Lead Score Case Study

Submitted by

Pasupulat Kamlesh kumar

Problem Statement:

X Education sells online courses to Industry professionals.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead

The company requires you to build a model wherein you need to assign a lead score 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.

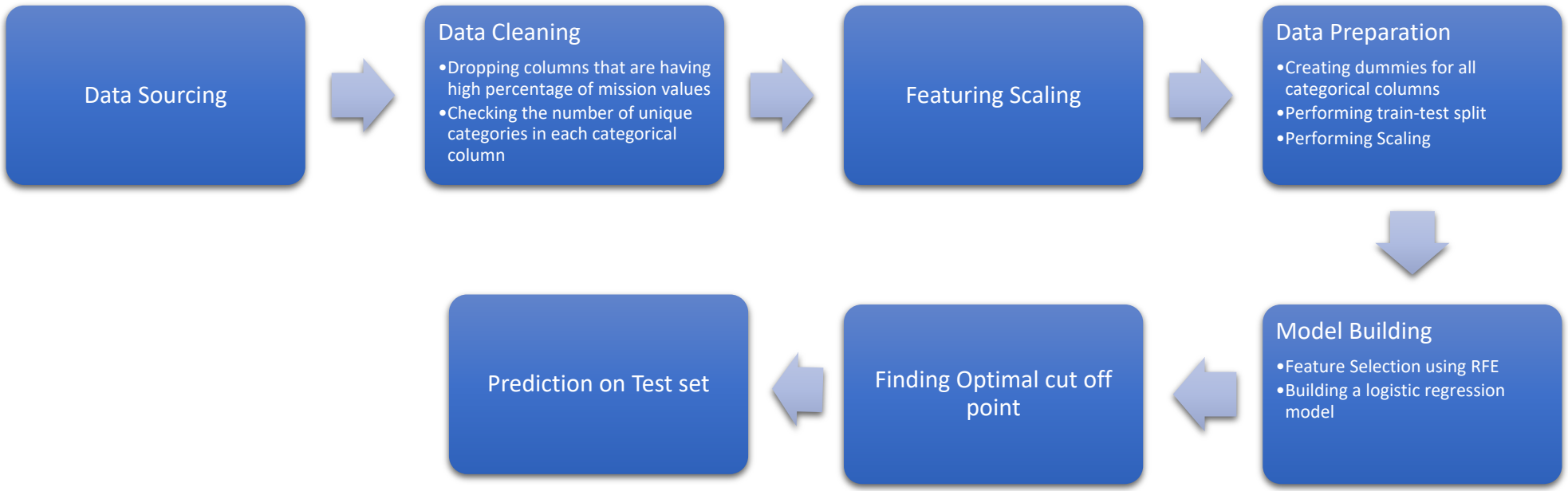
Business Goal:

The company requires you to build a model wherein you need to assign a lead score 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.

