Lead Scoring Case Study

Submitted by

Ankita Shukla

Roopak Ar

Problem statement

- An education company named X Education sells online courses to industry professionals.
- The company markets its courses on numerous websites and search engines like Google. After these people land on the website, they might glance the courses or fill up a form for the course or watch some videos.
- When these individuals fill up a form provided that their email address or phone number, they are categorized to be a lead. Additionally, the company also gets leads through past referrals.
- The typical lead conversion rate at X education is around 30%.

Business Objective

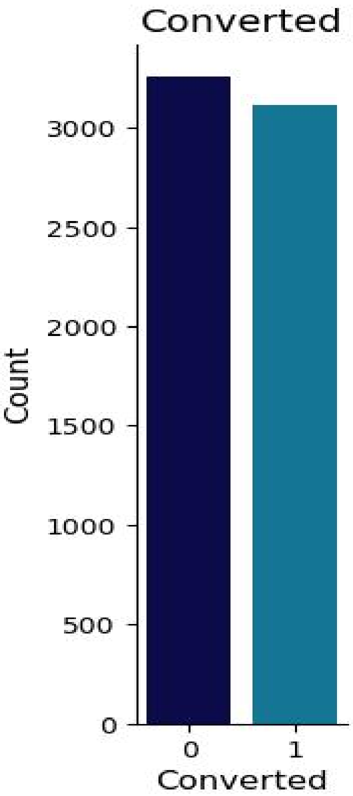
- The company requires to build a model where need to allocate a lead score to each one of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.
- The CEO, in precise, has given a ballpark of the target lead conversion rate to be around 80%.

Solution Methodology Used

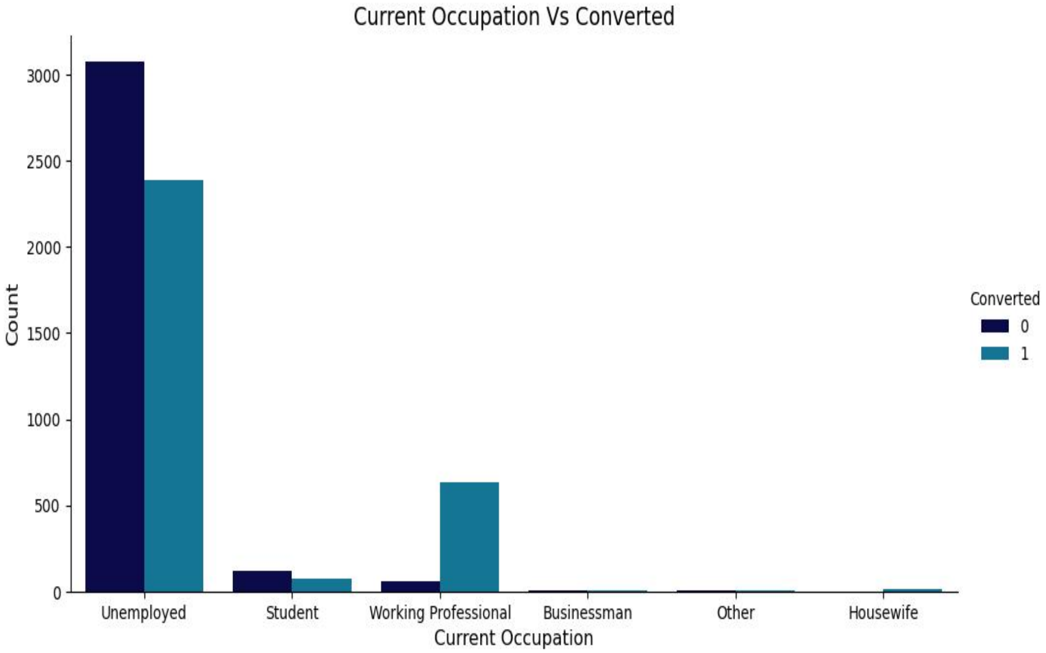
- Importing dataset
- Cleaning and preparing the dataset
- Exploratory Data Analysis
- Feature scaling
- Splitting the data into Test and Train dataset
- Building a Logistic Regression Model
- Evaluating the model by using different matrices
- Measuring the accuracy of the model and other matrices for evaluation
- Summary of the Model

