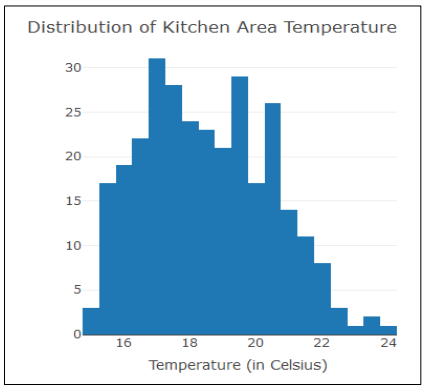
SIT 718: Assignment 3

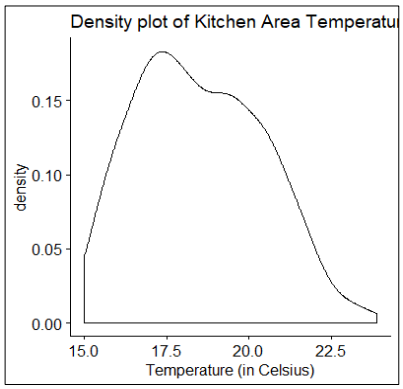
Shrutik Panchal
218412482

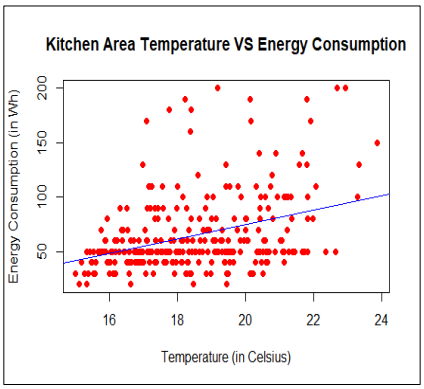
Task 1:

Histograms, scatterplots and summary tables for the given variables are as below:

X1: Temperature in Kitchen Area, in Celsius

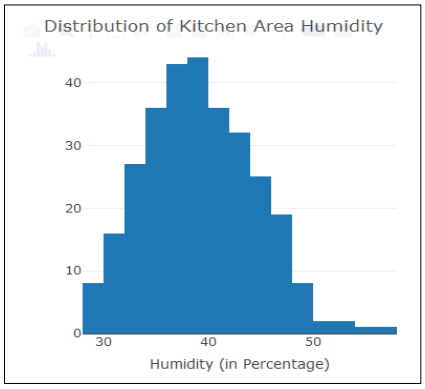


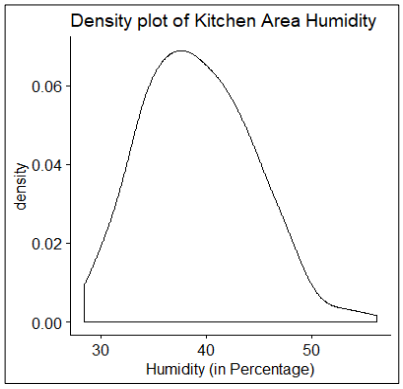


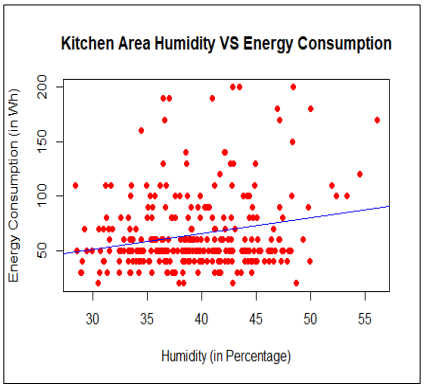


Based on the above histogram and density plot of kitchen area temperature, it is visible that the data is **slightly right skewed** with long tail to higher temperatures though three to four peaks are noticeable as well. For this variable, **mean** and **median** are **18.51** and **18.33** respectively hence **mean > median** clarifies that data is slightly right skewed as there is no major difference. **Pearson skewness score** is **0.285813** hence data is positively skewed. As per scatter plot diagram, we can denote that kitchen area temperature has **positive correlation** with energy consumption hence increase in temperature shows increase in energy consumption as well.

