Leeds University
Business School

UNIVERSITY OF LEEDS

Assessed Coursework Coversheet

For use with individual assessed work

Student ID Number:	2	0	1	7	1	0	3	4	5
Module Code:	LUBS5308M								
Module Title:	Business Analytics and Decision Science								
Module Leader:	Aritad Choicharoon								
Declared Word Count:	1476 + 1491 = 2967								

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.

Part 1: Autonomous Shipment Roll-out

We can apply the CRISP-DM framework to investigate the business problems mentioned in the "Part 1 Autonomous Shipment roll-out: autonomous delivery trial."

Task 1: Prototype Robot Recommendation

Business Understanding

Goal: In this part of the assignment, I am the analytic manager of "Autonomous Shipment", a new start-up venture based in Leeds. I have been tasked with providing a report to the management team. The report should address two questions:

1. Which prototype robot should be selected for the trial based on a set of requirements?

2. How many robots should be allocated across various stores to ensure that the goal and constraints of the trial are satisfied?

Motivation: The trial aims to determine whether employing autonomous drones for last-leg logistics is feasible, focusing on rapid delivery, cost-effectiveness and reaching a broad consumer base.

Constraints: Restricted budgets and a range of products offered by various stores focus on ensuring the trial addresses many possible participants and remains economical.

Data Understanding

We have developed four prototype robots, namely Robot A032 - Archer, Robot B23 - Bowler, Robot CJKL - Corner, and Robot DSXX – Deviant and the trial is expected to run for a **month**. We aim to cover as many potential customers across different store types as possible while ensuring the trial remains on budget.

The decision on which prototype to use should be based on the following criteria:

Criteria	Preference	Priority
Carrying Capacity	Higher	*
Battery Size	Higher	***
Average Speed	Higher	**
Cost per Unit	Lower	****
Reliability	Higher	*****

Data sources - Management_Priority.xlsx

Note: The team determined that the **cost per unit** should account for at least 25% of the total evaluation across all criteria.

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

Data sources - Robot_Info.csv

Data Preparation

Converting qualitative variables to quantitative variables:

1. Point Allocation: *(Godwin Odu, 2019)*

- Allocate 1.00 points across criteria, giving more to those deemed more important.
- Example: Reliability (0.30), Cost per Unit (0.25), Battery Size (0.20), Average Speed (0.15), Carrying Capacity (0.10).

2. Direct Proportional: *(Godwin Odu, 2019)*

- Assign numerical values to qualitative importance (e.g., 5 for "most important", 1 for "least important").
- Calculate normalized weights by dividing each value by the total.
- For example, 'Reliability' would have a normalized weight of 5/15 = 0.3333.

3. Salo Scale: *(Salo and Hämäläinen, 1997)*

- Rank criteria from most to least important.
- Assign scores of 9, 7, 5, 3, and 1, respectively.
- Normalize weights by dividing each score by the sum of all scores.
- For example, 'Reliability' would have a normalized weight of 9/25 = 0.36.

Robot Prototype	Point Allocation Method	Direct Proportional Method	Salo scale
Carrying Capacity	0.10	1/15	0.04
Battery Size	0.20	3/15	0.20
Average Speed	0.15	2/15	0.12
Cost Per Unit	0.25	4/15	0.28
Reliability	0.30	5/15	0.36

For our decision-making process, we have chosen to use the **Salo scale** method for weight assignment. This decision is based on the even distribution of the Salo scale, which implies an equal dispersion between each scale value.

Modelling

I am using the VIKOR method as a Multi-Criteria Decision Analysis (MCDA) technique. It stands out for its ability to select the most suitable option among various alternatives when dealing with conflicting and varied unit criteria to arrange the ranking of these options. The core aspect of VIKOR is the focus on how closely alternatives approach the ideal solution, comparing them based on this measure of proximity. Additionally, VIKOR distinguishes itself from the TOPSIS method by employing unique aggregating functions and normalisation approaches. *(KOKOÇ and ERSÖZ, 2019)*

