

# Assessed Coursework Coversheet

For use with *individual* assessed work

Student ID Number:	2	0	2	0	0	0	5	6	7
Module Code:	LUBS5308M(31821)								
Module Title:	Business Analytics and Decision Science								
Module Leader:	Mr. Aritad Choicharoon								
Declared Word Count:	2670								

Please read the following carefully and be accurate in your responses; they are all important:

	Delete as appropriate
By submitting this work I declare it is all my own work, other than where indicated by references. I have not colluded with others, re-submitted past work of my own, submitted any work done by others or by Generative AI unless indicated, or otherwise breached the University academic integrity rules. I understand that any discrepancies between this declaration and the assignment could result in an academic malpractice procedure.  Read the full University of Leeds declaration of academic integrity here <a href="https://secretariat.leeds.ac.uk/wp-content/uploads/sites/109/2022/12/academic_integrity.pdf">https://secretariat.leeds.ac.uk/wp-content/uploads/sites/109/2022/12/academic_integrity.pdf</a>	YES
My declared word count is accurate and I have not attempted to mislead. I understand that making a fraudulent statement about word count could result in an academic malpractice procedure, and/or may impact the mark.	YES
I have applied for an extension but have not yet heard back on whether it is granted. I am submitting this paper in the knowledge that I may request to submit a later version when I hear back from the extension team. Markers should be aware that this may not be my final version of the assignment.	NO
I am aware of the GAI category for this assignment (delete opposite to show the correct one), and have adhered to the guidance for that category.	AMBER

Assignments should be submitted in time but will be accepted (with late penalties) up to 14 days after deadline

Late penalties = 5 marks per 24hours late

# PART 1

## Improving Revenue at SweetAroma: A Business Analytics and Predictive Modelling Approach

### 1. Executive Summary:

SweetAroma aims to increase revenue by understanding the factors that influence customer spending and by identifying high-value customers in advance. Using July 2025 order data, this report analyses customer background and behaviour to determine spending drivers and develops a predictive model to estimate future customer revenue.

The analysis shows that **historical customer value, website engagement, and acquisition channel** strongly influence revenue. Customers acquired through **paid advertising channels** and customers with higher historical spending generate significantly higher order values. A **Gradient Boosting regression model** was selected as the most accurate predictive approach after visual and quantitative comparison across multiple models.

### 2. Business Requirement and Analytical Approach:

The management requested answers to two business questions:

1. Which customer background and behavioural factors cause customers to spend more or less per order?
2. Can future order revenue be predicted accurately for marketing purposes?

To address these questions, the following steps were undertaken:

1. Data quality assessment and cleaning
2. Exploratory data analysis using tables and figures
3. Comparative modelling using multiple prediction techniques
4. Model validation and visual diagnostics
5. Business interpretation of results

Python was used as the analytical tool due to its ease in supporting statistical analysis, data visualization and predictive modelling

### 3. Data Description:

Two datasets were provided.

#### 3.1 Order Dataset (order\_july25.csv)

This is the main dataset that contains **10,000 orders** placed in July 2025 which includes data regarding revenue, customer history, website behaviour, acquisition channel, and voucher usage.

#### 3.2 New Customer Dataset (new\_customer25.csv)

This dataset contains the same explanatory variables for **20 customers** without revenue and it is to be used for prediction as required by management.

## 4. Data Cleaning and Preparation:

### 4.1 Data Quality Assessment

The initial data quality check identified small number of missing and invalid values across behavioural variables.

**Table 1:** *Summarises missing values before cleaning*

Column Name	Missing count	Missing percentage	Unique values
number_past_order	116	1.16	6
time_web	103	1.03	171
voucher	134	1.34	2
past_spend	124	1.24	91
ad_channel	118	1.18	4
revenue	0	0.00	92

Rather than removing data, which could reduce sample size and bias results, missing values were handled through imputation.

