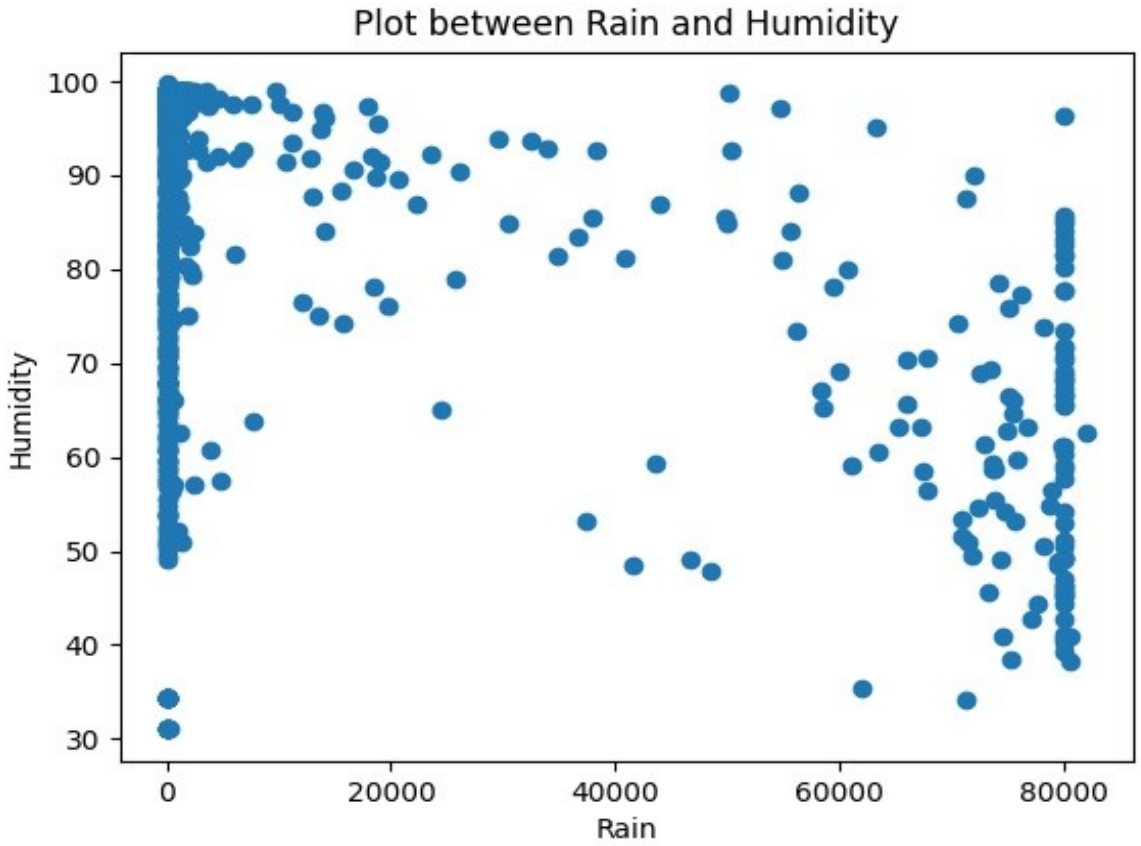
Correlation coefficient is = -0.10889328

The plot is scattered and is more dense in regions where rain is around 0mm and about 80000mm.

Negative value of correlation coefficient signifies that on increasing temperature, the value of rain decreases and vice versa.

Since the value of correlation coefficient is very low, this shows that the there is very low correlation between temperature and rain.

2. Scatter plot between Rain and Humidity:



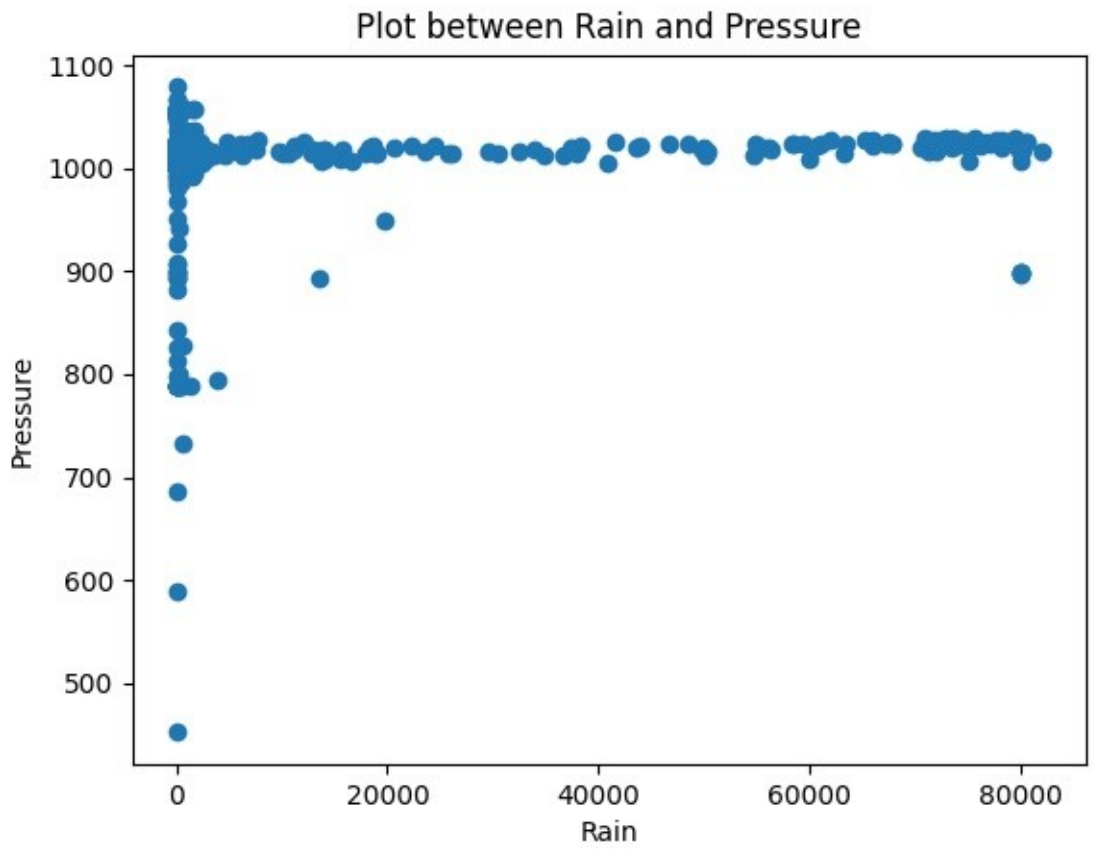
Correlation coefficient = -0.43491684

The plot is scattered and is more dense around regions with rainfall about 0mm and 80000mm. The top right corner of the graph is more dense than the bottom left which depicts negative correlation coefficient.

The negative value of correlation coefficent suggests that on increasing Humidity, amount of rainfall decreases and vice versa.

The value of correlation coefficient is moderately low which signifies there is not much correlation between the two quantities.

3. Scatter plot between Rain and Pressure:



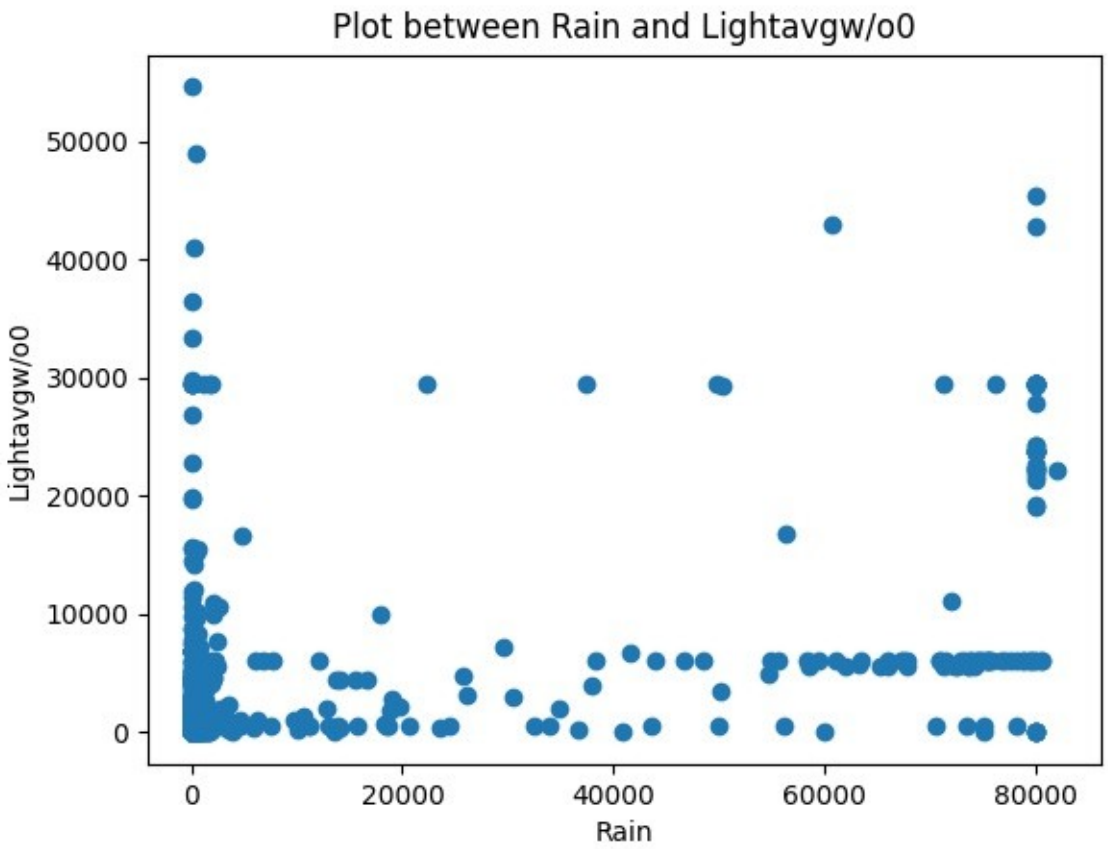
Correlation Coefficient = 0.07078467

The plot is most dense where the value of pressure is around 1000mb and also in region where rain is around 0mm.

The correlation coefficient is positive which suggests that increasing pressure also increases rain and vice-versa.

However, the magnitude is very low which shows that there is almost no correlation between the two quantities.

4. Scatter plot between Rain and Lightavgw/o0 :



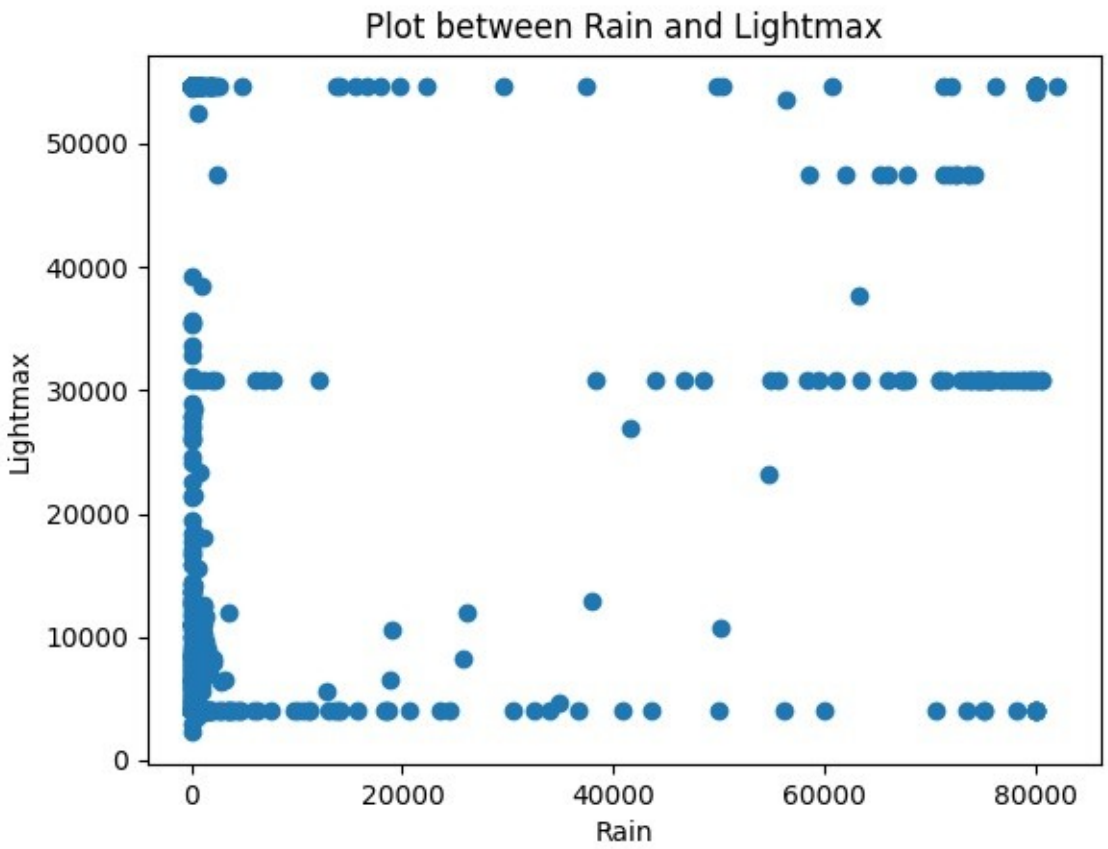
Correlation coefficient = 0.52749031

From the plot, we can see that most values are surrounded around the region with lightavg from 0-10000 lux units.

