

## **ASSIGNMENT – 1 , Report**

**IIT MANDI, IC-272**

**Name – Sumit Kumar**

**Roll No. – B19118**

**Mobile Number – 7549233722**

Q1. Finding the Mean, Median, Mode, Minimum, Maximum and Standard Deviation of the attributes except “dates” and “stationed”:-

a. For “temperature”:-

Mean is **21.214888105820105**

Median is **22.27273**

Mode is **12.727269999999999**

Maximum is **31.375**

Minimum is **7.6729**

Standard Deviation is **4.353512664832556**

b. For humidity :

Mean is **83.47993155555555**

Median is **91.38095**

Mode is **99.0**

Maximum is **99.72**

Minimum is **31.0**

Standard Deviation is **18.20042716285037**

c. For pressure :

Mean is **1009.008773798647**

Median is **1014.6778321678**

Mode is **789.3926923077**

Maximum is **1079.162**

Minimum is **452.0978873239**

Standard Deviation is **46.95561327534729**

d. For rain :

Mean is **10701.53837037037**

Median is **18.0**

Mode is **0.0**

Maximum is **82037.25**

Minimum is **0.0**

Standard Deviation is **24839.10246612766**

e. For lightavgw/o0 :

Mean is **4438.428453333333**

Median is **1656.88**

Mode is **4488.9103**

Maximum is **54612.0**

Minimum is **0.0**

Standard Deviation is **7569.154781086212**

f. For lightmax :

Mean is **21788.62328042328**

Median is **6634**

Mode is **4000**

Maximum is **54612**

Minimum is **2259**

Standard Deviation is **22053.315399022737**

g. For moisture :

Mean is **32.38605259259259**

Median is **16.7042**

Mode is **0.0**

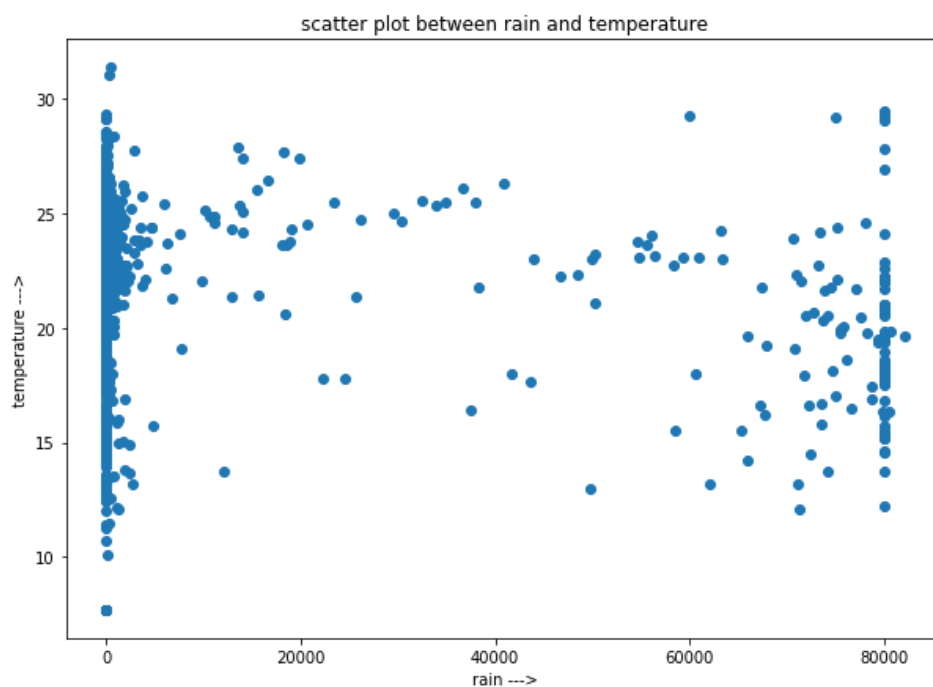
Maximum is **100.0**

minimum is **0.0**

Standard Deviation is **33.653244650764854**

- I have not used the inbuilt function to find the Mean, Median, Mode, Minimum, Maximum and Standard Deviation of the attributes therefore the values of the attributes may be slight differs after the decimal point.

Q2(a).

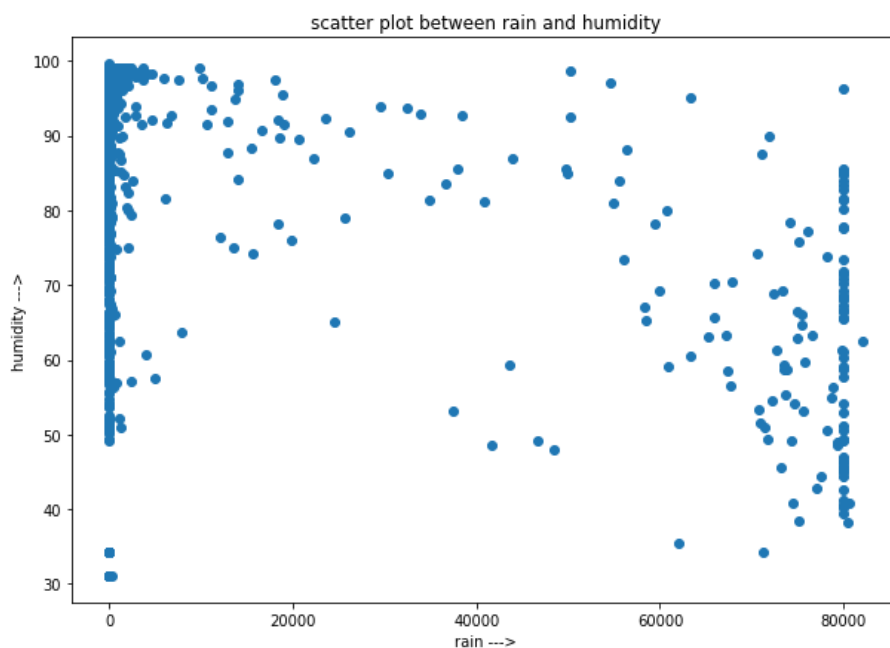


### OBSERVATION:-

- The graph is very dense when rain is around 0 and 80000.
- Most of the points are close to y-axis.
- The graph is not following a linear dependency because most of the point are densely packed near y- axis and points are non-uniformly scattered for rain between 20000-75000.
- For almost each values of the temperature, there is atleast one value of rain which is close to 0.
- correlation coefficient between rain and temperature is **-0.10877805105737656**
- We cannot use linear regression model to fit the data points.

