

LEAD SCORING CASE STUDY

SUBMITTED BY:

AMLESH DWIVEDI

HARSHVARDHAN SINGH JEENA

ANIRUDH YERRA

PROBLEM STATEMENT

- The typical lead conversion rate at X education is around 30%
- Company wants to identify the most potential leads, also know as 'Hot Leads'
- Identify the leads that are most likely to convert into paying customers(ball park figure 80%)
- Build a logistic regression model to assign a lead score between 0 and 100

STEPS INCLUDED IN THE PROCESS

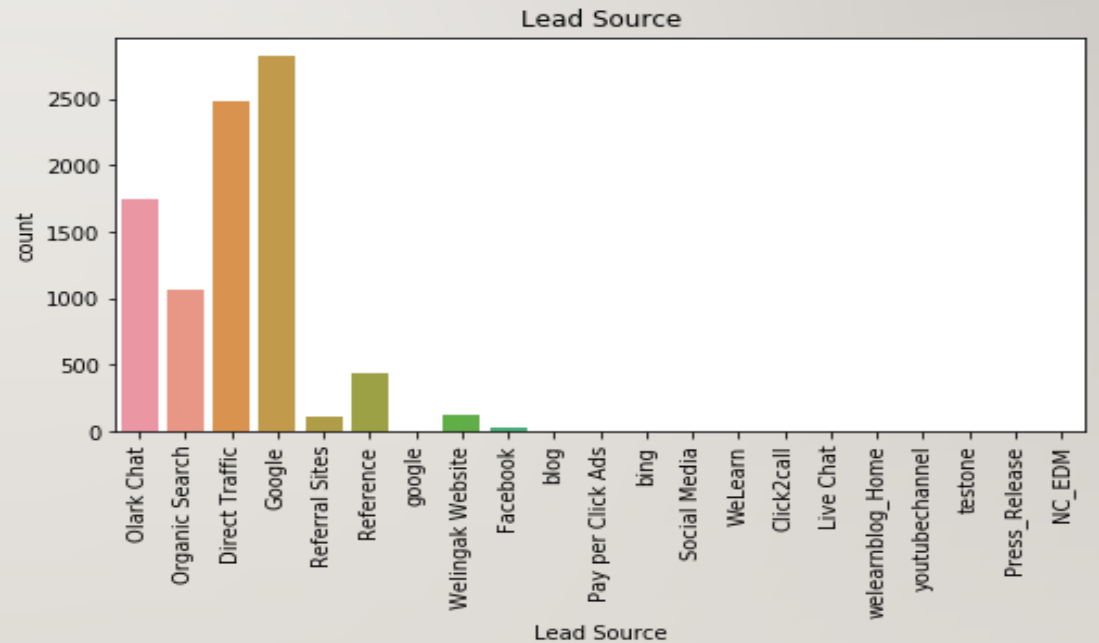
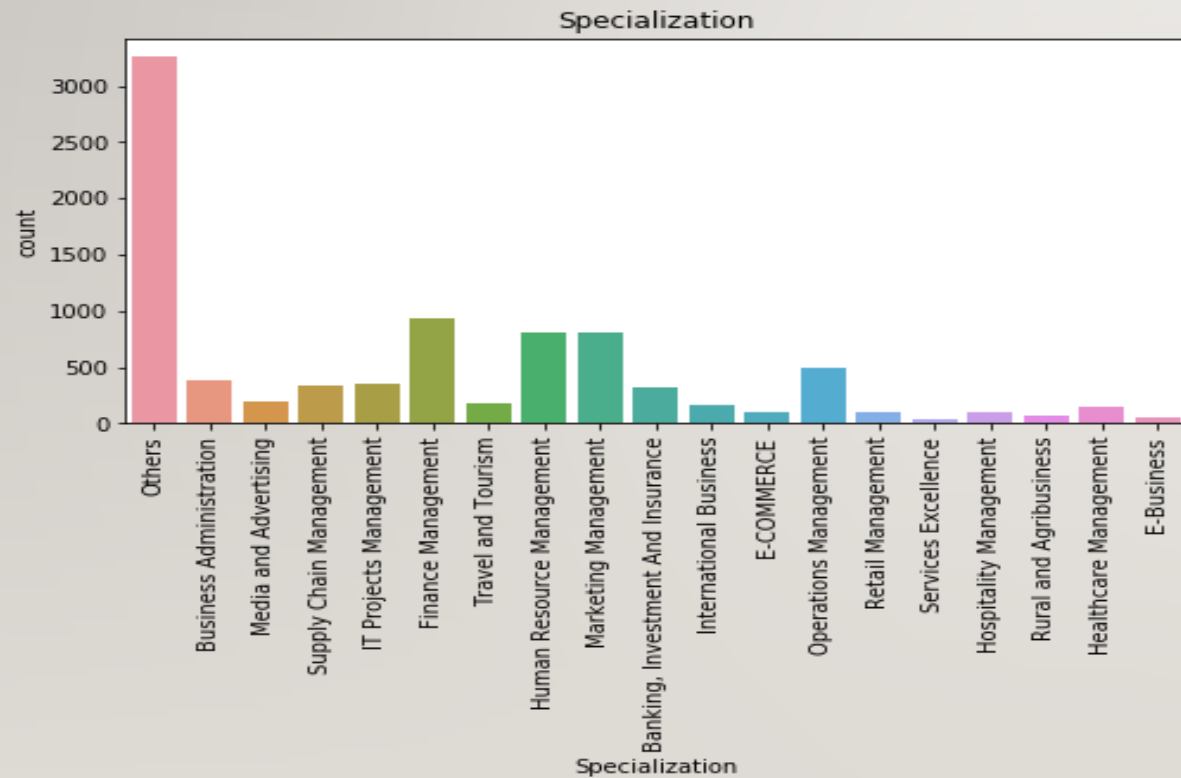
- Loading and understanding the data set
- Cleaning of the Data Frame
- EDA
- Data Preparation for Modelling
- Building The Model
- Model Evaluation
- Making Predictions on Test Set
- Calculating the Lead Score
- Feature Selection

