Leeds University
Business School

UNIVERSITY OF LEEDS

Assessed Coursework Coversheet

For use with *individual* assessed work

Student ID Number:	2	0	1	9	1	1	0	4	8
Module Code:	LUBS5308M								
Module Title:	BUSINESS ANALYTICS AND DECISION SCIENCE								
Module Leader:	Aritad Choicharoon								
Declared Word Count:	3076								

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 https://secretariat.leeds.ac.uk/wp-content/uploads/sites/109/2022/12/academic_integrity.pdf	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 heard yet whether it is granted. I am submitting this paper in the knowledge that I may request to submit a later version, if extension granted. Markers should be aware that this may not be my final version of the assignment. (Please indicate length of extension requested too, so we know when to expect updated submissions – delete two leaving the correct one visible)	NO
I am aware of the Generative AI category for this assignment (delete two, leaving the correct one visible), 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, down to a minimum of the bare pass mark (if pass standard).

Part 1: Business Analytics: Predicting Potential Return

1. Executive Summary

This research offers insights into client expenditure patterns and introduces a forecast model to aid Physical Sound's strategic objective. An examination of July 2024 order data reveals critical determinants affecting expenditure, including website engagement duration, historical spending patterns, coupon utilization and advertising channel interaction, providing pragmatic recommendations to improve marketing tactics and customer engagement.

A predictive model was created to predict customer's spending and is assessed using metrics like R^2 and Mean Absolute Error, with high accuracy. The model results then was used to forecast the spend for additional 20 customers, facilitating customer segmentation and bolstering expansion strategy. All these results and findings are totally data driven and decisions or recommendations based are based on the analysis done on the data and taking into consideration the business requirement to make sure better business performance and to manage the resources efficiently as well.

2. Introduction

The swift expansion of e-commerce has rendered the understanding of customer behavior and the optimization of expenditure which is essential for any online business success. Leveraging a data-driven approach is very crucial for Physical Sound to increase and enhance the customer engagement and stimulate growth. The report aims to address the two principal business problems raised by the company's management team:

1. **What factors influence customer spending on the websites?** Recognizing these elements will help enhancing the online shopping experience for a customer and will help facilitating tailored marketing strategies.

2. **How can customer spending be predicted?** An effective prediction model will allow the organization to assess the customer segments and facilitate future expansion decisions.

The report uses and combines exploratory data analysis (EDA) and prediction model techniques to answer these business questions. Using the results from the model and EDA, findings provide actionable insights which can help company to understand the spending behavior of customer and allow the management to estimate the spending for additional customers.

3. Data & Methodology

a. Data Sources

The analysis is based on two datasets provided by Physical Sound:

1. **Order_july24.csv:** This dataset contains comprehensive information regarding customer orders placed on the company's website during July 2024.
2. **new_customer24.csv:** This dataset contains information about 20 new customers for whom spending predictions are required. It includes variables like those in Order_july24 dataset, excluding the actual spending value.

b. Data Cleaning and Preparations

To ensure that the analysis is reliable and accurate, the datasets underwent a data cleaning and preparation process. Below are the detailed steps for the same with analysis results.

1. **Handling the missing values:**

The dataset consists of some missing values for different row combinations, below is the summary for the same.

Column Name	Missing Values Before Handling	Missing Values After Handling
past_spend	25	0
age	14	0
ad_channel	16	0
time_web	21	0
voucher	24	0
spend	27	0

Table 1.1 Missing Values in Order Data

For numerical values such as age, time_web, past_spend and spend were replaced with median of the respective column to prevent skewness induced by the outlier values.

For categorical values such as ad_channel and voucher were replaced with the mode (most frequent category) to maintain consistency in the data.

2. **Outlier Detection and Handling:**

Outliers in the numerical variables were identified using boxplots. Extreme values were checked and capped at a certain defined threshold. Below are the boxplots representing both before the removal of outliers and after removal of outliers for each numerical variable.

