Report

IC272

DS3 Assignment-1

Data visualization and statistics from data

Ques 1

Using pandas, Numpy, Statistics library, I calculated the values of mean, median, mode, minimum, maximum, and standard deviation for all the attributes.

The results obtained are as follow :

---------- temperature ----------

The Mean is : 21.214888105820105

The Median is : 22.27273

The Mode is : 12.727269999999999

The Min is : 7.6729

The Max is : 31.375

The Standard Deviation is : 4.353512664832514

---------- humidity ----------

The Mean is : 83.47993155555555

The Median is : 91.38095

The Mode is : 99.0

The Min is : 31.0

The Max is : 99.72

The Standard Deviation is : 18.200427162850616

---------- pressure ----------

The Mean is : 1009.008773798647

The Median is : 1014.6778321678

The Mode is : 789.3926923077

The Min is : 452.0978873239

The Max is : 1079.162

The Standard Deviation is : 46.95561327535368

---------- rain ----------

The Mean is : 10701.53837037037

The Median is : 18.0

The Mode is : 0.0

The Min is : 0.0

The Max is : 82037.25

The Standard Deviation is : 24839.102466127668

---------- lightavgw/o0 ----------

The Mean is : 4438.428453333333

The Median is : 1656.88

The Mode is : 4488.9103

The Min is : 0.0

The Max is : 54612.0

The Standard Deviation is : 7569.154781086207

---------- lightmax ----------

The Mean is : 21788.62328042328

The Median is : 6634.0

The Mode is : 4000

The Min is : 2259

The Max is : 54612

The Standard Deviation is : 22053.315399022664

---------- moisture ----------

The Mean is : 32.38605259259259

The Median is : 16.7042

The Mode is : 0.0

The Min is : 0.0

The Max is : 100.0

The Standard Deviation is : 33.63543398815152

For all the attributes in the given data set, I simply made 1 list for each and every attribute and by using these 7 lists I got the required values.

Ques 2 :

Scattered plot : The scattered plot tell us how much one variable (attribute) is effected by another. i.e. It mainly shows the effect caused by one attribute on the other one.

For the scattered plot between rain and other attributes, I had observed that if the values inside the list of the attributes and that of the list of rain increase together then the distribution of the scattered plot shows a positive association which mainly implies that the correlation coefficient is positive and if both decreases together then the distribution of the scattered plot shows a negative association which mainly implies that the correlation coefficient is negative.

If the values of the lists of attributes increases and that of rain decrease or vice versa, the scattering was so random that it becomes difficult to find any positive or negative correlation between them, which basically implies no correlation present.

- **Correlation between rain and other attributes are given as :-**

1 - Correlation (rain,temperature) : No correlation

2 - Correlation (rain,humidity) : No correlation

3 - Correlation (rain,pressure) : Positive correlation

4 - Correlation (rain,lightavgw/o0) : Negative correlation

5 - Correlation (rain,lightmax) : No correlation

6 - Correlation (rain,moisture) : Positive correlation

- **Correlation between temperature and other attributes are given as :-**

1 - Correlation (temperature,humidity) : No correlation

2 - Correlation (temperature,pressure) : Positive correlation

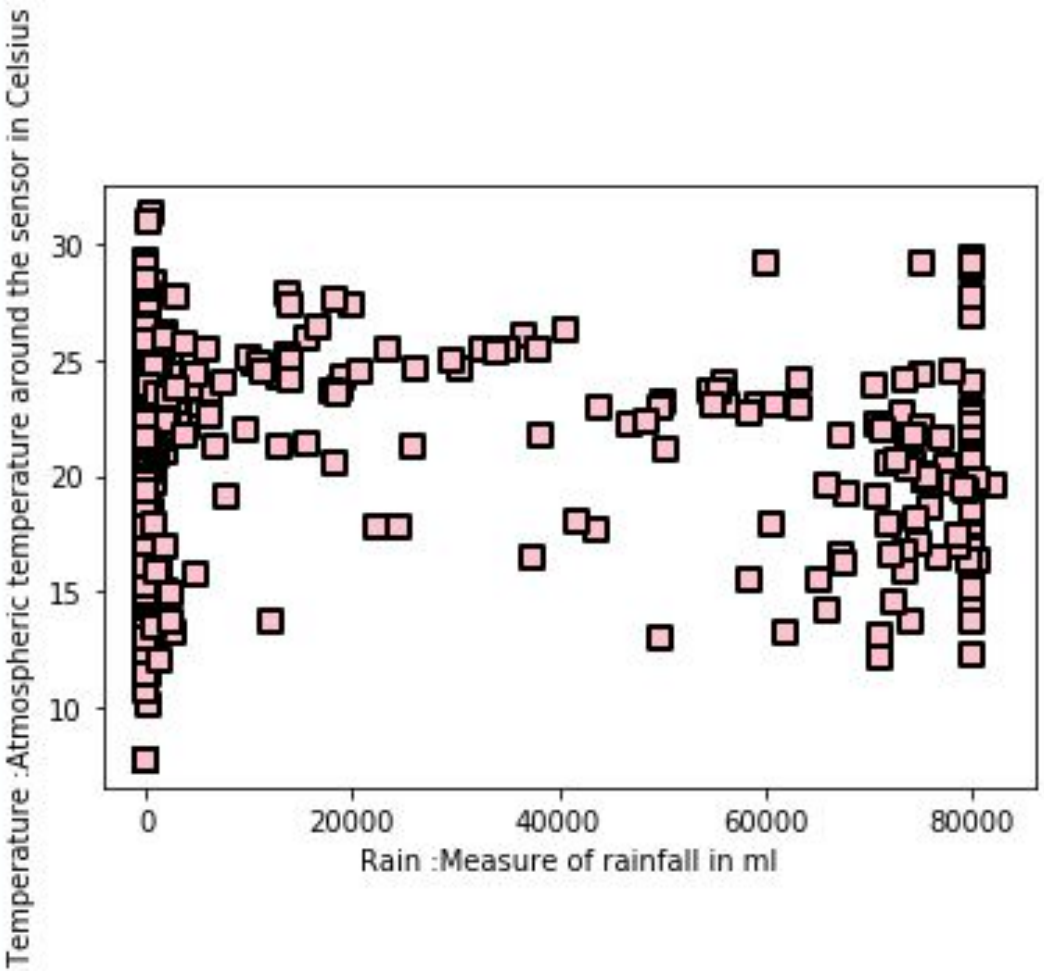
3 - Correlation (temperature,rain) : No correlation

4 - Correlation (temperature,lightavgw/o0) : Negative correlation

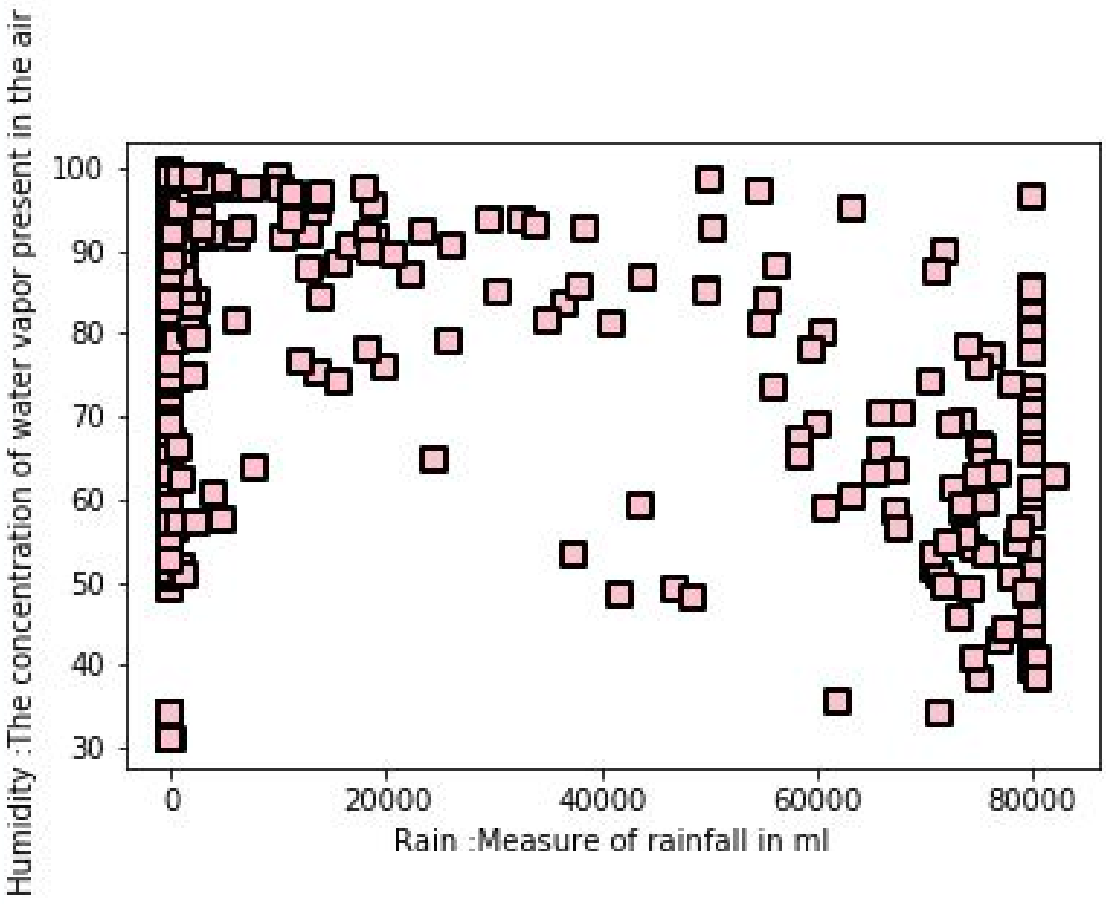
5 - Correlation (temperature,lightmax) : Negative correlation