Exploratory Data Analysis

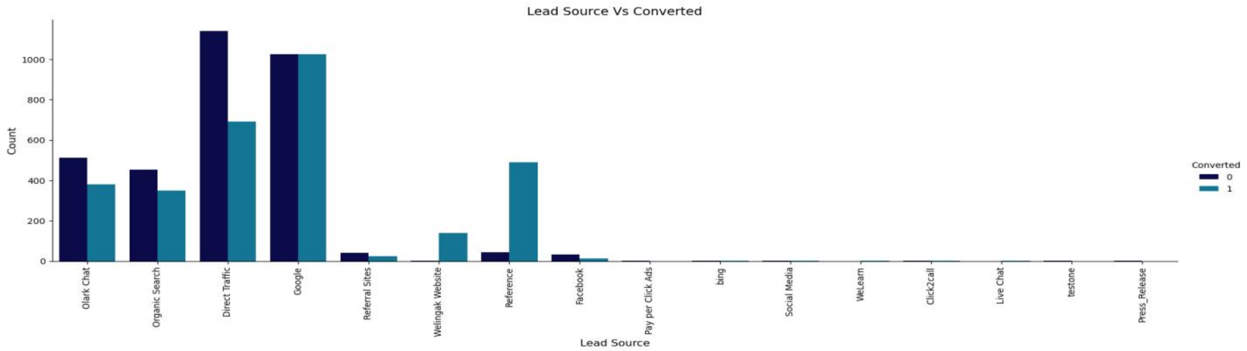
Overall Conversion Rate

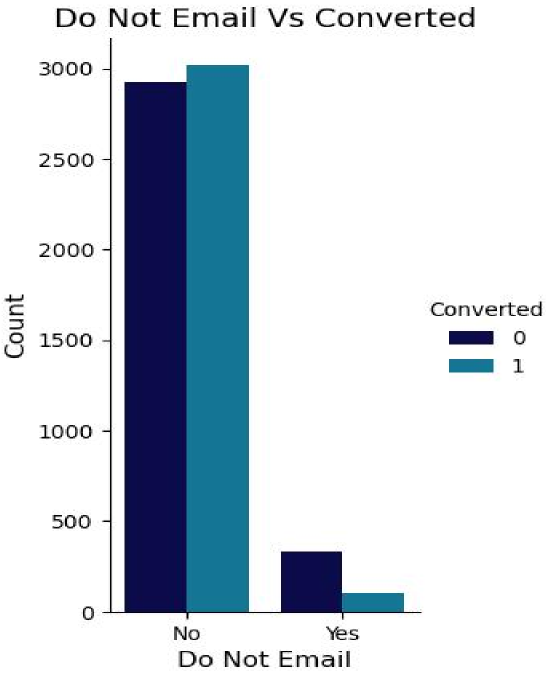


