Real World Analytics Assignment

Student Name: Dhananjay Pandya

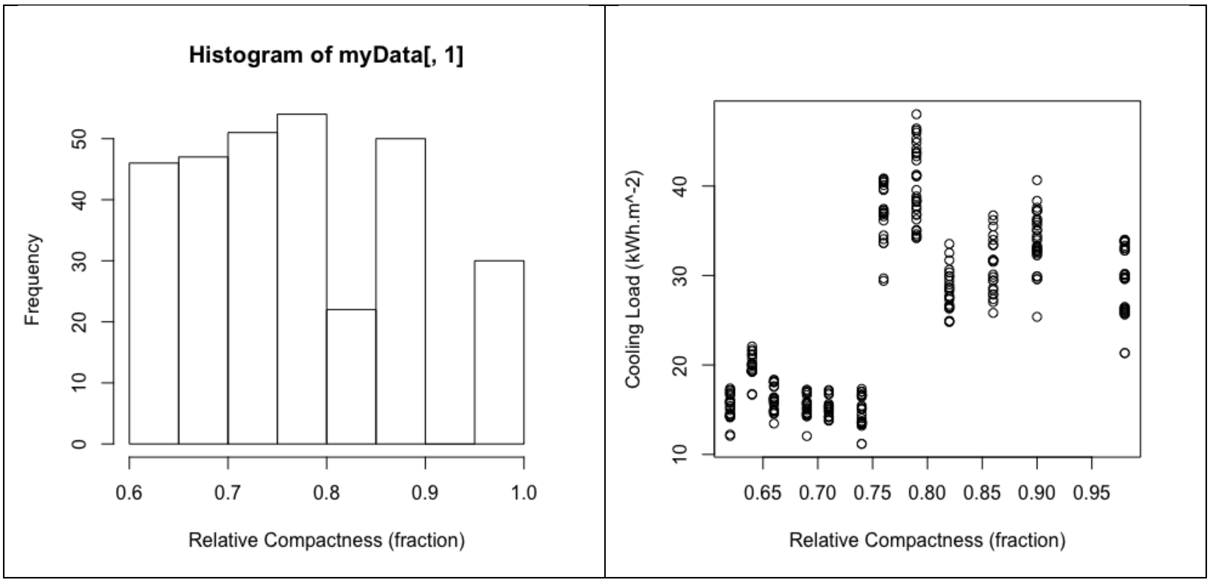
Student Number: 218202943

Part A

Question 1

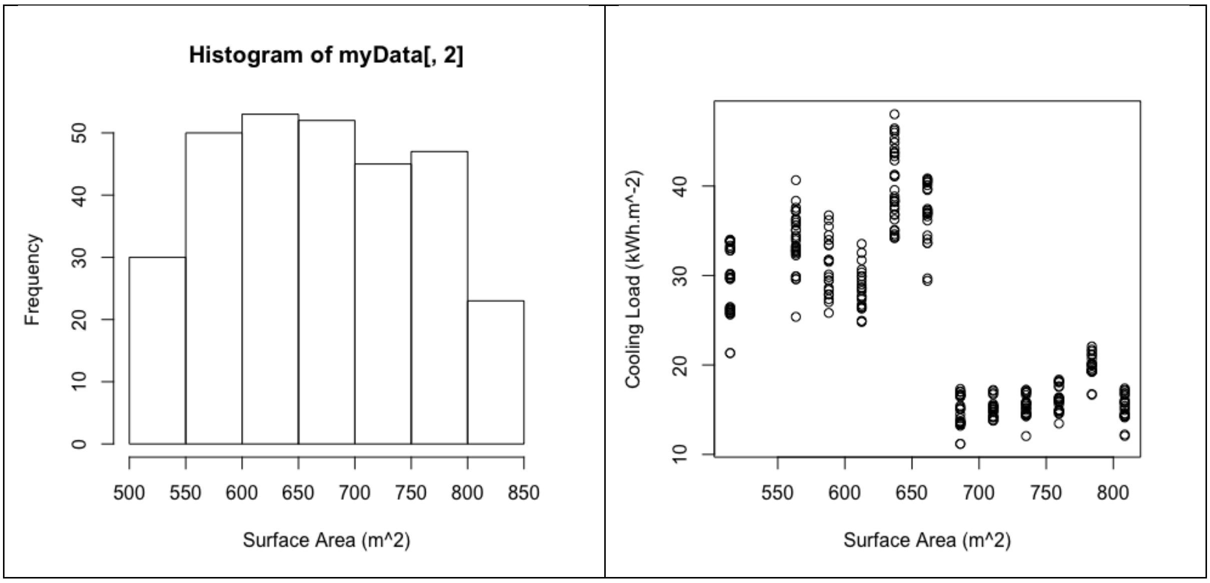
Part iv

Table 1 - Histogram of Relative Compactness (Left) & Cross-lot of Cooling Load Vs Relative Compactness (Right)