6 - Correlation (temperature,moisture) : No correlation

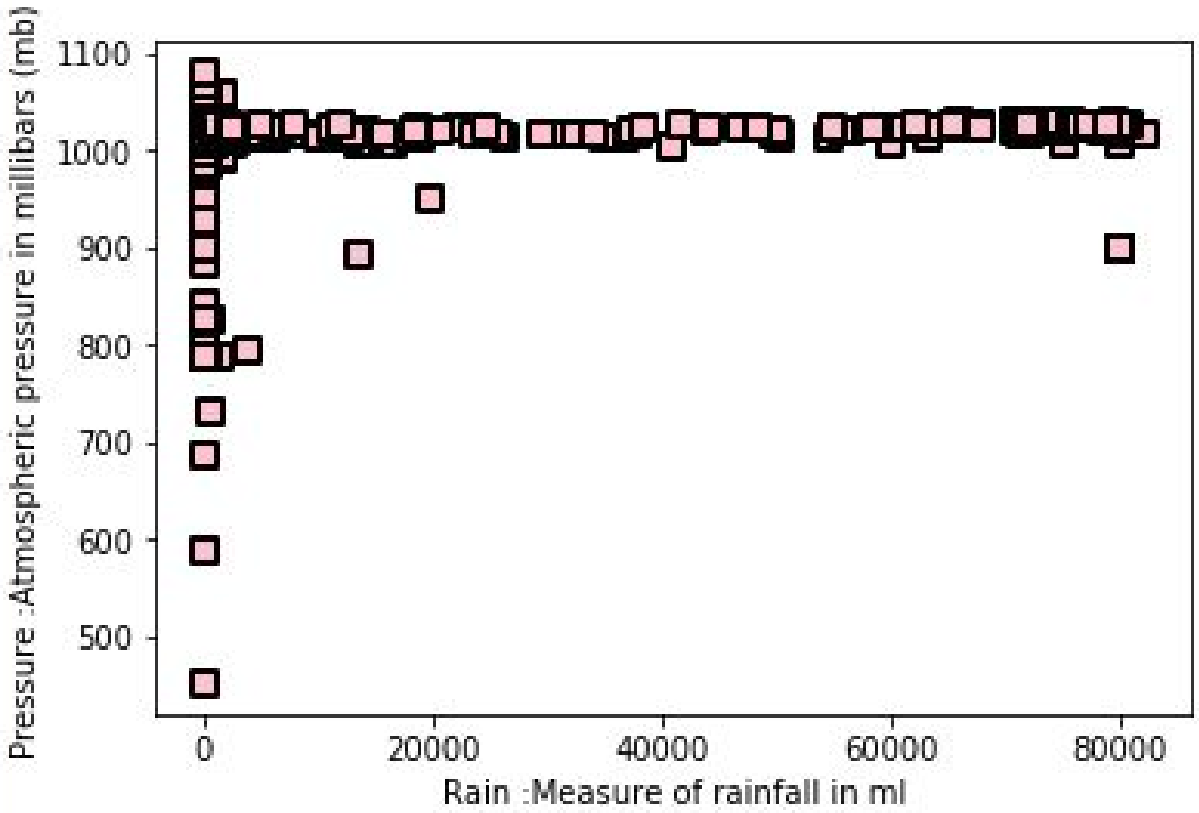
graph: 1



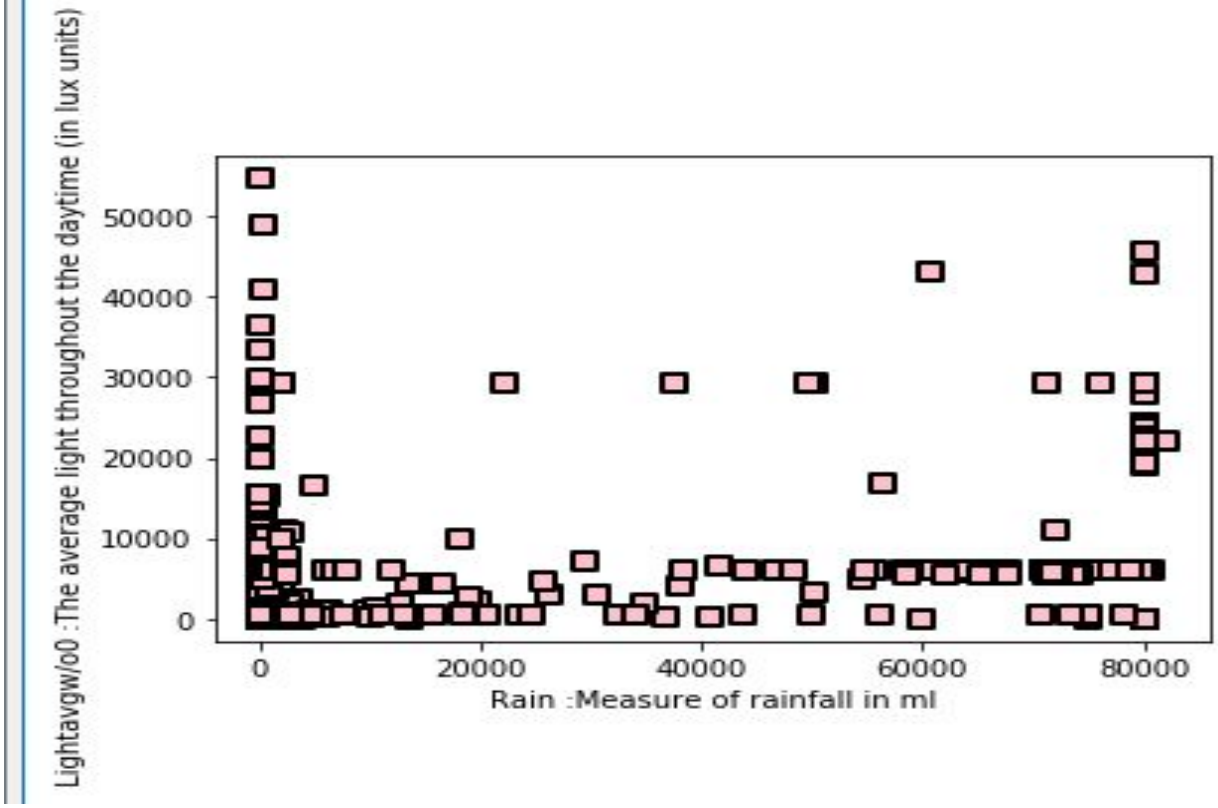
graph: 2


Humidity :The concentration of water vapor present in the air
100
90
80
70
60
50
40
30
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