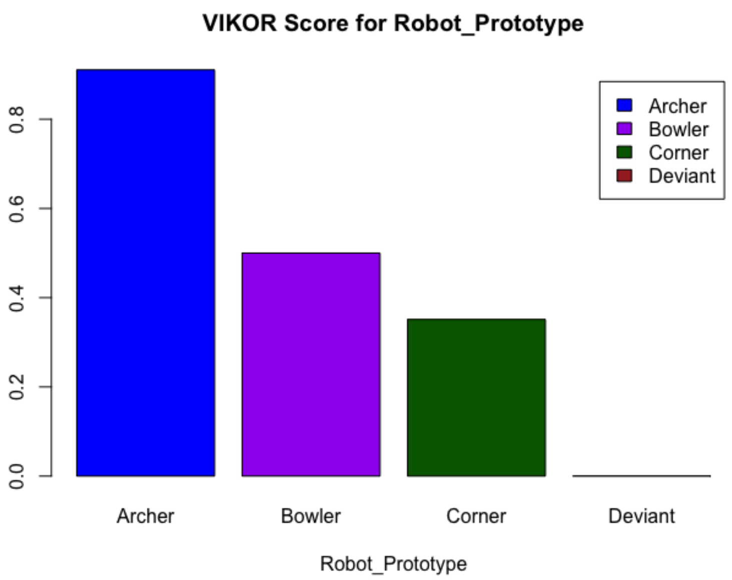
Why VIKOR is ideal for Autonomous Shipment: *(Caylor and Hammell II, 2021)*

- **Normalisation:** Uses linear normalisation which does not depend upon the unit of criteria.
- **Multiple Criteria:** Effectively handles the evaluation of robots based on carrying capacity, battery size, speed, cost, and reliability.
- **Balancing Trade-offs:** Optimizes the selection of a robot prototype by balancing performance and cost factors.
- **Quantitative Decision Making:** Provides a systematic approach using numerical data for each criterion.
- **Handling Conflicting Priorities:** Reconciles differing priorities (e.g., high capacity vs. low cost) to find the best collective solution.
- **Transparency and Rationality:** Offers clear decision-making based on a transparent algorithm.
- **Sensitivity Analysis:** This allows for exploring how changes in criteria weights and closeness to the ideal solution impact the outcome.

Now after implementing VIKOR, I got this final $\boldsymbol{Q_j}$ values lowest score and is considered the best option.

Robot Prototype	Archer	Bowler	Corner	Deviant
$\boldsymbol{Q_j}$ (v = 0.5)	0.911	0.550	0.413	0.000
				Best Choice

VIKOR's $\boldsymbol{Q_j}$ values for salo scale weight



Summary: Based on the VIKOR analysis, the **Deviant** prototype would be the recommended robot for the Autonomous Shipment trial, as it best meets the criteria established by the management team when considering both the utility and regret measures (**v = 0.5**).

Evaluation

Sensitivity analysis includes varying these weights and thresholds to understand how robust the method is to changes in these parameters.

Robot Prototype	Point Allocation Method	Direct Proportional Method	Salo scale
Archer	0.909	0.910	0.911
Bowler	0.500	0.500	0.550
Corner	0.334	0.351	0.413
Deviant	0.083	0.000	0.000

VIKOR's $\boldsymbol{Q_j}$ values for different weights with same v = 0.5

Robot Prototype	Archer	Bowler	Corner	Deviant
v = 0.1	0.982	0.190	0.162	0.000
v = 0.2	0.964	0.280	0.225	0.000
v = 0.3	0.947	0.370	0.287	0.000
v = 0.4	0.929	0.460	0.350	0.000
v = 0.6	0.894	0.640	0.475	0.000
v = 0.7	0.876	0.730	0.538	0.000
v = 0.8	0.858	0.820	0.601	0.000
v = 0.9	0.841	0.910	0.663	0.000
				Best Choice

VIKOR's $\boldsymbol{Q_j}$ values for different v at the same Salo scale weight

Sensitivity analysis suggests that the choice of the Deviant robot is a stable and robust decision across a variety of decision-making weights and parameters (v).

Conclusion

Prototype Selection: After a thorough evaluation using the VIKOR multi-criteria decision-making method, I recommend the Robot **DSXX – Deviant** for our autonomous delivery trial in Leeds. This prototype is costly but has the best battery size, average speed, and reliability. The Deviant consistently scored the highest in the VIKOR analysis and proved to be the most robust choice across various weighted scenarios.

