IC272 – Data Science 3 *Aryan Garg*

REPORT *B19153*

Lab Assignment – 1 *+91-8219383122*

1.

```
A1. ***  Attributes of numerical data  ***

----------Mean----------
temperature : 21.214888105820105
humidity : 83.47993155555555
pressure : 1009.0087738022222
rain : 10701.53837037037
lightavgw/o0 : 4438.428453333333
lightmax : 21788.62328042328
moisture : 32.38605259259259

----------Median----------
temperature : 22.27273
humidity : 91.38095
pressure : 1014.6778320000001
rain : 18.0
lightavgw/o0 : 1656.88
lightmax : 6634.0
moisture : 16.7042
```

Figure 2: Mean and Median

Figure 1: Mode and Min values

```
----------Mode----------
temperature : 0     12.72727
dtype: float64
humidity : 0     99.0
dtype: float64
pressure : 0     789.392692
dtype: float64
rain : 0     0.0
dtype: float64
lightavgw/o0 : 0     4488.9103
dtype: float64
lightmax : 0     4000
dtype: int64
moisture : 0     0.0
dtype: float64

----------Min value----------
temperature : 7.6729
humidity : 31.0
pressure : 452.09788729999997
rain : 0.0
lightavgw/o0 : 0.0
lightmax : 2259
moisture : 0.0
```

```
----------Max value----------
temperature : 31.375
humidity : 99.72
pressure : 1079.162
rain : 82037.25
lightavgw/o0 : 54612.0
lightmax : 54612
moisture : 100.0

----------Standard Deviation----------
temperature : 4.355817940432199
humidity : 18.210064667980546
pressure : 46.98047725024858
rain : 24852.255288053453
lightavgw/o0 : 7573.162806138657
lightmax : 22064.993088694104
moisture : 33.653244650764854
```

Figure 3: Max values and standard deviation

2. a) INFERENCES: (Rain vs. Others)

i. There is a very little correlation (negative) between temperature and rain. As the temperature dips, the rain increases. See figure 4.

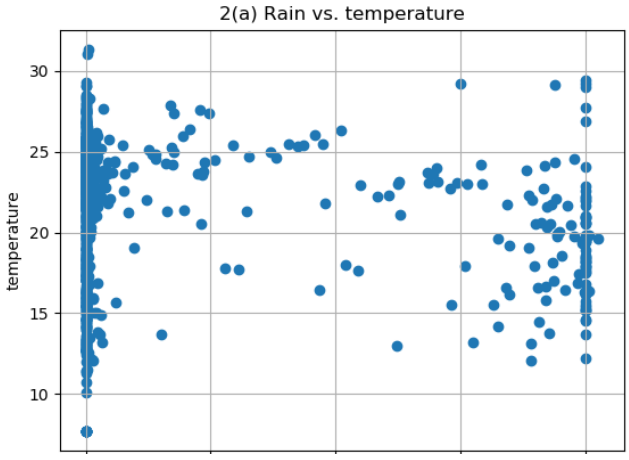

Figure 4

ii. Larger negative correlation between rain and humidity exists. As humidity decreases rain increases significantly. See figure 5.

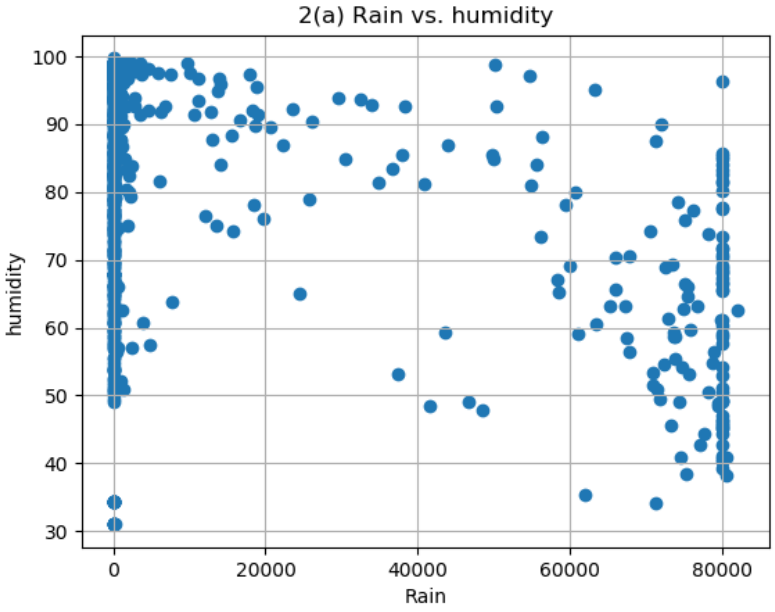

Figure 5

iii. There's seems to be no relation between air pressure and rain, which is valid because we know that air-pressure remains relatively constant unless there's air breakdown by lightning or very strong winds. See figure 6.