### CONCLUSION:-

- Very little or no rainfall occurs at all values of temperature.
- Very heavy rainfall occurs between temperature range of 15-25.
- Most of the places experience less rainfall than the average because most of the points are located on left of the average rain value **10701.53837037037**.
- Since the average temperature is **21.214888105820105**, we can observe that almost equal points are located on the top and bottom of the average temperature.
- The correlation coefficient is negative.

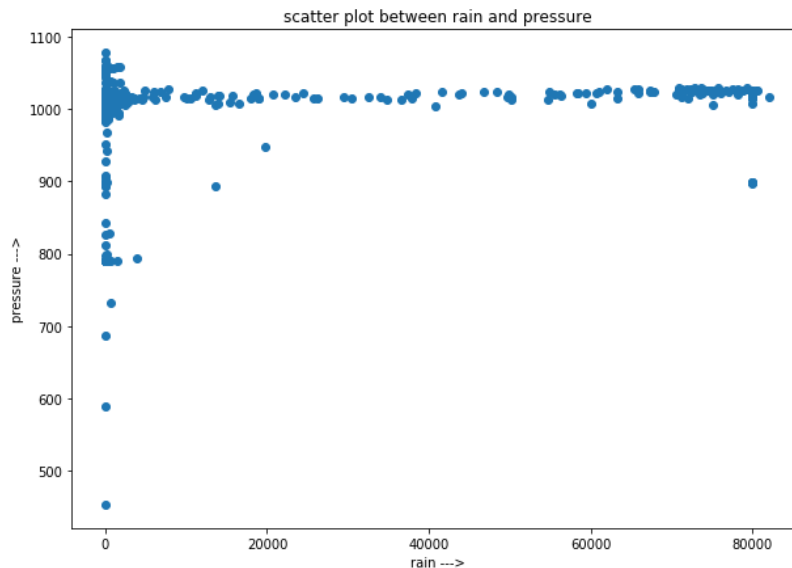


### OBSERVATION:-

- The graph is very dense when rain is around 0 and 80000.
- The points are uniformly scattered around rain between 70000-80000 but when rain is close to 80000, the points seems to be in a line parallel to y-axis.
- Very few points are lying between rain 20000-60000.
- The maximum value of the humidity is 99.72.
- correlation coefficient between rain and humidity is **-0.4344566152391602(negative)**.

### CONCLUSION:-

- It can be observed that the rain and the humidity are not linearly dependent.
- Zero or no rainfall is found to be between the humidity range of 50-99.72
- The points having no rainfall but high humidity is very worse climate and these are majority in numbers.
- We cannot use a linear regression model to plot the data points.

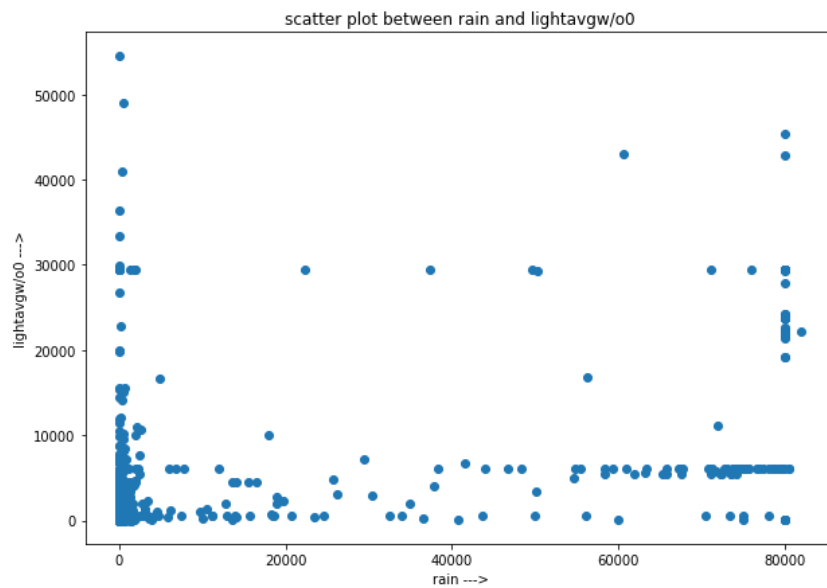


### OBSERVATION:-

- Almost all the points are found to be placed parallel to the x-axis.
- For all values of rain, the value of the pressure is nearly constant around 1000-1050
- Maximum humidity occurs for rain around 0-500.
- correlation coefficient between rain and pressure is **0.07070976966275146**

### CONCLUSION:-

- Since the data points are uniformly distributed parallel to the x-axis, we can say that the pressure is linearly dependent on the rain.
- The maximum value of the pressure is around 1100.
- Most of the points lie on the right side of the average value of rain i.e **10701.538**.
- From the graph it seems that most of the rainfall occurs when the pressure is high.
- The correlation coefficient is very low and here We cannot use a linear regression model to plot the data points because rain and pressure is linearly dependent.

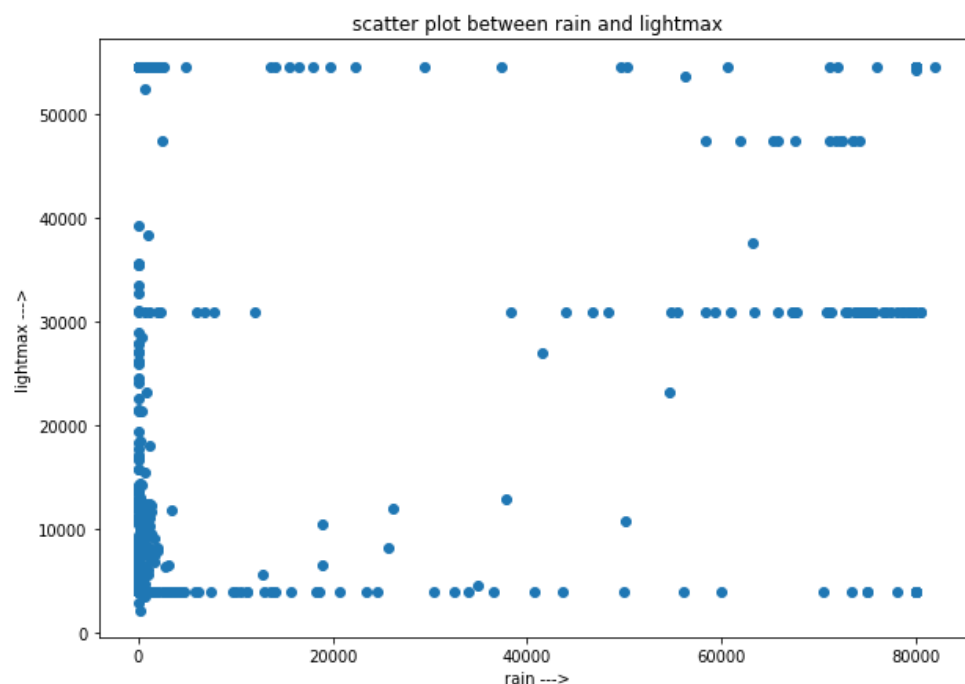


#### OBSERVATION:-

- The scatter points are highly dense around the lightavgw/o0 below 20000.
- There are only a few points above 20000.
- Most of the points are around 0 rain so the points are highly shifted around around 0-20000 rain.
- correlation coefficient between rain and lightavgw/o0 is **0.526932118836284**