Current Occupation vs Conversion

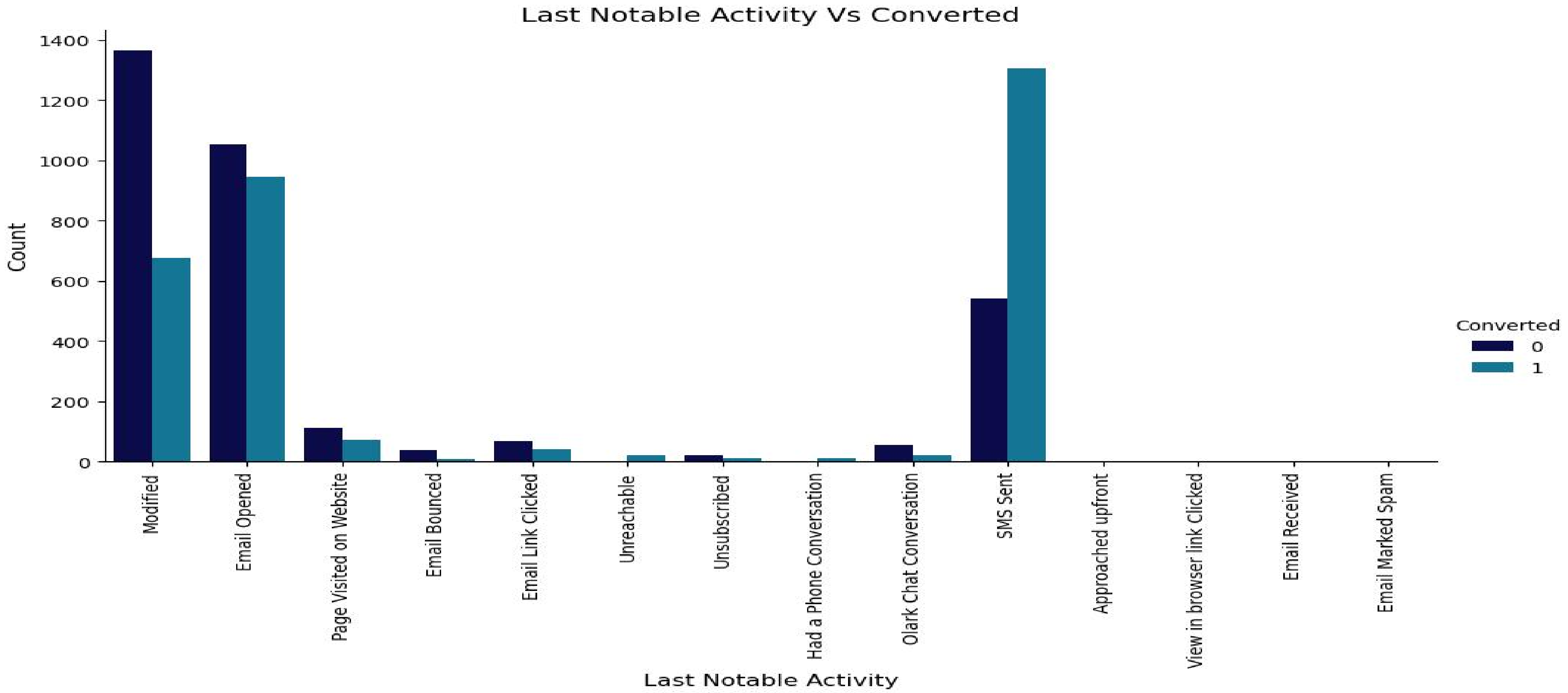


Lead Source Vs Conversion

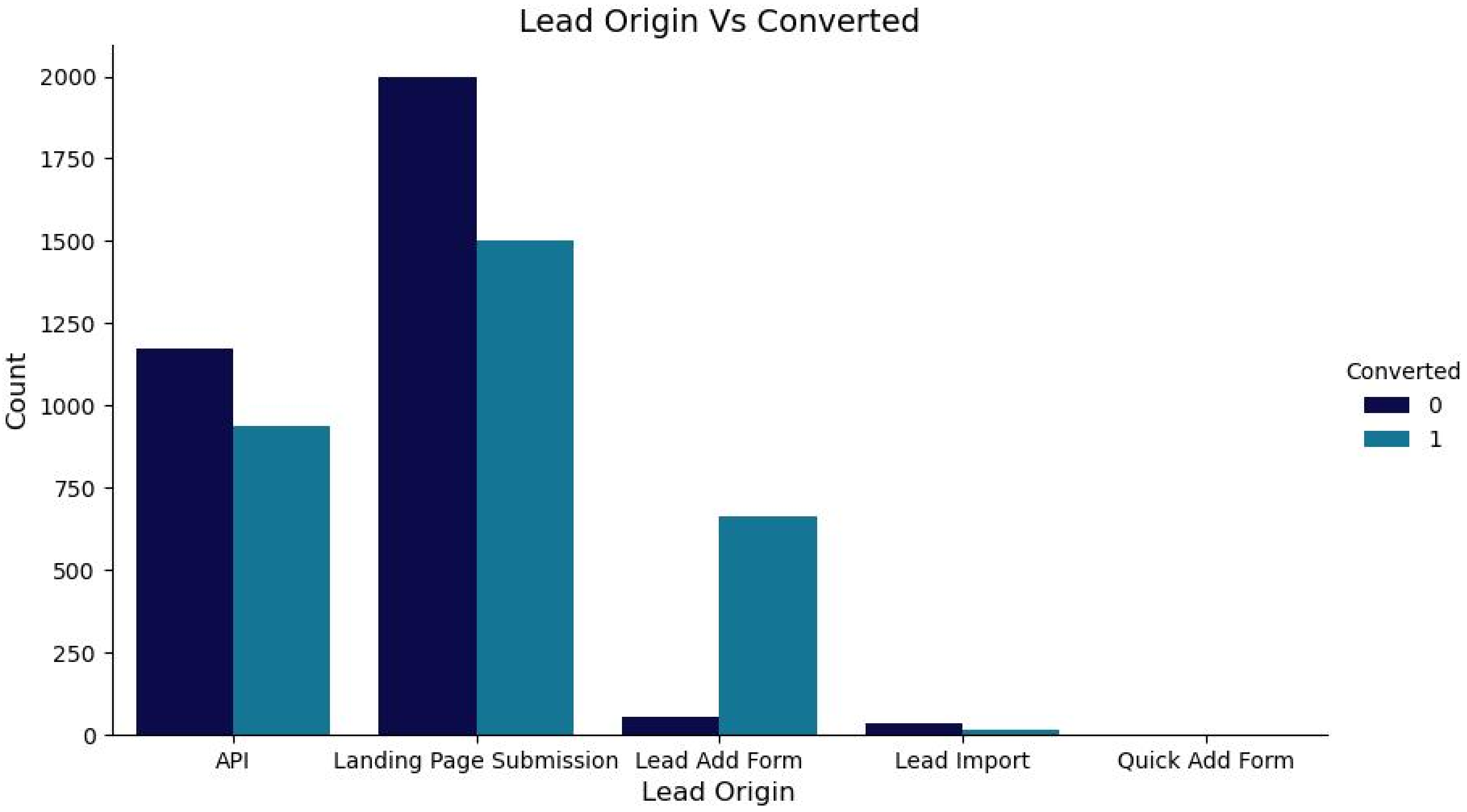




