Lead Scoring Case Study Solution Summary

- Loading and reading the data
- Data Cleaning
- EDA
- Removing Nulls from data
- Checking the other categorical variables
- Analysing the categorical columns
- Analysing the numerical columns
- Checking for outliers
- Removing the outliers
- Preparing the data for modelling
- Checking for correlation
- Preparing train and test data
- Feature scaling
- Feature Selection using RFE
- Building the Model
- Making predictions on the train set
- Drawing the ROC curve
- Determining the optimal cut off point
- Precision and recall curve
- Evaluating the model
- Making Predictions on test data
- Checking evaluating parameters on test dataset

Model Evaluation Parameters

- Accuracy - Train data: 92%, Test Data: 92%
- Sensitivity - Train data: 92%, Test Data: 91%
- Specificity - Train data: 92%, Test Data: 93%
- Precision - Train data: 87%, Test Data: 88%
- Recall - Train data: 92%, Test Data: 91%

The sensitivity and specificity metrics show that the model is capable of predicting hot and cold leads >90% of the time, therefore we can say that the model is good and gives better expected results vs the CEO's expectation of 80%.

Tags, Lead source and Time spent on website are the most important variables to identify potential leads