

Assessed Coursework Coversheet

For use with *individual* assessed work

Student ID Number:	2	0	1	8	0	1	2	9	5
Module Code:	LUBS5308M								
Module Title:	Business Analytics and Decision Science								
Module Leader:	Aritad Choicharoon								
Declared Word Count:	2959								

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).

Index

Business Question 1

1. Introduction
2. Data Understanding
 - Methodology Used for Task 1
 - Result
 - Methodology Used for Task 2
3. Conclusion

Business Question 2

1. Introduction
2. Data Understanding
 - Methodology used for task 1
 - Evaluation of model
 - Summary of Model
3. Data Understanding for task 2
 - Task 2 discussion

References

Introduction

Autonomous Shipment was approached by a start-up based outside of Leeds who were interested in using autonomous robot drones for faster delivery of goods to customers. As per the key decisions, we have been provided with four prototype robot drones by the company, tasking us to decide the highly effective robot that would participate in the trial based on the set of criteria. Additionally, we must also determine how many robots need to be allocated across various stores like grocery, clothing and sports to ensure the maximisation of orders per day while adhering to certain constraints. This report illustrates the outcome of a Multi-Criteria Decision Analysis to recommend the optimal robot drone for the trials phase.

Linear Programming is further applied in order to decide how many robots need to be allocated to all the specified stores.

Data Understanding

The company has provided us with four potential models that they've chosen:

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

The below table depicts the association between the scores of alternatives (robots) and the corresponding criteria around which the decision should be grounded:

Criteria	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

The priority of the criteria was not directly quantified, but from the qualitative statements from the management team, we can make out the importance of criteria to be the following:

Reliability > Cost per Unit > Battery Size > Average Speed > Carrying Capacity.

Methodology Used for Task 1

We will utilise the Weighted Sum Method among the various MCDA techniques to recommend the best possible robot prototype model based on a set of criteria. This approach involves assigning weights to various criteria, indicating their relative importance, and calculating a weighted sum for each alternative. In this method, each criterion is assigned a weight in such a way that all the criteria weights must add up to one. Each alternative is now assessed with respect to every criterion (Karandea et al, 2016). The alternative with highest weighted sum is regarded the most favourable outcome for maximizing utility across all attributes/criteria.

WSM is robust due to its simplicity and ease of application. It is one of the most widely applied decision-making methods.

A similar problem has earlier been resolved using WSM method (Budiharjo et al , 2017)

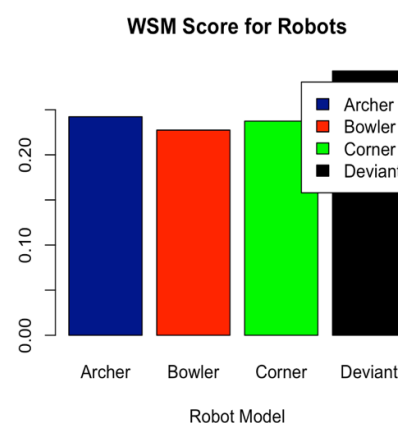
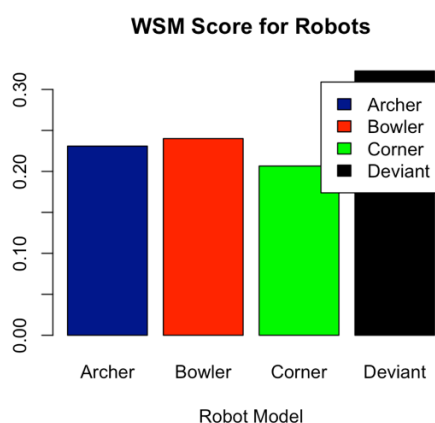
Hence this method provides additional justification for its use.

We will use a scale of 1-5 to quantify the order of criteria mentioned before, with 5 being the highest priority and 1 being the lowest. Next we normalise the weights. It is the process of dividing each weight with the total of all weights for e.g. $1/(1+3+2+4+5) = 0.06666667$ and so on. The reason for doing so is to make out the weights more robust and transparent resulting in an unbiased decision.

