IC272: Lab-1 Report

Jai Prakash Yadav • +91-9306871378 • b19247@students.iitmandi.ac.in

Inference from Task 1

Used the inbuilt function to calculate mean, median, mode, maximum, minimum, standard deviation of the given data set.

Here are the results obtained -

	Temperature	Humidity	Pressure	Rain	Lightavgw/o0	Lightmax	Moisture
Mean	21.215	83.480	1009.009	10701.538	4438.428	21788.623	32.386
Median	22.273	91.381	1014.678	18.000	1656.880	6634.000	16.704
Mode	12.727	99.000	789.393	0.000	4488.910	4000.000	0.000
Maximum	31.375	99.720	1079.162	82037.750	54612.000	54612.000	100.000
Minimum	7.673	31.000	452.098	0.000	0.000	2259.000	0.000
Standard deviation	4.356	18.210	46.980	24852.255	7573.163	22064.993	33.653

Inference from task 2 and 3.

(Task 2 & 3-To obtain Scatter plot and Correlation coefficient between different data set)

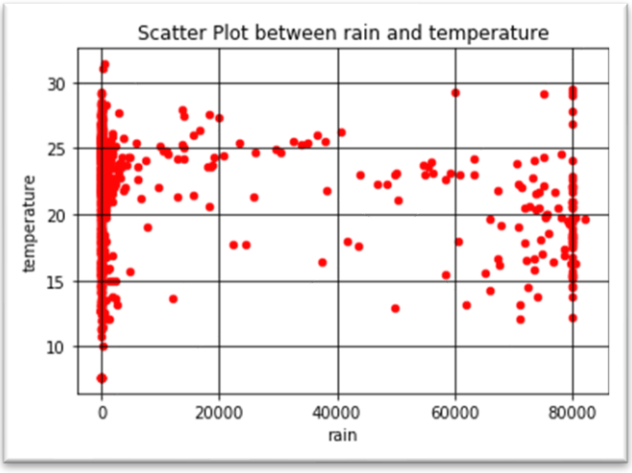
1. *Correlation and scatter plot between rain and different data sets.*

   (a) Rain vs Temperature
   Correlation coefficient: -0.108893

   From the correlation coefficient value between rain and temperature, we cannot draw any solid result because the value is too close to 0.

   It is also clear from the scatter plot that the points are scatter all over.

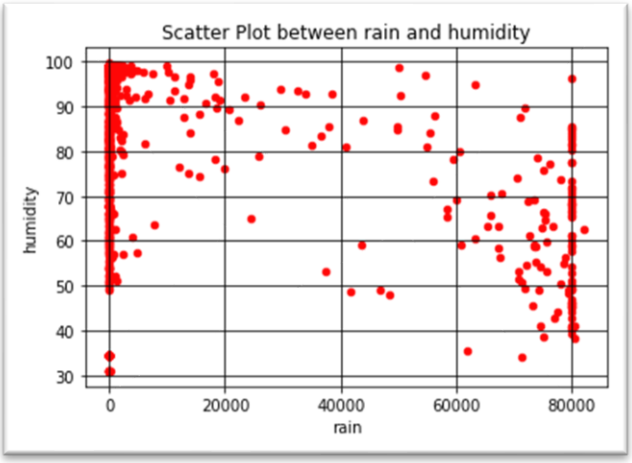
   Hence the correlation between rain and temperature is not very significant



(b) Rain vs Humidity
Correlation coefficient: -0.434917

From the correlation coefficient value, it can be inferred that when the humidity is high there is less chances of rain.