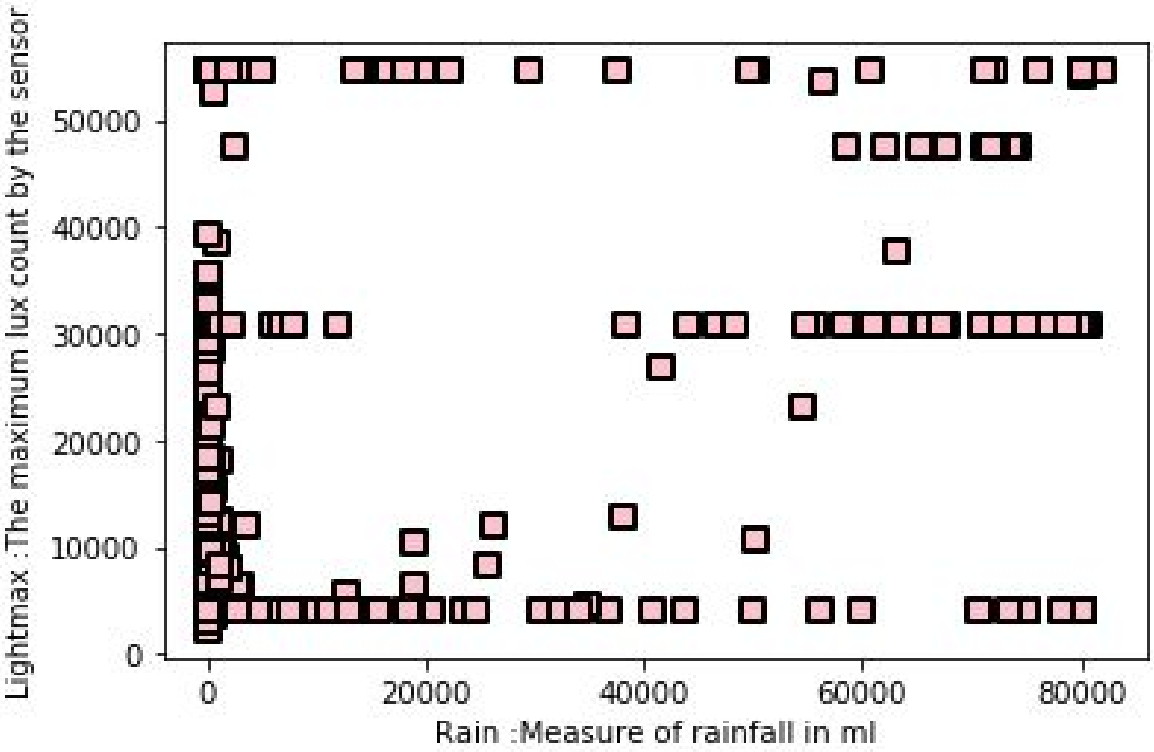
graph: 3


Pressure :Atmospheric pressure in millibars (mb)
1100
1000
900
800
700
600
500
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

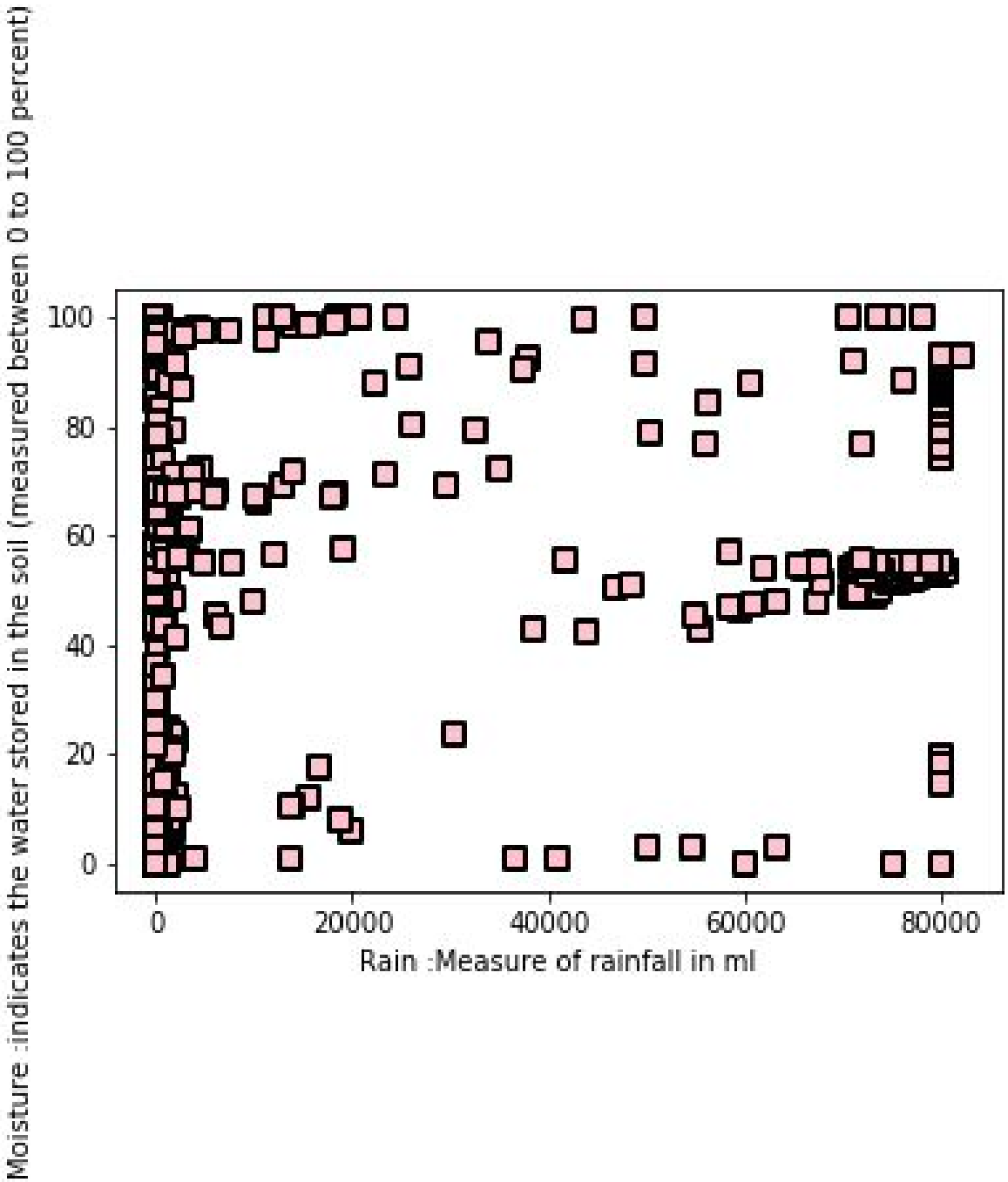
graph: 4


Lightavgw/o0 :The average light throughout the daytime (in lux units)
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

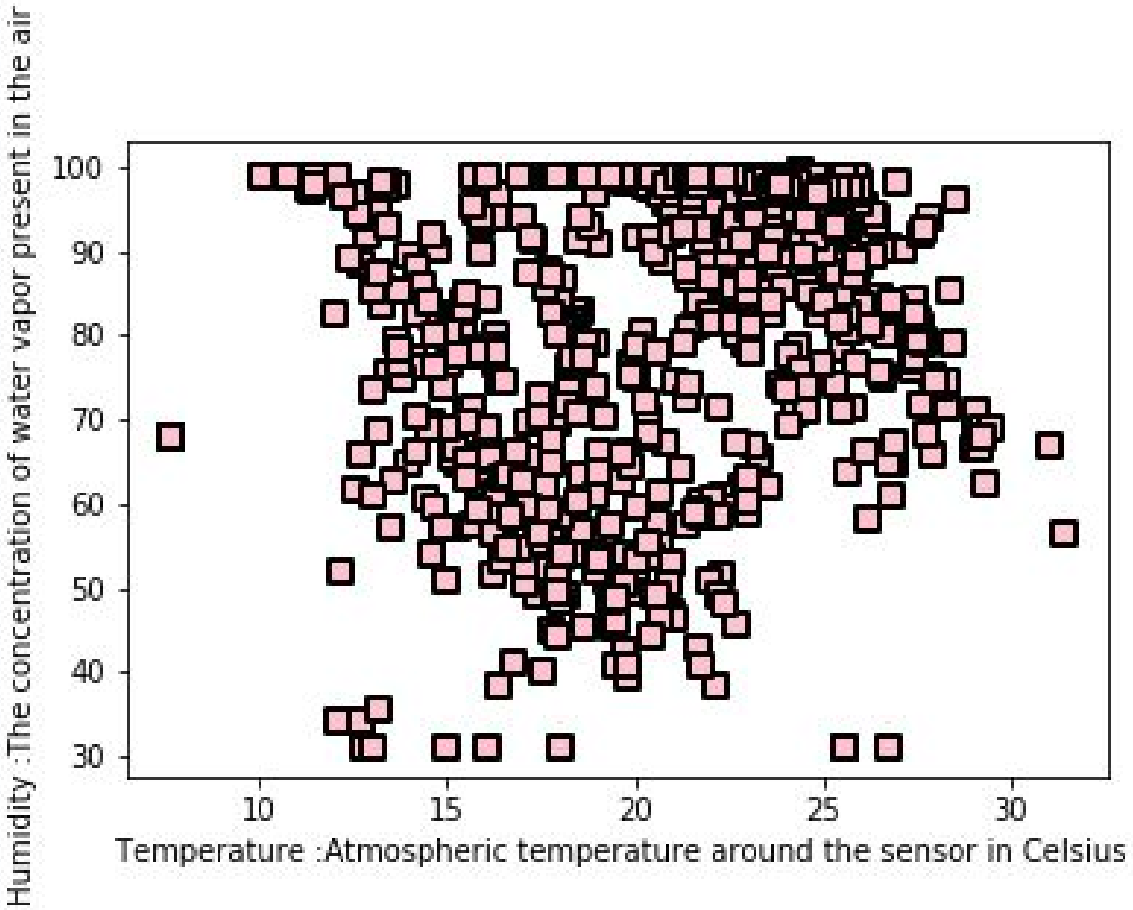
graph: 5


Lightmax :The maximum lux count by the sensor
50000
40000
30000
20000
10000
0
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

