IC272: Data Science III

Report of Lab 1: Data visualization and statistics from data

Name: Prashant Kumar

Registration Number: B19101

Mobile Number: 8700350173

Using python, we have calculated mean, median, mode, maximum and minimum value and standard deviation of each attribute.

The output of is shown in figure 1.

Attributes	Mean	Median	Mode	Standard Deviation	Max	Min
Temp.	21.2149	22.272	12.727	4.355818	31.375	7.673
Humidity	83.4799	91.381	99.000	18.21006	99.720	31.00
Pressure	1009.01	1014.7	789.39	46.98047	1079.16	452.1
Rain	10701.5	18.000	0.0000	24857.25	82037.2	0.000
Lightavgw/o0	4438.43	1656.8	4488.9	7573.162	54612.0	0.000
Lightmax	21788.6	6634.0	4000.0	22064.99	54612.0	2259
Moisture	32.3861	16.704	0.0000	33.65324	100.000	0.000

Figure 1

In 2nd we plot the scatter plot of rain and temperature with each attribute.

Scatter Plot of Rain with Other Attribute

A) Rain with temperature:

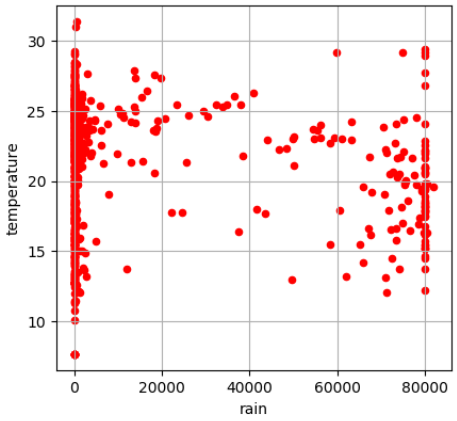

Figure 2(a)

The scatter plot between rain and temperature is show in figure 2(a).

Observation: From the graph in figure 2(a) it shows that there is not any linearity between rain and temperature. Therefore, there is no correlation between them and it is hard to form any regression line. Also, the correlation coefficient that we calculated between rain and temperature in third question is approx. -0.1089 that also shows that there is no correlation between them.

B) Rain with humidity:

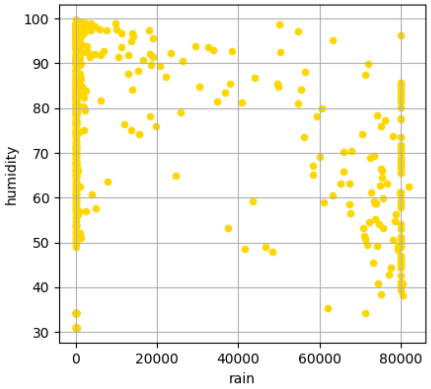

Figure 2(b)

The scatter plot between rain and humidity is show in figure 2(b).

Observation: From the graph in figure 2(b) it shows that there is not any linearity between rain and humidity. Therefore, there is no correlation between them and it is hard to form any regression line. However, the correlation coefficient that we calculated between rain and humidity in third question is approx. -0.4349 that shows that there is some correlation between them.

C) Rain with pressure

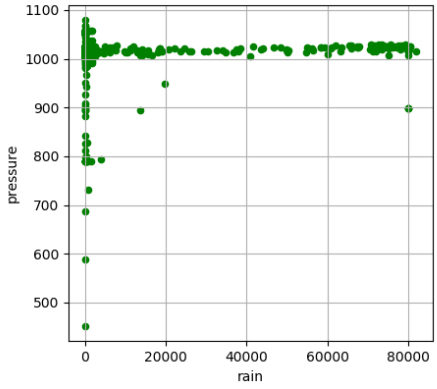

Figure 2(c)

The scatter plot between rain and pressure is show in figure 2(c).

Observation: From the graph in figure 2(c) it shows that there is not any linearity between rain and pressure. Therefore, there is no correlation between them However, a regression line is form that will remain constant or almost parallel to x-axis so that pressure is independent of rain. Also, the correlation coefficient that we calculated between rain and pressure in third question is approx. 0.071 that also shows that there is no correlation between them.