Criteria	Weights	Normalised Weights
Carrying Capacity (max)	1	0.06666667
Battery Size (max)	3	0.20000000
Average Speed (max)	2	0.13333333
Cost Per Unit(min)	4	0.26666667
Reliability (max)	5	0.33333333

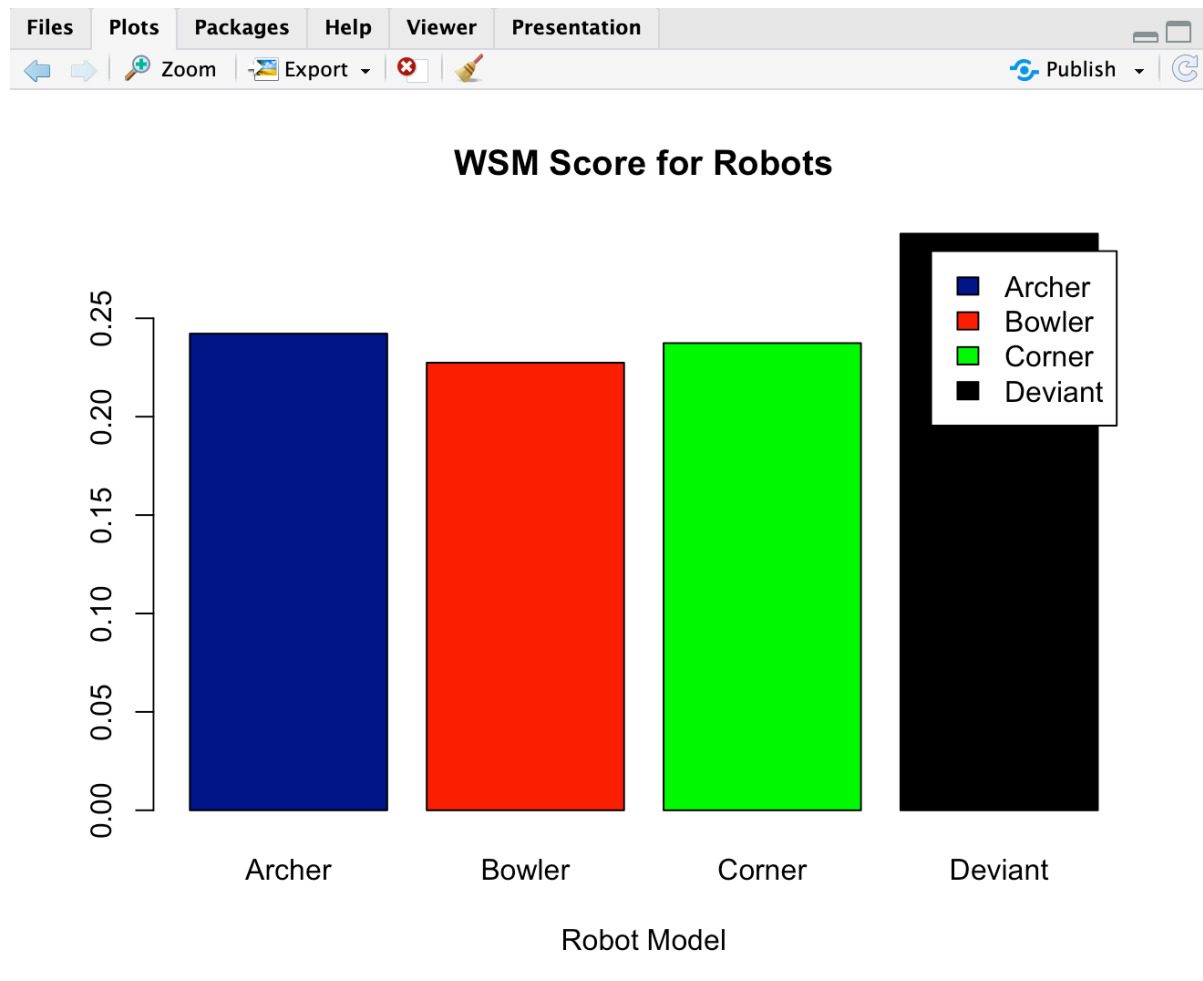
To apply WSM, all the criteria should either be increasing or decreasing utility order. In our case, all the criteria are maximising except for cost, which is minimizing in nature. To stabilise this situation, the scores for cost per unit criteria have been inverted to enable a fair comparison.

The figures below illustrates, how the ranks varies if we don't apply inversion:



We will then normalise the scores to make the scale consistent as the criteria have different measurement units.

The WSM scores have been visualized using a bar plot, providing a clear representation of how each robot model performed in the evaluation.