Task 2: Robot Allocation for Trial

Data Understanding

To decide how many robots to allocate across various stores, we should consider the following constraints as per different stores:

Store	Estimated Orders per Robot per Day	Operating Cost per Robot per Month	Technician Man Hours per Robot per Week
Grocery Store	9	1600 GBP	10
Clothing Store	6	1000 GBP	7
Sport Equipment Store	4	600 GBP	5

Also, consider the criteria and constraints for the robot trial:

Criteria	Constraint
Budget	<= 250,000 GBP
Number of robots per store	>= 5
Technician staff hours	<= 250 hours
Orders delivered per day	Maximize

Note: *The trial adheres to a strict budgetary policy of 250,000 GBP*

Data Preparation

Mathematical Formulation:

Decision Variables:

- x_1: Number of robots in the Grocery Store
- x_2: Number of robots in the Clothing Store
- x_3: Number of robots in the Sports Equipment Store
- d_1^+ , d_1^-: Deviation in budget.
- d_2^+ , d_2^-: Deviation in technician hours.

Objective Function: Maximize the total number of orders per day:

$$\text{Maximize } Z = 9x_1 + 6x_2 + 4x_3$$

Minimize: The sum of all deviation

$$d_1^+ + d_1^- + d_2^+ + d_2^-$$

Constraints:

1 Minimize deviation from the budget (including the cost of the selected robot from Task 1 and operating costs):

$$1600x_1 + 1000x_2 + 600x_3 + 7100(x_1 + x_2 + x_3) + d_1^+ - d_1^- \leq 250000$$

2 Make sure at least 5 robots in each store:

$$x_1 \geq 5, x_2 \geq 5, x_3 \geq 5$$

3 Minimize deviation from the technician staff hours:

$$10x_1 + 7x_2 + 5x_3 + d_2^+ - d_2^- = 250$$

4 Non-Negativity: $d_1^+, d_1^-, d_2^+, d_2^- \geq 0$

Modelling

Robot Allocation: By applying a **goal programming** model in the above equation, the robots were allocated as follows to maximize order fulfilment:

Store	Number of Robots	Estimated Orders per Robot per Day	Operating Cost per Robot per Month	Technician Man Hours per Robot per Week
Grocery Store	19	171	30,400 GBP	190
Clothing Store	5	30	5,000 GBP	35
Sport Equipment Store	5	20	3,000 GBP	25
Total	**29**	**221**	**38,400 GBP**	**250**

Total budget = 29 robot cost + total operating cost of robot = 244,300 GBP.

Conclusion

Robot Allocation: The total cost of acquiring 29 robots and their operating expenses, is within 97.72% of the allocated trial budget of 250,000 GBP. The allocation satisfies all the constraints outlined for the trial:

- Each store has at least 5 robots.
- The number of orders completed per day is maximized, with a total of 221 estimated orders per robot per day across all stores.
- The technician man-hours do not exceed the per-week limit set for the trial.
- The total cost is within the trial budget.

In summary, the trial's strategic approach, utilizing the Deviant robot, is set to significantly improve delivery services in Leeds. This initiative not only meets immediate operational goals but also positions Autonomous Shipment for future expansion and innovation in autonomous logistics.

Part 2: Value of customers

Task 1: Analysing Customer Spend Influencers

Business Understanding

In this part of the assignment, as an analyst at 'Drinks@home.uk', I have been provided with data on 400 customers. The data includes the revenue from the order that they made, the advertisement medium that brought them to the website, their age, their income, the time that they have spent on the website on average over a week, and if they have been presented with an online voucher pop-up in the past. My manager has tasked me to write a report answering two business questions:

1. What factor has significantly led to customers spending more or less money on the drinks@home website, based on the demographics and behaviour of past customers?