#### CONCLUSION:-

- The graph is mostly lying around the line  $y = 5000$  but it is not following a strong linear relationship between lightavgw/o0 and rain.
- The maximum lightavgw/o0 is above 50000 which occurs for very low or no rainfall.
- We can use a linear regression model to plot the points.

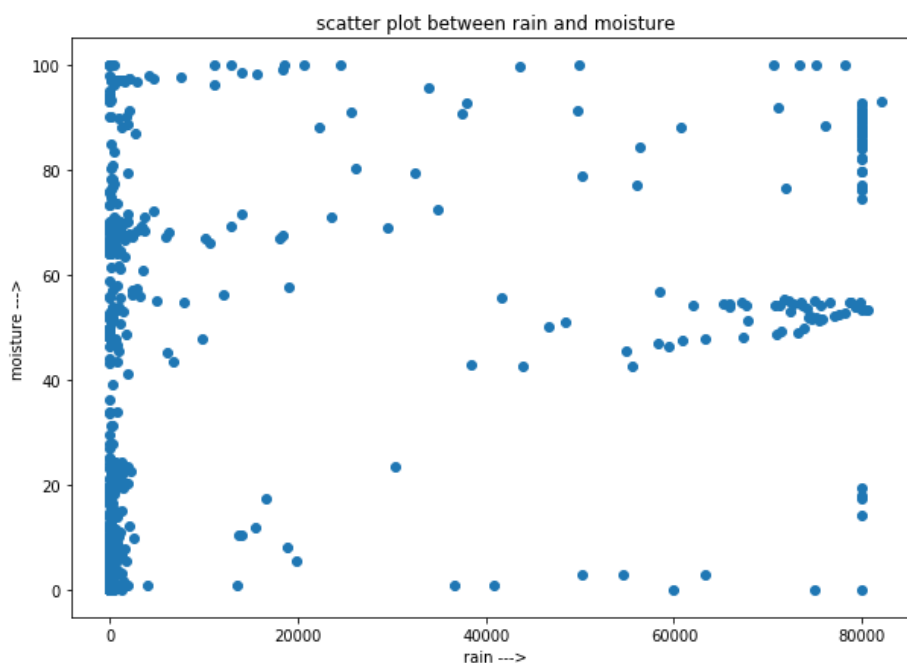


#### OBSERVATION:-

- Most of the points lie around low rainfall and high light.
- The maximum value of light intensity is 54612 and around 20 points which experience maximum light intensity.
- Light intensity around 2000-30000 are most in numbers.
- correlation coefficient between rain and lightmax is **0.31251168706558863**

#### CONCLUSION:-

- There is no any linear relationship between rain and lightmax.
- Mostly heavy rainfall occurs for high value of lightmax.
- The relation between the attributes is moderate.
- We cannot use linear regression model to plot the attributes.



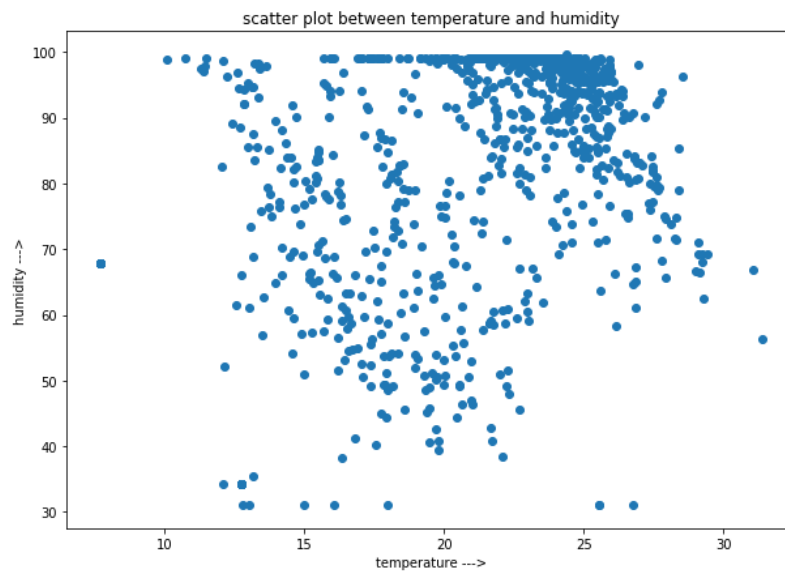
#### OBSERVATION:-

- The graph is densely paced near y-axis.
- Most of the points lie around the area of zero rainfall.
- Almost all values of moisture occurs for zero rainfall.
- For rainfall greater than 5000, the points are scattered everywhere.
- Maximum value of moisture is nearly hundred.
- correlation coefficient between rain and moisture is **0.42647615378681775**

#### CONCLUSION:-

- There is no any linear relationship between rain and moisture because the points are not lying in a uniform pattern.
- Most of the places experience rainfall less than the average rain.
- Most of the heavy rainfall occurs under high moisture content.
- Linear regression model cannot be used here.

Q2(b).