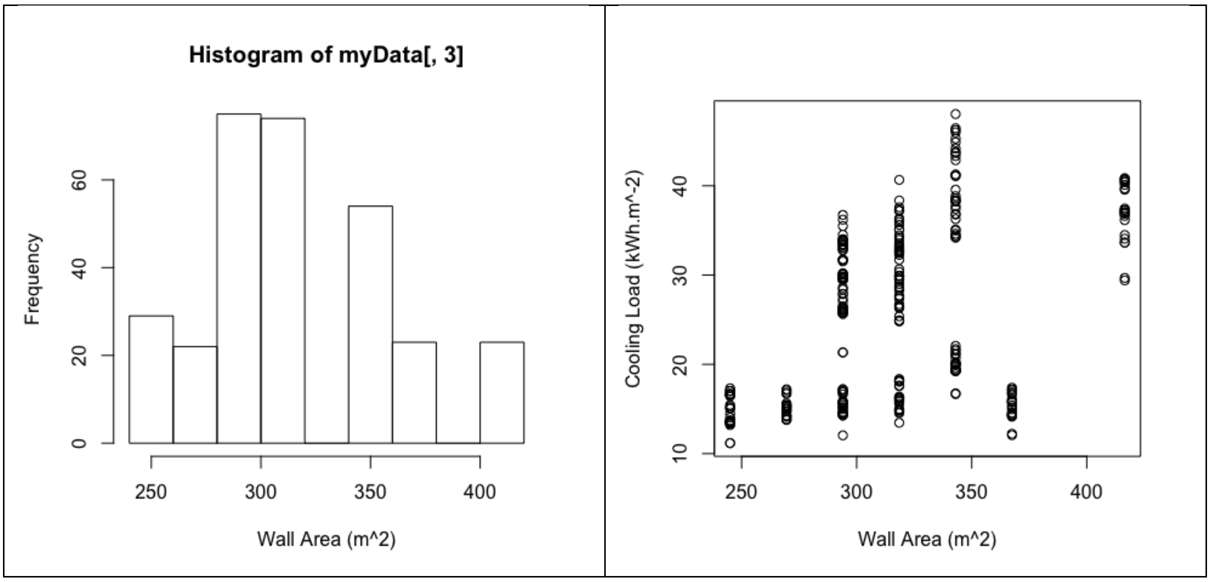
Relative compactness is pretty evenly distributed between values of 0.6 and 1.0. Relative compactness below 0.75 is correlated with relatively low cooling loads (most values are less than 20 kWh.m^-2). Relative compactness over 0.75 is correlated with relatively high cooling loads (most values are greater than 20 kWh.m^-2).

Table 2 - Histogram of Surface Area (Left) & Cross-lot of Cooling Load Vs Surface Area (Right)



