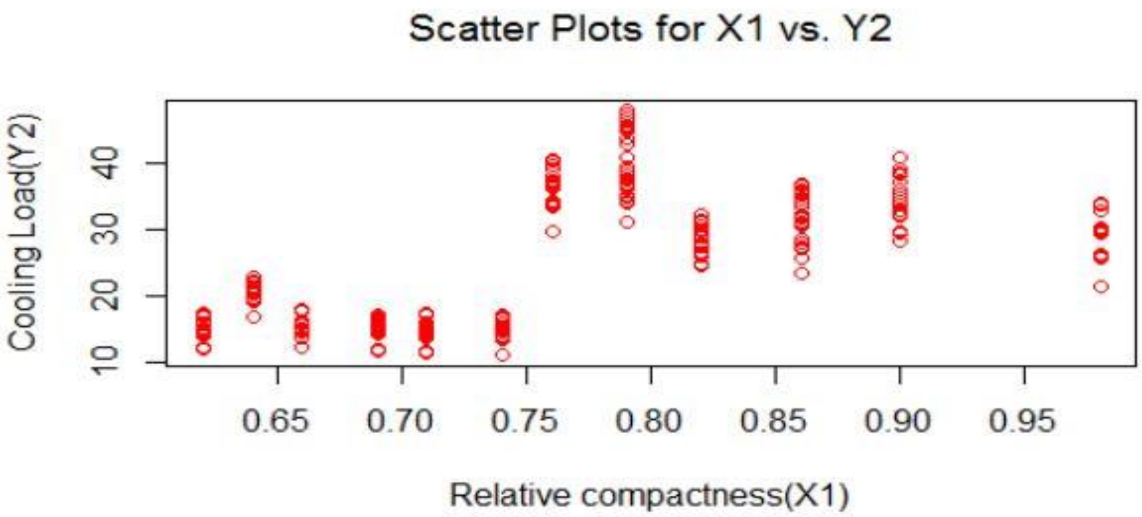
# SIT718 Real world Analytics Assignment

**Name:** Shantanu Gupta

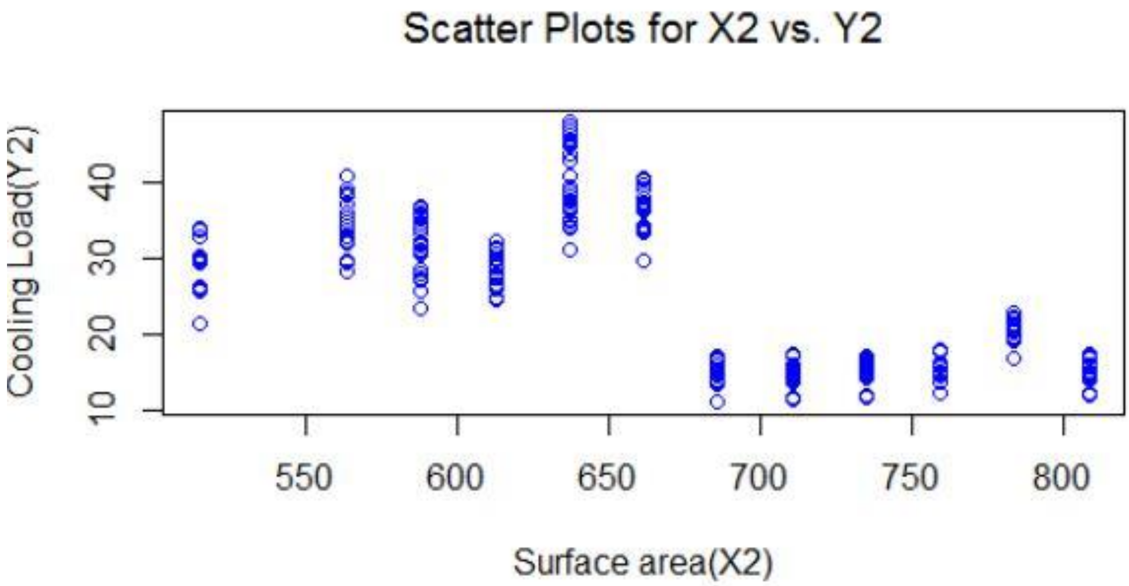
**Student ID:** 218200234

# PART-A

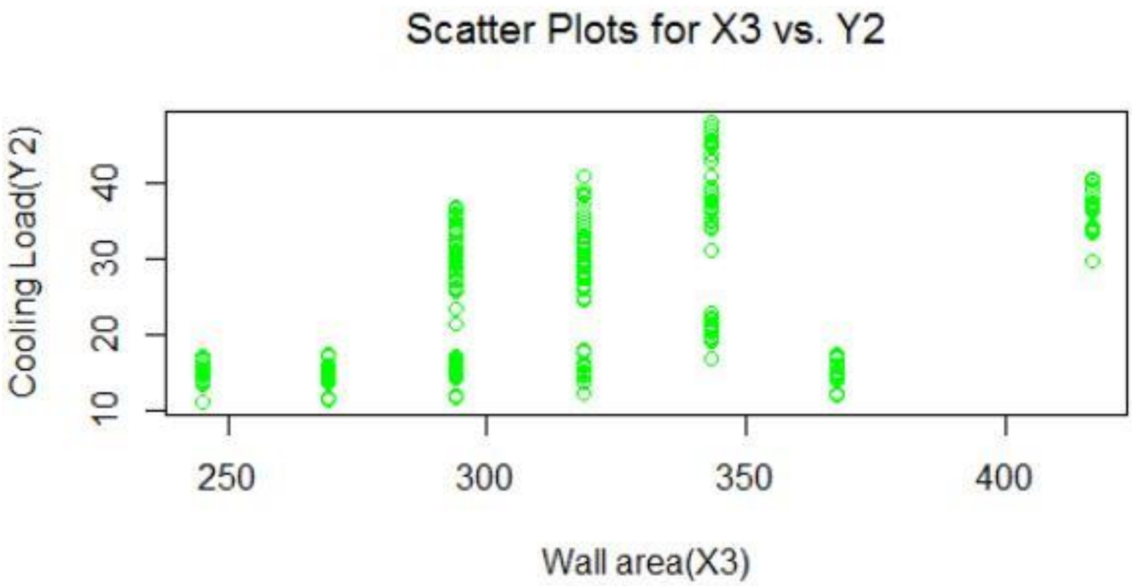
1. (iv)



