REAL WORLD ANALYTICS ASSIGNMENT -3

Understand the data:

The text file is downloaded and saved to the R working directory.

The data is assigned to the matrix by using the given R command

To investigate the variable of interest (Y) a subset of 300 data is generated using the given R command

The data is visualized using the scatter plots and histograms for the given variables including (Y)

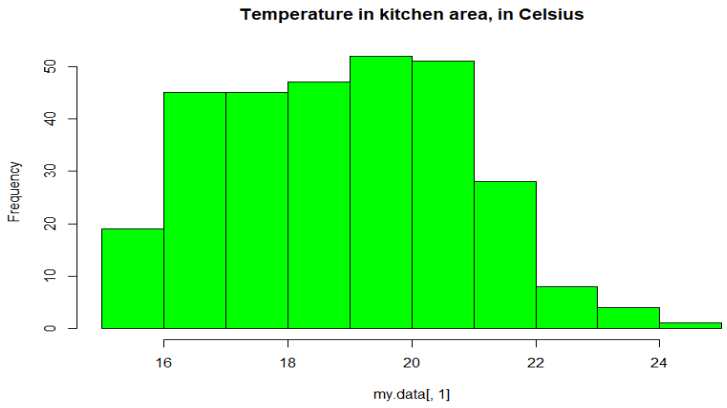

Fig 1

The temperature in the kitchen frequency is high 20 and it is low at 25. The highest frequency is 50.

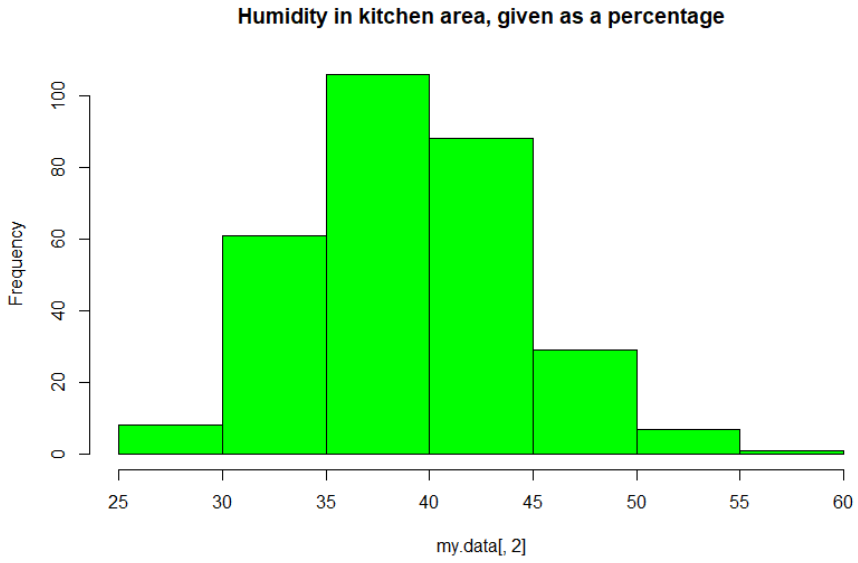

Fig 2

The frequency of humidity in the kitchen area is high at 40 and low at 60 the highest frequency is 100