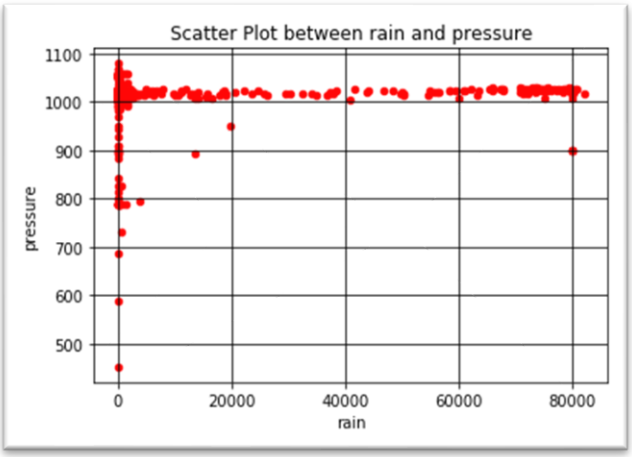
It is also clearly visible the from the scatter plot that there is notable relation between Rain and Humidity



(c) Rain vs Pressure
Correlation coefficient: 0.0707847

As the correlation coefficient value is very close to 0, there is no relation between rain and Pressure.

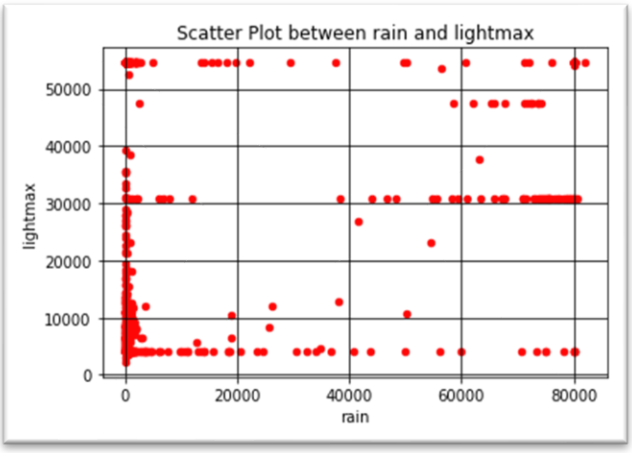
As the scatter plot remain almost constant for all values, we can infer that there is no significant relation between these two.



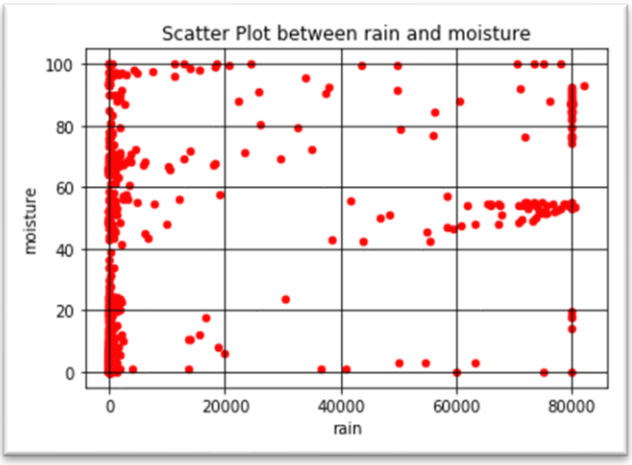
(d) Rain vs Lightmax
Correlation coefficient: 0.312843

As the correlation coefficient value is neither to close to 0 or 1, hence we can not draw any solid result.

Also from the scatter plot the point are scatter all over, but we can tell that for low intensity there is less rain, and same result can also infer from correlation coefficient value.



(e) Rain vs Moisture
Correlation coefficient: 0.426928

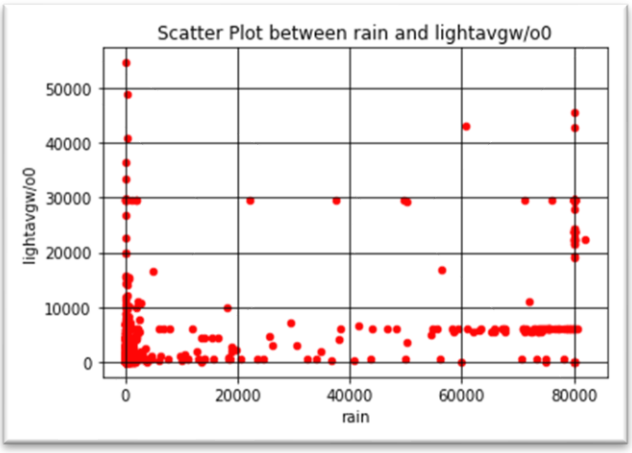


As the correlation coefficient value is almost in middle of 0 and 1, we can infer that almost half the time when the rain is high the moisture is high.