Results for Task 1

Based on the scores and visualizations, the recommended robot model for the trials phase is Deviant. This recommendation aligns with the specified decision criteria and weights. It is the best possible outcome as it has the best reliability which is our topmost concern and has the best overall performance across three criteria. i.e. reliability, battery size and average speed. The carrying capacity being the lowest is not that significant as it is the least of our concerns.

Hence, Weighted Sum Method of MCDA has provided us with valuable insights into our decision-making process. Through a systematic analysis of the criteria and the utilization of weighted scores, we have successfully identified optimal solution.

Methodology for Task 2

The second decision to make is to find out the number of robots to be allocated across all 3 stores i.e. grocery, clothing ,sports for maximising the number of orders per day.

We concluded that Deviant will be deployed for the trials, so the total cost for each store will be the sum of the cost of the Deviant and operating cost of the corresponding store. The resource utilisation details for all stores is provided below:

Goals	Grocery	clothing	sports
Cost	8700	8100	7700
Staff Hours	10	7	5
No Of Orders	9	6	4

The constraints and goal of the trial are as follows:

- All the three stores should have at least 5 robots during the trials.
- The total availability of the technician staff hours to support this trial is 250 hours per week.
- The operating cost and purchase cost must not exceed 250k GBP.
- The goal is to maximize the number of orders completed by robots on a daily basis.

To solve this task, we utilise Linear Programming technique. Linear programming, also known as linear optimization, is a method for seeking the best possible outcome (either maximised or minimised) in a situation where the requirements are subject to constraints. Here our main objective is to maximise the quantity of orders completed per day while adhering to certain constraints. Hence, LP is the ideal choice of optimisation technique to proceed with. This method has applications in various industries and fields, from production planning to logistics optimization, and is a powerful tool in making data-driven decisions that can lead to improved efficiency and cost savings ([Avcontentteam, 2023](#)).

Let the no of robots required in Grocery store be 'A', in Clothing store be 'B' and in Sports store be 'C'. Thus, the equations formed are:

No Of orders completed per day: $9*A + 6*B + 4*C \geq 95$

Operating Cost per store (including robot cost): $8700*A + 8100*B + 7700*C \leq 250000$

No Of Staff required per week: $10*A + 7*B + 5*C \leq 250$

The output of LP is as follows:

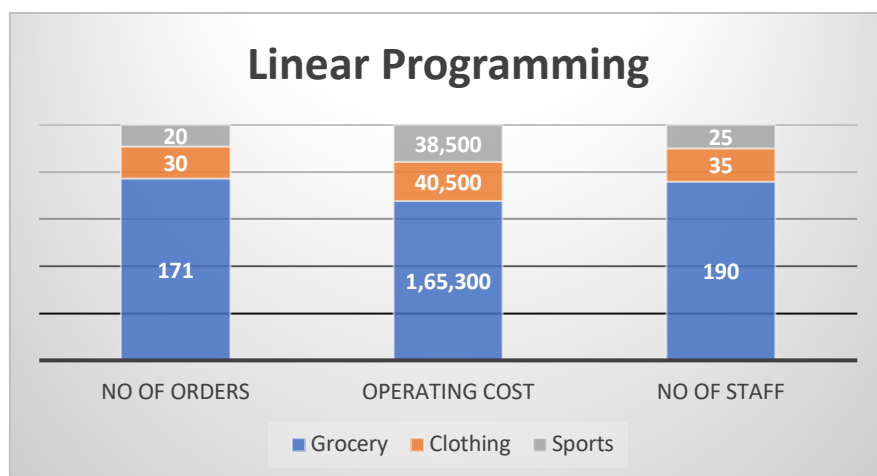
Goals	Grocery	Clothing	Sports	Actual	Target
NoOfOrders	9	6	4		
Operating Cost	8700	8100	7700	244300	250000
NoOfStaff	10	7	5	250	250
Minimum no of robots that have to be present in each store	5	5	5		
Maximum number of robots to be allocated	19	5	5		
				Maximum number of robots to be allocated	221

Stores	No Of Orders (per day)	Operating Cost (per month)	No Of staff hours (per week)
Grocery	171	165,300	190
Clothing	30	40,500	35
Sports	20	38,500	25
Actual Outcome	221	244,300	250

We conclude from the above table that maximum number of orders that can be completed in a day for all the three stores combined are 221, with a total allocation of 29 robots.