DATA CLEANING

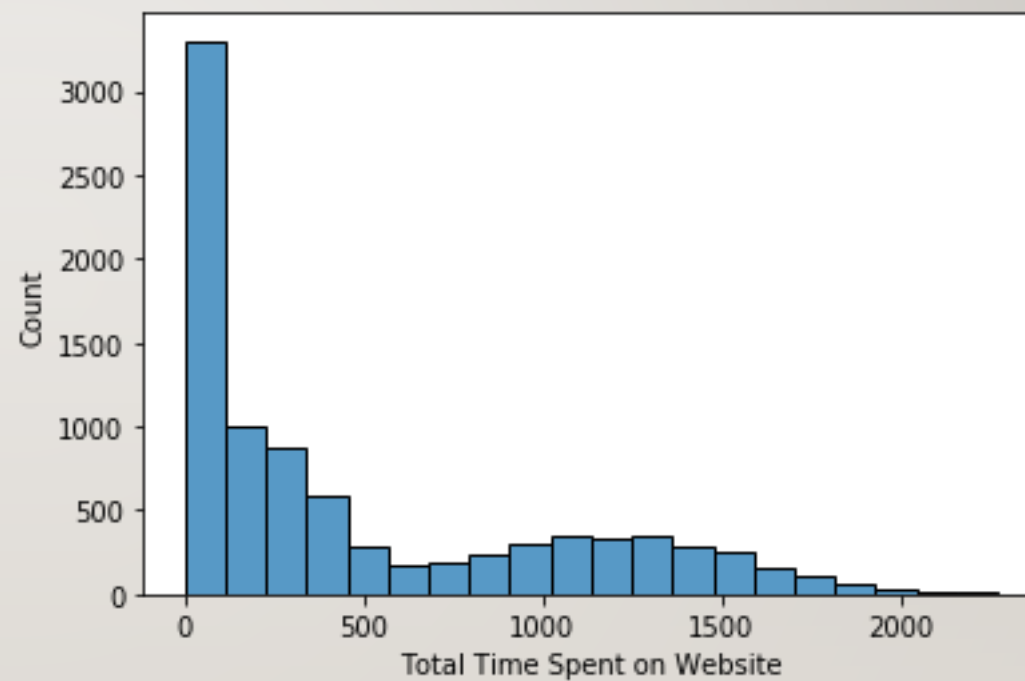
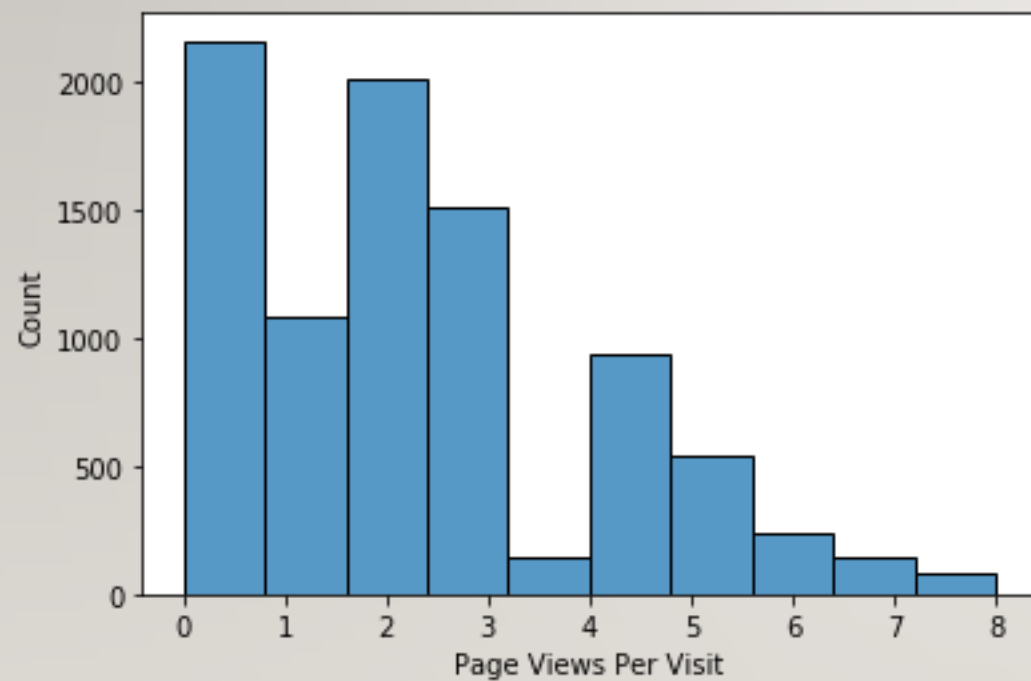
- We removed the unnecessary columns from the data set eliminating the data set with high % of null values
- The columns with skewed values were also removed
- After removing we were left with 14 columns to work ahead