The positive value of correlation coefficient shows that lightavg increases as rain increases and vice versa.

The value of correlation coefficient is moderate which shows that both the quatities are moderately correlated.

5. Scatter plot between Rain and Lightmax:



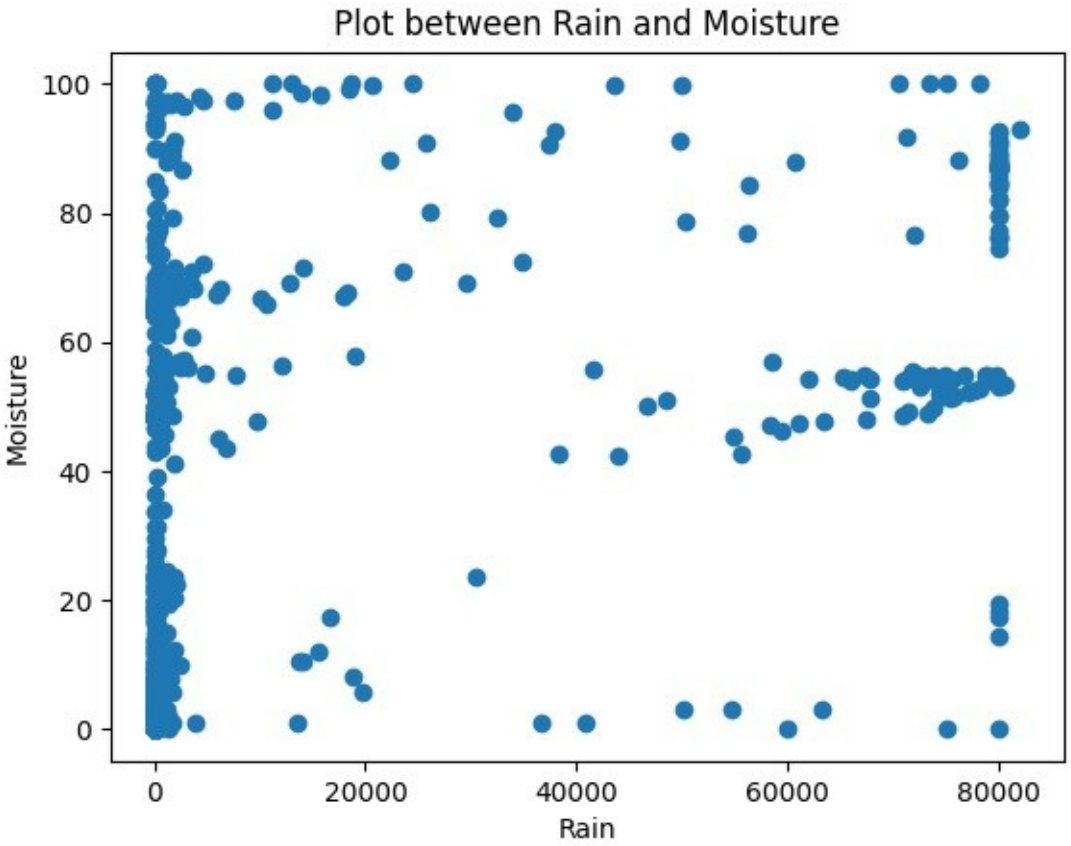
Correlation coefficient = 0.31284274

From the graph it can be observed that graph is dense around region of 0mm rainfall. Also the values where lightmax is about 2000, 30000 and 50000 lux units, the graph is comparatively concentrated.

The positive value of correlation coefficient suggests that increase in lightmax also increases rain and vice versa.

The magnitude of correlation coefficient is low hence there is very low correlation between the two quantities and the plot is scattered.

6. Scatter plot between Rain and Moisture:



Correlation coefficient = 0.42692793

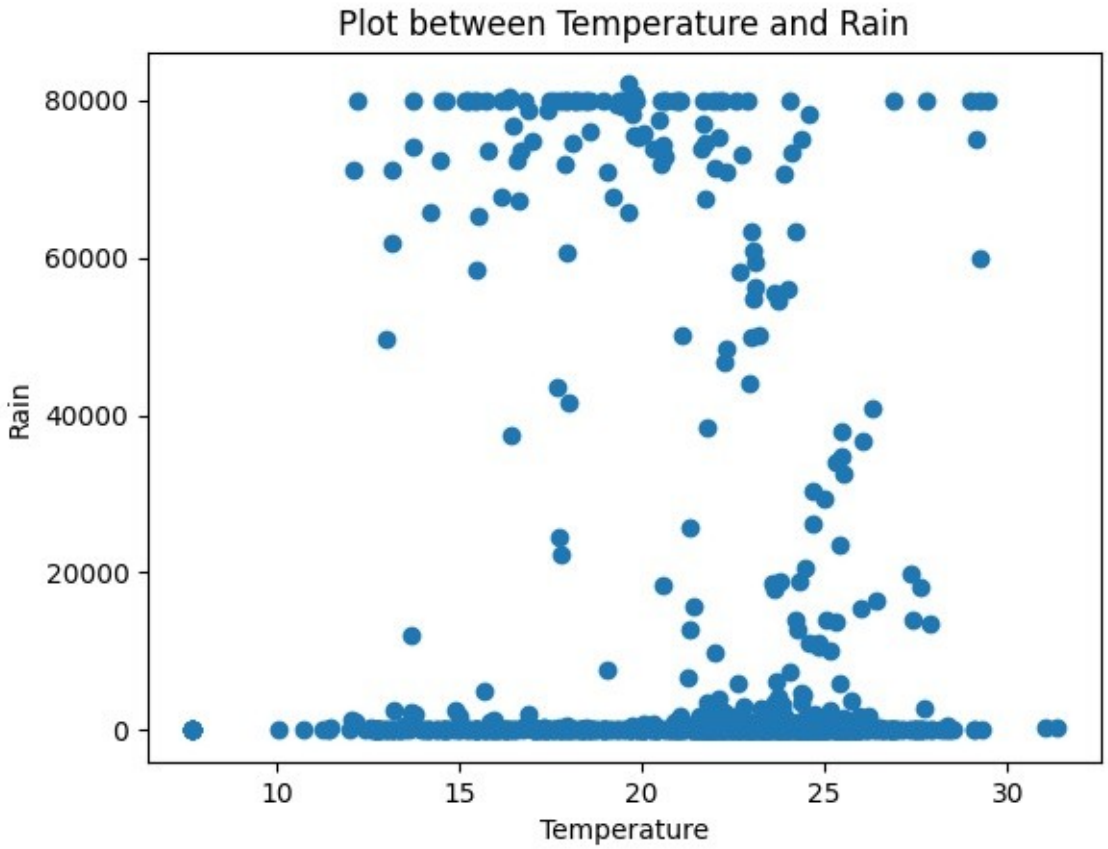
The graph is little crowded near values of rain around 0mm. Otherwise the plot looks fairly scattered.

The positive value of correlation coefficient suggests that increase in moisture also increases rain and vice versa.

The magnitude of correlation coefficient is moderately low hence there is low correlation between the two quantities and the plot is somewhat scattered.

b)

1. Scatter plot between Temperature and Rain:



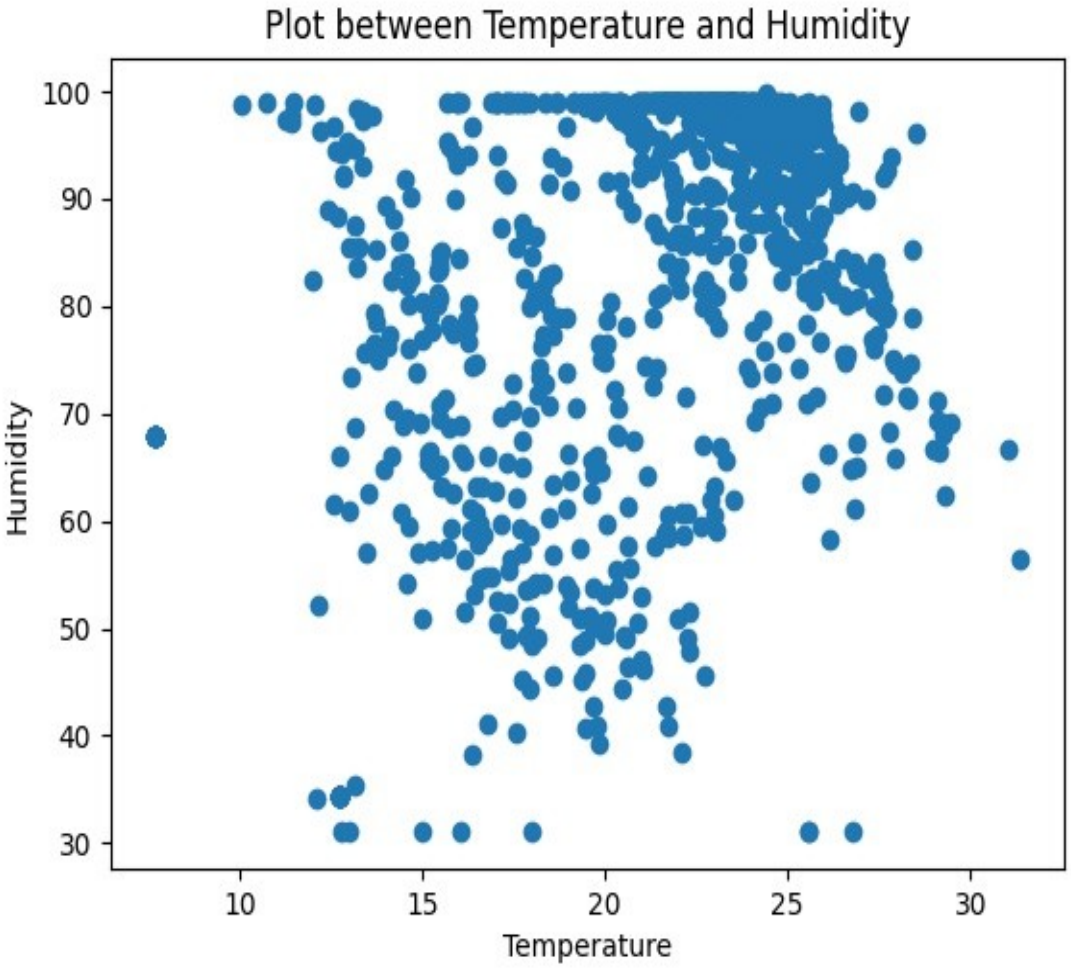
Correlation Coefficient = -0.10889328

The plot is scattered and is more dense in regions where rain is around 0mm and about 80000mm.

Negative value of correlation coefficient signifies that on increasing temperature, the value of rain decreases and vice versa.

Since the value of correlation coefficient is very low, this shows that there is very low correlation between temperature and rain.