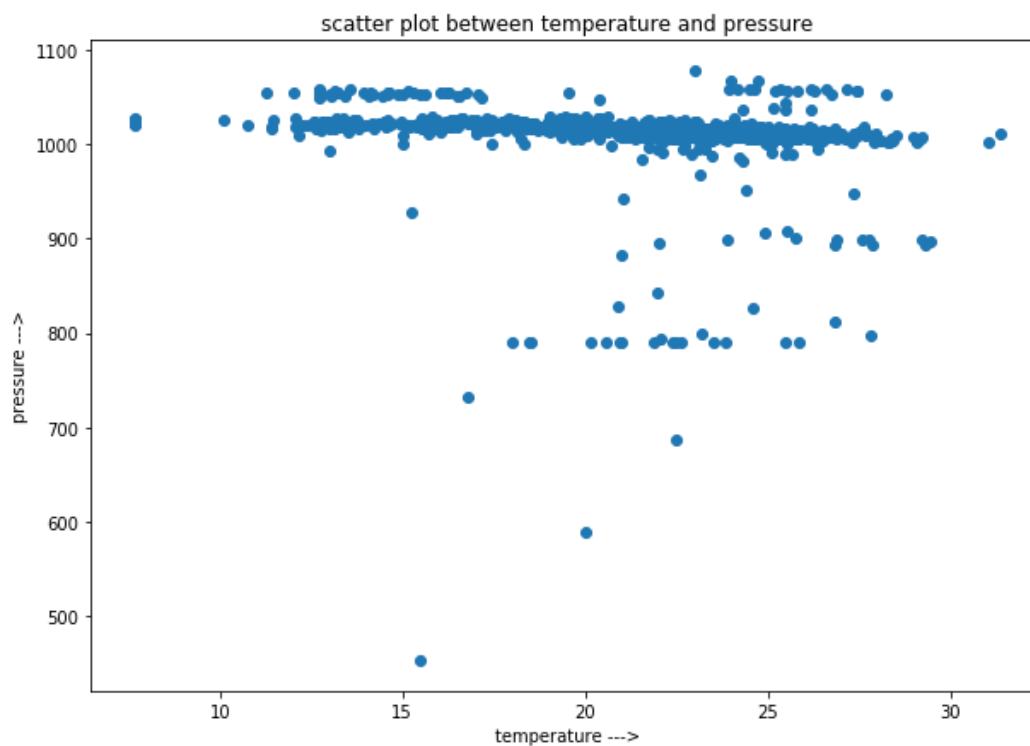


#### OBSERVATION:-

- The average value of the temperature is found to be is **21.2148**
- Most of the points are uniformly scattered for rainfall between 10-20 and the points are packed densely around the rainfall above 20.
- Most of the points are lying above the average temperature.
- correlation coefficient between temperature and humidity is **0.40114490623630455**
- The relation between the attributes is moderate.

#### CONCLUSION:-

- The plot is not following any linear relationship between rain and humidity.
- Most of the stations experience humidity of range 70-100.

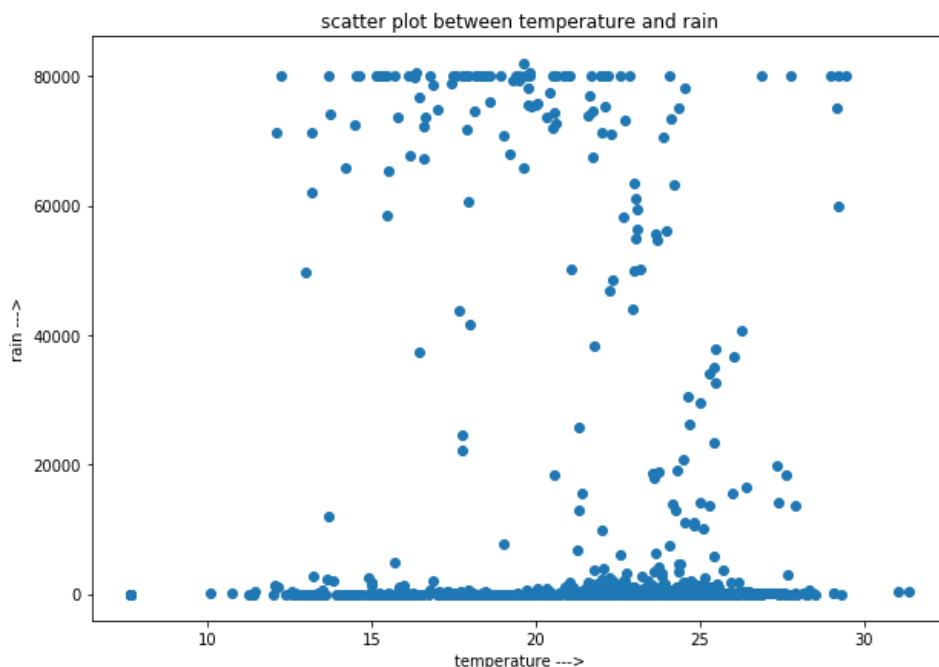


#### OBSERVATION:-

- The graph is dense around the pressure of 1000.
- Most of the points are lying around the pressure of 1000 for almost all values of temperature.
- The mean value of the pressure is 1009 and most of the stations experience average pressure.
- correlation coefficient between temperature and pressure is **-0.1811971291075043**

#### CONCLUSION:-

- There is not a very strong linear relationship between the temperature and the pressure because most of the points are scattered around the line parallel to x-axis ( $y = 1000$ ).
- Very few stations are there which experience the pressure below 1000.
- The mean of the pressure is nearly equal to the maximum value of pressure.
- We can use a linear regression model here.



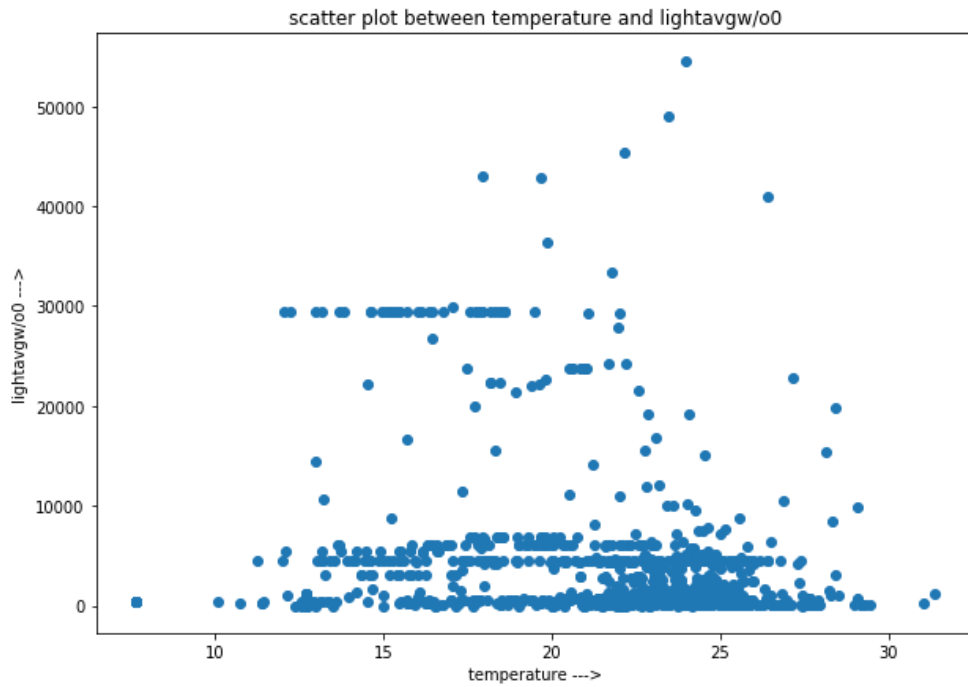
#### OBSERVATION:-

- As observed earlier, most of the stations experience low or zero rainfall.
- Most of the high rainfall occurs for a temperature range of 25-25.
- correlation coefficient between temperature and rain is **-0.1087780510573765**