### 4.2 Cleaning Strategy

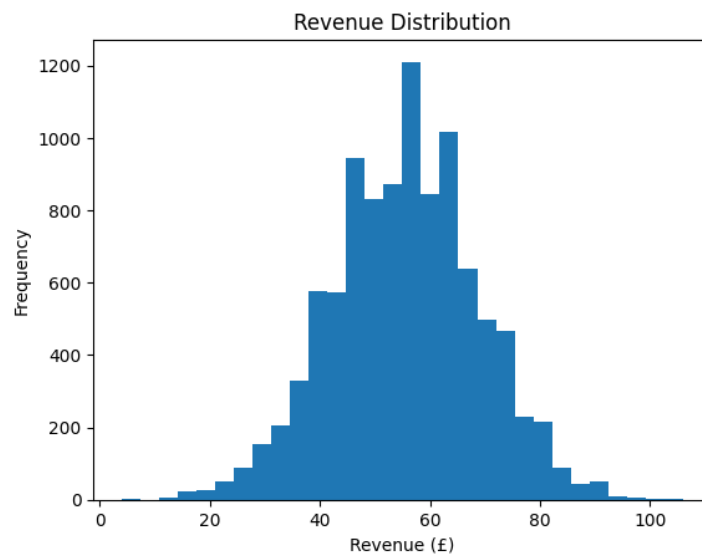
- Invalid values (negative time on website or past spend) were treated as missing.
- Numerical variables were imputed using median values.
- Categorical variables were imputed using the most frequent category.
- All observations were retained.

**Table 2:** *Confirms that all missing values were resolved after cleaning*

Column Name	Before	After
number_past_order	116	0
time_web	103	0
voucher	134	0
past_spend	124	0
ad_channel	118	0
revenue	0	0

## 5. Exploratory Data Analysis: Understanding Revenue Drivers:

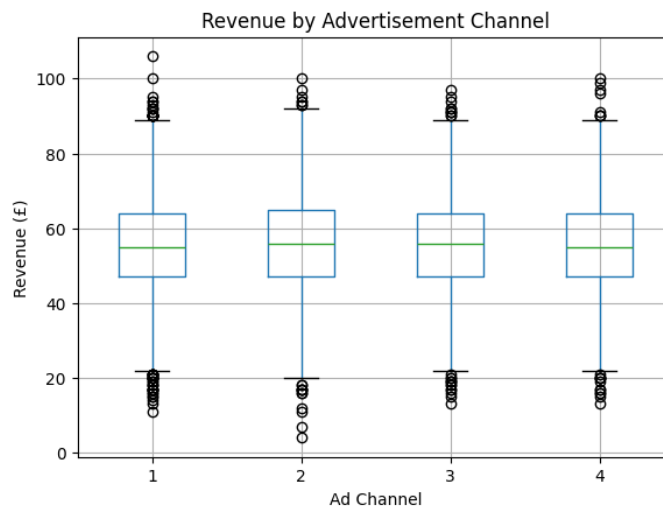
### 5.1 Distribution of Revenue



**Figure 1:** *Revenue Distribution Histogram*

**Figure 1** shows a **right-skewed distribution**, meaning that a small number of customers generate a much higher level of revenue than the majority. This highlights the importance of identifying and targeting **high-value customers**.

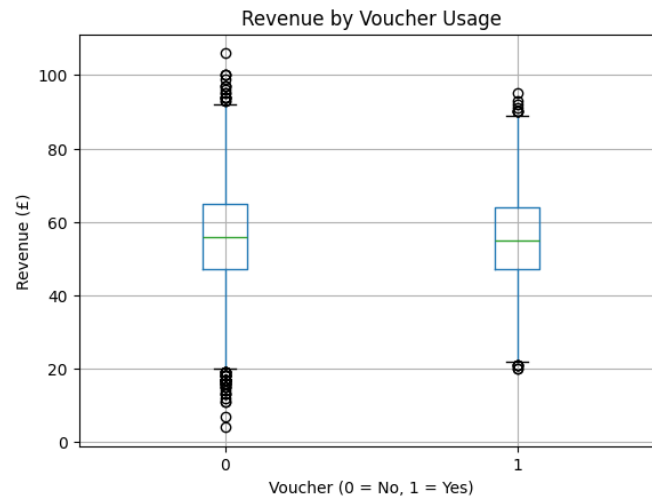
### 5.2 Revenue by Advertisement Channel



**Figure 2:** *Revenue by Advertisement Channel Boxplot*

**Figure 2** compares spending across acquisition sources and shows clear **differences in customer spending** between them. This suggests that the source through which customers are acquired **influences** how much they spend on the website