EDA

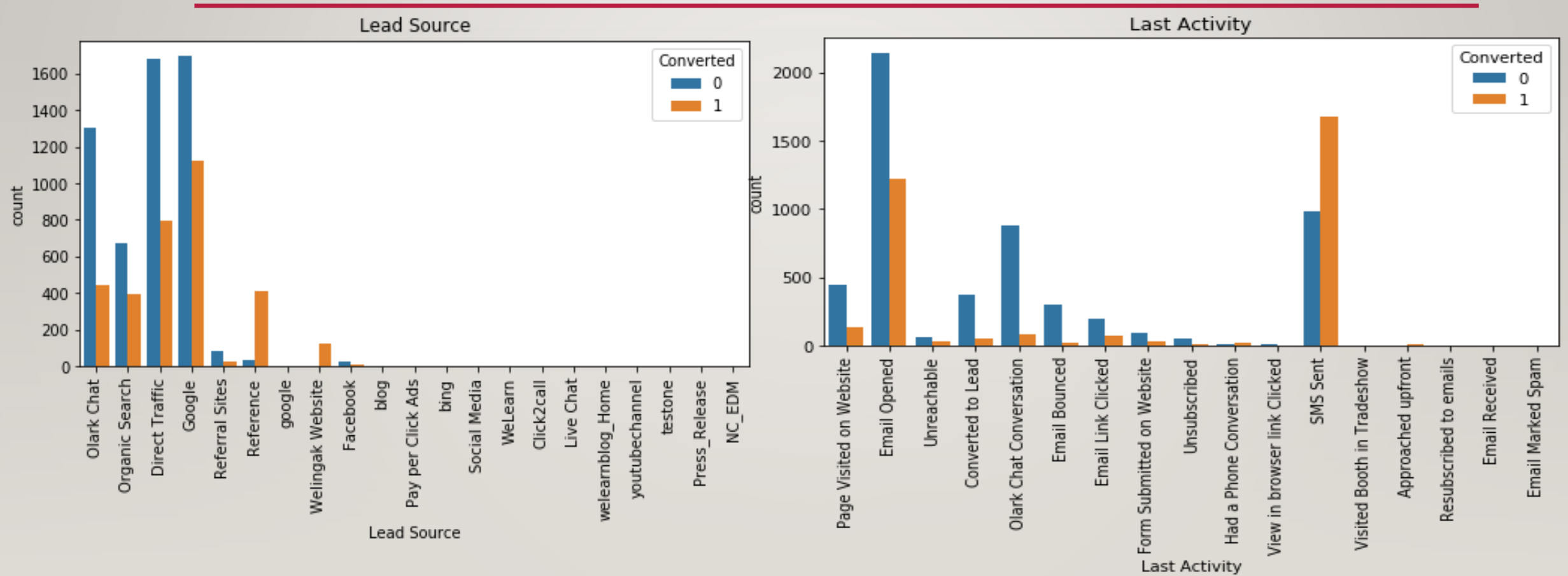
- Univariate and Bivariate analysis were performed on the data set