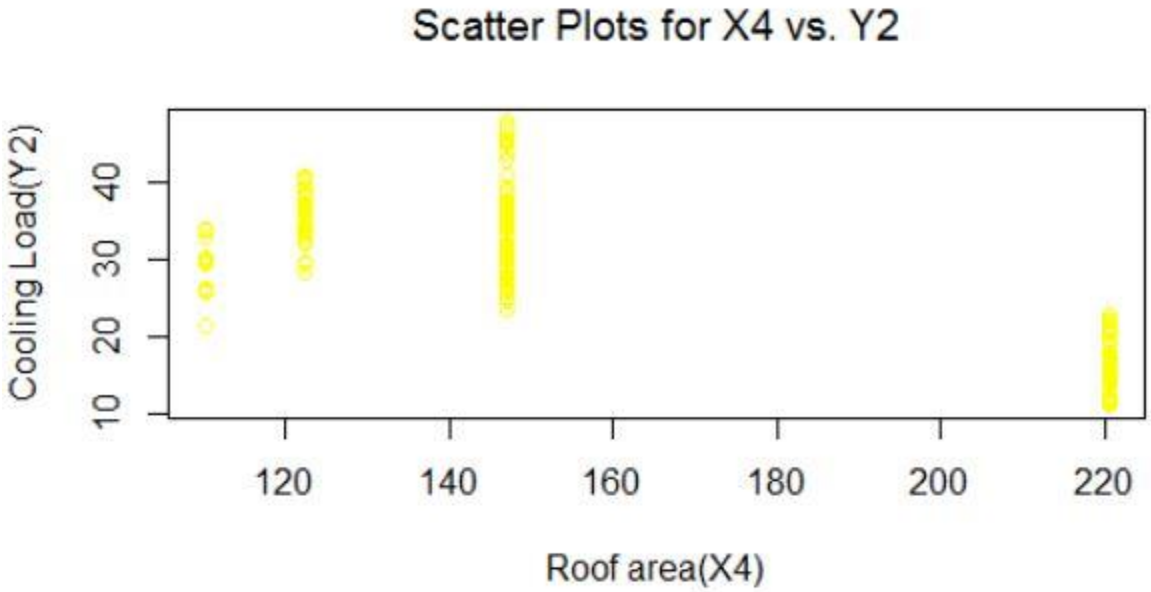

Our aim is to understand more about the relationship between the relative compactness conditions and the cooling load. We would expect that building should be as compact as possible, & will require less cooling and lead maintain comfortable conditions. After X1=0.75, the value of cooling load (Y2) becomes too high. The Relative compactness is showing the positive correlation.



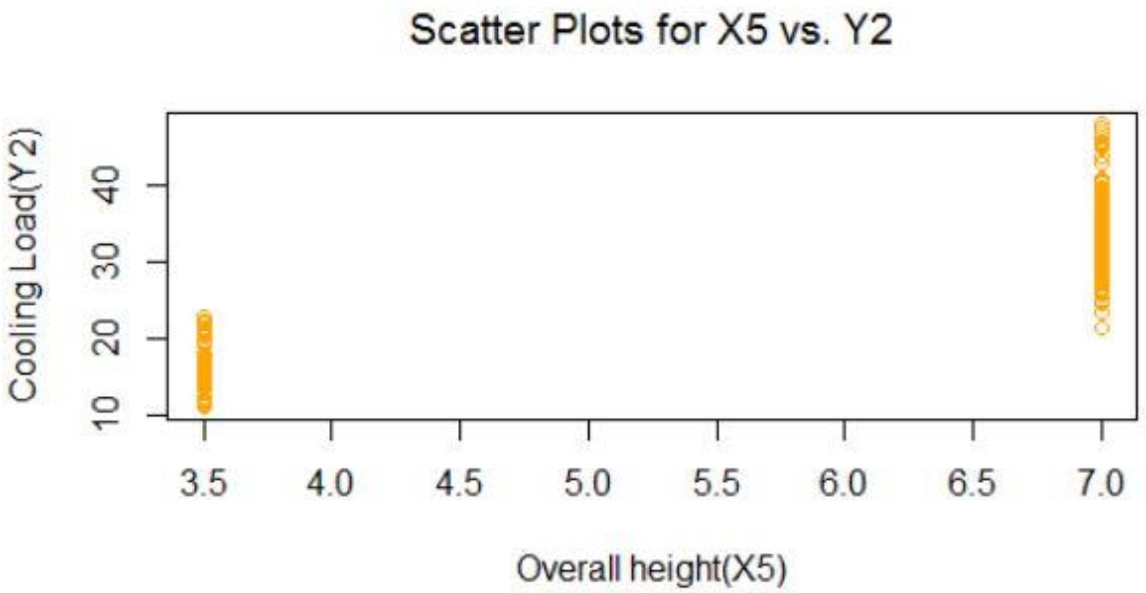

The surface area shows an approximate negative trend. This is what we might expect, given that if surface area is low that might be fairly uncomfortable for the cooling load.