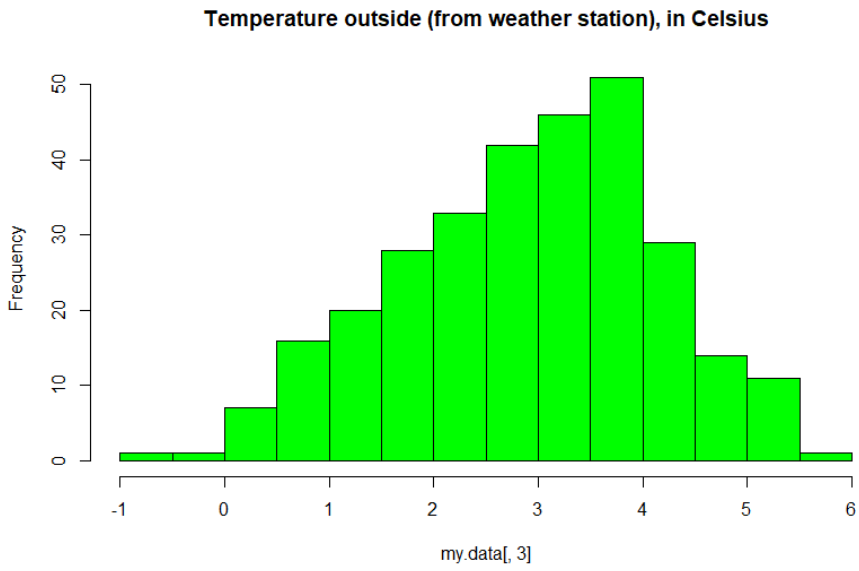

Fig 3

The frequency of temperature outside is high at 4 degree Celsius and low at 6 degree Celsius. The highest frequency is 50.

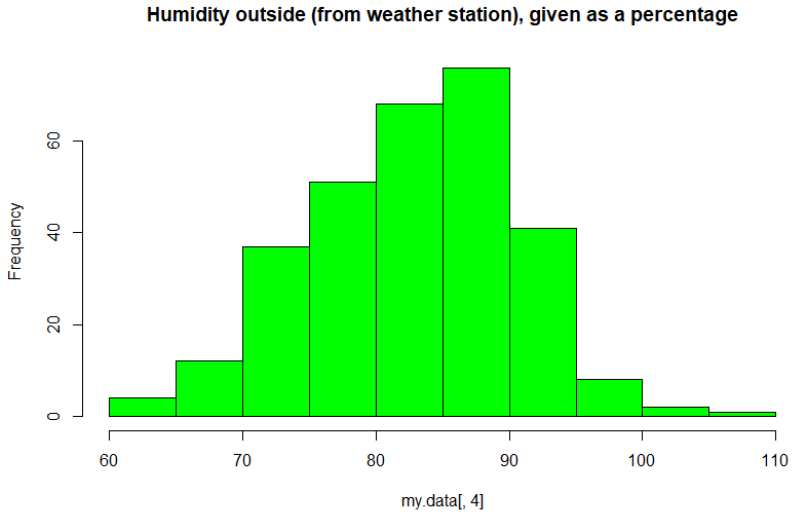

Fig 4

The frequency of humidity outside is high at 90 percent and at 110 percent. The maximum frequency is more than 60

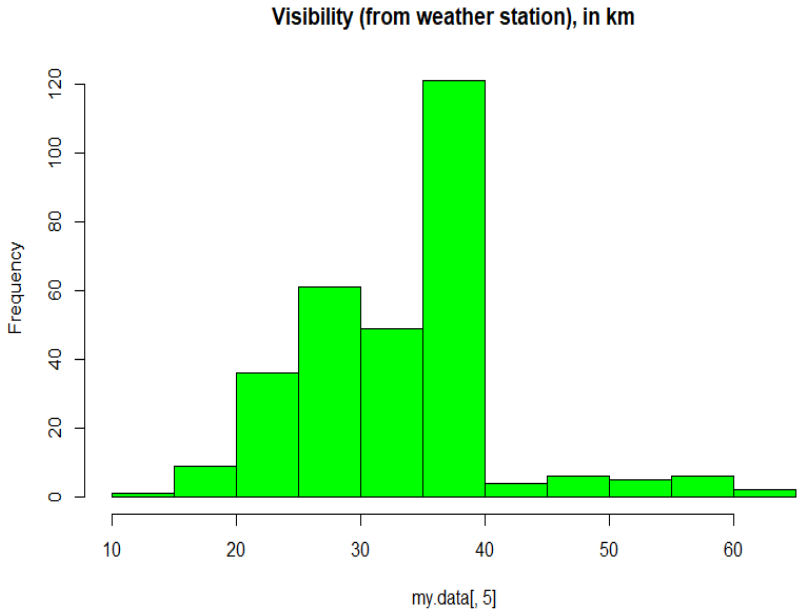

Fig 5

The highest visibility frequency is at 40 kilometers and the low at 60 kilometers