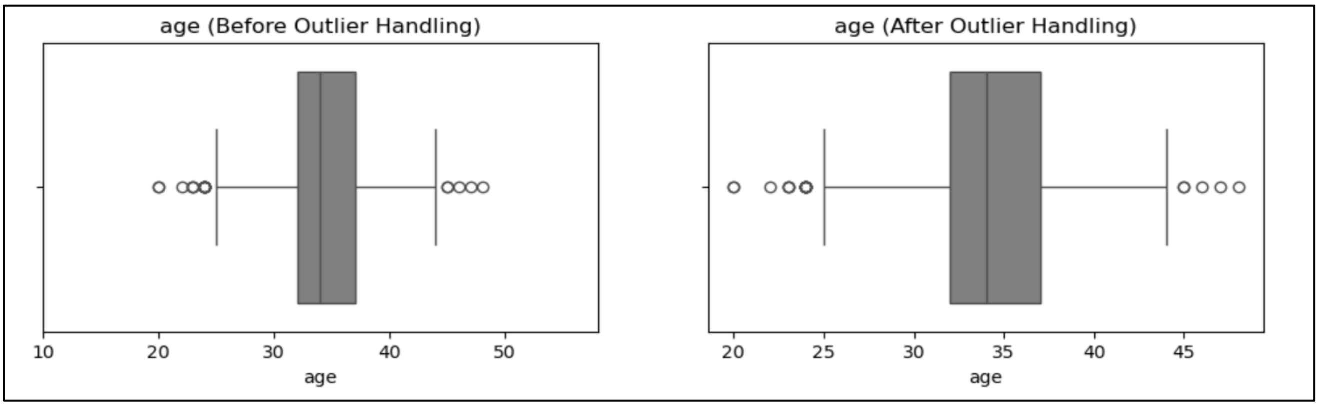

Fig 1.1.a. Box Plot Age (before and after outlier handling)

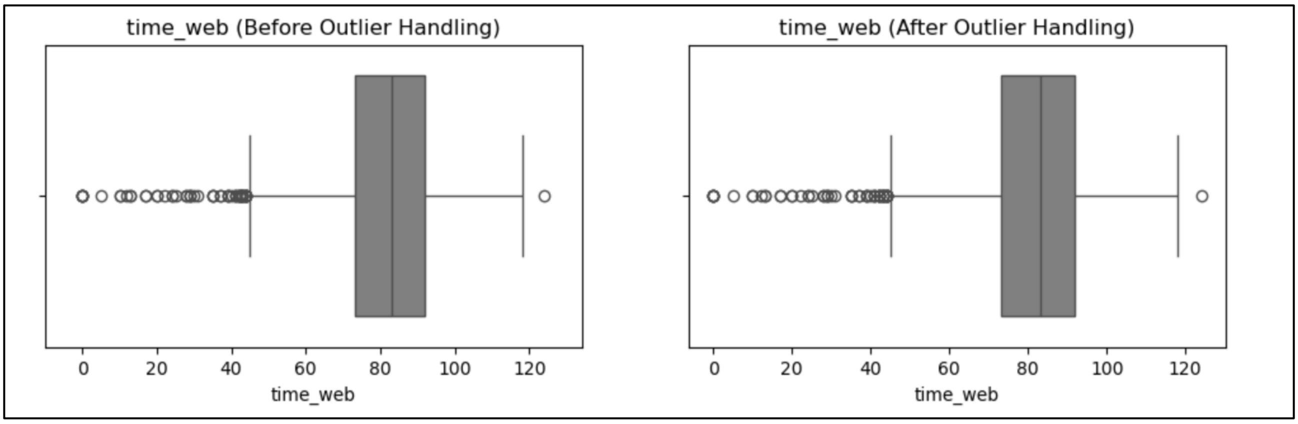

Fig 1.1.b. Box Plot Time Spent on Website (before and after outlier handling)

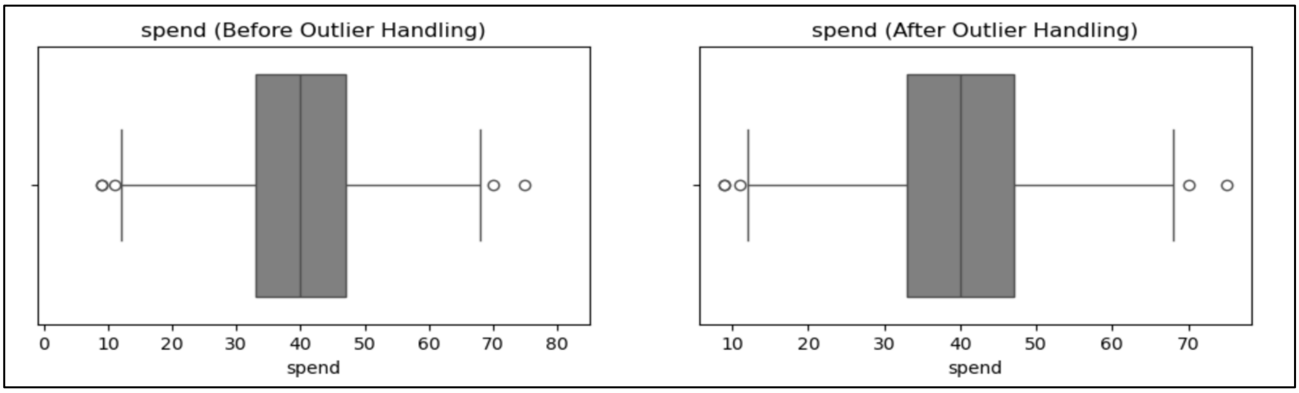

Fig 1.1.c. Box Plot Spend (before and after outlier handling)