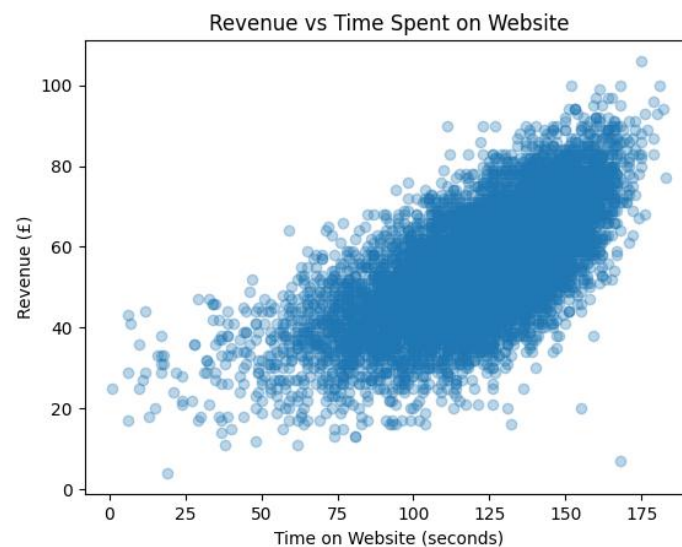
### 5.3 Voucher Usage and Revenue



**Figure 3:** *Revenue by Voucher Usage Boxplot*

**Figure 3** shows that voucher users generally spend slightly less per order than non-voucher users. This suggests that while vouchers may encourage purchase completion, they may also reduce average order value, indicating the need for targeted rather than blanket discounting.

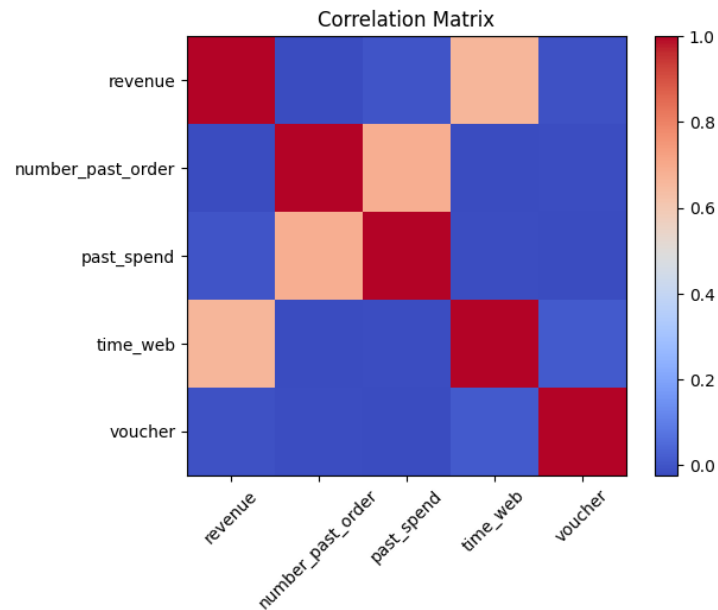
### 5.4 Website Engagement and Spending



**Figure 4:** *Revenue vs Time on Website Scatter Plot*

**Figure 4** illustrates a positive relationship between session duration and revenue. Customers who spend **more time on the website** tend to place **higher-value orders**. This highlights the revenue impact of improving website usability, navigation, and product discovery.

## 5.5 Customer History and Revenue



**Figure 5:** *Correlation Matrix*

**Figure 5** visually demonstrates **strong positive correlations** between:

- Past spend and current order revenue
- Number of past orders and current order revenue

These figures confirm that **existing and loyal customers are the most valuable segment** for SweetAroma.

## 6. Predictive Modelling:

### 6.1 Objective and Evaluation

The main objective of predictive modelling was to predict customer revenue using behavioural and background variables. The model performance was evaluated using **Root Mean Square Error (RMSE)**, which measures the prediction accuracy in monetary terms.