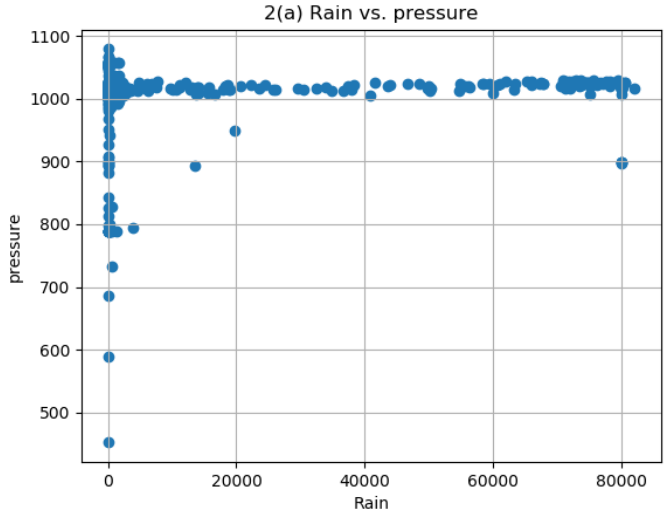

Figure 6

iv. It’s visually hard to tell the relation between light average and rain. The data might be polluted due to artificial lighting and rains at night as well. See figure 7.

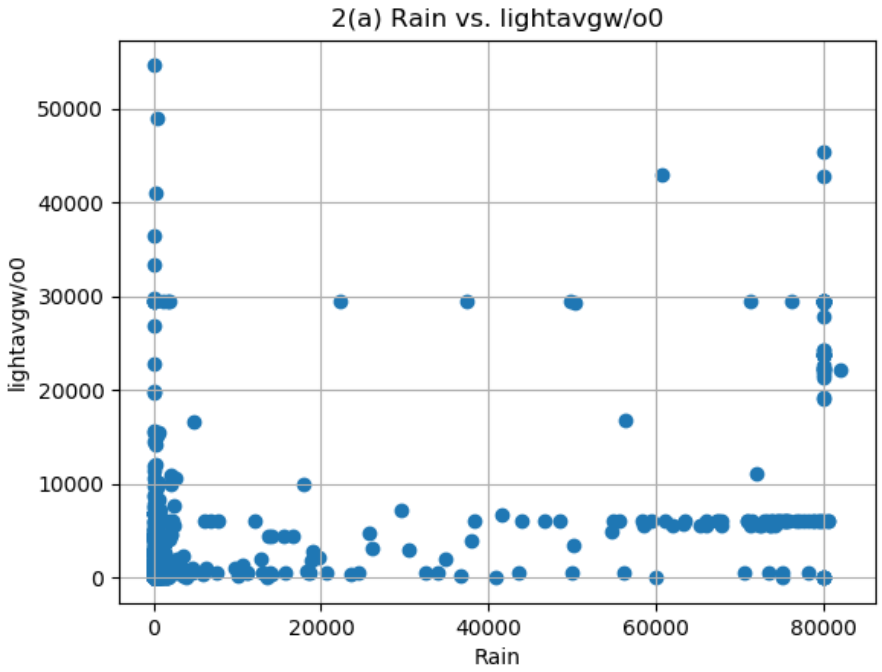

Figure 7

v. Same goes for rain and max light.

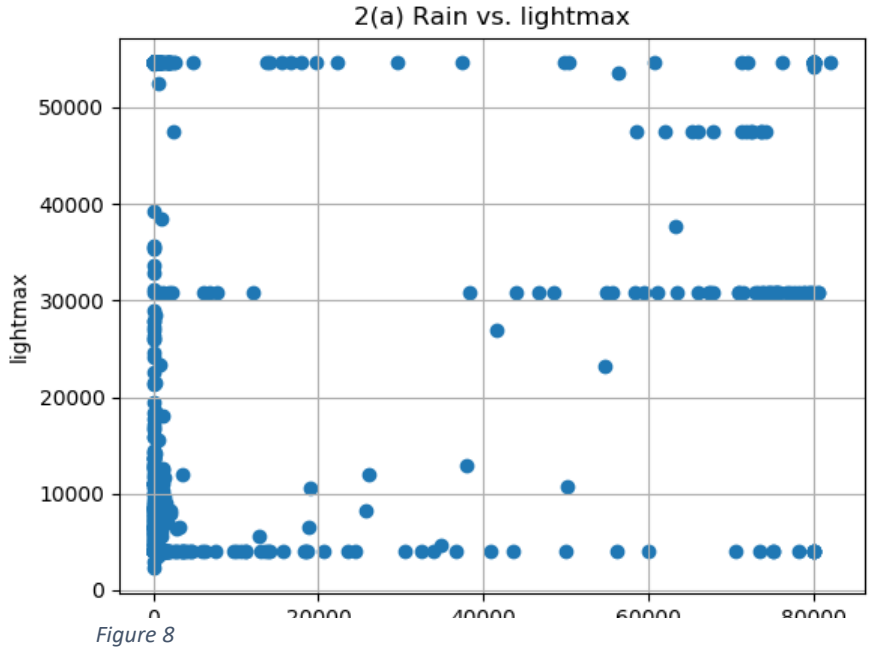

Figure 8

vi. Rain vs. Moisture is also very scattered due to water bodies nearby which might have a greater impact on the surroundings.