The operating cost consumption is below the budget of 250,000GBP by 5,700GBP, with full utilisation of technician hours.

A graphical representation of the linear programming result is shown below:



Conclusion

Among all the robots at our disposal, deviant is the best choice for deployment in the trial phase. As for maximum of 221 orders could be completed under the given time and budget constraints. This scenario requires the robots to be allocated in the following manner:

Grocery store – 19

Clothing shop – 5

Sports shop – 5

The total fund consumption will 244,300GBP and all technician hours will be utilised.

Part 2: Value of customers:

Introduction

'Drinks@home.uk' is an e-commerce website that operates in Great Britain offering both alcoholic and non-alcoholic drink selections all around the world. In this business problem, we have been provided with 400 customers information, that includes the revenue made from the latest orders, the advertisement medium that caused customers to visit the website, their personal information like age and income, along with time spent on the website on an average weekly basis. It also includes if the customers came across a voucher popup.

The following 2 tasks are assigned to us:

Task 1 : Determine the factors that significantly affect the customer spendings either more or less on the 'drinks@home.uk' website, by analysing the behaviour and demographics of past customers.

Task 2 : Based on the three choices that would be provided for the next marketing project, we have to recommend on which choice would be the most appropriate for increasing profits.

To solve these business questions, we'll be performing regression to determine the best solutions.

Data Understanding

The data used for analysis include about 400 customers information of the website.

The following are the variables that will be used for each customer :

Revenue(GBP)	The revenue obtained from the latest order placed by the customer
Advertisement Channel	The medium via which the customers were attracted to the website. <ul style="list-style-type: none">• Leaflet• Social Media• Search Engine• Influencer
Estimated Age	The age of the customers according to the tracking system of the website software.
Estimated Income(GBP)	The estimated income of the customer according to the tracking system of website software.
Time spent on website per week(seconds)	The average time the customer spends on the website
Seen Vouchers	If the customer has seen any voucher pop-ups on the website.

Methodology Task 1

The first task of the business problem, focuses on determining which factors have a positive/negative influence on the revenue generated on the website.

To analyse this task we use linear regression approach of supervised learning model.

Linear regression is useful tool for predicting a value of a variable based on value of another variable. The value of variable that we predict is called dependent variable, while the variable we use to predict the value of other variable is called independent variable.

The data consists of quantitative(numerical) as well as qualitative(categorical) value . Quantitative data is highly compatible with regression models but categorical data cannot be directly used hence we need to convert them to numerical data. To do this, we perform dummification of categorical variables, which creates dummy variables, and then assigns these unique integer to each category of data. The value of dummy variable is either 0 or 1, representing the availability or non-availability of that attribute.

Here, advertisement channel contains categorical data. Therefore, we convert the categorical variable of advertisement channel into a numerical variable or dummy variable .

Taking advertisement channel =1 as base condition , we create three dummy variables for advertisements through social media, search engine and through an influencer.

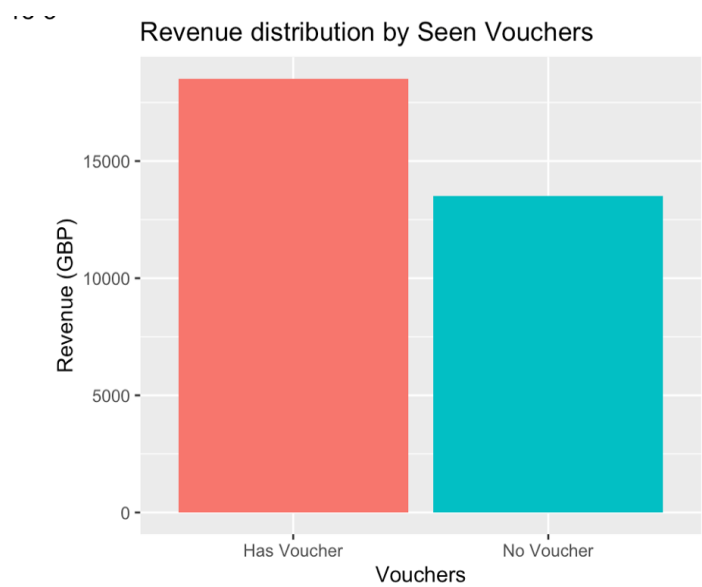
advertisement channel = 2 , create dummy variable for social media having values 1 or 0

advertisement channel = 3 , create dummy variable for search engine having values 1 or 0

advertisement channel = 4 , create dummy variable for influencer having values 1 or 0