2. Which of the three marketing projects is the best to go with for increasing profits?

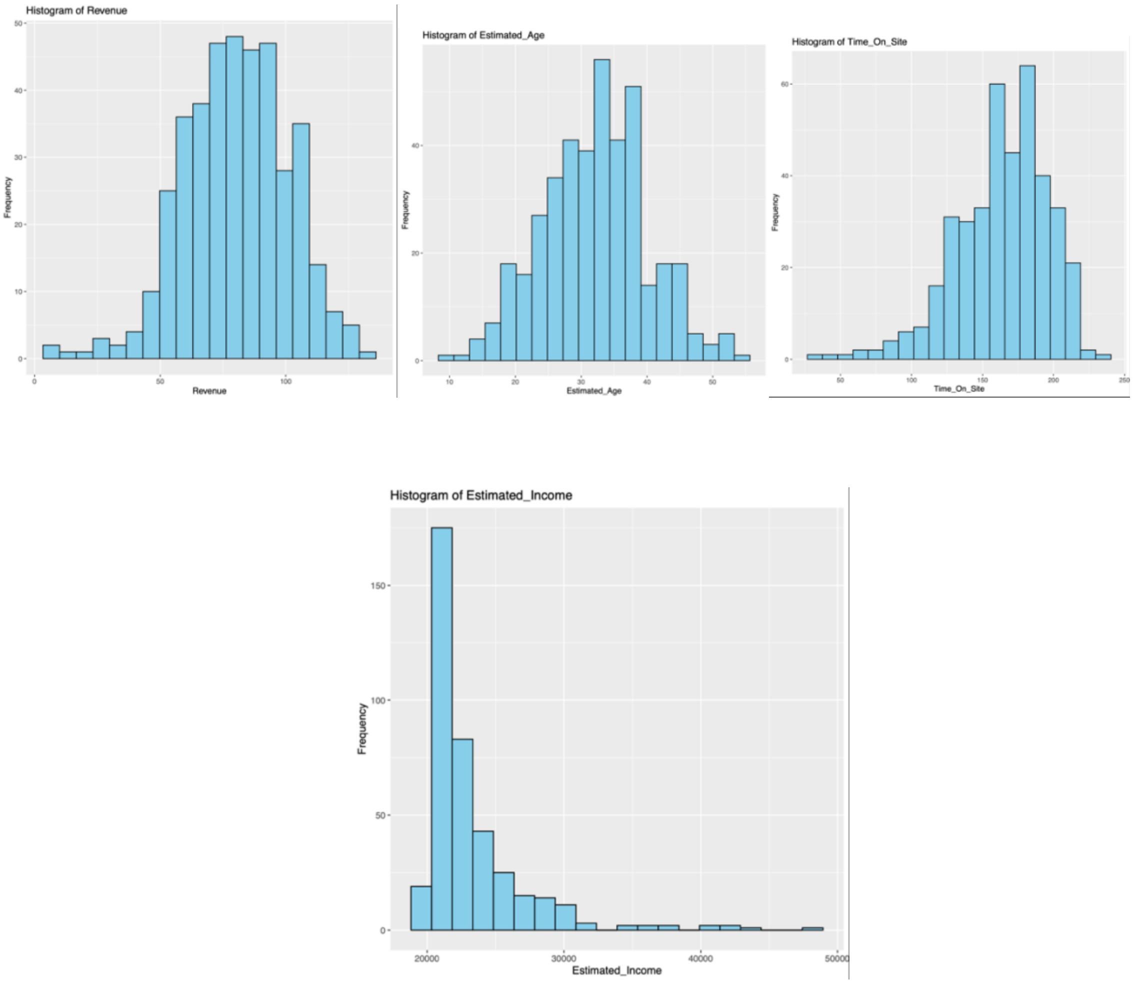
    i) Focus our advertising on customers over 45, as they are more likely to make purchases.
    ii) Offer a 20 GBP voucher to encourage future purchases.
    iii) Invest more in influencer marketing to reach a wider audience.

Data Understanding

Explantory variable **Response variable**

Variable	Data Type	Description
Revenue (GBP)	Ratio (Scalar)	Most recent order revenue
Advertisement Channel	Nominal (Categorical)	1: Leaflet 2: Social Media 3: Search Engine 4: Influencer
Estimated Age	Ratio (Scalar)	Customer's age
Estimated Income (GBP)	Ratio (Scalar)	Customer's income
Time on Website (Seconds)	Ratio (Scalar)	Average weekly website engagement
Seen Voucher	Nominal (Categorical)	0: Not seen discount voucher 1: Seen discount voucher

Exploratory Data Analysis (EDA)