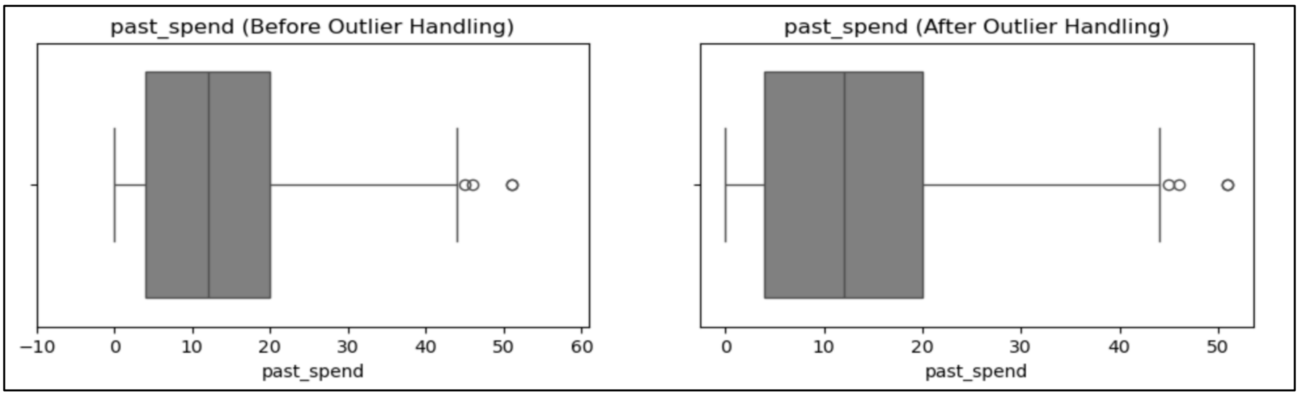

Fig 1.1.d. Box Plot Past Spend (before and after outlier handling)

Additionally, we also checked the data distribution before and after outlier removal and as per the curve (shown below) it is clearly visible that removal of outliers skews the distribution and hence outlier removal is not an efficient way to move ahead with this dataset.

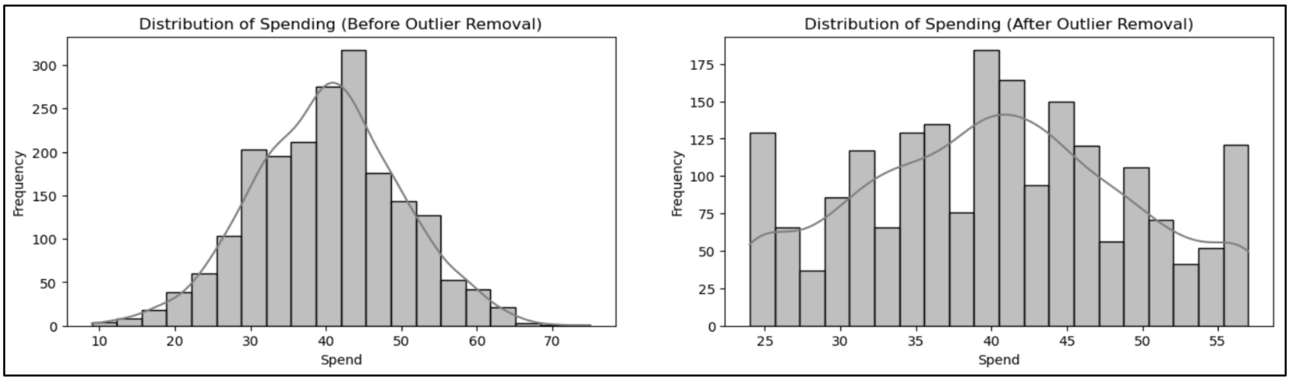

Fig 1.2. Distribution of Spend (before and after outlier handling)

3. **Data Type Adjustments:**

Categorical attributes present in data like ad_channel and voucher, were converted to factor data type to make sure it is compatible with model to be used. For this we have used “*label encoder”* function*.*

Numerical variables were cleaned and standardized wherever necessary to improve the performance of the regression model.

4. **Correlation Metric:-**

The correlation matrix is used to highlights primary factors influencing customer spending. Based on the chart below “Age”(0.62) and “time_web”(0.59) has strong positive correlation with expenditure, highlighting that older customers and the customers who spend more time on the website are more likely to spend more. Another matrix “Past_spend”(0.39) also shows a moderate correlation indicating a customer’s spending habits is also a pertinent reason for spending . Rest of the features like “voucher” and “ad_channel” have weak correlations with spend, showing very minimal impact on the expenditure. These findings underscore the significance of demographic and behavioral parameters, notably age and website involvement, as key drivers of customer spending. Below shared is the correlation matrix for the shared data.