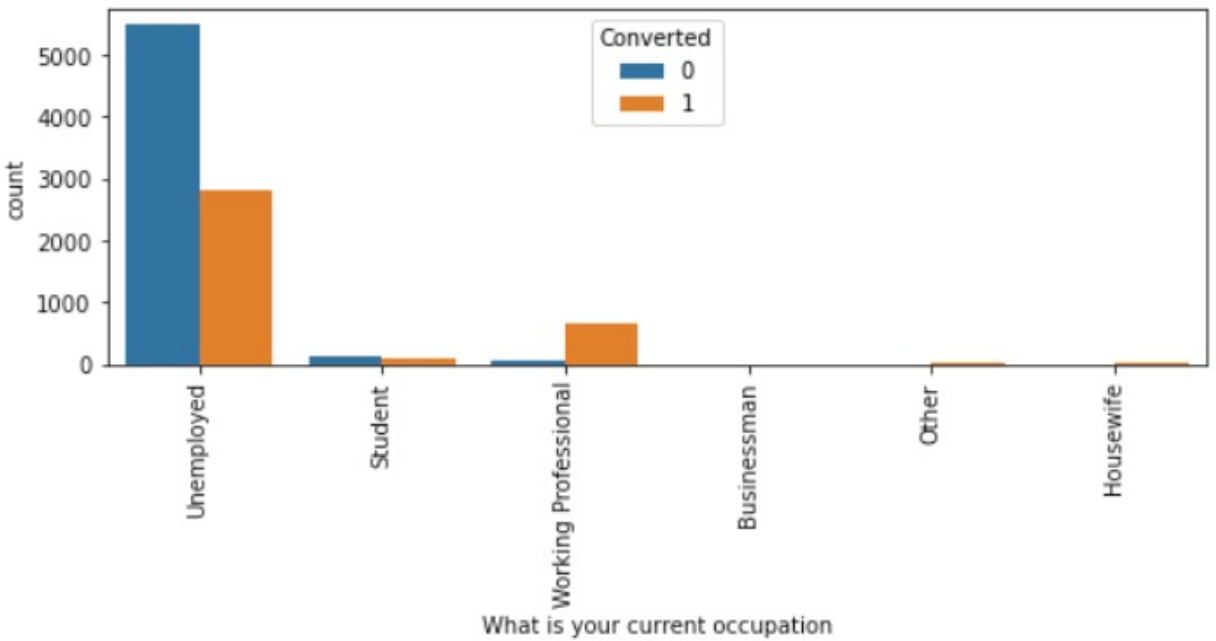
Strategy:

- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis
- Splitting the data into Test and Train dataset
- Building a logistic Regression model and calculate Lead Score
- Evaluating the model by using different metrics – Specificity and Sensitivity or Precision and Recall
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