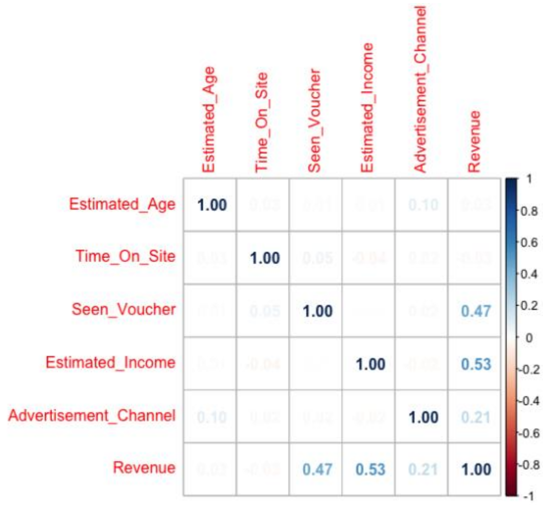


Histogram for scalar explanatory variables

Note: The histogram of Estimated Income shows a right-skewed distribution, indicating more individuals with lower incomes and fewer individuals with higher incomes.

Correlations

We'll analyse how each variable correlates with the 'Revenue' to identify potential influences.



Analysis Summary

Age and Time Spent: Our customers come in all ages, with an average of 32 years old. They visit our site for an average of 2.7 minutes per week however, some spend over 8 minutes, while some less than a minute. It appears that there is no correlation between their time spent and revenue.

Income and Vouchers: Customers with higher estimated incomes tend to spend more on their "drinks@home" purchases. Another great way to help them is by offering a coupon. Also, exposure to vouchers is highly correlated with higher spending.

Channel Choice: Reaching our customers matters as well. Even if the four ad channels appear to be equally popular, the particular channel may encourage them to make larger baskets. However, its correlation is less than that of income and vouchers.

Missing Nothing, Correlating Everything: Fortunately, there are no missing pieces in our data set, which simplifies analysis. The puzzle begins to fit together when we study the correlations. The two factors that most clearly drive revenue **are income and voucher exposure; advertising channels have a supporting role**. Age and time spent, however, appear satisfied to watch this play of spending in silence.

Next Steps: Understanding our customers is just the first step in the process. Now, with the help of regression models, we will be able to precisely measure each factor's effect on revenue.

Data Preparation

Log Transformation: We apply log transformation to ' Estimated Income' because the data is right-skewed. The log transformation helps stabilise variance and make the data more 'normal', which is good for multiple linear regression models. Improved model performance due to more normally distributed data, reduced influence of outliers, and allows for a better interpretation of regression coefficients.

Dummy Variables for Advertisement Channel: Created dummy variables for 'Advertisement Channel' with social media as the base category. This allows us to include this categorical variable in the regression analysis and understand the impact of different advertisement channels compared to social media.

Modelling

Now, applying multiple linear regression involves analysing the impact of various factors on customer spending at 'Drinks@home.uk'.

Factor	Coefficient	P-Value	Impact on Revenue
Seen Voucher	19.6955	< 2e-16	Significantly increases revenue by £19.70 per customer.
Estimated Income	0.0029	< 2e-16	Slight increase in revenue by 0.29 pence for every £1 increase in income.

Dummy: Leaflet	-6.8284	0.000783	Decreases revenue by £6.83 compared to social media.
Dummy: Influencer	6.1452	0.002567	Increases revenue by £6.15 compared to social media.
Dummy: Search Engine	1.2625	0.528702	No significant impact compared to social media.
Estimated Age	-0.0152	0.864718	No significant impact on revenue.
Time On Site	-0.0222	0.312467	No significant impact on revenue.

Summary statistics of multiple linear regression

Evaluation

Multiple R-squared: This shows that 59.31% of the variance in 'Revenue' is explained by the model, indicating a generally good fit.

Residual Standard Error (RSE): The RSE is 13.47, reflecting the average deviation of the observed values from the predicted ones in units of 'Revenue'.

Adjusted R-squared: Slightly lower at 0.5858, it adjusts for the number of predictors, providing a more accurate measure of the model's fit.