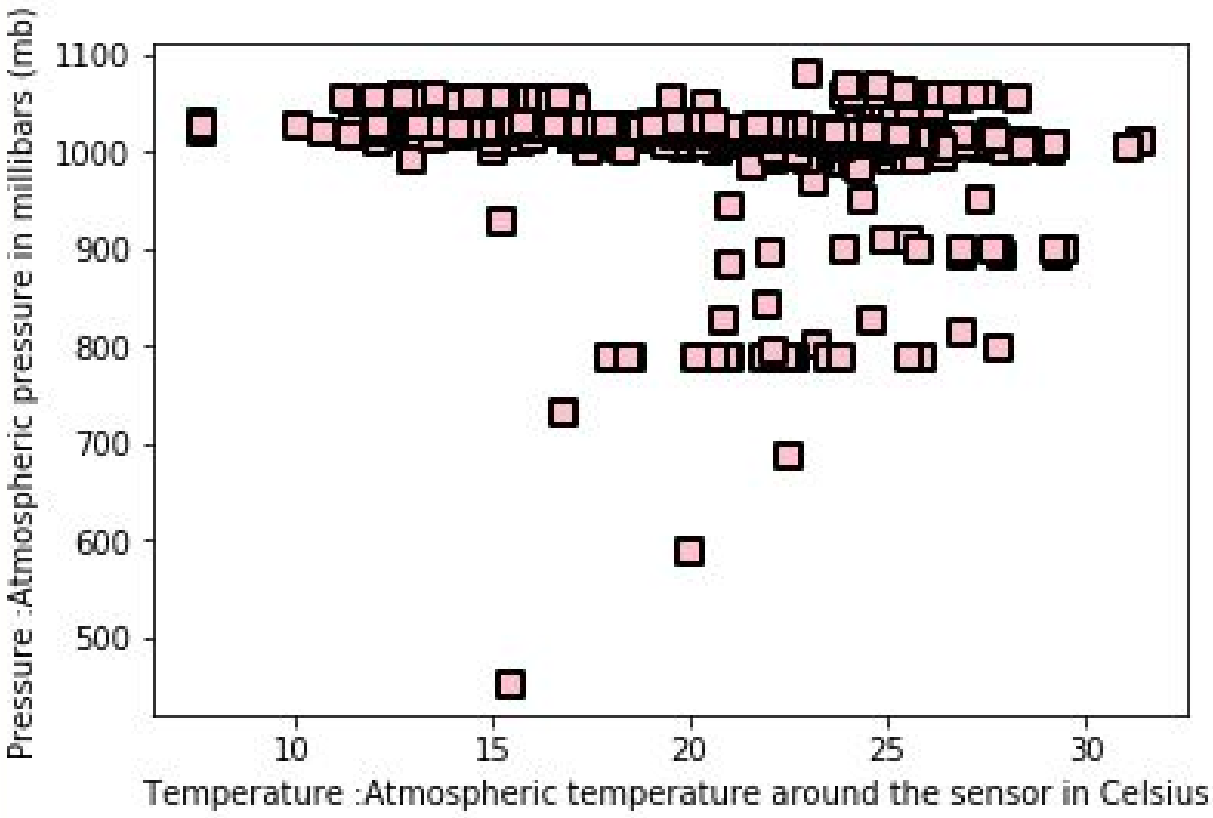
graph: 6


Moisture :indicates the water stored in the soil (measured between 0 to 100 percent)
100
80
60
40
20
0
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

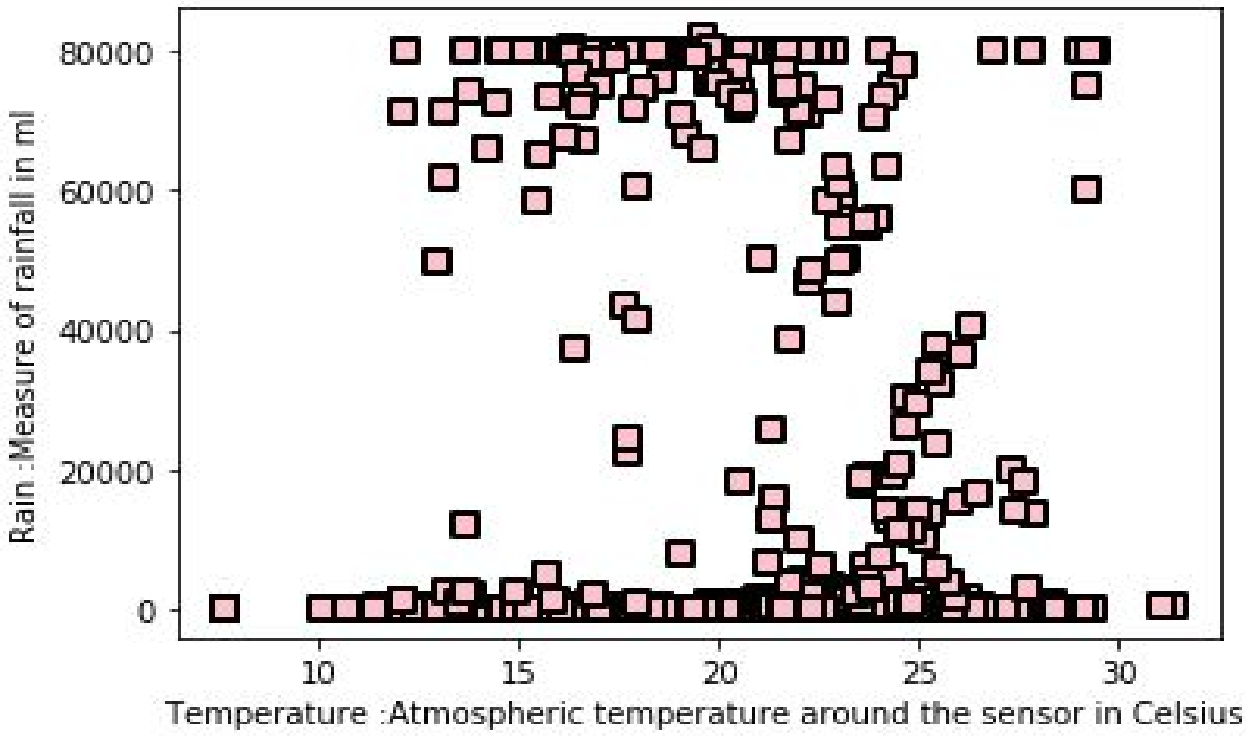
graph: 7


Humidity :The concentration of water vapor present in the air
100
90
80
70
60
50
40
30
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