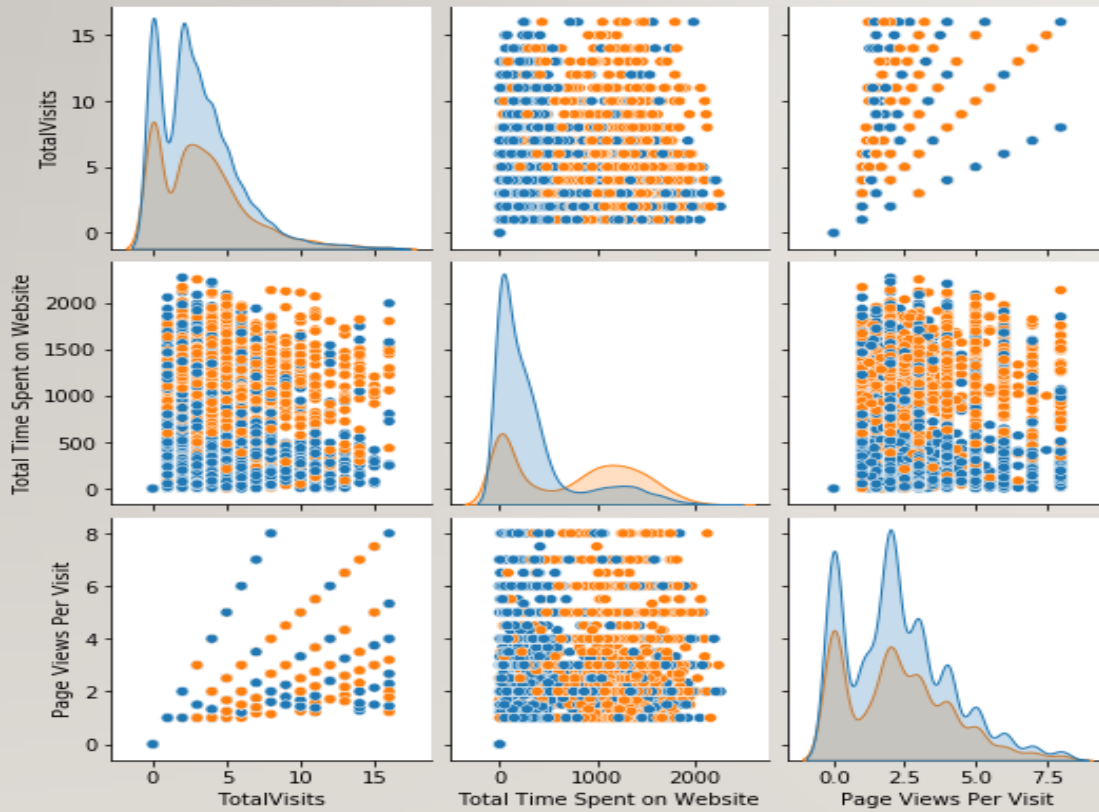
EDA



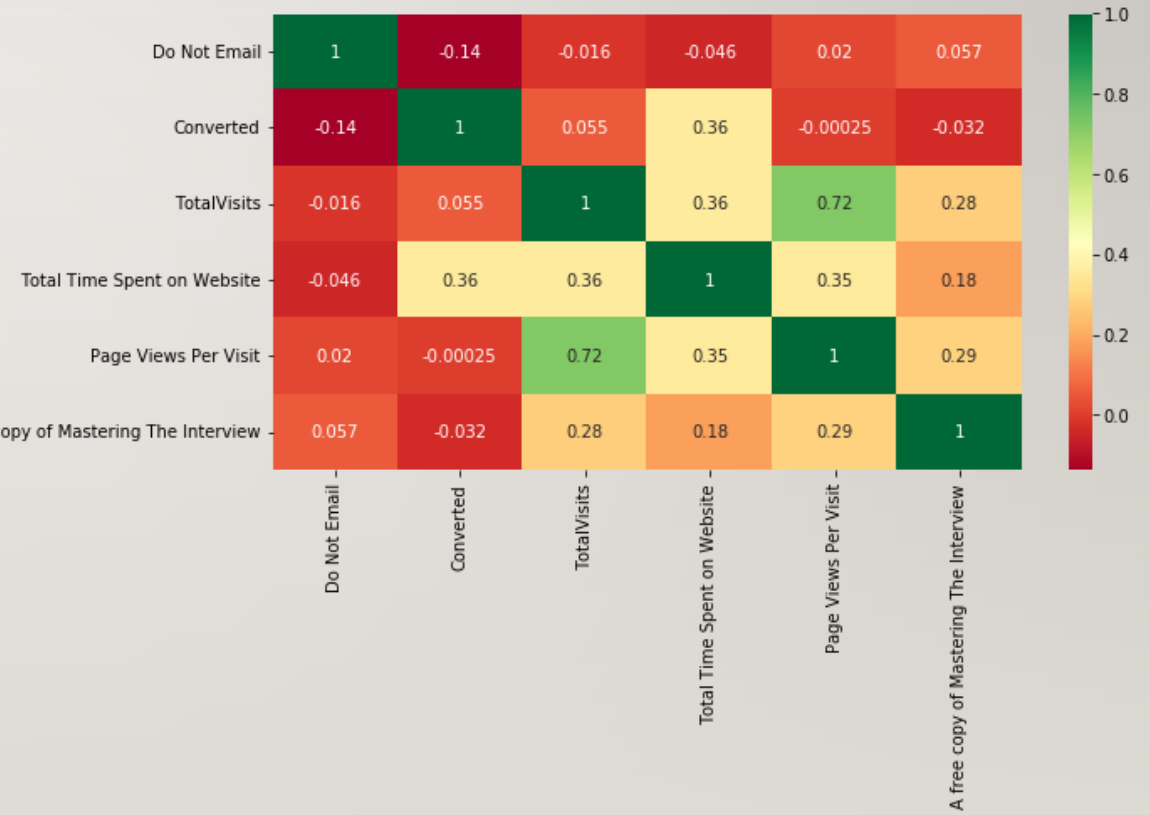
EDA



EDA



A free copy of Mastering The Interview

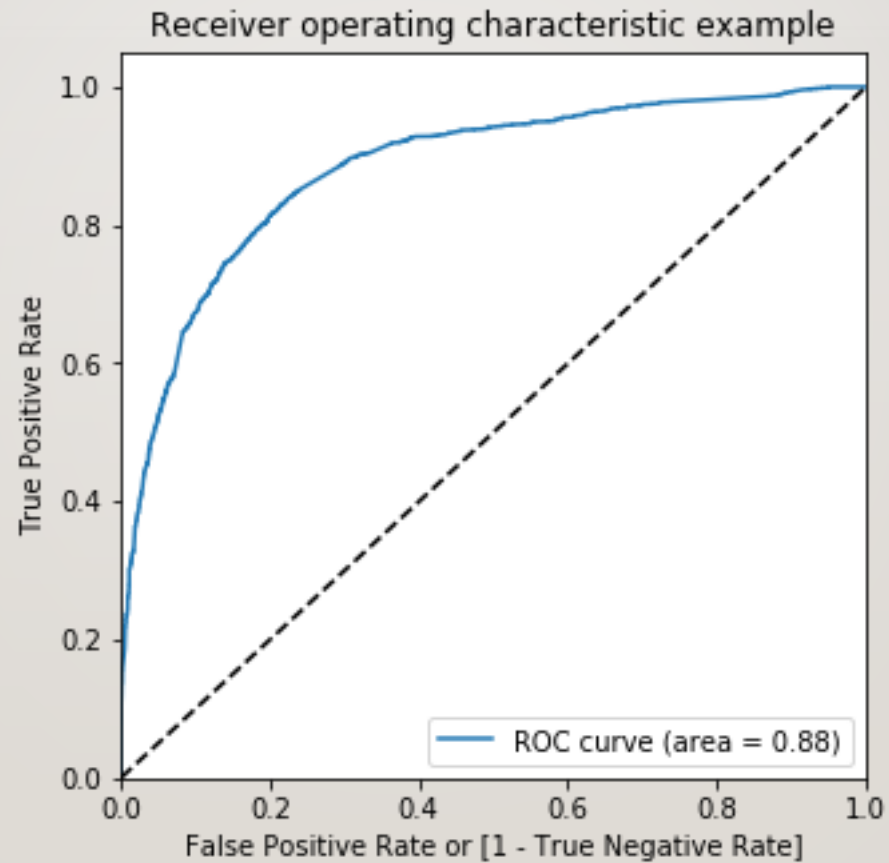


