REPORT ON AUTONOMOUS SHIPMENT ROLL-OUT'S PROTOTYPES ROBOT TRAILS

SUMMARY

This is a detailed report of the one-month trial autonomous shipment in Leeds, this report aims to provide the identification of the best robots for deployment and optimal operational strategy.

Task 1:

The optimal robot for deployment, using the pre-specified criteria by the management team is **'Deviant'**. Using the MCDA model Topsis the Deviant had an overall well-balanced score.

Task 2: Operational Strategy

The ideal allocation for stores like Grocery, Clothing, and sports equipment are **19, 5, and 5** respectively. This solution was found using Lexicographic Goal Programming, which would ideally use **244,300** of the budget with **221 orders**, without any change to technician staff hours required.

INTRODUCTION TO BUSINESS PROBLEM

This report is based on the analysis done for Autonomous Shipment Company. The purpose of this report is to identify behaviour key aspects of the one-month trial that the management has planned to execute in Leeds. Autonomous Shipment is a new start-up venture that is funded by several venture capital investors and by the United Kingdom's government, the automation of the company would help consumers enjoy faster delivery.

There are multiple options to consider as well as multiple criteria. The task here is to deploy the robots for different stores like groceries, clothing, sports, and tech stores. There are four types of robots which are Robot A032 – Archer, RobotB23 - Bowler, Robot CJKL – Corner, and Robot DSXX – Deviant to consider out of which one will be trailed for a month. This report addresses 2 tasks that the management team is required to fulfil.

1. The decision on a prototype robot that would participate in this trial is based on a set of requirements.
2. The decision of how many robots to allocate across various stores ensures that the goal and constraints of the trial are satisfied.

TASK 1

INTRODUCTION

We have identified 4 robots for this trial rollout. To understand which robot is to be used for this trial run, the robots are categorized into four categories. They are Carry capacity, Battery size, Average speed, Cost per unit, Reliability, software, and other factors are not taken into account while this analysis is done. The details of this are elaborated in the table (Robots 1.1).

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

Table 1.1 Robots

DECISION CRITERIA'S

The management team has come up with several pre-specified criteria that needed to be considered, these criteria had different weights to them, and this was done based on the ranking that they had suggested. These weights are used to prioritize these criteria which ever was considered the most weighted criterion was ranked the most with the most percentage value and vice versa for the least weighted. The weights are later considered for the MCDA models, which are used to identify the best model out of these.

SL No	Weights	Category	Details
1	30%	Reliability	This is the average amount of time that occurs between malfunctions.
2	25%	Cost per Unit	The price per unit (robot) is in British pounds. The robot with the lowest cost per unit is what the corporation would desire.
3	20%	Battery Size	Each autonomous delivery robot's battery capacity is expressed in operating hours. The robot with a bigger battery is preferred.
4	15%	Average Speed	This is the average speed of every robot, expressed in km/h. The robot with a greater average speed is what a company would desire.
5	10%	Carrying Capacity	Given that some products might be enormous, the company would prefer to utilize a larger robot.

Table 1.2 Weights

DATA ANALYSIS USING MCDA

The model that I have identified to be best for this is **Topsis** (*Technique for Order of Preference by Similarity to Ideal Solution*). Topsis is the method in which the model identifies two points which ideal in nature these points are unachievable in reality but this would help us find two points that can later be used to identify the Positive ideal number and Negative ideal number. Topsis taken because of the multiple criteria's that each of the robots has. This allows us to incorporate multiple weights to the criteria specified by the management team, and this model allows us to use sensitive analysis to identify the optimal solution without bias which adds to the understanding of the best alternative.

ASSUMPTIONS

Here the weights of the criteria are not specified in numbers, the weights are given as ranks and approximate values. From that, we have derived the table (Weights table 1.2) which in total is 100% of the criteria.

