Brief Summary Report

1. **Approach to proceed with the assignment :**

    a. Understood the problem statement and the requirement of the business.
    b. Understood the business domain by looking at the data and the variable definitions.
    c. Analysed the format of each variable to identify the categorical and numerical variable in the data.
    d. Checked for null values in the data and handled them by imputing the values using mean, mode and median method or dropping the columns/rows based on the percentage of null values in the columns/rows.
    e. Handled the outliers in numerical variable by capping them to a specific limit.
    f. Performed exploratory data analysis to analyse each variable to identify the impact of each variable on the target variable and then identify the important variables that have significant impact on target variable.
    g. Prepare the inferences as part of exploratory data analysis.
    h. Prepared the data for model building:
        a. Convert categorical variable to dummy variables.
        b. Scaling the numerical variables.
        c. Split the data in to training and test set.
    i. Once the data is prepared, then we got into the mixed model building on training data set:
        a. Used Recursive Feature Selection technique to select 15 features.
        b. Then built the first model with these 15 selected features .

    c. Dropped the features that are statistically insignificant based on p-value / variable inflation factor (VIF) to reach the final features that are statistically significant for the logistic regression model

j. Predicted the probabilities on training data set using the regressor that is trained on training data set.
k. Evaluated the final model by calculating metrics like sensitivity, specificity , precision and recall for different probabilities using confusion matrix.
l. Plotted the ROC curve to see the trade of between sensitivity and specificity.
m. Found the optimal cut off point using the probabilities.
n. Now made the prediction on test data set using following steps:
    a. Select the features in the test data set based on what's been identified as statistically significant features from the final model .
    b. Scale the non-categorical features using standard scaler
    c. Use transform method on the test data set
    d. Predict the probabilities and label coverted/non-converted in terms of binary form using the optimal cut off point on the test data set
o. Evaluated the sensitivity and specificity for the test data set using the confusion matrix
p. Calculate the score for each prospect id using the formula : probability *100 .

2. Learnings gathered :

a. Understood that how to reach out to business problem and finalize the solution approach
b. Learnt the practical implementation of Exploratory Data Analysis and make inferences.
c. Learnt to identify the important features for model building.
d. Learnt how to prepare the data for training the model.
e. Learnt how we can evaluate our final model to see if it is statistically significant for the data or not.

f. Learnt the different matrix like confusion matrix, specificity, sensitivity etc used in Logistic regression.
g. Learnt how to make short and simplified presentation to present the overall analysis to business.