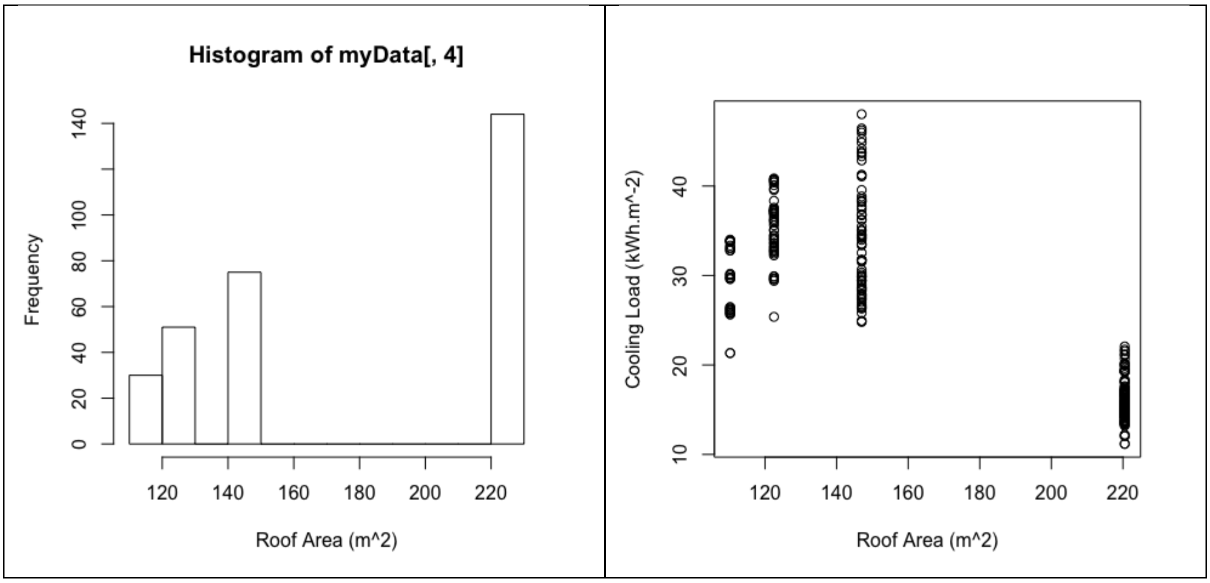
Surface area's distribution appears to be pretty even across all values apart from the highest and lowest. Cooling loads are high (greater than 20kWh.m^-2) when surface area is low, whereas cooling loads are low when surface area is high (most values less than 20kWh.m^-2).

Table 3 - Histogram of Wall Area (Left) & Cross-lot of Cooling Load Vs Wall Area (Right)



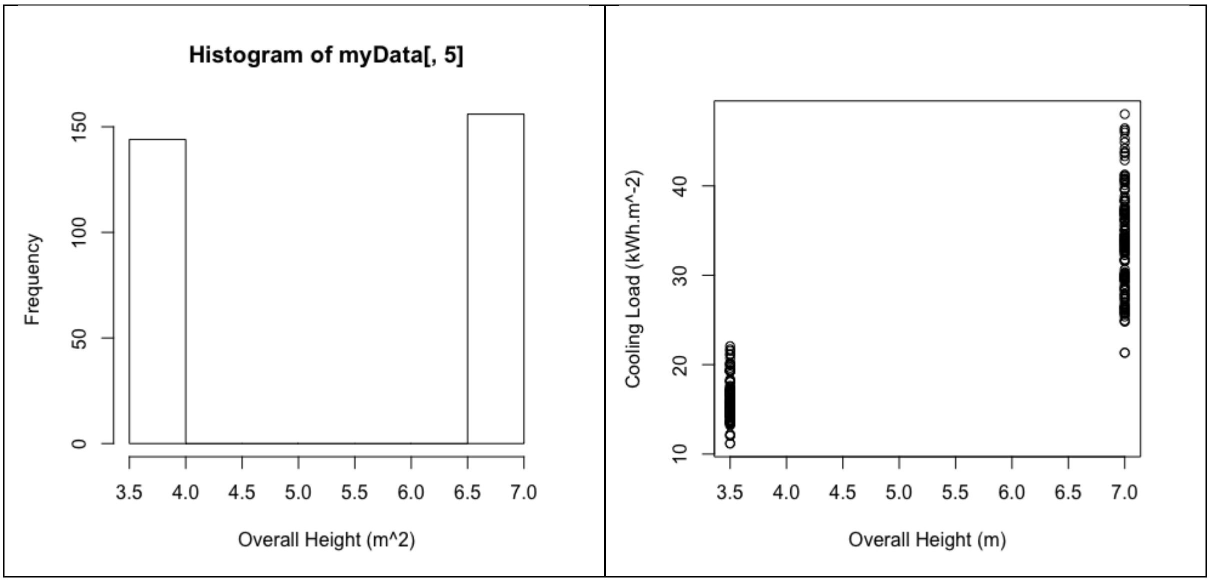
Wall Area's distribution is quite random although most of the values are between 290 and 350 m^2. On the cross plot, it appears that Wall Area does not seem to have any correlation with Cooling Load because wall areas between 290 and 350 m^2 can lead to cooling loads as low as ~12 kWh.m^-2 and as high as ~48 kWh.m^-2.