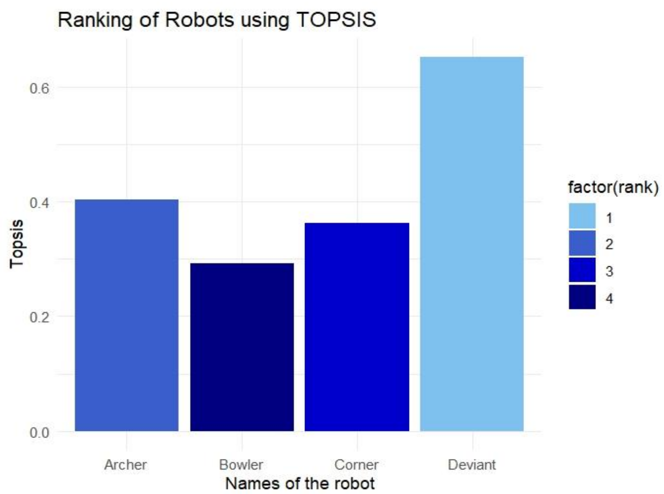
RECOMMENDATIONS

Based on the criteria pre-specified by the management team, using Topsis the optimal choice for the robot to be used in the trial is **'Deviant'**. The scores that each of the robots exhibited when compared with each other are shown in the table (1.3 results)

Rank	Name	Values
1	Deviant	0.654382
2	Archer	0.406787
3	Corner	0.36372
4	Bowler	0.299851

- The deviant prototype is the most well-balanced choice, having a clear difference when compared to other robots
- With the pre-specified criteria in check, the top three being Reliability, Carry cost per unit, and Battery size with a combined score of almost 75%.

LIMITATIONS



One of the limitations of this MCDA model is that the weights of the criteria that were pre-specified by the management team are subjective and if changed the result might also change.

Here this is not true as the Robot Deviant outranks all the robots when a sensitive analysis with all the criteria given as 20%, the result were the same, but when it comes to the other ranks like Archer and Corner some weightage may prove to have a different outcome.

Overall, the Deviant is the optimal solution for this trial and the Bowler is the worst option.

While considering the weightage we are only considering the quantifiable metrics like cost per unit, carrying capacity, or battery size, here there is no importance given to the qualitative aspects like customer experience or the use of robots in an urban environment.

TASK 2

INTRODUCTION

The utilization of robots that can deliver which the need of human interference would make the optimization of the company's logistics. This task is to identify the best coordination of the trial deployment issue of the autonomous robots across stores such as Grocery, Clothing, and Sports Equipment stores. For this one-month test run we have already identified the robot to be used as 'Deviant', given that we are moving forward with this trail with the above-discussed robot, the objective is this trail more optimization rather than profit-oriented.

CHALLENGES AND OBJECTIVES

There are some constraints and targets to achieve,

1. Each store must have at least 5 robots.
2. The number of orders that are not specific but to have a successful full trial run the maximum number of orders is to be achieved.
3. The budget and the working order of the technicians must not exceed expectations.

DECISION CRITERIA'S

The management team identifies a few constraints,

The A, B, and C here denote each of the Grocery, Clothing stores, and Sports Equipment Stores. Each of these stores must have at least 5 robots.

The objective is to maximize the **number of orders,** within the constraints. These constraints are operating costs, which are Cost per unit (of Deviant) added to the cost of operations in each store, and finally the number of technician staff's working hours. These constraints are made into a linear programming equation to find the optimal solution.

RANK	CATEGORY	CONSTRAINT	EQUAL TO
1	Orders	9*A + 6*B + 4*C	>= 0
2	Budget	8700*A + 8100*B + 7700*C	<= 250000
3	Man hours	10*A + 7*B + 5*C	= 250

METHODOLOGY

Since the goal here is to identify the best solution, keeping in mind the multiple constraints the MCDA model used for this analysis is **Lexicographic goal programming (LGP).** In this method, the constraints are introduced into a hierarchy of priorities then the best solution is found sequentially. Here the constraints are ordered as to their priorities,

1. Number of orders
2. Cost of operations
3. Man-hours

The MCDA model here identifies the priority and then finds an optimal solution to the constraints as a multiple linear equation. Here the LGP is an ideal choice as there are multiple constraints and the priorities can be ordered in a hierarchical manner which can satisfy the constraints. LGP can help to identify a structured approach to handle multiple constraints by addressing their importance which is satisfying the highest priority to the lowest.

