Leeds University

Business School

UNIVERSITY OF LEEDS

Assessed Coursework Coversheet

For use with *individual* assessed work

Student ID Number:	2	0	1	6	6	9	0	9	0
Module Code:	LUBS5308M								
Module Title:	Business Analytics and Decision Sciences (31821)								
Module Leader:	Aritad(Alan) Choichoroon								
Declared Word Count:	2850								

Please Note:

Your declared word count must be accurate, and should not mislead. Making a fraudulent statement concerning the work submitted for assessment could be considered academic malpractice and investigated as such. If the amount of work submitted is higher than that specified by the word limit or that declared on your word count, this may be reflected in the mark awarded and noted through individual feedback given to you.

It is not acceptable to present matters of substance, which should be included in the main body of the text, in the appendices ("appendix abuse"). It is not acceptable to attempt to hide words in graphs and diagrams; only text which is strictly necessary should be included in graphs and diagrams.

By submitting an assignment you confirm you have read and understood the University of Leeds **Declaration of Academic Integrity** (http://www.leeds.ac.uk/secretariat/documents/academic_integrity.pdf).

PART 1: AUTONOMOUS SHIPMENT ROLLOUT: DELIVERY TRIAL

INTRODUCTION TO THE BUSINESS PROBLEM:

Autonomous Shipment, a brand-new startup company based in Leeds plans to employ autonomous robot drones for last-mile delivery of various things to client's doorsteps. It is thought that automation would help customers by enabling speedier delivery, and that the business will gain from cost savings and increased efficiency. The UK government is backing this new enterprise in addition to several venture capital firms. The company has created four robot prototypes that they plan to test and the technical specifications and the importance for each criterion decided on July 2023 has been considered for analysis. The management has proposed a one-month trial period. The following challenges must be solved in the report:

- To choose a prototype robot that meets certain criteria to take part in this testing.
- To choose how many robots to distribute throughout the different stores based on trials objective and restriction.

TASK 1: CHOOSING THE PROTOTYPE ROBOT

Understanding the task:

It is acknowledged that the initial duty is to select a prototype using standards set by management. The following table lists each robot's technical specifications (see Table 1.1):

Robot Prototype	Unit	Archer	Bowler	Corner	Deviant
Carrying Capacity	litres	45	50	60	40
Battery Size	hours	18	18	12	24
Average Speed	km/hr	6	4	4	10
Cost Per Unit	GBP	5210	6250	4500	7100
Reliability	hours	22	24	24	32

Table 1.1

Methodology:

The multi-attribute method has been used since the task requires a numerical output score for each alternative to determine which is the most efficient. One of the greatest multi-attribute strategies for choosing the best option is the Weighted Sum Method. The Weighted Sum Method assigns weights to different performance factors, and individual scores are multiplied by these weights to calculate a weighted sum. The formula used is as follows:

$$A_i = \sum_{j=1}^{n} w_j a_{ij} \quad for\ i = 1, 2, \ldots, m$$

Where n is the number of performance factors and w is the weight for each criterion j and a is the decision variable for the alternative i and criterion j. A is the final score for the alternative i.

Performing factors and weights:

Carrying capacity receives 10% weight because it is the least significant performance component based on management priorities. The average speed is the second least important factor as per the management and it is given 15% of the total weight. The battery size is given 20% of the total weight as it is the third most important factor as per management priorities. Because that is how the management desired it, the cost per unit is given 25% of the weight. Finally, reliability is the most important factor considered by the management and has been provided 30% of the total weight. The overview for the weights and performance factors is provided in the table 1.2 below:

Performance Factor	Weight
Carrying Capacity	0.10
Battery Size	0.20
Average Speed	0.15
Cost per Unit	0.25
Reliability	0.30

Table 1.2

Calculation:

The weights are applied for the performance factors on each alternative. In the case of cost per unit, the least value indicates the best, whereas the largest value indicates the best for other parameters. Hence, the cost per unit values are inversed. The range of values is normalized to 0 to 1 and the final weighted sum for the alternatives are calculated using the formula mentioned in the methodology. The scores are listed below in the figure 1.1.

