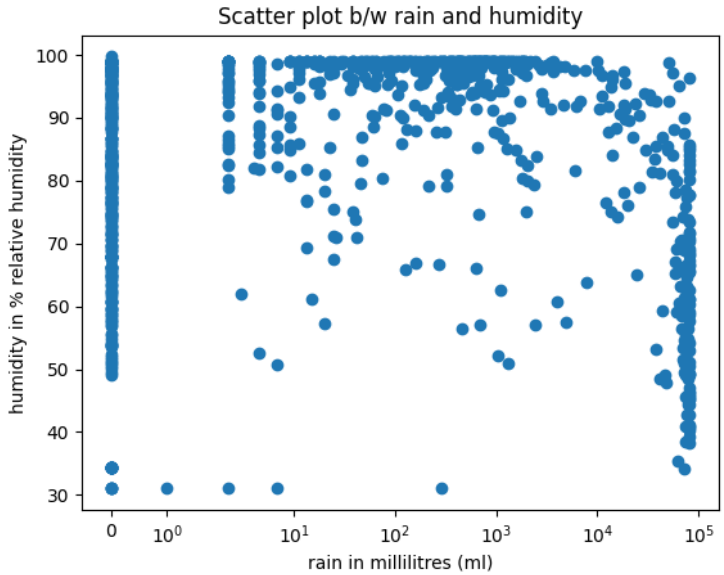
IC272-Lab1-Report

Name : Gajraj Singh Chouhan
Roll No. B19130

- **Q1**

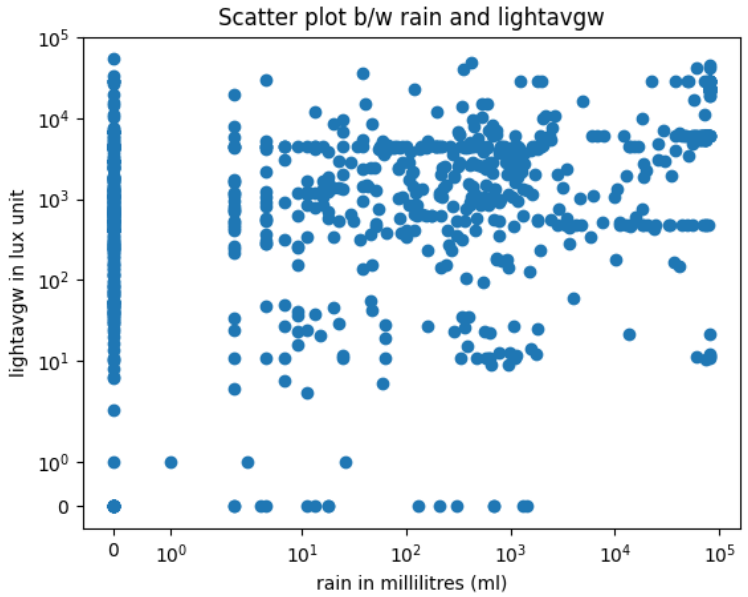
attribute	mean	median	mode	minimum value	maximum value	standard deviation
temperature	21.214	22.272	12.727	7.672	31.375	4.353
humidity	83.479	91.380	99.0	31.0	99.72	18.200
pressure	1009.008	1014.677	789.392	452.097	1079.162	46.955
rain	10701.538	18.0	0.0	0.0	82037.25	24839.102
lightavgw	4438.428	1656.88	4488.910	0.0	54612.0	7569.154
lightmax	21788.623	6634.0	4000	2259	54612	22053.315
moisture	32.386	16.704	0.0	0.0	100.0	33.635

