REPORT (LAB 1)
Rashmi, B19218

1. Mean of:
- temperature   21.214888 °C
- humidity   83.479932 %
- pressure   1009.008774 mb(millibars)
- rain   10701.538370 ml
- lightavgw/o0   4438.428453 lx(lux)
- lightmax   21788.623280 lx
- moisture   32.386053 %

Median of:
- temperature   22.272730 °C
- humidity   91.380950 %
- pressure   1014.677832 mb
- rain   18 ml
- lightavgw/o0   1656.88 lx
- lightmax   6634 lx
- moisture   16.7042 %

Mode of:
- temperature   12.72727 °C
- humidity   99.0 %
- pressure   789.392692 mb
- rain   0.0 ml
- lightavgw/o0   4488.9103 lx
- lightmax   4000 lx
- moisture   0.0 %

Minimum of:
- temperature   7.672900 °C
- humidity   31 %
- pressure   452.097887 mb
- rain   0.0 ml
- lightavgw/o0   0.0 lx
- lightmax   2259.0 lx
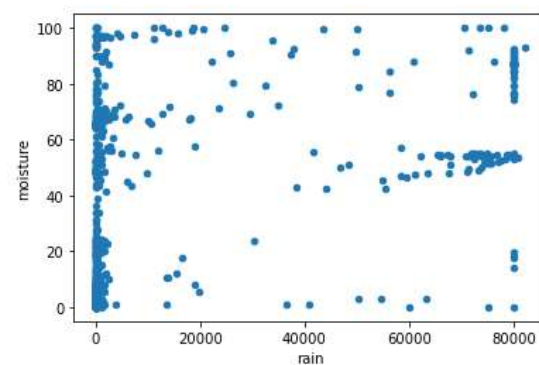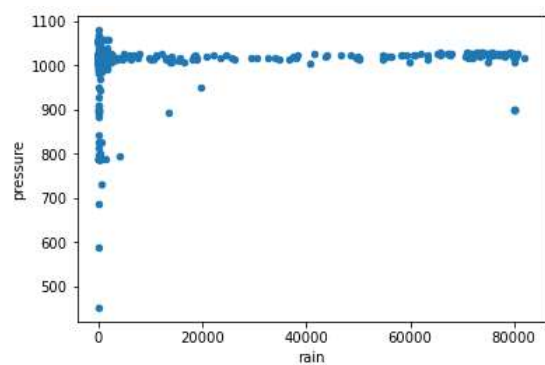- moisture   0.0 %

Maximum of:
- temperature   31.375 °C
- humidity   99.720 %
- pressure   1079.162 mb
- rain   82037.250 ml