MODEL PREPARATION

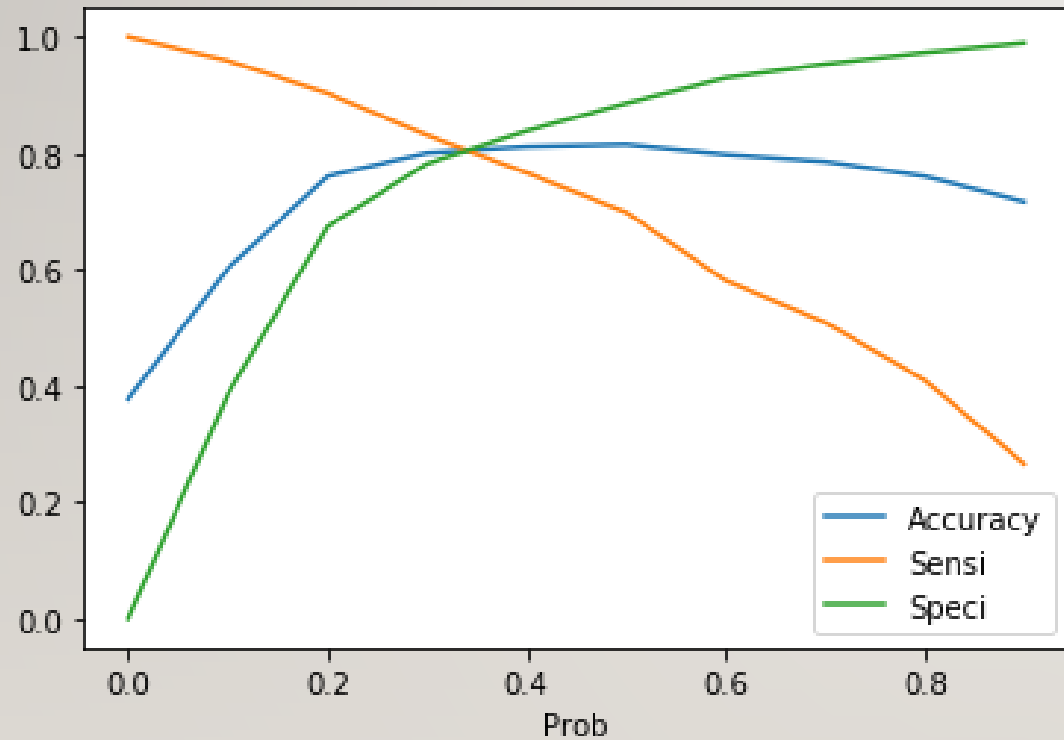
- Dummy variables were created for the categorical columns
- Scaling of the data was done
- The test train split on the data was carried out
- Features selection was done on the data set with the help of RFE
- VIF was also used to check the optimal features
- p -value 0.05 and $RFE < 5$