- **Q2. Scatter Plots:**
  - ***Rain and Humidity***

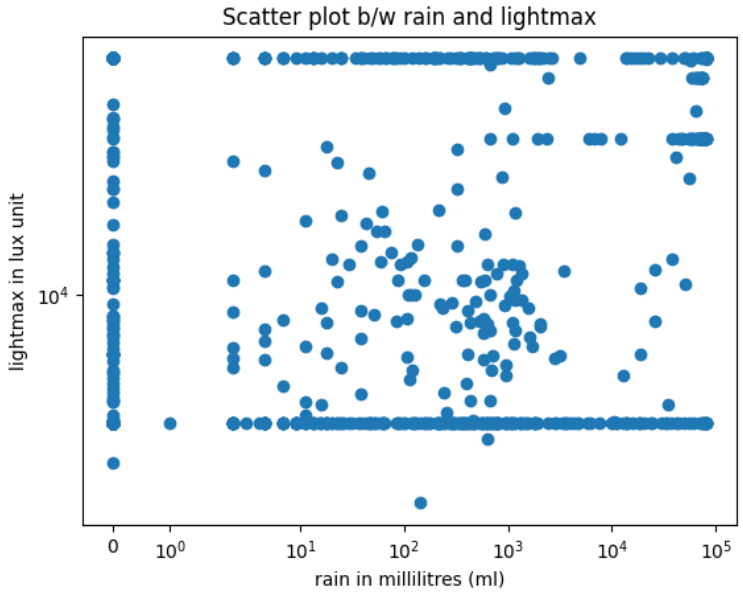


  - Initially when the amount of rain is low (around 10-4000 ml) the humidity is high (>=95%), but as the amount of rain increases the humidity decreases. We can see the shift around 10^4 millilitres of rain, there is a decrease in humidity.
  - We can infer that the correlation between these two attributes would be negative, due to decrease in humidity while increasing the rain.

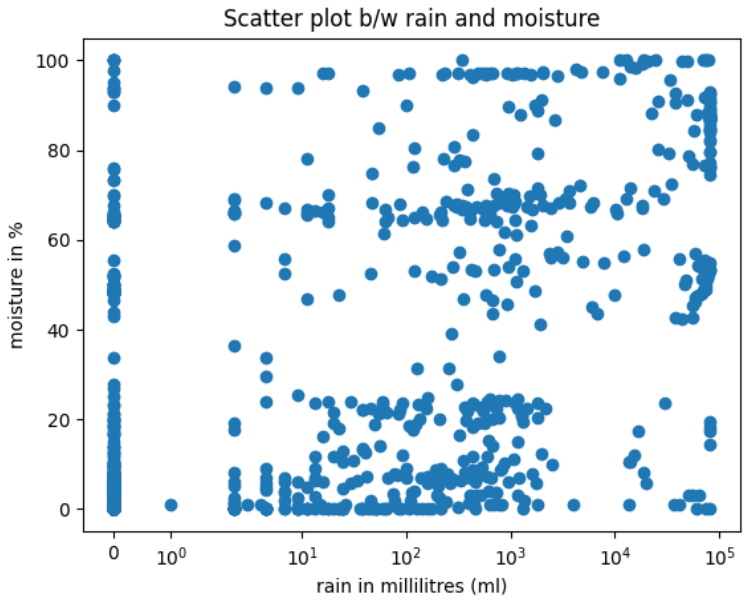
- The correlation comes out as -0.434917. This means that they have a weak linear relationship.
- ***Rain and lightavgw***



- Lightavgw is increasing as the rain is increasing.
- The correlation should be positive.
- The linear relationship would be a weak one, due to high variance of light for low amount of rainfall.
- The correlation comes out be 0.52
- ***Rain and lightmax***

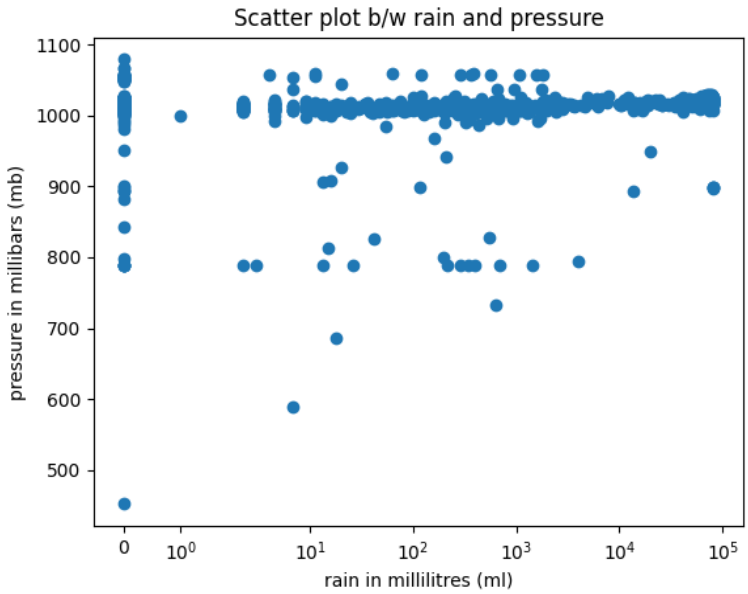


- The lightmax is remaining constant at both high and low values (mostly) across the x axis.
- There isn't a strong linear relationship as the y axis value mostly has a horizontal line over x axis.
- The correlation is -0.14.
- ***Rain and moisture***