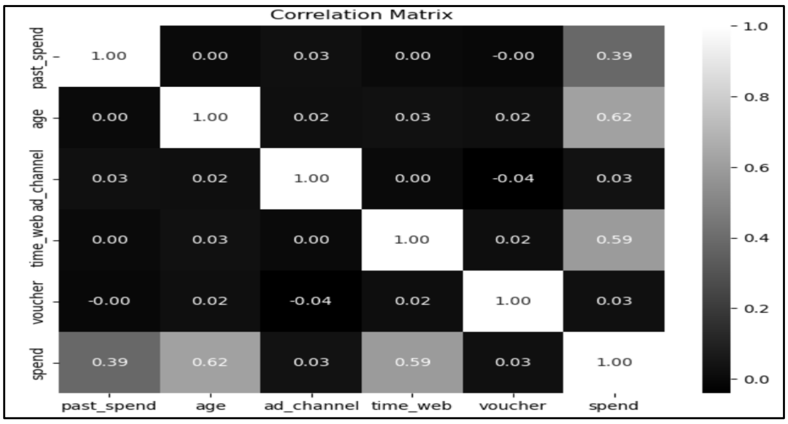

Fig 1.3 Correlation Plot

5. **Data Validation:**

The data was scrutinized for any inconsistency such as duplicity in the records or any invalid rows (like negative spending or ages below zero).

The dataset is refined with data cleaning and data preparation steps which results in high quality input suitable for robust analysis and modeling.

c. Methodology & Analysis Result

1. **Feature Validation:**

Once the features are selected using the correlation analysis (Age, past_spend and time_web). These features are then analyzed to check if these have any linear relationship with spend before modeling. Below are the scatter plots to show the same.

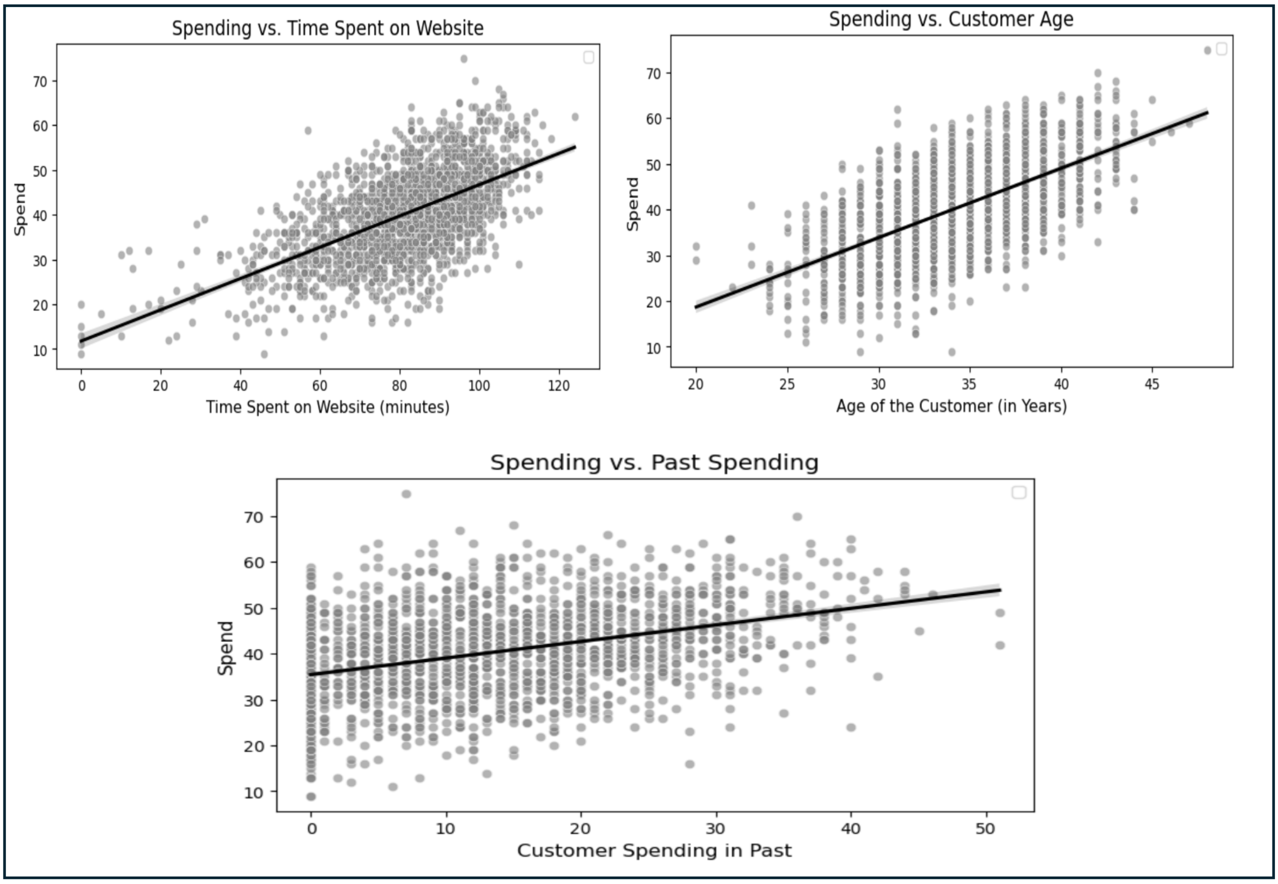

Fig 1.4 Spend vs Selected Feature (To Check Linear Relation)

Based on the all the three trends, the graphs show positive linear relationship with spend. Hence, improving website interaction while focusing on customers with past spending habits and targeting a specific age group can drive the spend higher.

