Question 1

1 Business Understanding

This report addresses two key questions to support our expansion strategy:

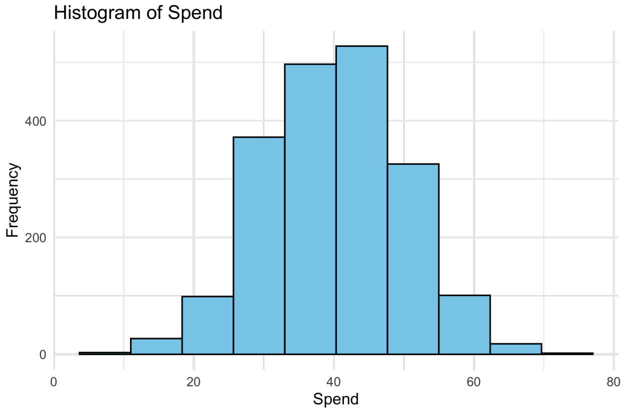
1. What factors influence variations in customer spending on our platform?
2. How can we predict customer spending to evaluate and target specific customer groups effectively?

The analysis begins with understanding customer behaviour through exploratory analysis of the provided datasets. Key patterns and drivers of spending are identified to inform marketing strategies. Subsequently, a predictive model is developed, validated, and evaluated to forecast spending trends. The results will provide actionable insights to support data-driven decision-making, enabling resource optimisation and strategic targeting of high-value customer segments.

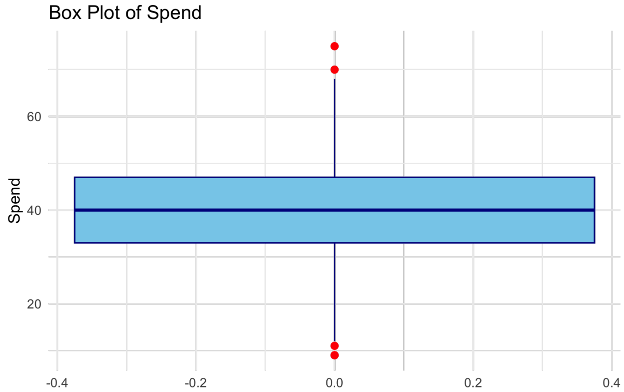
2 Data Understanding

Statistic	**Value**
Minimum	9.00
1st Quartile (Q1)	33.00
Median	40.00
Mean	40.15
3rd Quartile (Q3)	47.00
Maximum	75.00
Missing Values (NA's)	27
Standard Deviation	9.82
Variance	96.42

Table 1: Summary Statistics for Target Variable (`spend`)



(a) Spend Histogram



(b) Spend Box Plot

Figure 1: Visualisations for the Spend Variable

2.1 Target Variable: Spend

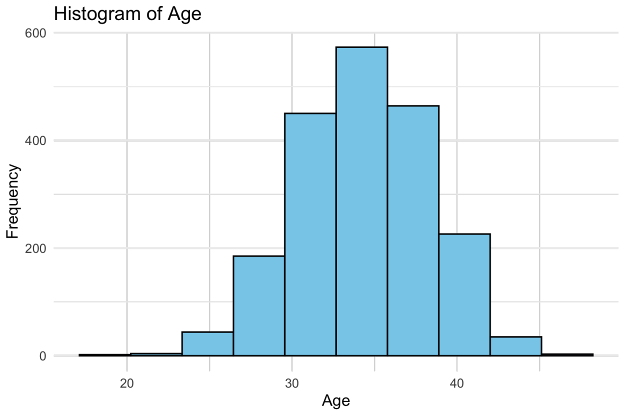
The histogram (Figure 1a) shows a symmetric distribution, with values clustered around the mean. The standard deviation and IQR indicate moderate variability. Outliers beyond the upper quartile may represent high-value customer segments.

2.2 Feature Variables

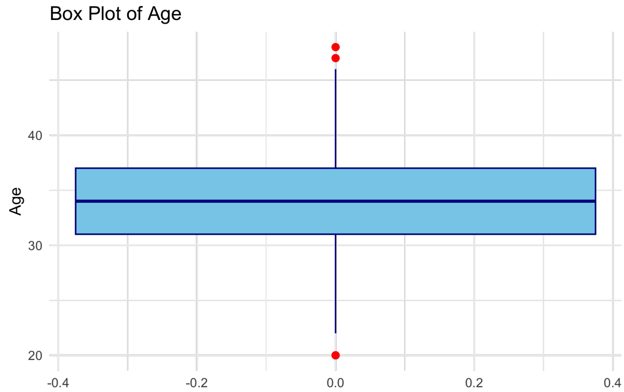
Variable	Min	1st Qu.	Median	Mean	3rd Qu.	Max	NA's	SD	Variance
Past Spend	0.00	4.00	12.00	13.02	20.00	51.00	25	10.49	110.08
Age	20.00	31.00	34.00	34.14	37.00	48.00	14	3.97	15.78
Time on Website	0.00	73.00	83.00	81.21	93.00	124.00	21	16.59	275.30

Table 2: Summary statistics for Past Spend, Age, and Time on Website

2.2.1 Age



(a) Age Histogram

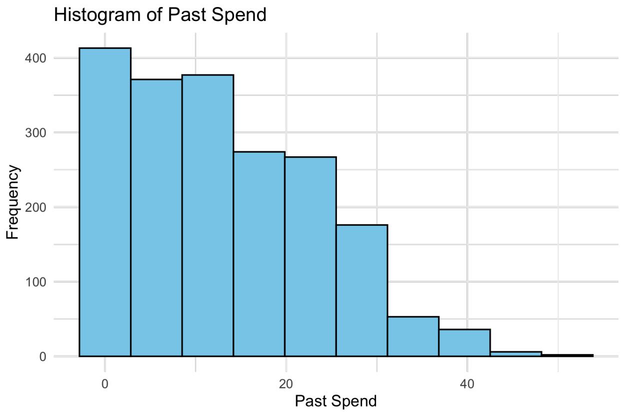


(b) Age Box Plot

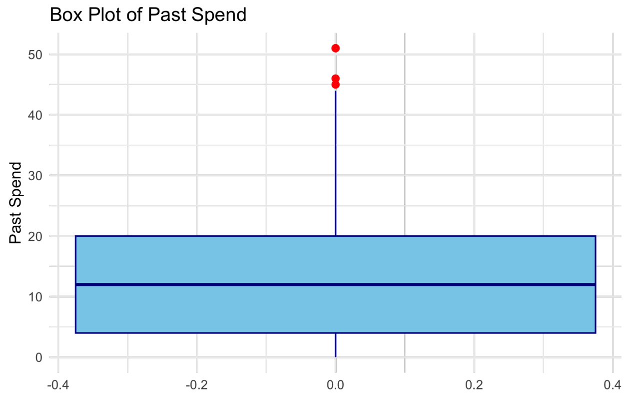
Figure 2: Visualisations for the Age Variable

Most customers fall within 31–37 years (IQR). This narrow range likely reflects a targeted demographic, warranting further analysis of spending trends by age.