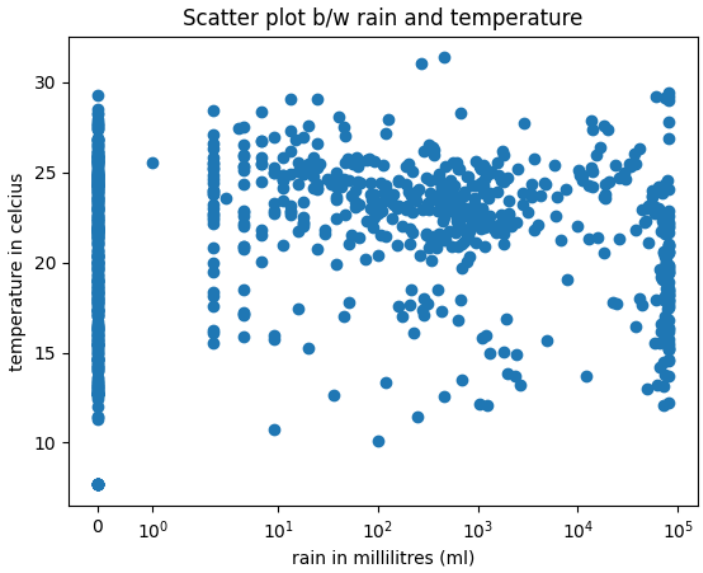


- As the amount of rainfall is increasing, after 10^2ml the moisture starts to shift upward to higher values. Initially when rain was $10^2 - 10^3$ ml, moisture was low (<=20%), but then it increased to upto >=60%.
- We can infer there should be a positive correlation between the two attributes.
- According to the calculations the coefficient is 0.426928.

- ***Rain and pressure***

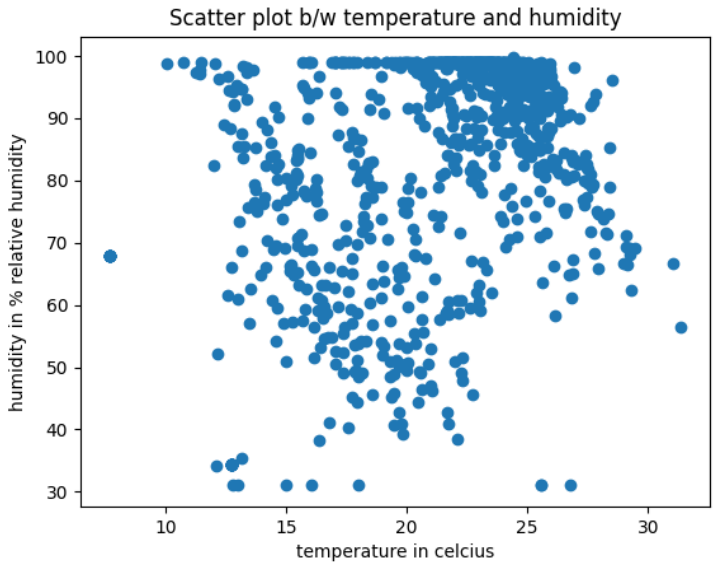


- Pressure initially when the rain is 0ml is ranging from 800-1100 mb.
- The pressure doesn't seem to change much when the amount of rain is starting to increase (>10ml), though sometimes it has changed, most of the time it has remained around 1000 mb.
- Since there isn't much change of pressure the attributes should be having no relationship. There the correlation should be very close to 0.
- The correlation comes out to be 0.070785 which is very low indeed.
- ***Rain and temperature***



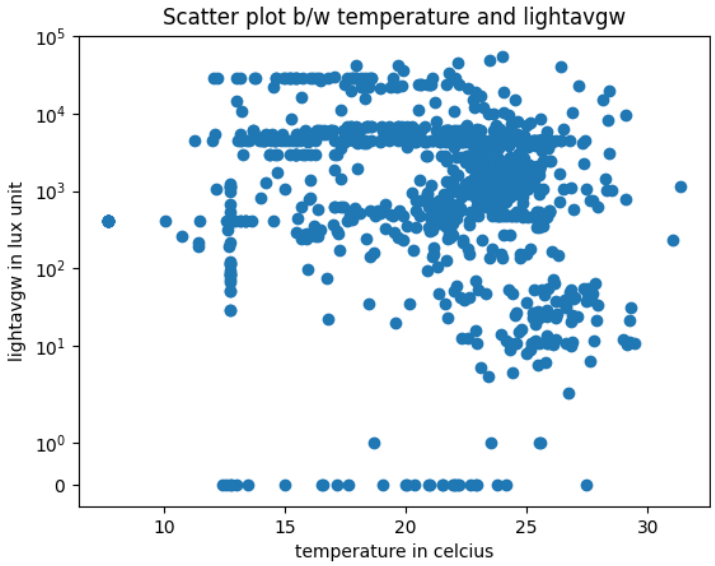
- Temperature is decreasing from 25 upto ~10C as the rain increases upto ~82,000.