X2: Humidity in Kitchen Area, in Percentage



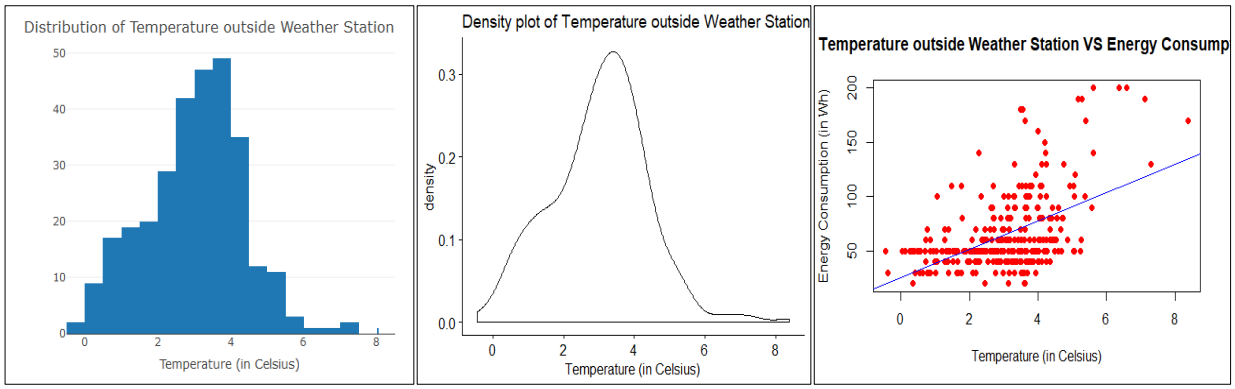




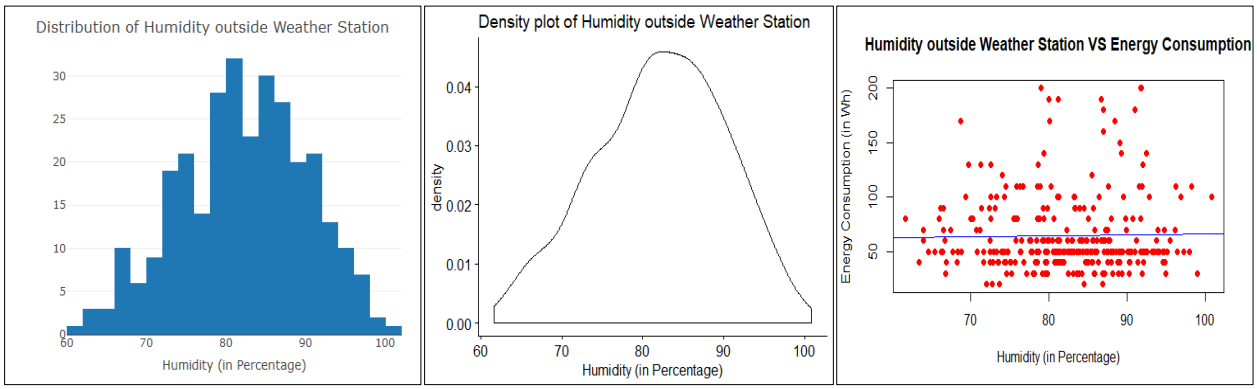
Based on the above histogram and density plot of humidity in kitchen area, it is **slightly right skewed** though majority of the distribution seem to follow bell shaped curve. **Mean** and **median** are **39.18** and **38.76** respectively that denotes data is rightly skewed though there is no huge difference. **Pearson skewness** score is **0.2376266** hence it is positively skewed. The scatter plot diagram denotes that the kitchen area humidity has **low positive correlation** with energy consumption.

X3: Temperature outside Weather Station, in Celsius

Based on the histogram and density plot of temperature outside weather station below, it is noticeable that the distribution has a peak though data around the peak is not normally or evenly distributed. **Mean** and **median** are **3.0115** and **3.1407** respectively hence data seem to have **slightly left skewness**. **Pearson skewness** score is **-0.2832601** that shows the data is **negatively skewed**. The scatter plot shows the **high positive correlation** between both the variables. It is visible in histogram that data has some negative values.

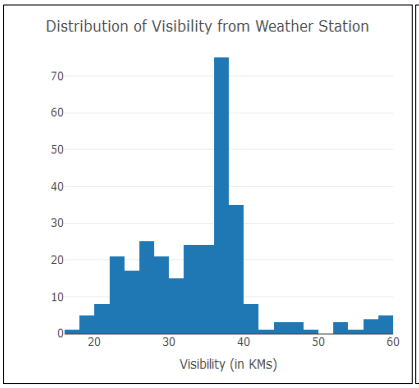


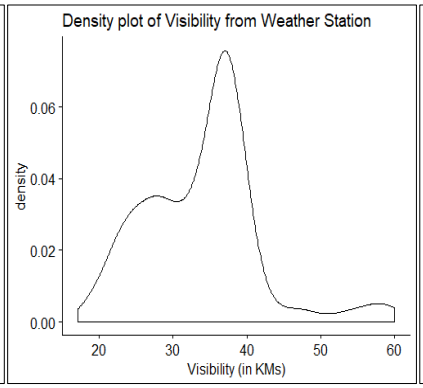
X4: Humidity outside from Weather Station, in Percentage