Same result is also inferred from the plot as high moisture there is high chances of rain.

Hence there is remarkable relation between Rain and Moisture.

(f) Rain vs Lightavgw/o0
Correlation coefficient: 0.527490

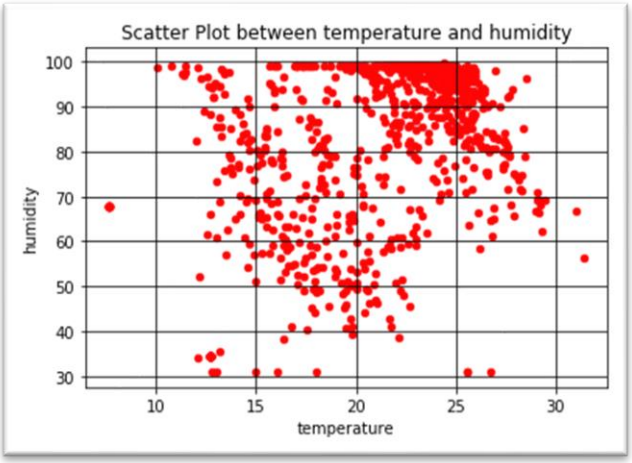


As the correlation value is very close to 0.5, we can infer that there is solid relation between rain and lightavgw/o0.

But from the graph the as we see that points are condensed around the corner hence nothing can be said.

2. *Correlation and scatter plot between temperature and different data sets.*

(a) Temperature vs Humidity
Correlation coefficient: 0.40157



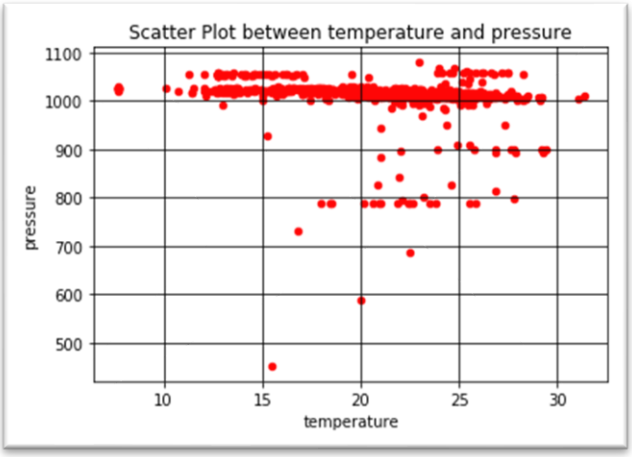
As the corelation coefficient value is near to 0.5 hence there is fairly a relation between two. As temperature increase the humidity also increases.

From the scatter plot , most the points are concentrated in the temperature range between 20 -30 , where the humidity increases the most with temperature.

(b) Temperature vs Pressure
Correlation coefficient: -0.181389

As the correlation coefficient is almost close to 0 hence no major conclusion can be drawn.

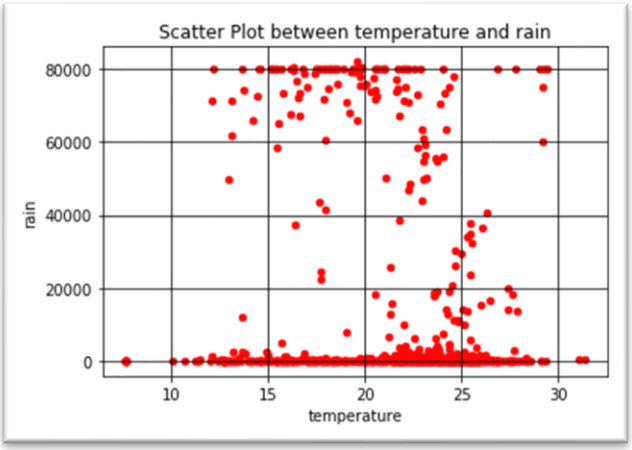
From the scatter plot it is clear that the relation between two remain almost constant for the complete range of values. Hence pressure is very much less corelated.