2.2.2 Past Spend



(a) Past Spend Histogram

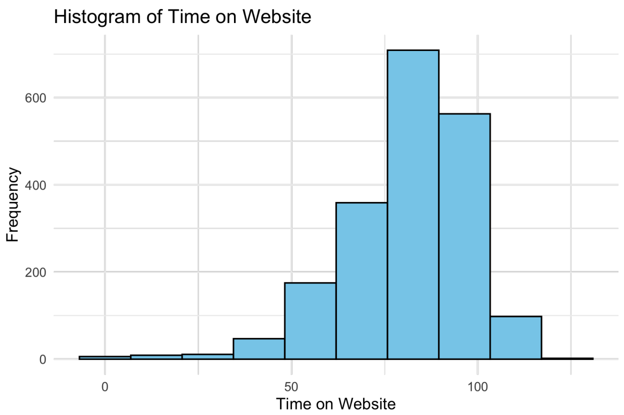


(b) Past Spend Box Plot

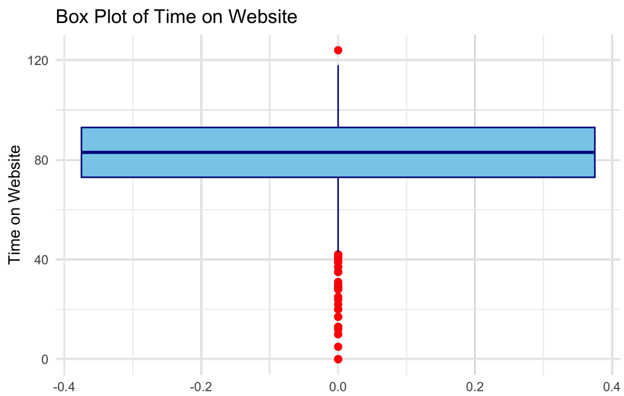
Figure 3: Visualisations for the Past Spend Variable

The histogram (Figure 3a) shows a right-skewed distribution, with high-value outliers suggesting loyal customers.

2.2.3 Time on Website



(a) Time Spent Histogram



(b) Time Spent Box Plot

Figure 4: Visualisations for the Past Spend Variable

Most customers spend 73–93 seconds (IQR) on the website. Outliers represent quick purchases. Five rows correspond to a time spent on the website of 0 seconds. This raises questions about potential data collection errors. Further investigation is needed to determine if this is a valid value or an issue with data recording.

2.2.4 Advertisement Channel

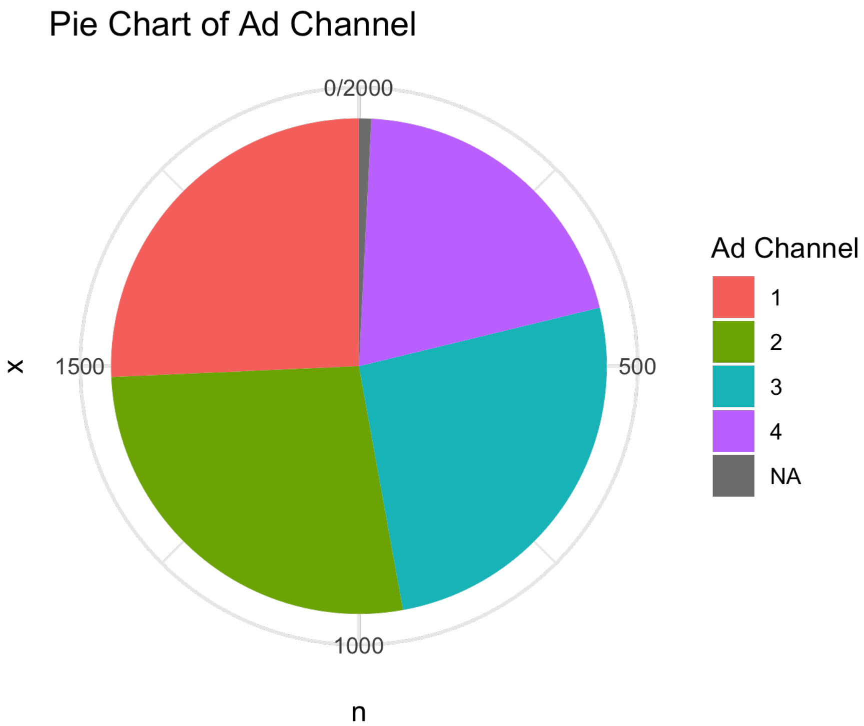


Figure 5: Visualisation of AD Channel variable distribution

Channels are evenly distributed, with slight under representation of Leaflets (`ad_channel1`).

