

# Summary Report

## Introduction

The Leads Score Assignment involved generating a prediction model that would rank leads in terms of the probability of their conversion to customers. The goal was to analyse the lead dataset, determine the most important factors that would influence conversion of a lead, and come up with a predictive model that would serve as an indicator of the likelihood of any lead that is generated being converted.

## Data Loading and Exploration

The first task of the project started with the uploading of the data set that came with several lead-related parameters such as, Prospect id, Lead number, Lead source, Total time spent on a website and Tags. This phase examining the first few rows of the data, checking for missing values, and understanding the distribution of the variables under consideration that were both numerical and categorical.

## Data Preprocessing

**Handling Missing Values:** Missing values were handled by determining the extent to which values were missing and either eliminating columns with a large portion of missing values or filling in the blanks.

**Encoding Categorical Variables:** Categorical encoding was achieved using one hot encoding for the variables in question to be model-ready.

**Scaling Numerical Features:** The use of StandardScaler resolved the issue of having all numerical features on the same scale by appropriately scaling them.

## Model Building and Evaluation

The activities that resulted in the model building included:

**Splitting Data:** The data comprising the dataset was partitioned as training data and testing data.

**Logistic Regression:** Building an initial logistic regression model to predict the conversion probability.

**Feature Selection:** Using Recursive Feature Elimination (RFE) to select the most relevant features.

**Final Model:** Building a final model with the selected features and evaluating its performance using metrics such as accuracy, precision, recall, and ROC-AUC.

## Key Findings and Learnings

**Top Contributing Variables:** The top variables for lead conversion are Total Time Spent on Website, Lead Source (especially Welinkak Website), and specific tags like Will revert after reading the email, Ringing, and Closed by Horizzon.

**Importance of Engagement Metrics:** Metrics such as Total Time Spent on Website and Page Views Per Visit are highly predictive of lead conversion, emphasizing the importance of lead engagement.

## **Conclusion**

The Leads Score Assignment offered valuable insights into factors influencing lead conversion and the importance of data-driven decision-making in sales. By using logistic regression and feature selection, the model accurately predicted conversion probabilities. These learnings can enhance lead conversion rates and optimize sales strategies.