2. **Train – Test split:**

The dataset was split into a training set (80%) for building the model and test set (20%) for validating the model's accuracy and generalizability.

3. **Predictive Modeling:**

The analysis focuses to address two primary business problems:

1. To identify the factors influencing customer spending on the company's website.
2. To build and calculate a prediction model for customer expenditure and use this to predict spend for new customers.

The linear regression model was chosen to predict the customer spending. The target variable as spending and features included past_spend, age, time_web.

Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-square (R^2) were calculated to assess the model's accuracy and performance. Spending predictions for 20 additional customers were generated and saved for evaluation.

```
Model Evaluation Metrics:
Mean Absolute Error (MAE): 2.76
Root Mean Squared Error (RMSE): 3.60
R-squared (R2): 0.86
```

In addition to this, have also checked the R^2 for both test and train split given below.

```
Model Performance:
Training R-squared (R²): 0.85
Testing R-squared (R²): 0.86
```

Also, below is the trend chart for predicted vs actual based on the prediction model. The trend shows that the prediction is close and more accurate to the training data.

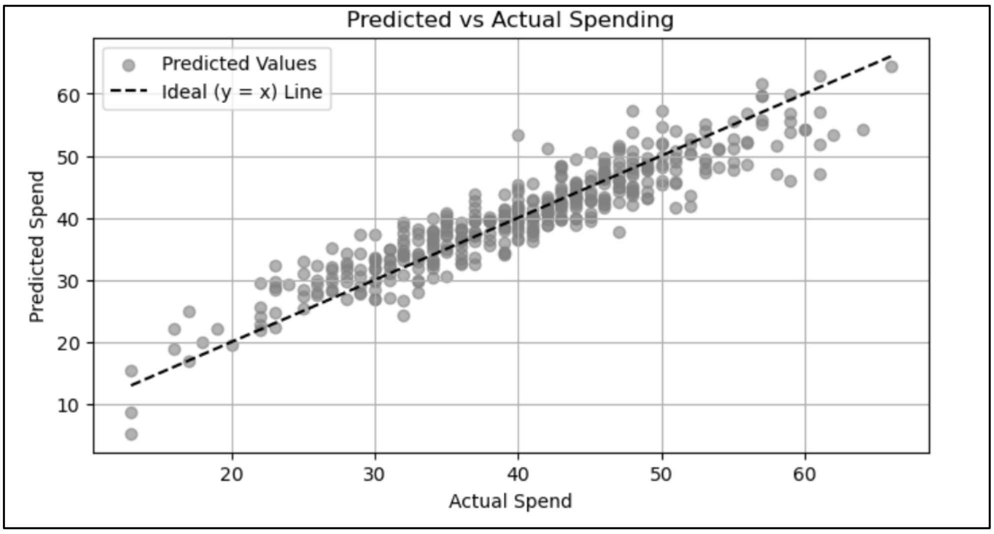

Fig 1.5 Actual vs Predicted Spend Plot

4. Discussion and Business Implications

The analysis and predictive model results provide valuable insights into customer spending behavior, identifying key factor such as age, time spent on the website, and past spendings as critical predictors. These insights enable the company to better understand customer profiles and implement targeted strategies. High spenders can be prioritized with exclusive offers, while low and medium spenders can be engaged with incentives and personalized campaigns to drive higher spending.

The business impact of these finding is significant. By utilizing the model to segment customers, the company can optimize resource allocation, reduce marketing inefficiencies, and improve ROI. The ability to predict spending for new customer also supports market expansion by focusing on high-value segments, allowing the company to scale efficiently.

Incorporating these insights into decision-making strengthens the company's ability to adapt to changing market conditions, improve customer retention, and drive sustainable revenue growth. This data driven approach not only ensures the operational efficiency but also enhances the company's competitive edge by fostering customer-centric strategies.

5. Recommendations & Conclusion

a. Recommendations:

Based on the analysis and model results, it is recommended that the company prioritize customers who exhibits high spending potential. Customers with high past spending, significant time spent on the website, and belonging to specific demographic groups should be targeted with tailored marketing campaigns and exclusive offers to maximize profitability. Additionally, introducing loyalty programs or personalized incentives can encourage repeat purchases and strengthens customer relationships.