2.2.5 Voucher

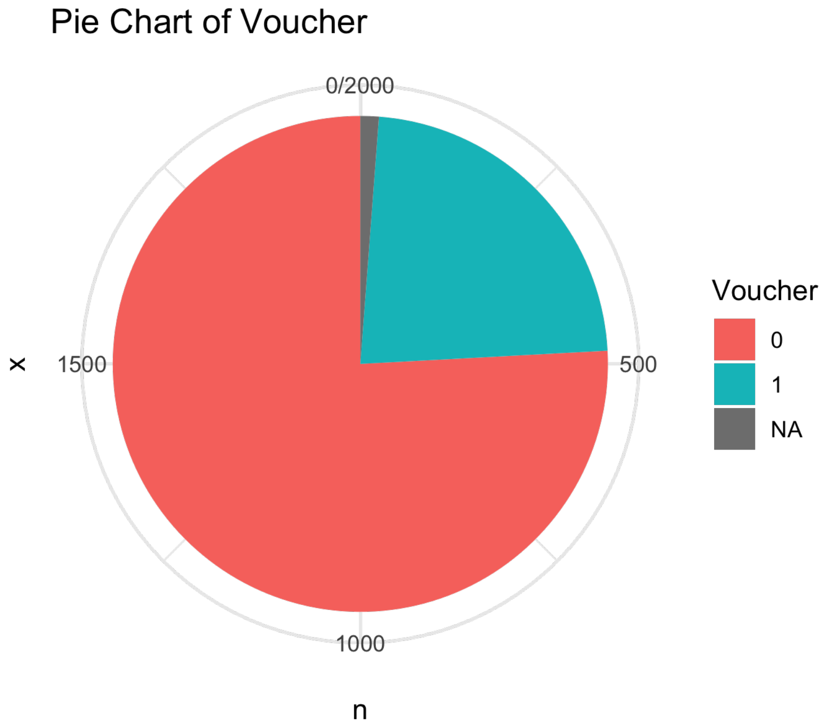


Figure 6: Visualisation of Voucher Variable Distribution

Around 22.95% of customers used a 5% discount voucher (Figure **??**) showing a lack of engagement.

3 Data Preparation

3.1 Missing Data

Variable	**Missing Values**	**Total Values**	**Missing Percentage (%)**
Past Spend	25	2000	1.25
Age	14	2000	0.70
Advertisement Channel (`ad_channel`)	16	2000	0.80
Time on Website	21	2000	1.05
Voucher	24	2000	1.20
Spend	27	2000	1.35

Table 3: Missing data statistics for each variable.

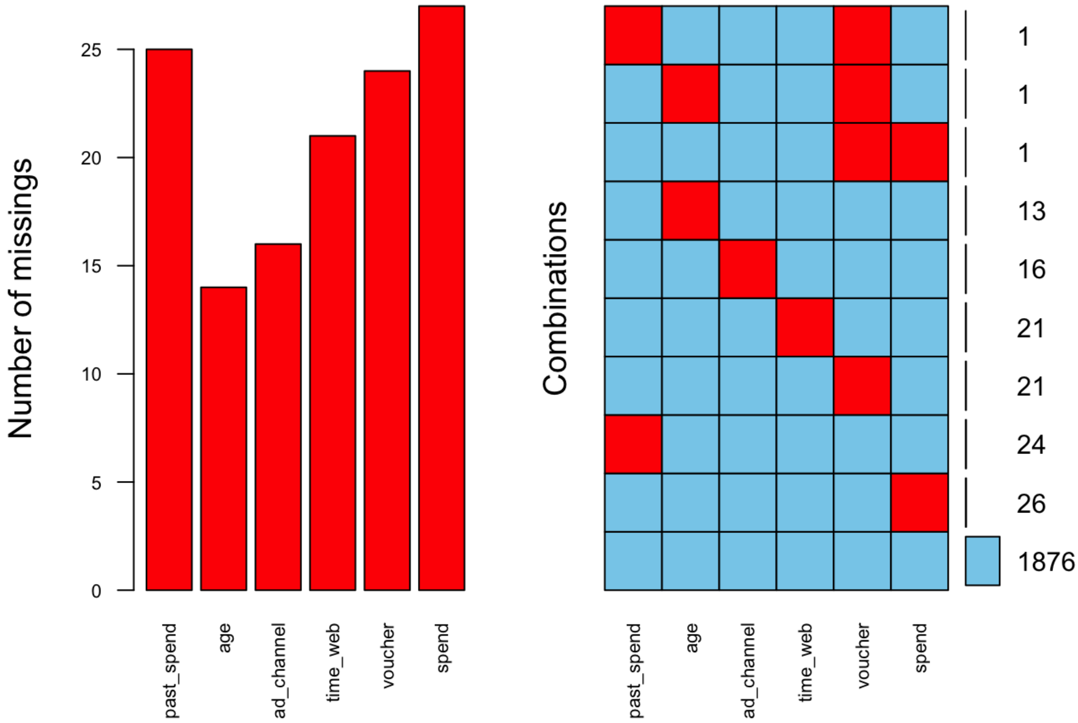


Figure 7: Visualisation of Missing Variables Location

Analysis revealed missing values in all columns, with 124 total rows (6.2%) incomplete. Given the low proportion, a deletion method was applied, resulting in a clean dataset of 1,876 observations. Minimal changes in the summary statistics confirmed that this method preserved the integrity of the data.

3.2 Dummy Encoding

The categorical variable `ad_channel` was dummy-encoded, with Leaflet (`ad_channel1`) as the baseline. Boolean columns were created for Social Media, Search Engine, and Influencer channels, enabling the model to assess their relative effects.

3.3 Correlation Analysis

Correlation analysis highlighted:

- **Strong correlations:** Age (0.62) and `time_web` (0.60) positively correlate with spend, indicating older customers and those engaging longer with the website tend to spend more.
- **Moderate correlation:** `past_spend` (0.39) suggests historical spending somewhat predicts future spending.
- **Negligible correlations:** `voucher` (0.02) and `ad_channel` dummy variables exhibit minimal impact but may contribute indirectly in advanced models.

This analysis emphasises `age`, `time_web`, and `past_spend` as key predictors while noting potential multivariate contributions of low-correlation features.