Table 4 - Histogram of Roof Area (Left) & Cross-lot of Cooling Load Vs Roof Area (Right)



Values of roof area are distributed across low and high regions, with no intermediate values. Across the low values for roof area, cooling loads generally increase as roof area increases. For the high value, cooling loads are low despite roof area being high.

Table 5 - Histogram of Overall Height (Left) & Cross-lot of Cooling Load Vs Height (Right)



There are only two values of overall height and they are pretty evenly split between the low and high value. Low overall height correlates with relatively low cooling loads and high overall height correlates with relatively high cooling loads.

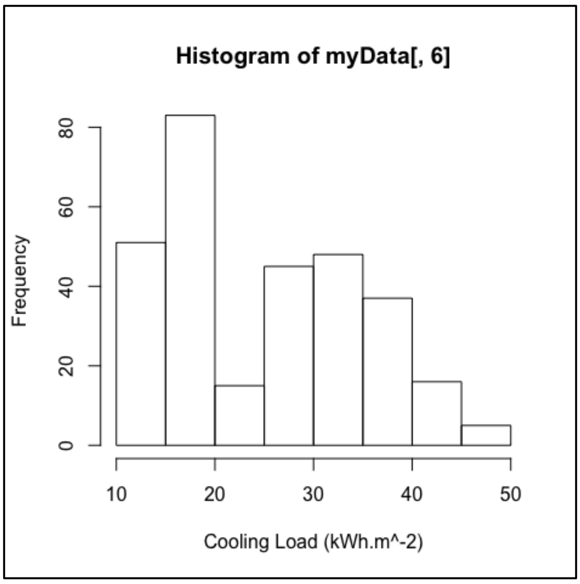


Figure 1 - Histogram of Cooling Load

Cooling loads are distributed pretty randomly. No distinct distribution appears to apply to this cooling loads.

Question 2

Part ii

The variables that I chose were Surface Area, Wall Area, Roof Area and Overall Height. For each of these, I simply fit them to the unit interval using the formula shown below. I did not see any need to make complex transformations because none of the frequency distributions indicated any noticeable patterns (normal distributions, skewed distributions etc.). Note, this also applies to the cooling load.

$$\begin{aligned}&\text{Let } x \text{ be the variable of interest,}\\&\text{and } T(x) \text{ be the function that fits } x \text{ to the unit interval}\\&\\&\text{Then,}\\&T(x) = \frac{x - \min(x)}{\max(x) - \min(x)}\end{aligned}$$

Question 3

Part iii

Table 6 - Error Measures for Fitting Functions

	RMSE	**Av. Abs error**	**Pearson Correlation**	**Spearman Correlation**
WAM	0.156470446010064	0.123805028034003	0.900105595422177	0.871710487999059
PM (p=0.5)	0.107083708750492	0.0783766780304178	0.917370402546387	0.869775378799302
PM (p=2)	0.214711657441631	0.166606420699038	0.7647067758472	0.809230686096816
OWA	0.221053439285888	0.176066196418882	0.705426067083965	0.750275010874028
Choquet	0.0980479571274367	0.0730064659070377	0.928691607569856	0.871710487999059

See next page for table containing weights and parameters for fitting functions.

Table 7 - Weights & Parameters for fitting functions

Function	WAM	PM (p=0.5)	PM (p=2)	OWA	Choquet
w_1	0	0.0785409541369554	0	0.779706999457406	N/A
w_2	0.461729064930518	0.14663767966966	0.736357340720217	0	N/A
w_3	0.0550732501356484	0.203035566777422	0	0	N/A
w_4	0.483197684933833	0.571785799415962	0.263642659279783	0.220293000542593	N/A
Orness	N/A	N/A	N/A	0.220293000542593	0.36143704407063
Shapley i_1	N/A	N/A	N/A	N/A	0.0862723819859653
Shapley i_2	N/A	N/A	N/A	N/A	0.239177518538591
Shapley i_3	N/A	N/A	N/A	N/A	0.120207089889727
Shapley i_4	N/A	N/A	N/A	N/A	0.554343009585813
Binary Number 1	N/A	N/A	N/A	N/A	0
Binary Number 2	N/A	N/A	N/A	N/A	0.175393380358087
Binary Number 3	N/A	N/A	N/A	N/A	0.175393380358087
Binary Number 4	N/A	N/A	N/A	N/A	0.101804123711335
Binary Number 5	N/A	N/A	N/A	N/A	0.101804123711335
Binary Number 6	N/A	N/A	N/A	N/A	0.175393380358101
Binary Number 7	N/A	N/A	N/A	N/A	0.175393380358101
Binary Number 8	N/A	N/A	N/A	N/A	0.404286489419351
Binary Number 9	N/A	N/A	N/A	N/A	0.404286489419351
Binary Number 10	N/A	N/A	N/A	N/A	0.741182854042256
Binary Number 11	N/A	N/A	N/A	N/A	0.741182854042256
Binary Number 12	N/A	N/A	N/A	N/A	0.404286489419351
Binary Number 13	N/A	N/A	N/A	N/A	0.663103635377247
Binary Number 14	N/A	N/A	N/A	N/A	0.741182854042201
Binary Number 15	N/A	N/A	N/A	N/A	1.0000000000001