RESULTS & CONCLUSIONS

Based on the analysis, Lexicographic goal programming the best solution is to roll out **221 orders**. With a deviation to the **cost of operation** which is **244300** and with the technician's hours is **250 hours per week**. With the given constraints this is the most ideal solution. Here the budget does not exceed the 250,000 limit, and the maximum number of orders can be achieved with a reduction of 5700.

Even when the budget is given priority with the technician hours given as 250 as the second priority and then to maximize the order the same values are achieved by the LGP.

- The objective function's value achieved is 0, which means the solution meets all the pre-defined constraints.
- The optimal solution found by the model was,
    - A = 19 (Grocery shop)
    - B = 5 (Clothing stores)
    - C = 5 (Sports Equipment stores)

Orders that can achieved in a day

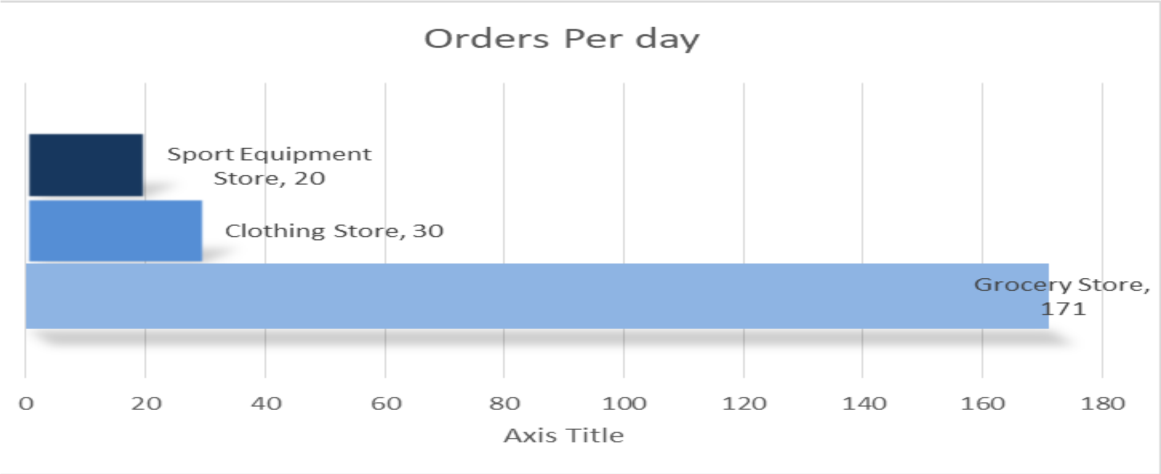


Figure (1.1)

Budget of the total operations

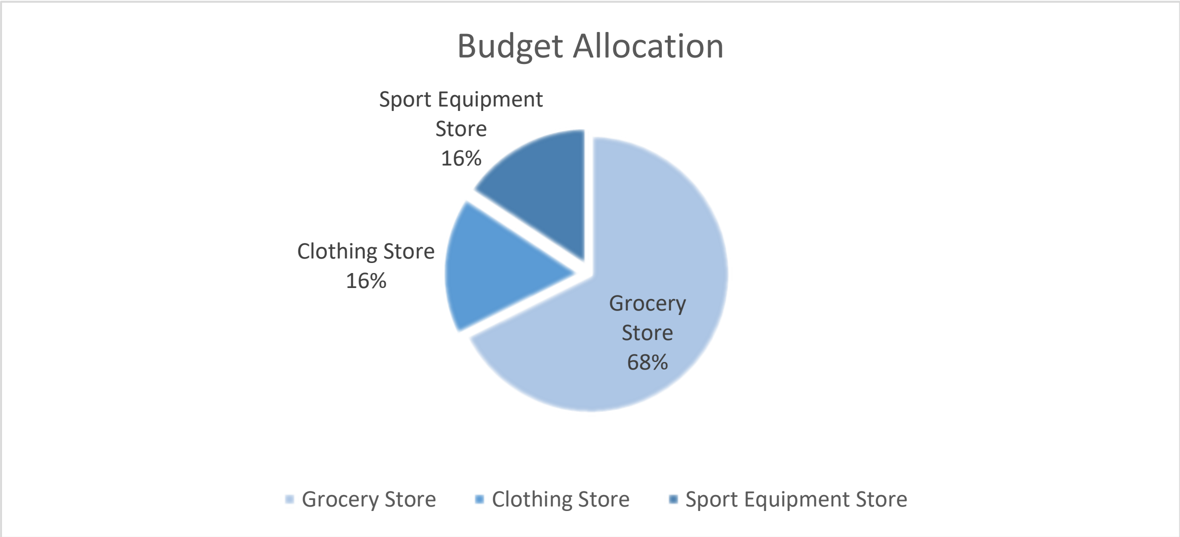


Figure (1.2)

