Table of Contents

Part 1

Introduction

Autonomous Shipment is a Leeds based start-up looking to utilise robot drones to automate the last phase of the logistics to efficiently deliver various products directly to customers. The underlying belief is that it will improve the delivery speed for customers and the company will benefit from the optimisation. This innovation is supported by several venture capitalists and funded by UK government.

The purpose of this report is to advise the management team on two key decision areas:

- **KD1** - Select a robot prototype to participate in the trial phase based on a set of requirements.
- **KD2** - Decide on the quantity of robots to allocate across various stores while ensuring that the goals and constraints of the trial are satisfied.

The primary focus of the trial phase is not profit but to deepen the understanding for future implementation of autonomous robots. The report will explore the use of Multi-Criteria Decision Analysis (MCDA) and Linear Optimisation (LP) methodologies to support the findings.

Data Overview and Understanding

We have four potential robot prototypes in the options pool. They are as follows:

- Robot A032 – Archer
- Robot B23 – Bowler
- Robot CJKL – Corner
- Robot DSXX – Deviant

And the list of all criterions for evaluation associated with each robot are explained below:

Criterion	Definition	Unit	Preference
Carrying Capacity	The amount of weight a robot can carry safely	Litres	Higher is better
Battery Size	The amount of time a robot can operate for one delivery without shutting down	Hours	Higher is better
Average Speed	The movement speed of a robot	Km/h	Higher is better

Acquisition Cost	The cost price of a robot	Great Britain Pounds	Lower is better
Reliability	The time for which a robot can operate without breaking down	Hours	Higher is better

It is important to note that none of these criteria are qualitative, all are quantitative in nature. Clear scores have been assigned to each robot based on these criteria. However, the importance of these criteria is not quantified, instead, the criteria are ranked qualitatively. As for the allocation decision, our targets are 3 types of stores which have their own individual requirements in terms of labour hours, operative cost and per day order completion.

Methodology for KD1

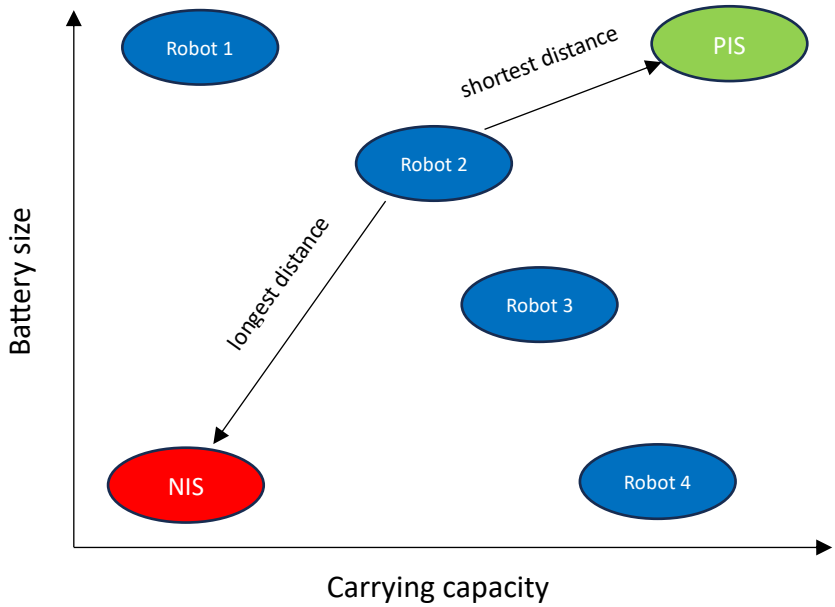
MCDA is a decision-making concept that considers multiple criteria or factors when evaluating different alternatives or solutions. MCDA techniques assess the problem in a structured approach and aims to address the conflicts among the criteria and derive a compromise. Assignment of correct weights to criteria is an important step in any MCDA problem. In our case we first need to compute the formal weightage of each criterion for which we will apply the Pairwise Comparison Method. This method determines the rank of alternatives or criterions by comparing them subjectively in a 'head-to-head' manner e.g. Ferrari is better than Mustang in terms of Top speed. For this comparison, it uses a nine-point scale called Saaty Scale. This scale assesses the relative importance of each criterion/alternative pair. It varies from 1-9, where '1' indicates equal importance between two criteria/alternatives and '9' indicates the absolute importance of one criterion/alternative over another.