#### CONCLUSION:-

- There is no any linear relationship between the rain and the temperature.
- Low rainfall occurs when the temperature is below average.
- Most of the points lie around the temperature range of 20-30.
- The minimum value of temperature is 7.6729 and the maximum value of the temperature is around 31.
- Linear regression model cannot be used here.



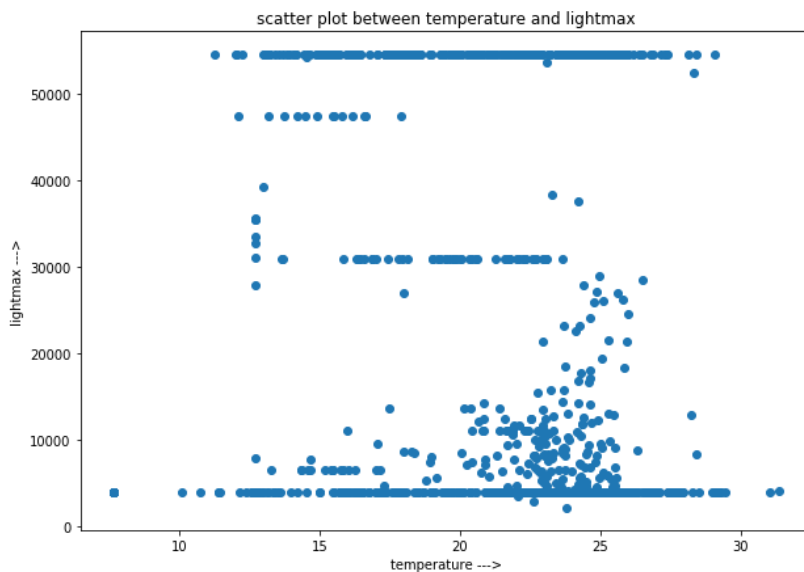


#### OBSERVATION:-

- The graph is very dense around the lightavgw/o0 value of 0-10000 which denotes that most of the stations experience low light intensity.
- There are only few stations which experience light intensity greater than 10000.
- The maximum value of the lightavgw/o0 is **54612** and the minimum value is **0**.

#### CONCLUSION:-

- There is no any linear relationship between temperature and lightavgw/o0.
- Almost equal points lie below and above the average temperature.
- correlation coefficient between temperature and lightavgw/o0 is **-0.1812080058457248**
- correlation coefficient is negative and here we can use the linear regression model for the attributes.

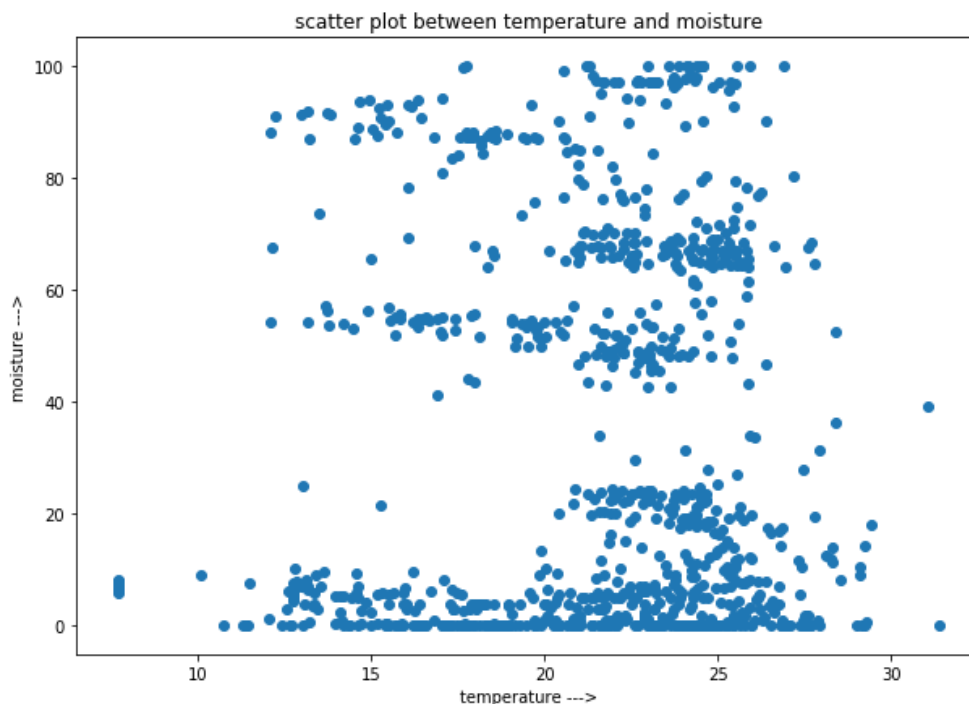


#### OBSERVATION:-

- The points are densely packed near the x-axis.
- The average value of lightmax is **21788.623** and the maximum value is **54612**.
- Most of the points are lying around the average temperature with low value of lightmax.
- correlation coefficient between temperature and lightmax is **-0.14572914001778584**

#### CONCLUSION:-

- Most of the points are lying below the lightmax value less than 20000 which conclude that most stations experience less lightmax.
- There is no any linear relationship between lightmax and temperature.
- Linear regression model cannot be used here.