2. Scatter plot between Temperature and Humidity:



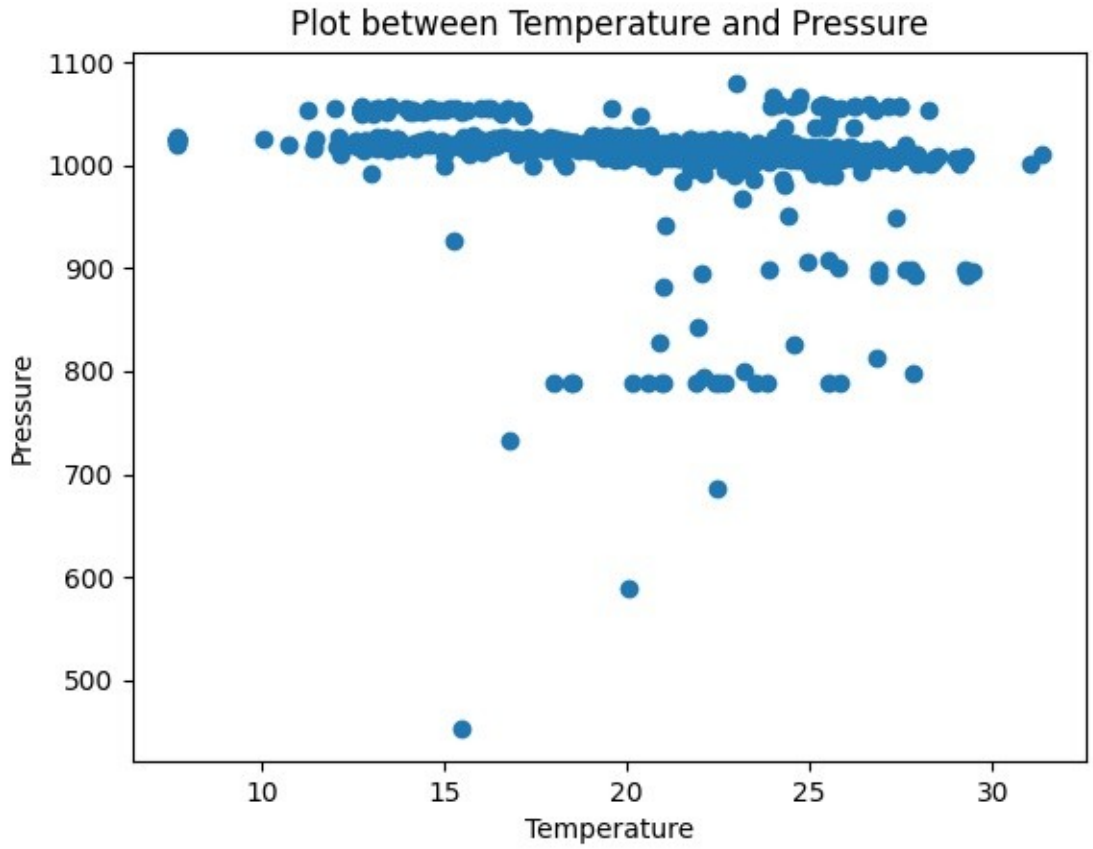
Correlation coefficient = 0.40156985

The plot looks scattered and more concentrated at the top middle .

The positive value of correlation coefficient suggests that increase in Humidity also increases Temperature and vice versa.

The magnitude of correlation coefficient is moderately low hence there is low correlation between the two quantities and the graph is scattered.

3. Scatter plot between Temperature and Pressure:



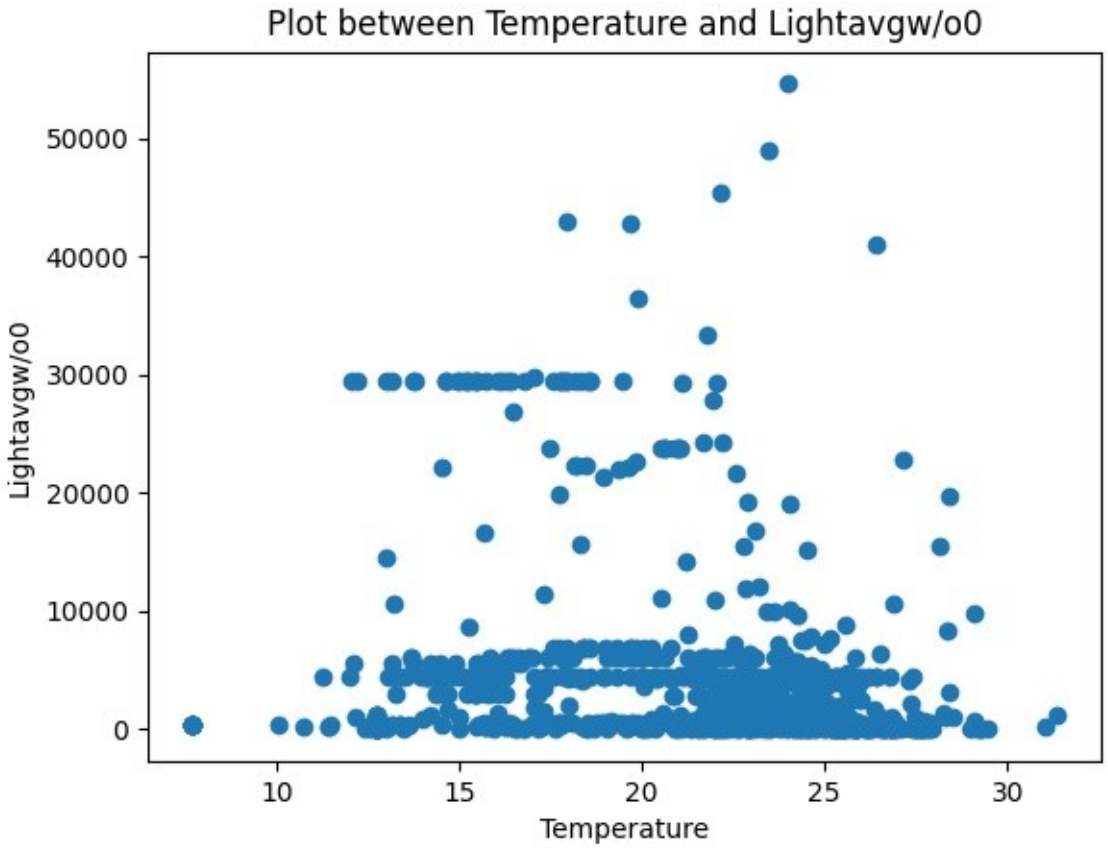
Correlation coefficient = -0.18138908

The plot is mostly concentrated around Pressure equal to about 1000mb.

The negative value of correlation coefficient signifies that on increasing temperature, the value of pressure decreases and vice versa.

Since the magnitude of correlation coefficient is very low, this shows that there is very low correlation between temperature and pressure. The two quantities are almost negligibly linearly dependent.

4. Scatter plot between Temperature and Lightavgw/o0 :



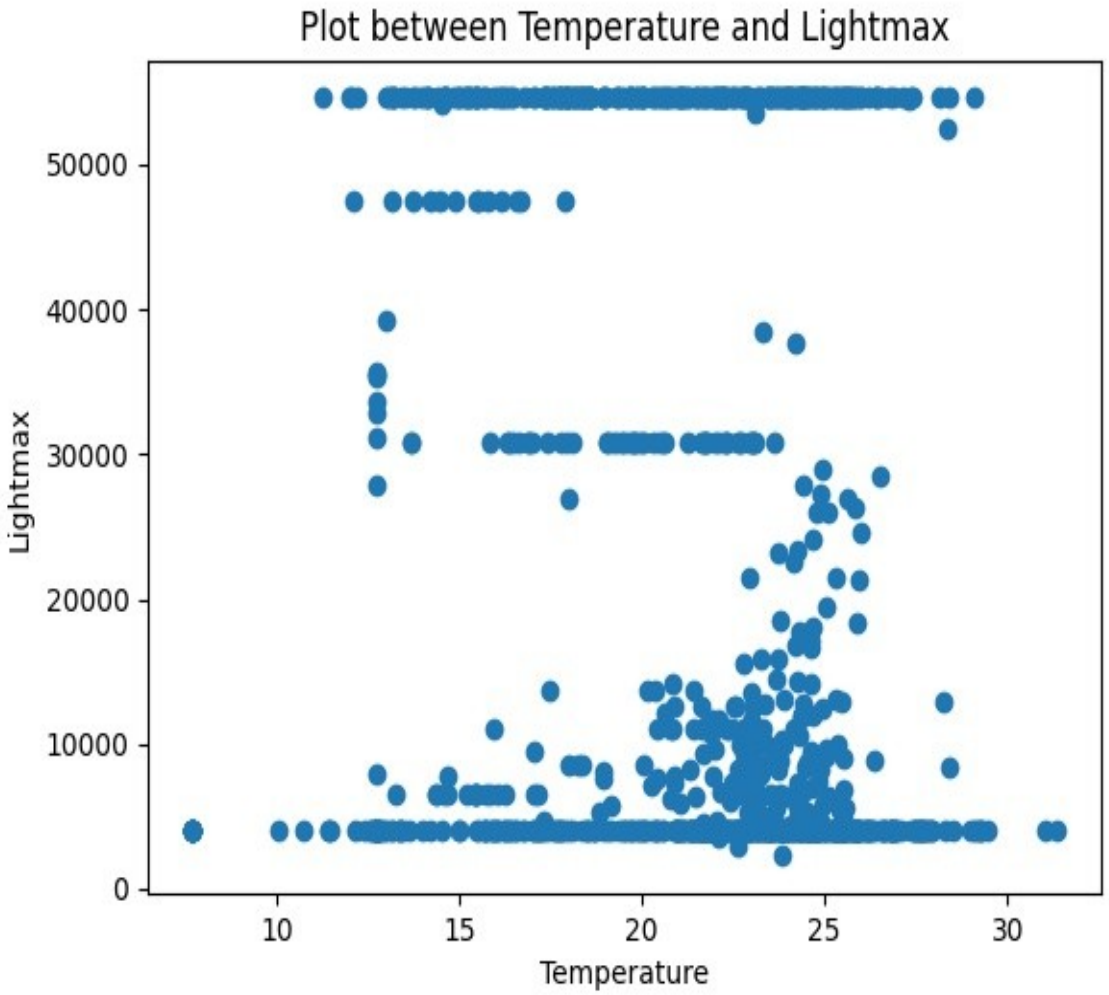
Correlation coefficient = -0.18139996

The plot is mostly concentrated around lightavg equal to about 0-10000 lux units. It is mostly dense in the bottom part of the plot.

The negative value of correlation coefficient signifies that on increasing temperature, the value of lightavg decreases and vice versa.

Since the magnitude of correlation coefficient is very low, this shows that there is very low correlation between temperature and lightavg. The two quantities are almost negligibly linearly dependent.