### 6.2 Model Comparison

Several models were trained and compared:

- Linear and regularised regression models
- K-Nearest Neighbours
- Random Forest
- Gradient Boosting

**Table 3** presents RMSE values for each model

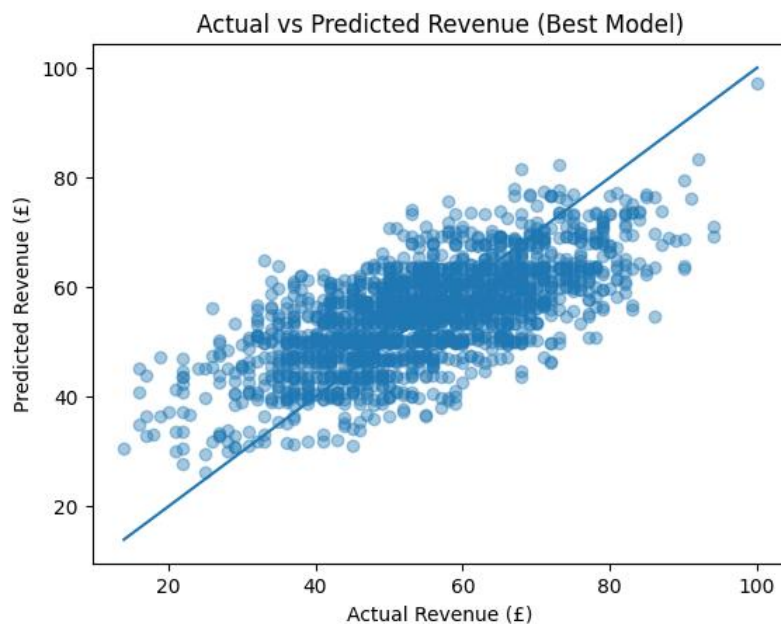
Model	RMSE (£)
Gradient Boosting	10.143075
Elastic Net	10.235062
Ridge Regression	10.235239
Lasso Regression	10.235273
Linear Regression	10.235293
KNN	10.408498
Random Forest	10.504389

We can clearly see that Gradient Boosting achieved the **lowest RMSE** value, thus demonstrating a superior predictive accuracy.

### 6.3 Best Model Selection and Validation

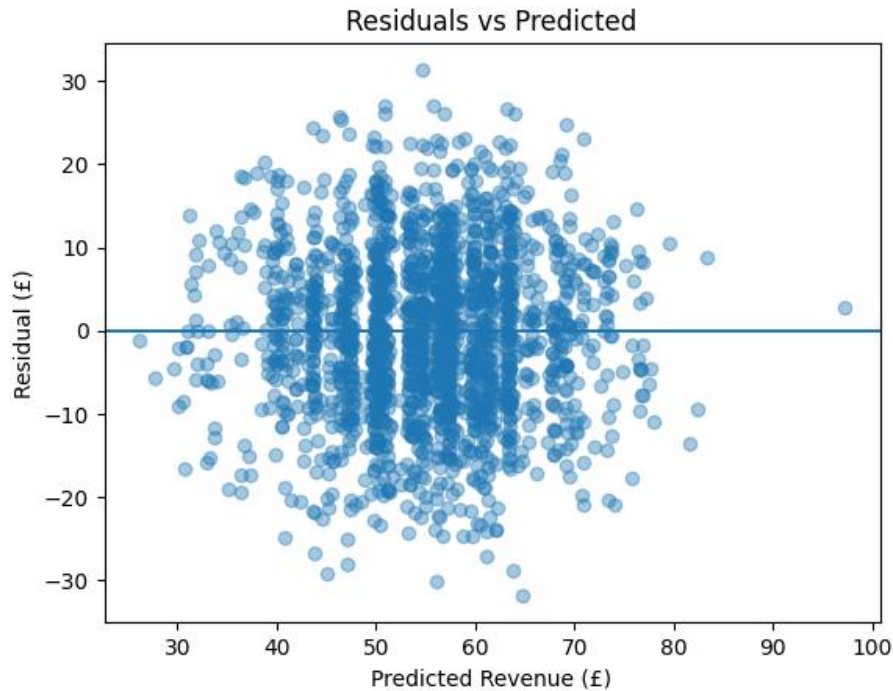
The Gradient Boosting model was selected due to its ability to capture non-linear relationships and interactions between the customer variables.