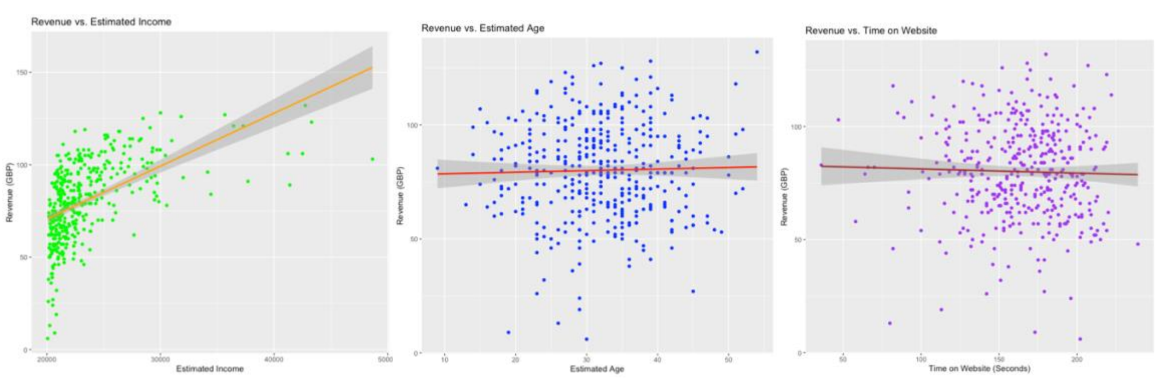
F-statistic: The overall model is statistically significant ($p < 2.2e-16$).

Assumption tests

Applying assumption tests in multiple linear regression is crucial for ensuring the validity and reliability of your model.

1. Linearity

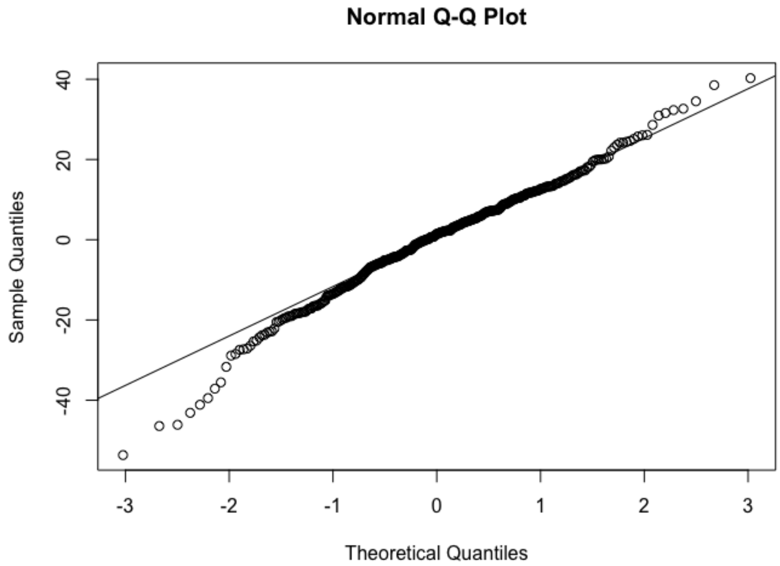
The relationship between the independent variables and the dependent variable should be linear. This can be checked with scatter plots.



Scatter plot for scalar independent variables and revenue

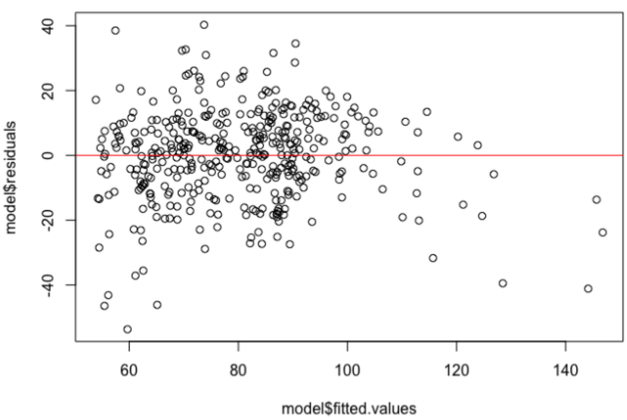
2. Normality of Residuals

The residuals should be normally distributed. This can be checked with a Q-Q plot.



3. Homoscedasticity

The variance of error terms should be constant. You can visually inspect this using a residuals vs fitted values plot.