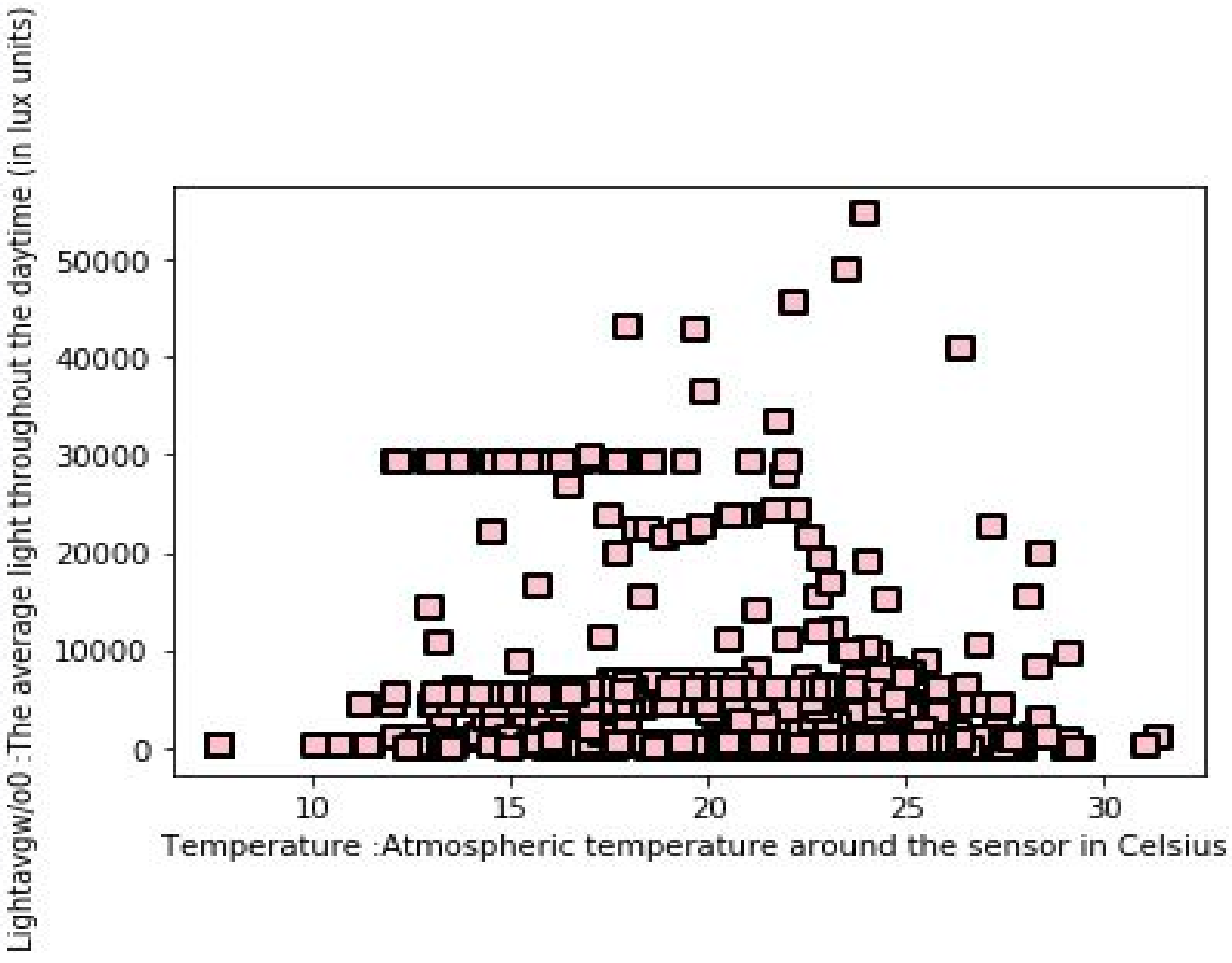
graph: 8


Pressure :Atmospheric pressure in millibars (mb)
1100
1000
900
800
700
600
500
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

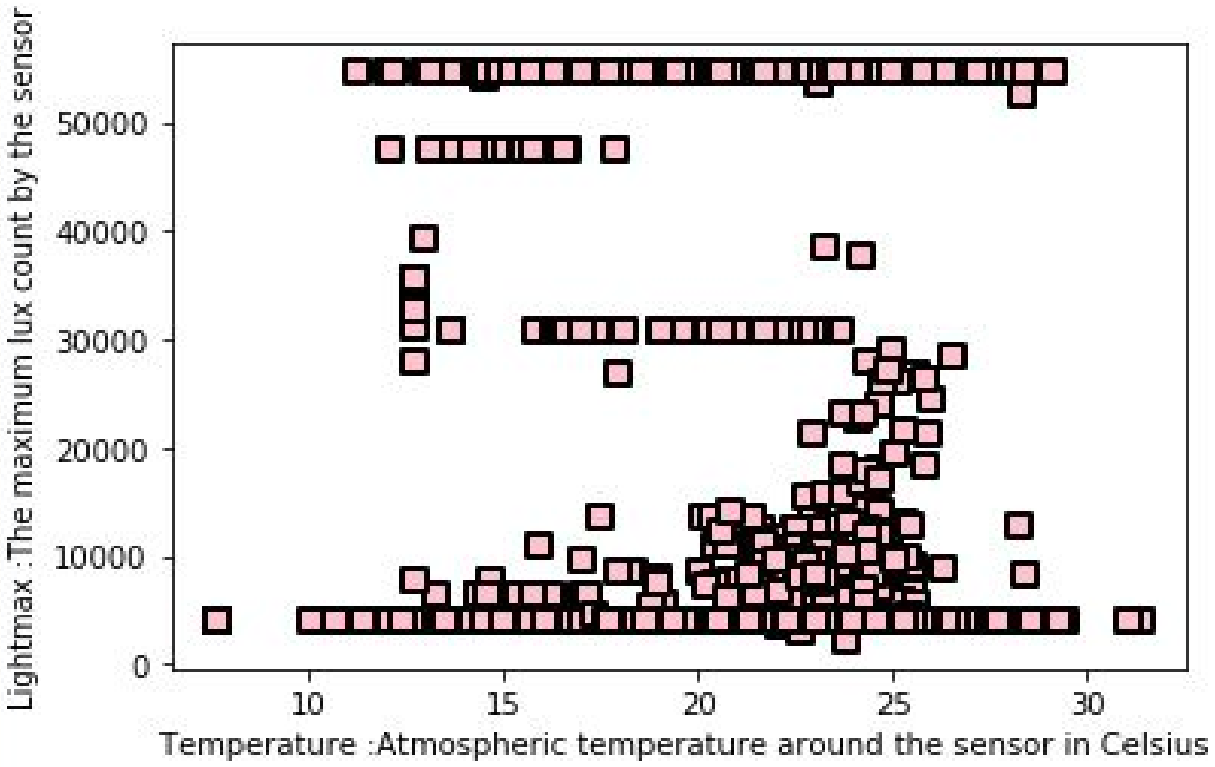
graph: 9


80000
60000
40000
20000
0
Rain :Measure of rainfall in ml
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