- lightavgw/o0   54612 lx
- lightmax   54612 lx
- moisture   100.0 %
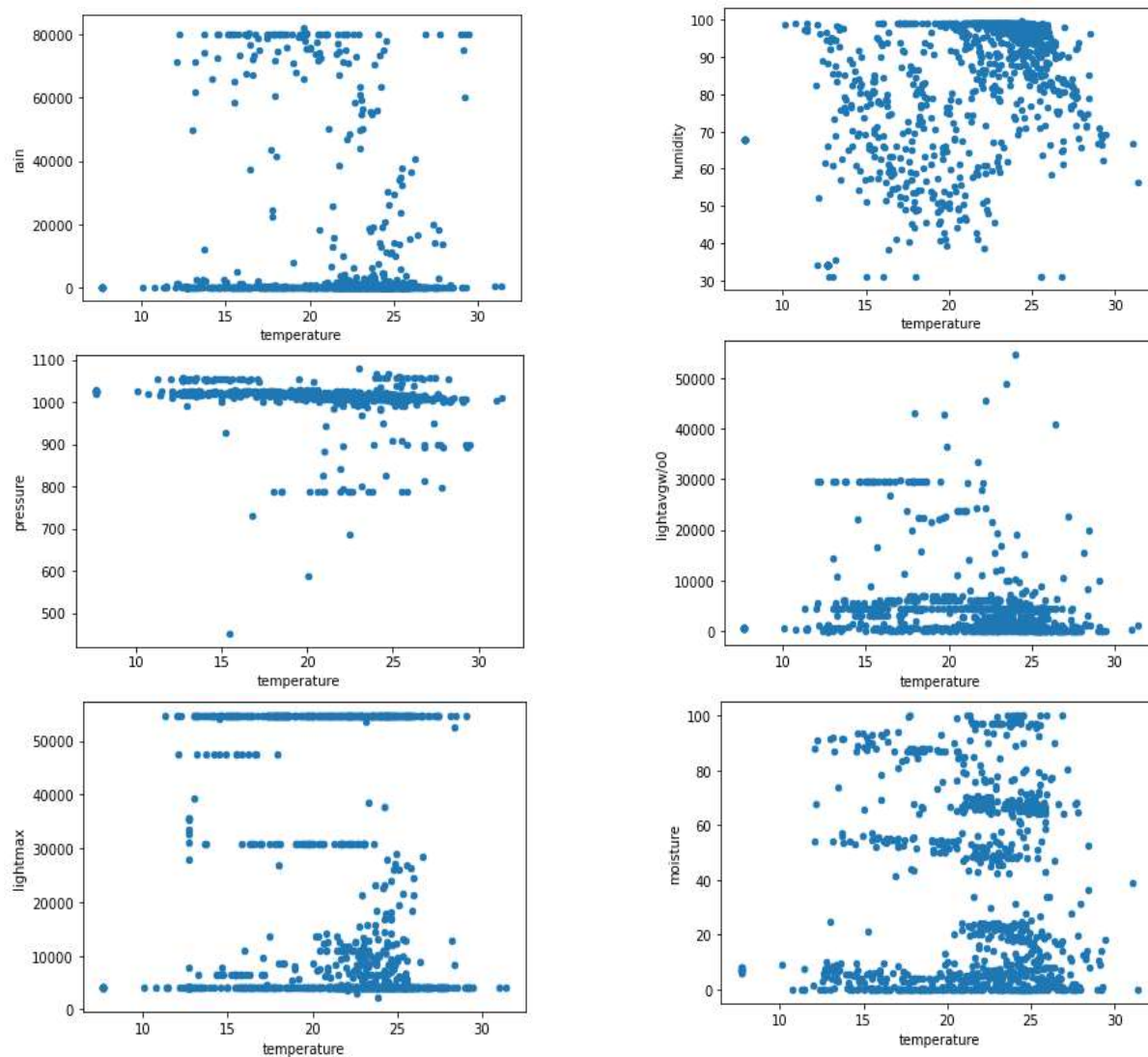
Standard Deviation of:

- temperature   4.355818 °C
- humidity   18.210065 %
- pressure   46.980477 mb
- rain   24852.255288 ml
- lightavgw/o0   7573.162806 lx
- lightmax   22064.993089 lx
- moisture   33.653245 %

2.(a) Scatter Plots between **'rain'** and each of the other attributes:
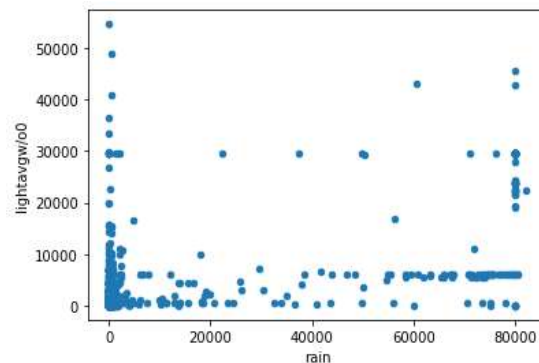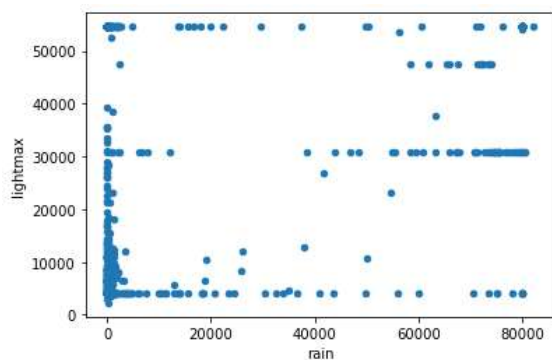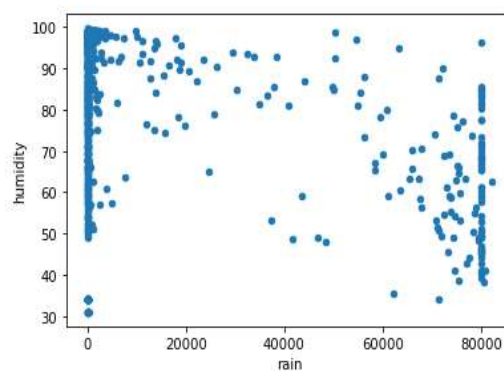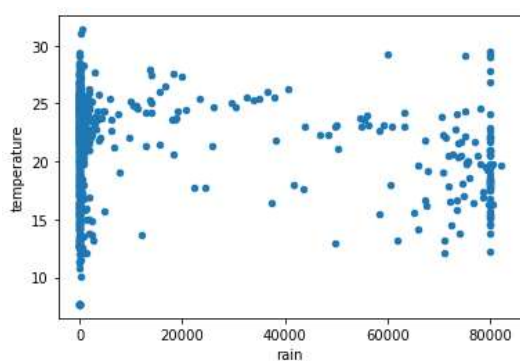
2.(b) <u>Scatter Plots</u> between **'temperature'** and each of the other attributes:

3. (a) <u>Correlation Coefficient</u> (Pearson) of **'rain'** with:

- temperature   -0.108893
- humidity   -0.434917
- pressure   0.070785
- lightavgw/o0   0.527490
- lightmax   0.312843
- moisture   0.426928

(b) <u>Correlation Coefficient</u> (Pearson) of **'temperature'** with:

- humidity   0.401570
- pressure   -0.181389
- rain   -0.108893
- lightavgw/o0   -0.181400
- lightmax   -0.145884
- moisture   0.080660

Inference from above <u>Scatter Plots</u> and corresponding <u>Correlation Coefficients:</u>

In 2(a) scatter plots between 'rain' and each of the other attributes is shown and in 2(b) scatter plots between 'temperature' and each of the other attributes is shown above.

- <u>Positive</u> value of correlation coefficient shows that the two attributes are positively correlated i.e. if one increases, other will also <u>increase</u> and vice-versa.
- <u>Negative</u> value of correlation coefficient shows that the two attributes are negatively correlated i.e. if one increases, other will <u>decrease</u> and vice-versa.
- If value is 0, then there is no correlation between them and they may be independent of each other.
- Value of correlation coefficient varies from -1 to 1. Higher the magnitude of correlation coefficient is, stronger the correlation.

The scatter plot between 'rain' and 'pressure' is almost linear and parallel to x-axis indicating that they are very weakly correlated. This can also be inferred from their correlation value i.e. 0.070785 It is almost 0 but still positive.

In 2(a) maximum correlation is of 'rain' and 'lightavgw/o0' i.e. 0.527490 which shows that these are strongly correlated in comparison with others. Otherwise rest of all have less than this and their plots have too much randomness, no particular pattern is shown overall.

The scatter plot of 'temperature' and 'humidity' is having pattern with some slope. It can be inferred that they have good correlation and also it is proved by their correlation value which is highest in 2(b) i.e. 0.401570

On the other hand, 'temperature' and 'moisture' have minimum correlation value i.e. 0.080660 that's why their scatter plot is linear and almost parallel to x-axis. This shows their weak correlation and their independency.