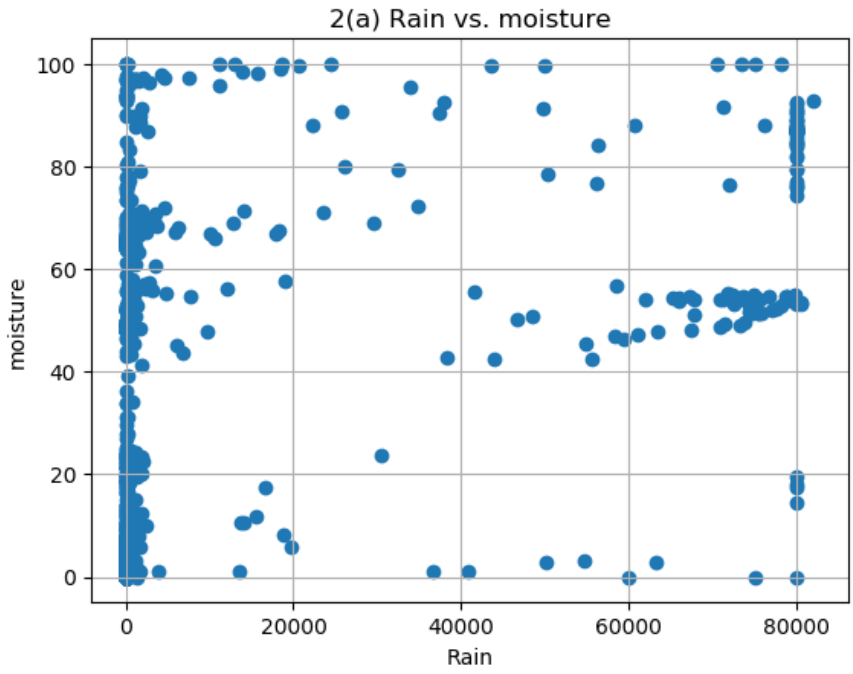

Figure 9

b) INFERENCES: (Temperature vs. Others)

i. On warmer days, there's more humidity.

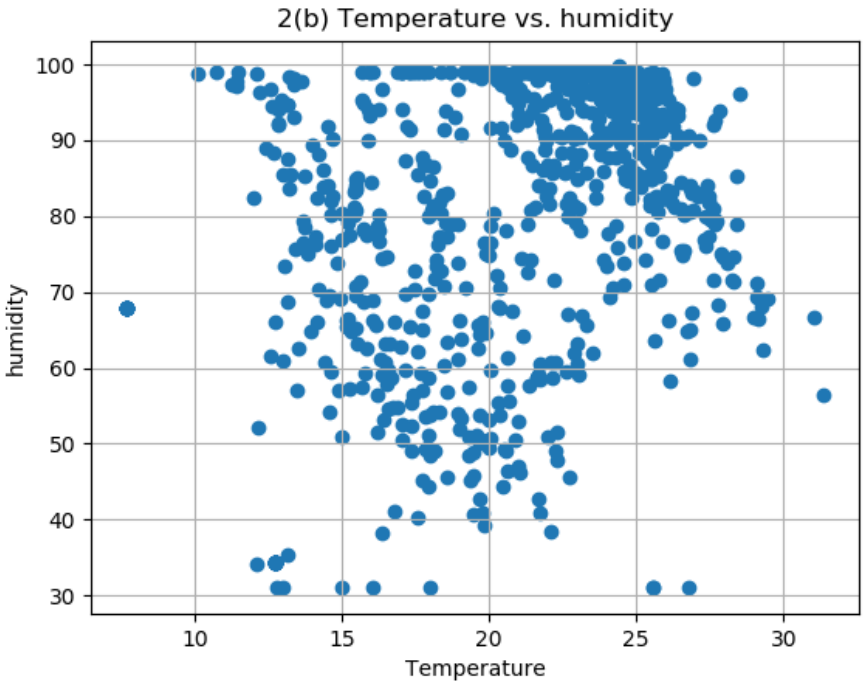

Figure 10

ii.

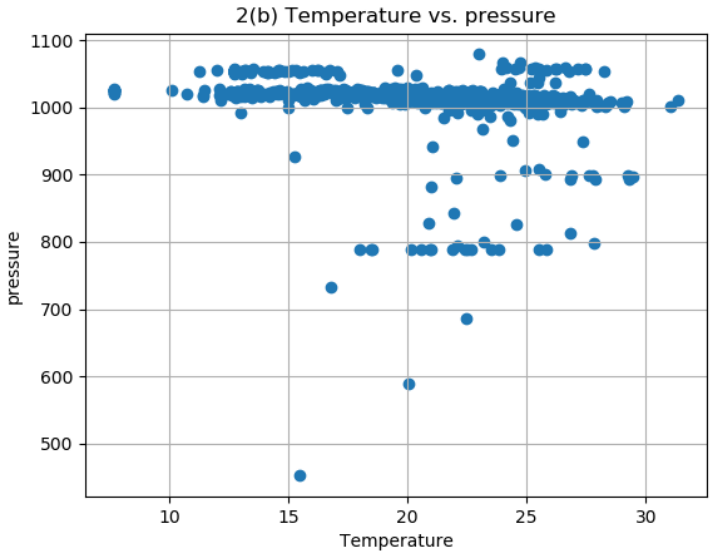

Figure 11: No visible relation.

iii. Same as above.