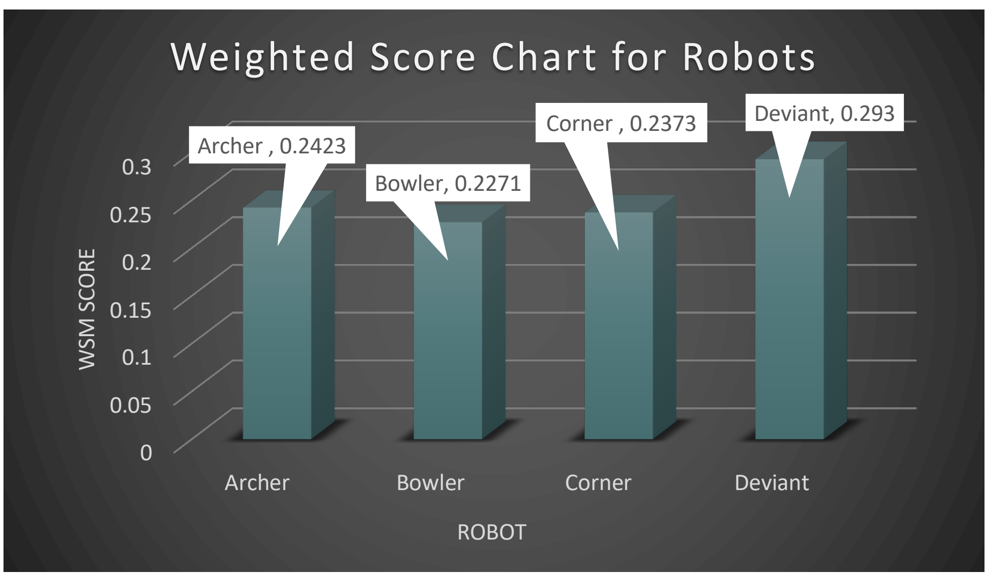

Figure 1.1

Verdict:

From the weighted score chart (fig 1.1), deviant scored the highest among the alternatives if the management priorities are truly considered. Therefore, among the options, the deviant robot is highly suggested for greatest utility in terms of both quantity and quality.

TASK 2: CHOOSING THE QUANTITY OF ROBOTS FOR EACH STORE DISTRIBUTION

Understanding the Task:

It is acknowledged that the number of robots assigned to each store must be allocated according to the following objectives and limitations:

- For the trial, a minimum of 5 robots must be present in each store.
- The trial should have robots complete as many orders as possible per day.
- The total number of technician staff hours available to support this trial is 250 hours per week.
- Both acquisition and operating costs must stay within the allocated funds. (250000 GBP)

Methodology:

Optimization is the best approach to address this problem as it involves finding the ideal values within the limitations that have been provided. One of the best optimization strategies is goal programming, which seeks to achieve every goal with the fewest possible deviations. The objectives must be expressed as a formula, and the target variables are computed using linear optimization to determine the optimal solution.

Equation 1 for the trial: (Based on number of orders per day)

It is clear from the information provided that the grocery store can deliver nine orders per robot each day. Additionally, a clothes store's robots can deliver six orders each day. Furthermore, the sport equipment store can fulfill four orders per robot per day. The total number of orders per day is initially aimed for 200 because the total number of orders per day has not been specified but has been asked to be as high as possible. So the following equation should be:

$$9X1+6X2+4X3 \pm D1 = 200$$

Where,

X1 → Number of deviant robots for grocery store

X2→ Number of deviant robots for clothing store

X3→ Number of deviant robots for sport equipment store

D1→ Deviation variable

Equation 2 for the trial: (Based on total cost)

Operating cost of a robot per month for grocery store is 1600 GBP. Similarly for clothing store , operating cost of a robot per month is 1000 GBP. Moreover for sport equipment store , operating cost of a robot per month is 600 GBP. The price of deviant robot is added to the equation to compute the total cost . The total cost is aimed for 250000 GBP as per the limitation. So the following equation should be:

$$(7100+1600)X1 +(7100+1000)X2+ (7100+600)X3 \pm D2=250000$$

$$\Rightarrow \quad 8700\,X1+ 8100\,X2 + 7700X3 \pm D2= 250000$$

Where,

X1 → Number of deviant robots for grocery store

X2→ Number of deviant robots for clothing store

X3→ Number of deviant robots for sport equipment store

D2→ Deviation Variable

Equation 3 for the trial: (Based on working hours)