- For 0C, the temperature is spread across 10C-30C.
- They should have a negative correlation due to decrease in temperature.
- But they have a very weak relationship due to high spread of temperature for large amounts of rain.
- The correlation comes out to be -0.108893.
- ***Temperature and humidity***

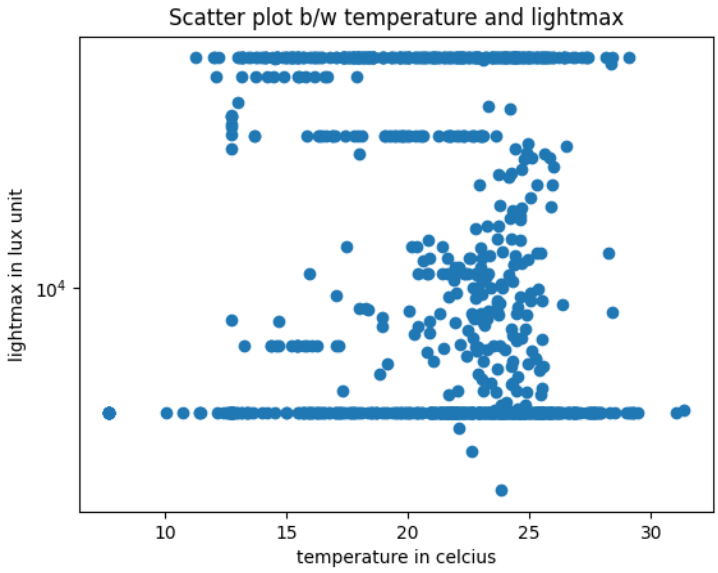


- As the temperature increases, the humidity is increasing and goes upto 100% around 25C and then decreases when the temperature increases past 25C.
- Accordingly the correlation coefficient should be a positive number.
- It comes out to be 0.401570.
- They have a weak relationship.

- ***Temperature and lightavgw***

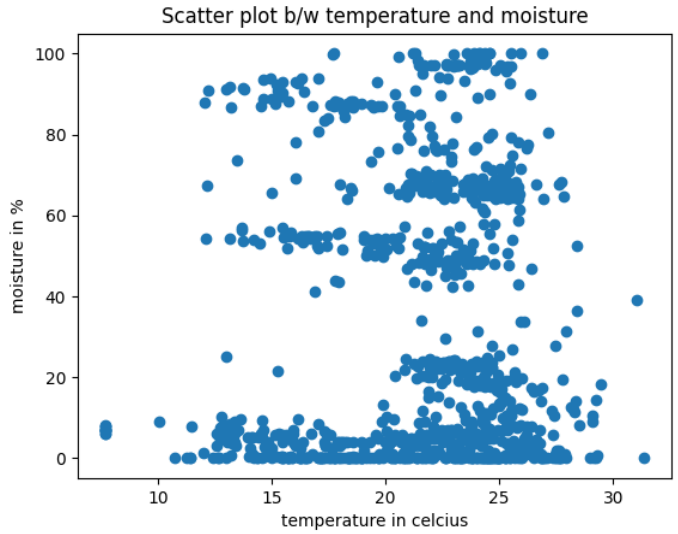


- Lightavgw is decreasing with increase in the temperature.
- The correlation coefficient would be negative.
- Due to high variance we can say the correlation coefficient would show a weak relationship.
- The correlation coefficient is -0.181400.
- ***Temperature and lightmax***

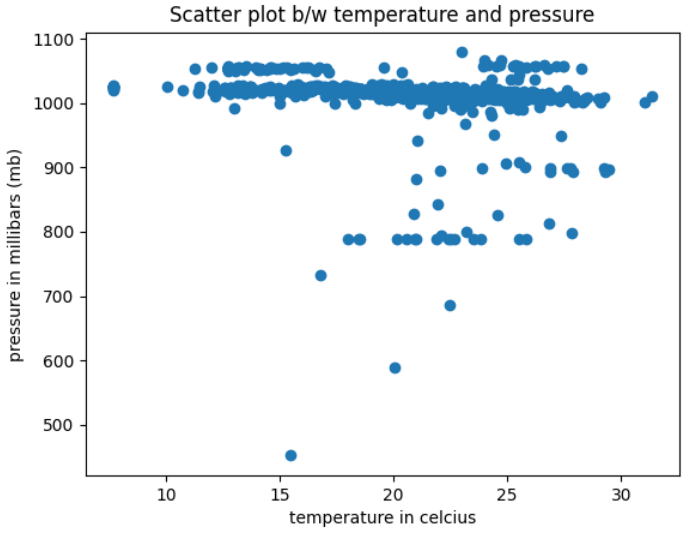


- This graph consists of horizontal lines for different values of y and are spread across the x axis.
- Although with increase of temperature there is some cluster of increased light in y axis.