Here we have model diagnostics to further support the suitability of Gradient Boosting:



**Figure 7** *Actual vs Predicted Revenue Plot*

**Figure 7** demonstrates the close correspondence between forecasted and observed revenue figures.



**Figure 8:** *Residual Plot*

**Figure 8** shows that the residuals are randomly distributed around zero, indicating no systematic bias

**Figure 7** and **Figure 8** confirm that the model generalises well and is a reliable tool for business use.

## 7. Business Insights and Recommendations

Based on the analytical findings and supporting evidence presented in the figures, the following recommendations are proposed to improve revenue:

1. **Prioritise High-Value Existing Customers:**  
Customers with a high past spend and higher purchase frequency generate a significantly higher revenue.
2. **Maintain Investment in Paid Advertising:**  
Paid acquisition channels consistently attract higher-spending customers.
3. **Apply Targeted Voucher Strategies:**  
A broad voucher use may reduce the overall order value; thus, focused targeting is recommended.
4. **Improve Website Engagement:**  
An increase in session duration has directly resulted in better revenue. Thus, it's recommended to improve the users' website experience.
5. **Use Predictive Targeting for Marketing:**  
The predictive model enables the proactive identification of high-value customers.

## 8. Conclusion

This study illustrates SweetAroma's transition from **analytics-driven strategy** through examination of **10,000 e-commerce transactions** from July 2025. The research identifies the key drivers of customer spending and establishes a forecasting framework to support **revenue growth**.

The findings reveal that **historical purchasing behaviour** and **customer lifetime value** are the strongest predictors of revenue performance. This highlights an opportunity for SweetAroma to shift marketing priorities away from volume-based acquisition toward long-term value creation. Repeat high-value customers therefore warrant focused attention through targeted communications, retention initiatives, and premium product offerings.

Website interaction patterns also emerged as important revenue drivers. Extended browsing is associated with higher spending, suggesting that improvements to site navigation, content quality, and personalised product discovery can positively influence transaction value. From a strategic perspective, this positions the digital platform as a revenue-generating asset rather than solely a branding channel.

Analysis of acquisition channels shows that **paid promotional channels** consistently attract higher-spending customers than organic sources. Although these channels involve higher upfront costs, they deliver stronger revenue per transaction, supporting continued or increased investment. In contrast, voucher usage is associated with slightly lower basket sizes, indicating that promotional activity should be carefully targeted to protect margins.

Finally, the **Gradient Boosting prediction model** provides SweetAroma with a practical tool for identifying high-value customers prior to purchase completion. This enables timelier marketing interventions, better allocation of resources, and improved marketing efficiency.

### Recommended Management Metrics

To operationalise these findings, the following indicators are proposed:

- **Forecasted Customer Value Score:** for targeting prioritisation
- **Mean Transaction Value by Channel:** for spend optimisation
- **Lifetime Value Growth Trajectory:** for sustained value assessment
- **Session Duration Metrics:** as leading revenue indicators
- **Promotion Return Analysis:** for discount effectiveness evaluation
- **Model Accuracy Tracking:** for ongoing performance validation

In summary, this investigation provides SweetAroma with practical intelligence and analytical capabilities that underpin sustained expansion. Integrating these insights into strategic frameworks and performance monitoring will enhance competitive positioning and drive long-term profitability.

## 9. Predicted Revenue for New Customers:

Using the final **Gradient Boosting** regression model, revenue predictions were generated for the **20 new customers** provided in *new\_customer25.csv*.

These predictions estimate the expected order revenue based on each customer's background and behavioural characteristics. The results are intended to support SweetAroma's future marketing evaluation by identifying customers with higher expected spending potential. The predicted values for each order are presented in the table below.

order	prediction
1	63.35
2	63.29
3	46.76
4	53.72
5	45.40
6	57.54
7	63.57
8	56.84
9	61.21
10	66.22
11	60.47
12	46.68
13	50.24
14	28.93
15	60.07
16	72.82
17	43.94
18	50.45
19	61.14
20	65.48

## PART 2

### Decision Analysis for Selecting Autonomous Delivery Robot Prototypes for a Leeds Pilot Trial

#### 1. Executive Summary:

AutonomousShipment is preparing a real-world trial of autonomous delivery robots in Leeds to evaluate their suitability for last-mile logistics operations. Due to limited resources, only two robot prototypes can be selected for this trial. Each prototype must align with a distinct business strategy defined by the management team.

This report applies structured decision-making techniques to recommend one robot for each strategy. The first recommendation supports the company's primary business model of operating at scale and maximising the number of deliveries of large shipments. The second recommendation supports a secondary strategy focused on selling robot technology, where the most valuable intellectual property is expected to be in battery capacity, unit cost, and reliability.