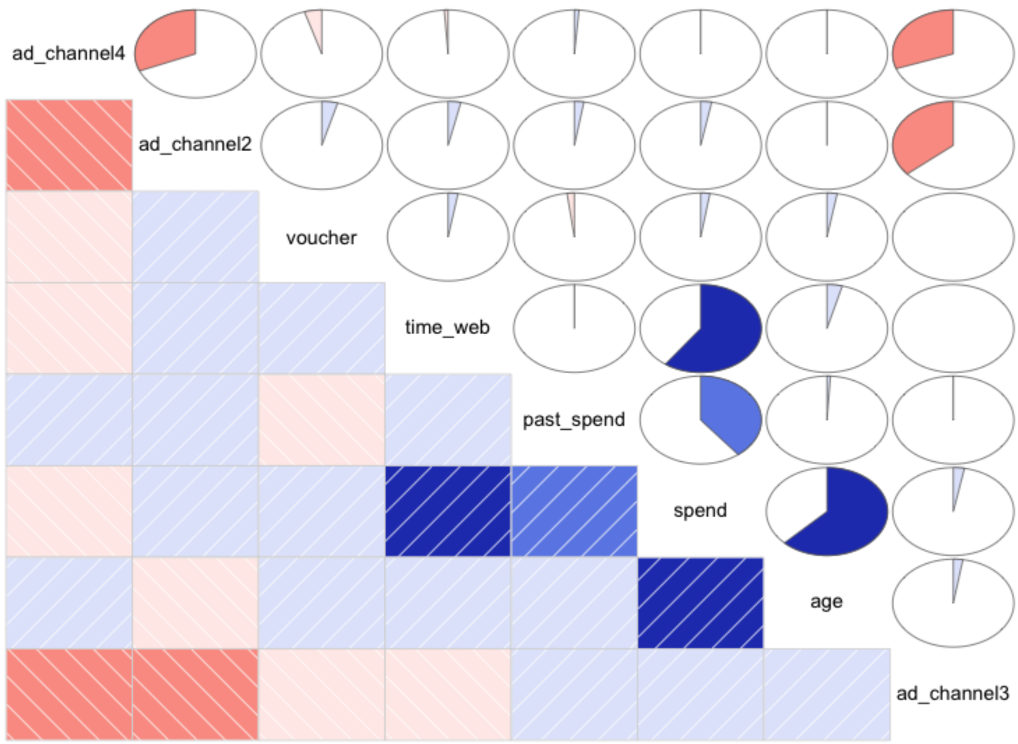


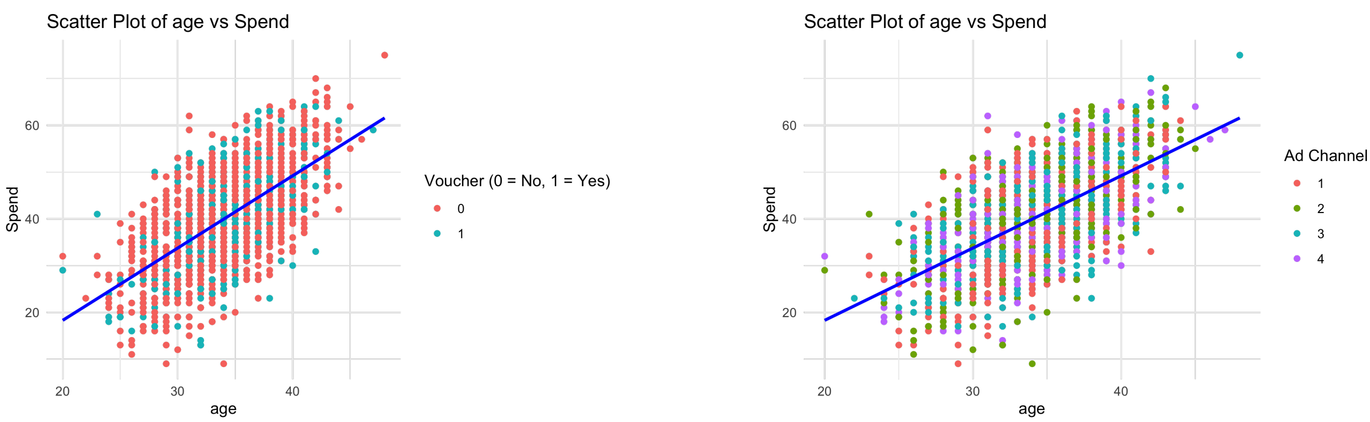
Figure 8: Visualisation of Correlations between Variables

4 Modelling

The target variable (`spend`) is scalar, making this a regression problem. Linear regression was selected for its simplicity and interpretability. To ensure suitability, the following assumptions were evaluated:

4.1 Linearity

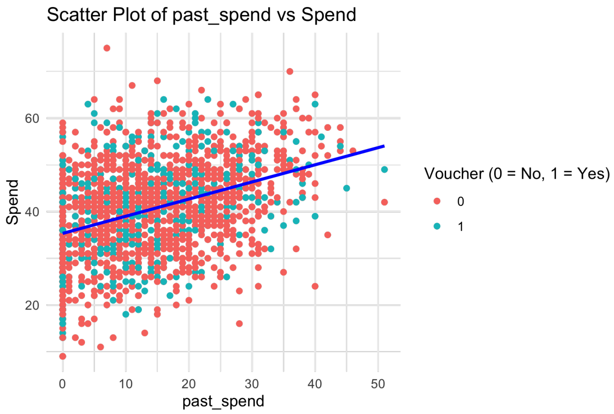
Scatter plots of numerical variables (`age`, `past_spend`, and `time_web`) against `spend` confirmed linear relationships. Categorical variables (`voucher`, `ad_channel`) were tested using jittered scatter plots for increased differentiability between points.