D) Rain with lightavgw/o0

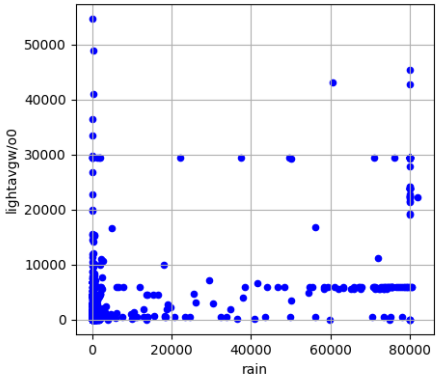

Figure 2(d)

The scatter plot between rain and lightavgw/o0 is show in figure 2(d).

Observation: From the graph in figure 2(d) it shows that there is no linearity between rain and lightavgw/o0. Therefore, there is no correlation between them as lightavvge/o0 is independent of rain. However, the correlation coefficient that we calculated between rain and humidity in third question is approx. 0.5275 that shows that there is some correlation between them.

E) Rain with lightmax

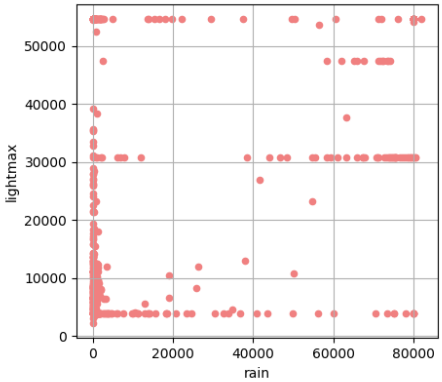

Figure 2(e)

The scatter plot between rain and lightmax is show in figure 2(e).

Observation: From the graph in figure 2(e) it shows that there is not any linearity between rain and lightmax. Therefore, there is no correlation between them and it is hard to form any regression line. Also, the correlation coefficient that we calculated between rain and lightmax in third question is approx. 0.312 that also shows that there is not much correlation between them.

F) Rain with moisture

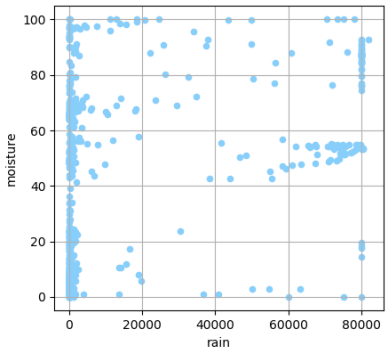

Figure 2(f)

The scatter plot between rain and moisture is show in figure 2(f).

Observation: From the graph in figure 2(f) it shows that there is not any linearity between rain and moisture. Therefore, there is no correlation between them and it is hard to form any regression line. However, the correlation coefficient that we calculated between rain and moisture in third question is approx. 0.4269 that shows that there is some correlation between them.

Scatter Plot of Temperature with Other Attributes

A) Temperature with humidity

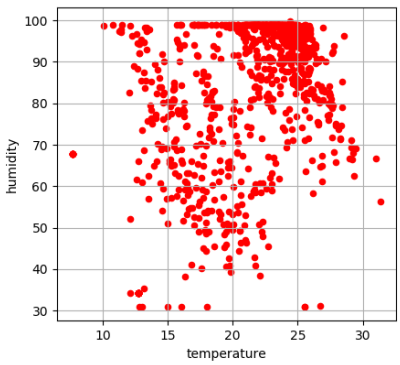

Figure 3(a)

The scatter plot between temperature and humidity is show in figure 3(a).

Observation: From the graph in figure 3(a) it shows that there is not any linearity between temperature and humidity. Therefore, there is no correlation between them and it is hard to form any regression line. However, the correlation coefficient that we calculated between rain and humidity in third question is approx. 0.4016 that shows that there is small relation between them.