Staff hours per week for grocery store is ten hours. Similarly for clothing store, staff hours for clothing store is seven hours per week. Moreover for sport equipment store, staff hours per week is five hours. The total available staff hours for a week is 250. The following equation is:

$$10X1 + 7X2 + 5X3 \pm D3 = 250$$

Where,

X1 → Number of deviant robots for grocery store

X2→ Number of deviant robots for clothing store

X3→ Number of deviant robots for sport equipment store

D3→ Deviation Variable

Calculation:

The equations are calculated using goal programming, and a summary is shown in figure 1.2 below.

```
Problem formulation:
Orders : 9*x1 + 6*x2 + 4*x3 = 200 | 1 1 | 1# 1#
Cost : 8700*x1 + 8100*x2 + 7700*x3 = 250000 | 1 1 | 1# 1#
Hours : 10*x1 + 7*x2 + 5*x3 = 250 | 1 1 | 1# 1#

Objective function value: 20

Solution:
   value    type
x1    17 integer
x2     5 integer
x3     8 integer

Deviations:
       d- d+ w- w+ p- p+
Orders  0 15  1  1  1  1
Cost    0  0  1  1  1  1
Hours   5  0  1  1  1  1
```

Figure 1.2

Observation:

According to the summary, the solution obtained is 17 deviant robots for the grocery store, 5 deviant robots for the clothing store, and 8 deviant robots for the sporting equipment store. The total budget used in this scenario is 250000 GBP, and the total staff hours used is 245 hours. According to the deviation, the total number of orders that can be delivered in a day is 215.

Evaluation:

Now the obtained solution is evaluated with the provided constraints. The validation is provided in the table 1.3 below.

Expected Goal		Achieved Solution	Condition Pass/Fail
Number of Deviant Robots per store	Grocery Store ≥ 5	Grocery Store = 17	Pass
	Clothing Store ≥ 5	Clothing Store = 5	
	Sport Equipment Store ≥ 5	Sport Equipment Store = 8	
Total Staff Hours ≤ 250		Total Staff Hours = 245	Pass
Total Cost ≤ 250000 GBP		Total Cost = 250000	Pass
Total orders per day should be as many as possible(Aimed for 200)		Total orders per day = 215	Pass

Table 1.3

Verdict:

From the evaluation, it is clear that the achieved optimal solution has satisfied all the goals and constraints provided.

SUMMARY OF RESPONSE TO MANAGEMENT OF THE COMPANY:

Overall from the analysis, deviant is strongly recommended as the most effective robot among the alternatives for the trial. Moreover, thirty deviant robots should be bought and seventeen should be supplied to the grocery store, five should be supplied to the clothing store and eight should be supplied to the sport equipment store to satisfy all the management constraints and goals. With this approach, the trial has high chances to get succeed for the Autonomous Shipment. The solution is tailored to enhance the cost effectiveness and maximize the resource utilization.

BUSINESS UNDERSTANDING:

"Drinks@home.uk" is an online store that sells beverages directly to customers. It works in the Great Britain region and handles both alcoholic and non-alcoholic beverage products. The data supplied includes the amount of money made from the order they placed, the advertising medium that directed them to the website, their age, their income, the average amount of time they have spent on the website over the course of a week, and whether they have previously encountered an online voucher pop-up. The report mainly focuses on two goals for the next marketing campaign:

1. To identify the factor that has substantially influenced consumers decision to spend more money on the "drinks@home" website.
2. To determine which of the three choices should be used for the upcoming campaign in order to maximize profits. The choices are listed below:
    - To target consumers aged over 45 with an advertisement since they are more likely to be high rollers.
    - To give consumers a coupon voucher for 20 GBP off their next purchases.
    - To increase the investment on influencer marketing.

DATA UNDERSTANDING:

Records of 400 customers' revenue, the advertising medium that brought them to the website, their age, income, the average amount of time they spend on the website each week, and whether or not they have ever seen an online voucher pop-up are all included in the dataset which can be useful for exploratory analysis. Luckily the dataset didn't have any missing values or incomplete records and almost remains perfect with respect to the data quality. However there is a small noise which can be noted in figure 1.3 as a customer aged 9 has ordered drinks for 81 GBP which practically seems impossible.