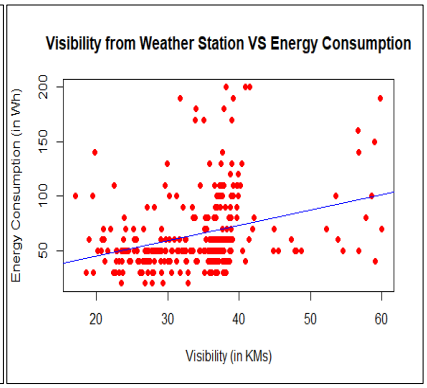


Based on the histogram and density diagram above, it is noticeable that data is **slightly left skewed** with long tail. **Mean** and **median** are **82.17** and **82.56** respectively hence **mean < median** that shows **negative skewness**. **Pearson skewness** score is **-0.1438816** that denotes data is **slightly negatively skewed**. The scatter plot shows that there is **no correlation** between these variables as there is no change in energy consumption with the change in Humidity.

X5: Visibility from Weather Station, in km

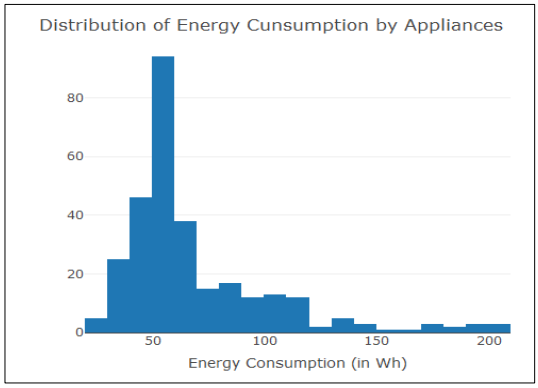


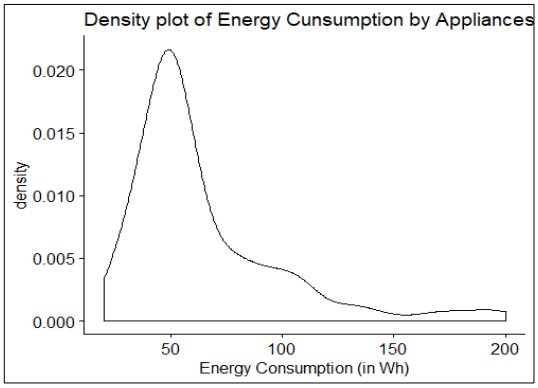




Based on the histogram and density diagram above, it is noticeable that the data is **unimodal** however majority for the distribution is on the left side of the peak hence based on **Pearson Skewness** score which is **-0.6606924**, data is **negatively skewed**. Mean and median are **33.87** and **35.60** respectively and slightly **positive correlation** is visible.

Y: Energy use of Appliances, in Wh





Based on the histogram and density diagram above, it is visible that the data is **right skewed with long tail** and **Pearson Skewness** score is **1.27098** hence data is **positively skewed**. **Mean** and **median** are **64.8** and **50.0** respectively.

Summary of the variables:

Variables	Minimum	Median	Mean	Maximum	Pearson Skewness
X1	15.00	18.33	18.51	23.88	0.285813
X2	28.43	38.76	39.18	56.13	0.2376266
X3	-0.4411	3.1407	3.0115	8.3832	-0.2832601
X4	61.65	82.56	82.17	100.82	-0.1438816
X5	17.11	35.60	33.87	59.97	-0.6606924
Y	20.00	50.00	64.8	200.00	1.27098

Task 2:

Here, transformations are applied to all the variables however X1, X2, X3 and X4 will be used with Variable Y for the further analysis.

Variable X1: As data is slightly right skewed, Log transformation has been used.

Variable X2: As data is slightly right skewed, Log transformation has been used.

Variable X3: There are negative values in data hence rescaled them using "value + 1 – minimum (column)" then applied power transformation with power 1.5 (getting as close as possible to normal distribution).

Variable X4: As data is slightly left skewed, Power transformation has been used with power 2 (Square Transformation).

Variable X5: As data is negatively skewed, Power Transformation has been used with power 3 (Cube Transformation).

Variable Y: As data has high right skewness, first log transformation has been used and later, power transformation with power 0.001 (as there was no any significant difference by decreasing the value more) has been applied.

After the transformation applied, data were normalized using formula:
(Data Column – minimum (Data Column)) / (maximum (Data Column) – minimum (Data Column))

Summary of the variables after the transformations and normalization:

Variables	Minimum	Median	Mean	Maximum	Pearson Skewness
X1	0.0000	0.4312	0.4411	1.0000	0.1340443
X2	0.0000	0.4558	0.4584	1.0000	0.04024345
X3	0.0000	0.2956	0.2931	1.0000	-0.05190983
X4	0.0000	0.4738	0.4739	1.0000	0.002103771
X5	0.0000	0.19039	0.19192	1.0000	0.02756271
Y	0.0000	0.4680	0.5220	1.0000	0.8386953

Only X1, X2, X3, X4 and Y will be used in further analysis.