The relation between order and the technicians available

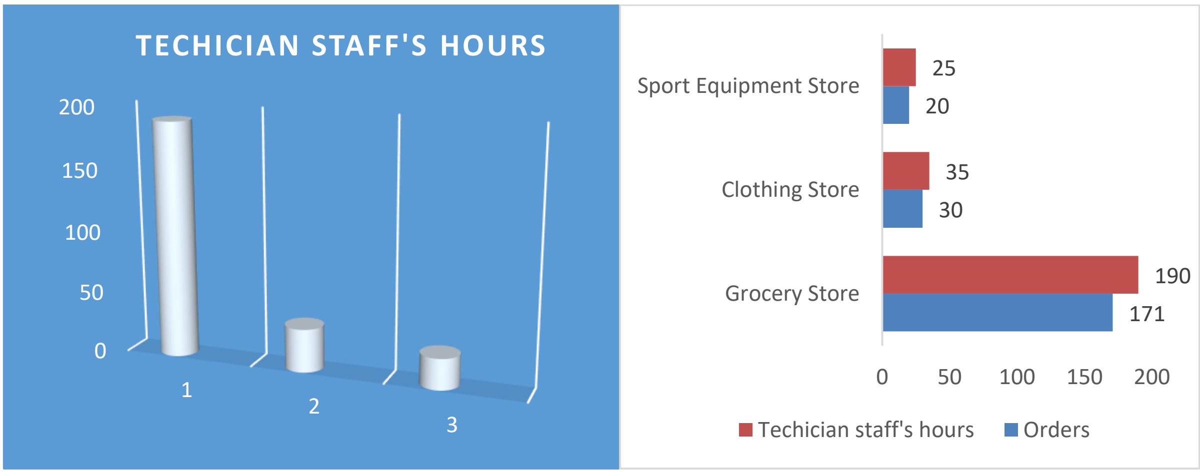


Figure (1.3)

REPORT ON DRINKS@HOME CUSTOMER BASE AND MARKETING STRATEGIES

SUMMARY

This report is based upon the findings for drinks@home an e-commerce business. This report deals with the key factors of customer behaviour and demographics when compared with revenue generated on the site. Analysing factors such as the customer database age, Income, and how traffic was generated from the marketing campaigns.

From the analysis independent variables such **as Seen voucher, Estimated income, Social media, Search engine, and Influencer** have a positive influence in notably increasing the revenue.

To identify the optimal recommendation for the marketing strategy when analysed with factors such as Influencer, Social media marketing and even Search engine, offering a 20 GBP discount as a **voucher** has the most significant increase in Revenue generation.

INTRODUCTION

This is a report based on Drink@home.uk which is a United Kingdom-based company that operates in the e-commerce business, with direct sales through their website. The product line consists of both alcoholic and non-alcoholic drinks that are delivered from throughout the world. The data of over 400 customers are used in this report to identify the following,

1. To identify the factors that lead the customer to spend money more or less on the website. With the use of the demographic and past behaviours of the customer.
2. Recommend the best option for the next marketing project, from the 3 options to increase the profit.

DATA UNDERSTANDING

The data of 400 customers have been used in this analysis, this data is given in 5 categories Estimated Age, Time on Site, Seen Voucher, Estimated Income, Advertisement Channel, and Revenue.