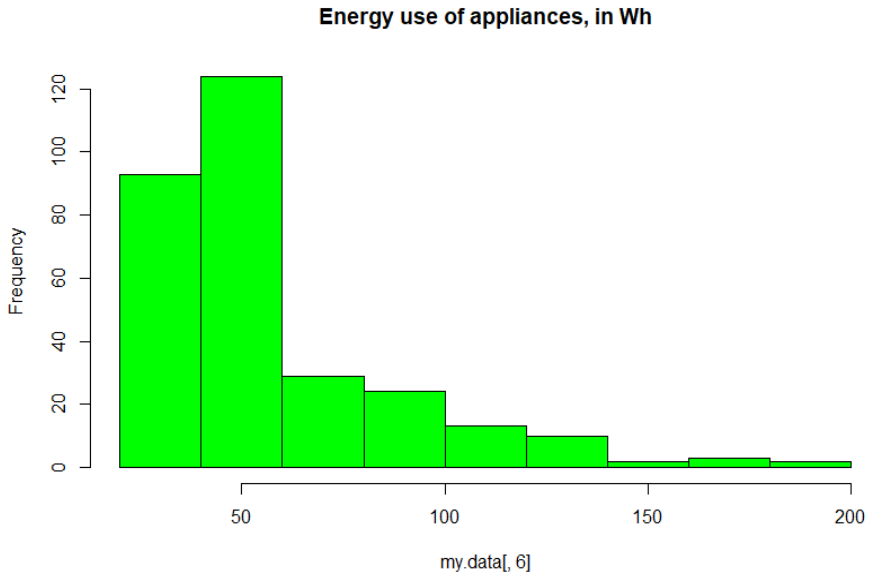

Fig 6

The frequency of energy used by appliances is more at 50 watts and low at 200 watts

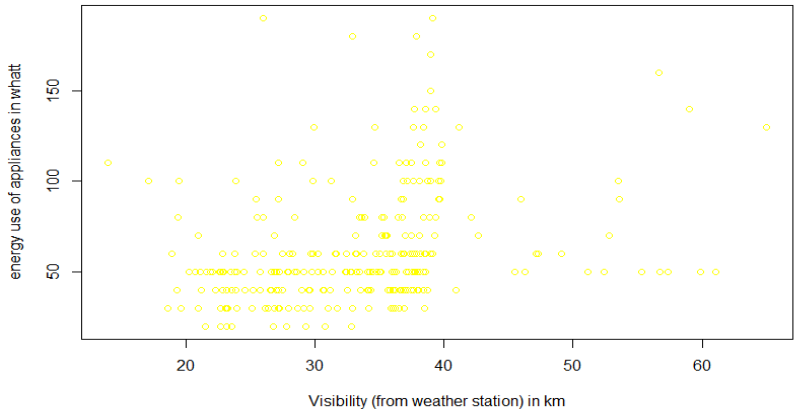

Fig 2.1

The visibility is more between 20 km to 30km and energy use of appliances is more at 50watts. The relation is weak between visibility and energy usage of appliances.

