Summary

Analysis should be performed to find leads with higher probabilities.

The following steps were done:

1. Cleaning data:

Null values were taken care of in this step.

2. EDA:

Exploratory Data Analysis was performed to understand the data and see if any outliers exist in our data. To our observation no outliers were found but we were able to figure out various variables that were of no use to us.

3. Dummy Variable Creation & Standardisation:

Dummy variables were created for all the required categorical features and numeric variables were standardised using StandardScaler.

4. Train-Test split:

The data was divided into Train (70%) and Test (30%) sets respectively.

5. Model Building:

Initially RFE was used to retrieve top 15 relevant variables. Later, variables were removed manually depending based on their VIFs and p-values.

6. Model Evaluation:

A confusion matrix was made. Later on, the optimum cut off value (using ROC curve)

was used to find the accuracy, sensitivity and specificity which came to be around 90%

each.

7. Prediction:

Prediction was done on the test data frame and with an optimum cut off as 0.5 with accuracy,

sensitivity of 90%. and specificity of 96%.

8. Precision – Recall:

This method was also used to recheck and a cut off 0.5 was found with Precision around

93.69% and recall around 90.34% on the test data frame.