5. Scatter plot between Temperature and Lightmax :



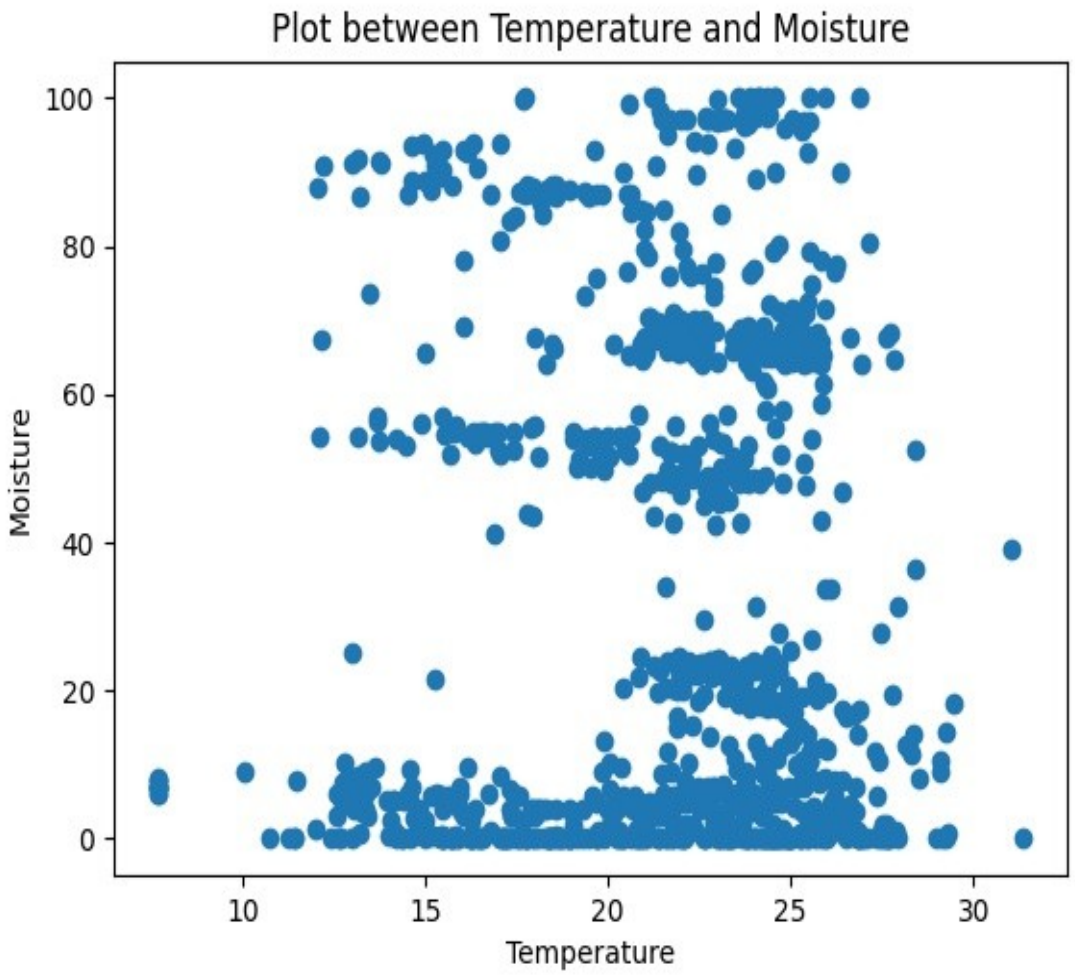
Correlation Coefficient = -0.14588351

The plot is mostly concentrated around lightmax equal to about 0-10000, 30000 and 50000 lux units.

The negative value of correlation coefficient signifies that on increasing temperature, the value of lightmax decreases and vice versa.

Since the magnitude of correlation coefficient is very low, this shows that there is very low correlation between temperature and lightmax. The two quantities are almost negligibly linearly dependent. The plot looks fairly scattered.

6. Scatter plot between Temperature and Moisture:



Correlation Coefficient = 0.08066020

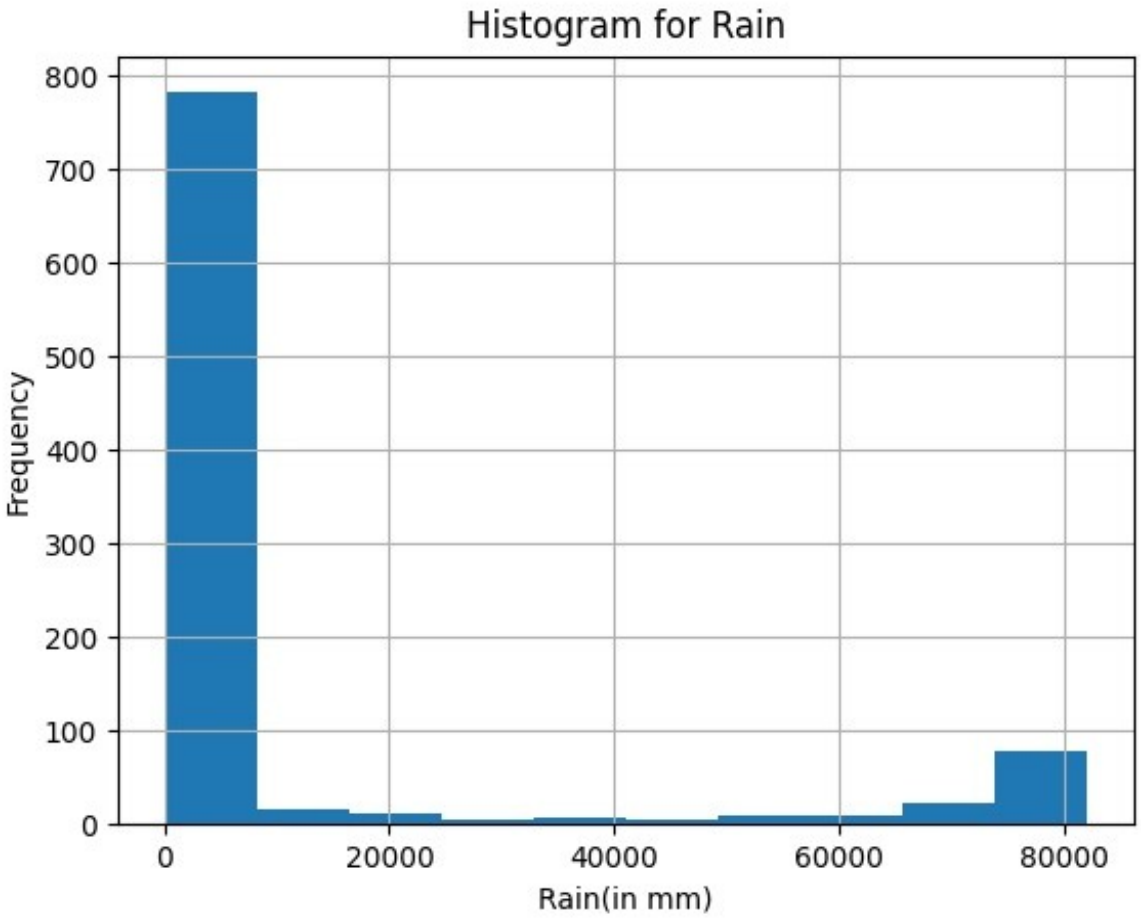
The graph looks completely scattered which ensures very low magnitude of correlation coefficient.

The positive value of correlation coefficient suggests that increase in Moisture also increases Temperature and vice versa.

Magnitude of correlation coefficient is very low which confirms that the strength of correlation is negligible and plot will be scattered.

Q4)

1. Histogram for Rain:



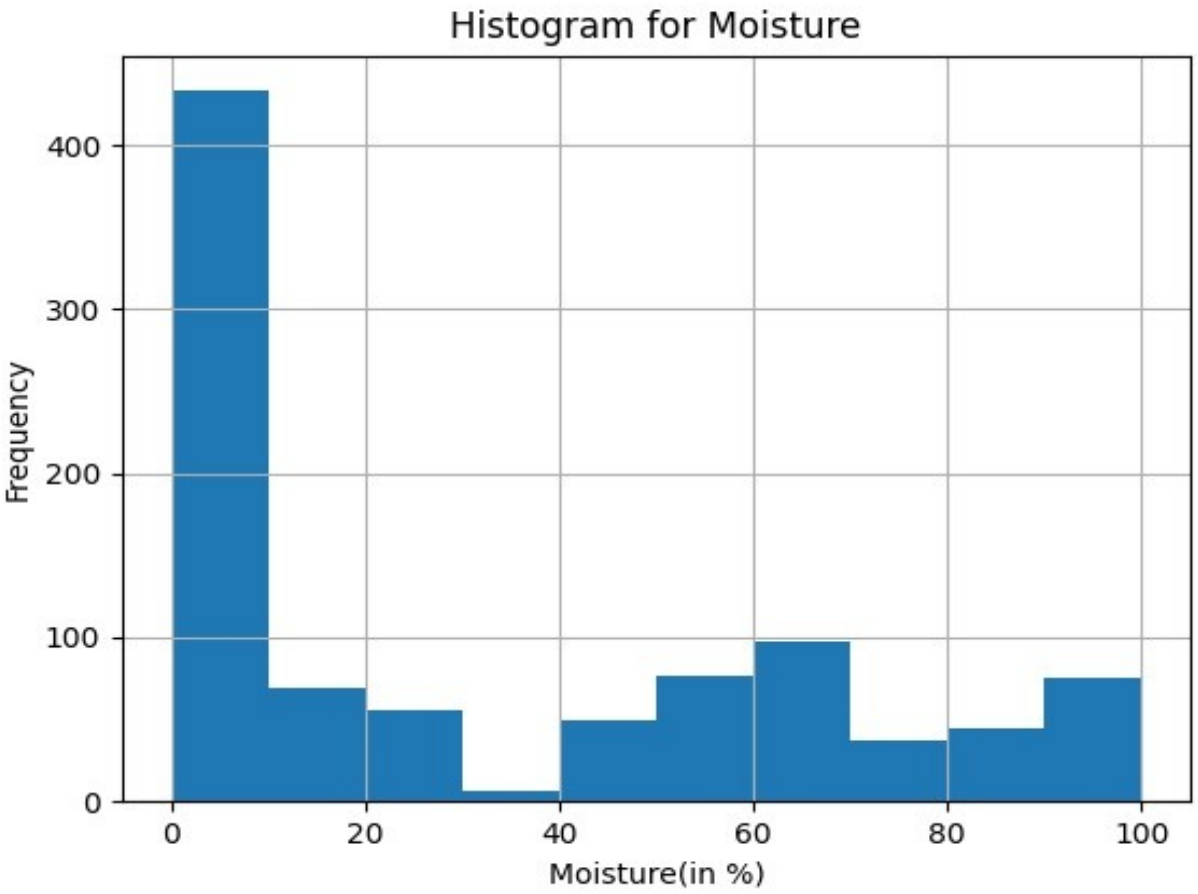
The Histogram helps us to properly visualize the data and analyse the statistics.

The histogram shows frequency of rainfall.

Minumum value of rainfall is around 0mm and maximum is around 80000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 800 at 0mm.

2. Histogram for Moisture:



The Histogram helps us to properly visualize the data and analyse the statistics.

The histogram shows frequency of moisture.

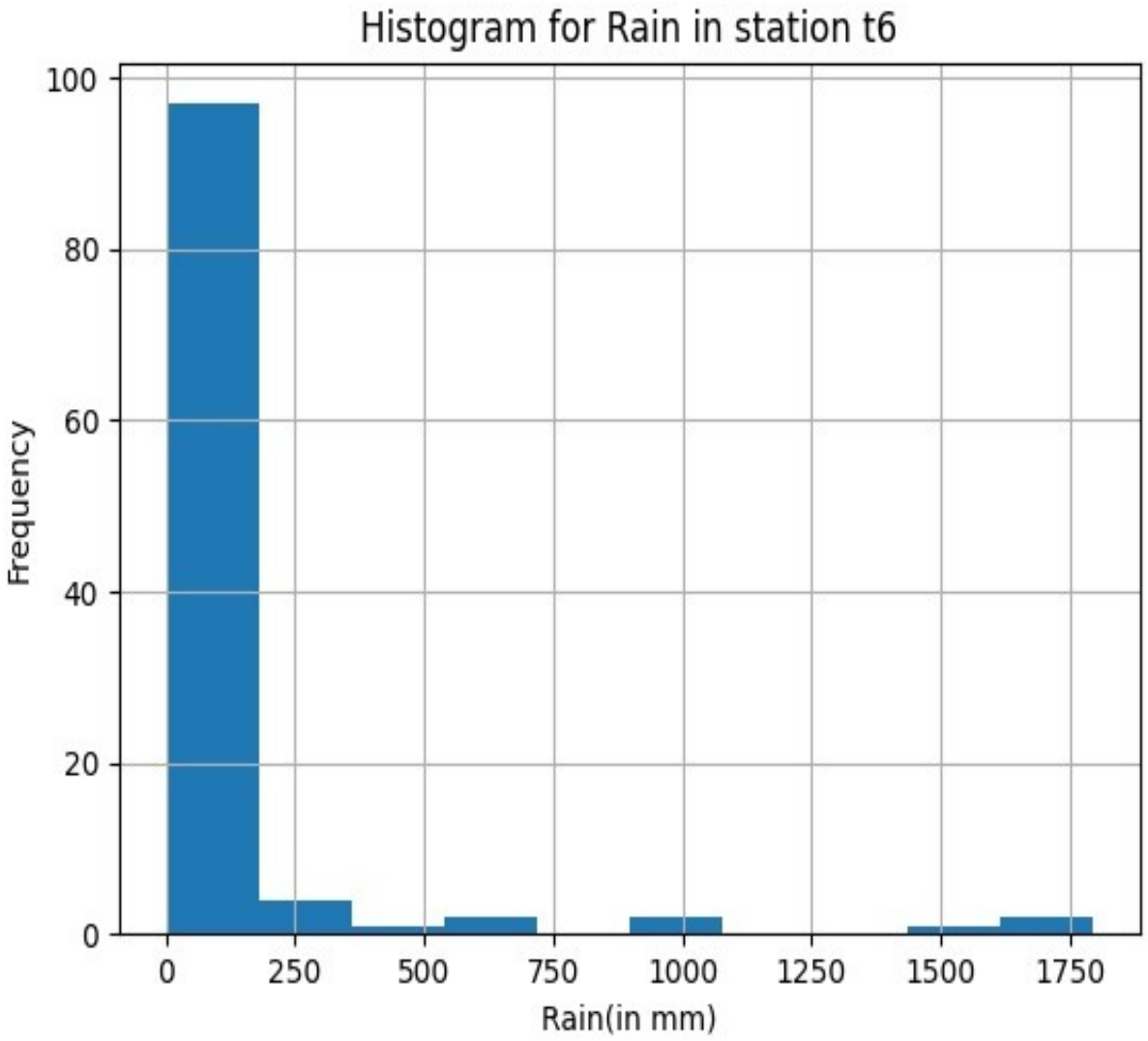
Minumum value of moisture is around 0% and maximum is around 100%.