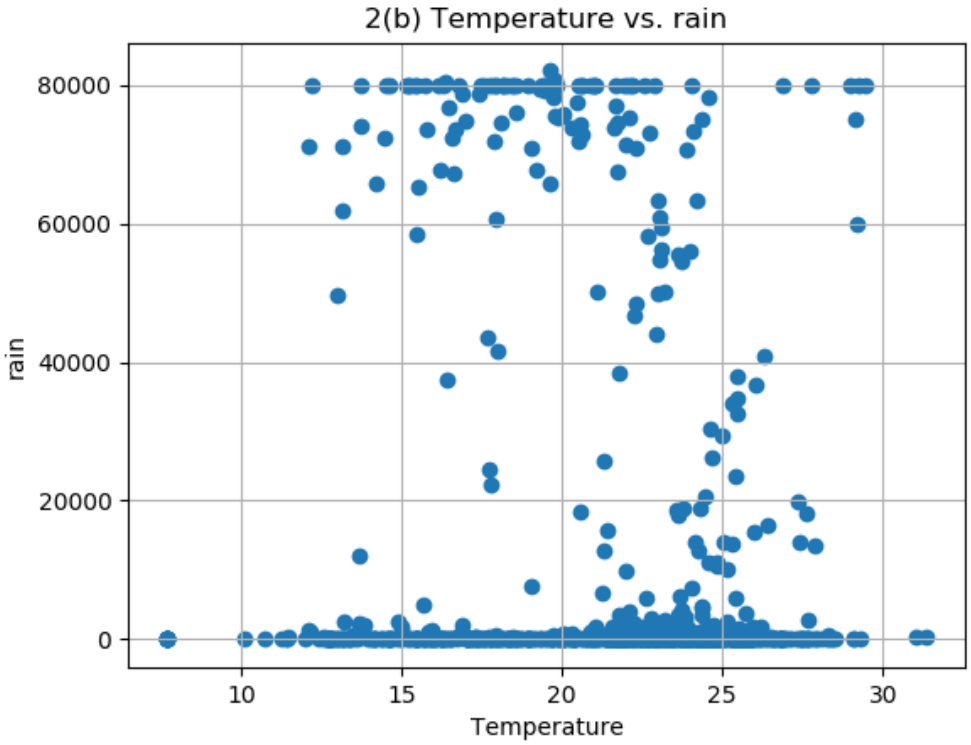

Figure 12

iv. Same as above.

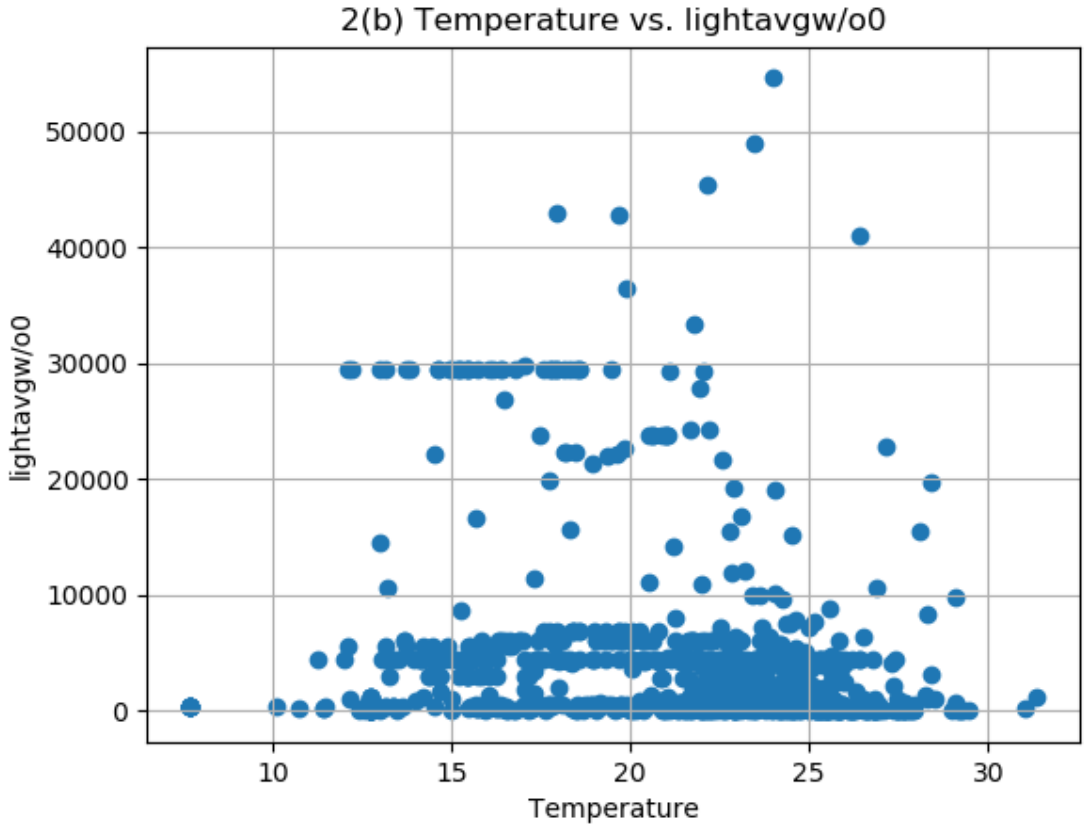

Figure 13

v.