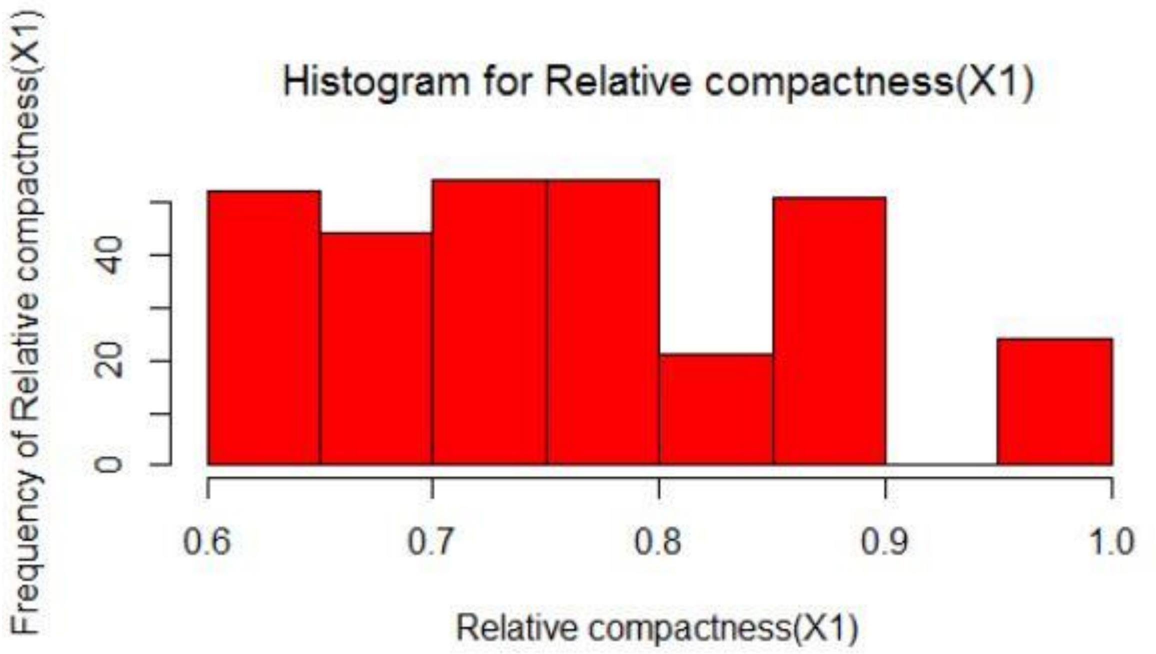
The wall area shows a positive linear trend, although for the few observations of wall area above 400, the cooling load does not seem to decrease.



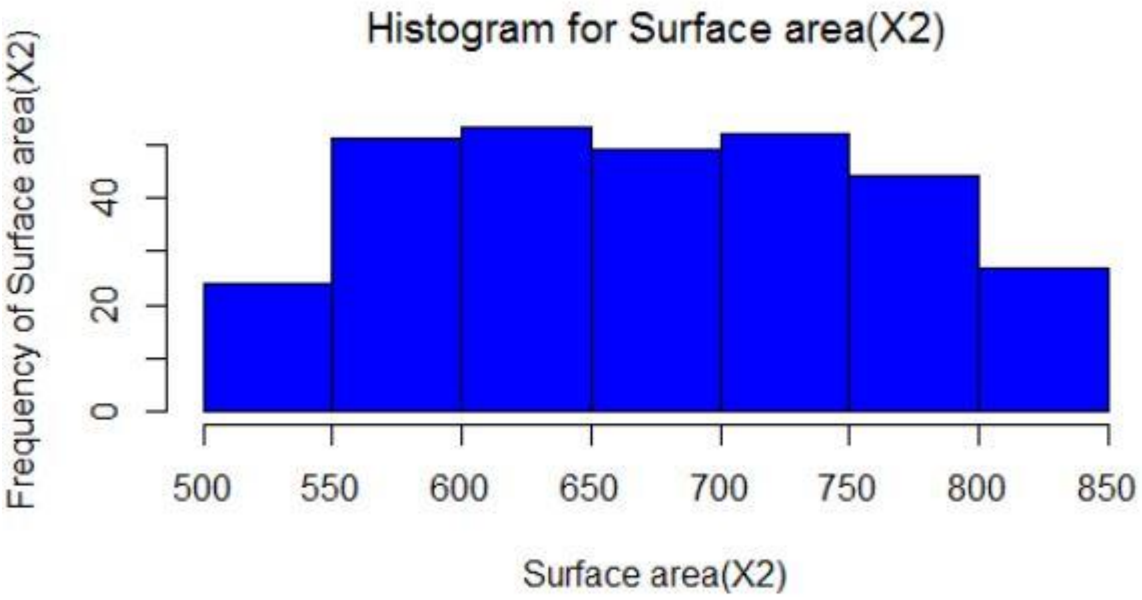

The roof area(X4) shows an approximate negative trend. What we expect from the graph that if roof area is between100 to 150, the cooling load (Y2) needs to work hard.