Furthermore, the company should continuously refine the predictive model by incorporating new data from ongoing operations as well as including the competitor's data as well, as this will give the management confidence about their product and provide better results including the current market scenario as well. In future expansion, the predictive model can be used to evaluate customer spending in different regions or segments, allowing the company to make informed decisions regarding resource allocation and scaling operations.

b. Conclusion:

The model has provided actionable insights into customer behavior and spending patterns, highlighting the key drivers of high spending. The model successfully identifies factors such as past spending, age, and time on the websites as critical indicators of customer spending, enabling effective customer segmentation and targeted strategies. The predictions for new customers further validated the model's applicability, demonstrating its potential to support the company's growth and market expansion plans.

<u>Part 2: Business Decision-Making: Trial of Delivery Robot</u>

1. Introduction & Business Understanding

The recent growth in demand for bringing automation in the logistics using autonomous robots has prompted many companies to explore the innovative options to increase the efficiency and profit. In the current use case, the company wants to conduct a trial to evaluate the performance of the autonomous robots shared in the data (might be available in the market) based on certain parameter shared in the dataset in real life situations.
Using very limited resources, company must strategically decide on two robots for two different trials. One selection is to align with the primary business requirement (prioritizing large scale operations and delivering maximum number of deliverables). Other selection is to focus on an alternative to sell the intellectual technology, emphasizing about the battery, cost, capacity and reliability.
Additionally, to support the decision, company must have a detailed analysis addressing the decision-making problems, addressing the business requirements and justify the recommendations based on decision science methods.

2. Data & Methodology

a. Data Sources

There are two main datasets mentioned below:

Robo_Info.csv:- This file contains the detailed information of each of the prototypes as assessed in the preliminary stage by the design and development team.

Management Priority:- This dataset contains stakeholder weighted priorities of each criterion based on their primary business model.

b. Data Cleaning and Preparations

As the dataset only has very limited records. There was not much data processing required (at least in the Robot_Info.csv file).

For Management_Priority dataset, had to create another column as priority ranking (table 1.0) to rank the priority of each criterion based on the statements shared by the stakeholders, which can be used to create weights for further analysis.

Additionally, none of the datasets had any missing values in them so no further process required to handle the missing values.

c. Analysis Methodology

1. For Primary Strategy:-

For primary strategy, objective is to maximize the operational efficiency. To attain the objective, will be using the rankings created in the Management_Priority using the information shared by stakeholders in the file to create the weighted factors to score and rank the robots. Explaining the process for application of the weighted factors and the criteria for the ranking of the robots for prioritization. This method ensures that the robot tops the business requirement like having high delivery volume and is reliable is prioritized, providing a transparent basis for selecting the most suitable robot for the trial.

2. For Alternative Strategy:-

For the alternative strategy, objective is to get much profit by selling the robots. Robot prototypes are analyzed based on the criteria shared in the problem statement i.e., battery capacity, cost and capacity. As these features defines the value of the robot in the market and talks about the whether it is good in the practical use.

To complete the objective, method we can use is the normalization of the standardized values of the features mentioned for all the prototypes. Once the standardization is done, each robot is then scored by combining the normalized values of the features selected for this problem statement. This approach will allow us to identify the robot which can offer the best combination of technical aspect of the automation at the same time is cost-effective as well, which aligns the company's goal to commercialize the prototype and get profit from it.

3. Visual Representation:-

To describe the rankings (prioritization based on rankings) and analysis results, below are some tables and plots used to summarize the results and methods mentioned in the above steps for both primary and alternative strategies.

I. *Weights distribution Chart:-*

Feature	**Ranking_Description**	**Ranking**
Carrying Capacity	This is the most important consideration according to the management team and this is clearly favoured over all other criteria.	1
Battery Size	The battery size is not as important as speed or mobility but more important than reliability.	3
Speed	After careful deliberation, the speed is rated 3 out of 5 in term of importance. They believe that it is an important criterion but not as important as some others and that this would likely improve with better tech.	4
Mobility	The mobility is as important as speed.	4
Aesthetic	The aesthetic is a factor but should be the lowest in the consideration.	6
Cost Per Unit	The cost is more important than any other criteria except for the capacity. One of the management team considered it to be at least 20% of total consideration amongst all criteria.	2
Reliability	This is the 2nd least important criterion according to most of the management team.	5

Table 2.1 Ranking table for Primary Strategy

A pie chart is created to visualize the weights assigned to each of the required features present in the dataset. Started with providing rank to each criterion based on the ranking description present in the Management_Priority file. Once the ranking is done using min-max criteria created scores between 1-5 (taking 1 as lowest priority and 5 as highest priority).