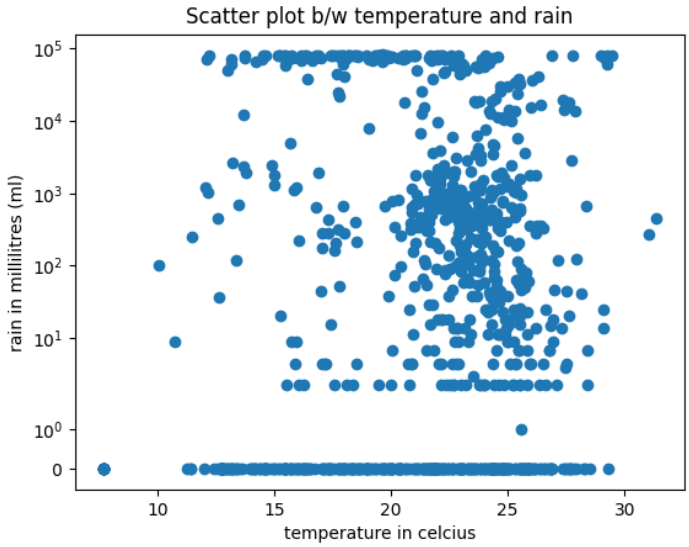
- We can say y axis doesn't depend strongly on temperature as it is taking some values across the temperature.
- The correlation coefficient comes out to be 0.312.
- We can say they don't have a strong linear relationship.
- ***Temperature and moisture***



- Initially when the temperature is low around 10C, the moisture is around 0-10% mostly. In some cases the moisture is also high (around 60% and 80%). As the temperature is increasing, the higher moisture content is also increasing, while low moisture is still remaining there.
- Their relationship doesn't seem to be linear, as initially the amount of high moisture is lower.
- The correlation comes out to be 0.08
- ***Temperature and pressure***



- The pressure remains constant around 1000 mb as the temperature is increasing. Though there is a small decrease in pressure as the temperature is increasing.
- They could have a very weak/negligible linear relationship as the amount of pressure mostly remains unchanged.
- Correlation comes out to -0.18. They show very weak linear relationships.
- ***Temperature and rain***



- As the temperature is increasing, the amount of rainfall is decreasing.
- However temperature also increases if the rainfall has been very low <10 ml.
- We can say they would have a negative correlation due to decrease of rain when the temperature is increasing.
- Though the linear relationship would be very weak as decrease in rain is only for some range of temperature mostly.
- The correlation comes out to be -0.10.

- **Q3. Correlation Coefficients**

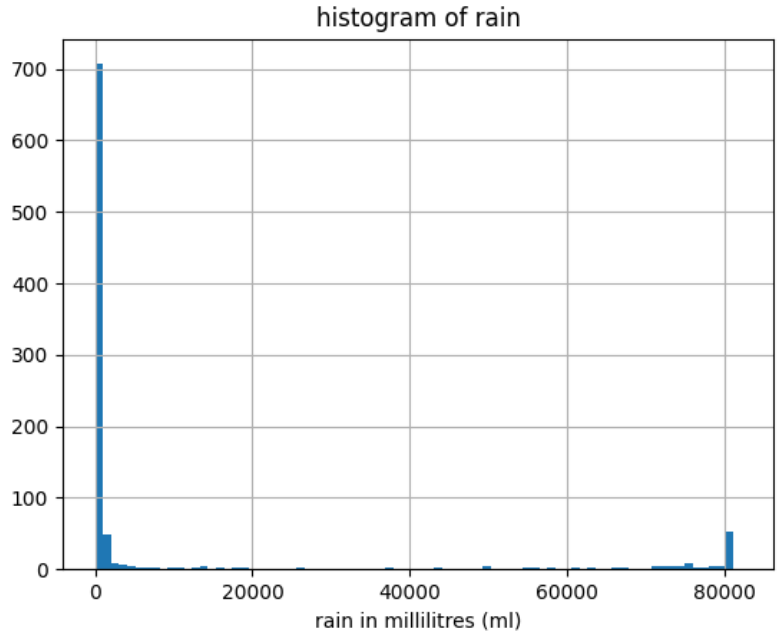
	rain	temperature
temperature	-0.108893	1.000000
humidity	-0.434917	0.401570
pressure	0.070785	-0.181389
rain	1.000000	-0.108893
lightavgw	0.527490	-0.181400
lightmax	0.312843	-0.145884

moisture	0.426928	0.080660

Observations :

