Lead Scoring Summary

An education company named X Education want to select the most promising leads using logistic regression model. The main goal of this case study is Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

To build a model we first imported the data and then checked the missing values and then used different imputation techniques to reduce the missing values and checked for outliers and done outlier treatment.
After that done the EDA to analyze the relation between dependent and independent variables.
Then did the Data Preparation -
(Created Dummies for categorical variables) done before model building then split the data into Train and Test data.
Rescale the data using Standard Scalar.
Then built a Stats model and using RFE, VIF and eliminated the unwanted features.
After getting low VIF values we derived the probabilities, Lead Score, Predictions on Train Data and then formed a confusion matrix to calculate the Accuracy, Sensitivity, Specificity and Precision.
After that we made prediction on test data and formed the confusion matrix to compare the Accuracy, Sensitivity, Specificity and Precision of model on test data.
Here we used initially 0.5 cutoff based on Accuracy-
Sensitivity- Specificity curve we changed the cutoff to 0.3(we can use from 0.3 to0.4).

The result received from Test Data:
Accuracy: 86.8%
Sensitivity: 82.3%
Specificity: 92.1%
Precision: 80.8%

The result received from Training Data:
Accuracy: 81.4%
Sensitivity: 83.4%
Specificity: 80.2%
Precision: 70.69%