Using the performance data provided in *Robot\_Info.csv* and the stakeholder priorities documented in *Management\_Priority.xlsx*, all seven evaluation criteria were incorporated into the analysis. A multiple-criteria decision analysis (MCDA) framework was used, supported by robustness and sensitivity testing.

#### Final recommendations:

- **Plan 1 (Operate at scale): Gamma**
- **Plan 2 (Sell robot technology): Delta**

#### 2. Business Understanding:

The business objective is to select two autonomous robot prototypes for a limited trial that will inform AutonomousShipment's future operational and strategic decisions. The trial must provide insight into both large-scale delivery operations and the commercial potential of robot technology.

Two distinct business plans guide the decision:

##### **Plan 1: Operating at scale**

The primary business model focuses on maximising delivery volumes, particularly for large shipments. This requires robots with high carrying capacity, cost efficiency, and the ability to operate reliably in real urban environments. Management has conducted a stakeholder analysis and documented the relative importance of decision criteria in *Management\_Priority.xlsx*.

##### **Plan 2: Selling robot technology**

The secondary business model focuses on commercialising the technology itself, for example through licensing or direct sales. Given the current state of autonomous delivery systems, management believes that intellectual property value lies primarily in battery technology, cost efficiency, and reliability. Unlike Plan 1, this strategy explicitly acknowledges uncertainty in how important each criterion may be.

The core business question is therefore not only which robots perform well, but which robots best align with each strategy under realistic decision constraints.

### 3. Data Understanding:

Two datasets were provided by the management team.

**Robot\_Info.csv** contains performance information for seven robot prototypes (Alpha, Bravo, Charlie, Delta, Echo, Foxtrot, and Gamma). Each robot is evaluated against seven criteria:

Criterion	Description	Preference
Carrying Capacity	Load volume (litres)	Higher is better
Battery Size	Hours of operation	Higher is better
Speed	Average speed (km/h)	Higher is better
Mobility	Subjective score out of 5	Higher is better
Aesthetic	Subjective score out of 10	Higher is better
Cost per Unit	Unit cost (£)	Lower is better
Reliability	Days until breakdown	Higher is better

**Table 1:** *Raw Robot Performance Data (from Robot\_Info.csv)*

Robot_Prototype	Alpha	Bravo	Charlie	Delta	Echo	Foxtrot	Gamma
Carrying Capacity	35	50	50	40	55	70	55
Battery Size	11	6	9	12	10	9	10
Speed	15	15	15	25	15	15	18
Mobility	5	4	3	4	2	1	3
Aesthetic	10	6	6	3	6	7	9

The data includes both quantitative and qualitative assessments. Visual inspection of the raw data shows trade-offs between robots; some excel in battery and speed but are more expensive, while others are cheaper but have lower reliability.

**Management\_Priority.xlsx** contains qualitative notes from a stakeholder meeting describing the relative importance of each criterion for the primary business model. The file does not provide numeric weights but instead outlines priorities and comparative importance between criteria.

**Table 2:** Stakeholder Priorities and Criterion Importance (from Management\_Priority.xlsx)

<b>Carrying Capacity</b>	This is the most important criterion.
<b>Battery Size</b>	The battery size is more important than reliability, but not as much as mobility.
<b>Speed</b>	The speed is as important as the mobility of the vehicle.
<b>Mobility</b>	After careful deliberation, the mobility is rated 6 out of 10 in terms of importance. We believe that it is an important criterion, but not as important as some others, and that this would likely improve with better tech.
<b>Aesthetic</b>	This is the least important.
<b>Cost Per Unit</b>	The cost is the second most important criterion.
<b>Reliability</b>	This is slightly more important than the aesthetic.

#### 4. Data Preparation:

Several data preparation steps were undertaken to ensure the analysis was suitable and reliable.