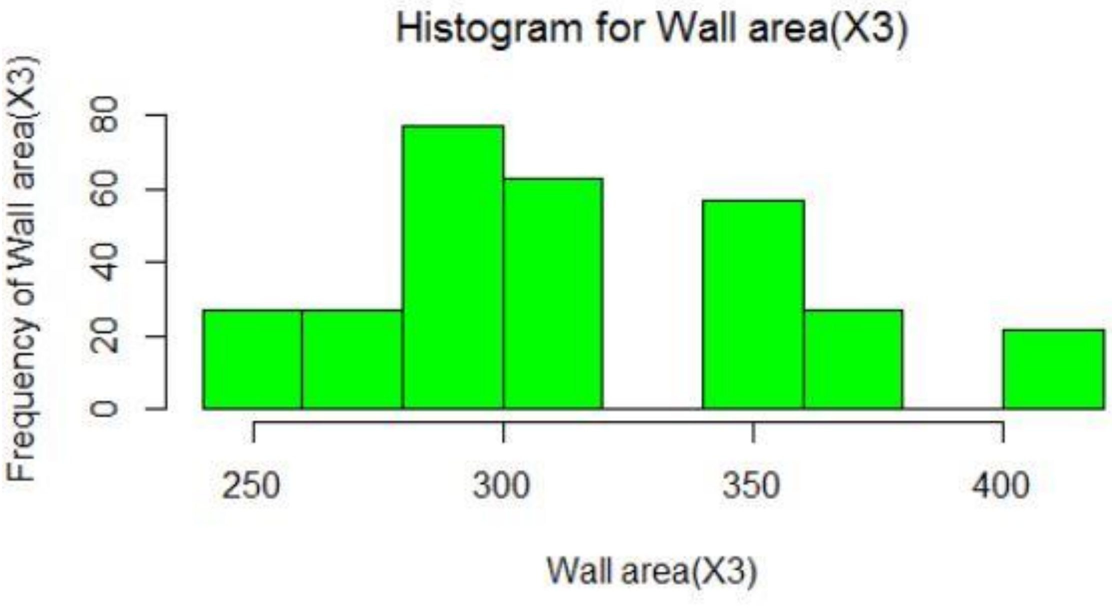
Although overall height is categorical, we can see that as the overall height becomes large, the cooling load will go high. Although there are two potential labels, there was no data instance that was recorded between 3.5 & 7.



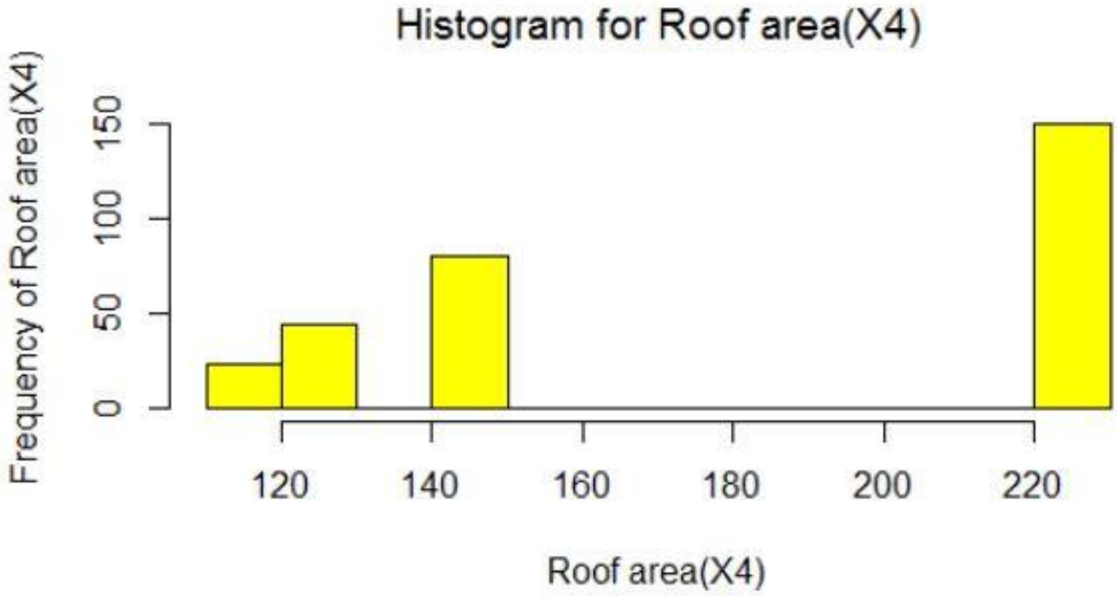

With the majority of the Relative compactness(X1) performed their sprint in between 0.6 and 0.8.On the other hand, if the X1 beyond 0.9 then much smaller than everyone else, but we might not want this score in just one variable to have an undue influence on the final result. Positively Skewed Distribution, so transformations like log, or *x^p* with *p* less than 1 could be useful.