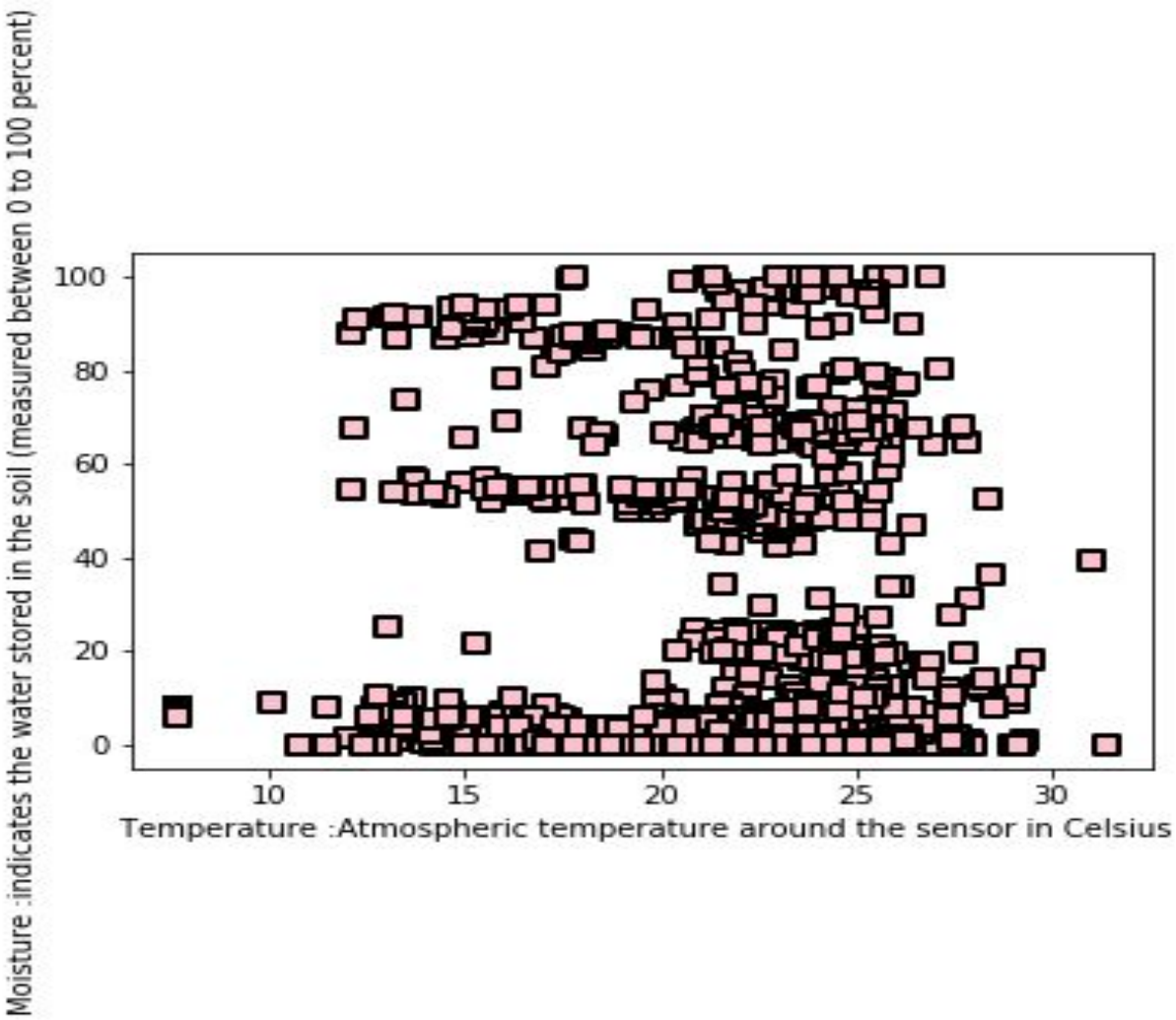
graph: 10


Lightavgw/o0 :The average light throughout the daytime (in lux units)
50000
40000
30000
20000
10000
0
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

graph: 11


Lightmax :The maximum lux count by the sensor
50000
40000
30000
20000
10000
0
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

graph: 12


Moisture :indicates the water stored in the soil (measured between 0 to 100 percent)
100
80
60
40
20
0
10
15
20
25
30
Temperature :Atmospheric temperature around the sensor in Celsius

- **Pearson's correlation coefficient is a measure of the strength of the association between the two variables.**

Ques 3(a) :

Pearsons correlation between rain and temperature:

-0.109

Pearsons correlation between rain and humidity:

-0.435

Pearsons correlation between rain and pressure:

0.071

Pearsons correlation between rain and lightavgw/o0:

0.527

Pearsons correlation between rain and lightmax:

0.313

Pearsons correlation between rain and moisture:

0.427

Ques 3(b) :

Pearsons correlation between temperature and humidity:

0.402

Pearsons correlation between temperature and pressure:

-0.181

Pearsons correlation between temperature and rain:

-0.109

Pearsons correlation between temperature and lightavgw/o0:

-0.181

Pearsons correlation between temperature and lightmax:

-0.146

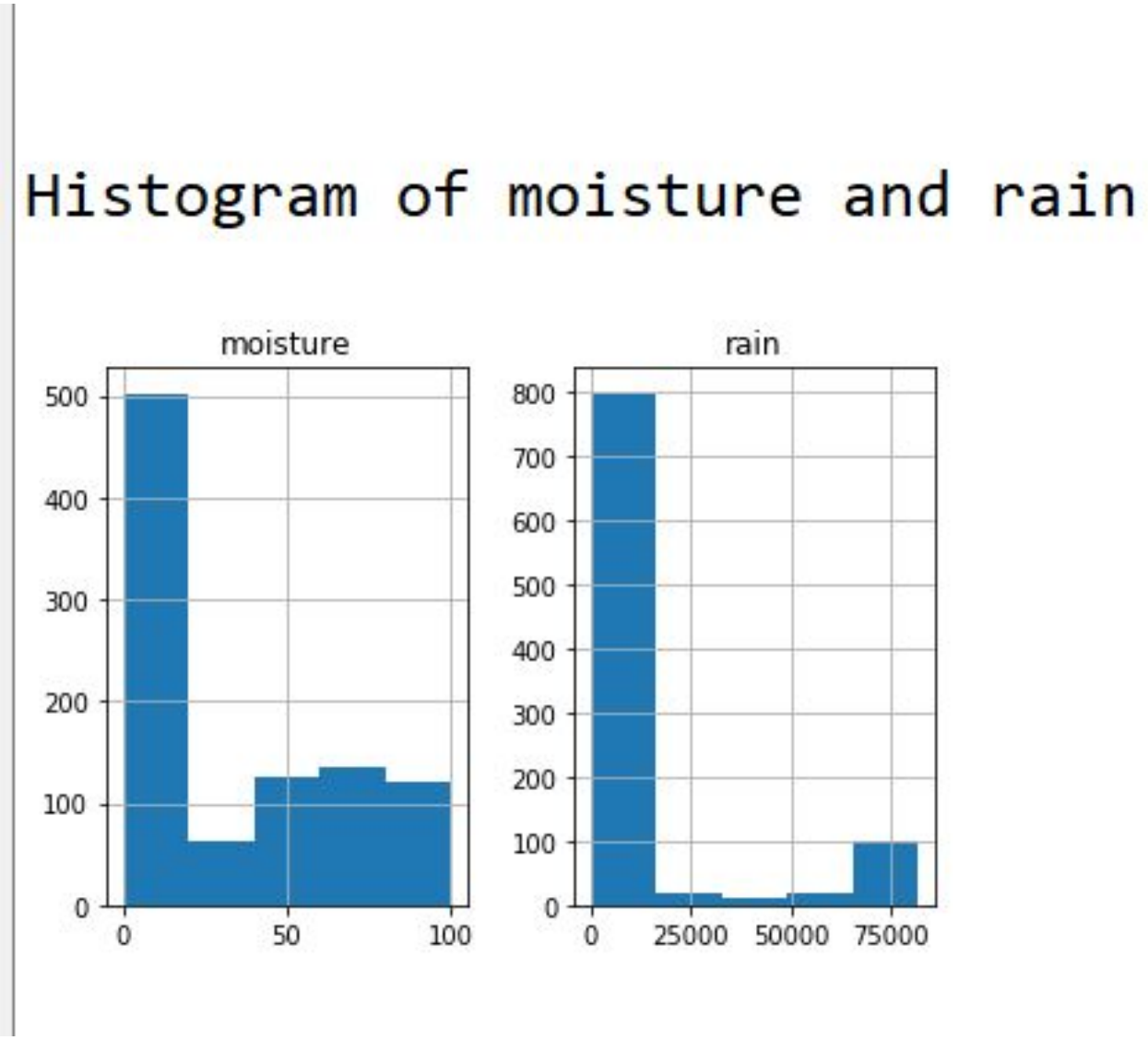
Pearsons correlation between temperature and moisture:

0.081

Ques 4 :

- **Histogram : It graphically summarize the distribution of a given data set.**

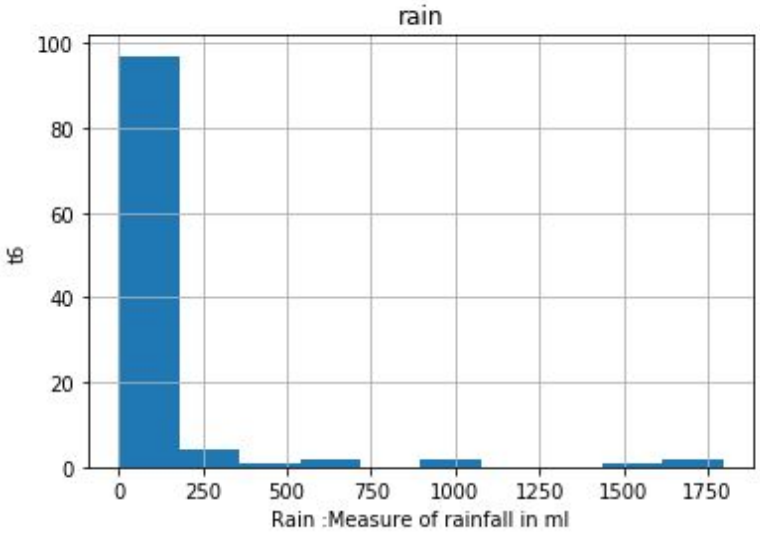
Histogram for the attributes ‘rain’ and ‘moisture’ is given as :