4. Independence of Errors

Residuals should be independent, implying no correlation between consecutive residuals. This can be checked using the Durbin-Watson test. *(Zach, 2020)*

DW = 2.1346, p-value = 0.9095

5. Absence of Multicollinearity

Independent variables should not be too highly correlated. This can be checked with the Variance Inflation Factor (VIF). *(Zach, 2019)*

Factor	VIF
Seen Voucher	1.007126
Estimated Income	1.028477
Dummy: Leaflet	1.537666
Dummy: Influencer	1.549565
Dummy: Search Engine	1.515089
Estimated Age	1.011279
Time On Site	1.007985

Interpreting Results

- **Linearity:** Random patterns in the residuals vs. fitted values plot indicate linearity, also the scatter plot shows linearity.
- **Normality:** No deviations from the straight line in the Q-Q plot suggest normal residuals.
- **Homoscedasticity:** Consistent spread around the horizontal line indicates homoscedasticity.
- **Independence:** A Durbin-Watson statistic near 2 suggests independence of errors.
- **Multicollinearity:** Low VIF values indicate no multicollinearity.

Conclusion

Vouchers and Income: The most influential factors. Voucher visibility greatly increases spending, and higher income correlates with higher spending.

Advertisement Channels: Influencer marketing positively impacts revenue, while leaflet advertising negatively impacts it, compared to the baseline of social media. Search engine advertising does not show a significant difference from social media.

Age and Time on Site: These factors do not significantly influence customer spending.

Task 2: Strategic Marketing Recommendations

To solve Task 2 for 'Drinks@home.uk', we need to make a recommendation from three proposed marketing strategies based on the findings from Task 1. Let's evaluate each option using the insights gained:

Option 1: Target Customers Older Than 45

Analysis Findings: Our regression analysis in Task 1 reveals that age (Estimated_Age) has no significant impact on spending. This suggests solely age-based targeting may not be the most effective strategy.

High Revenue, Low Volume Segment: While the 46-60 age group boasts the highest average revenue, the customer count is considerably lower. Targeting them could yield high per-customer revenue, but its impact on total revenue might be limited due to their smaller demographic size.

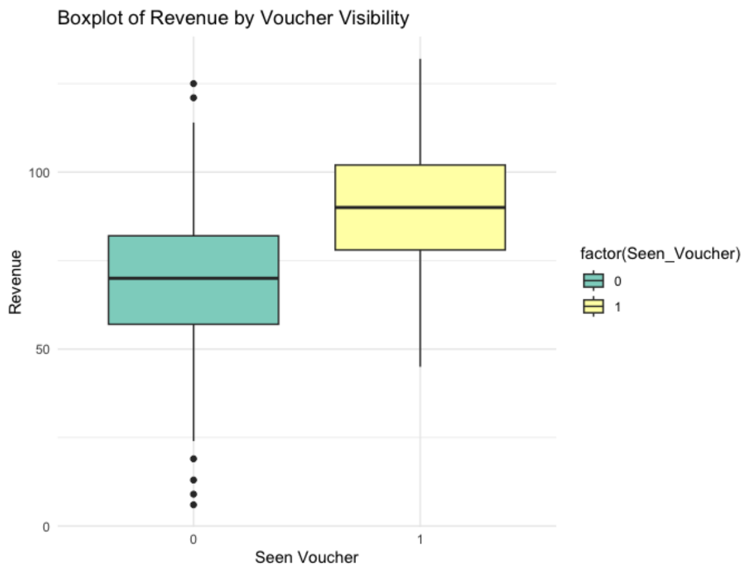
Age Group	Count	Average Revenue	Total Revenue
Under 18	21	80.3	1687
18 - 30	147	78.3	11504
30 - 45	214	80.8	17294
45 - 60	18	86.3	1553

Age bucketing analysis

Market Understanding: We recognise that older customers might possess distinct purchasing habits. However, the absence of a direct correlation between age and higher spending weakens the data-driven support for targeting this segment solely based on age.

Option 2: Provide a £20 Voucher

Analysis Findings: Task 1 revealed that the visibility of a voucher (Seen_Voucher) had a significant positive impact on revenue. Customers who saw a voucher were likely to spend substantially more.

