Summary Report: Lead Scoring Case Study

- Understood the business problem. We went through the data and data dictionary.
- In Jupiter notebook, imported all the required libraries and load the data. Then started data inspection. i.e. shape of the data, null values in each column, and statistical values of numeric columns.
- Dropped the columns. Replaced Yes/No values with 1/0. Imputed the median value in place of the null value. After that checked the higher conversion rate of the categorical variable and identified the outlier variables using EDA.
- Created a dummy variable and dropped the first column. Split the variables in test and train sets. Before building the model standardize all the variables i.e. Total Time Spent on Website. Checked the correlation between variables.
- Built model. Added/removed features from the model and monitored their P-value and VIF. Based on the P value and VIF identified the best model.
- Plotted Roc curve. Find out the optimal cutoff point of the train model. Calculated accuracy, sensitivity, specificity, precision, and recall of the train model. Plotted precision and recall trade-off on train set.
- Plotted Roc curve of test data. Find out the optimal cutoff point of the test model. Calculated accuracy, sensitivity, specificity, precision, and recall of the test model. Plotted precision and recall trade-off on the test set.