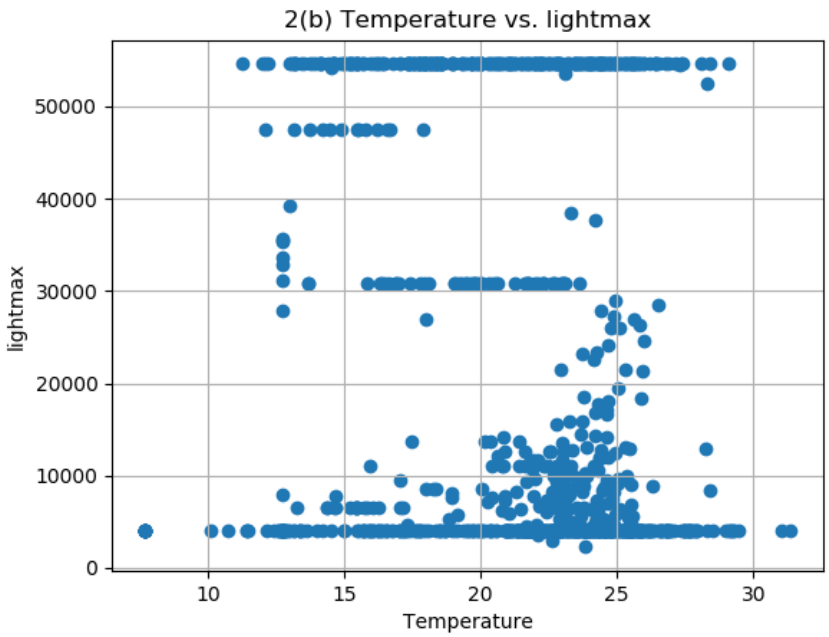

Figure 14

vi. Same as above. No visible relation. Majorly due to external factors.

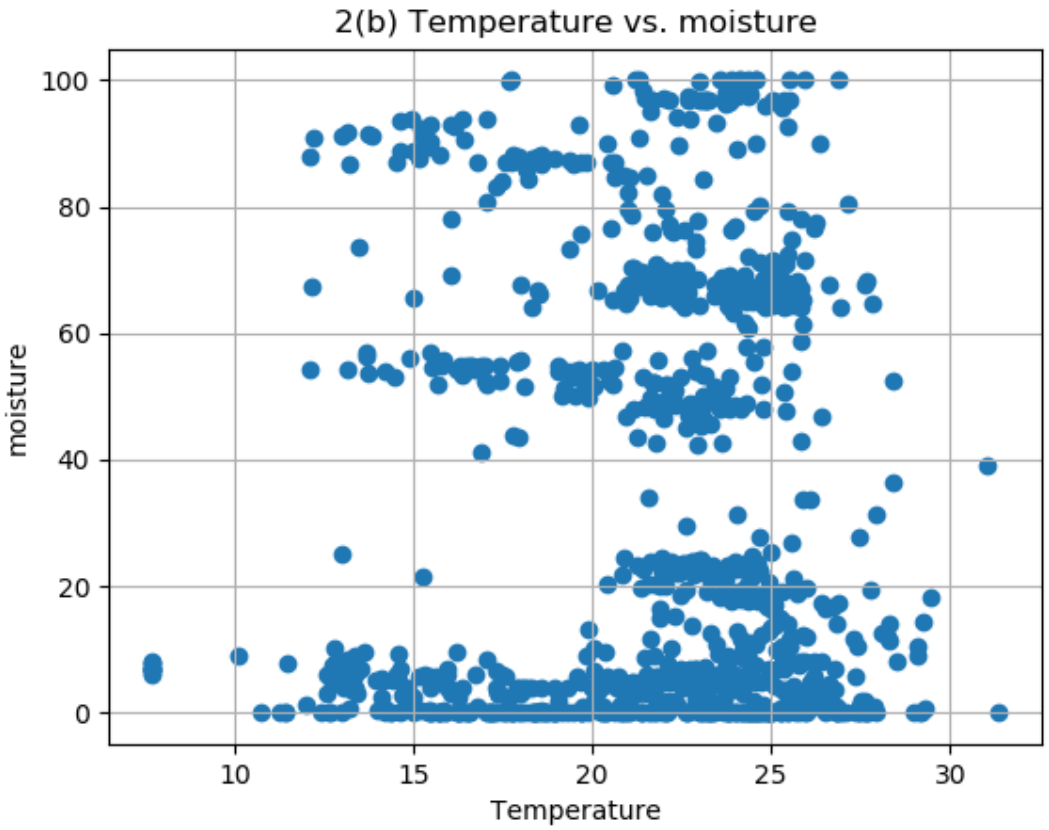

Figure 15

3.

```
A3. Finding Pearson's correlation coefficient...

For rain & others:
Corr(rain,temperature): -0.10889328204367477
Corr(rain,humidity): -0.4349168447044639
Corr(rain,pressure): 0.07078467382175054
Corr(rain,rain): 1.0
Corr(rain,lightavgw/o0): 0.527490309640136
Corr(rain,lightmax): 0.3128427375815485
Corr(rain,moisture): 0.42692792937345597

For temperature & others
Corr(temperature,temperature): 1.0
Corr(temperature,humidity): 0.40156984787421673
Corr(temperature,pressure): -0.18138907534693857
Corr(temperature,rain): -0.10889328204367477
Corr(temperature,lightavgw/o0): -0.18139996347905046
Corr(temperature,lightmax): -0.14588351410680925
Corr(temperature,moisture): 0.08066019795561973
```

Figure 16: Correlation values of the scatter plots obtained before

To analyse and then predict rainfall in the region, the parameters that should be taken into account should be, humidity, light average, and moisture. We can discard temperature, pressure and light max (we have a better parameter: light avg.).

However, to predict the temperature of a day, we should be more inclined to use only humidity however those with R values above 0.18 could also help but they pose the danger of over-fitting the data. Rain and moisture parameters should certainly be discarded for this prediction.

4. The rain histogram indicated relatively more dry spells. It also indicates that it rains a lot when it happens. The place seems to have extreme conditions as that seen in the mountains. The moisture data seems to be in close agreement with the speculation.