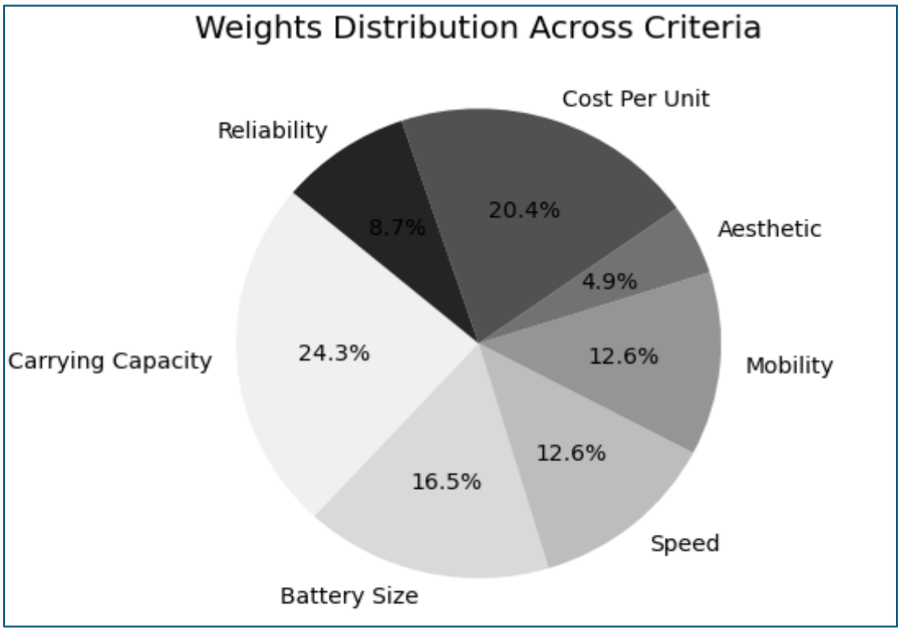

Fig 2.1: Weight Distribution Across Criterion

II. Ranking table for the Primary Strategy:-

Below present is a comprehensive table to show the weighted score for each robot for the primary business strategy. In the plot, each bar line represents a robot with height based on the weighted factors.

Robot Prototype	Weighted Score	Rank	Carrying Capacity	Battery Size	Speed	Mobility	Aesthetic	Cost Per Unit	Reliability
Aura	0.66	1.00	0.57	0.67	0.30	0.50	0.86	0.91	1.00
Deviant	0.57	2.00	0.14	1.00	1.00	0.75	0.00	0.36	0.90
Eva	0.53	3.00	0.57	0.67	0.00	0.25	0.43	0.82	0.75
Grant	0.51	4.00	1.00	0.50	0.00	0.00	0.57	0.45	0.75
Comer	0.46	5.00	0.43	0.00	0.00	0.75	0.43	1.00	0.45
Bowler	0.45	6.00	0.43	0.50	0.00	0.50	0.43	0.68	0.45
Fleur	0.31	7.00	0.00	0.83	0.00	1.00	1.00	0.00	0.00

Table 2.2: Ranking table for Primary Strategy

The table displays each prototype’s attributes (in columns after transposing the data) alongside their weighted scores respectively along with their overall ranking (for prioritization). This table can help stakeholders to compare the performance of each robot based on their features in parallel and understand the contribution of the features to result in the final ranking.

III. *Bar chart for weighted scores:-*

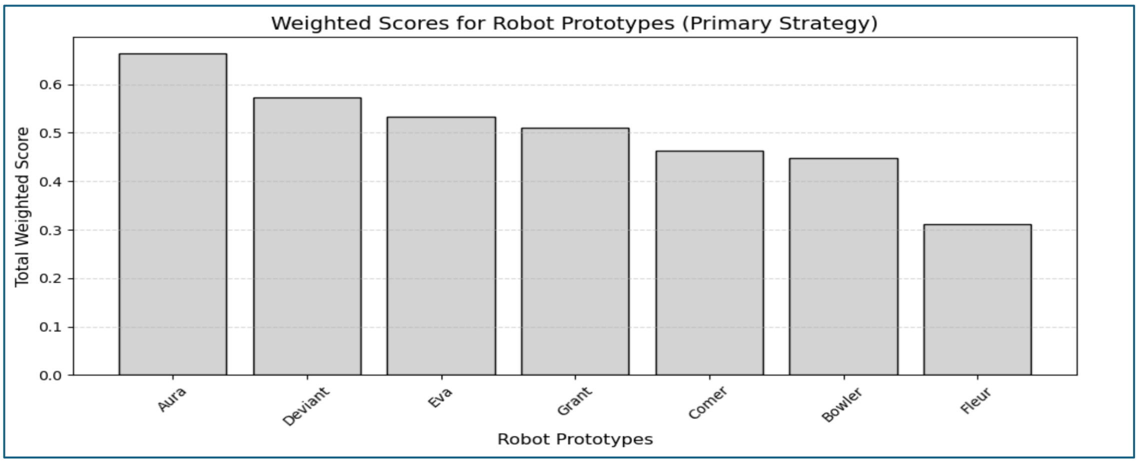

Fig 2.3: Weight Score for Prototypes Based on Priority Ranking

IV. *Normalization Calculation and Result for Alternative strategy:-*

This method is used to convert the values of all the features and bring them to the same scale (generally between 0 to 1). For this case we have used min-max normalization, this is done to make the values comparable, mostly while evaluating feature values with different ranges (like in this case). Formula used is mentioned below-