(c) Temperature vs Rain
Correlation coefficient: -0.108893

Due to very less corelation coefficient value we can infer that there is no significant relation between temperature and rain.

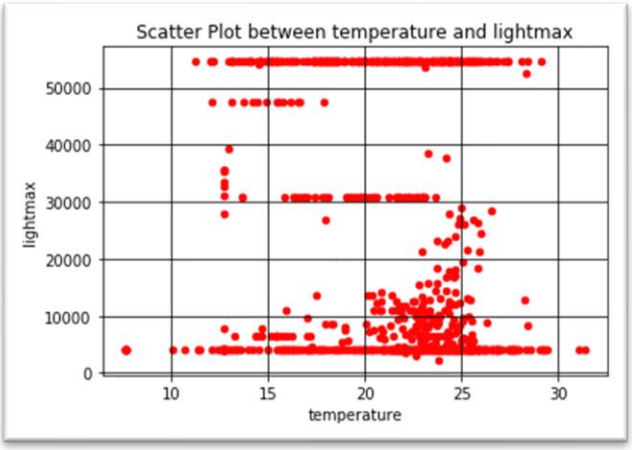
The same result also seen from the scatter plot as the points are varied all over.



(d) Temperature vs Lightmax
Correlation coefficient: -0.145884

From the correlation coefficient value between lightmax and temperature, we cannot draw any solid result because the value is too close to 0.

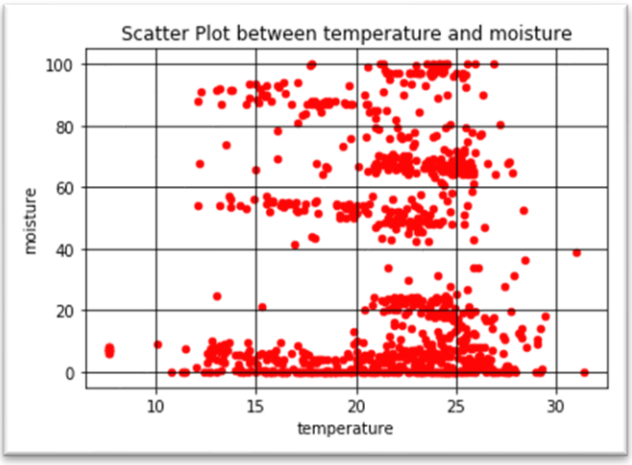
From the plot due to such a vast amount of data, the plot dots has been overlapped and no good relation could be found.



(e) Temperature vs Moisture
Correlation coefficient: 0.0806602

The corelation coefficient is very much less as value almost tends to 0. There is very insignificant relation between two.

From the scatter plot, the points are spread all over the plot. Hence no relation can be drawn.

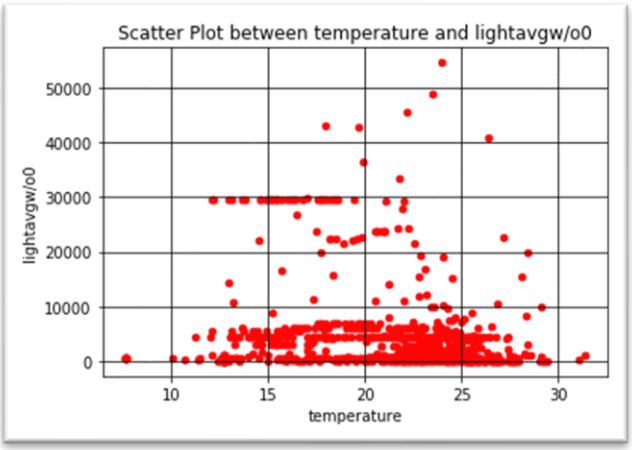


(f) Temperature vs Lightavg/o0
Correlation coefficient: -0.1814

From the correlation coefficient value between two, we cannot draw any solid result because the value is too close to 0.