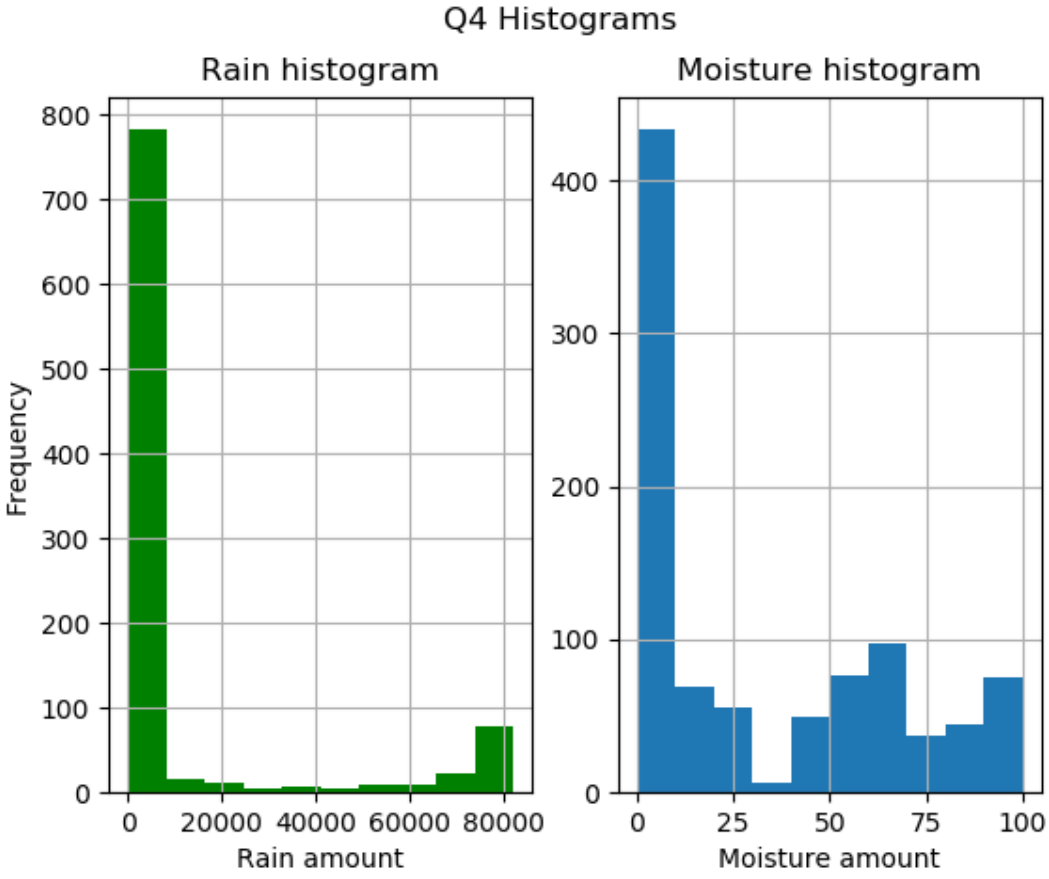

Figure 17

5. Histograms of rain in each of the observation stations.

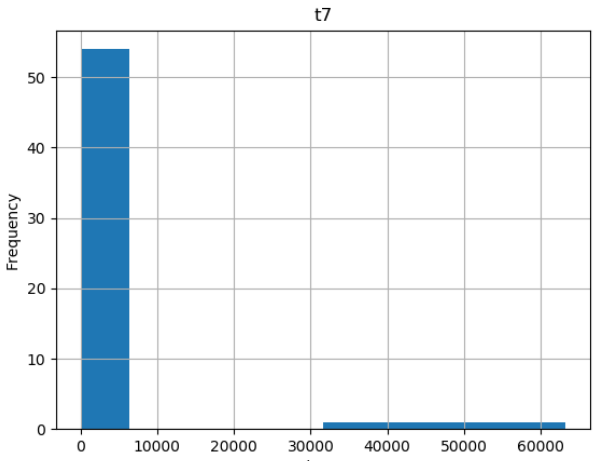

Figure 21

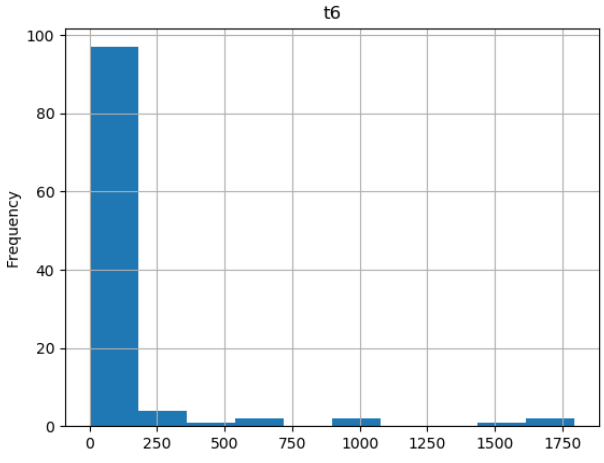

Figure 20

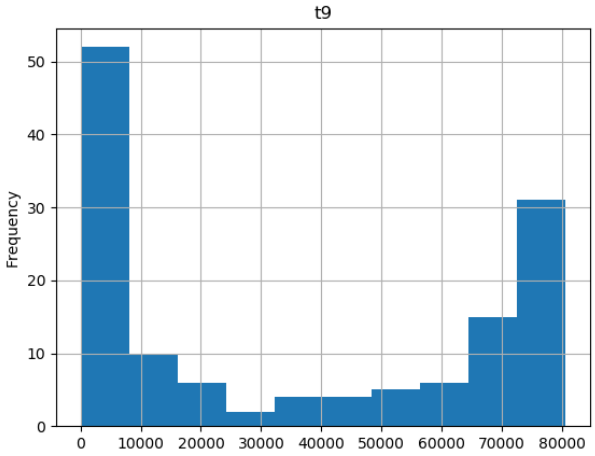

Figure 19

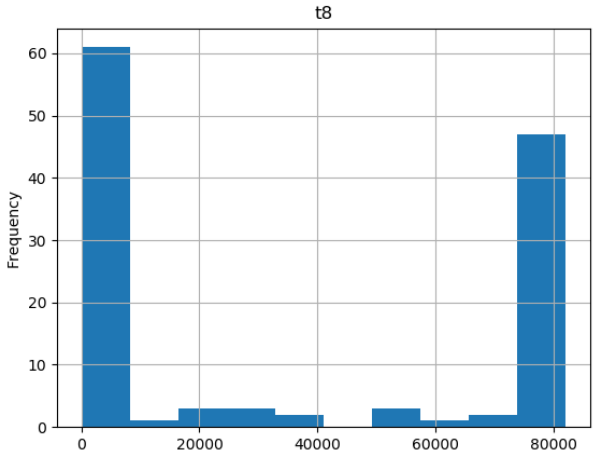