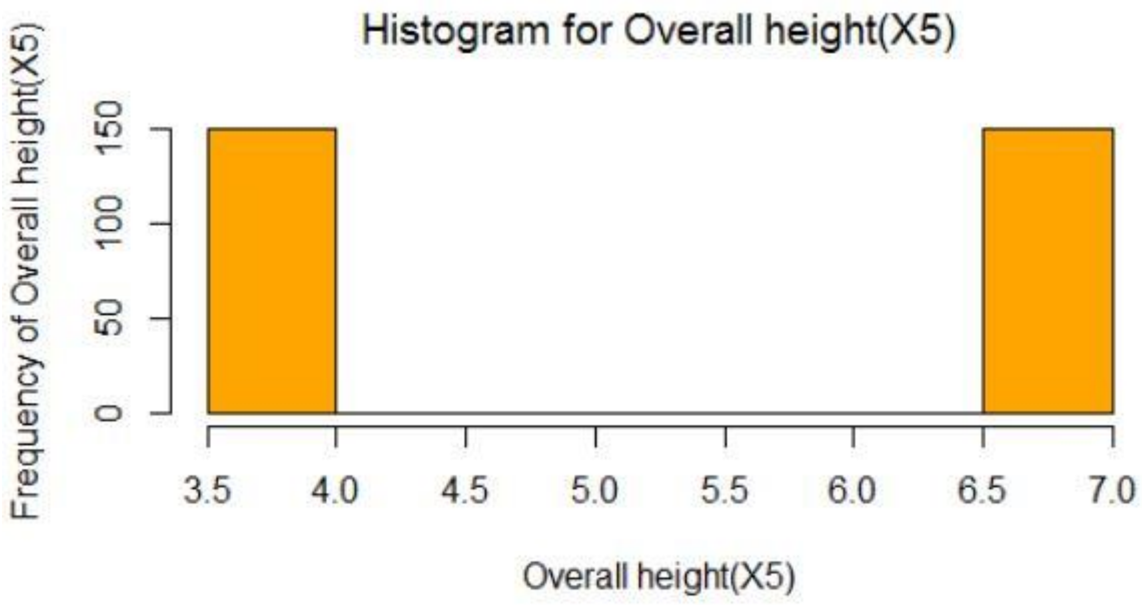

Both of the variables are reasonably symmetric in terms of their distribution. Surface Area(X2) is following the Normal distribution.



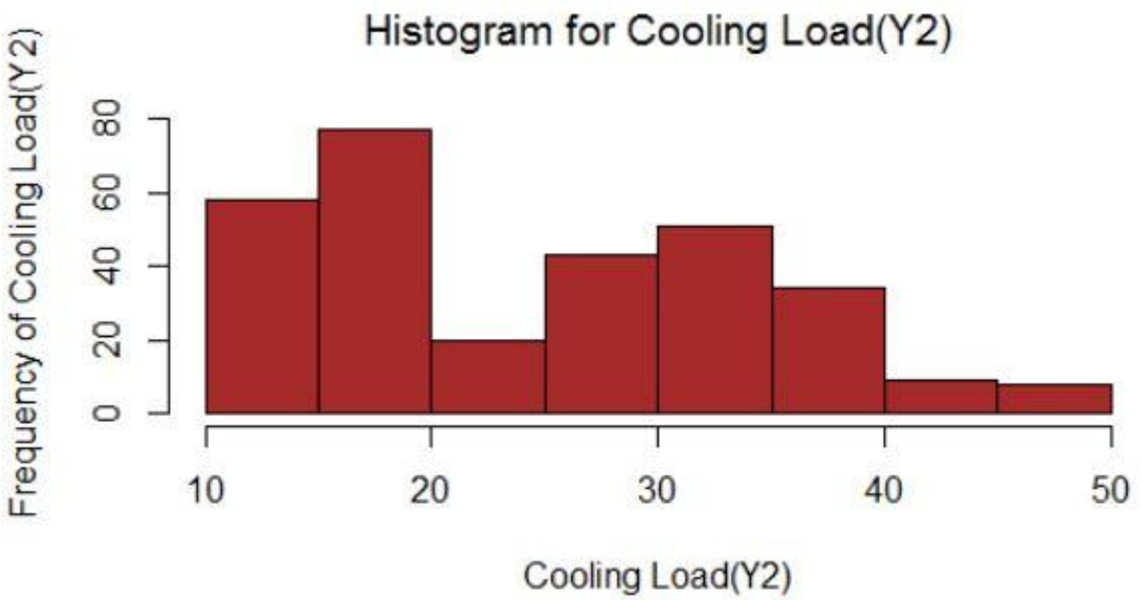

Wall area(X3) is somewhat follow normal distribution. It is not following the strict normal distribution. There is some skewness is there.




The roof area(X4) of the building has highest frequency in 220 square metre. There is some increasing exponential distribution is followed by the graph. Although the general difference between the values is somewhat maintained, values toward the lower end of the scale become more spread out while higher values are pushed closer together. So, may be log or polynomial transformation would be better.




The overall height of the building is either fall in two categories either in 3.5-4.0 metre or 6.5-7.0 metre.




There is some increasing exponential distribution is followed by the graph .Positively Skewed Distribution, so transformations like log, or *x^p* with *p* less than 1 could be useful. I think data is not much skewed so, best would be the Polynomial function transformation can have a similar (but less drastic) effect, & make our skewed distributions to become more symmetric

2. (ii)

**Transformations:**

**Cooling Load (Y2):**

The Cooling Load (Y2) can be transformed using linear feature scaling with the maximum Cooling Load (Y2) recorded at 47.59 and the minimum equal to 10.94.

**Overall Height(X5):**

A simple negation and transformation to the unit interval (indicating suitability for cooling load) can be achieved using so that overall height=3.5 is allocated a score of=1, while overall height=7.0 is allocated a score=0.

**Roof Area(X4):**

The log transformation makes for the outliers to become less pronounced allowing skewed distributions to become more symmetric. Then apply a scaling and negation to make our data fall in the range of 0 to1.

**Wall Area(X3):**

This variable can be transformed using standardization. Both of the variables are reasonably symmetric in terms of their distribution. We can also see from the distributions that both variables have more values closer to zero, so transformations could be useful.

**Surface Area(X2):**

Polynomial functions can help data to make more symmetric. Then apply a scaling and negation to make our data fall in the range of 0 to1.

**Relative Compactness(X1):**

No transformed is done because the range lies between 0 to 1.