#### OBSERVATION:-

- Most of the stations experience less moisture content .
- The minimum value of the moisture is found to be around 0 and the maximum value is 100.
- There are very few stations which experience temperature below 10.
- Most of the stations experience average temperature.
- correlation coefficient between temperature and moisture is **0.08057484324879301**

#### CONCLUSION:-

- There is no any linear relationship between moisture and the temperature.
- Most of the points lie around the temperature range of 20-30.
- The linear regression model cannot be used here,

Q3(a).

correlation coefficient between rain and temperature is **-0.10877805105737656**

correlation coefficient between rain and humidity is **-0.4344566152391602**

correlation coefficient between rain and pressure is **0.07070976966275146**

correlation coefficient between rain and lightavgw/o0 is **0.5269321188362847**

correlation coefficient between rain and lightmax is **0.31251168706558863**

correlation coefficient between rain and moisture is **0.42647615378681775**

- The correlation coefficient describes the strength of relationship between two variables.
- The correlation coefficient between rain and temperature, rain and humidity is found to be negative. The negative values of correlation coefficient states the extent to which two variables move in opposite directions. For ex:- an increase in value of rain is associated with the decrease in the value of humidity.
- A positive value of the correlation coefficient signifies that both the variables move in the same direction.
- The relation between rain and lightavgw/o0 is very strong here as compared to others whereas the relation between the rain and the pressure is very weak.

Q3(b).

correlation coefficient between temperature and humidity is **0.40114490623630455**

correlation coefficient between temperature and pressure is **-0.18119712910750438**

correlation coefficient between temperature and rain is **-0.10877805105737656**

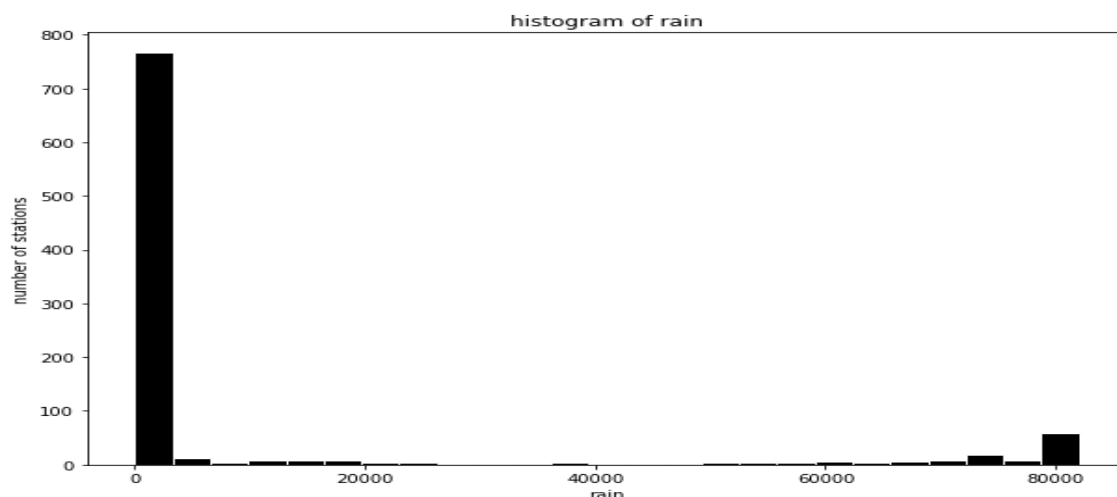
correlation coefficient between temperature and lightavgw/o0 is **-0.18120800584572488**

correlation coefficient between temperature and lightmax is **-0.14572914001778584**

correlation coefficient between temperature and moisture is **0.08057484324879301**

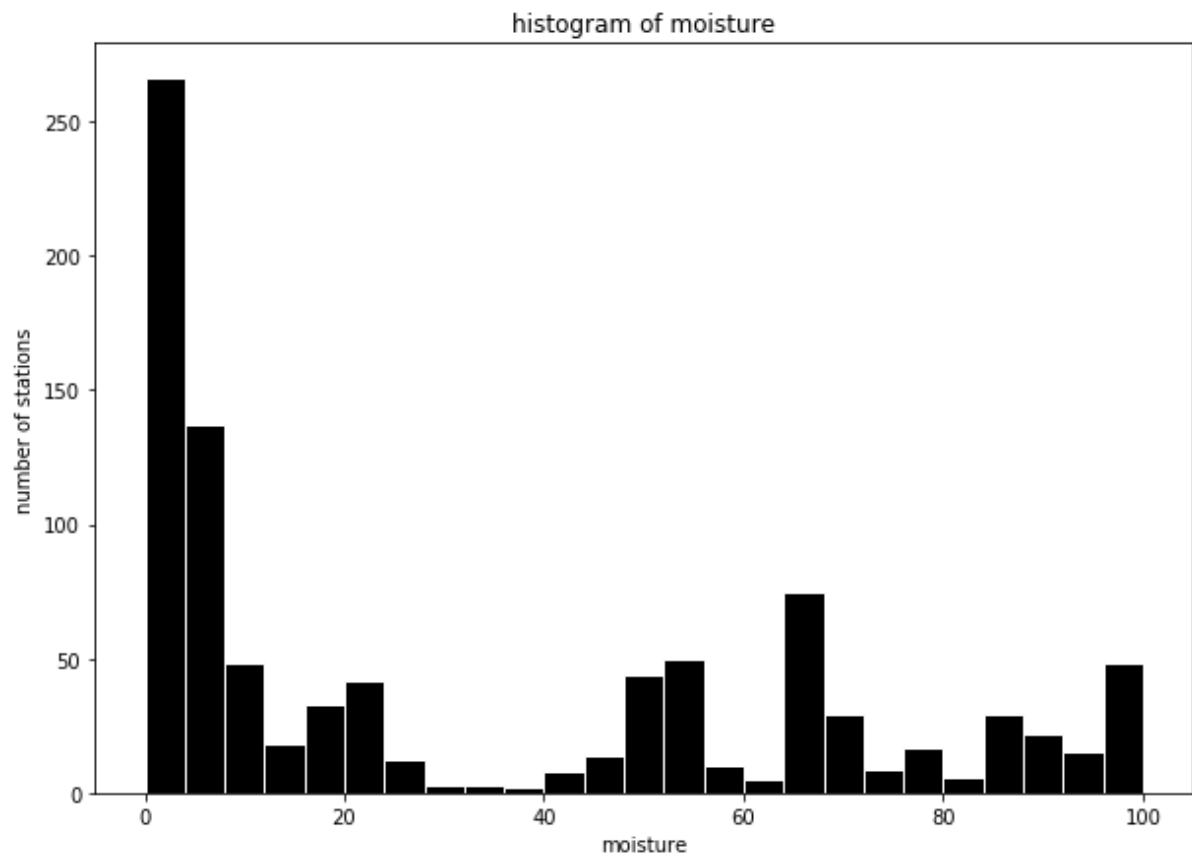
- The correlation coefficient between temperature and pressure, temperature and rain, temperature and lightavgw/o0, temperature and lightmax is negative.
- The correlation coefficient between temperature and humidity, temperature and pressure is positive.
- The relation between temperature and humidity is strong as compared to others whereas the relation between temperature and moisture is very weak.