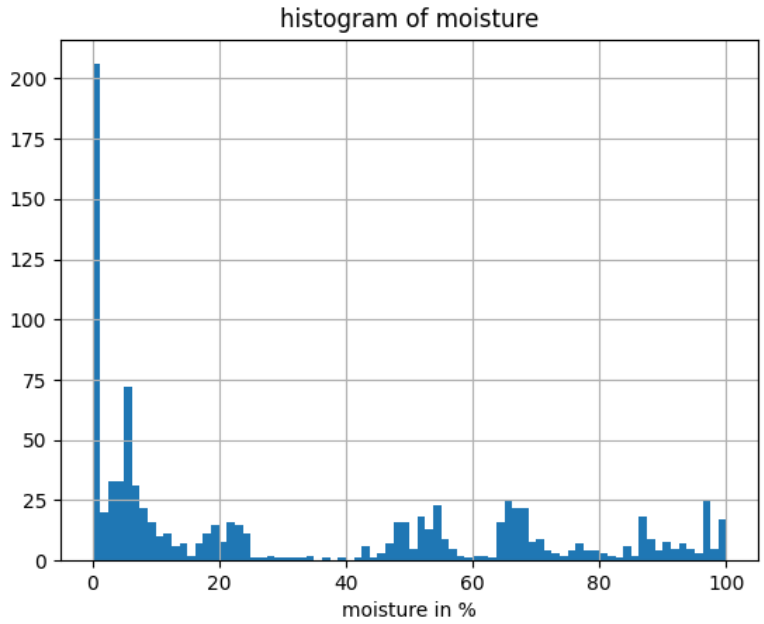
- “humidity”, “lightavgw”, “lightmax”, “moisture” show a weak correlation with “rain”.
- “temperature” and “pressure” show negligible correlation with “rain”.
- “Humidity” shows a weak correlation with “temperature”.
- The rest of attributes are showing negligible correlation with “temperature”.

- We can see as the spread (variance, standard deviation)of a variable increases the correlation will decrease.
- We can’t tell the linear relationship between the variables.
- We may have to use the spearman correlation coefficient to check for some kind of non-linear relationship.
- **Q4a. Histogram of Rain**

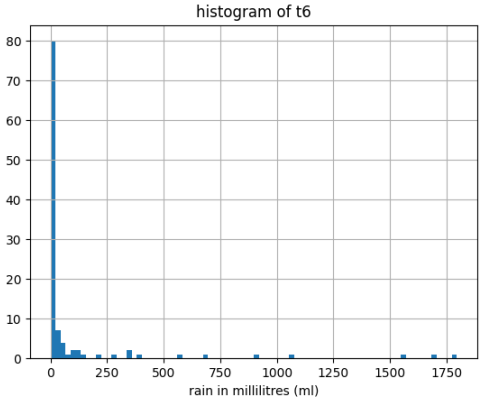


- Rainfall of low amounts has most occurrences in the histogram.
- Mean > Median > Mode
- The data is right skewed or has positive skewness.

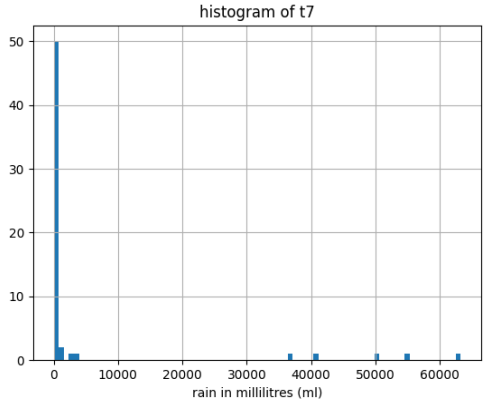
- **Q4b. Histogram of moisture**



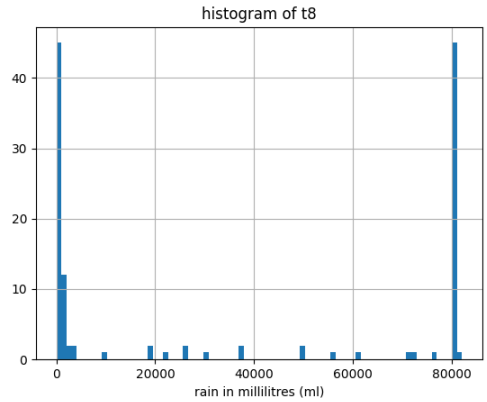
  - Moisture of 0% is having most occurence in the histogram.
  - Then the occurrence starts to decrease as moisture content is increasing.
  - Mean > Median > Mode
  - We can say data is right skewed or has positive skewness.
- **Q5. Histogram of Stations**
  - Histogram of t6