Steps followed to solve

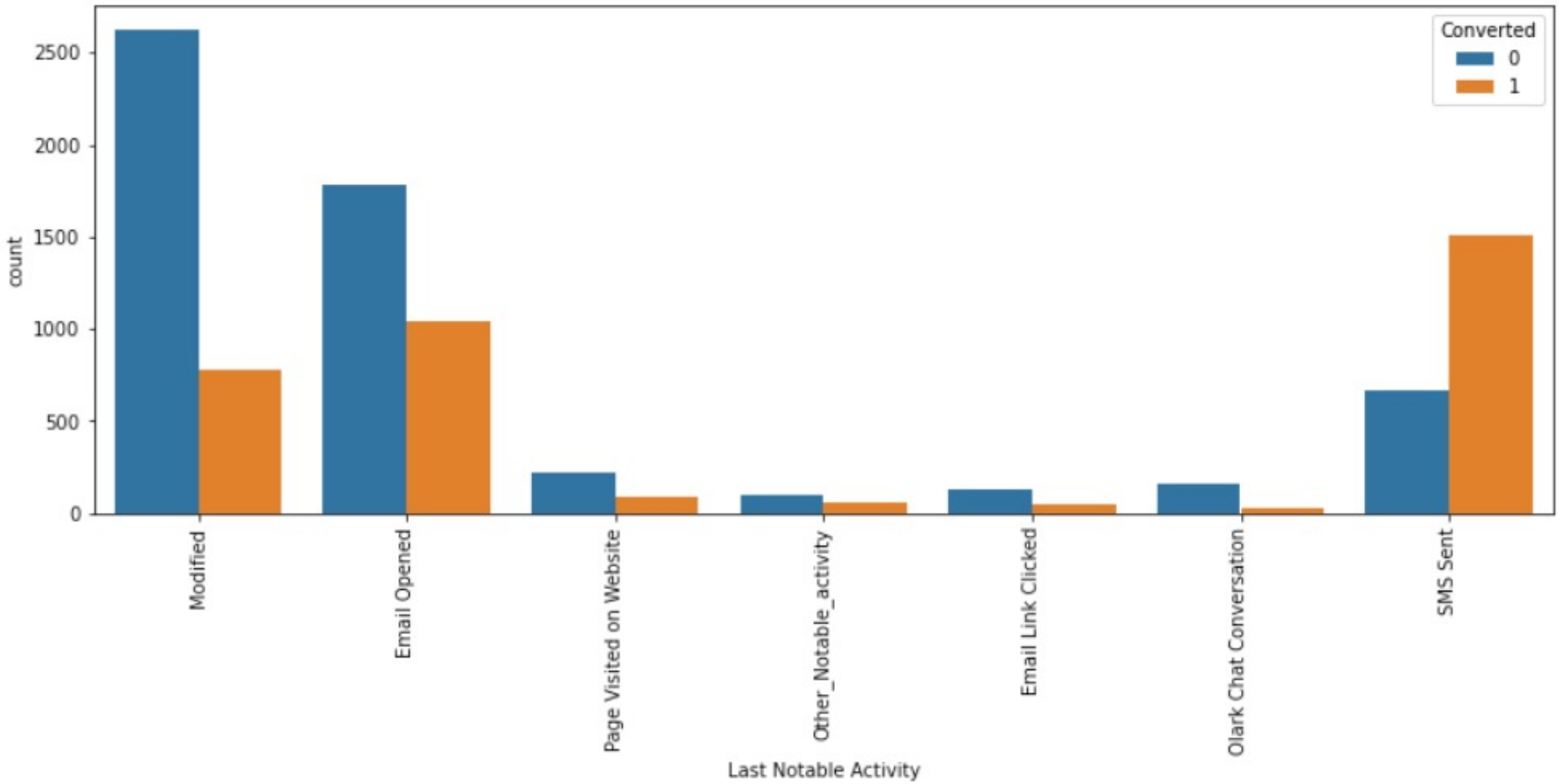


Exploratory Data Analysis

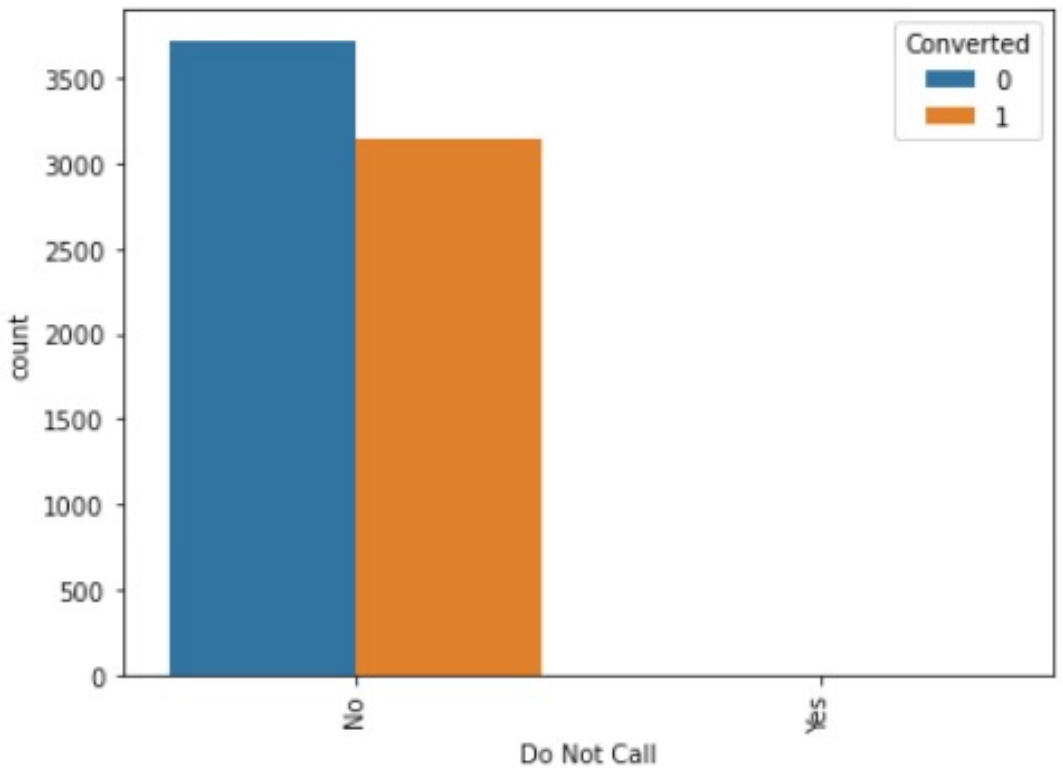