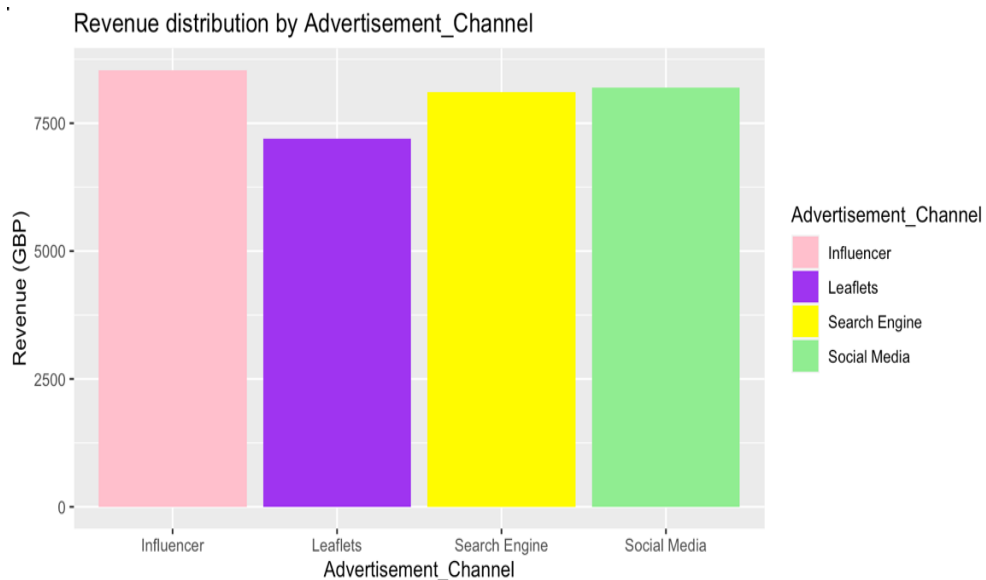
After transforming the data, we analyse the relationship between the factors that affect the revenue generated by the customers from their latest transactions.

1. Seen Voucher



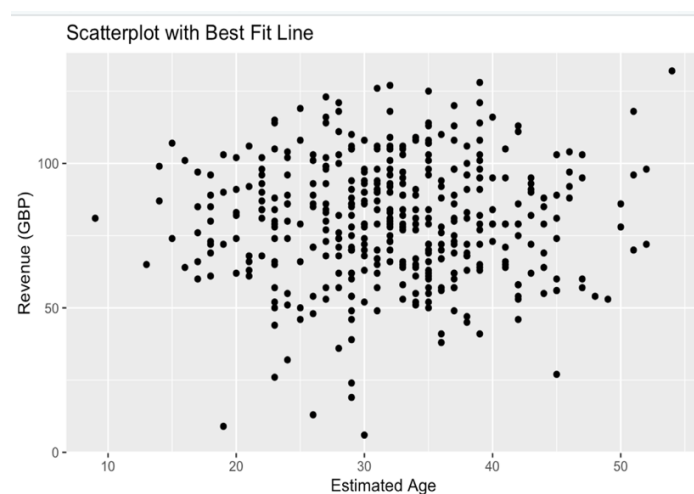
The graph shows the distribution of revenue by voucher interaction. It can be observed that people who have seen the voucher corresponds to the increased revenue.