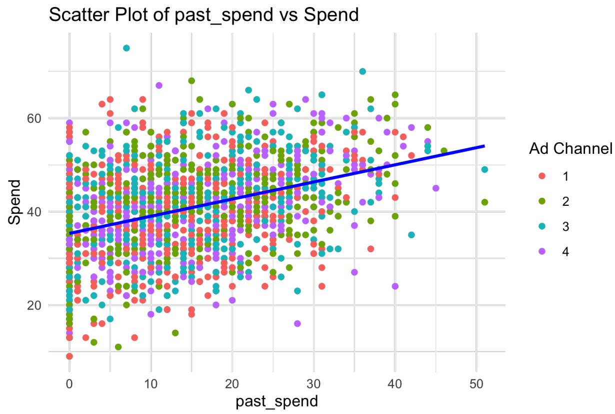
(a) Age vs Spend Scatter with Voucher

(b) Age vs Spend Scatter with AD Channel

Figure 9: Scatter Plots for Age vs Spend

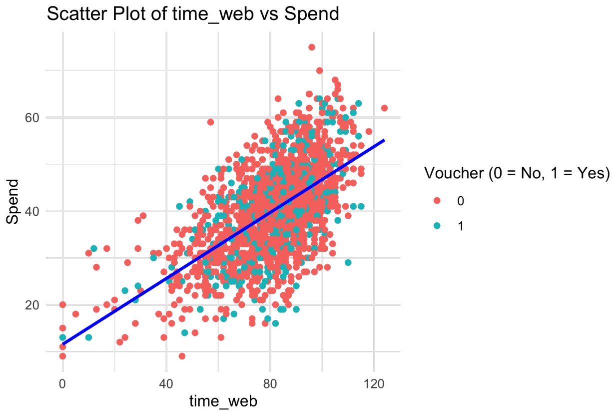


(a) Past Spend vs Spend Scatter with Voucher

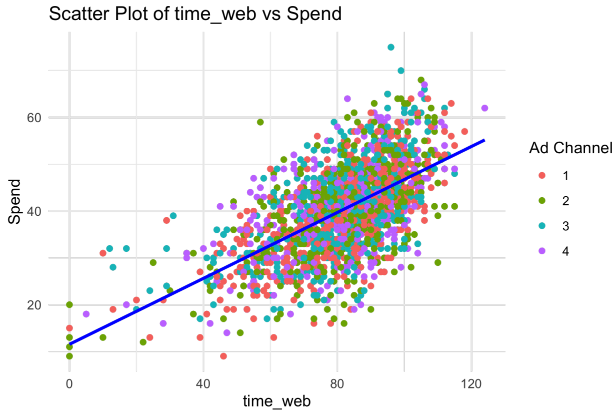


(b) Past Spend vs Spend Scatter with AD Channel

Figure 10: Scatter Plots for Past Spend vs Spend

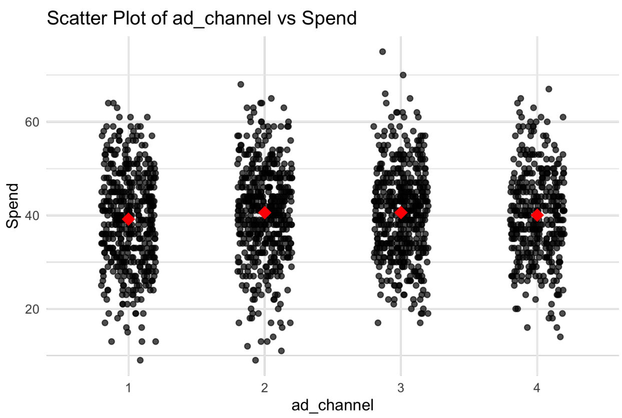


(a) Time Web vs Spend Scatter with Voucher

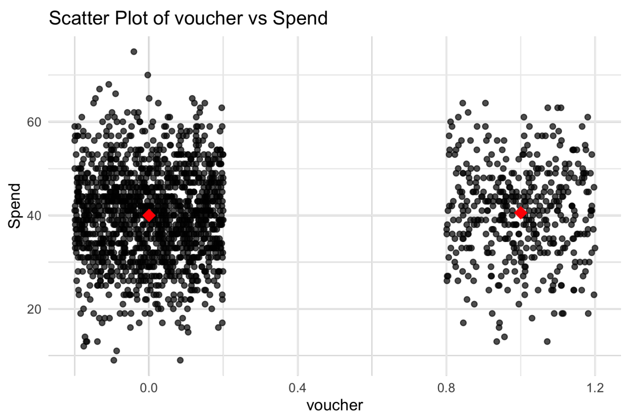


(b) Time Web vs Spend Scatter with AD Channel

Figure 11: Scatter Plots for Time Web vs Spend



(a) AD Channel vs Spend Scatter with Voucher



(b) Voucher vs Spend Scatter

Figure 12: Scatter Plots for AD channel and Voucher vs Spend

4.2 Homoscedasticity

Scatter plots showed consistent variance in `spend` across all numerical variables, with no evidence of increasing or decreasing variance (i.e., no "funnel shape"). For categorical variables, jittered plots indicated consistent variance within categories.

4.3 Independence of Errors

Residuals showed no discernible patterns, supporting the independence of errors.

4.4 Conclusion

The data meets all assumptions for linear regression, validating its use as a viable modelling approach.

5 Evaluation

The model was trained on the clean data and used to evaluate the explanatory value each feature variable had on spending.

5.1 Key Predictors

5.1.1 Age

- **Estimate:** 1.48
- **Significance:** Highly significant ($p < 0.001$).
- **Interpretation:** For every additional year of customer age, their spending increases by an average of £1.48. Age is the strongest predictor, indicating that older customers tend to spend more on the platform.

5.1.2 Time Spent on Website (`time_web`)

- **Estimate:** 0.34
- **Significance:** Highly significant ($p < 0.001$).
- **Interpretation:** Each additional second spent on the website is associated with an average increase in spending of £0.34. This highlights the importance of keeping customers engaged on the platform to drive spending.

5.1.3 Past Spending (`past_spend`)

- **Estimate:** 0.36
- **Significance:** Highly significant ($p < 0.001$).
- **Interpretation:** Customers who have spent more in the past are likely to spend more in the future, with every additional £1 in past spending contributing £0.36 to current spending. This underscores the importance of loyalty and retention strategies.

5.2 Weak or Insignificant Predictors

5.2.1 Voucher

- **Estimate:** 0.01
- **Significance:** Not significant ($p = 0.96$).
- **Interpretation:** The minimal and statistically insignificant effect suggests that offering vouchers does not strongly influence customer spending. This could indicate the need to reevaluate the design or targeting of voucher campaigns.

5.2.2 Advertisement Channel

- **Ad_Channel2:**
  - **Estimate:** 0.25
  - **Significance:** Not significant ($p = 0.27$).
  - **Interpretation:** Customers using this channel spend £0.25 more than those in the reference channel (`ad_channel1`), but the effect is not statistically significant.
- **Ad_Channel3:**
  - **Estimate:** 0.40
  - **Significance:** Weak significance ($p = 0.08$).
  - **Interpretation:** Customers in this channel spend £0.40 more than those in `ad_channel1`. Although weakly significant, this could indicate some potential differences in spending across advertisement channels worth further exploration.
- **Ad_Channel4:**
  - **Estimate:** 0.08
  - **Significance:** Not significant ($p = 0.73$).
  - **Interpretation:** No meaningful impact on spending compared to `ad_channel1`.