Q4. Plotting the histogram of the attribute "rain" :-



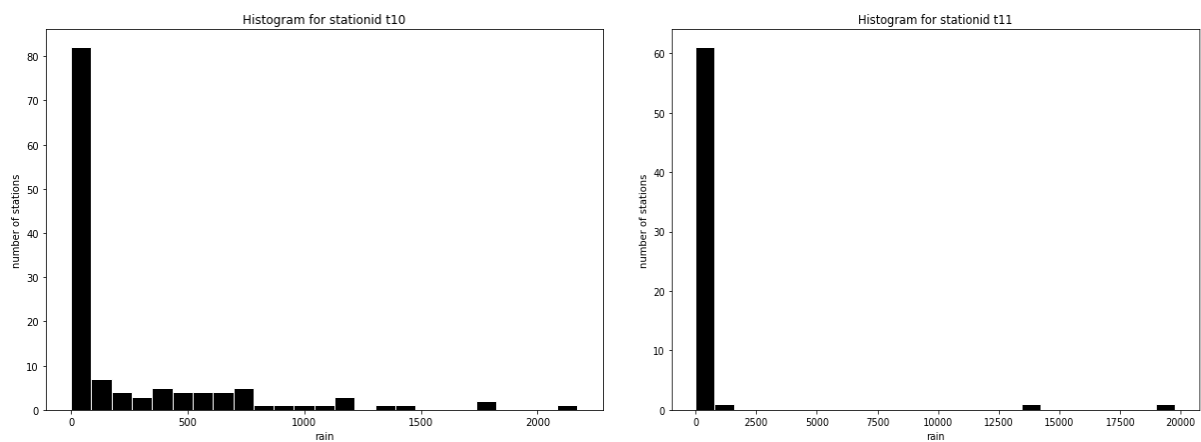
- Almost 85% of the station receives rainfall between 0-10000.
- The outliers can be clearly seen in the range 0-10000.
- The graph is skewed rightwards as there is a clear spike in the histogram bar therefore the data is noisy.

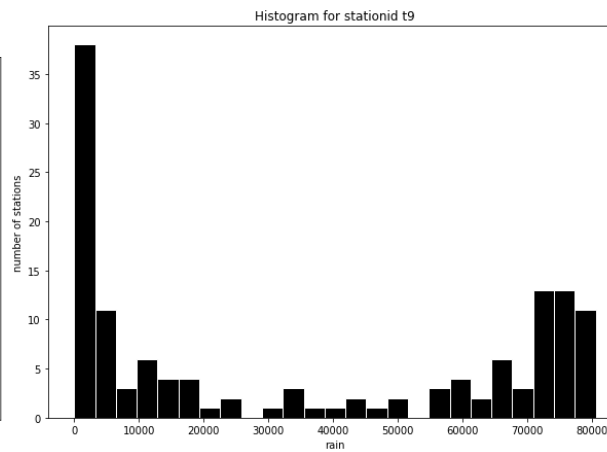
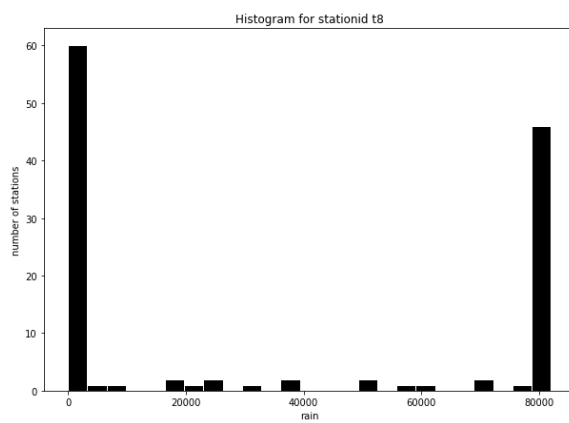
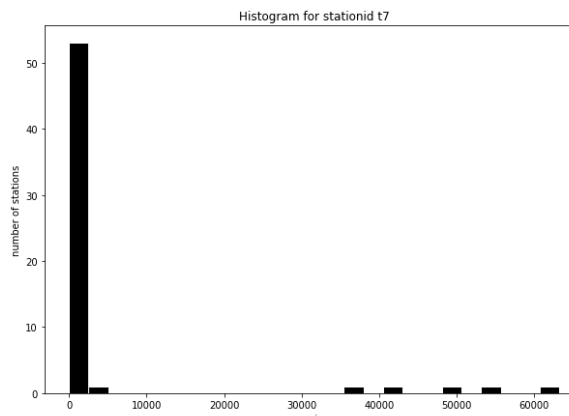
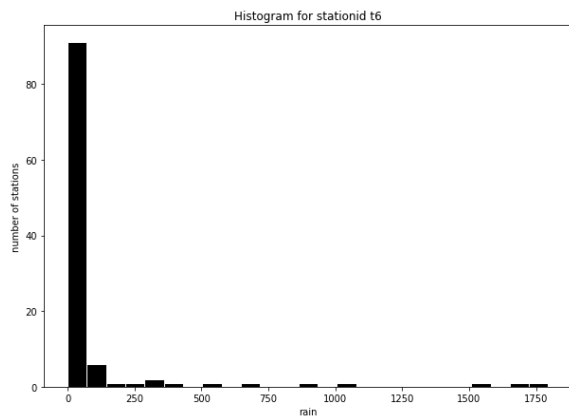
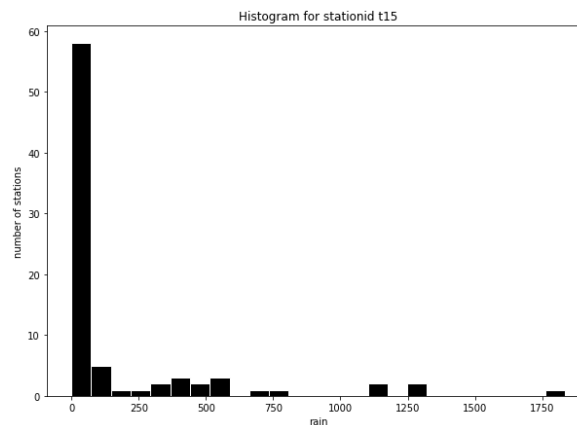
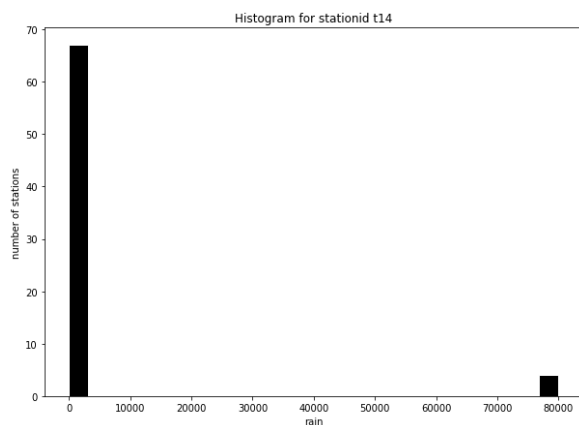
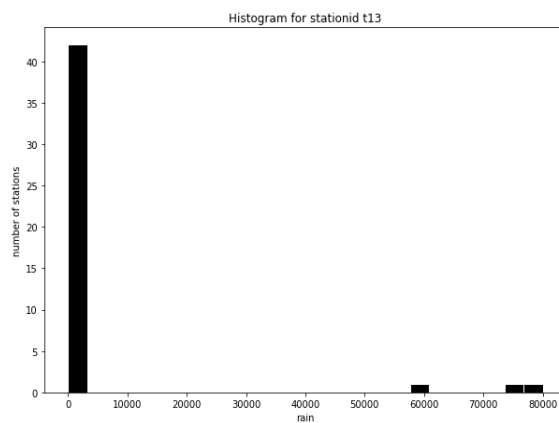
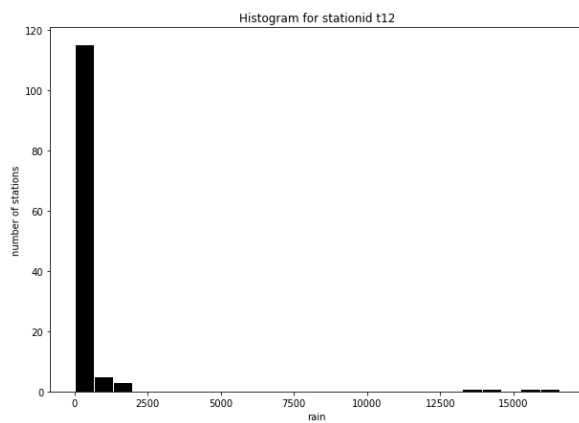
Plotting the histogram of the attribute “moisture” :-