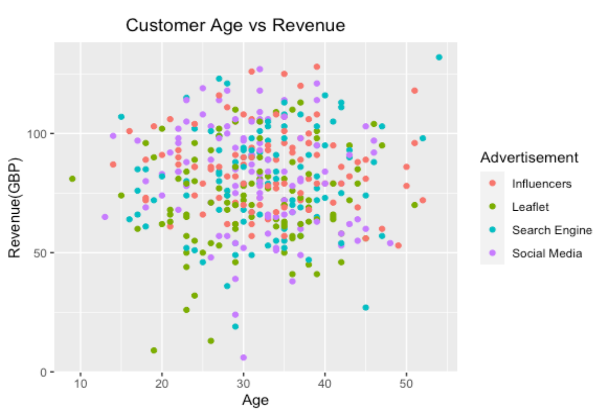

Figure 1.3

Figure 1.4

From figure 1.3 , it is understood that most of the customers are aged from 20 to 40(can also be interpreted in figure 1.5). Also there is no evidence to prove that high paying customers are aged above 45. Moreover customers influenced by different advertisement channels are equally distributed. We can observe that all the customers who got influenced by influencers have paid more than 50 GBP which shows the importance of influencers in the sales. From figure 1.4, it is observed that most of the customers has spent time between

100 to 200 seconds on the website. Moreover it is evident that customers who has seen the voucher popup has spent more money than who hasn't seen. These inputs can be considered for choosing the ideal choice for the next marketing campaign.

Figure 1.5

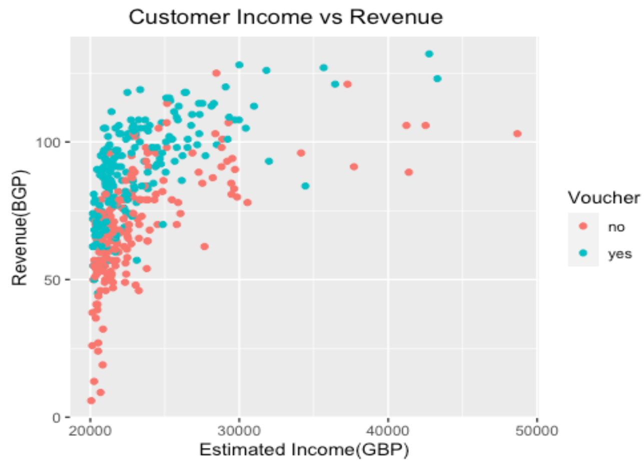

Figure 1.6

Figure 1.6 conveys that most of the customers income is between 20000 to 30000 GBP per annum. However a small trend is observed that increase in annual income affects the revenue in a positive manner. This can be a crucial input for modelling.

Note: To find ideal choice for marketing campaign, please refer page no 10.

DATA PREPARATION:

From the data understanding, customer's estimated is considered as one of the factors that affect the revenue. An additional proof is shown in the figure 1.7. Moreover the correlogram of the dataset is displayed in figure 1.8 to find the crucial factors for modelling.