For Pearson Skewness Calculation following alternate formula has been used:

Formula: **(3 * (Mean (Data Column) – Median (Data Column))) / (standard deviation (Data Column))**

Task 3:

Table 1: Summary - Error Measures and Correlation Coefficients

Fitting Functions	Error Measures		Correlation	
	RMSE	Av. abs error	Pearson	Spearman
WAM	0.194740172731476	0.153654543353769	0.433233614925839	0.346231055994134
WPM (p=0.5)	0.206112828715406	0.163235784775009	0.410126554684282	0.323022848533575
WPM (p=2.0)	0.182314623978056	0.14253411290233	0.470528818479753	0.389007873832748
OWA	0.172190856114445	0.134261743517928	0.469526551906281	0.391473005463871
Choquet	0.160204530673553	0.123758655452769	0.579475393733826	0.502371311116107

Table 2: Weights and Parameter Summary

Fitting Functions	Weights/Shapley			
	X1	X2	X3	X4
WAM	0.305375491560073	0.308596932933558	0.121829040122349	0.26419853538402
WPM (p=0.5)	0.29721463307916	0.308385541210519	0.0967081227169895	0.297691702993331
WPM (p=2.0)	0.33095280439716	0.260490877599195	0.214322423152443	0.194233894851202
OWA	0.165837710911044	0.17401593212446	0.120051692066584	0.540094664897913
Choquet	0.242381384168655	0.104961388844636	0.479558654585187	0.173098572401512

As per the table 1, it is clearly visible that Choquet performs the best in the given scenario comparted to other fitting functions. Choquet has the lowest error measure values for Root Mean Square Error as well as Average Absolute Error though model is least accurate with Weighted Power Mean where p=0.5. OWA seems to perform better than WAM and WPM functions though there is noticeable difference between Choquet accuracy and OWA. Pearson and Spearman correlation values are far better with Choquet fitting compared to other functions though it shows significant difference where other functions have Correlation below 0.5, Choquet function has Pearson and Spearman correlation values greater than 0.5 that shows high positive relationship.

As per the table 2, it is noticeable that all the fitting functions assign different weights to different attributes however if we compare our scatter plot diagrams (from task 1) to this weights assigned, Choquet function seem to be doing better job applying respected weights to attributes as X3 has high positive regression with Y hence logically it should be given higher weights and following it X1 should receive the second highest weight. Compared to this, other fitting functions doesn't seem to provide sensible outputs/weights to attributes.

X3 and X1 covers more than 70 percent of weightage according to Choquet function. As per the Choquet fuzzy measures below, we can say that there are nearly 4 measures with value as 1.

As per the fuzzy measures collected with Choquet function,
X1 and X2 are additive however relationship with other variable relations such as X1 to X3, X1 to X4, X2 to X3, X2 to X4 and X3 to X4 are somewhat redundant.

Orness score for Choquet and OWA is 0.753750530061303 and 0.678134436983788 respectively hence our model favours the higher inputs as both the scores are higher than 0.5.

Choquet Fuzzy Measures				
1: 0.647386170104526	4: 0.730216876904541	7: 0.999999999999999	10: 0.53291257311322	13: 1
2: 0.354665086496711	5: 1.00000000000001	8: 0.53291257311322	11: 0.649564608305823	14: 0.975907973457021
3: 0.647386170104526	6: 0.925758283550086	9: 0.64956460830582	12: 0.975907973457025	15: 0.99999999999999

Task 4:

For the considered data, model seems to perform the best with Choquet compared to other fitting functions hence Choquet function has been used for the further analysis.

After considering the given input, appropriate transformation and normalization were applied hence received normalized output is 0.4809877. After undoing normalization and transformations, the final Energy Consumption is 51.48105.

Considering “Kitchen Area Temperature VS Energy Consumption” and “Temperature outside Weather Station VS Energy Consumption” scatter plot diagrams, Y is approximately 60 for X1 = 18 and Y is approximately 75 for X3 = 4 hence 75 + 60 = 135 though as discussed in Task 3 that X1 and X3 seems to have more than 70 of weightage of distribution hence (135/2) * 0.7 = 47.25. This denotes that theoretically based on X1 and X3 our Energy consumption should be around 47.25 and our predicted output is approximately 51.5 Wh. Hence, model performance seems reasonable.

As per the scatter plot diagrams and weightages received, we can say that X4 does not have noticeable relation with energy consumption though X1, X2 and X3 seem to have positive relation with the energy consumption hence increase in the values of these variables will affect and increase the Energy consumptions. To get a low energy usage of appliances, temperature in the kitchen area (recommended to have less than 16) and outside weather station (recommended to have less than 2) should be least as possible and kitchen area humidity (recommended to have less than 35) should be low as well.

----------X----------