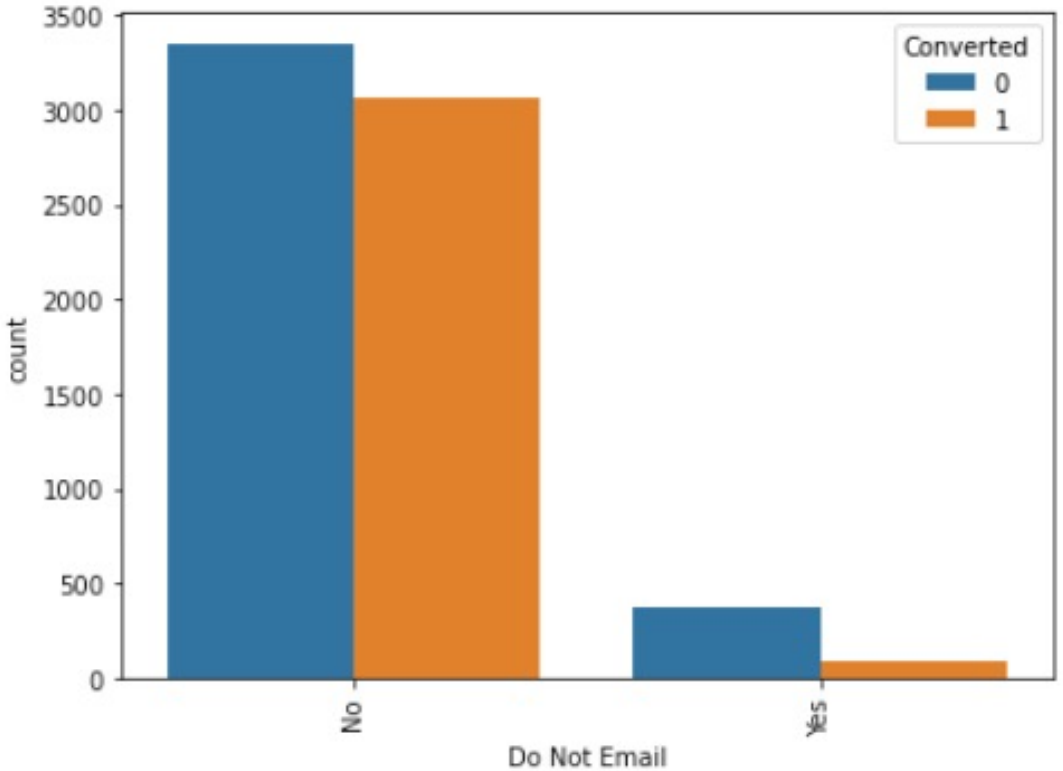
Unemployed and working professional are tends to take courses. Working professionals has better rate conversion