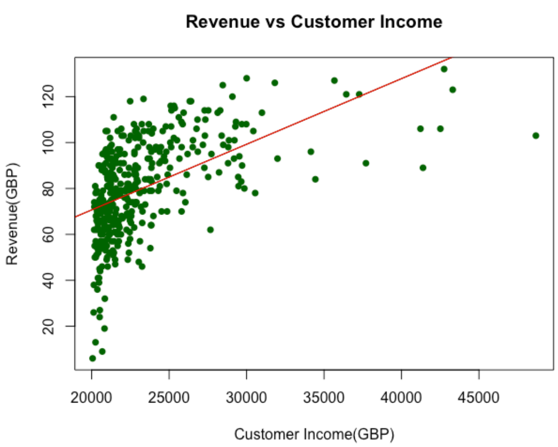

Figure 1.7

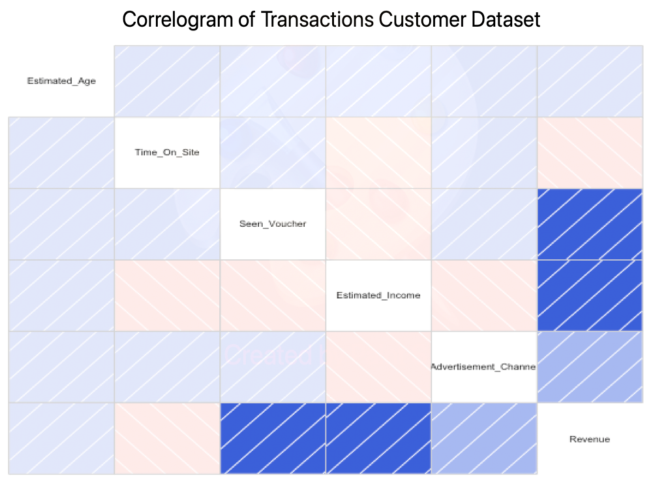

Figure 1.8

From figure 1.8 , Estimated Income, Seen Voucher and Advertisement Channel has been identified as the crucial factors that affects the revenue in positive manner. The estimated income variable is transformed into logarithmic values as it gives a better form to the model. The identified variables are finalized as the independent variables and revenue as the dependent variable for the model. The seen voucher variable and advertisement channel variable has been considered as factors as they belong to categorical variables. Now the independent and dependent variables are fixed and ready for modelling.

MODELLING:

The aim of the model is to predict the revenue with the help of estimated income, seen voucher and the advertisement channel and make sure that the variables affect the revenue in a positive or negative way. The dataset is split into two parts called the train dataset(70% of overall) and the test dataset(30% of overall). To predict the revenue, Linear Regression under supervised learning is chosen as the ideal choice for modelling as the revenue is a numerical variable. The model variables are represented below in the table 1.4:

Dependent Variable	Independent Variable
Revenue	log(Estimated Income)
	Seen Voucher(As Factor)
	Advertisement channel(As Factor)

Table 1.4

The model is fit in the dataset and the summary is shown in the figure 1.9 below:

```
Call:
lm(formula = Revenue ~ log(Estimated_Income) + Seen_Voucher +
    Advertisement_Channel, data = traindata)

Residuals:
    Min      1Q  Median      3Q     Max
-48.744  -7.787   0.536   7.669  34.739

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -844.4302    57.3987 -14.712  < 2e-16 ***
log(Estimated_Income)   89.9097     5.7104  15.745  < 2e-16 ***
Seen_Voucher            19.9330     1.5260  13.062  < 2e-16 ***
Advertisement_Channel    4.2357     0.6848   6.186 2.22e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.76 on 276 degrees of freedom
Multiple R-squared:  0.627,     Adjusted R-squared:  0.623
F-statistic: 154.7 on 3 and 276 DF,  p-value: < 2.2e-16
```

Figure 1.9

The final derived model is stated below:

Revenue = 89.9*log(Estimated_Income) + 19.933*Seen_Voucher + 4.23*Advertisement_Channel - 844.43

EVALUATION:

The multiple R squared value and adjusted R squared value is found to be 0.627 and 0.623(see figure 1.9) respectively. This means that model can explain 62% of variation within revenue. The t- value for the coefficients lies outside the 95% CI range [-1.96,+1.96]. The p value for all the coefficients lie below 0.05 which is quite efficient for 95% confidence level. The model is now used to predict the revenue of test data and the root mean square error is 14.9582 which looks good fit for a model. The root mean square error calculation is given in the figure 1.10.

```
> rmse<-sqrt(mean((prediction-testdata$Revenue)^2))
> rmse
[1] 14.95862
```

Figure 1.10

This proves that the model's independent variables are capable enough to predict the revenue which indirectly means that they strongly affect the dependent variable.

IDEAL CHOICE FOR NEXT MARKETING CAMPAIGN:

Option 1: Targeting Customers over 45

From figure 1.3 and 1.5 , it is already understood that most of the customers are aged between 20 to 40. Moreover there is no evidence to prove that customers who are elder than 45 are high paying customers. Also customer age is not a good independent variable for the model as per the correlogram (see figure 1.8). So this will not be an ideal choice for the next marketing campaign.

Option 2: Providing 20 GBP Vouchers

The customers who has seen the pop up voucher has given more revenue than customer who hasn't seen. A supporting proof for this claim is presented in the boxplot given below(figure 1.11):

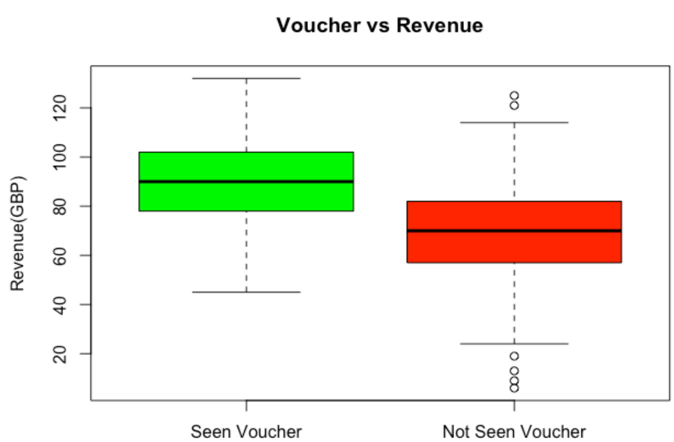

Figure 1.11

However there is no proof that all the customers who has seen the voucher in the pop up has used it. The statement provided only interprets that the customers has seen it and not used it. So unfortunately this option should be ruled out from considering in the next campaign.