Figure 18

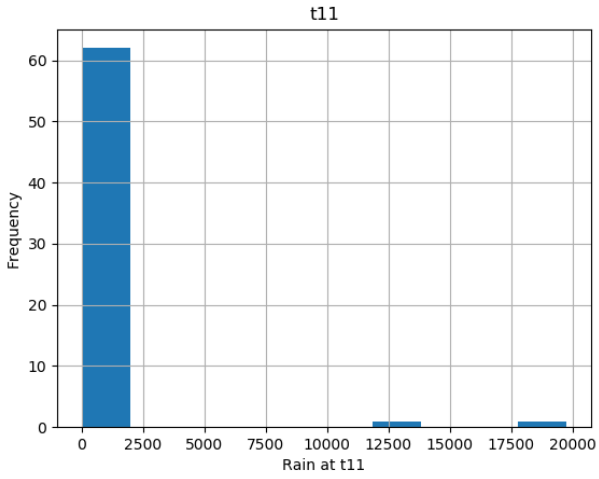

Figure 22

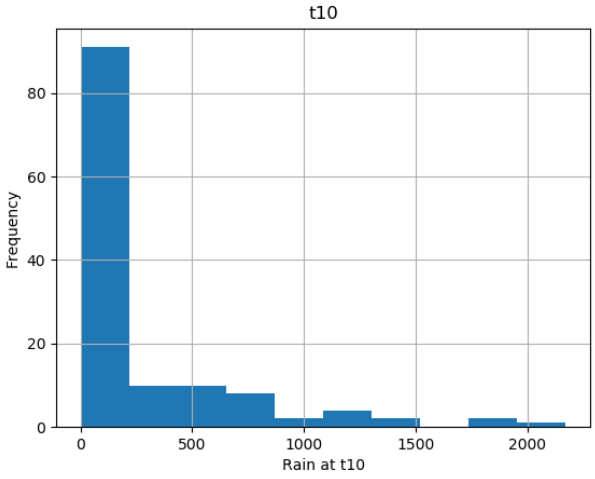

Figure 23

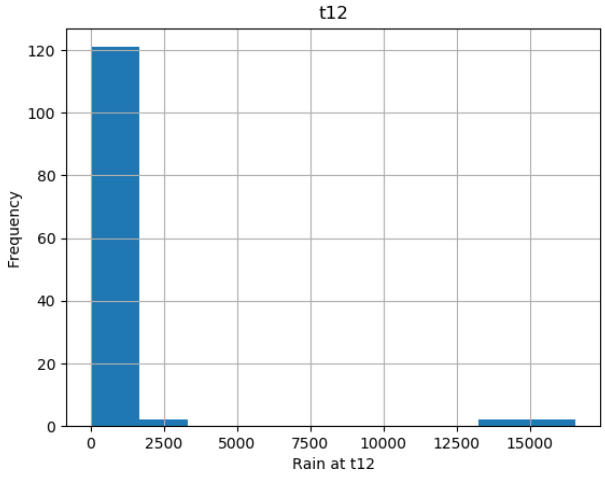

Figure 27

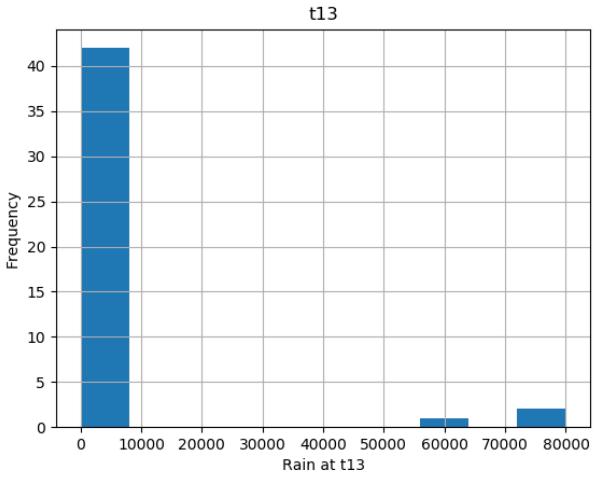

Figure 26

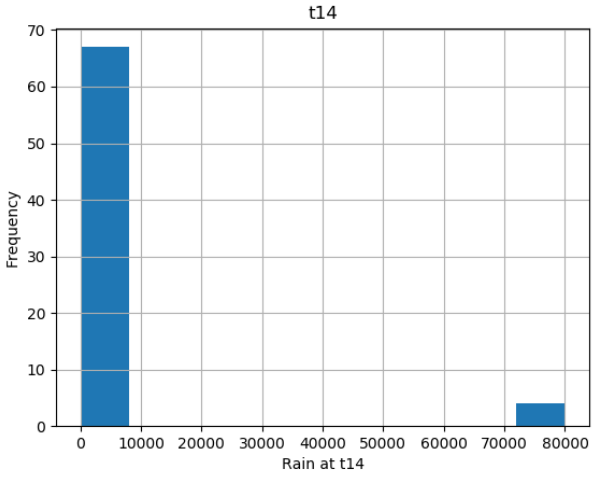

Figure 25