Numerical value	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very Strong importance
9	Extremely importance

A matrix is constructed using these pairwise comparisons, called reciprocal matrix. To ensure that our judgement of weights is logical, we calculate a statistic called the 'Consistency Ratio' (CR), which gives a measurement of robustness and reliability of the reciprocal matrix. The industry standard threshold for CR is <= 0.1. Finally, we convert the matrix into normalised scores using a mathematical technique called geometric mean.

Next, we use TOPSIS, which stands for Technique for Order of Preference by Similarity to Ideal Solution to calculate the final scores for each robot prototype. TOPSIS is an ideal point MCDA method that finds out the optimal solution which has shortest distance from the positive ideal solution (PIS) and longest distance from negative ideal solution (NIS) (Behzadian et al., 2012). PIS/NIS are imaginary alternatives having best/worst possible scores

in all criteria. In the diagram below, TOPSIS will give us robot 2 as the preferred solution over other alternatives.



TOPSIS is effective for resolving small to medium-sized decision problems like the one in hand, when the number of alternatives and criteria is limited. In our case, some robot prototypes excel in certain criteria but fall short in other criteria e.g. Deviant has the best reliability, yet it is the costliest. So, we must decide on a prototype which is necessarily not the best but the most optimal in every aspect so that a poor performance in one criterion can be compensated by a good performance in another. Among all the criteria present, reliability, battery size, average speed, carrying capacity are maximising favourable criteria while per unit cost is minimising favourable criterion. Hence, TOPSIS is suitable for the problem.

Methodology for KD2

Our main objective is to maximise the number of orders completed per day, which is subject to various budgetary, labour, allocation constraints such as:

- The total cost of the trial must not exceed the allocated budget of 250000 GBP.
- The total no of technician staff hours available per week are 250.
- Each store should have at least 5 robots.

Hence Linear programming is the most suitable approach to this problem. It is a mathematical technique to find the best possible solution (either maximum or minimum) for an objective function which is subject to a set of constraints (Dantzig, 2002). In problem context, the objective function is number of orders completed per day.

Results

Utilizing the information provided on criteria importance, we derive the order of importance to be the following:

Reliability > Cost per unit > Battery size > Average speed > Carrying capacity

The reciprocal matrix obtained after applying pairwise comparison:

```
##                    Reliability Cost Per Unit Battery Size Average Speed
## Reliability          1.0000000     3.0000000    5.0000000     7.0000000
## Cost Per Unit        0.3333333     1.0000000    3.0000000     5.0000000
## Battery Size         0.2000000     0.3333333    1.0000000     3.0000000
## Average Speed        0.1428571     0.2000000    0.3333333     1.0000000
## Carrying Capacity    0.1111111     0.1428571    0.2000000     0.3333333
##                    Carrying Capacity
## Reliability                        9
## Cost Per Unit                      7
## Battery Size                       5
## Average Speed                      3
## Carrying Capacity                  1
```

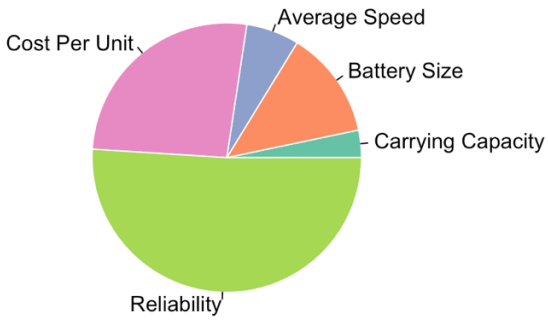
We can understand from the above matrix that:

- Reliability is 3 times more important than Cost per unit.
- Battery size is 5 times more important than Carrying capacity.

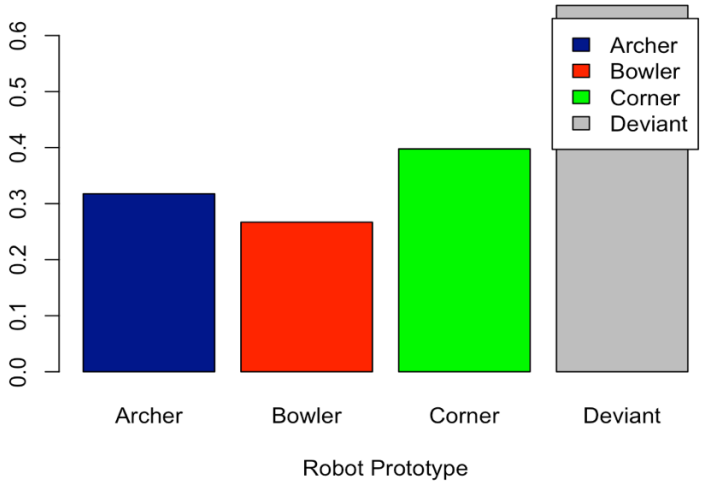
and so on.