5.3 Model Performance

- **Adjusted R^2:** 0.8664
- **Residual Standard Error:** 3.59
- **F-statistic:** 1738 ($p < 0.001$)

The adjusted R^2 value indicates that approximately 86.6% of the variation in `spend` is explained by the independent variables, suggesting a strong model fit. The statistically significant F-statistic confirms that the model provides meaningful explanatory insights.

5.4 Predictive Modelling

The dataset was then split into training (80%) and testing (20%) sets to ensure robust evaluation. The model's performance is as follows: **RMSE:** 3.558, indicating a low average error in predictions.

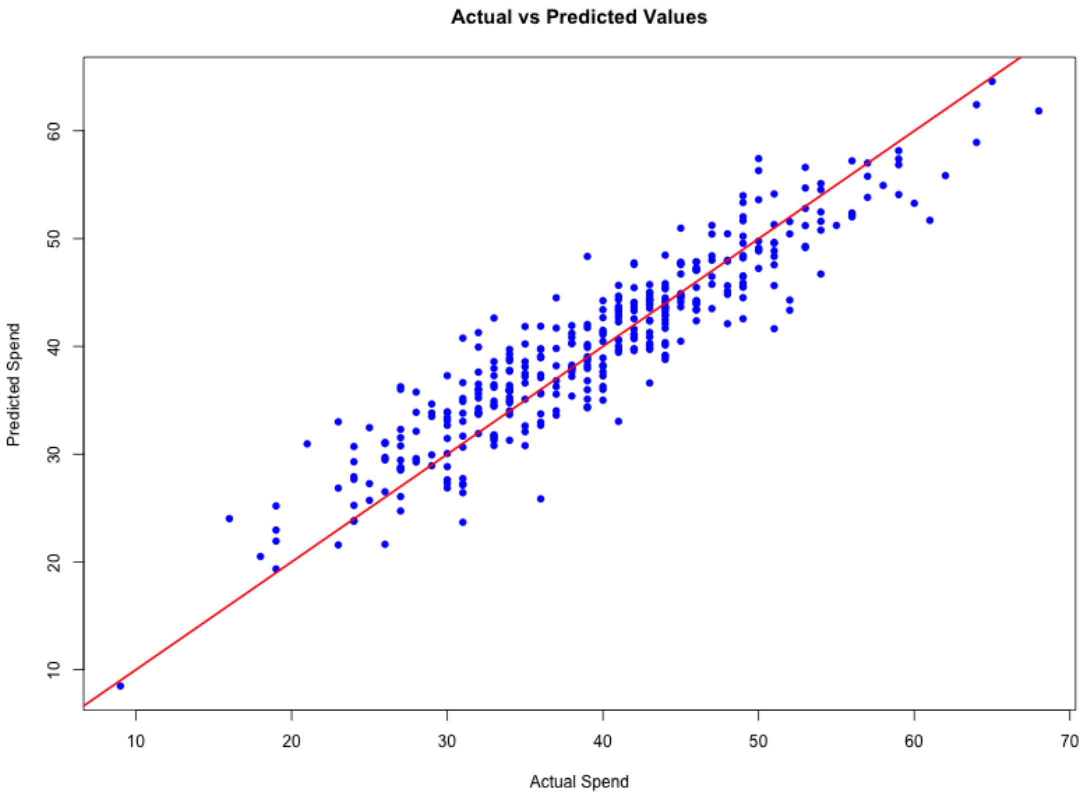


Figure 13: Actual values vs Predicted Values

The actual vs predicted scatter plot (Figure 13) shows strong alignment along the diagonal, demonstrating the model's reliability across spending values.

5.5 Interpretation and Application

The model effectively captures spending variation and offers strong predictive power for operational use. Retrained on the full dataset, the model is prepared to forecast spending for new customer data, with the predictions available in the accompanying table.

6 Actionable Insights

The analysis highlights three key drivers of customer spending: `age`, `time_web`, and `past_spend`. To maximise revenue, efforts should focus on:

- Engaging older customers.
- Fostering loyalty among high spenders.
- Enhancing website engagement through features like personalised recommendations and interactive content.

6.1 Advertisement Channels

- **Search Engine Advertising (`Ad_Channel3`):** Shows the strongest potential among advertisement channels, warranting resource prioritisation and further optimisation through A/B testing.
- **Leaflets (`Ad_Channel1`), Social Media (`Ad_Channel2`), and Influencer Partnerships (`Ad_Channel4`):** Exhibit minimal impact, suggesting these strategies should be re-evaluated.

6.2 Vouchers

Vouchers show limited influence on spending, indicating a need to reassess their structure and targeting.

Order	**Prediction**
1	17.03519
2	45.18834
3	46.31305
4	46.48792
5	34.88865
6	44.41080
7	47.13945
8	54.43254
9	50.98928
10	28.79869
11	49.44404
12	24.39662
13	37.54445
14	40.60553
15	37.21583
16	49.23343
17	45.76117
18	48.68407
19	32.12940
20	55.65403

Table 4: Predictions for customer spending for each order.

Question 2

7 Introduction

This report provides an analysis and recommendation for the selection of two autonomous robot prototypes to be included in the upcoming large-scale trial in Leeds. Due to resource and cost constraints, one prototype will support the primary business model of scaling operations and maximising deliveries of large products, while the other will align with the alternative technology-focused strategy prioritising battery capacity, cost, and reliability.

Using decision science methodologies, I have evaluated the seven prototypes listed in the `Robot_Info.csv` file. For the primary strategy, the rankings from the `Management_Priority.xlsx` file were applied, and for the alternative strategy, rankings were inferred based on the strategic focus provided by the management team.

The analysis identifies the prototypes best aligned with each strategy, providing clear recommendations to guide the trial phase and support future decision-making.

8 Business Understanding

As part of the trial, management has outlined two distinct strategies for evaluating the robot prototypes:

1. **Primary Business Model**: Focused on scaling operations and maximising deliveries of large products, this strategy prioritises attributes such as carrying capacity, cost efficiency, speed, and mobility to meet the company's profitability and customer satisfaction goals.