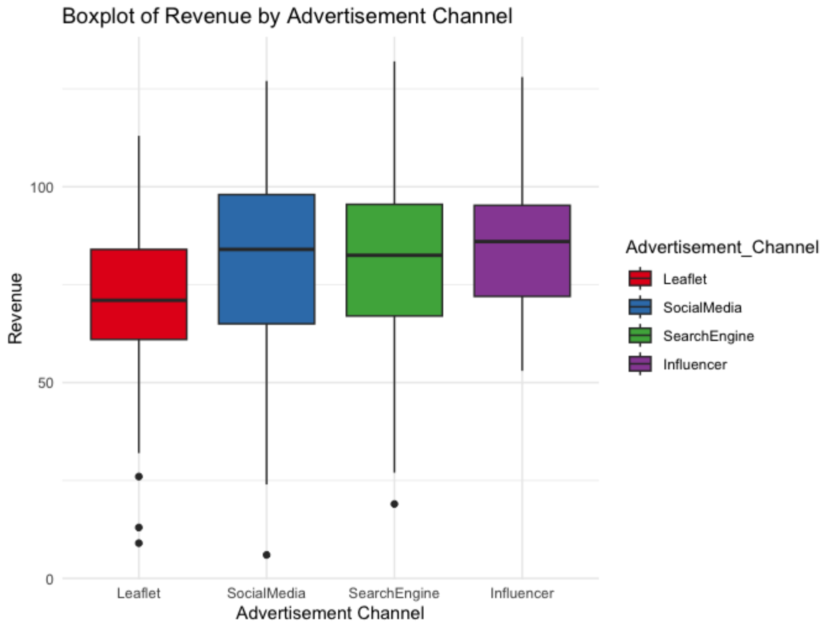


Customer Psychology: Vouchers can incentivize new purchases, increase the size of the average order, and encourage repeat business.

Cost-Benefit Analysis: The cost of offering a £20 voucher needs to be weighed against the potential increase in order size and customer acquisition/retention.

Option 3: Spend More on Influencer Advertising

Analysis Findings: The regression analysis indicated that influencer advertising (dummy_Influencer) positively influenced customer spending compared to the baseline of social media advertising.



Market Trends: Influencer marketing can be highly effective in today's digital age, potentially reaching a broad and engaged audience.

Target Audience Alignment: If the influencer's audience aligns well with 'Drinks@home.uk's target demographic, this could be a lucrative strategy.

Conclusion

'Drinks@home.uk' should prioritize **£20 Voucher** promotions to maximize immediate revenue gains. However, it's also advisable to consider a blended approach with **Influencer Marketing,** given its positive impact, allocating a portion of the budget to influencer marketing could maximize reach and effectiveness.

This blended approach capitalizes on the positive impact of vouchers and leverages influencer marketing for long-term brand growth.

Reference

1. Caylor, J.P. and Hammell II, R.J. (2021). Utilization of Multi-Criteria Decision-Making for Emergency Management. *Computación y Sistemas*, 25(4). doi: https://doi.org/10.13053/cys-25-4-4102

2. Godwin Odu (2019). Weighting methods for multi-criteria decision making technique. [online] ResearchGate. Available at: https://www.researchgate.net/publication/335806989_Weighting_methods_for_multi-criteria_decision_making_technique

3. KOKOÇ, M. and ERSÖZ, S. (2019). Comparison of AHP-TOPSIS and AHP-VIKOR Methods in Product Selection in terms of Inventory Management. *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, pp.163–172. doi: https://doi.org/10.29137/umagd.391359

4. Serafim Opricović and Tzeng, G. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research, [online] 156(2), pp.445–455. doi: https://doi.org/10.1016/s0377-2217(03)00020-1

5. Salo, A. and Hämäläinen, R.P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, [online] 6(6), pp.309–319. doi: https://doi.org/10.1002/(sici)1099-1360(199711)6:6%3C309::aid-mcda163%3E3.0.co;2-2

6. Sudha, S., Martin, N. and Smarandache, F. (n.d.). Neutrosophic Sets and Systems Neutrosophic Sets and Systems State of Art of Plithogeny Multi Criteria Decision Making Methods State of Art of Plithogeny Multi Criteria Decision Making Methods State of Art of Plithogeny Multi Criteria Decision Making Methods. Neutrosophic Sets and Systems, [online] 56, p.2023. Available at: https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=2342&context=nss_journal

7. Zach (2019). *How to Calculate Variance Inflation Factor (VIF) in R - Statology*. [online] Statology. Available at: https://www.statology.org/variance-inflation-factor-r/
8. Zach (2020). *How to Perform a Durbin-Watson Test in R - Statology*. [online] Statology. Available at: https://www.statology.org/durbin-watson-test-r/