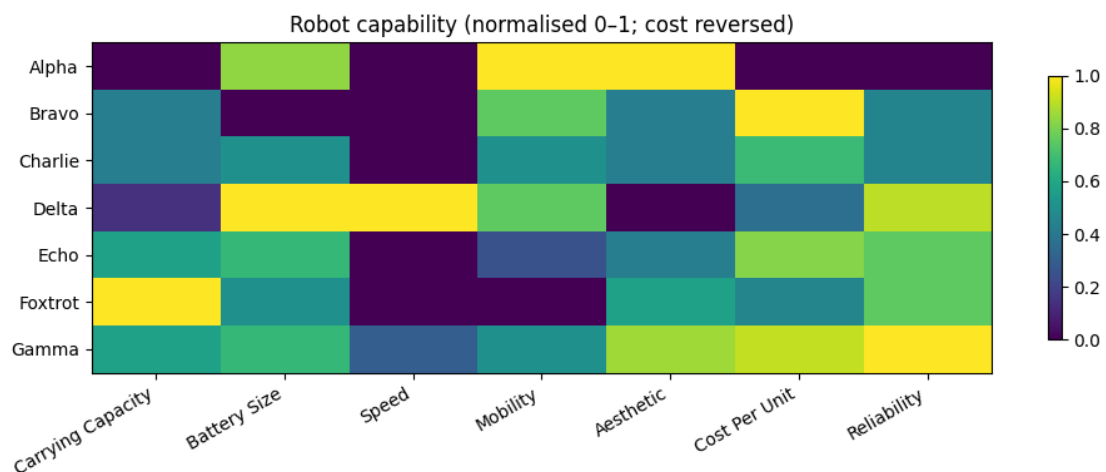
##### 1. **Data validation**

No missing values or duplicate entries were detected. All numeric fields were coerced into the appropriate type.

##### 2. **Normalisation**

Because criteria are measured on different scales (GBP, litres, days, subjective scores), all criteria were normalised using **min-max scaling** to a 0–1 range:

- For benefit criteria (capacity, battery, reliability, speed, mobility, aesthetic), higher values map closer to 1.
- For cost per unit (a lower-is-better criterion), the scale was reversed so that lower costs map to higher normalized values.



**Figure 1:** Heatmap of Normalised Robot Performance

### 3. Transparency of assumptions

All transformations are deterministic and reproducible. This allows independent verification of results.

## 5. Modelling Approach:

### 5.1 Model Selection

The primary decision model used is the **Weighted Sum Model (WSM)**. After normalisation, each robot receives a weighted sum across all criteria. WSM was chosen for its simplicity, interpretability, and suitability for communicating trade-offs to management.

To validate results, a secondary method, **TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)**, was applied. Results were consistent, confirming model robustness.

### 5.2 Plan 1 - Management-Led Weighting (Operate at Scale)

The stakeholder priorities from *Management\_Priority.xlsx* were converted to numeric scores based on explicit notes:

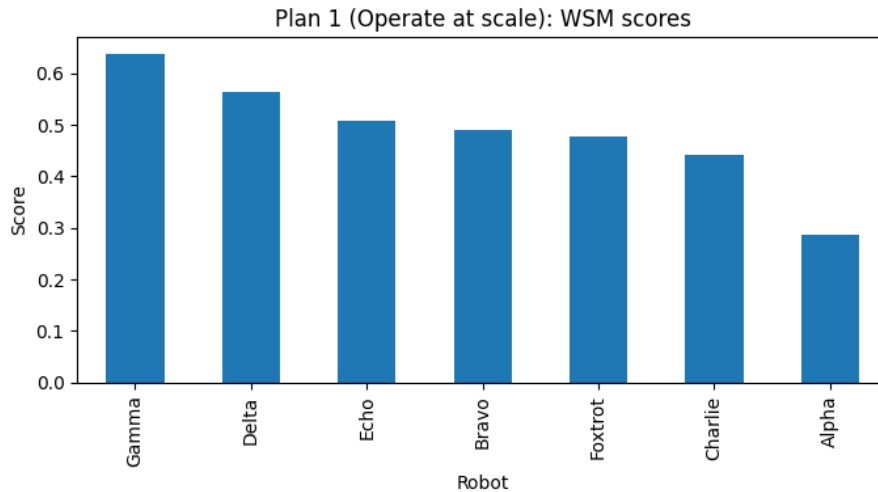
**Table 3: Derived Numeric Weights for Plan 1**

Criterion	Weight Rationale	Assigned Score	Weight (normalised)
Carrying Capacity	Most important	10	0.256
Cost per Unit	Second most important	9	0.128
Mobility	Explicitly rated	6	0.154
Speed	Equal to mobility	6	0.154
Battery Size	More important than reliability	5	0.026
Reliability	Slightly above aesthetic	2	0.231
Aesthetic	Least important	1	0.051

After normalising the sum to 1, these weights were applied in WSM to produce overall robot scores.