2. **Alternative Strategy**: Focused on technology sales, this strategy emphasises battery capacity, cost per unit, and reliability to maximise the value of intellectual property.

This report provides recommendations for one robot prototype for each strategy to ensure alignment with the company's strategic objectives and to support successful trials.

9 Data Understanding

The `Robot_Info.csv` file provides detailed information on seven robot prototypes, with each evaluated across several attributes that are critical for the company's strategic goals. The attributes and their importance are ranked for each strategy as follows:

9.1 Primary Business Strategy Rankings

1. Carrying Capacity
2. Cost per Unit
3. Mobility and Speed (equal ranking)
4. Battery Size
5. Reliability
6. Aesthetic

9.2 Alternative Business Strategy Rankings

1. Battery Capacity
2. Reliability
3. Cost per Unit
4. Carrying Capacity
5. Speed and Mobility (equal ranking)
6. Aesthetic

These rankings form the basis for calculating attribute weights and evaluating robot performance in subsequent stages.

10 Data Preparation

To prepare the data for analysis, the following steps were undertaken:

10.1 Rank Sum Method

The rank sum method was applied to convert the rankings for each business strategy into weights. The formula for calculating the weight of an attribute is:

$$\text{Weight} = \frac{n - \text{Rank} + 1}{\sum_{i=1}^{n} i}$$

Where:

- n is the total number of attributes.
- Rank is the rank of the attribute.
- The denominator is the sum of all ranks: $\sum_{i=1}^{n} i$.

10.1.1 Explanation of Mobility and Speed Ranking

In cases where two attributes share the same rank, the rank sum method averages their ranks. For example:

- For Strategy 1, Mobility and Speed were both ranked 3, so their adjusted rank was 3.5 (the average of ranks 3 and 4).
- For Strategy 2, Mobility and Speed were both ranked 5, so their adjusted rank was 5.5 (the average of ranks 5 and 6).

This adjustment allows both attributes to share the same weight without disrupting the overall ranking system.

10.1.2 Weights for the Primary Business Strategy

Attribute	**Rank**	**Weight Formula**	**Weight**
Carrying Capacity	1	$\frac{7}{28}$	0.250
Cost per Unit	2	$\frac{6}{28}$	0.214
Mobility and Speed	3.5	$\frac{4.5}{28}$	0.161
Battery Size	4	$\frac{3}{28}$	0.107
Reliability	5	$\frac{2}{28}$	0.071
Aesthetic	6	$\frac{1}{28}$	0.036

Table 5: Weights for the Primary Business Strategy

10.1.3 Weights for the Alternative Business Strategy

Attribute	**Rank**	**Weight Formula**	**Weight**
Battery Capacity	1	$\frac{7}{28}$	0.250
Reliability	2	$\frac{6}{28}$	0.214
Cost per Unit	3	$\frac{5}{28}$	0.179
Carrying Capacity	4	$\frac{4}{28}$	0.143
Speed and Mobility	5.5	$\frac{2.5}{28}$	0.089
Aesthetic	6	$\frac{1}{28}$	0.036

Table 6: Weights for the Alternative Business Strategy

10.2 Transforming Cost per Unit

To align *Cost per Unit* with the maximisation objective, its values were inverted using the reciprocal:

$$\text{Transformed Value} = \frac{1}{\text{Cost per Unit}}$$

This ensures that all attributes contribute positively to the utility score.

10.3 Normalisation

Each attribute was normalised to account for the varying scales of measurement. This was achieved by dividing each value by the sum of all values in its attribute column:

$$\text{Normalised Value} = \frac{\text{Value}}{\sum(\text{All Values in Attribute})}$$

This normalisation approach ensures that the sum of normalised values for each attribute equals 1, making the attributes directly comparable.

11 Modelling

The Weighted Sum Method (WSM) was employed to calculate an expected utility score for each robot prototype. This method evaluates each robot by summing the weighted contributions of its attributes, with the highest score indicating the optimal choice.

11.1 Score Calculation

The formula for calculating the (`Utility`) score for each robot is:

$$\text{Utility Score} = \sum_{i=1}^{n}(\text{Weight}_i \times \text{Normalised Score}_i)$$

Where:

- n is the total number of attributes.
- Weight_i is the weight assigned to attribute i, derived from the Rank Sum Method.
- $\text{Normalised Score}_i$ is the normalised value of the robot's score for attribute i.

Each robot's utility score is computed by summing the products of its normalised scores and the corresponding attribute weights. This approach ensures that the utility score reflects the relative importance of each attribute as defined by the business strategy.

12 Comparison of Robots

Graphs were generated to visualise the utility scores for each robot under the two business strategies. These visualisations provide a clear comparison of the prototypes and how well they align with the strategic priorities of each strategy.

The robot with the highest utility score for each strategy is recommended as the optimal choice to meet the respective business goals.

13 Evaluation

The utility scores for each robot prototype were calculated using the Weighted Sum Method (WSM) and are presented in the accompanying graphs. These visualisations clearly identify the top-performing robots for each strategy:

- **Primary Business Strategy**: The robot with the highest utility score is recommended based on its ability to maximise deliveries of large products.
- **Alternative Business Strategy**: The robot with the highest utility score is recommended for its strong alignment with technological attributes, including battery capacity, reliability, and cost efficiency.