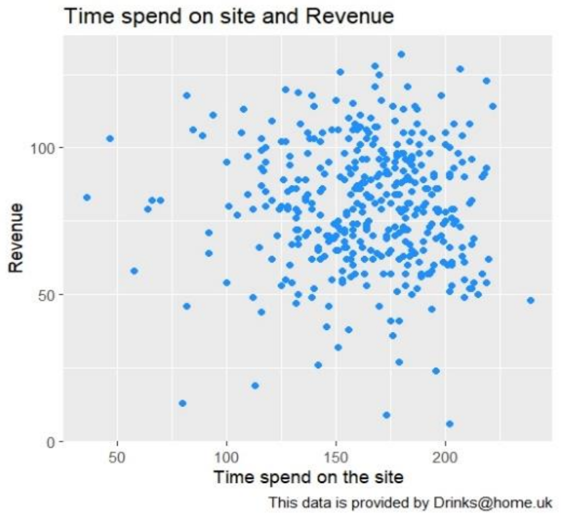
Category	Details	Classified as
Estimated age	Based on the tracking software their age	
Time on site	Average time spent per week	Seconds
Seen voucher	Used a discount voucher or not	Categorical data
Estimated income	Based on the tracking software customer's income	GBP
Advertisement channel	As multiple had done in the past and currently ongoing the advertisement channel is again divided into 4, they are Leaflet, Social Media, Search Engine, Influencer	Categorical data
Revenue	The last order made on-site	GBP

Here the revenue is taken as the dependent variable and the rest of them are taken as an independent variable.

To understand the factors that affect the customer's spending habits the relation between the customer's behaviour and demographics better. Here we are plotting each of the data categories to revenue the customer had spent on the site. This is to identify the strengths and direction of the relation between 2 variables.

Here are the following relations that can be established,

- Time spent on the site has a slight relation to revenue as the customer spends about 150-200 seconds the customer tends to buy a lot more (Figure 2.1).
- The income of the customer has a positive effect on the customer's spending or revenue generated on the site (Figure 2.4).
- The estimated age does not a much of a relation but has a slight relation as the age increases we can see that the customer the more they spent but this drops down as the customer's age is more than 45 years.(Figure 2.2)
- Seen voucher has a positive relation with the revenue. As we can see the chance of purchase is higher when a seen voucher is used. (Figure 2.3)



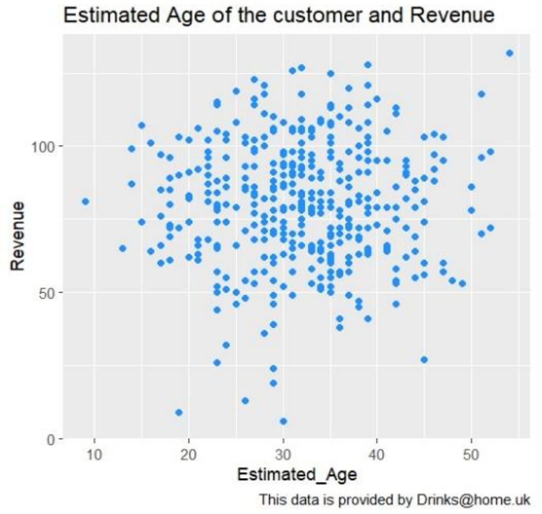
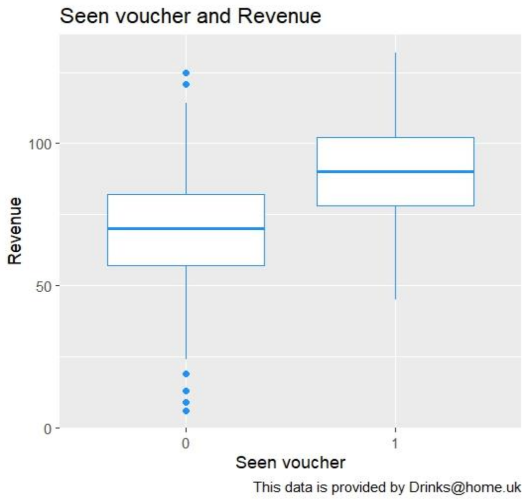