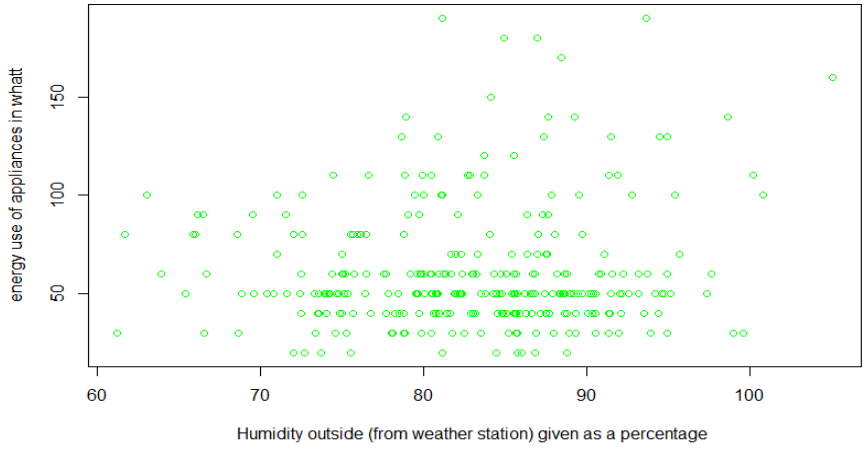

Fig 2.2

There are wide outliers, so we can say that relation in weak between energy usage and humidity.

The energy usage of appliances is more at 50when the humidity ranges between 80 to 90

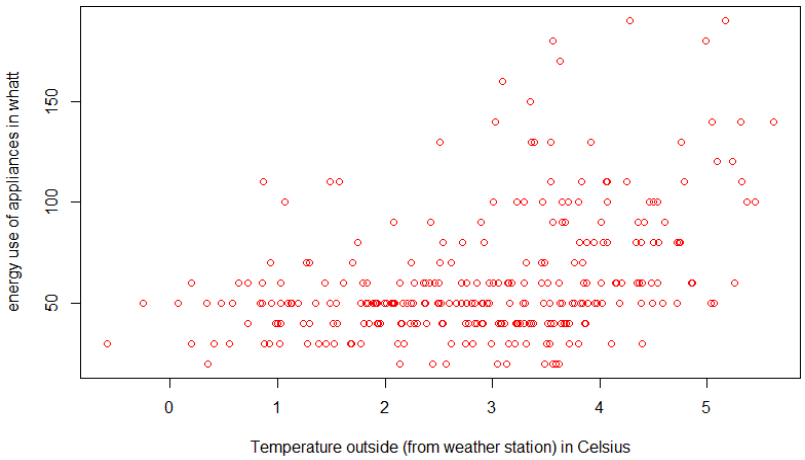

Fig 2.3

There are many outliers so the relation between energy and temperature outside is weak. Energy usage is more when the temperature is between 2 and 4

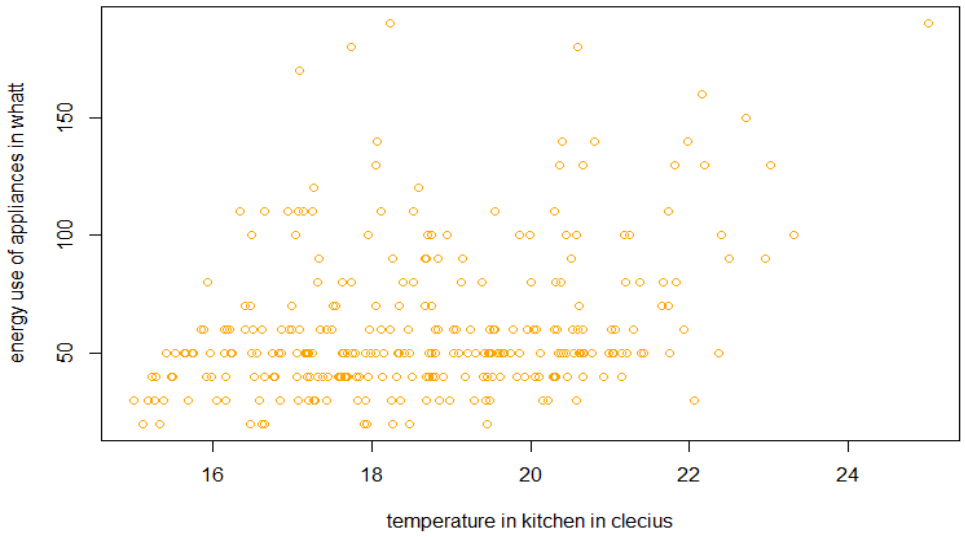

Fig 2.4

The general relation between temperature in kitchen and energy use of appliances is weak it has many outliers. Energy is more at 50 and temperature in kitchen is more between 16 to 20