- I have used 25 bins so that the graph looks better and easy to be observed.
- There are nearly 260 stations which experience little or no moisture.
- The graph is skewed rightwards therefore the mean is greater than the median.

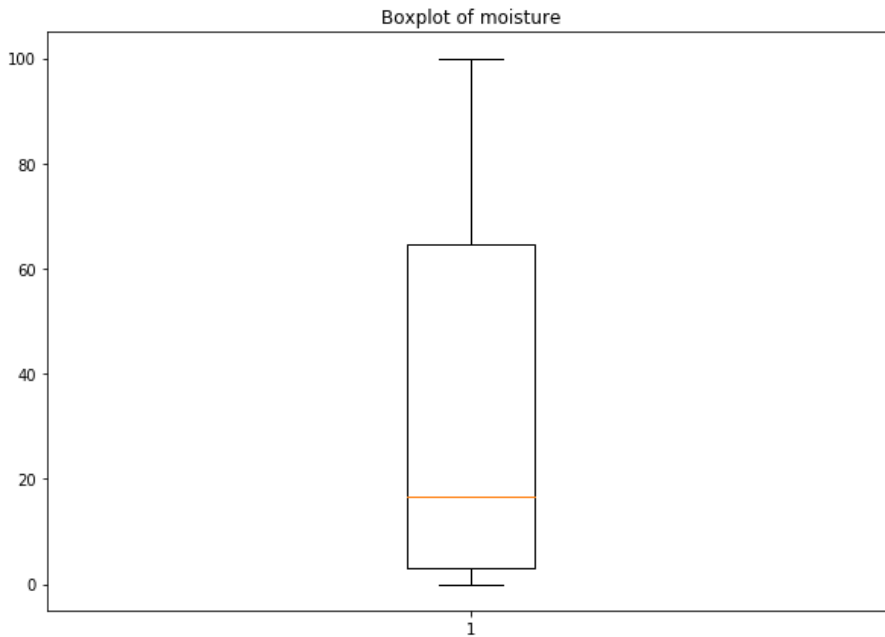
Q5. Plotting the histogram of rain for all of the 10 stations : -





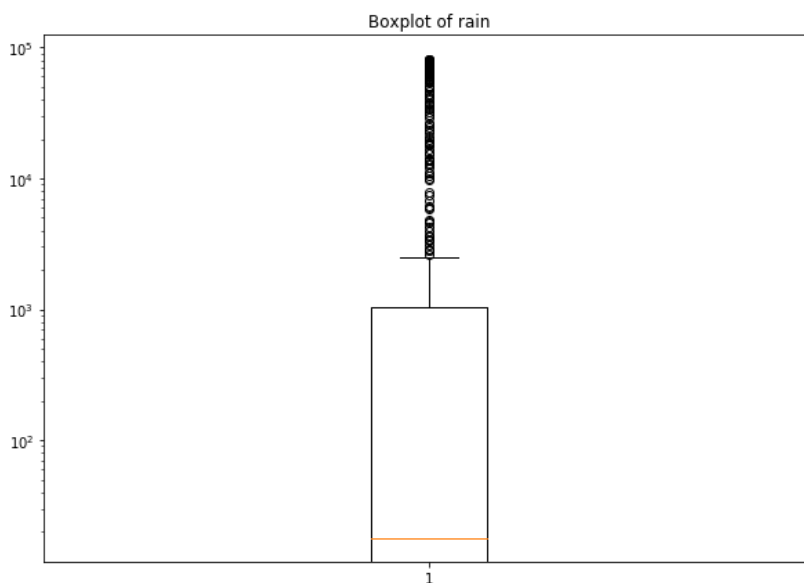
- The histograms with station ID t6, t7, t10, t11, t12, t13, t14, t15 are all skewed rightwards which represents that the data is very noisy.
- There is a clear spike in all the histograms for zero or no rainfall.
- For station ID t8, there is a spike in the bar at 0 rain and around 8000 rain.
- The outliers can be seen clearly in all the histograms.

Q6. Boxplot of “moisture” :-



- The median of the data is close to the first quartile.
- The data is not noisy due to the absence of outliers.
- The interquartile range is around 60.

Boxplot of “rain” :-



- The outliers can be seen clearly in the boxplot which represents that the data is very noisy.
- To improve the aesthetics of the boxplot, I have used `plt.yscale('log')`.