Exploratory Data Analysis

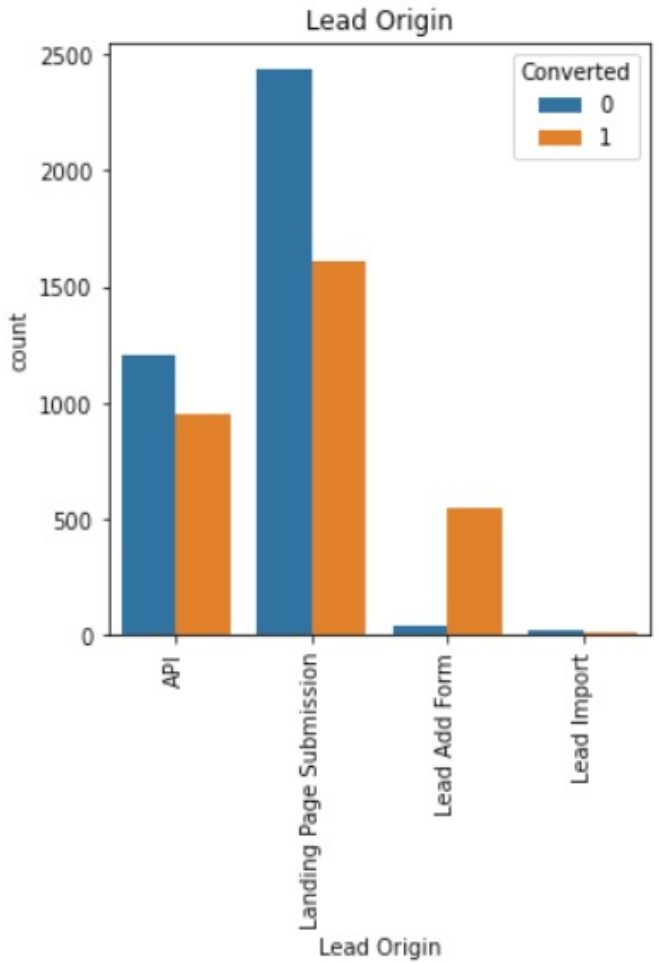


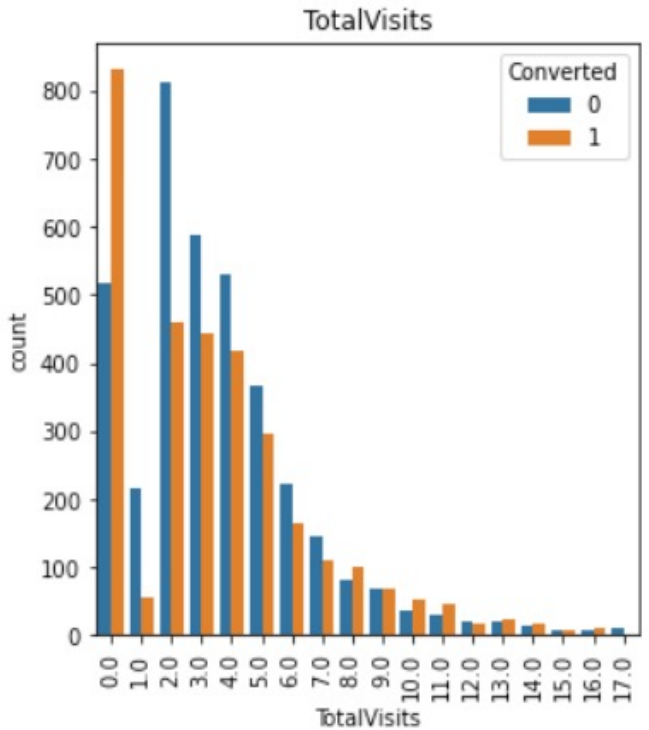
Exploratory Data Analysis



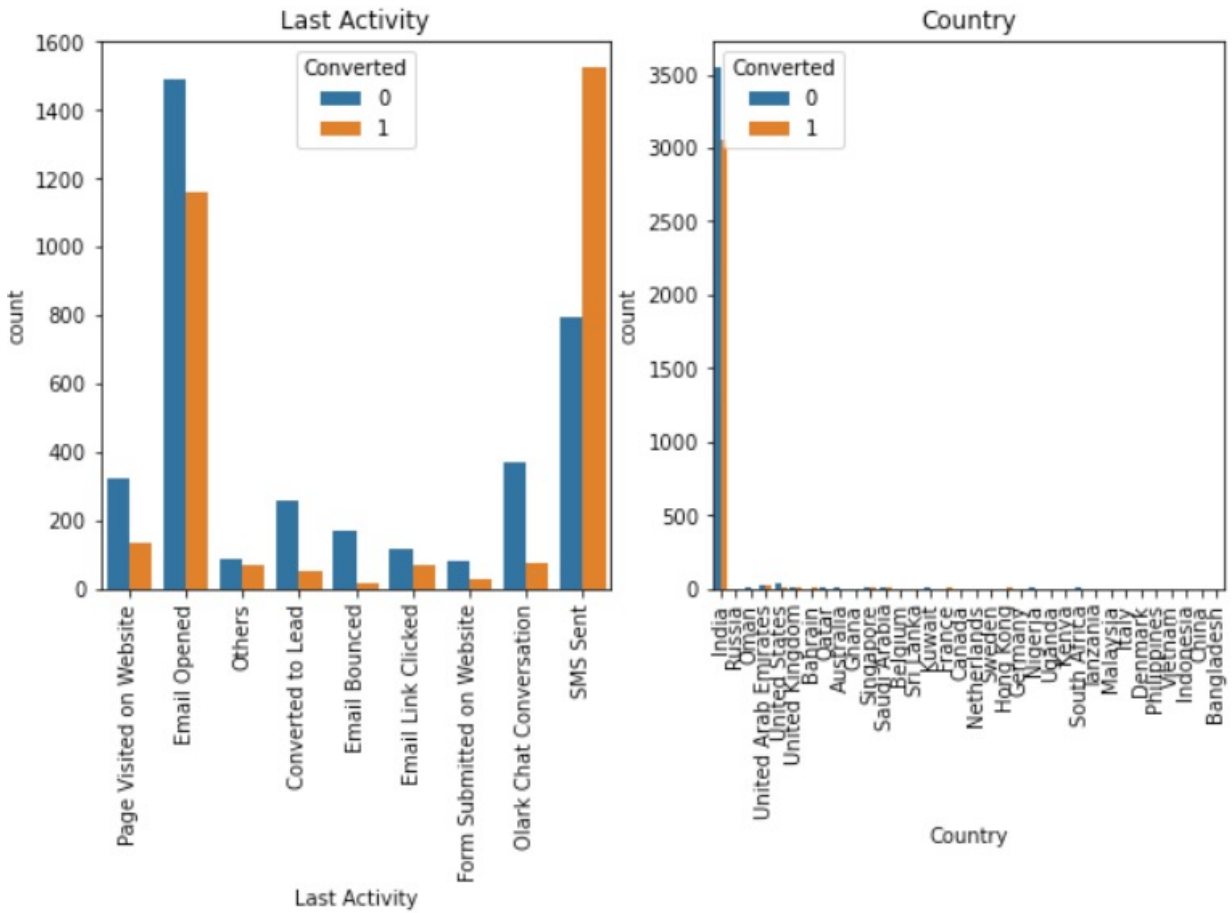