Figure (2.1) **Figure(2.2)**



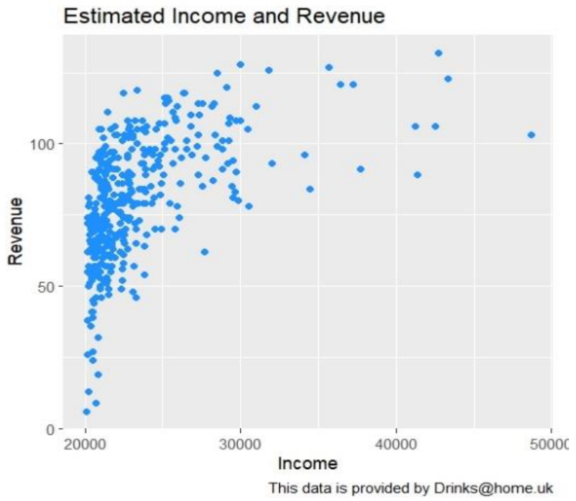
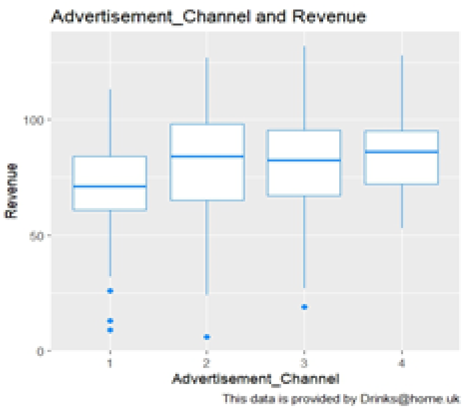


Figure (2.3) **Figure(2.4)**



Figure(2.5)

DATA PREPARATION

Data used here is in the file **'Transactions_Customer.csv'.** This data is loaded into R and the data is checked for any missing values, and then checked for any cleaning or any transformation is needed.

CHECKING FOR ANY MISSING DATA

When checked, the data does not have any missing data, all 400 datasets are present. The sum(complete.cases(data) is used to count all the number of cases in the dataset. Here a complete set is referred to as there are no missing values in a row. The sum of the complete rows is given if the row is complete.

DATA TRANSFORMATION

The data loaded into R has no missing values and is now ready for the transformation. The Advertisement channel has been divided into four as per Drinks@home.uk, they are Leaflets, Social media, Search engine, and Influencer. These are the different marketing activities that generate traffic to the website.

To better understand the advertisement channels' impact on the customer, we can use dummy variables to separate each of the channels. This will introduce each channel into a new format of representation using 0, 1 which indicates their presence and absence. This would provide a more detailed perspective of the different channels and whether they have an influence or not.

MODELING

CORRELATION ANALYSIS

To identify the correlation of each other factor to revenue. We can plot them to find any correlation in a graph. Here the **corrgram plot** is taken to plot the data. This graph is with the dummy variables implemented to identify the best marketing strategy as well.

Figure (2.6)

The correlation of the graphs shows the following,

The analysis showed that seen voucher and estimated income with 0.531 and 0.465 respective scores means the two factors have a strong correlation to revenue. Other factors like Influencer have only moderate correlation with a much weaker score of 0.147.

REGRESSION ANALYSIS

For this analysis linear regression is used to identify the factors like R squared, Adjusted R, P value, T-test, and coefficients. The linear regression model is used to predict the revenue based on independent variables like Time on site, Seen voucher, Income, and other factors.

Here the intercept is 0.38772 which denotes the value of the dependent variable when all the independent variables are set to zero. The analysis shows that seen voucher, estimated income, social media, search engine, and influencers has a positive strong association with revenue.

The analysis conducted with t-test models the significance of the differences of the different datasets, which will emits the impact of the independent variable to the dependent variable revenue.