B) Temperature with pressure

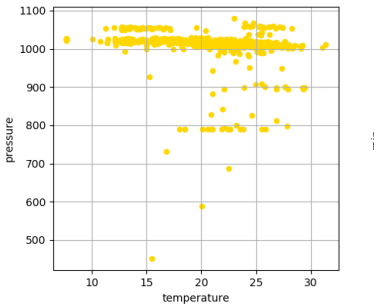

Figure 3(b)

The scatter plot between temperature and pressure is show in figure 3(b).

Observation: From the graph in figure 3(b) it shows that there is no linearity between temperature and pressure. Therefore, there is no correlation between them. However, a regression line is form that will remain constant or almost parallel to x-axis so that pressure is independent of temperature. Also, the correlation coefficient that we calculated between temperature and pressure in third question is approx. -0.1814 that also shows that there is not much relation between them.

C) Temperature with rain

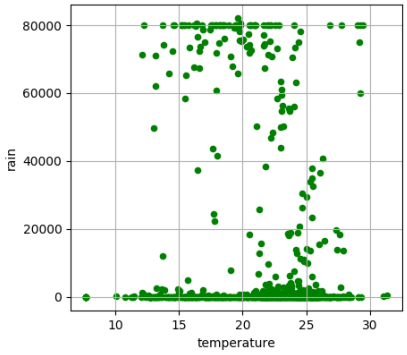

Figure 3(c)

The scatter plot between rain and temperature is show in figure 3(c).

Observation: From the graph in figure 3(c) it shows that there is not any linearity between rain and temperature. Therefore, there is no correlation between them. Also, the correlation coefficient that we calculated between rain and temperature in third question is approx. -0.1089 that also shows that there is no correlation between them.

D) Temperature with lightavgw/o0

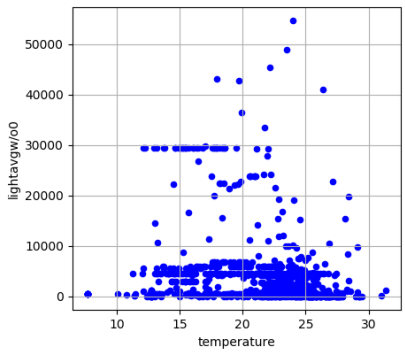

Figure 3(d)

The scatter plot between temperature and lightavgw/o0 is show in figure 3(d).

Observation: From the graph in figure 3(d) it shows that there is not any linearity between temperature and lightavgw/o0. Therefore, there is no correlation between them. Also, the correlation coefficient that we calculated between temperature and lightavgw/o0 in third question is approx. -0.1814 that also shows that there is no correlation between them.

E) Temperature and lightmax

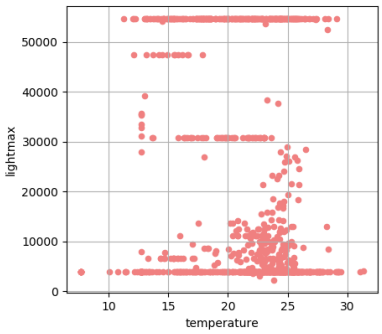

Figure 3(e)

The scatter plot between temperature and lightmax is show in figure 3(e).

Observation: From the graph in figure 3(e) it shows that there is not any linearity between temperature and lightmax. Therefore, there is no correlation between them and it is hard to form any regression line. Also, the correlation coefficient that we calculated between temperature and lightmax in third question is approx. -0.1459 that also shows that there is no correlation between them.

F) Temperature and moisture

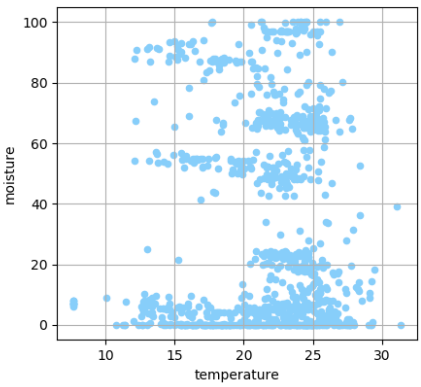

Figure 3(f)

The scatter plot between temperature and moisture is show in figure 3(f).

Observation: From the graph in figure 3(f) it shows that there is not any linearity between temperature and moisture. Therefore, there is no correlation between them and it is hard to form any regression line. Also, the correlation coefficient that we calculated between temperature and moisture in third question is approx. 0.0807 that also shows that there is no correlation between them.