2. Advertisement channels



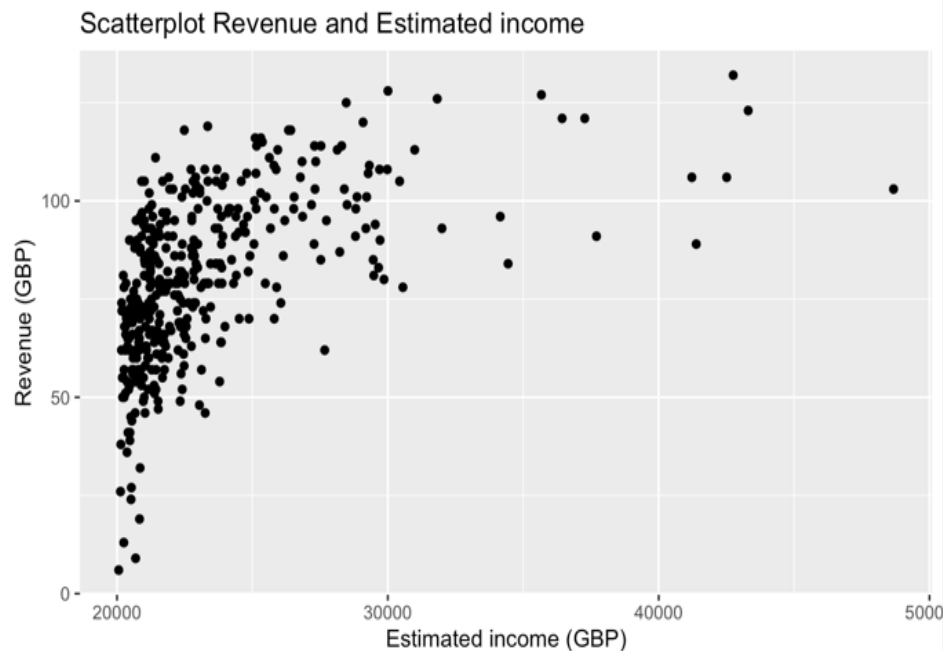
The above graph shows the distribution of revenue by advertisement channel. Influencer advertisement has generated the highest revenue, followed by social media and search engine. Leaflets has the lowest share in revenue generation, hence we had used this as the base criteria for dummification.

3. Estimated Age



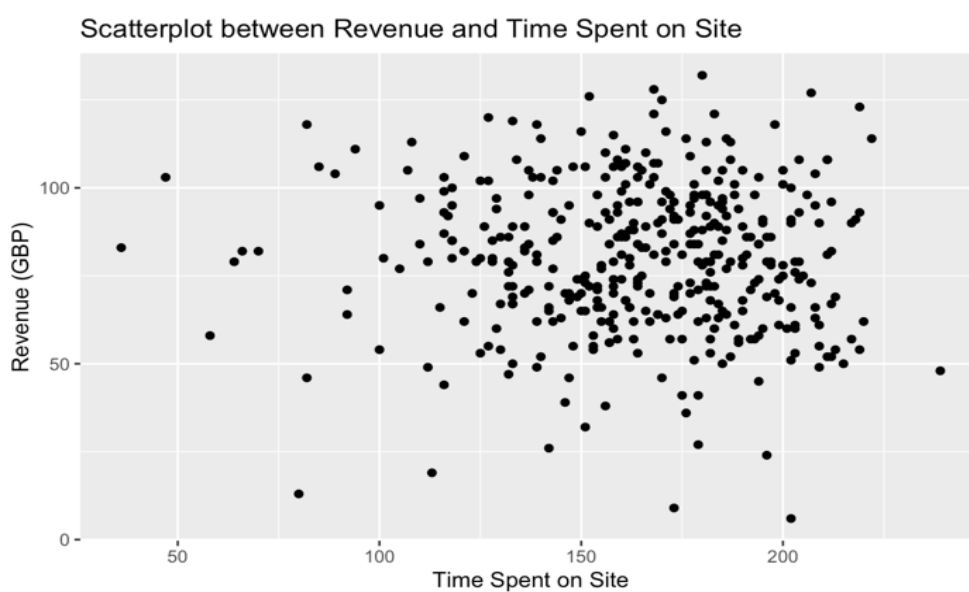
The scatterplot depicts that there isn't much association between the estimated age of the customer and the revenue they generate from their latest order.

4. Estimated Income.



The graph provide a valuable insight that the customers whose income falls between 20,000 - 30,000GBP, contribute to maximum revenue generation.

5. Time Spent



The above graph shows that the time spent on the website by the customer per week doesn't have much connection with the increase in revenue.

Since this is an explanation task and not prediction we will be limited to t-test and R square, which will help in confirming the relationship between independent and dependent variables.

```
summary(overall1)
```

```
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 ***
Channel_SocialMedia	6.8284251	0.000783 ***
Channel_SearchEngine	8.0909325	6.28e-05 ***
Channel_Influencer	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
```

Evaluation of model –

In Model Evaluation we focus on the Multiple R-Squared and Adjusted R-squared values.

Multiple R-Squared are values between 0 and 1, that gives a measure of how well our model fits the given data.