Transform the data

I have chosen four variables X1, X3, X4, X5 and variable of interest Y to apply transformations. I have applied polynomial transformation for X1, X3, X5, and variable of interest Y and scaling transformation for X4 and the obtained values are aggregated and predicted the variable of interest. I have assigned

the transformed data along with the transformed variable of interest to an array which has 300 rows and 5 columns, and I have saved it by name “pranav-transformed.txt” by using the given line of code.

X1 vs Y in this transformation we came to know that temperature in kitchen and energy used by appliances

X3 vs Y in this transformation we came to know that temperature outside and energy used by appliances

X4 vs Y in this transformation we came to know that humidity and energy used by appliances

X5 vs Y in this transformation we came to know that visibility and energy used by appliances

Build models and investigate the importance of each variable

I have downloaded the AggWaFit718.R file to my working directory and loaded into the R workspace using the given R command

I have applied all the specified fitting functions onto the transformed data.

1. Weighted Arithmetic Mean (WAM)
2. Weighted Power Mean with p = 0.5 (WPM05)
3. Weighted Power Mean with p = 2 (WPM2)
4. Ordered Weighted Averaging Function (OWA)
5. Choquet Integral (CI)

Error Measures and correlation:

	WAM	**PM05**	**PM2**	**OWA**	**Choquet**
RMSE	9.741835477347	9.74183547734	9.7418354773	9.6171851948	9.6171851946
Av. abs error	8.523949495851	8.52394958520	8.5239494958	8.4060200844	8.4060200843
Pearson correlation	0.048443660763	0.48443660763	0.0484436607	0.3907220759	0.3907220758
Spearman correlation	0.000311859303	0.00031185930	0.0003118593	0.2877380008	0.2875494040
Orness				1.0000000000	0.9999999999

Weights Parameters(w_i):

w_i	**WAM**	**PM05**	**PM2**	**OWA**	**CI Shapely Val**
X1	0	0	0	0	0.249999999
X3	0	0	0	0	0.249999999
X4	0.99999999999	0.99999999999	0.99999999999	0	0.249999999
X5	0	0	0	1.000000000	0.249999999

Number	Choquet Weight
1	0.999999999987197
2	2 0.999999999999959
3	3 0.999999999983006
4	4 0.999999999992564
5	5 0.999999999840136
6	6 1.00000000001561
7	7 0.999999999637285
8	8 0.999999999995471
9	9 0.999999999977221
10	10 0.99999999999964
11	11 0.999999999982537
12	12 0.999999999995743
13	13 0.999999999821108
14	14 1.00000000001223
15	15 0.999999999615081
16	0.999999999987197

I have included the two tables and additional required information required

I have compared all the variables and among those choquet variables are best suitable for these model and I have also tested with other models from which it is clear that choquet value is between 0 and 0.5 this is the suitable model for the prediction of this data.

Among the variables X1 gives the temperature it tells about the temperature it tells about temperature in kitchen

X3 tells about the temperature in outside kitchen and we can predict the temperature using this variable

X4 tells about the humidity of in the given data we can see the humidity by seeing X4 values

Y tells about the energy usage of the appliances from which we can predict the energy usage of the appliances

The X1 and X5 interaction is complementary because they have higher values and X3 has the low values because it is redundant.

Use your model for Prediction

Choquet model and weighted arithmetic mean is the best fitting model for predicting the energy use of appliances for the given inputs

Low energy use of appliances will occur when the temperature is less and thee humidity is more low energy use of appliances will occur