Histogram of Rain and Temperature

A) Histogram of Rain

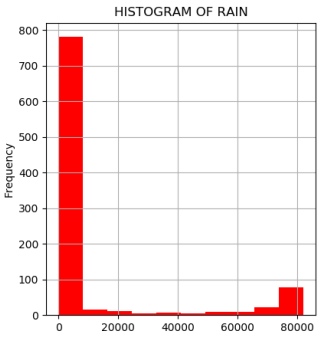

Figure 4(a)

Observation: In the histogram of rain as shown in figure 4(a) shows that the rainfall between 0-8000(in ml) have high frequency that shows that in each station there is mostly 0-8000ml of rainfall happen and there is less rainfall between 8000-80000ml approximately.

B) Histogram of Moisture

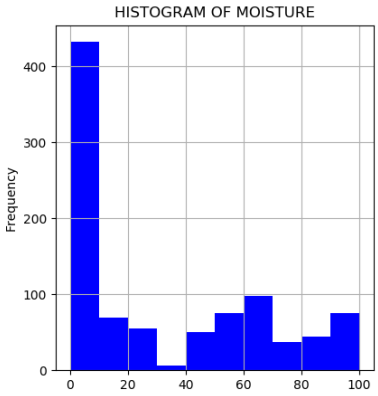

Figure 4(b)

Observation: In the histogram of moisture that indicates the water stored in the soil (measured between 0 and 100 percent) as shown in figure 4(b) shows that there is high frequency of moisture of up to 10 percent and almost zero frequency for 30-40%. Other range of percent have least frequency of moisture.

Histogram of rain for each station

A) Station T10

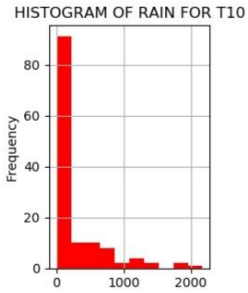

Figure 5(a)

Observation: In the histogram of rain for station T10 as shown in figure 5(a) that the rainfall between 0-200(in ml) approx. has high frequency as compared to 200-2100 ml of rainfall.

B) Station T11

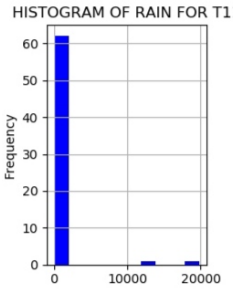

Figure 5(b)

Observation: In the histogram of rain for station T11 as shown in figure 5(b) that the rainfall between 0-200(in ml) approx. has high frequency as compared to 200-20000 ml of rainfall.

C) Station T12

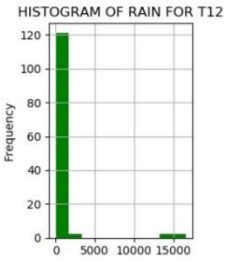

Figure 5(c)

Observation: In the histogram of rain for station T12 as shown in figure 5(c) that the rainfall between 0-2000(in ml) approx. has high frequency as compared to 2000-18000 ml of rainfall at station T12.

D) Station T13

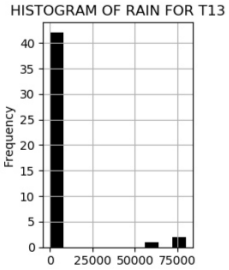

Figure 5(d)

Observation: In the histogram of rain for station T13 as shown in figure 5(d) that the rainfall between 0-10000(in ml) approx. has high frequency as compared to 10000-75000 ml of rainfall at station T13. For 10000-52000 has almost zero frequency.

E) Station T14

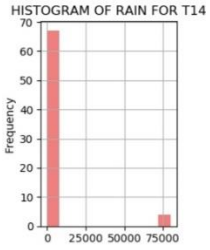

Figure 5(e)

Observation: In the histogram of rain for station T14 as shown in figure 5(e) that the rainfall between 0-10000(in ml) approx. has high frequency as compared to 10000-75000 ml of rainfall at station T14. For 10000-73000 has zero frequency.