$$Normalised\ value\ = \frac{Value - \ Min\ Value}{Max\ Value - \ Min\ Value}$$

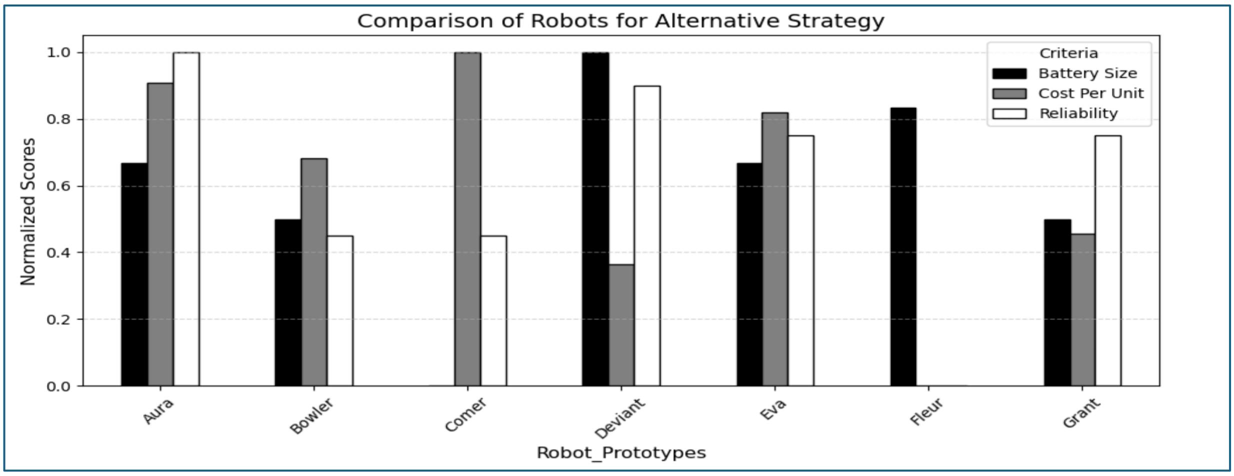

Fig 2.4. Comparison plot for Alternative Strategy

Below present is the table with calculated value based on the formula and using weights (apparently there is no prioritization given for this case so considering same weights for all the attributes i.e., Battery Size, Cost Per Unit, and Reliability)

Robot_Prototype	Battery Size Normalized	Cost Per Unit Normalized	Reliability Normalized	Alternative Score
Aura	0.667	0.909	1.000	**0.859**
Bowler	0.500	0.682	0.450	**0.544**
Comer	0.000	1.000	0.450	**0.483**
Deviant	1.000	0.364	0.900	**0.755**
Eva	0.667	0.818	0.750	**0.745**
Fleur	0.833	0.000	0.000	**0.278**
Grant	0.500	0.455	0.750	**0.568**

Table 2.3. Normalized Score and Total Score for Alternative Strategy

***Additionally, as the cost per unit is a metric which has inverse relation with the cost effectiveness, while normalizing the value for this metric for both the strategies, have taken the inverse impact of the cost to make sure the prototype with low cost has high score.*

$$Normalised\ value\ (\ Cost\ Per\ Unit) = \frac{Max\ Value\ -\ Value}{Max\ Value -\ Min\ Value}$$

4. Recommendation & Conclusion

Below shared is the summary table based on the results from the analysis done as per the business requirement and management priorities

Strategy	Selected Prototy	Key Criteria	Justification
Primary	Aura	carrying capacity, cost per unit, age, reliability, speed, aesthetic, mobility	Aura has highest weighted score amongst all the prototypes based on the priorities. Its high carrying capacity and cheaper cost make it an ideal selection for the operation process.
Primary	Deviant		Deviant has the second highest weighted score after Deviant. As its highest speed and reliability factor amongst all the prototypes
Alternative	Aura	battery size, cost per unit, reliability	Aura outperforms other prototypes on the technical attributes making it most suitable for the commercial usage by selling this prototype.

Table 2.4. Result Summary for Both Strategies

Using the weighted scoring for the primary approach and normalized scoring for the alternate strategy, the analysis effectively assessed the robot prototypes performance in relation to important business factors. The result support the company's two goals of increasing the marketability of its intellectual property and improving operational efficiency.

The organization can increase market positioning, operational results, and resource utilization by implementing Aura for both initiatives. The organization is positioned for success in both operational and commercial endeavors because to strict decision-making process, which guarantees that the suggestions are data driven and in-line with business goals.