3. (iii)

| p | p=0.5(WPM) | p=1(WAM) | p=2(WPM) | OWA | Choquet |
|---|---|---|---|---|---|
| **RMSE** | 0.14337 | 0.12857 | 0.19400 | 0.15460 | 0.144719 |
| **Av. Abs error** | 0.0971581 | 0.09109 | 0.131374 | 0.115484 | 0.0979 |
| **Pearson Correlation** | 0.836790 | 0.8688 | 0.8097190 | 0.811219 | 0.8312033 |
| **Spearman Correlation** | 0.6688906 | 0.82430 | 0.668890 | 0.76639 | 0.668890624 |

## Error Measures

| p | p=0.5(WPM) | p=1(WAM) | p=2(WPM) | OWA | Choquet |
|---|---|---|---|---|---|
| **W1** | 0.2838712 | 0 | 0 | 0.0702579 | 0.1895 |
| **W2** | 0 | 0.79290 | 0.087313 | 0.359495 | 0.1895 |
| **W3** | - | 0.1630779 | 0 | 0.546480 | - |
| **W4** | 0.59394865 | 0 | 0.894188 | 0 | 0.774 |
| **W5** | 0.12218011 | 0.044012 | 0.0184983 | 0.0237663179 | 0 |

## Weights Parameters

## (iv)

Overall we can see that the Weighted Arithmetic Mean (WAM) has the best fitting performance because of the lowest RMSE & average absolute error and so it seems to be the best fitting function. The weighted arithmetic mean (WAM) is best in all of the power means on the table. This implied that the output doesn’t tend toward high or low values.

From the weights for the weighted power means, we see that Overall height(X5) was by far the most important variable for making predictions, followed by the Surface Area(X2), Roof Area(X4), Wall Area(X3) and then Relative compactness(X1).

The second best fitting function would be Choquet integral or weighted power mean (p=0.5) as both of them have similar type of error. Besides, they have pretty much the similar correlation. However, Choquet integral is not surprising since it is defined with respect to 16 parameters rather than 4 and so it is more flexible. All of the models can be interpreted as having an average error in the range of between 10–16%. If we knew the expected variable like Relative compactness(X1), Surface Area(X2), Wall Area(X3), Roof Area(X4), Overall height(X5) for a particular building, we could hope that our model would predict the number of casual users within about 15% (provided our model holds for unseen data).For predictions corresponding with Cooling Load (Y2), this amounts to about 3 unit ($kWh.m^{-2}$ per annum).

The OWA did not fit as well as some of the means like (Quadratic Mean (p=2)), from which we can infer that the source of the inputs is more important for making predictions than their relative size. Here the OWA is almost exactly the same as the median, with 0.546 allocated to the middle-weight. If we compare the power means (p=2) & OWA we find that the OWA had slightly better performance in terms of RMSE, average absolute error and correlation. Weighted power model(p=2) that fit worse had slightly higher allocations of weight to Roof Area(X4), since as weights become closer to 1 the function behaviour will be less influenced by the value of the parameter *p*.

v ({1, 2,3,4})=1

v ({1,2,3})=0.744230

v ({1,2,4})=0.189521

v ({1,3,4})=0.7442303

v ({2,3,4})=1

v ({1,2})=0.189521

v ({1,3})=0.744230

v ({1,4})=0.189521

v ({2,3})=0.744230

v ({2,4})=0.189521

v ({3,4})=0.7442303

v ({1})=0.189521

v ({2})=0.189521

v ({3})=0.744230

v ({4})=0

v (null)=0

Looking at the fuzzy measure we can spot a few interesting relationships. The Complementary/positive synergy relationship for one of the triplets simply suggests that if all are high then it is likely that the output should be high, whereas in some cases. v({2})=0.1895, v({3})=0.74423,and v({4})=0(binary number 2=0010, binary number 3=0011,and binary number 4=0100 respectively),however together their weight is v({2,3,4})=1 (binary number 14=1110).This suggests a super-additive relationship. Perhaps sometimes we are finding cooling load are more similar to each other, both having good scores does not suggest much about Cooling load (Y2).Take the example of variable 1 and 4 together are fairly additive(no interaction). Similar cases like variable 2 &4, and third one is with 3 & 4.Overall height(X5) is worth nothing either by themselves or as a pair, and combining either or both with Roof Area(X4) & Surface Area (X2) does not increase the weight of the measure. We can spot a large number of redundant interaction. An overall interpretation of this fuzzy measure is that we need Surface

Area(X3), Roof Area(X4) *and* Overall height(X5) in order for there to be a correct Cooling Load (Y2).

v ({1,2,3})= Redundant Interaction

v ({1,2,4})= Redundant Interaction

v ({1,3,4})= Redundant Interaction

v ({2,3,4})=Complementary Interaction

v ({1,2})=Redundant Interaction

v ({1,3})= Redundant Interaction

v ({1,4})=No Interaction

v ({2,3})= Redundant Interaction

v ({2,4})= No Interaction

v ({3,4})= No Interaction

The orness values of close to the 0.4(Choquet model-0.48 and the OWA-0.38) summarizes this tendency follows toward the lower inputs.