Multiple R-squared: 0.5547

Adjusted R-squared: 0.5467

From the above values, we can observe that our model can describe 55% of variation within revenue generated using the independent variables. This means that there are certain other factors which have influenced a customer's spending on their latest order, which would explain the remaining variation.

Variable Evaluation –

Estimate – represents the change in dependent variable by one unit of increase/decrease in independent variable

Pr(>|t|) – represents the probability of the estimate occurring if the null hypothesis is that there is no relationship between dependent and independent variable.

According to the industry standard, if **pr(>|t|)** ≤ 0.05 , the relationship is deemed as significant.

In our case,

Null Hypothesis : There is no significant relationship between revenue and other factors.

Independent Variable	Impact on revenue	Impact value (estimates)	Probability of this estimate happening by chance
Estimated age	negative	0.015	86% as there is no significance.
Time on Site	negative	0.022	32% with no significance
Seen Voucher	positive	19.69	Is almost negligible if there is no relationship between seen voucher and revenue.
Estimated Income	positive	0.002	Is almost negligible if there is no relationship between income and revenue.
Social Media	positive	6.82	0.078% and has significant relationship
Search Engine	positive	8.09	0.006% with significance
Influencer	positive	12.97	Is almost negligible if there is no relationship between advertisement channel influencer and revenue.

Summary of Model

- Seen Voucher is statistically significant related to revenue with an increase in viewership of a voucher resulting in an increase of 19.6954714 in revenue on average.
- Estimated Income is statistically related to revenue with an increase in 1GBP of income, resulting in an increase of 0.0028609 in revenue on average.
- Advertisement through Social Media is statistically related to revenue with an increase in a customer attracted resulting in an increase of 6.8284251 in revenue on average.
- Advertisement through Search Engine is statistically related to revenue with an increase in a customer attracted resulting in an increase of 8.0909325 in revenue on average.
- Advertisement through influencer is statistically related to revenue with an increase in a customer attracted resulting in an increase of 12.9736091 in revenue on average.
- Estimated age and time on site are not statistically related to revenue, as they have negative impact on revenue of -0.0152422 and -0.0221743.

Data for task 2

This task requires us to present “drinks@home.uk” with a recommendation by choosing between three options, that they are taking into consideration in order to increase profit on the website. The options provided are :

1. Run an advertisement targeting customers who are older than 45 years old as they are likely to spend more money.
2. Provide a voucher for 20GBP off their next orders.
3. Spend more money on advertising with an influencer

Task 2 discussion

Revenue is the total income generated without excluding the other expenses and costs in a business. Profit is the net amount that remains after all operational deductions are made.

We have to choose between the three options provided that could possibly increase the profit.

Analysing options:

- Customers aged over 45 years, would not significantly impact the profit as there is no notable relationship between age and revenue. The revenue vs age graph shows that there are less customers above the age of 45, hence running an advertisement for them would not be a profitable option.
- We know, one view of seen voucher causes a revenue generation of 19.69gbp. Let's consider, the 20gbp voucher as expenses. The profit generated per person would nearly become (-0.304gbp), which is loss.

Therefore, even though seen voucher is the most significant independent variable and has the highest positive value, proving a voucher of 20gbp, would not make any profit and eventually the essence of voucher gets lost. Also, loss will increase with each additional voucher provided. Thus, this option is also not profitable.

- Customers through advertisement channel Influencer tend to increase revenue by 12.97gbp each. If the cost of influencer advertising is lower than the revenue generated, then the website would make profit. We need to make sure that the number of customers through influencer should cross the break-even point, so that maximum profit can be generated. Thus, this option has the most potential to increase the profits of 'drinks@home.uk'. Hence, I recommend to proceed with this option for the campaign.

Reference:

1. Karande, P., Zavadskas, E. and Chakraborty, S., 2016. A study on the ranking performance of some MCDM methods for industrial robot selection problems. *International Journal of Industrial Engineering Computations*, 7(3), pp.399-422.
2. Tunas Bangsa Pematangsiantar, S.T.I.K.O.M., 2017. Comparison of weighted sum model and multi attribute decision making weighted product methods in selecting the best elementary school in Indonesia. *International Journal of Software Engineering and Its Applications*, 11(4), pp.69-90.
3. Avcontentteam 2023. What is linear programming? Definition, methods and problems for data scientists. Analytics Vidhya. [Online]. Available from: <https://www.analyticsvidhya.com/blog/2017/02/introductory-guide-on-linear-programming-explained-in-simple-english>.