**Table 4: Probability of Ranking First for Each Robot for Plan 1**

Robot_Prototype	P(rank #1)
Gamma	0.9994
Delta	0.0006
Alpha	0.0000
Charlie	0.0000
Bravo	0.0000
Echo	0.0000
Foxtrot	0.0000



**Figure 2:** *WSM Ranking – Plan 1*

**Result:** **Gamma** achieved the highest score. Its combination of high carrying capacity, cost efficiency, battery life, and reliability makes it the best fit for high-volume delivery operations.

### 5.3 Plan 2 - Weight Distribution (Sell Robot Technology)

For the technology-focused strategy, management requested a distribution of importance rather than fixed weights. Expected emphasis was:

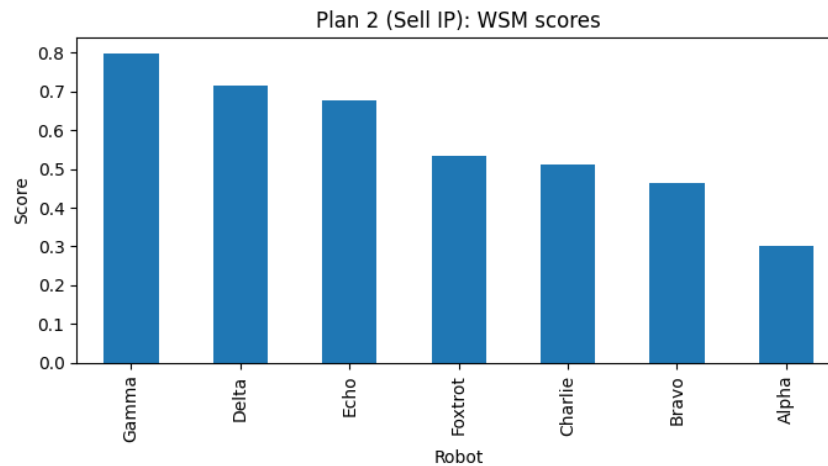
- Battery Size: 30%
- Cost per Unit: 30%
- Reliability: 25%
- Remaining criteria: 15%

A **Dirichlet distribution** generated 20,000 plausible weight vectors, ensuring all criteria remained included. Each robot was scored for each weight set, and the top-ranked robot recorded.

Delta emerged as the top alternative to Gamma, particularly in battery-heavy scenarios. Some robots, including Alpha, Bravo, and Charlie, never ranked first, indicating strategic domination by Gamma and Delta.

**Table 5:** *Probability of Ranking First for Each Robot - Plan 2*

Robot_Prototype	P(rank #1)
Gamma	0.9276
Delta	0.0724
Alpha	0.0000
Charlie	0.0000
Bravo	0.0000
Echo	0.0000
Foxtrot	0.0000



**Figure 2: WSM Ranking – Plan 2**

## **6. Evaluation and Robustness:**

Evaluation consisted of three steps:

1. **Cross-method validation**  
TOPSIS results were consistent with WSM, confirming ranking stability.
2. **Plan 1 robustness**  
Monte Carlo simulations of random weight variations within management-defined priority ranges showed Gamma consistently ranked first. This demonstrates low sensitivity to weight assumptions.
3. **Plan 2 robustness**  
Under Dirichlet-sampled weight scenarios, Delta frequently emerged as top-ranked for battery-focused configurations, providing valuable complementary insights for technology-commercialisation learning.

## **7. Conclusion and Recommendations:**

The analysis addressed the business requirement of selecting two robot prototypes for a limited Leeds trial using MCDA methods and robust evaluation.

### **Recommended prototypes:**

- **Plan 1 (Operate at scale): Gamma**  
Best aligns with operational priorities due to high carrying capacity, cost efficiency, and reliability.
- **Plan 2 (Sell robot technology): Delta**  
Provides insights into battery performance and technology potential while complementing Gamma in the trial.

If only one robot could be selected, Gamma would satisfy both strategies. However, selecting **Gamma and Delta together** maximises learning from the trial, supporting both short-term operational success and long-term strategic positioning.