In the next step, we compute the CR which came out to be approximately 0.053, indicating a favourable outcome.

The resultant weightage of criteria after applying normalisation is as follows:



The final scores after applying TOPSIS are:



It can be observed that Deviant outperforms other robots by a considerable margin. Reliability is the most important concern, and it has the best reliability score. It is the costliest, but this is negated by its supremacy in speed and battery size. Hence it is the most favourable outcome in overall aspect.

For KD2, assuming:
RG - Robots allocated to Grocery Store
RC - Robots allocated to Clothing Store
RS - Robots allocated to Sports Equipment Store

Considering Deviant is chosen for trial phase and from other information provided on budget, labour and time, the following equations can be formulated:
Orders delivered per day: 9*RG + 6*RC + 4*RS >= 95
Operational Cost for 1 month (GBP) including robot's cost: 8700*RG + 8100*RC + 7700*RS <= 250000
Staff Hours required per week (hours): 10*RG + 7*RC + 5*RS <= 250

Where RG, RC, RS >= 5
The result of the business problem using LP is:

Goals	**RS**	**RC**	**RS**	**Actual**	**Target**	
Operative cost	8700	8100	7700	244300	250000	
Staff Hours	10	7	5	250	250	
Orders completed per day	9	6	4			
Minimum number of robots	5	5	5			
Recommended robots allocation	**19**	**5**	**5**			
				Maximum number of orders per day:		221

Summary
We can summarise the following from the above research and findings:

- Based on the superior performance across evaluated criteria, Deviant should be the ideal choice of deployment for the trial phase.
- The recommended number of robots that can be allocated without exceeding budgetary, technician hours and allocation constraints are - 19 to Grocery store, 5 to Clothing Store, 5 to Sports Equipment Store.
- The analysis indicates that, with the recommended robots allocation, the maximum number of orders that can be completed per day for all stores are 221, and the corresponding budget consumption in this scenario will be 244300 GBP, with full utilisation of technician hours.

Discussion

It is important to note that the MCDA method used in the report only provides a base framework for robot selection under current evaluation criteria and does not provide the absolute result. It is sensitive to the subjectivity in criteria weights and the rank of robot prototypes may be affected if a new option is added or subtracted in the future. Hence to make the decision making more robust, reliable, and transparent, we can conduct sensitivity analysis to assess the impact of changing weights on the final rankings.

Part 2

Introduction

"Drinks@home.uk" is an e-commerce platform specializing in business-to-consumer (B2C) operations, operating exclusively within the Great Britain region. The website offers a variety of both alcoholic and non-alcoholic beverages sourced globally.

The objective is to provide inputs on 2 key decision areas and aid the next marketing campaign:

- **Task 1 -** Based on the customer's demographic and behaviour data, what are the key factors that influences one's spending on the website.
- **Task 2-** What would be best approach among the 3 given choices to include in the next marketing campaign that would increase the website profit.

Data Overview and Understanding

We have the data of 400 customers in the following dimensions:

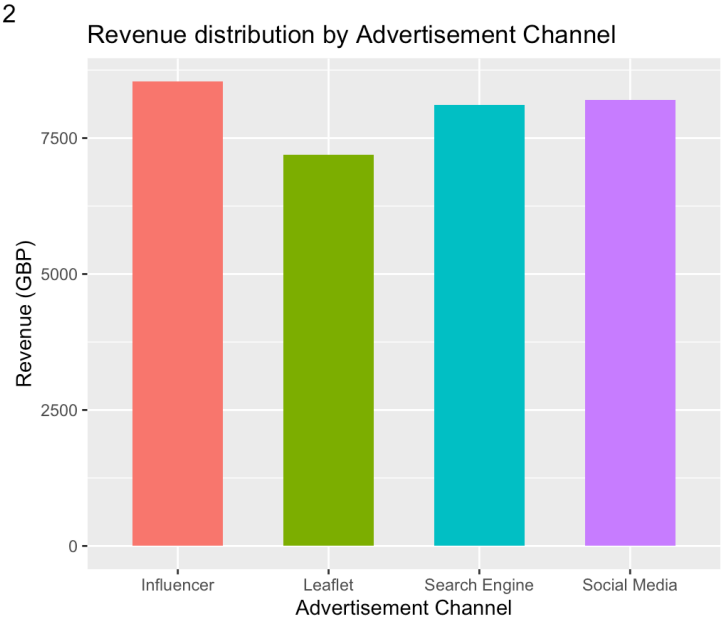
- Revenue (GBP) – The revenue generated from the customer's latest purchase.
- Advertisement Channel – The medium through which the customer got directed to the website. It includes options such as - leaflet, social media, search engine, influencer.
- Estimated Age – The customer's age as estimated by the website's tracking software.
- Estimated Income (GBP) – The customer's income as estimated by the website's tracking software.
- Time on website per week (seconds) – The average time spent by the customer on the website per week.
- Seen Voucher – A bit which indicates whether the customer has seen any voucher in the past.

From the bar chart below, it can be observed that customers who have come across a discount voucher in the past have generated more revenue through their latest orders compared to those who haven't.

Figure 1

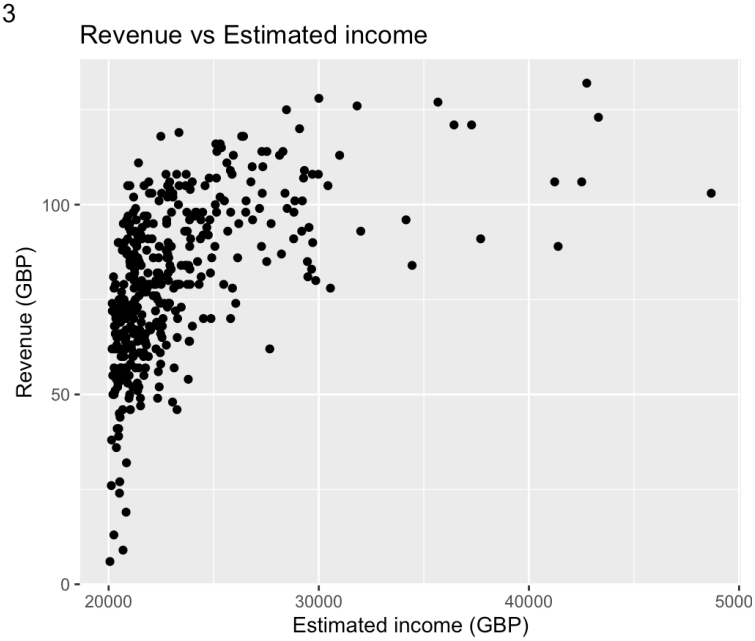
Looking at the plot below, customers who were introduced to the website through influencer advertisement have higher combined revenue compared to those who were directed by other mediums.

Figure 2



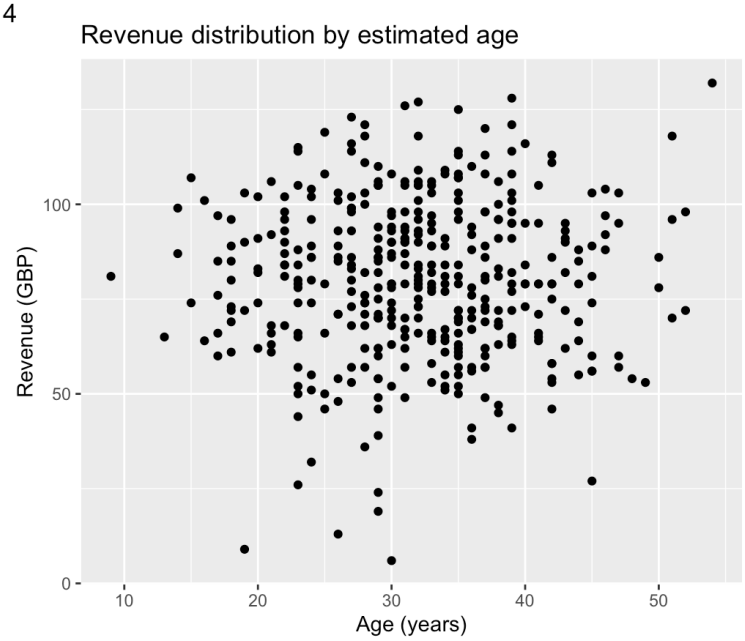
Revenue increases with increase in the estimated income of the customer in an exponential manner. From the graph it is visible that most of the revenue has been generated by the customers having 20-25k GBP estimated income, and the graph thins out after 30k mark.