R squared value denotes that close to 55% of the model is fit and the adjusted R is 0.54 this value denotes the model's effectiveness when the independent variables are adjusted.

```
Residuals:
    Min      1Q  Median      3Q     Max 
-53.677  -7.657   1.429   8.967  40.283 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.3877205  6.3658721   0.061 0.951465    
Estimated_Age   -0.0152422  0.0894058  -0.170 0.864718    
Time_On_Site    -0.0221743  0.0219252  -1.011 0.312467    
Seen_Voucher    19.6954714  1.4145999  13.923  < 2e-16 ***
Estimated_Income 0.0028609  0.0001838  15.567  < 2e-16 ***
SocialMedia      6.8284251  2.0170930   3.385 0.000783 ***
SearchEngine     8.0909325  1.9997523   4.046 6.28e-05 ***
Influencer      12.9736091  2.0003277   6.486 2.66e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.09 on 392 degrees of freedom
Multiple R-squared:  0.5547,	Adjusted R-squared:  0.5467 
F-statistic: 69.74 on 7 and 392 DF,  p-value: < 2.2e-16
```

EVALUATION

CORRELATION ANALYSIS

IDENTIFIED FACTOR: SEEN VOUCHER & INCOME

Among the categories of data analysed for correlation, it is noticed that only income and seen vouchers have displayed a correlation between revenue spent on the site. The channels like seen voucher and estimated income have a very strong correlation to revenue indicated by their dark blue shade. Other factors like estimated age and time on site are not identified as having a strong enough correlation to revenue which is indicated in Figure (2.6). Hence we can say that estimated income and seen voucher has a positive impact on revenue.

OTHER FACTORS

There is a lack of solid correlation to no correlation in the different categories like estimated age and time on site.

REGRESSION ANALYSIS

Based on the analysis, there are several factors to be considered that are strongly associated with the dependent variable revenue, Variables such as Seen voucher with 19.69, Estimated income with 0.002, Social media with 6.828, Search engine with 8.09, and Influencer with 12.97 has both exceeding 1.96 for t-value and p-value less than 0.05, which represents a positive relationship that influences the revenue.

CONCLUSION AND RECOMMANDATION

Upon the analysis, the factors that influence Revenue on the website are identified. The linear regression model has identified variables like seen voucher, influencer, social media, search engine, and estimated income all had a positive impact on revenue, and variables like time on site, and estimated age had little to negative impact on revenue.

- When considering the fact that most of the customers are from the income group of 20000-25000 GBP which generates 50 to 100 GBP (Figure 2.4).
- Seen voucher shows a positive impact when the customer utilizes it.

- Various marketing campaigns like social media, search engine, influencer has positive effect on the revenue as more traffic can be generated.

Upton the evaluation of the all analysis, there a prominent impact that seen voucher when compared with the other variables is clear. The analysis suggest that the voucher for 20 GBP off their next purchase would have a more positive impact to the other marketing strategies. I propose prioritizing the implementation of this strategy as a primary recommendation. By integrating this approach and allocating additional funds towards influencer marketing, there is considerable potential for enhanced revenue generation. This dual approach would not only broaden the reach of the voucher but also amplify the brand image across a more expansive audience.

APPENDIX

PART 1

TASK 1 TOPSIS

```
install.packages("./MCDA_0.1.0.tar.gz")
install.packages(c('Rglpk','triangle','plyr','ggplot2','glpkAPI','combinat'))
install.packages('topsis')
library('ggplot2')
library('topsis')
# Data prep ####
data <- read.csv('Robot_Info(topsis).csv')
data
performanceTable <- data[c(2,3,4,5),c(2,3,4,5,6)]
performanceTable
class(performanceTable)
# Heading to Rows ####
row.names(performanceTable) <- c("Archer","Bowler","Corner","Deviant")
performanceTable

#Criteriaweights ####
weights <- unlist(data[c(1),c(2,3,4,5,6)])
names(weights) <- colnames(performanceTable)
weights

# criteriaMinMax ####
criteriaMinMax <- c("+", "+", "+","-", "+")
names(criteriaMinMax) <- colnames(performanceTable)
criteriaMinMax

# Final report ####
class(performanceTable)
performanceTable <- as.data.frame(as.matrix(performanceTable))
performanceTable <- sapply(performanceTable, as.numeric)
class(performanceTable)

result <- topsis(performanceTable, weights, criteriaMinMax)
result

# Changing names of the alt row ####
robot_names <- c("Archer","Bowler","Corner","Deviant")
row.names(result) <- robot_names
```

```
#Visual ####

# Plotting using barplot ####

# Create a bar plot for ranks of robots
ggplot(result, aes(x = factor(robot_names), y = score, fill = factor(rank))) +
  geom_bar(stat = "identity") +
  labs(title = "Ranking of Robots using TOPSIS", x = "Names of the robot", y = "Topsis") +
  scale_fill_manual(values = c("grey0", "grey15", "grey42", "grey")) +
  theme_minimal()
```