Option 3: Increasing the investment on influencer marketing

All the customers who has ordered the drinks through influencers has spent more than 50 GBP. A supporting proof is already provided in figure 1.3(see page no 7) and figure 1.12 below.

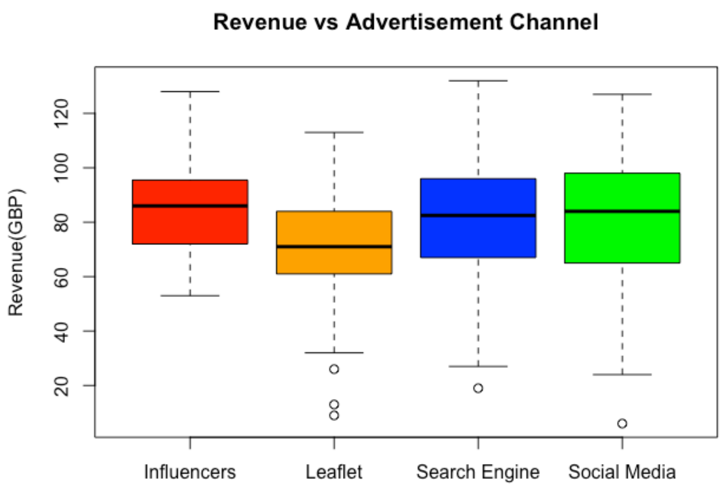

Figure 1.12

So this option will be the ideal choice for the next marketing campaign and holds high chances to increase the profits.

CONCLUSION:

From the overall analysis, the two main business questions have been answered. The brief summary of the business questions and their respective answers are given below:

1. Which factor influences the customer to spend more money on the Drinks@home.uk website?
   Ans:
   - Customer's annual income
   - Visibility of pop up vouchers
   - Advertisement channels that influence them

2. Which option will be the ideal choice for the next marketing campaign?
   Ans: Spending more money on the influencers will be the perfect choice for the next marketing campaign as it makes the customers to spend more money on the website.

APPENDIX: (NOT INCLUDED IN THE WORD COUNT)