It is also clear from the scatter plot that the points are scatter all over.

This data can be properly evaluated in combination with other attributes.

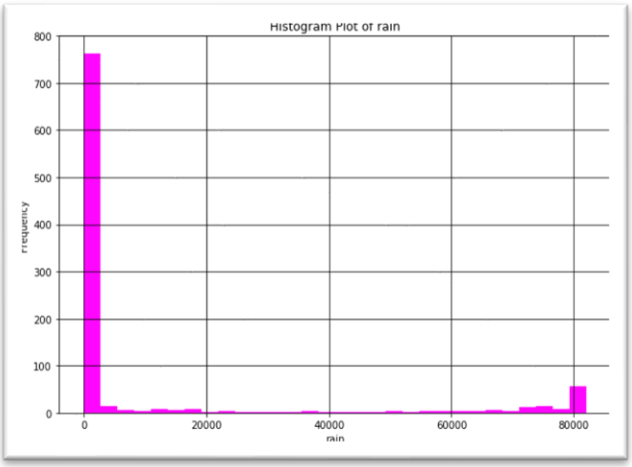


Inference from task 4

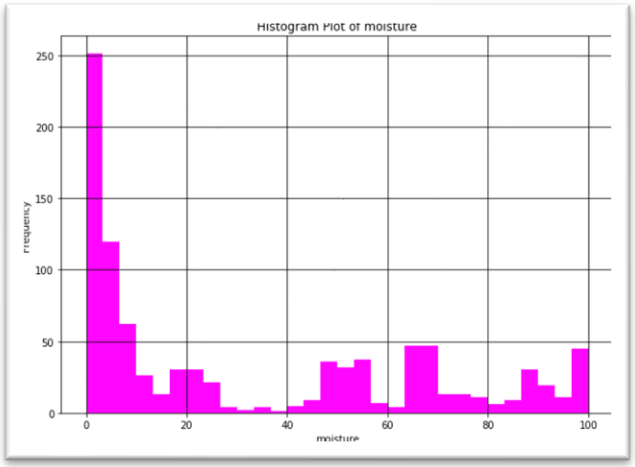
(task 4 – to obtain histogram for both Rain and Moisture.)

(a) Histogram for Rain

From the graph it is inferred that for zero rainfall the frequency is very high. And for the larger rainfall is frequency is less. There is little frequency when the rain is near to 8000 mm, hence from all these points we conclude that the rain is seasonal with heavy rainfall in that season.



(b) Histogram for moisture



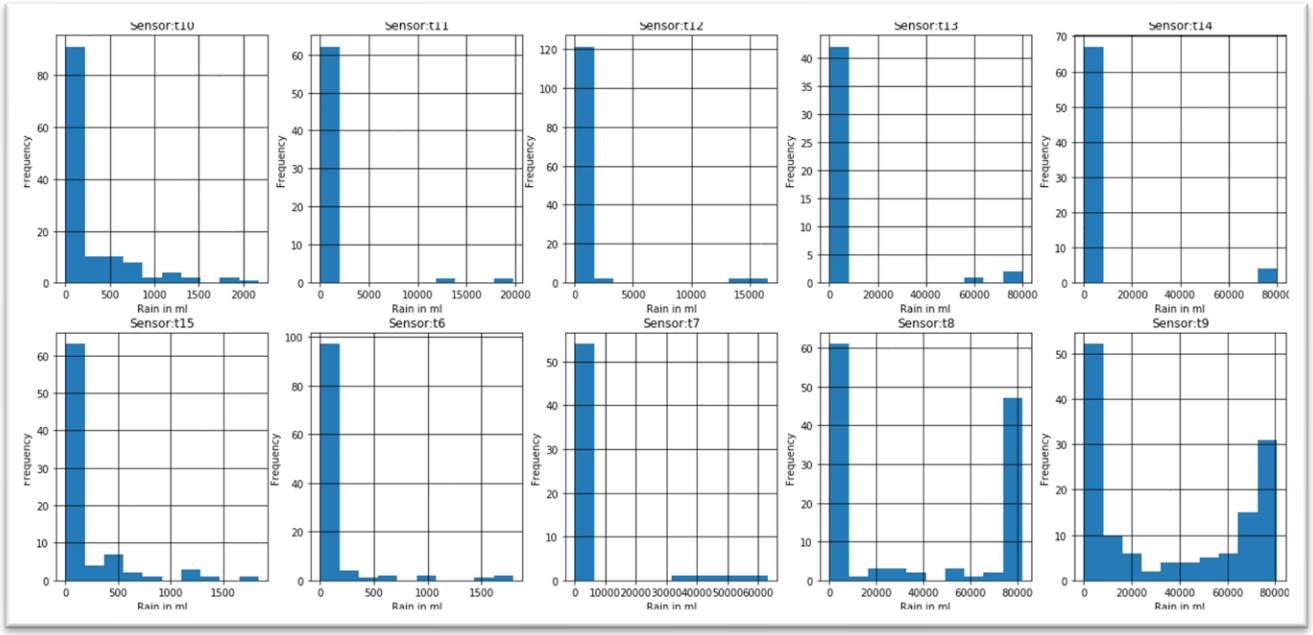
As from the plot we can infer that the frequency is very much high for less moisture content percentage.