Figure 3



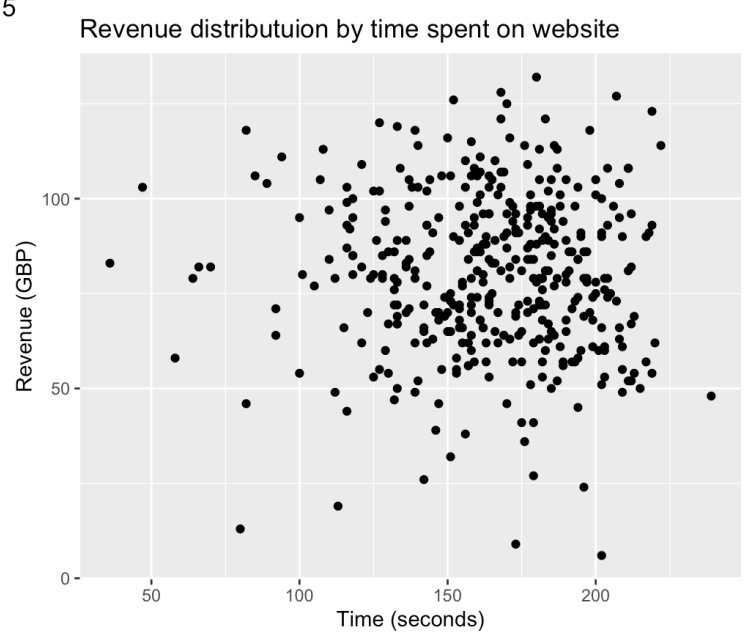
There is no significant increase in the revenue with the increase in customer's age as seen in the visualisation below. However, it is noteworthy that among all customers, those within the age range of 15-45 years have contributed to the most part of total revenue.

Figure 4



Minimum time spent by customers on website is 39 seconds and the maximum is 239 seconds, with the average being 164 seconds. Overall, there is some sort of increase in revenue with increase in time spent on website.

Figure 5



Data Transformation

As discussed, the problem in hand is a regression problem and our unit of analysis is an individual customer because each row in the dataset corresponds to a specific customer and

the columns provide information about their characteristics and interactions with the website.
Firstly, we will convert all categorical data into numerical data. Due to the non-numerical nature of categorical variables, we cannot include them directly in our regression model. To achieve this, we will create dummy variables. A dummy variable is a variable which can take either 0 or 1. They are used to represent categorical variables. For variable having more than 2 categories such as Season, Marital status, Caste etc, we define one of the categories as the base level and create N-1 dummy variables, where N is the number of categories that variable represents.
In our dataset, there 2 are categorical data. The **Seen Voucher** is already dummy coded, but we will create 3 dummy variables for **Advertisement Channel**, taking Leaflet as the base level.

Advertisement channel	Social_Media_Advert	Search_Engine_Advert	Influencer_Advert
Leaflet	0	0	0
Social media	1	0	0
Search engine	0	1	0
Influencer	0	0	1

Methodology for Task 1

We must explore the relationship between a customer's latest order revenue and his income, age, time spent on website, whether he/she has seen a voucher, and advertisement channel through which he/she came to know about the website and determine which of them have a positive/negative affect on revenue. So, the ideal choice would be to utilise multiple regression for analysis. It is statistical technique used to analyse the effect of multiple independent variables on one dependent variable (Kaya et al.,2013).
In problem context, **revenue** is the dependent variable while the rest are independent variables.
All the independent variables will be included in our analysis as we must understand the relationship of **revenue** with each one of them.