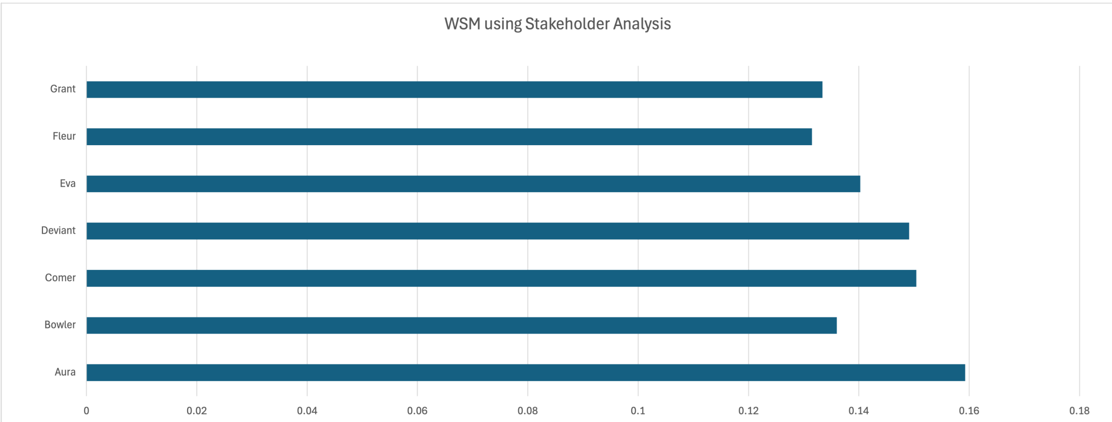


Figure 14: WSM based on Primary Strategy

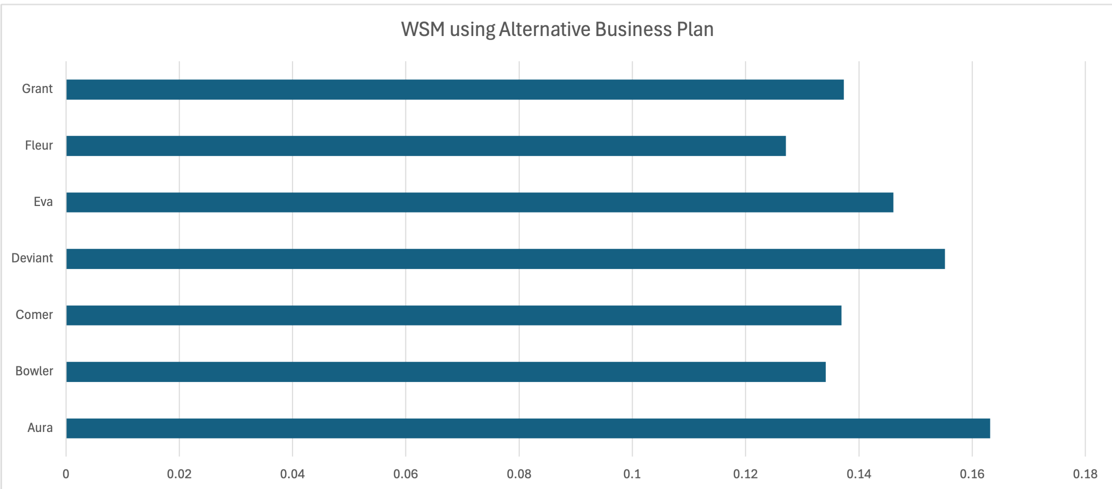


Figure 15: WSM based on Alternate Strategy

13.1 Primary Business Strategy (Stakeholder Analysis)

The graph titled *WSM using Stakeholder Analysis* illustrates the utility scores for each robot under the primary business strategy. Based on the graph:

- Aura achieved the highest utility score, making it the top recommendation for this strategy.
- Deviant and Bowler followed closely but performed significantly lower than Aura.
- Aura's performance is driven by its strong scores in **Carrying Capacity**, **Cost per Unit**, **Mobility** and **Speed**, which align with the strategy's focus on maximising deliveries of large products.

13.2 Alternative Business Strategy

The graph titled *WSM using Alternative Business Plan* shows the utility scores for each robot under the alternative strategy. Key observations include:

- Aura also achieved the highest utility score in this strategy, closely followed by Eva.

- Aura's strong performance in **Battery Capacity** and **Reliability**, coupled with a competitive **Cost per Unit**, aligns well with the alternative strategy's emphasis on technological attributes.
- While Eva was a strong contender, its slightly lower scores in **Carrying Capacity** and **Speed** contributed to its second-place ranking.

13.3 Recommendations

- **Primary Business Strategy**: Aura is recommended for this strategy due to its dominant utility score, demonstrating alignment with the priorities of maximising scalability and efficiency in large-product deliveries.
- **Alternative Business Strategy**: While Aura is also the top performer for this strategy, Eva could be considered a viable alternative if diversification of prototypes is deemed beneficial. Eva's strengths in **Battery Capacity** and **Reliability** directly address the technological focus of this strategy.

14 Sensitivity Analysis for Business Strategy 2

The table below explores the impact of varying the rankings for the top three attributes (Battery Size, Cost per Unit, Reliability) on the robot selection for the alternative strategy. The Weighted Sum Method (WSM) was recalculated for each permutation of the rankings:

Permutation	Rank 1	Rank 2	Rank 3	Top Robot
1	Battery Size	Cost per Unit	Reliability	Aura
2	Battery Size	Reliability	Cost per Unit	Aura
3	Cost per Unit	Battery Size	Reliability	Aura
4	Cost per Unit	Reliability	Battery Size	Aura
5	Reliability	Battery Size	Cost per Unit	Aura
6	Reliability	Cost per Unit	Battery Size	Aura

Table 7: Sensitivity Analysis for Business Strategy 2

14.1 Interpretation

The results indicate that **Aura** consistently achieved the highest utility score across all permutations of the top three rankings. This demonstrates the robustness of the recommendation for the alternative strategy.

14.2 Note on Business Strategy 1

Sensitivity analysis was not conducted for Business Strategy 1, as the rankings were predetermined by the management team. These fixed rankings reflect stakeholder priorities and ensure alignment with the strategy's objectives, making further exploration unnecessary.

15 Limitations

1. **Subjective Rankings for the Alternative Strategy:**
   - The rankings for Business Strategy 2 were inferred based on a broad description of priorities, which introduces a degree of subjectivity. While the analysis demonstrates robustness through sensitivity testing, further refinement of these rankings based on stakeholder feedback could improve accuracy.
2. **Potential Bias in Subjective Scores:**
   - Attributes such as Mobility, Aesthetic, and Reliability are based on subjective judgements. Any bias in these evaluations could affect the utility scores and the final recommendations.
3. **Exclusion of Real-World Data:**
   - The analysis is based solely on the provided prototype data (`Robot_Info.csv`) and does not account for real-world performance metrics, such as maintenance costs, unforeseen failures, or customer feedback during the trial phase.