F) Station T15

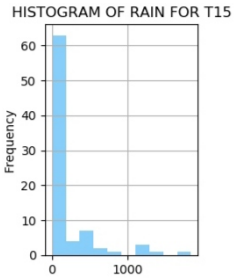

Figure 5(f)

Observation: In the histogram of rain for station T15 as shown in figure 5(f) that on station T15 there is a rainfall between 0-182(in ml) approx. has high frequency and there is very less rainfall at station T14 beyond 190 ml.

G) Station T6

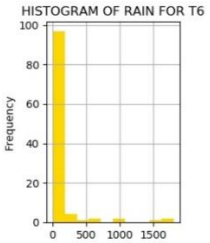

Figure 5(g)

Observation: In the histogram of rain for station T6 as shown in figure 5(g) that on station T6 there is a rainfall between 0-200(in ml) approx. has high frequency and there is very less rainfall at station T6 beyond 200 ml as its frequency is very less as shown in figure 5(g).

H) Station T7

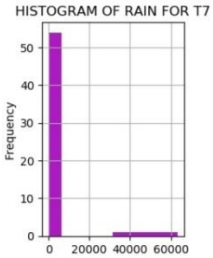

Figure 5(h)

Observation: In the histogram of rain for station T7 as shown in figure 5(h) that the rainfall between 0-8000(in ml) approx. has high frequency as compared to 8000-61000 ml of rainfall at station T7.

I) Station T8

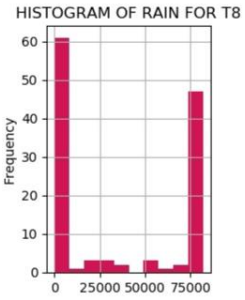

Figure 5(i)

Observation: In the histogram of rain for station T8 as shown in figure 5(i) that the rainfall between 0-8333(in ml) and rainfall between 75000-80000(in ml) approx. has high frequency as compared to 833375000 ml of rainfall at station T8.

J) Station T9

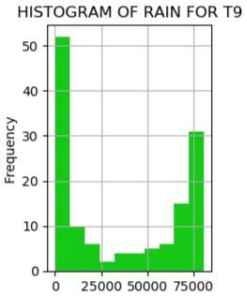

Figure 5(j)

Observation: In histogram of rain for station T9 as shown in figure 5(j) that the rainfall between 0-8333(in ml) approx. has high frequency and also the rainfall between 74000-77000(in ml) approx. has also good frequency as compared to other ranges of rainfall. Hence, at station T9 we can say that there is high rainfall as compared to other stations.

Box Plot of Rain and Moisture

A) Box plot of rain

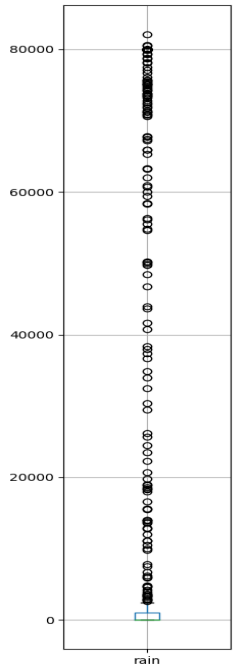

Figure 6(a)

Observation: The box plot of rain as shown in figure 6(a). It is clear that there is too much outliners in the box plot of rain which implies that the data of rain that we get from different sensers from each station is too noisy.

B) Box plot of moisture

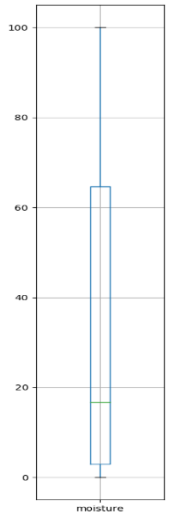

Figure 6(b)

Observation: The box plot of moisture as shown in figure 6(b) that there are no outliners in the box plot of moisture due to which data is not noisy. This also happen because the moisture was read in percentage which ranges between 0-100 but Q1 and Q2 is lie below i.e. 50% 0f data has value less than 20% which indicates that in most of the station water stored in the soil is less than 20%.