Part iv

RMSE and Av. Abs error are negatively oriented meaning that lower values are better (Medium 2016, para. 6). Based on these, the Choquet integral performs best for this model. Low values for both measures indicate that the predicted (output) and input cooling loads are have similar values. The Pearson correlation provides a measure of the linearity of the relationship between two variables (Minitab 2016, para. 2). The Spearman correlation provides a measure of the monotonicity of the relationship between two variables (Minitab 2016, para. 4). The Choquet integral has values closest to 1 for both measures indicating that the predicted cooling loads and input cooling loads have a strongly positive linear relationship. Together, the RMSE, Av. Abs error, Pearson and Spearman correlation's respective values indicate that the Choquet integral provides the best fit for this data. This is because the Choquet integral uses 16 parameters instead of just the four that are used for other fitting functions. The next best method is the Power Mean (p=0.5) taking into account all four factors, followed by WAM, Power Mean (p=2) and OWA respectively.

For most functions, the overall height seems to be the most important input for predicting cooling loads. This definitely applies to the Choquet integral. Based on the Shapley values, the Choquet integral assigns highest importance to overall height followed by wall area, roof area and surface area respectively.

There does not appear to be a strong complementary or redundant relationship between any of the variables.

Better models favour values because the OWA assigns most of the weight to the smallest input and the Choquet integral 'orness' is less than 0.5 (value is approx. 0.36).

Question 4

Part i

Predicted Value - 34.48 kWh.m^(-2)

Comment – This value is reasonable. This is based on the observation that the cooling load vs input scatter plots do not indicate that this value would be an outlier (i.e. the intersection of the cooling load and the input sits within in existing cloud of data rather than being an anomaly).

Part ii

Low cooling loads can be expected when:

- Relative compactness is less than 0.75 (less compact means more spacious)
- Surface area is greater than 675 m^2
- Wall area is less than 375 m^2
- Roof area is 220 m^2
- Overall height is 3.5 m