RMSE 0.144719299527487

Av. abs error 0.0979425235512706

Pearson Correlation 0.831203338892168

Spearman Correlation 0.668890624713162

Orness 0.47239641704501

i Shapley i

1 0.0631736784533495

2 0.148430223270693

3 0.70313955345861

4 0.0852565448173426

binary number fm.weights

1 0.189521035360053

2 0.189521035360055

3 0.189521035360054

4 0.744230365547971

5 0.744230365547971

6 0.744230365547972

7 0.744230365547967

8 0

9 0.189521035360053

10 0.189521035360055

11 0.189521035360054

12 0.744230365547971

13 0.744230365547971

14 1

15 0.999999999999995

File:Stat_choq.txt

4. (i)

X=<0.82, 612.5, 318.5, 147, 7>

Apply Transformation

X'=<0.82, 0.79, 0.072, 0.66, 0>

| **Function** | **WPM(p=0.5)** | **WAM(p=1)** | **WPM(p=2)** | **OWA** | **Choquet** |
|---|---|---|---|---|---|
| **Model output** | 0.5468284 | 0.637416 | 0.666296 | 0.4061627 | 0.7264 |
| **Cooling Load(Y2)** | 30.98126 | 34.30129 | 35.359 | 25.8258 | 37.56256 |

Depending on the model we use, we would predict that cooling load (Y2) for the input is about 30-38. The OWA gives a low result than the other functions because it is the only one that doesn't take into account the source of the inputs.

(ii)

**Ideal Conditions:**

To estimate the required cooling capacities, one has to have information regarding the design indoor and outdoor conditions, specifications of the building, and specifications of the conditioned space (such as the occupancy, activity level, various appliances and equipment used etc.) and any special requirements of the particular application.

**Cooling Load (Y2):**30 kWh.m$^{-2}$ per annum

**Overall Height(X5):**7 metre

**Roof Area(X4):**140-160 square metre

**Wall Area(X3):**300-350 square metre

**Surface Area(X2):**650-750 square metre

**Relative Compactness(X1):**0.75

# PART-B

1.

```
/*********************************************
 * OPL 12.8.0.0 Model
 * Author:Shantanu
 * Creation Date: May 3, 2018 at 8:43:55 AM
 *********************************************/
dvar float+ x;  //Number of litres used of product JA to produce special juice
dvar float+ y;  //Number of litres used of product JB to produce special juice
dexpr float Cost = 6*x + 5*y; //Objective Function
minimize Cost;

subject to {        //constraints

 //Carrot concentration in 100 litres of juice
 4*x + 8*y <= 6;

 //Orange concentration in 100 litres of juice
 6*x + 3*y >= 3.5;

 //Apple concentration in 100 litres of juice
 3*x + 6*y >= 4;

 x >= 0;      //Non-negativity constraints
 y >= 0;      //Non-negativity constraints

 }
```

Solution

```
// solution (optimal) with objective 4.5
// Quality There are no bound infeasibilities.
// There are no reduced-cost infeasibilities.
// Maximum Ax-b residual              = 0
// Maximum c-B'pi residual            = 8.88178e-16
// Maximum |x|                        = 0.5
// Maximum |slack|                    = 0.666667
// Maximum |pi|                       = 0.777778
// Maximum |red-cost|                 = 0
// Condition number of unscaled basis = 2.5e+01
//

x = 0.33333;
y = 0.5;
```

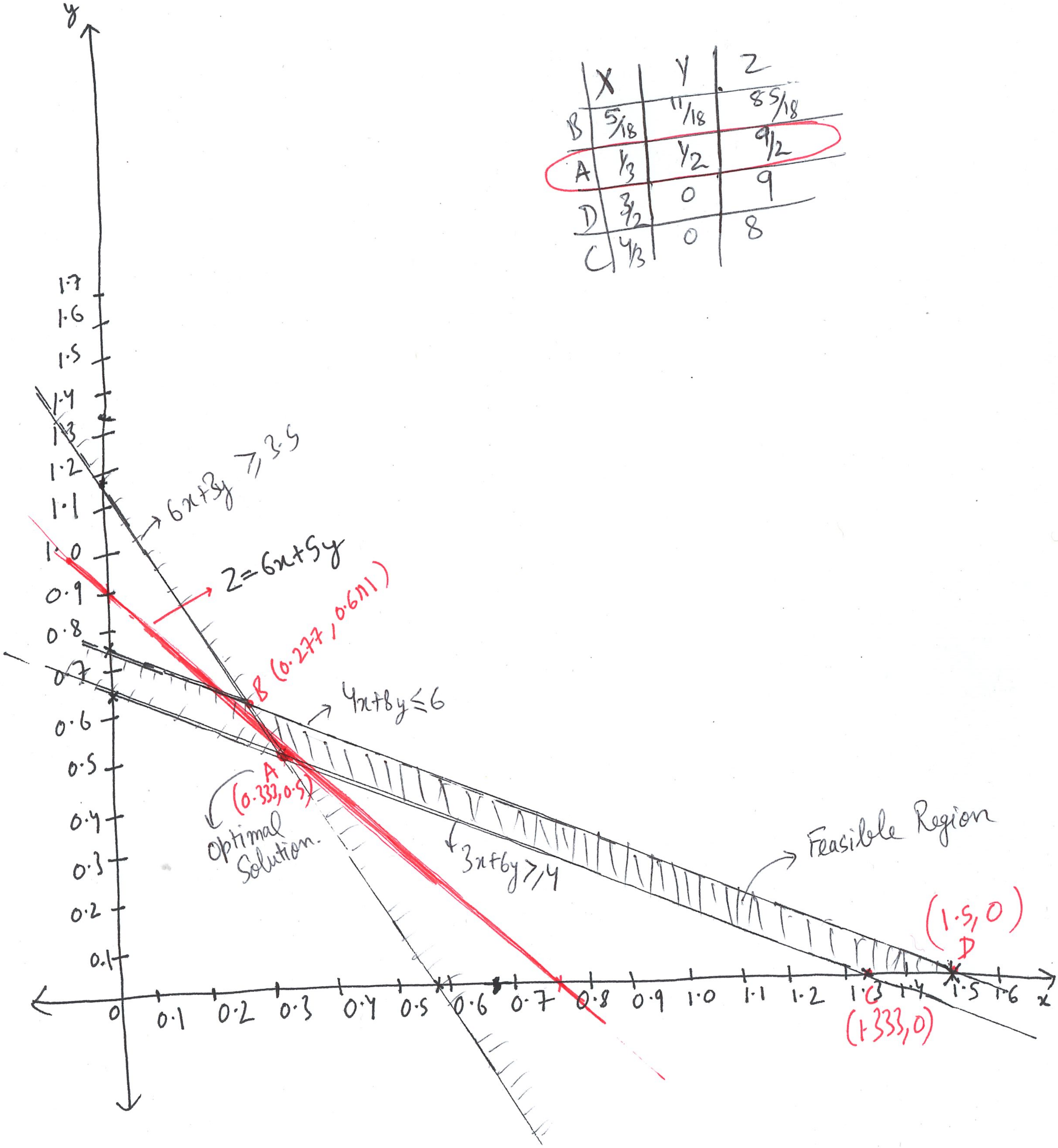
Minimize $Z = 6x + 5y$ // objective function