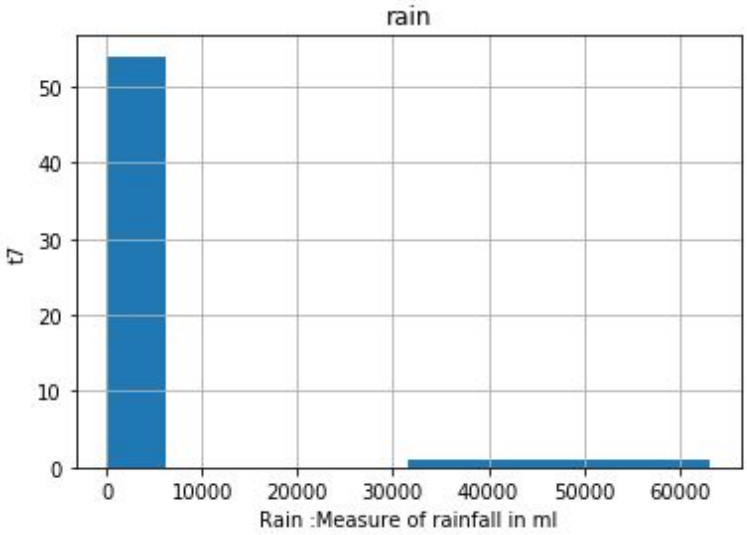


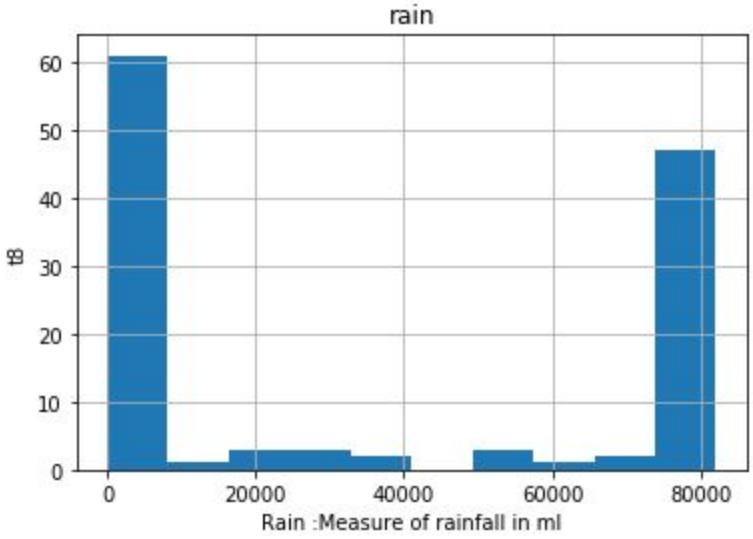
Ques 5 :

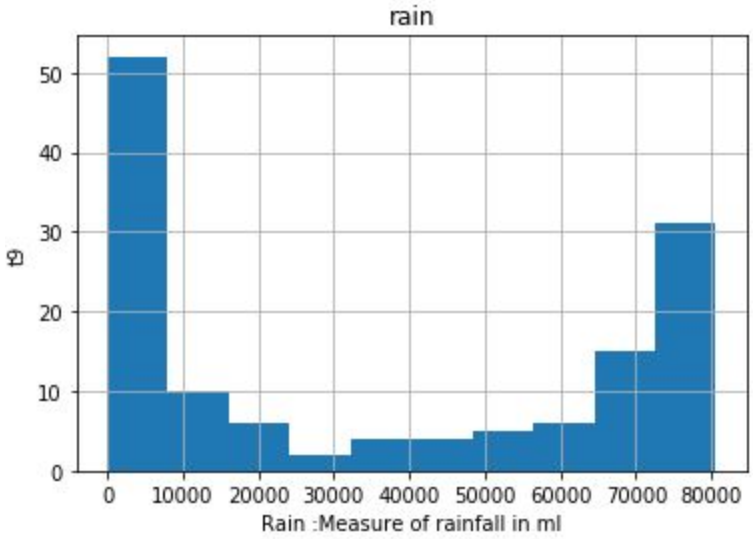
Histogram of attribute ‘rain’ for each of the 10 stations is given as :

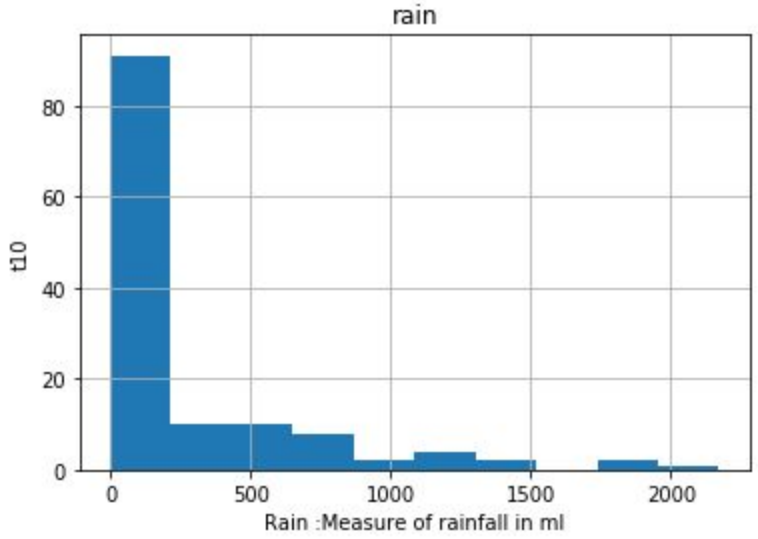
Using this plot we can easily get to know the occurrence of an outcome(here its measure of rainfall) for each individual sensor.

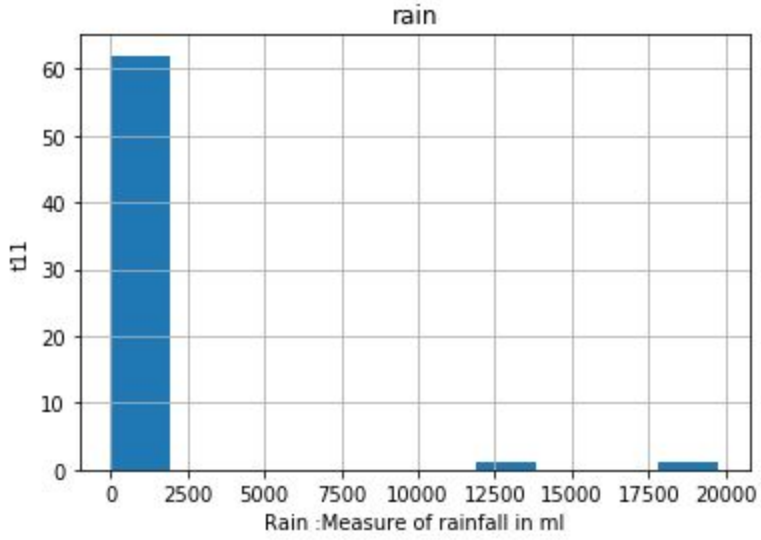

rain
100
80
60
40
20
0
t6
0
250
500
750
1000
1250
1500
1750
Rain :Measure of rainfall in ml

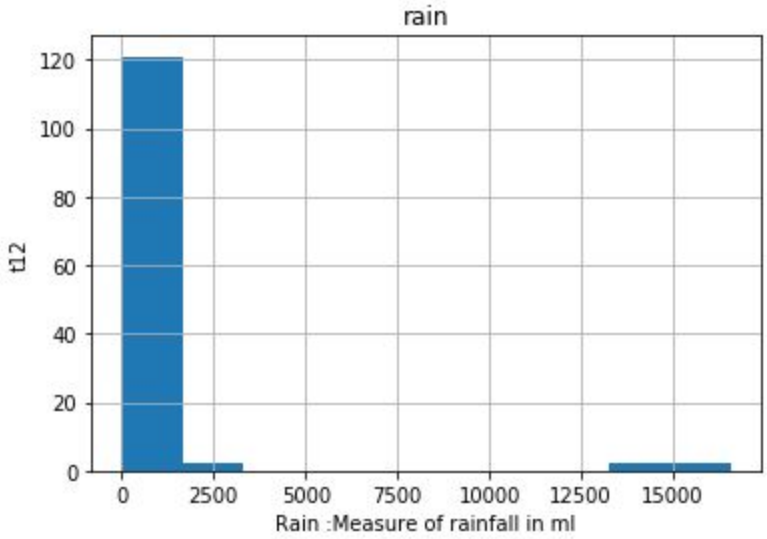

rain
50
40
30
20
10
0
t7
0
10000
20000
30000
40000
50000
60000
Rain :Measure of rainfall in ml