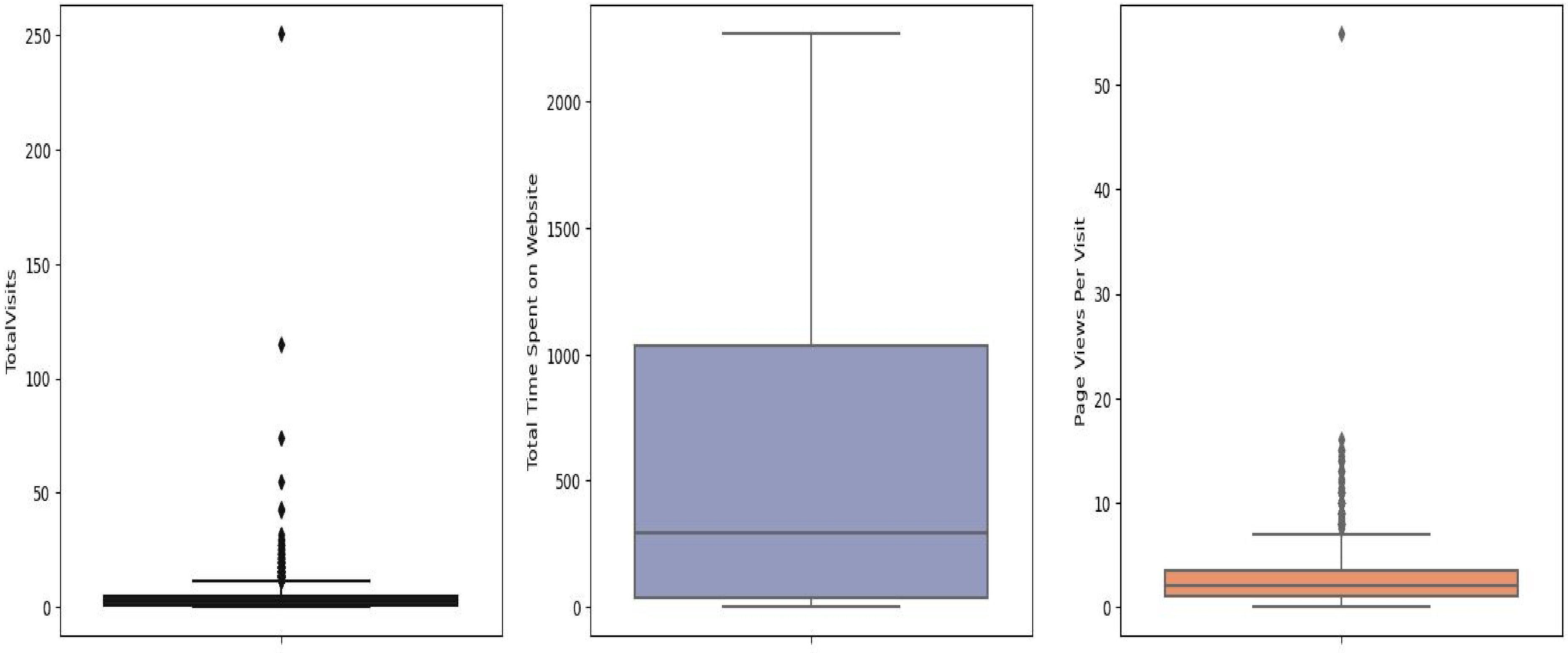
Do Not Email Vs Converted



Last activity of ‘**SMS Sent**’ has more conversion rates.