Exploratory Data Analysis



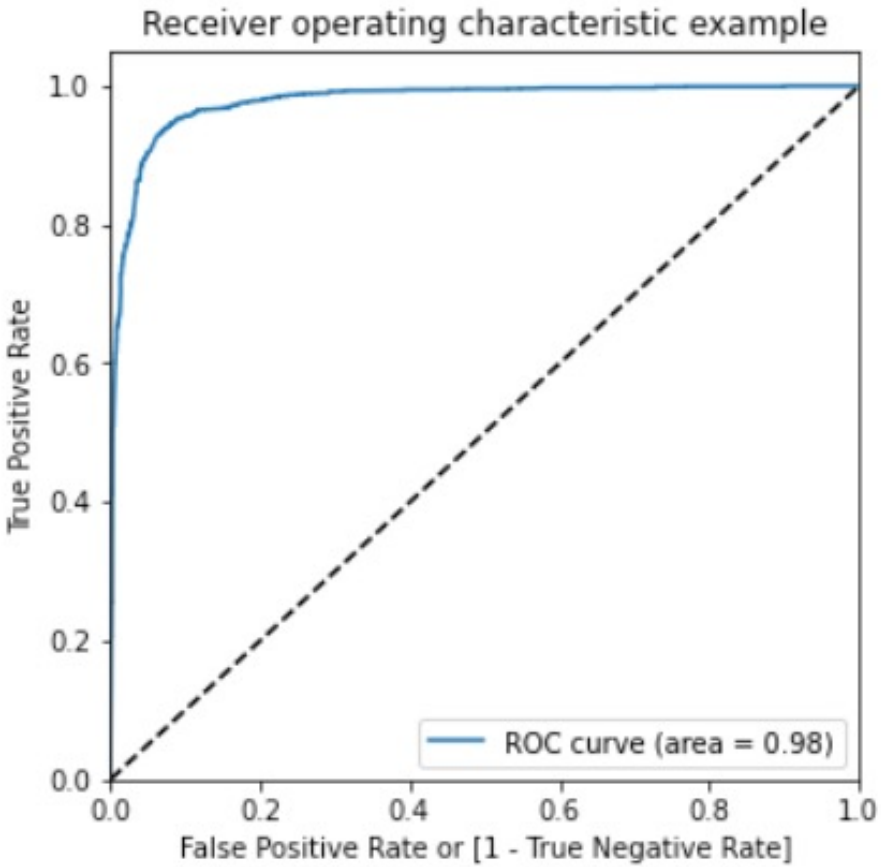


Exploratory Data Analysis



SMS and Email are more effective in conversion lead. Most of the enquiries are from India