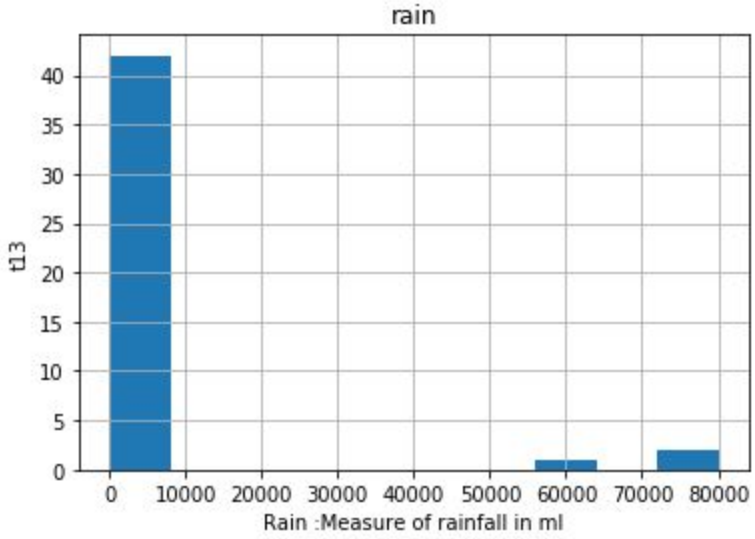

rain
60
50
40
30
20
10
0
t8
0
20000
40000
60000
80000
Rain :Measure of rainfall in ml

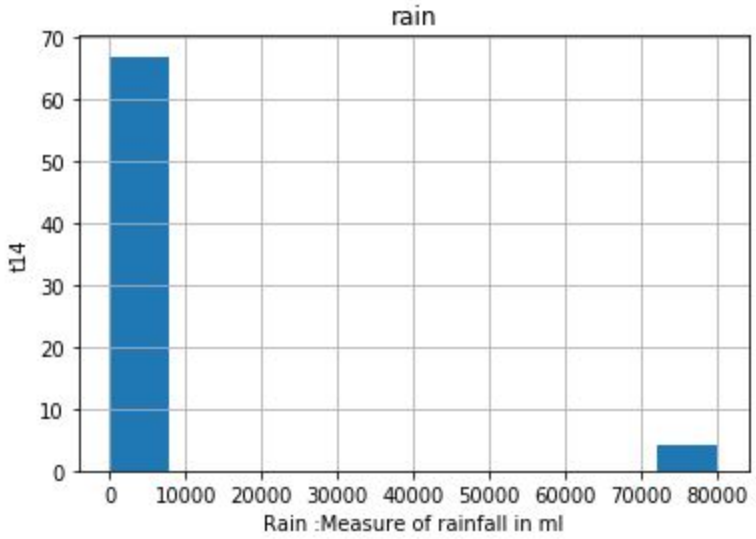

rain
50
40
30
20
10
0
t9
0
10000
20000
30000
40000
50000
60000
70000
80000
Rain :Measure of rainfall in ml

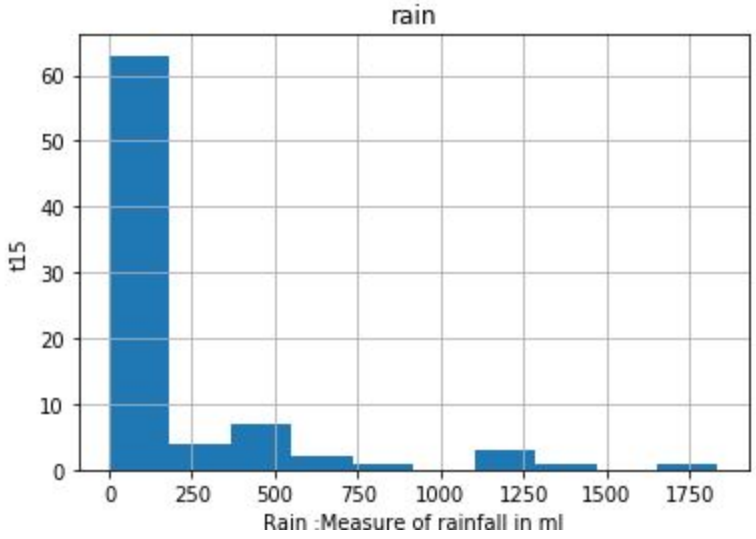

rain
t10
0
20
40
60
80
0
500
1000
1500
2000
Rain :Measure of rainfall in ml


rain
t11
0
10
20
30
40
50
60
0
2500
5000
7500
10000
12500
15000
17500
20000
Rain :Measure of rainfall in ml


rain
t12
120
100
80
60
40
20
0
0
2500
5000
7500
10000
12500
15000
Rain :Measure of rainfall in ml


rain
t13
40
35
30
25
20
15
10
5
0
0
10000
20000
30000
40000
50000
60000
70000
80000
Rain :Measure of rainfall in ml


rain
t14
0
10
20
30
40
50
60
70
0
10000
20000
30000
40000
50000
60000
70000
80000
Rain :Measure of rainfall in ml

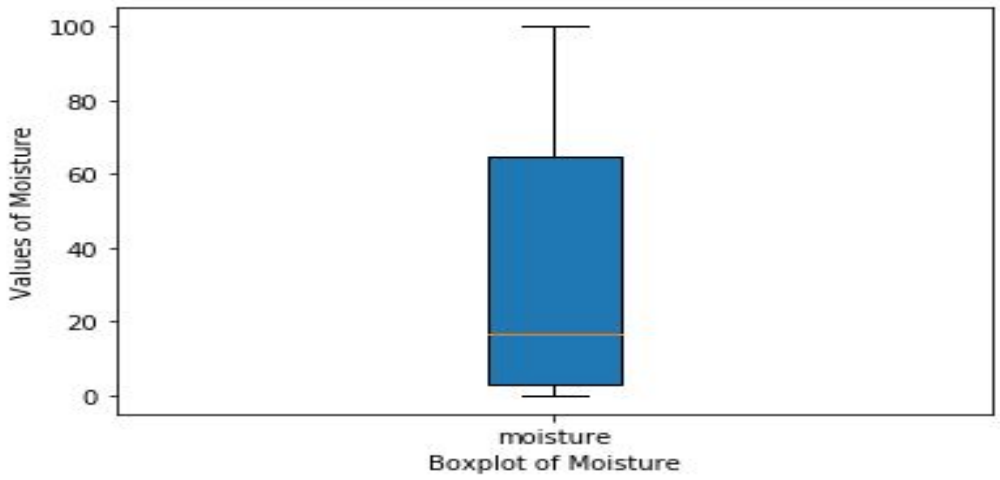

rain
t15
0
10
20
30
40
50
60
0
250
500
750
1000
1250
1500
1750
Rain :Measure of rainfall in ml

Ques 6 :

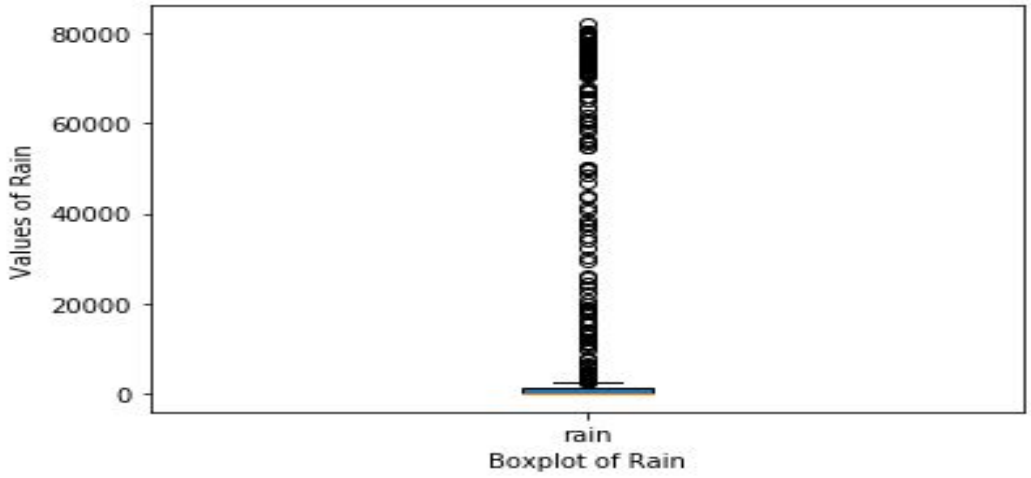
Boxplot for the attributes ‘rain’ and ‘moisture’ is given as :

Boxplot: It is basically a method for graphically representing groups of numerical data through their quartiles.

So from this graphical representation we can easily able to get our 1st quartile, 3rd quartile and interquartile range for a given data set.



Boxplot of Moisture



Boxplot of Rain