$x$ = no. of litres of JA used to produce Special Juice
$y$ = " " " " JB " " " "

$$\left.\begin{aligned} 4x + 8y &\le 6 \\ 6x + 3y &\ge 3.5 \\ 3x + 6y &\ge 4 \end{aligned}\right\}$$ // constraints

$x, y \ge 0$

|   | X | Y | Z |
|---|---|---|---|
| B | $5/18$ | $11/18$ | $85/18$ |
| A | $1/3$ | $1/2$ | $9/2$ |
| D | $3/2$ | 0 | 9 |
| C | $4/3$ | 0 | 8 |



```
/************************************************
 * OPL 12.8.0.0 Model
 * Author: Shantanu
 * Creation Date: May 3, 2018 at 9:00:07 AM
 *********************************************/
dvar float+ Xc1; // Number of tons of Cotton used in making Summer materials
dvar float+ Xc2; // Number of tons of Cotton used in making Autumn materials
dvar float+ Xc3; // Number of tons of Cotton used in making Winter materials
dvar float+ Xw1; // Number of tons of Wool used in making Summer materials
dvar float+ Xw2; // Number of tons of Wool used in making Autumn materials
dvar float+ Xw3; // Number of tons of Wool used in making Winter materials
dvar float+ Xv1; // Number of tons of Viscose used in making Summer materials
dvar float+ Xv2; // Number of tons of Viscose used in making Autumn materials
dvar float+ Xv3; // Number of tons of Viscose used in making Winter materials

// expressions
dexpr float SalesPrice = 50*(Xc1 + Xw1 + Xv1) + 55*(Xc2 + Xw2 + Xv2) + 60*(Xc3 +
Xw3 + Xv3);

dexpr float ProductionCost = 4*(Xc1 + Xw1 + Xv1) + 4*(Xc2 + Xw2 + Xv2) + 5*(Xc3 +
Xw3 + Xv3);

dexpr float PurchaseCost = 30*(Xc1 + Xc2 + Xc3) + 45*(Xw1 + Xw2 + Xw3) + 40*(Xv1 +
Xv2 + Xv3);

dexpr float TotalProfit = SalesPrice - ProductionCost - PurchaseCost;

// maximise function

maximize TotalProfit;     //Objective Function

subject to{         //Constraints

// Demand constraints
Xc1 +Xw1 + Xv1 <=4500;
Xc2 +Xw2 + Xv2 <=4000;
Xc3 +Xw3 + Xv3 <=3800;

// Cotton Proportion constraints
Xc1 >= 0.6*Xc1 + 0.6*Xw1 + 0.6*Xv1;
Xc2 >= 0.6*Xc2 + 0.6*Xw2 + 0.6*Xv2;
Xc3 >= 0.4*Xc3 + 0.4*Xw3 + 0.4*Xv3;

// Wool Proportion constraints
Xw1 >= 0.3*Xw1 + 0.3*Xc1+ 0.3*Xv1;
Xw2 >= 0.3*Xw2 + 0.3*Xc2+ 0.3*Xv2;
Xw3 >= 0.5*Xw3 + 0.5*Xc3+ 0.5*Xv3;

}
```

---

Solution

```
// solution (optimal) with objective 184250
// Quality There are no bound infeasibilities.
// There are no reduced-cost infeasibilities.
// Maximum Ax-b residual              = 1.13687e-13
// Maximum c-B'pi residual            = 2.66454e-15
// Maximum |x|                        = 3150
```

```
// Maximum |slack|                       = 450
// Maximum |pi|                          = 17.5
// Maximum |red-cost|                    = 10
// Condition number of unscaled basis = 2.7e+00
//

Xc1 = 3150;
Xw1 = 1350;
Xv1 = 0;
Xc2 = 2800;
Xw2 = 1200;
Xv2 = 0;
Xc3 = 1900;
Xw3 = 1900;
Xv3 = 0;
```

# References:

James (2016) ***An Introduction to Data Analysis using Aggregation Functions in R.*** Deakin Library [Online] Available at: https://link-springer-com.ezproxy-f.deakin.edu.au/content/pdf/10.1007%2F978-3-319-46762-7.pdf (Accessed 5 May 2018)

Ramgopal(2009) ***Refrigeration and Air Conditioning (Web)*** Available at: http://nptel.ac.in/courses/112105129/ (Accessed 5 May 2018)