TASK 2 LEXICOGRAPHIC GOAL PROGRAMMING

```
install.packages('goalp')
library('goalp')

# Assigning values ####
goals <- " Man hours: 10*A + 7*B + 5*C = 250 | #3
Cost Of Operations: 8700*A + 8100*B + 7700*C <= 250000 | #2
Number Of Orders: 9*A + 6*B + 4*C >= 0 | #1
A lBound 5
B lBound 5
C lBound 5"
goals
gp <- goalp(goals)
summary(gp)
```

PART 2

TASK 1 CORRELATION, REGRESSION MODEL

```
# Installing packages ####
install.packages("corrgram")
install.packages("fastDummies")
install.packages("VIM",dependencies = T)


# Loading the packages ####
library(dplyr)
library(tidyr)
library(ggplot2)
library(corrgram)
library("fastDummies")
library("VIM")

# Loading the Data into R ####
data <-read.csv("Transactions_Customer.csv",stringsAsFactors = TRUE)

# Data preperation ####
# Checking of any missing values
sum(complete.cases(data))
sum(!complete.cases(data))
dataplot <-data


# Data understanding ####
# Creating visual preparsentations of the different factors with revenue
# Age
ggplot(data = dataplot) +
  geom_point(aes(x = Estimated_Age, y = Revenue, color =  "Data points")) + scale_color_manual(values = "dodgerblue") +
  labs(
    title = "Estimated Age of the customer and Revenue",
    caption = "This data is provided by Drinks@home.uk", x = "Estimated_Age", y = "Revenue"
  )
# Time on site
ggplot(data = dataplot) +
  geom_point(aes(x = Time_On_Site, y = Revenue, color = "Data Points")) + scale_color_manual(values = "dodgerblue") +
  labs(
    title = "Time spend on site and Revenue",
    caption = "This data is provided by Drinks@home.uk", x = "Time spend on the site", y = "Revenue"
  )
```

```
# Seen Voucher
dataplot$Seen_Voucher <-factor(dataplot$Seen_Voucher)
ggplot(data = dataplot) +
  geom_boxplot (aes(x = Seen_Voucher, y = Revenue, color = "Data Points")) + scale_color_manual(values = "dodgerblue") +
  labs(
    title = "Seen voucher and Revenue",
    caption = "This data is provided by Drinks@home.uk", x = "Seen voucher", y = "Revenue"
  )
#Estimated_Income
ggplot(data = dataplot) +
  geom_point(aes(x = Estimated_Income, y = Revenue,color = "Data Points")) + scale_color_manual(values = "dodgerblue") +
  labs(
    title = "Estimated Income and Revenue",
    caption = "This data is provided by Drinks@home.uk", x = "Income", y = "Revenue"
  )
 #Advertisement_Channel

ggplot(data = dataplot) +
  geom_boxplot(aes(x = Advertisement_Channel, y = Revenue,color = "Data Points")) + scale_color_manual(values = "dodgerblue") +
  labs(
    title = "Advertisement_Channel and Revenue",
    caption = "This data is provided by Drinks@home.uk", x = "Advertisement_Channel", y = "Revenue"
  )

#Co relations
cor(data)
corrgram(data)

# Creating dummy varible ###
prepdata <- dummy_cols(data, select_columns = 'Advertisement_Channel',remove_first_dummy = TRUE)
prepdata$Advertisement_Channel<-NULL
prepdata
```

```
#Renaming the new column headings into different channels
colnames(prepdata)[colnames(prepdata)== 'Advertisement_Channel_2'] <- 'SocialMedia'
colnames(prepdata)[colnames(prepdata)== 'Advertisement_Channel_3'] <- 'SearchEngine'
colnames(prepdata)[colnames(prepdata)== 'Advertisement_Channel_4'] <- 'Influencer'
prepdata

cor(prepdata)
corrgram(prepdata)

# Data Modeling ####
model <-lm(Revenue~.,data = prepdata)
summary(model)
```