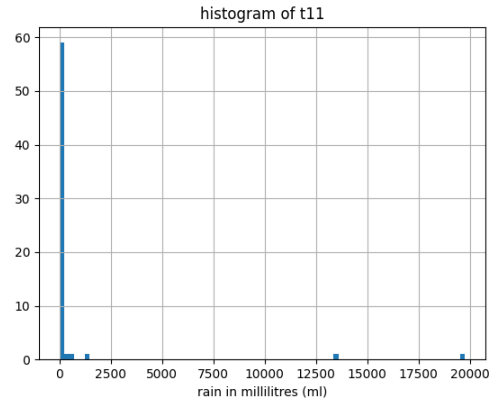
  - Histogram of t7



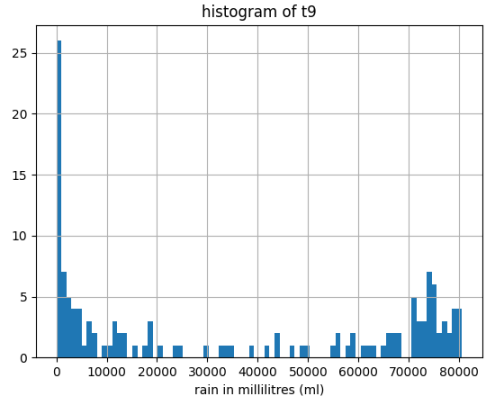
- Histogram of t8



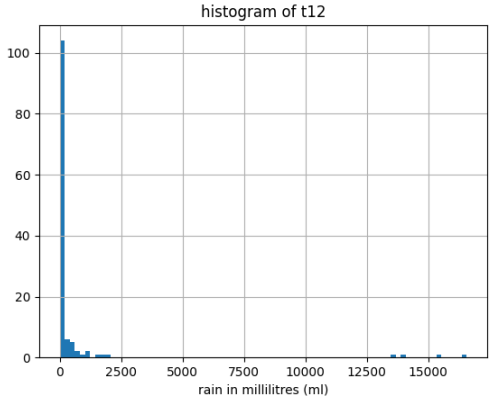
- Histogram of t11



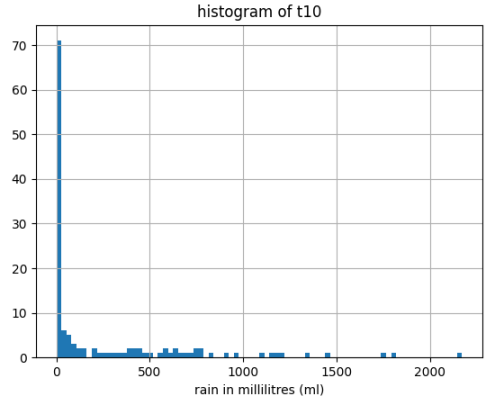
- Histogram of t9



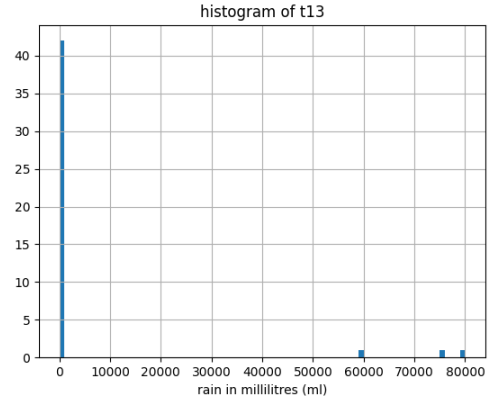
- Histogram of t12



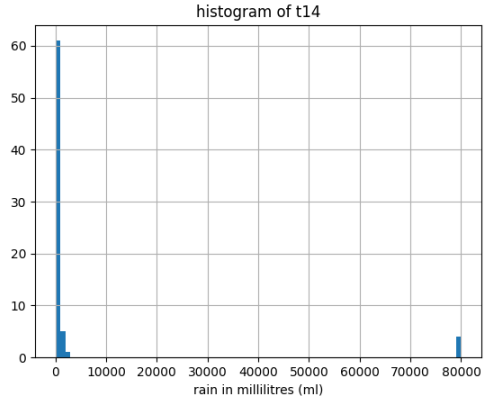
- Histogram of t10



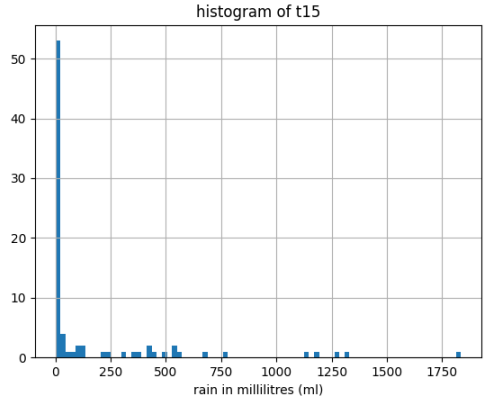
- Histogram of t13



- Histogram of t14

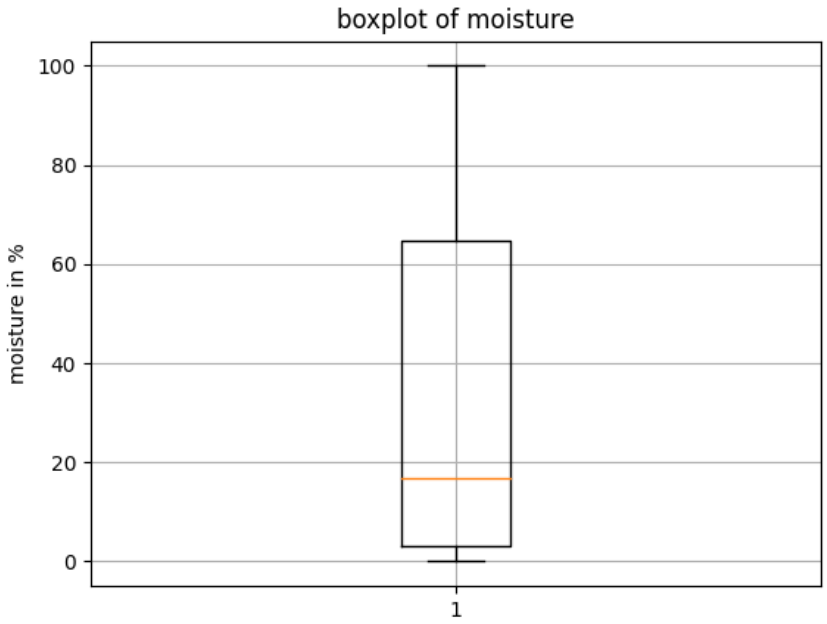


- Histogram of t15



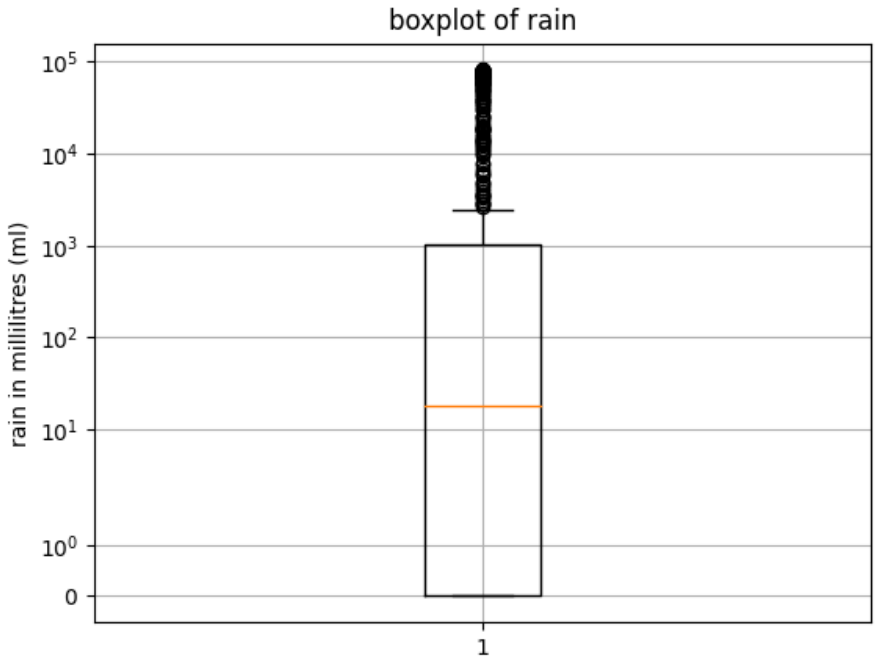
- All of these plots are showing positive skewness.
- We can see rainfall of 0ml has the maximum occurrence in every graph. Except station t9 and t8, rainfall of large amounts has very low occurrence.
- In case of station t8, it receives high rainfall as well as low rainfall equally.
- In case of station t9, it receives higher rainfall (more than 10,000 ml) than other stations though low rainfall is still dominating.
- Some of the stations receiving high rainfall are t14, t13, t9, t8, t7.
- Rest of them receive less rainfall.
-

- **Q6a. Box-Plot of moisture**



  - Largest Observation = 100
  - Smallest Observation = 0
  - Q1 = 2.993, Q2 = 16.704, Q3 = 64.6715
  - IQR = Q3 - Q1 = 61.6785
  - Since no element is larger/smaller than their respective quartiles, there are no outlier in the plot.

- **Q6b. Box-Plot of rain**



  - Smallest Observation = 0.0
  - Largest Observation = 2470.5
  - Q1, Q2, Q3 = 0.0, 18.0, 1044.0
  - IQR = Q3 - Q1 = 1044
  - Since a lot of points are larger than 1.5 x IQR, they are to be considered as outliers.