Evaluation

The model can be seen below:

```
Call:
lm(formula = Revenue ~ ., data = cust_info)


Coefficients:
                          Estimate     Pr(>|t|)
(Intercept)               0.3877205    0.951465
Estimated_Age            -0.0152422    0.864718
Time_On_Site             -0.0221743    0.312467
Seen_Voucher             19.6954714    < 2e-16  ***
Estimated_Income          0.0028609    < 2e-16  ***
Social_Media_Advert       6.8284251    0.000783 ***
Search_Engine_Advert      8.0909325    6.28e-05 ***
Influencer_Advert         12.9736091   2.66e-10 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.5547,  Adjusted R-squared:  0.5467
```

- **Estimate** is the expected change in dependent variable when the independent variable increases by 1 unit.
- **Pr(>|t|)** is a measure of the significance of the corresponding independent variable. In other words, how important it is, for explaining the variation within revenue. The lower the value, higher the significance. In research, p-value below 0.05 is usually considered to be significant.
- **Asterisk (*)** denotes the level of significance.

The intercept estimate shows that the expected value of revenue generated by a customer with an age of zero, no income, zero time spent on website, and no exposure to voucher and advertised through a leaflet is approximately 0.39 GBP. This is a hypothetical scenario and holds no practical meaning. The relationship of the independent variables can be summarised as:

- Customer's estimated age and time spent by them on website per week does not have a significant relationship with the revenue. Revenue decreases by **0.015 GBP** for 1 year increase in age, while it decreases by **0.022 GBP** for each second increase in time spent.
- Voucher encounter is significantly related to revenue and the customers who have come across a voucher, have spent **19.69 GBP** more on average, in their latest order.
- Income of customer has a significant relationship with revenue. Customers with higher income tend to spend **0.0028 GBP** more on average, in their latest order purchase.
- Social media advertisement significantly impacts the revenue. Customers captured through social media have spent **6.82 GBP** more on average, in their latest order.
- Search engine advertisement significantly impacts the revenue. Customers captured through search engine have spent **8.09 GBP** more on average, in their latest order.
- Influencer advertisement significantly impacts the revenue. Customers captured through influencer have spent **12.97 GBP** more on average, in their latest order.

The R-squared and adj. R-squared value indicates that approx. 55% of variation within revenue is explained by the independent variables. The model has a moderate level of explanatory power, and it provides evidence that there may be additional factors influencing the revenue which account for the unexplained variance.
The relationship of each of the factors with revenue is as follows:

Factor	Impact on Revenue and importance
Advertisement Channel	Positive and significant
Estimated Income	Positive and significant
Estimated Age	Negative and insignificant
Seen Voucher	Positive and significant
Time spent on website per week	Negative and insignificant

Recommendation for Marketing Campaign

The regression analysis conducted provides a robust framework for the marketing team to filter out the factors which influence customer spending and assess the strength of their impact. This will help in tailoring the marketing strategy to cater the needs of customers which will help in increasing the 'Drink@home.uk' revenue.

Below is an analysis of all proposed options for increasing the website profits:

- **Target customers older than 45 years**
  As per our analysis, estimated age of customer has no significant relationship with revenue. The impact of estimated age of the customer on the revenue of his/her latest order is purely by chance, and there is insufficient evidence to support the causality. Moreover, an increase of 1 year in customer's age will result in 0.015 GBP decrease in revenue. Hence, this option is **not recommended**.
- **Provide a discount voucher of 20 GBP**
  Customer seeing the voucher significantly impacts the revenue generation with 1 voucher view increasing revenue by 19.695 GBP which is the highest among all factors. However, assuming discounts are a part of operational cost, and cost incurred for providing a 20 GBP voucher is greater than the revenue generated by 0.304 GBP, which leaves no margin for profit. Moreover, this difference will increase with each voucher provided, thus negatively impacting profitability. Hence this option is also **not recommended**.
- **Invest in influencer advertisement**
  Each customer brought in through influencer advertisement increases revenue by 12.973 GBP. This means that allocating funds to influencer advertisement has the potential to substantially boost revenue. We don't have the cost details of the process, but we can assume that operational cost for the entire marketing campaign is such that the revenue generated is more than the cost incurred. The marketing team would have to run trials to determine the break-even point to avoid any loss. Hence, I **recommend** the higher management to proceed with this option for marketing campaign.

References:

Behzadian, M., Otaghsara, S.K., Yazdani, M. and Ignatius, J., 2012. A state-of the-art survey of TOPSIS applications. *Expert Systems with applications*, *39*(17), pp.13051-13069.

Dantzig, G.B., 2002. Linear programming. *Operations research*, *50*(1), pp.42-47.

Uyanık, G.K. and Güler, N., 2013. A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, *106*, pp.234-240.