For the Lead Scoring Case study, following are the summary of steps we did:

1. Importing and Exploring Data: After reading the data, we did basic inspection and cleanup of data.

2. Inspecting the data frame: We started with checking for missing values and Data Cleaning including Dropping columns. Dropping columns having greater than 40% missing values. After that we calculated the unique categories for the categorical variables. Dropped the columns with more than 90% unique categories. Replaced the missing values in categorical variables with appropriate values using Mean etc.

3. Data Preparation: We started with Converting the binary variables appropriately. Created the dummy variables for the categorical variables. and added those to the data frame.

4. Test Train Split: Split the overall dataset into Test and train.

5. Feature Scaling: implemented feature scaling so the data is uniformly scaled.

6. Correlations: Reviewed Correlations using heatmap plots.

7. Model Building: Started with logistical regression model.

8. Feature Selection Using RFE: completed feature selection using RFE. Trained the model and assessed the model using Stats model. Got the predicted values for training data. Checked VIFs: Check for the VIF values of the feature variables. dropped relevant columns and re ran the model using selected variables. Predicted the training set using the selected training variables. Checked the overall accuracy of the model. found it to be 89%. finalized the variables further using VIF analysis. checked confusion Matrix: checked Specificity, false positive rate, positive predictive value,

9.Plotted the ROC curve
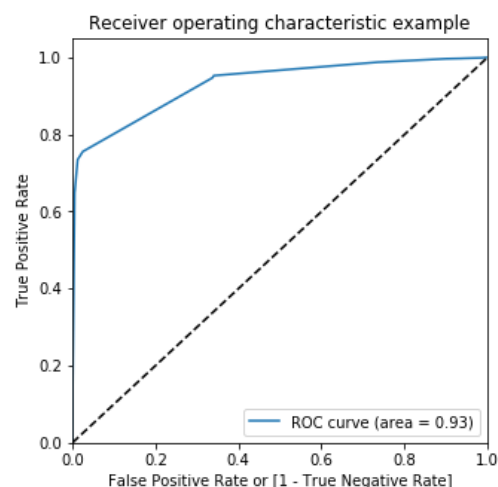
10. Found the optimal cut off point. calculate accuracy sensitivity and specificity for various probability cutoffs. Checked overall accuracy. checked specificity, false positive rate, positive/negative predictive values. Calculated precision and Recall
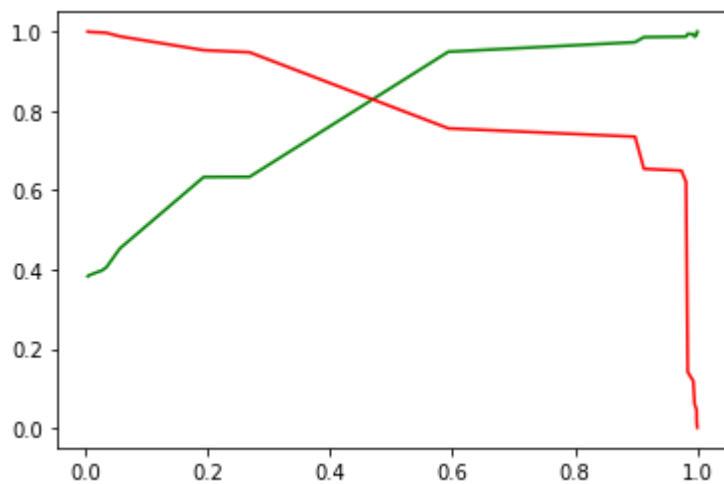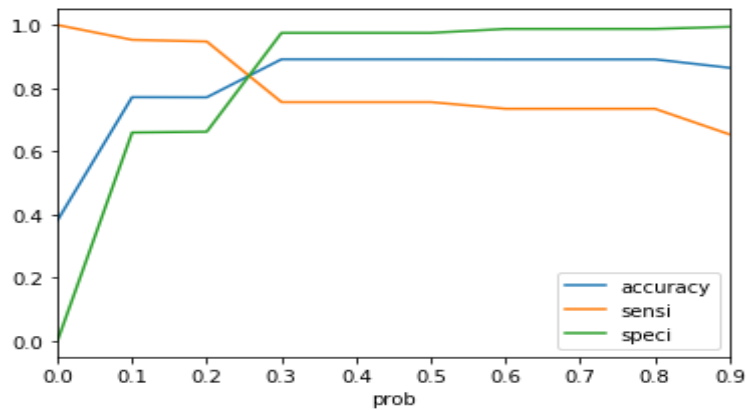


12. Made predictions on the test set based on the selected variables. Checked the accuracy, sensitivity, specificity