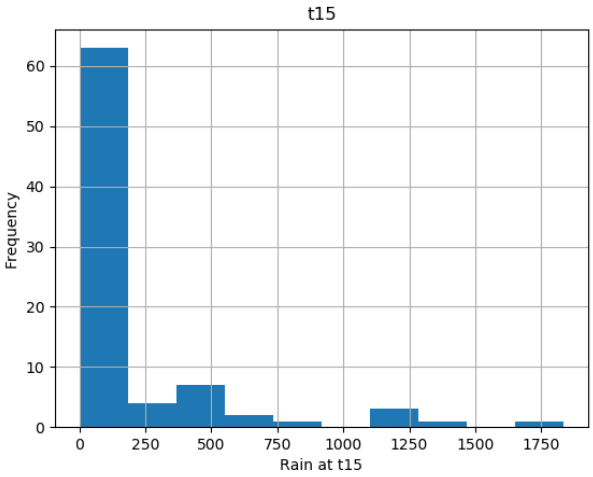

Figure 24

Observation:

Dry Stations:

T7, T6, T11, T12, T13, and T14

Moderate Rain Stations:

T15 and T10

Wet Stations:

T8 and T9

6. Rain data is vey sparse in nature.
Generally, there's low moisture in this area.

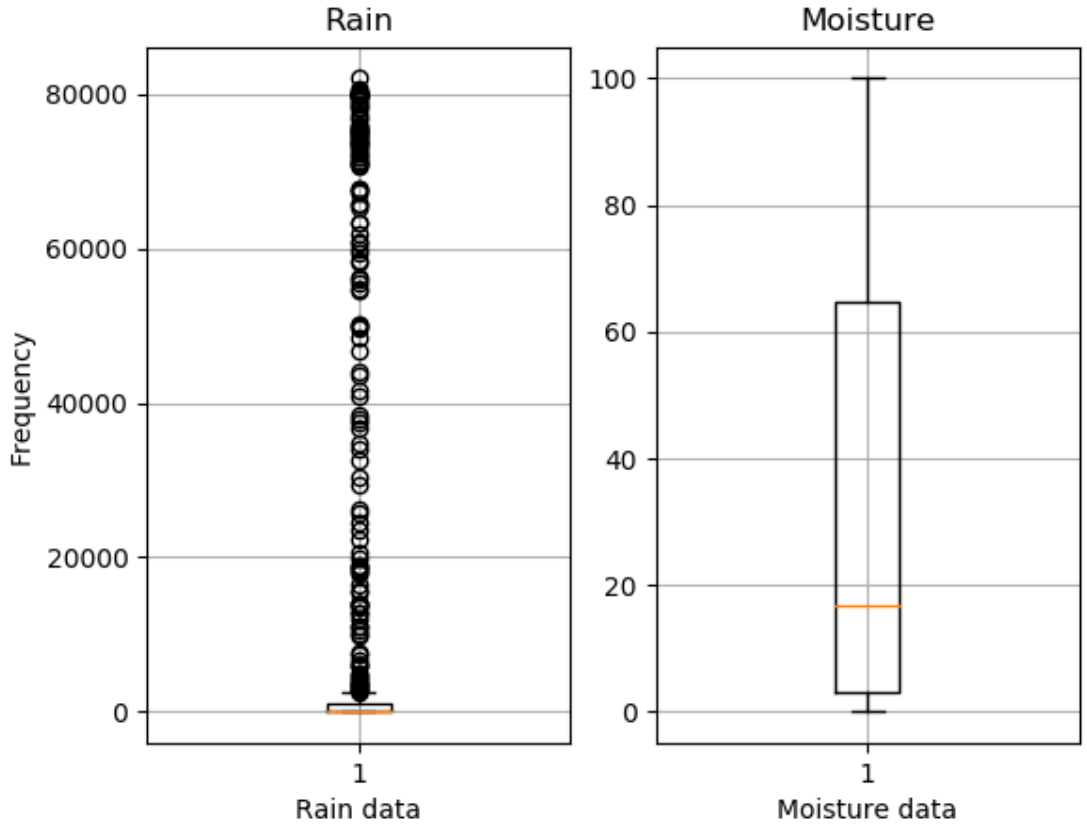

Figure 28: Box-plot