MODEL EVALUATION

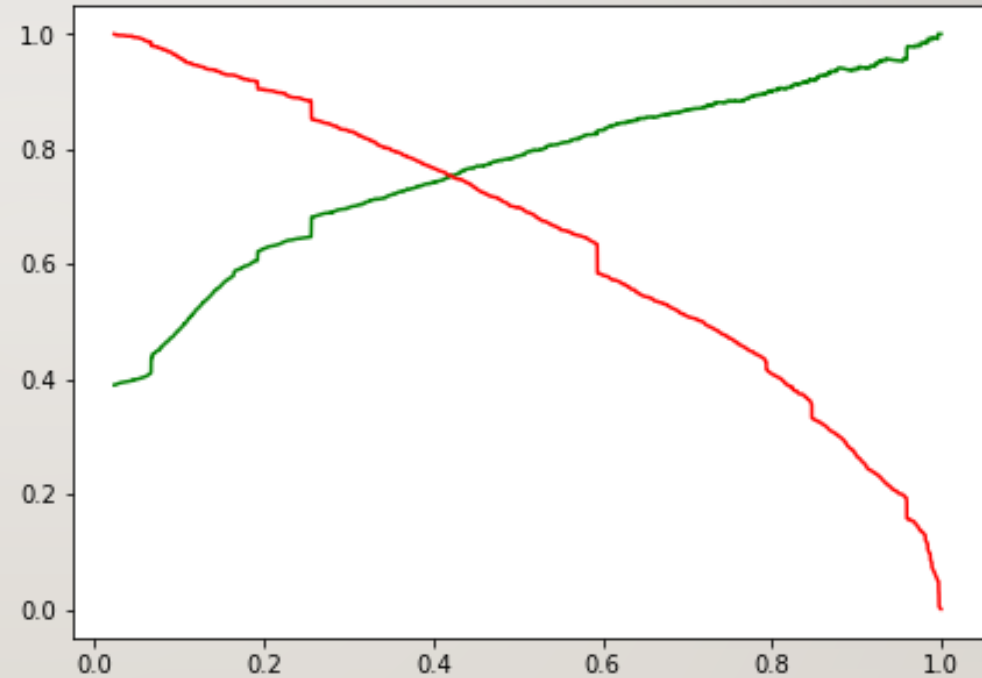
- ROC Curve



MODEL EVALUATION TRAIN

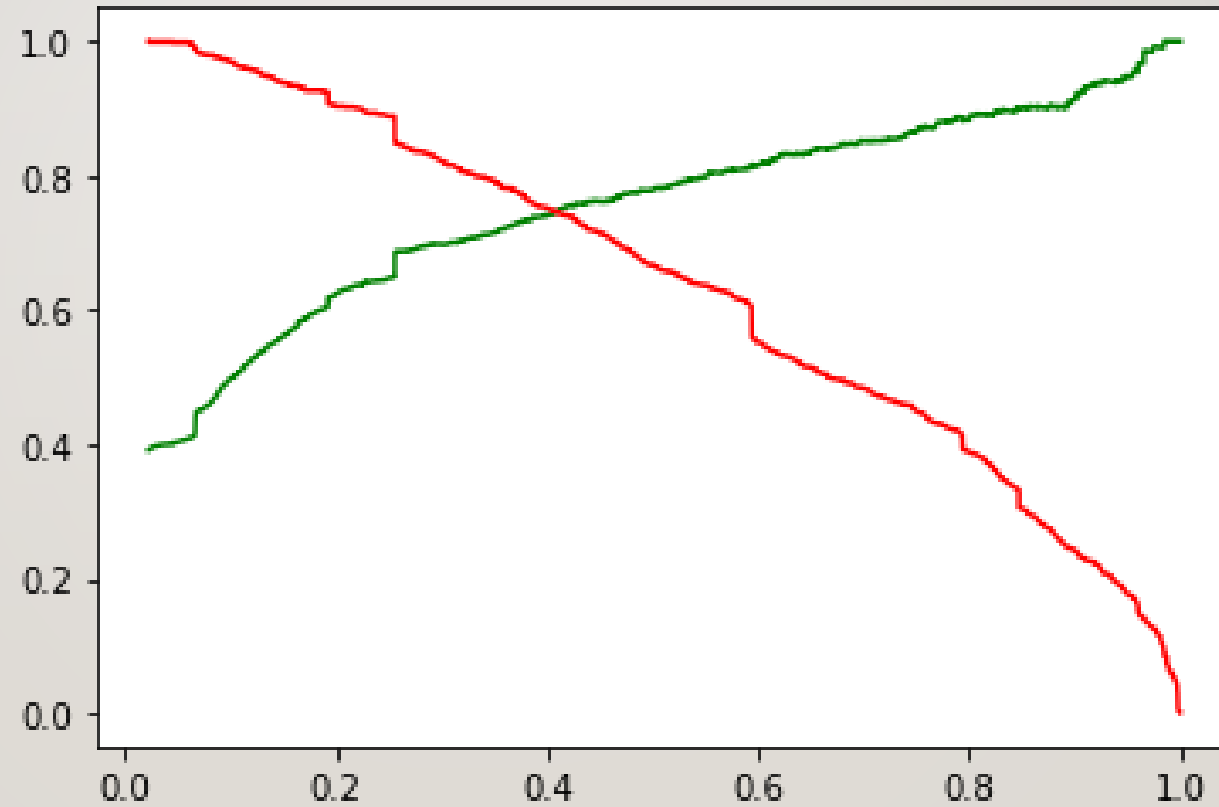


1. Acuracy: 81.49 2. Sensitivity: 69.84 3.
Specificity: 88.55



MODEL EVALUATION TEST

- Accuracy: 80.06%
- Sensitivity: 80%
- Specificity: 79.48%



HOT LEADS BASED ON SCORE

Prospect Id	Converted	Converted_Prob	Final_Predicted	Lead_Score
37	1	0.846540	1	85
64	1	0.967691	1	97
76	1	0.866202	1	87
91	1	0.866519	1	87
112	1	0.980587	1	98

FEATURE SELECTION

Total Time Spent on Website	4.56
Lead Origin_Lead Add Form	4.40
What is your current occupation_Working Professional	2.77
Lead Source_Welingak Website	2.62
Lead Source_Olark Chat	1.62
TotalVisits	1.51
Last Activity_Olark Chat Conversation	-1.21
Last Notable Activity_Email Opened	-1.45
Do Not Email	-1.61
Last Notable Activity_Olark Chat Conversation	-1.80
Last Notable Activity_Modified	-1.81
Last Notable Activity_Page Visited on Website	-1.94
Last Notable Activity_Email Link Clicked	-2.01
dtype: float64	

THANK YOU