The R code for the part 1 and part 2 is provided below for better understanding:

```
##Part-1-Task 1-MCDA-Weighted Sum Method

library("MCDA")

robot<-read.csv("Robot_Info.csv",header=TRUE)

management<-read.csv("Management_Priority.csv",header=TRUE)

pftable<-t(robot)

pftable<-pftable[c(2:5),]

pftable<-data.frame(pftable)

pftable<-sapply(pftable,as.numeric)

pftable[,c(4)]<-pftable[,c(4)]^-1

weight<-c(0.10,0.20,0.15,0.25,0.30)

pftable<-pftable/colSums(pftable)[col(pftable)]

overall<-weightedSum(pftable,weight)

names(overall)<-colnames(robot[,c(2:5)])

barplot(overall,main = "WSM MCDA Result for robots",xlab = "Robot model",ylab="WSM
Score",col=c("darkblue","red","green","cyan"),legend=names(overall))

##TOPSIS(An Alternative approach for cross verification)

ptable<-t(robot)

ptable<-ptable[c(2:5),]

ptable<-data.frame(ptable)

ptable<-sapply(ptable,as.numeric)

rownames(ptable)<-names(overall)

criteria<-c("max","max","max","min","max")

names(criteria)<-robot[,c(1)]

overallT<-TOPSIS(ptable,weight,criteria)

names(overallT)<-names(overall)

barplot(overallT)
```

```
##Task 2-Goal Programming

library(goalp)

goals<-"Orders:9*x1+6*x2+4*x3=200

    Cost:8700*x1+8100*x2+7700*x3=250000

    Hours:10*x1+7*x2+5*x3=250"

opt<-goalp(goals)

summary(opt)

##Part 2-CRISP-DM

customers<-read.csv("Transactions_Customer.csv",header = TRUE)

customers$Voucher<-ifelse(customers$Seen_Voucher==1,"yes","no")

customers$Advertisement<-
ifelse(customers$Advertisement_Channel==1,"Leaflet",ifelse(customers$Advertisement_Channel==2,"Soci
al Media",ifelse(customers$Advertisement_Channel==3,"Search Engine","Influencers")))

ggplot(data=customers)+geom_point(aes(Estimated_Income,Revenue,color=Voucher))+labs(title = "
Customer Income vs Revenue",x="Estimated Income(GBP)",y="Revenue(BGP)")

ggplot(data=customers)+geom_point(aes(Revenue,Estimated_Age,color=Voucher))+labs(title = "
Customer Age vs Revenue",x="Revenue",y="Age")

ggplot(data=customers)+geom_point(aes(Revenue,Estimated_Income,color=Advertisement))+labs(title = "
Customer Income vs Revenue",x="Revenue",y="Estimated Income")

ggplot(data=customers)+geom_point(aes(Estimated_Age,Revenue,color=Advertisement))+labs(title = "
Customer Age vs Revenue",x="Age",y="Revenue(GBP)")

ggplot(data=customers)+geom_point(aes(Revenue,Time_On_Site,color=Adverstisement))+labs(title = "
Time Spent On Site vs Revenue",x="Revenue",y="Time Spent")

ggplot(data=customers)+geom_point(aes(Time_On_Site,Revenue,color=Voucher))+labs(title = "      Time
Spent On Site vs Revenue",x="Time Spent(sec)",y="Revenue(GBP)")

customers$Voucher<-as.factor((customers$Voucher))

customers$Advertisement<-as.factor(customers$Advertisement)

ggplot(data=customers)+geom_point(aes(Estimated_Income,Estimated_Age,color=Advertisement))+labs(t
itle = "      Customer Age vs Customer Income",x="Income",y="Age")

influencers<-customers[which(customers$Advertisement=="Influencers"),]

leaflet<-customers[which(customers$Advertisement=="Leaflet"),]

searchengine<-customers[which(customers$Advertisement=="Search Engine"),]
```

```
socialmedia<-customers[which(customers$Advertisement=="Social Media"),]

yesvoucher<-customers[which(customers$Voucher=="yes"),]

novoucher<-customers[which(customers$Voucher=="no"),]

boxplot(influencers$Estimated_Age,leaflet$Estimated_Age,searchengine$Estimated_Age,socialmedia$Esti
mated_Age,col = c("red","orange","blue","green"),names= c("Influencers","leaflet","Search Engine","Social
Media"),ylab="Age",xlab="Mode of Advertisement")

boxplot(yesvoucher$Estimated_Age,novoucher$Estimated_Age,col = c("green","red"),names = c("Seen
Voucher","Not Seen Voucher"))

hist(customers$Estimated_Age,col = "darkgreen",main = "Histogram of Customer Age",xlab = "Age")

boxplot(influencers$Revenue,leaflet$Revenue,searchengine$Revenue,socialmedia$Revenue,names =
c("Influencers","Leaflet","Search Engine","Social Media"),col =
c("red","orange","blue","green"),ylab="Revenue(GBP)",main="Revenue vs Advertisement Channel")

boxplot(yesvoucher$Revenue,novoucher$Revenue,col = c("green","red"),names = c("Seen Voucher","Not
Seen Voucher"),ylab="Revenue(GBP)",main="Voucher vs Revenue")

corrgram(customers)

corrgram::corrgram(customers)

plot(log(customers$Estimated_Income),customers$Revenue,main = "Revenue vs Customer Income",xlab =
"Log(Customer Income(GBP))",ylab = "Revenue(GBP)",pch=16,col="darkgreen")

abline(model,col="red")

datasplit<-customers

datasplit$id<-1:nrow(datasplit)

traindata<-datasplit%>%sample_n(280)

testdata<-anti_join(datasplit,traindata,by="id")

traindata$Voucher<-as.factor((traindata$Voucher))

traindata$Advertisement<-as.factor(traindata$Advertisement)

model<-lm(Revenue~log(Estimated_Income)+Seen_Voucher+Advertisement_Channel,data=traindata)

prediction<-predict(model,newdata = testdata)

rmse<-sqrt(mean((prediction-testdata$Revenue)^2))
```