Highest frequency is around 0% which means that the mode for moisture content is 0%. The frequency of moisture content at 0% is about 450.

The minimum frequency is from 30-40% and is about 0.

Q5)

1. Histogram for Rain in station t6 :

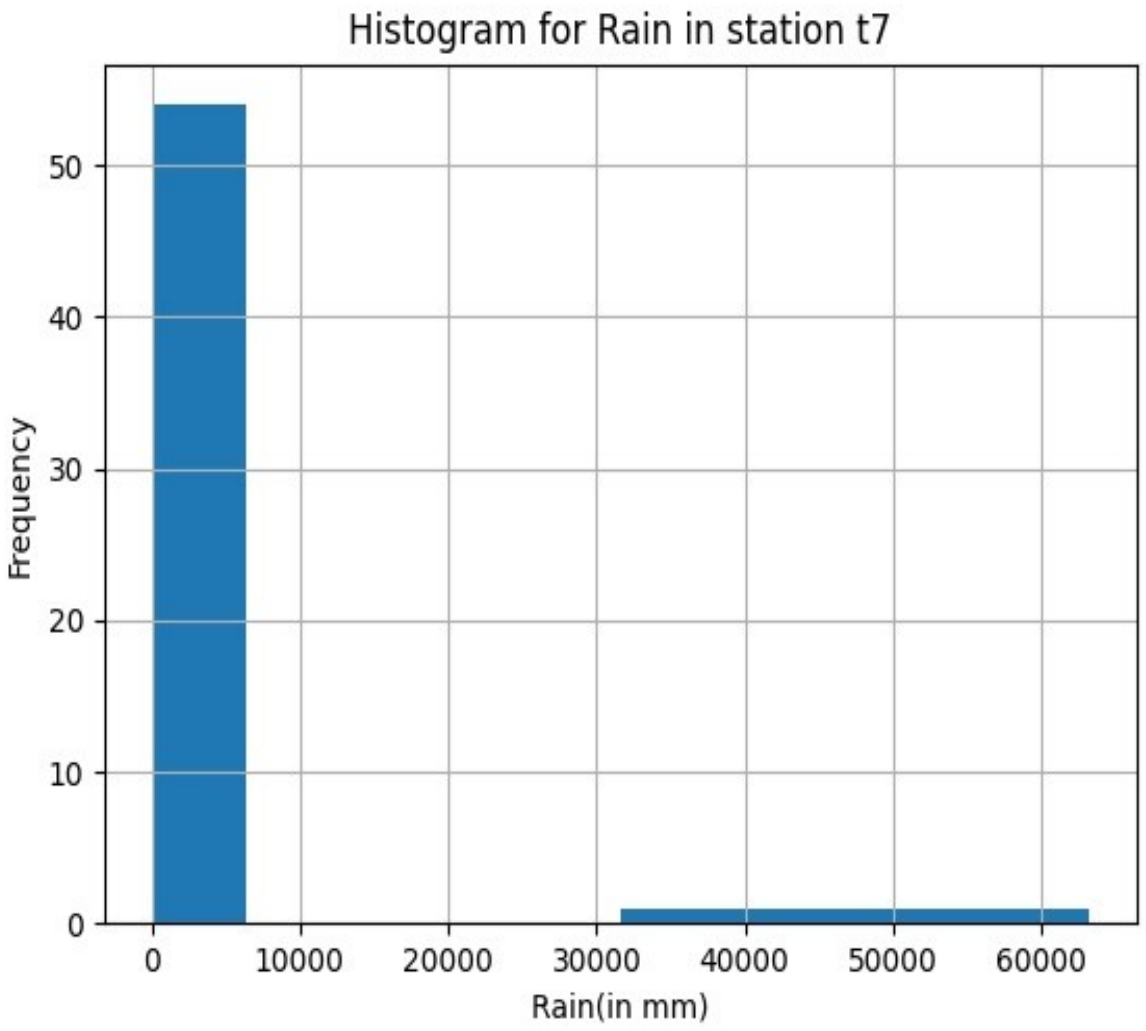


The histogram shows frequency of rainfall in station t6.

Minumum value of rainfall is around 0mm and maximum is around 1750mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 100 at 0mm.

2. Histogram for Rain in station t7 :

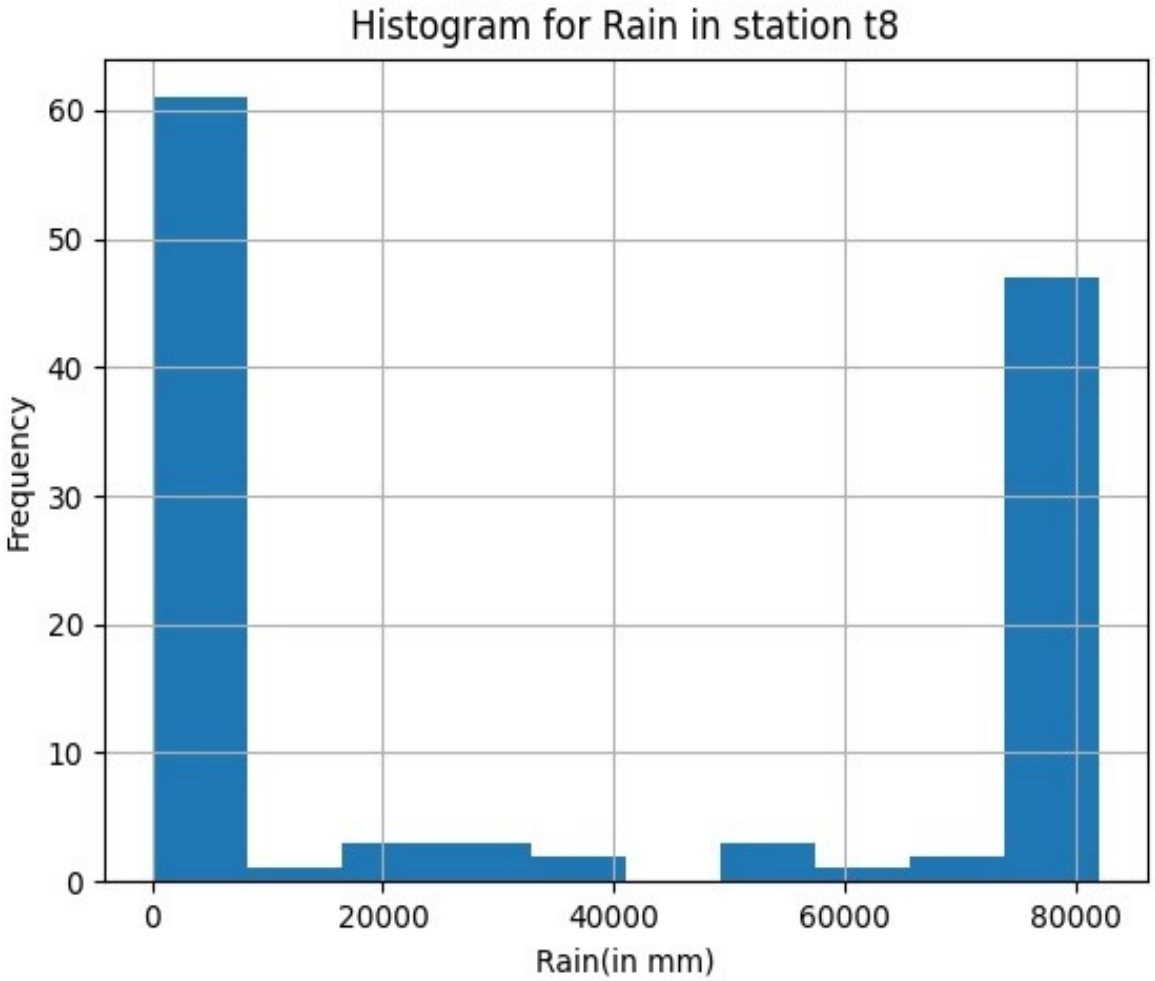


The histogram shows frequency of rainfall in station t7.

Minumum value of rainfall is around 0mm and maximum is around 60000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 55 at 0mm.

3. Histogram for Rain in station t8 :

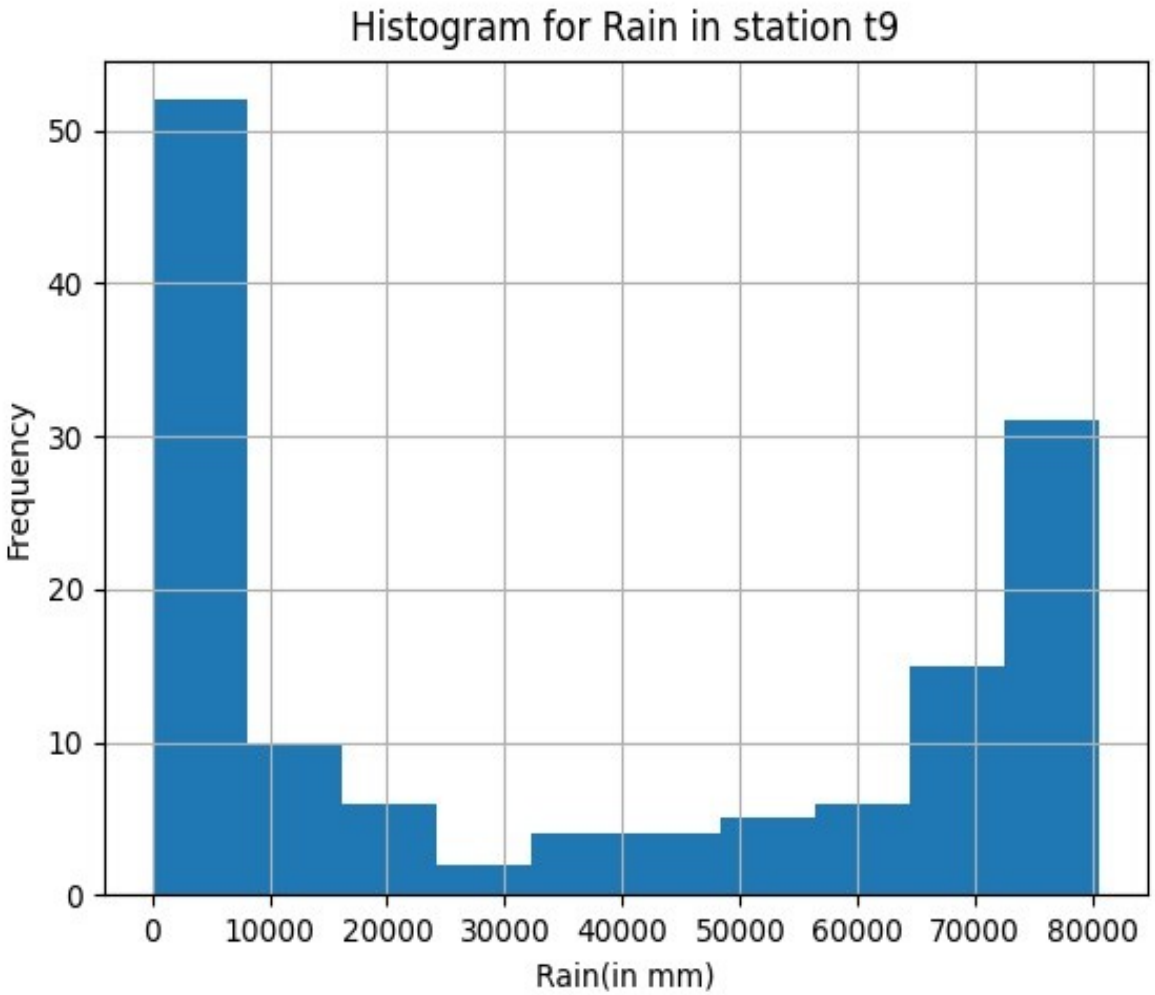


The histogram shows frequency of rainfall in station t8.