Part B

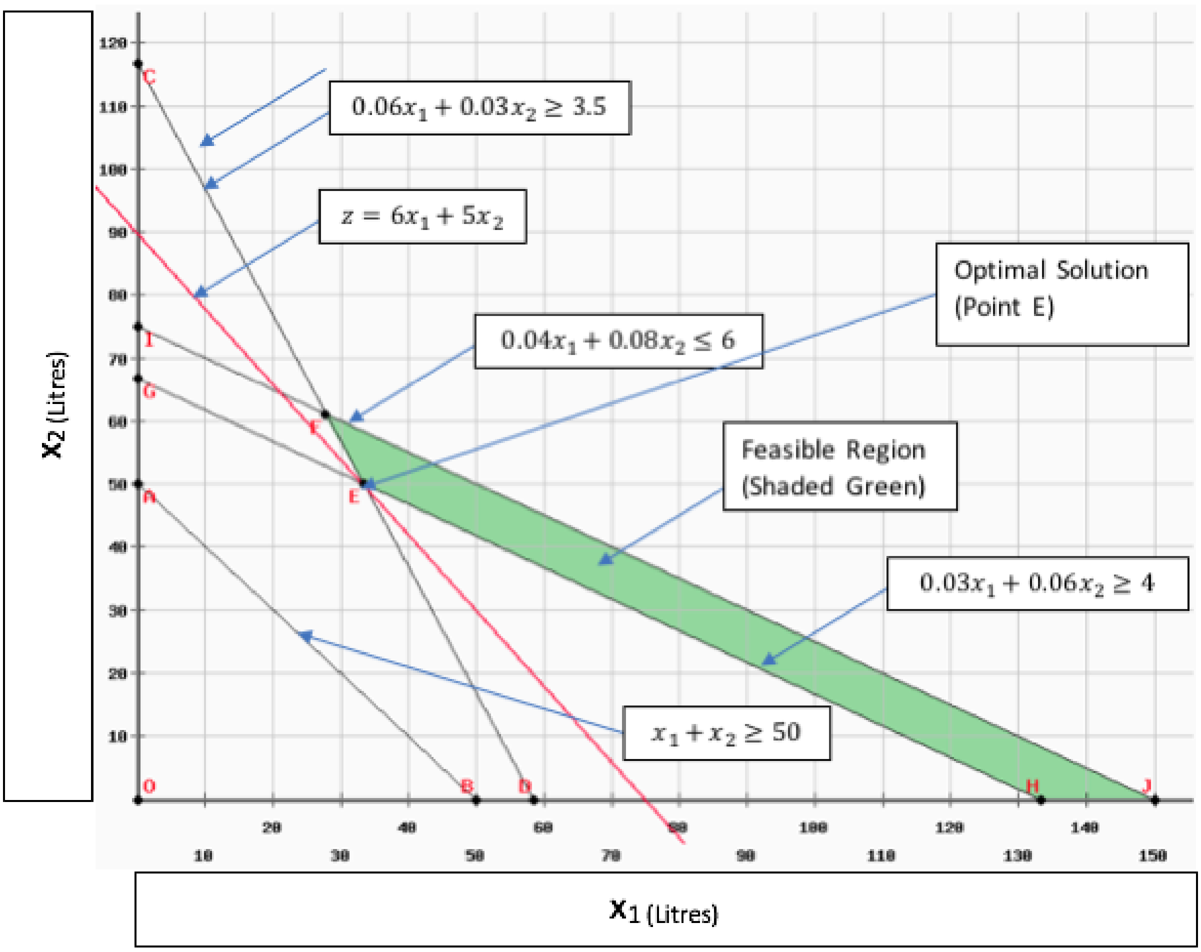
Question 1

LP Model:

$$\min(z = 6x_1 + 5x_2)$$
$$s.t.\ x_1 + x_2 \geq 50$$
$$0.06x_1 + 0.03x_2 \geq 3.5$$
$$0.03x_1 + 0.06x_2 \geq 4$$
$$0.04x_1 + 0.08x_2 \leq 6$$
$$x > 0$$
$$y > 0$$

Optimal Solution: JA = 33.33 Litres; JB = 50 Litres; Cost = $450

Figure 2 - Graphical solution for linear programming



Question 2

LP Model:

$$Let\ x_{ij}\ be\ the\ amount\ of\ material\ i\ used\ daily\ to\ produce\ fabric\ j, for\ i = 1,2,3; j = 1,2,3.$$

$$Amount\ of\ materials\ used\ (in\ tons)$$
$$Cotton: x_{11} + x_{12} + x_{13}$$
$$Wool: x_{21} + x_{22} + x_{23}$$
$$Viscose: x_{31} + x_{32} + x_{33}$$

$$Amount\ of\ fabrics\ produced\ (in\ tons)$$
$$Summer: x_{11} + x_{21} + x_{31}$$
$$Autumn: x_{12} + x_{22} + x_{32}$$
$$Winter: x_{13} + x_{23} + x_{33}$$

$$Revenue\ from\ fabric\ sales$$
$$50 * (x_{11} + x_{21} + x_{31}) + 55 * (x_{12} + x_{22} + x_{32}) + 60 * (x_{13} + x_{23} + x_{33})$$

$$Cost\ of\ purchasing\ materials$$
$$30 * (x_{11} + x_{12} + x_{13}) + 45 * (x_{21} + x_{22} + x_{23}) + 40 * (x_{31} + x_{32} + x_{33})$$

$$Cost\ of\ producing\ fabrics$$
$$4 * (x_{11} + x_{21} + x_{31}) + 4 * (x_{12} + x_{22} + x_{32}) + 5 * (x_{13} + x_{23} + x_{33})$$