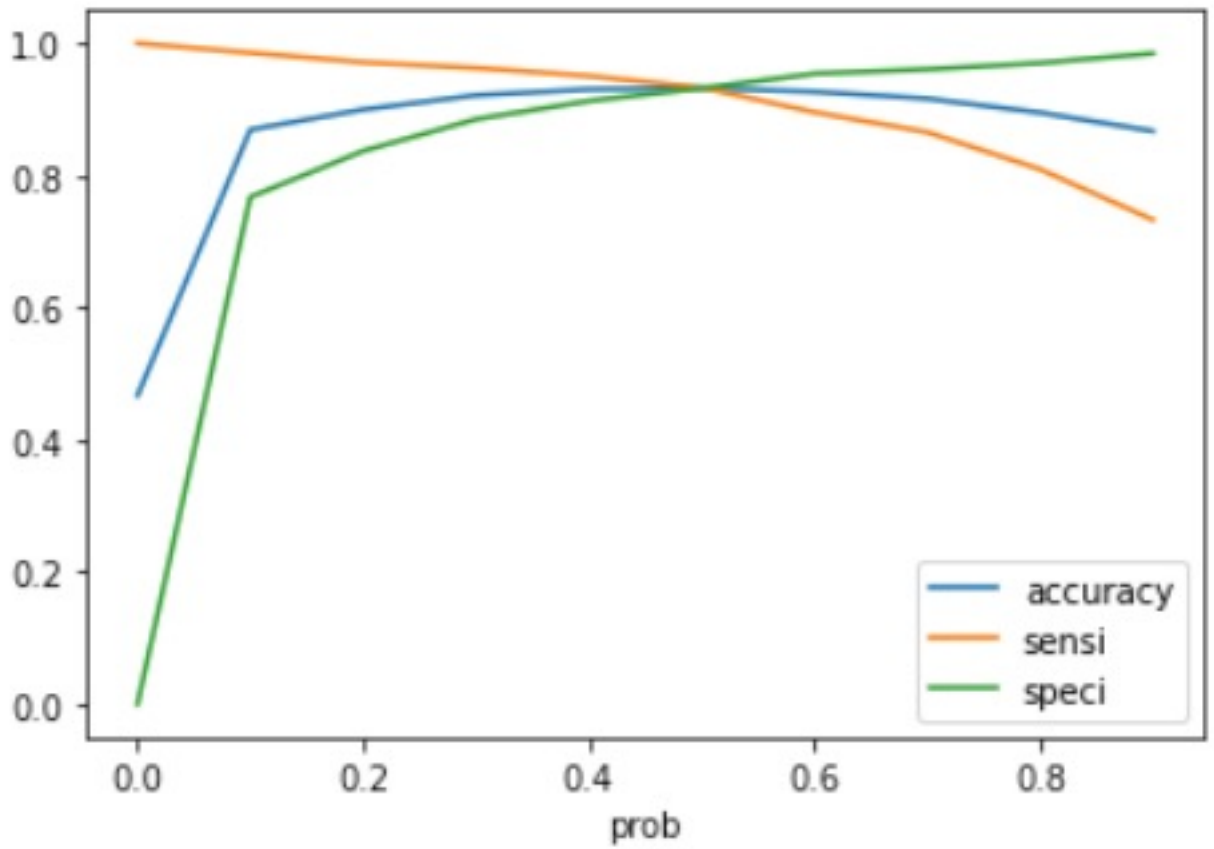
Model Evaluation



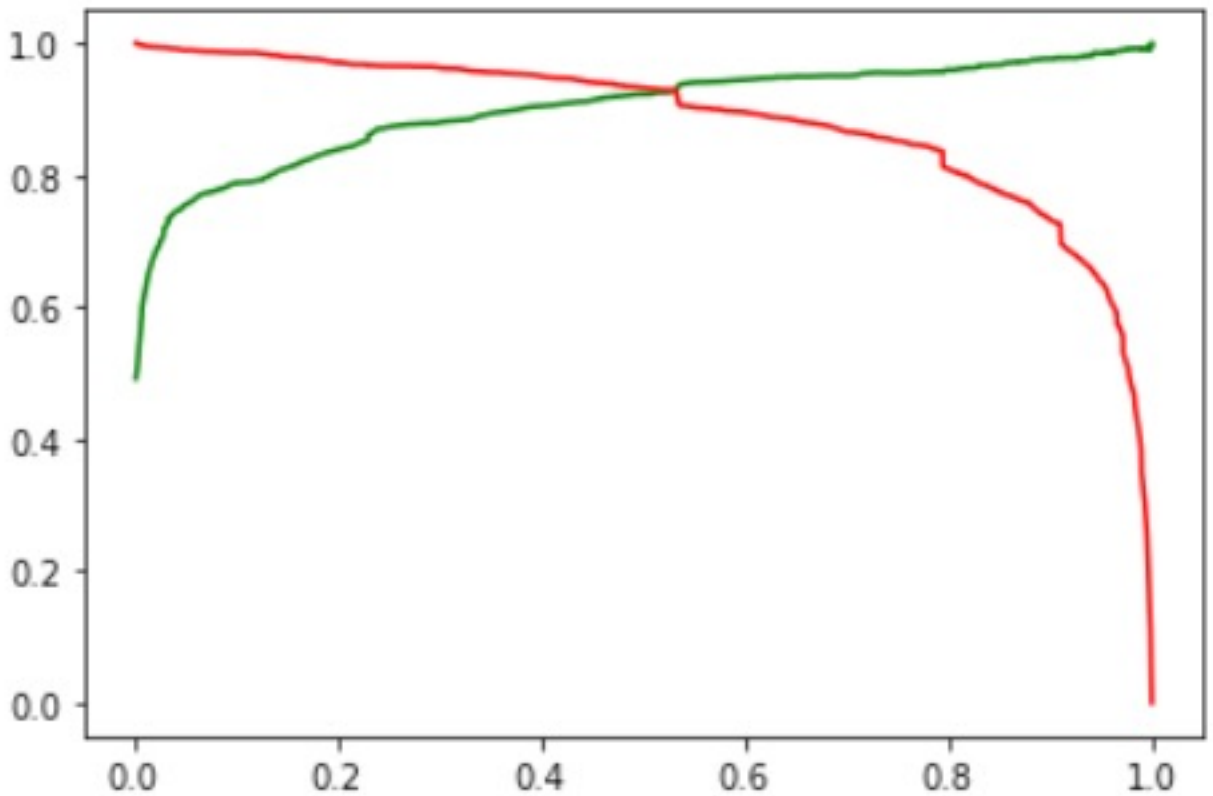
The ROC curve has a value of 0.98

- Accuracy : 92.12%
- Sensitivity : 96.26%
- Specificity : 88.50%

Model Evaluation



Model Evaluation



The graph depicts an optimal cut off of 0.56 bases on Precision and Recall

Model Evaluation

The top 3 variables that contribute for lead getting converted in the model are

- Tags
- Lead Origin
- Total Time Spend on Website

With the Accuracy, Sensitivity and Specificity values of test set are around 91%, 95% and 88% respectively, and values of train set are around 92%, 96% and 88% respectively model seems to perform good