Minumum value of rainfall is around 0mm and maximum is around 80000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 60 at 0mm.

4. Histogram for Rain in station t9 :

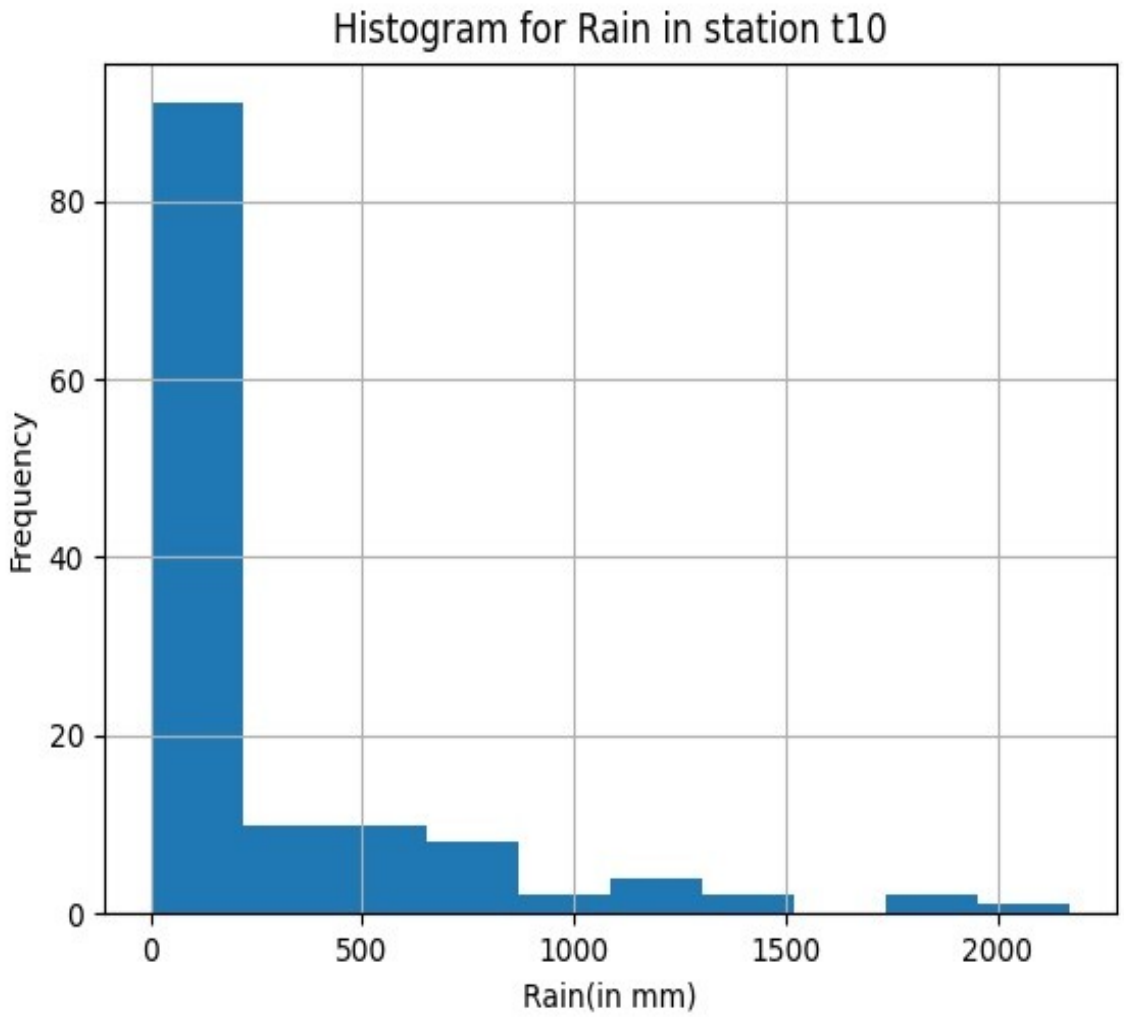


The histogram shows frequency of rainfall in station t9.

Minumum value of rainfall is around 0mm and maximum is around 80000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 55 at 0mm.

5. Histogram for Rain in station t10 :

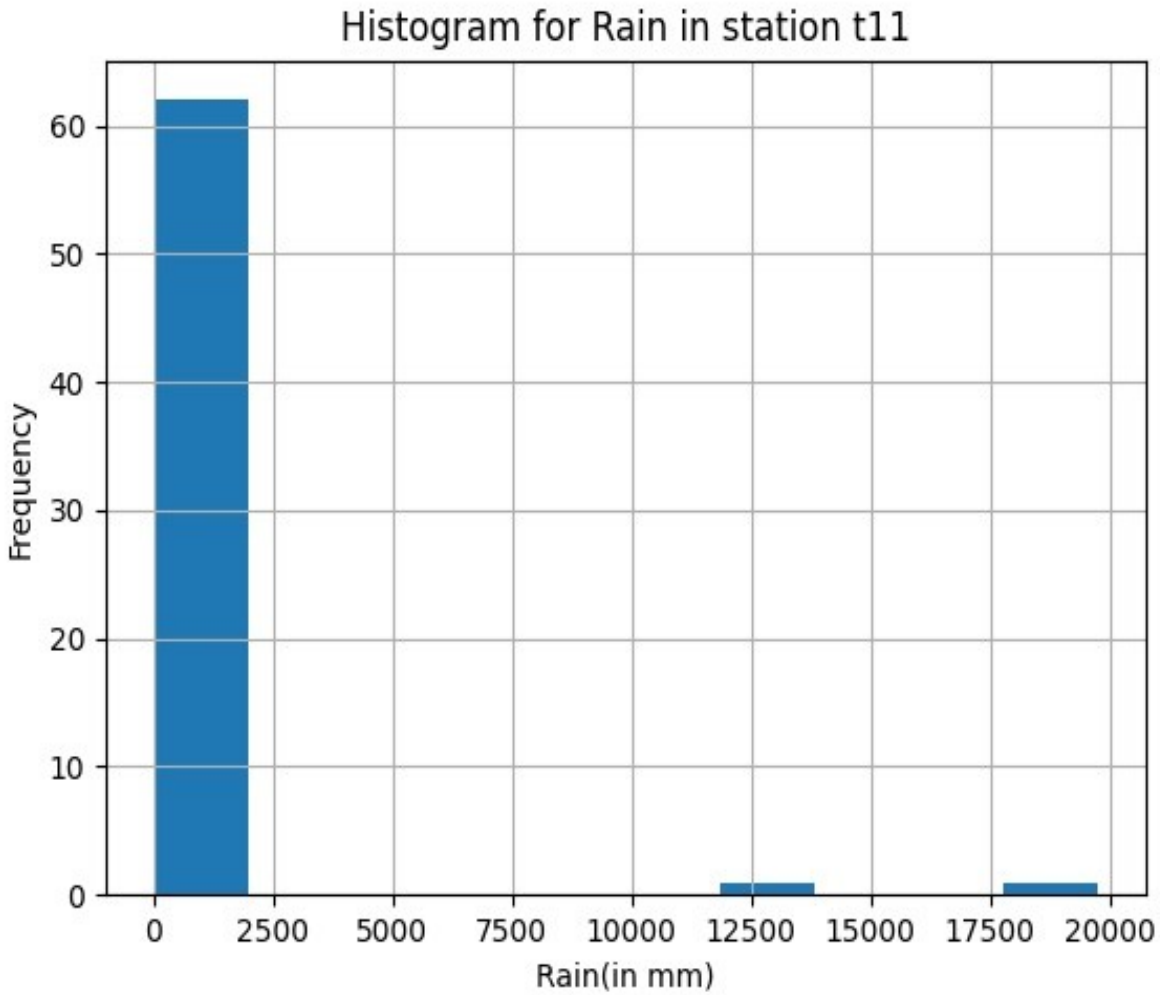


The histogram shows frequency of rainfall in station t10.

Minumum value of rainfall is around 0mm and maximum is around 2000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 85 at 0mm.

6. Histogram for Rain in station t11 :

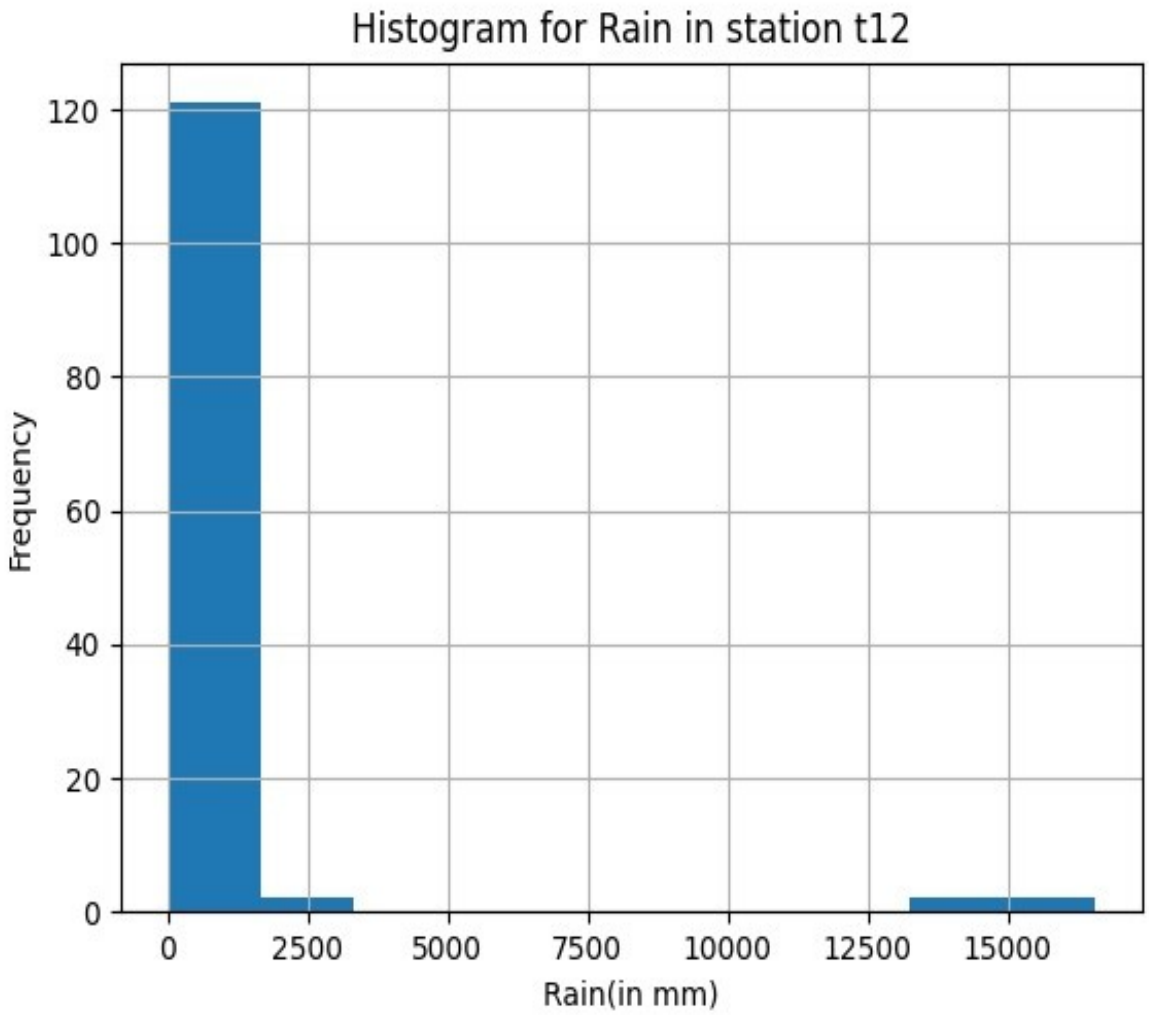


The histogram shows frequency of rainfall in station t11.