$$profit = Revenue\ from\ fabric\ sales - Cost\ of\ purchasing\ materials - Cost\ of\ producing\ fabrics$$
$$maximize\ z = 50 * (x_{11} + x_{21} + x_{31}) + 55 * (x_{12} + x_{22} + x_{32}) + 60 * (x_{13} + x_{23} + x_{33}) - 30 * (x_{11} + x_{12} + x_{13}) - 45 * (x_{21} + x_{22} + x_{23}) - 40 * (x_{31} + x_{32} + x_{33}) - 4 * (x_{11} + x_{21} + x_{31}) - 4 * (x_{12} + x_{22} + x_{32}) - 5 * (x_{13} + x_{23} + x_{33})$$
$$= 16x_{11} + x_{21} + 6x_{31} + 21x_{12} + 6x_{22} + 11x_{32} + 25x_{13} + 10x_{23} + 15x_{33}$$
$$s.t.$$
$$x_{11} + x_{21} + x_{31} \leq 4500\ (Summer\ demand\ constraint)$$
$$x_{12} + x_{22} + x_{32} \leq 4000\ (Autumn\ demand\ constraint)$$
$$x_{13} + x_{23} + x_{33} \leq 3800\ (Winter\ demand\ constraint)$$
$$x_{11} \geq 0.6 * (x_{11} + x_{21} + x_{31})\ (Summer\ minimum\ cotton\ proportion\ constraint)$$
$$x_{12} \geq 0.6 * (x_{12} + x_{22} + x_{32})\ (Autumn\ minimum\ cotton\ proportion\ constraint)$$
$$x_{13} \geq 0.4 * (x_{13} + x_{23} + x_{33})\ (Winter\ minimum\ cotton\ proportion\ constraint)$$
$$x_{21} \geq 0.3 * (x_{11} + x_{21} + x_{31})\ (Summer\ minimum\ wool\ proportion\ constraint)$$
$$x_{22} \geq 0.3 * (x_{12} + x_{22} + x_{32})\ (Autumn\ minimum\ wool\ proportion\ constraint)$$
$$x_{23} \geq 0.5 * (x_{13} + x_{23} + x_{33})\ (Winter\ minimum\ wool\ proportion\ constraint)$$

Optimal Profit = $184250

Decision Variables:

Cotton for Summer = 3150 tons
Cotton for Autumn = 2800 tons
Cotton for Winter = 1900 tons

Wool for Summer = 1350 tons
Wool for Autumn = 1200 tons
Wool for Winter = 1900 tons

Viscose for Summer = 0 tons
Viscose for Autumn = 0 tons
Viscose for Winter = 0 tons

CPLEX Code

```
dvar float+ x11; //amount of Cotton used to produce Summer
dvar float+ x12; //amount of Cotton used to produce Autumn
dvar float+ x13; //amount of Cotton used to produce Winter

dvar float+ x21; //amount of Wool used to produce Summer
dvar float+ x22; //amount of Wool used to produce Autumn
dvar float+ x23; //amount of Wool used to produce Winter

dvar float+ x31; //amount of Viscose used to produce Summer
dvar float+ x32; //amount of Viscose used to produce Autumn
dvar float+ x33; //amount of Viscose used to produce Winter

dexpr float revenue = 50*(x11+x21+x31)+55*(x12+x22+x32)+60*(x13+x23+x33);
dexpr float purchaseCost = 30*(x11+x12+x13)+45*(x21+x22+x23)+40*(x31+x32+x33);
dexpr float productionCost = 4*(x11+x21+x31)+4*(x12+x22+x32)+5*(x13+x23+x33);
dexpr float profit = revenue - (purchaseCost + productionCost);

maximize profit;

subject to {
//Demand constraints
x11+x21+x31<=4500;
x12+x22+x32<=4000;
x13+x23+x33<=3800;
//Cotton proportion constraints
x11>=0.6*(x11+x21+x31); //proportion of Cotton in Summer
x12>=0.6*(x12+x22+x32); //proportion of Cotton in Autumn
x13>=0.4*(x13+x23+x33); //proportion of Cotton in Winter
//Wool proportion constraints
x21>=0.3*(x11+x21+x31); //proportion of Wool in Summer
x22>=0.3*(x12+x22+x32); //proportion of Wool in Autumn
x23>=0.5*(x13+x23+x33); //proportion of Wool in Winter
}
```

References

Minitab 2016, *A Comparison Of The Pearson And Spearman Correlation Methods*, retrieved 29 Apr. 2018. http://www.minitab.com/en-us/

Medium 2016, *MAE And RMSE — Which Metric Is Better?*, retrieved 29 Apr. 2018. https://medium.com/