And when the moisture percentage is high the frequency is not very significant.

Hence, we conclude that the soil predominantly remains dry.

Inference from task 5

(Task 5 - Rain Histograms grouped by stationid)



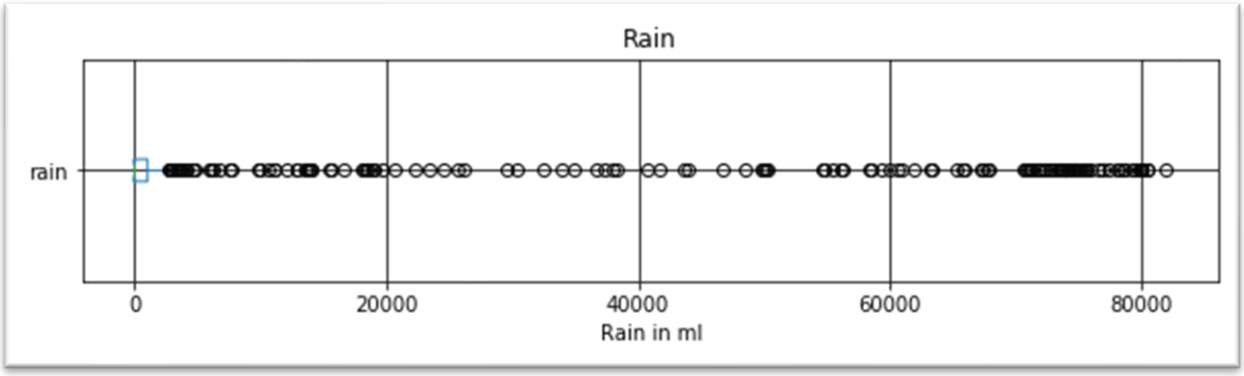
We can infer same result from all plots of all station id except T9. Rest of the stations have got their histogram concentrated on the left side only, with high frequencies of zero or low rainfall.

Station T9 histogram is spread all over the plot. From T9 station it is clear that there is quite significant rainfall because frequency is not diminishing with quantity in rainfall. The contrasting result for station id maybe due to its location at higher altitude.

Inference from task 6

(Task 6 - Obtain the boxplot for the attributes ‘rain’ and ‘moisture’)

(a) Boxplot for Rain



From the above boxplot for rain we can conclude that it is highly concentrated towards zero and same this result can also be verified from the histogram of rain for various station id where frequency is almost concentrated towards lower side. From these lots if outliers we can conclude that it is due to irregular and substantial rain.

(b) Boxplot for moisture

From the graph, it is evident that the median is nearly around 16 and this can be verified from the task 1 and also from the histogram where we can find that plot is almost condensed around it. In this graph we find no outliers which signifies that data was well distributed. The distribution for moisture also seems to be skewed towards 0. This also backs the high correlation value between rain and moisture.