Minumum value of rainfall is around 0mm and maximum is around 20000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 60 at 0mm.

7. Histogram for Rain in station t12 :

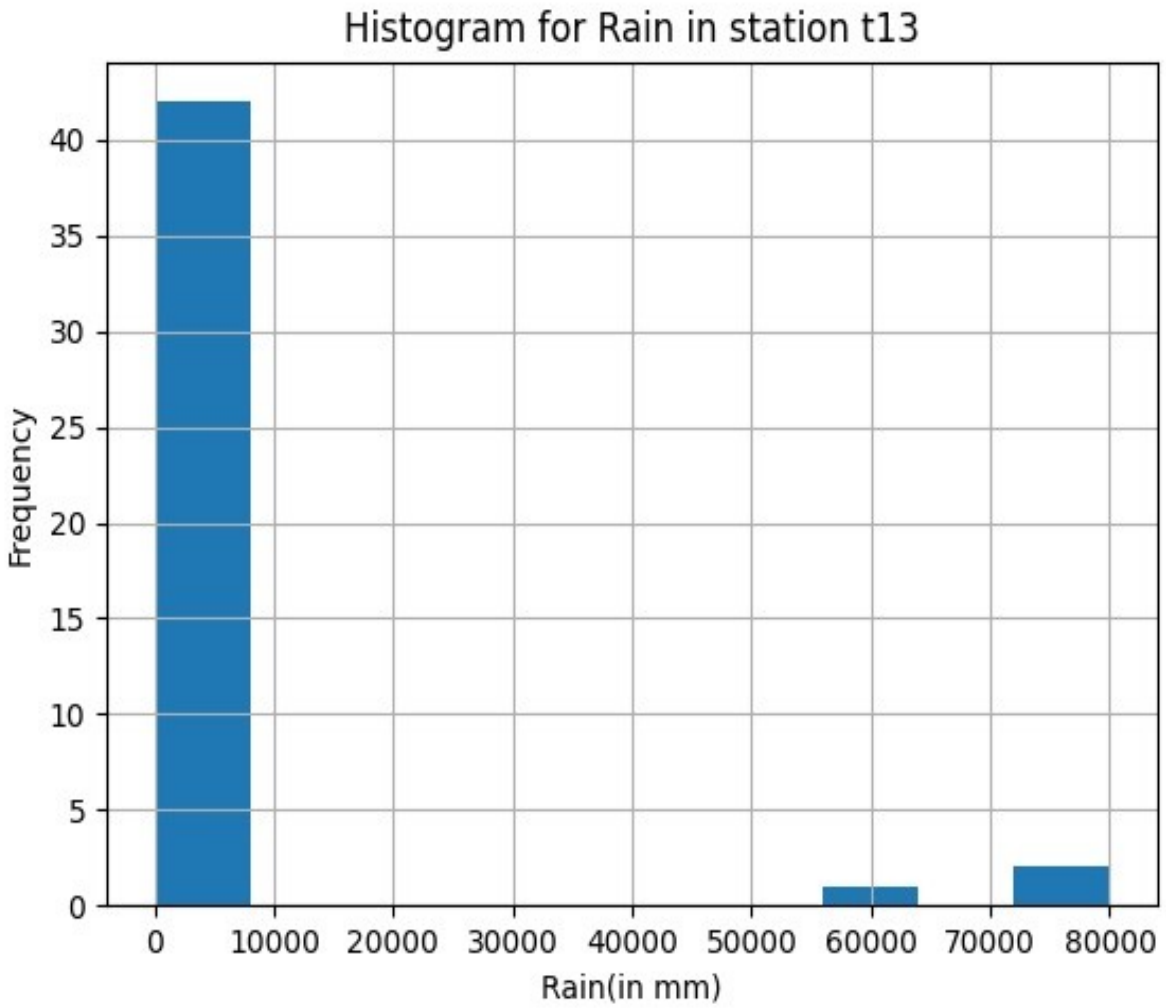


The histogram shows frequency of rainfall in station t12.

Minumum value of rainfall is around 0mm and maximum is around 15000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 120 at 0mm.

8. Histogram for Rain in station t13 :

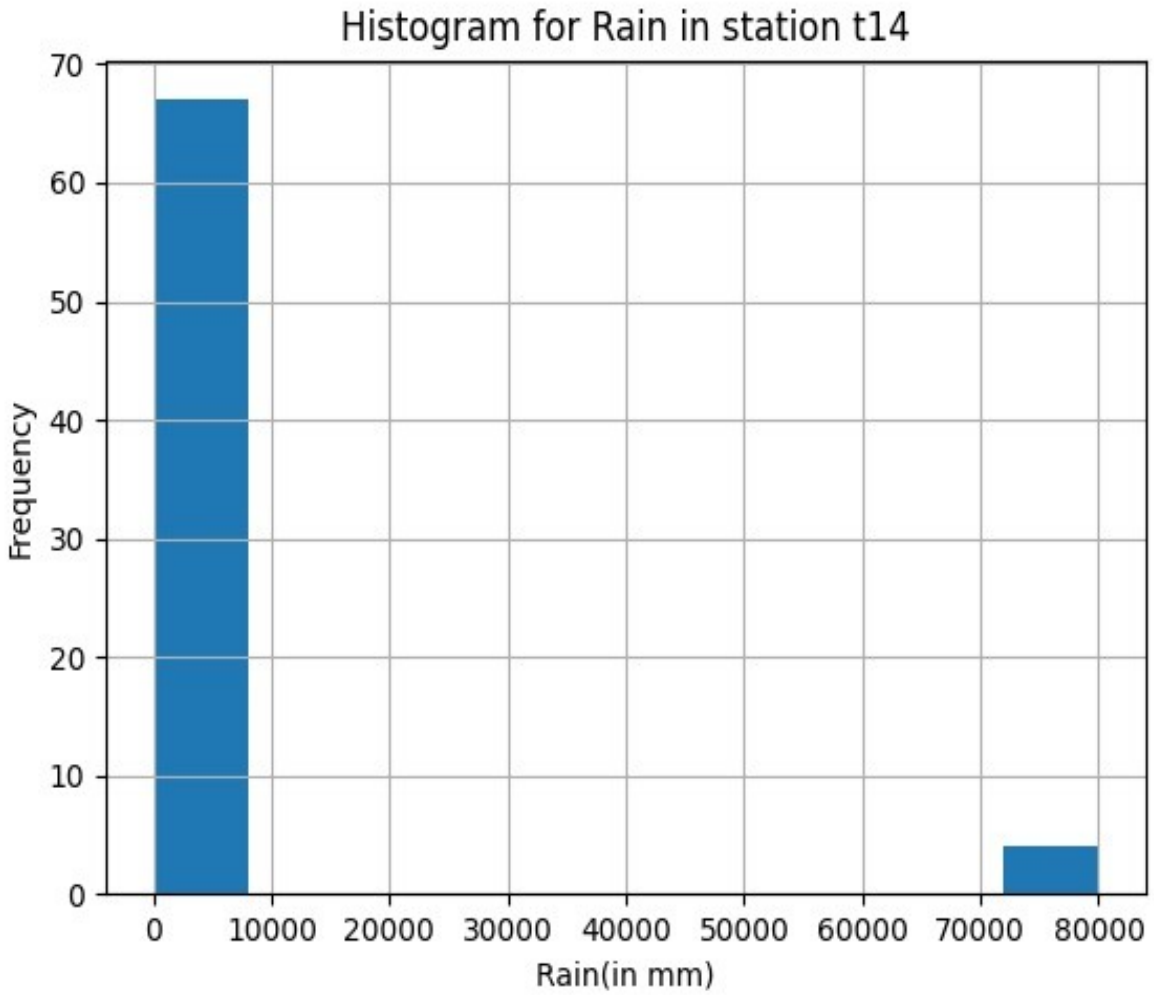


The histogram shows frequency of rainfall in station t13.

Minumum value of rainfall is around 0mm and maximum is around 80000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 42.5 at 0mm.

9. Histogram for Rain in station t14 :

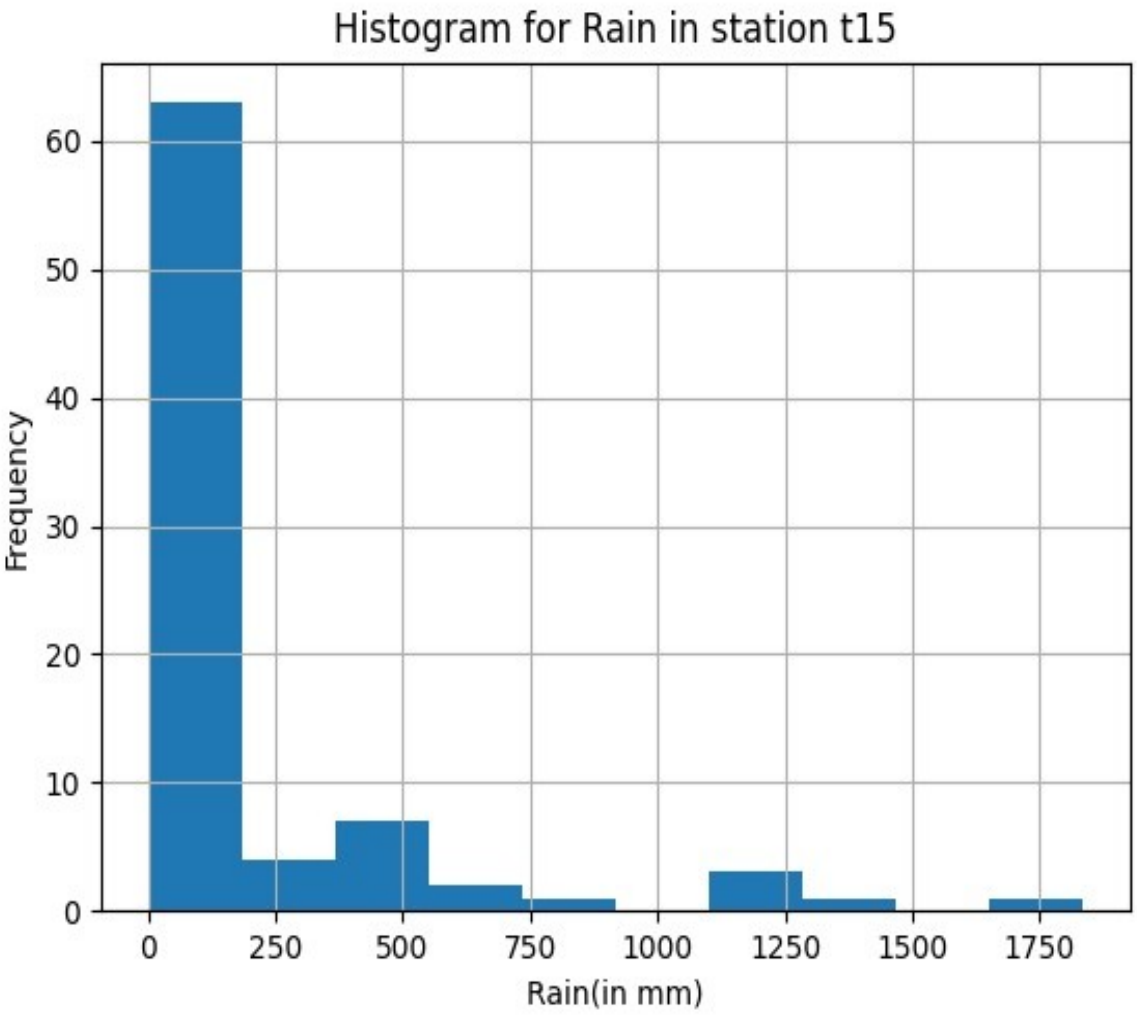


The histogram shows frequency of rainfall in station t14.

Minumum value of rainfall is around 0mm and maximum is around 80000mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 65 at 0mm.