Almost all scatter plots have too much randomness and it's difficult to infer a pattern in them without correlation value. Patterns are visible in shorter range only.
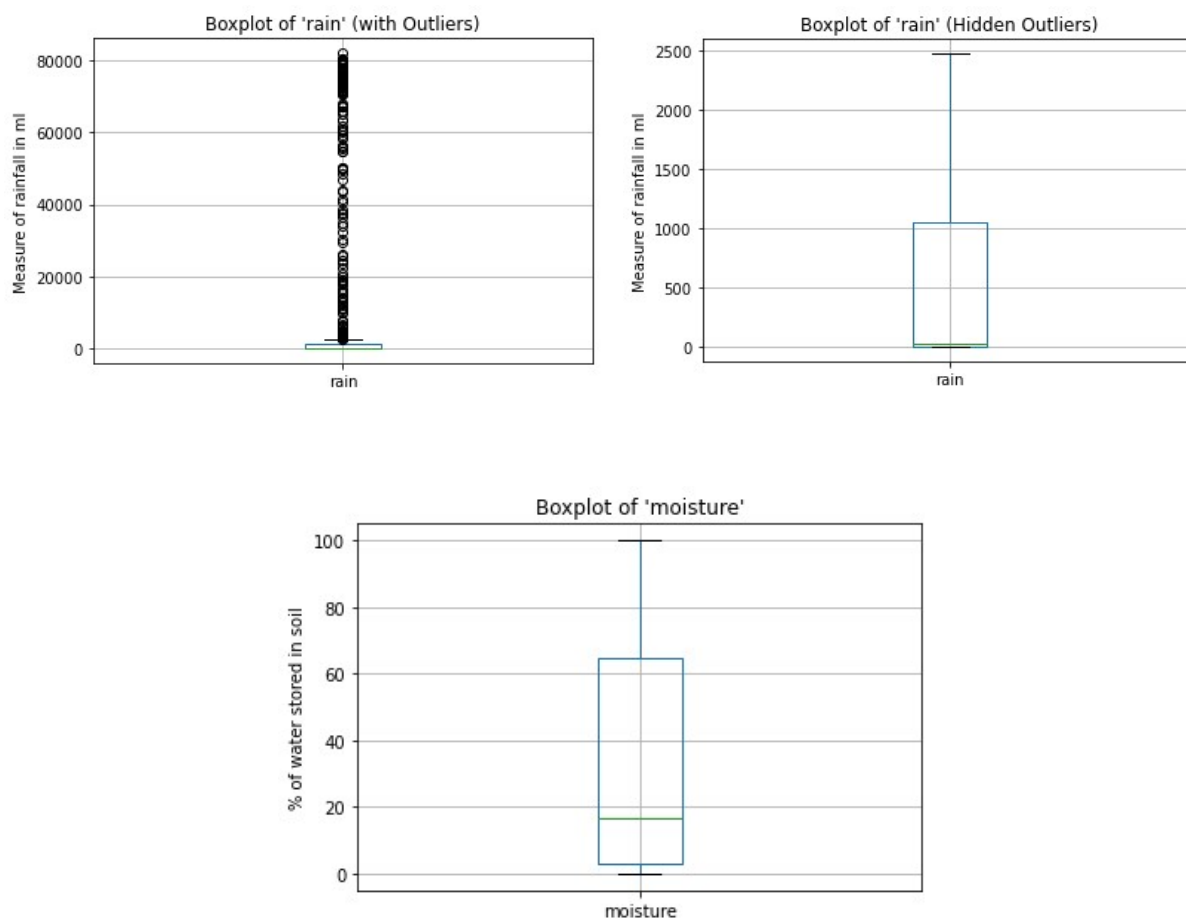
4. <u>Histogram</u> of **'rain'** and **'moisture'**:
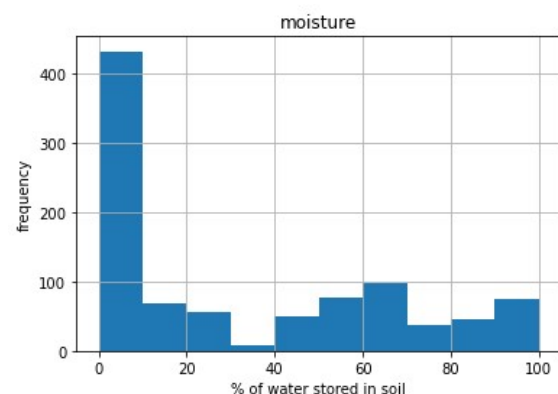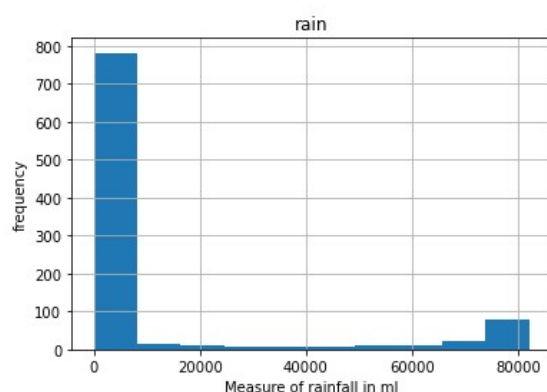
5. <u>Histogram</u> of **'rain'** for each of the 10 stations (t10, t11, t12, t13, t14, t15, t6, t7, t8, t9):
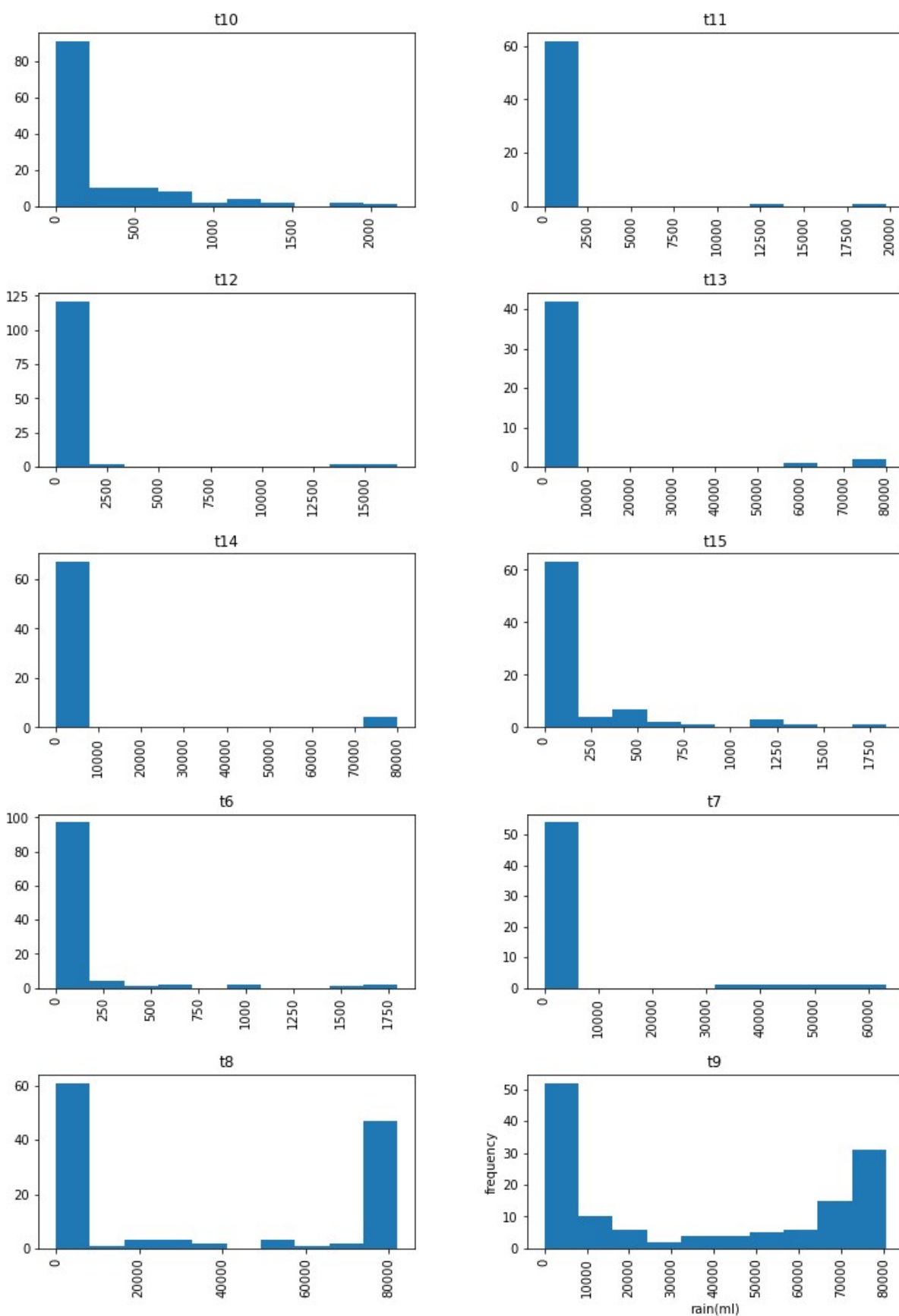
6. <u>Boxplots</u> of **'rain'** and **'moisture'**:

Inference from above <u>Histograms</u> and <u>Boxplots:</u>

In 4, histogram of 'rain' and 'moisture' is shown for all of the 10 stations in 1 plot whereas in 5, histograms of 'rain' for 10 stations are shown separately.

In 4, frequency of rainfall less than 10000 ml is higher than rest of the values. And frequency of water stored in soil less than 10% is highest.

In 5, for (t9, t8) frequency of rainfall <10000ml and >80000ml is higher compared to others; for (t13, t14, t7) frequency of rainfall <10000ml is higher; for (t6, t10, t15) frequency of rainfall <250ml is higher; for (t11, t12) frequency of rainfall <2500ml is higher.

In 6, Boxplot of 'rain' has a lot of outliers indicating too much noise in the data. In boxplot with hidden outliers, we can see that median shown by green line is too close to Q1(first quartile) i.e. <u>positively skewed</u>. The bottom whisker is at 0ml and top whisker is at ~2500ml.

In boxplot of 'moisture' there is no outliers and median is shifted towards Q1(first quartile) i.e. <u>positively skewed</u>. The bottom whisker is at 0% and top whisker is at 100%.

******end******