10. Histogram for Rain in station t15 :



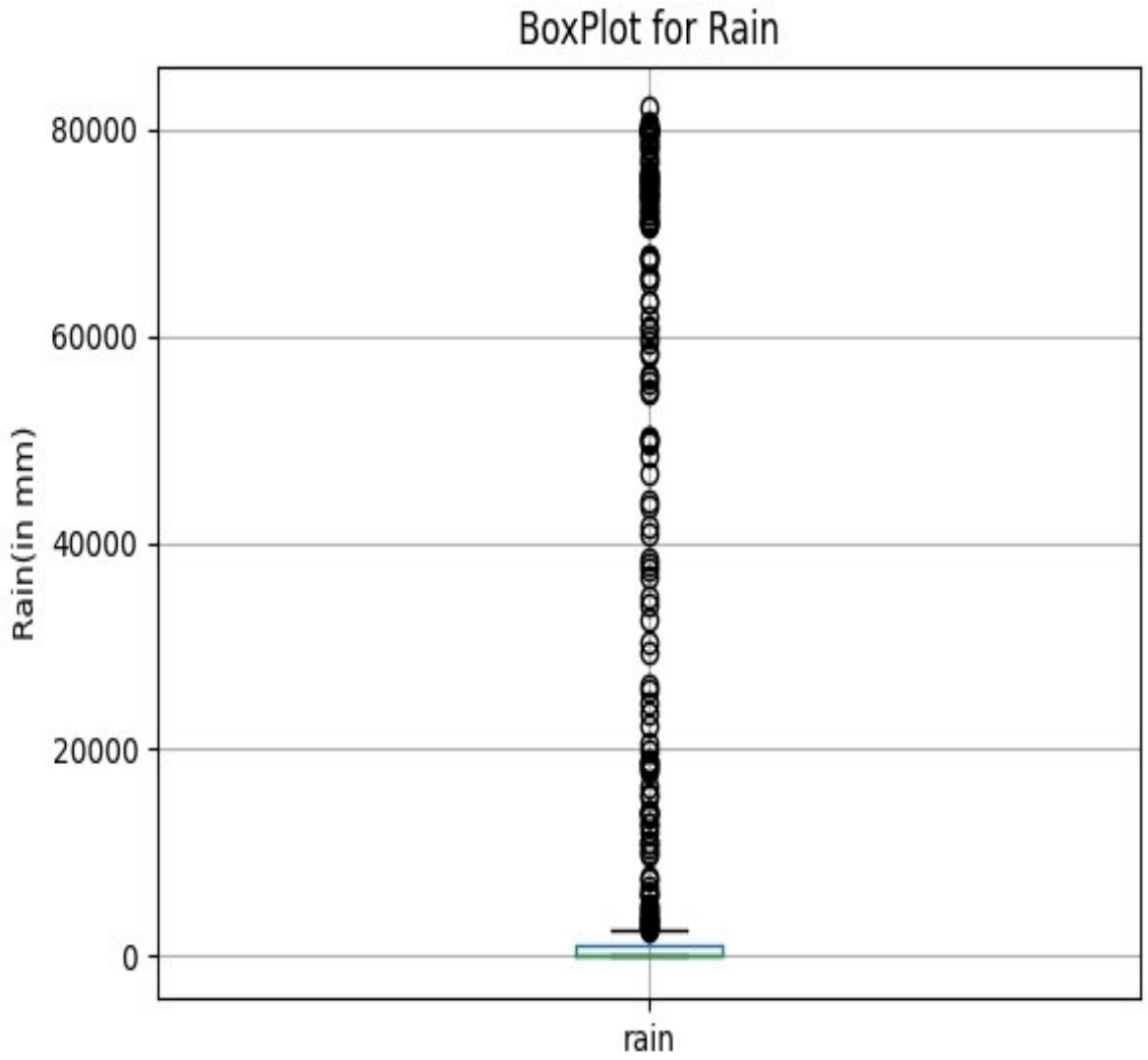
The histogram shows frequency of rainfall in station t15.

Minumum value of rainfall is around 0mm and maximum is around 1750mm.

Highest frequency is around 0mm which means that the mode for Rain is 0mm. The frequency of rainfall from graph is about 65 at 0mm.

Q6)

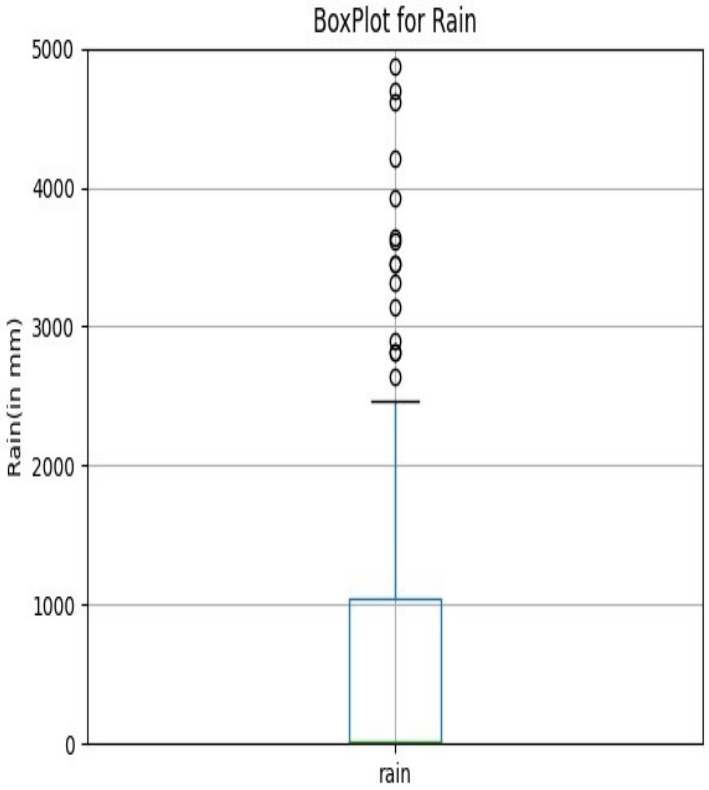
Boxplot for Rain :



Boxplots display the distribution of data based on five number summary (minimum, first quartile(Q1), median, third quartile(Q3) and maximum)

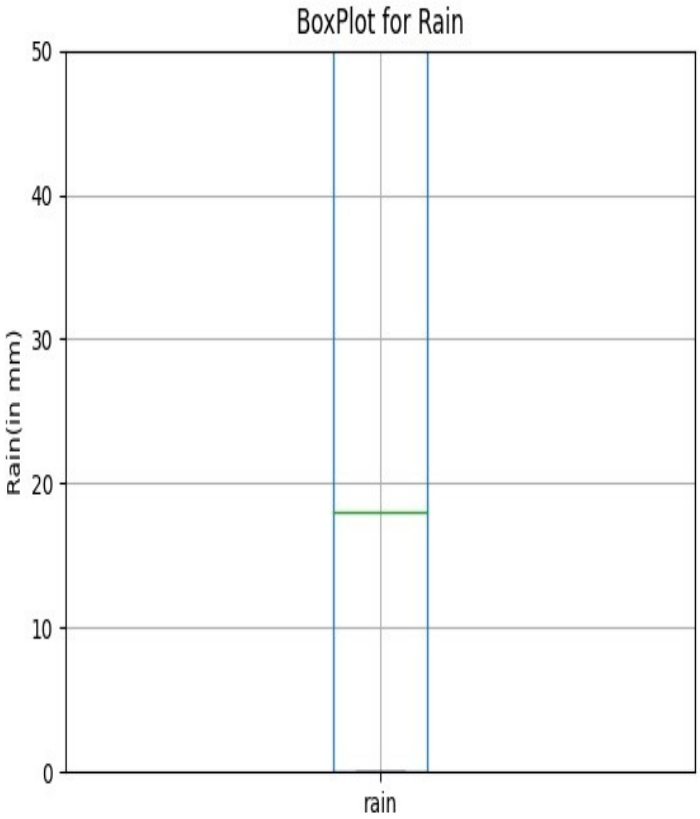
The black circles represent outliers. From the above boxplot, we can see that there are many outliers present.

For a clear look at the plot, we are looking in range (0-5000) and (0-50) on Y-axis in order to analyze other quantities clearly.



This plot shows:

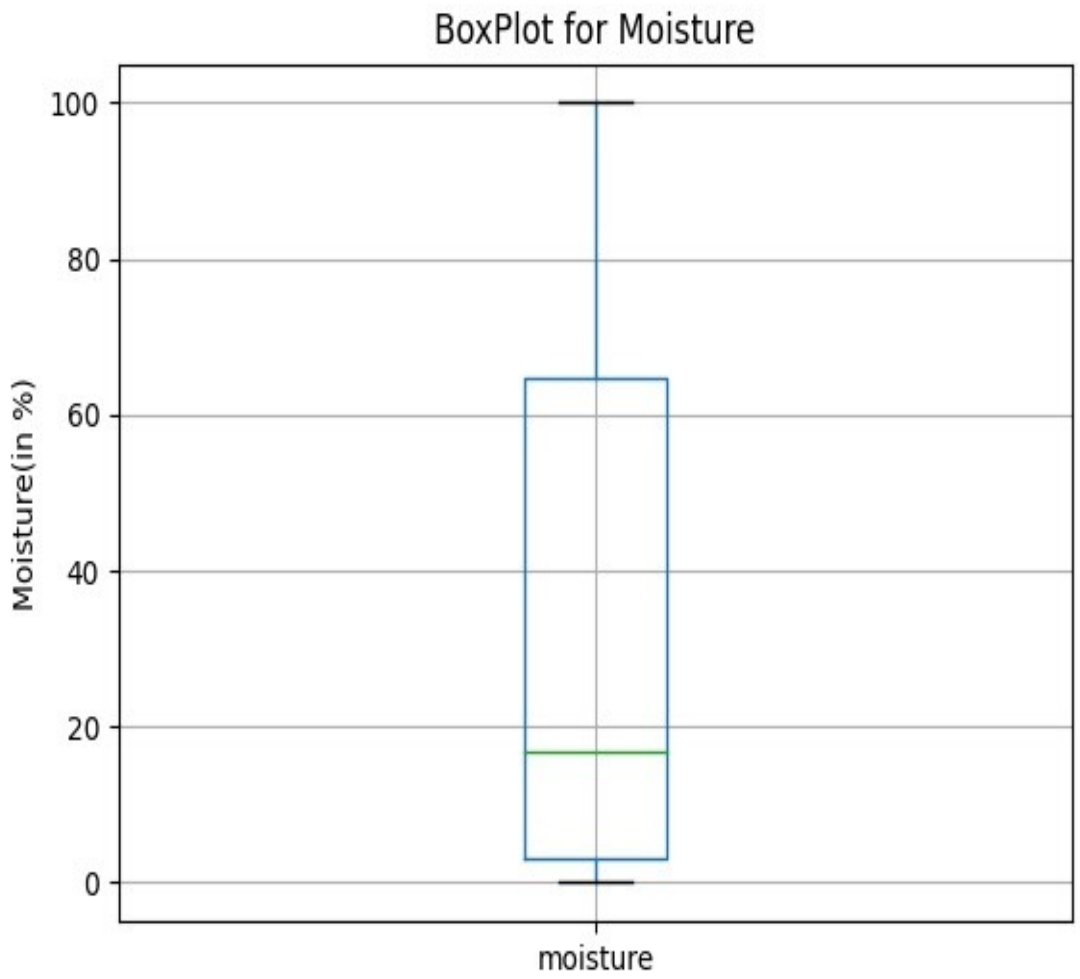
- Maximum value(Q1+1.5*IQR) which is the top whisker is around 2500.
- The Interquartile Range (IQR) is about 1000.
- Third Quartile Q3 is around 1000.



From this plot, we can infer:

- Median is approximately 17.
- First Quartile Q1 is about 0.
- Minimum value i.e. bottom whisker is also 0.

Boxplot for Moisture :



- Maximum value(Q1+1.5*IQR) which is the top whisker is 100.
- Third Quartile Q3 is approximately 65.
- First Quartile Q1 is approximately 3.
- The Inter-Quartile Range = Q3-Q1 = 62.
- Median is approximately 16.
- Minimum value (Q1-1.5*IQR) which is the Bottom whisker is 0.
- There are no outliers present in this plot.