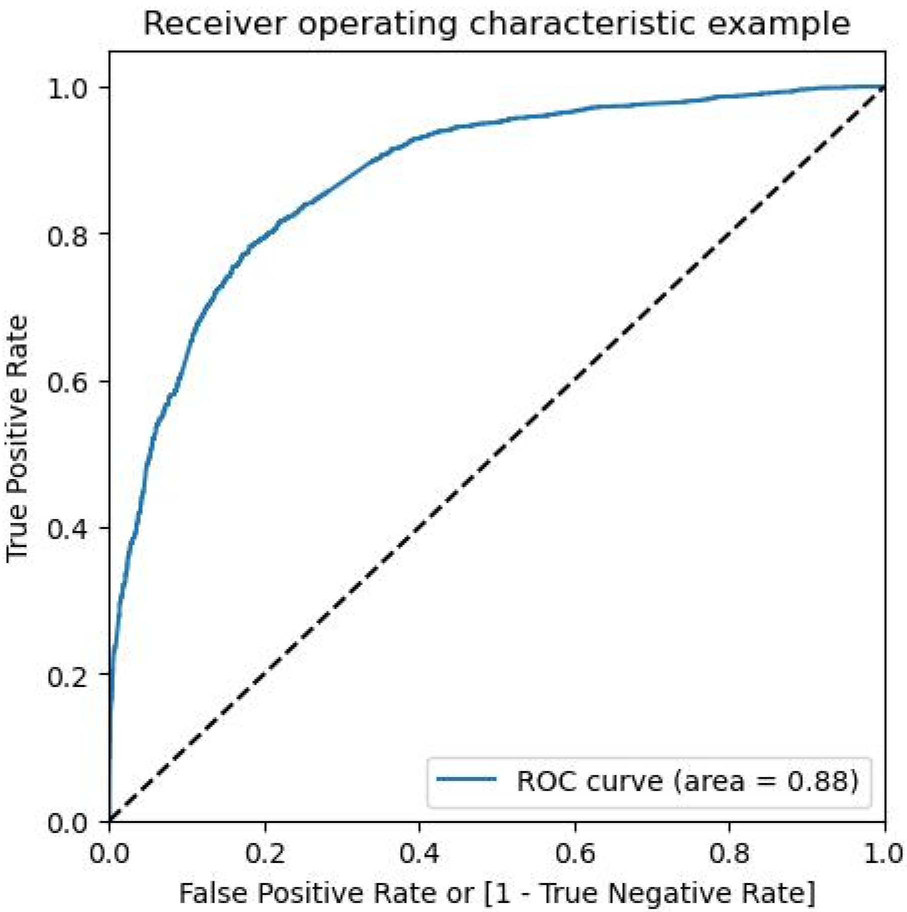


Lead origin maximum conversion happened from landing page submission



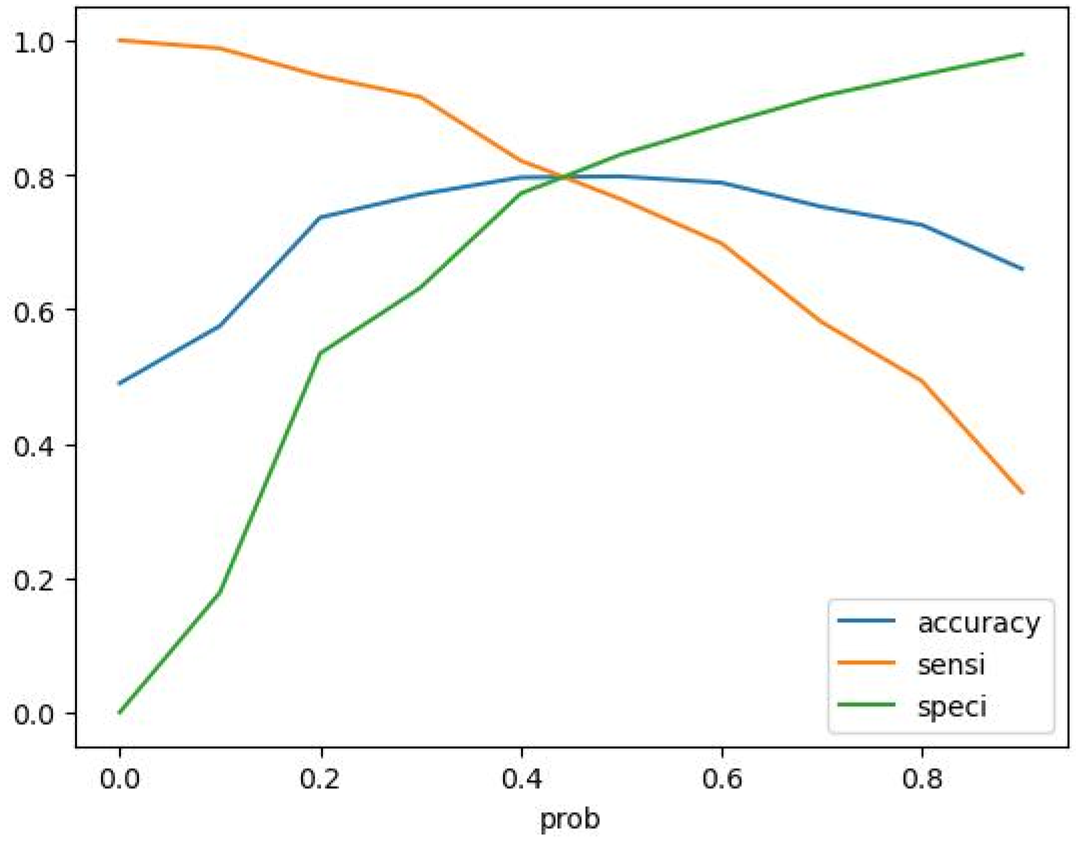
Conversion Rates for Total Visits, Total Time Spent on Websites and Page veiws per visits.

Model Evaluation- Sensitivity and Specificity on Train and Test Dataset



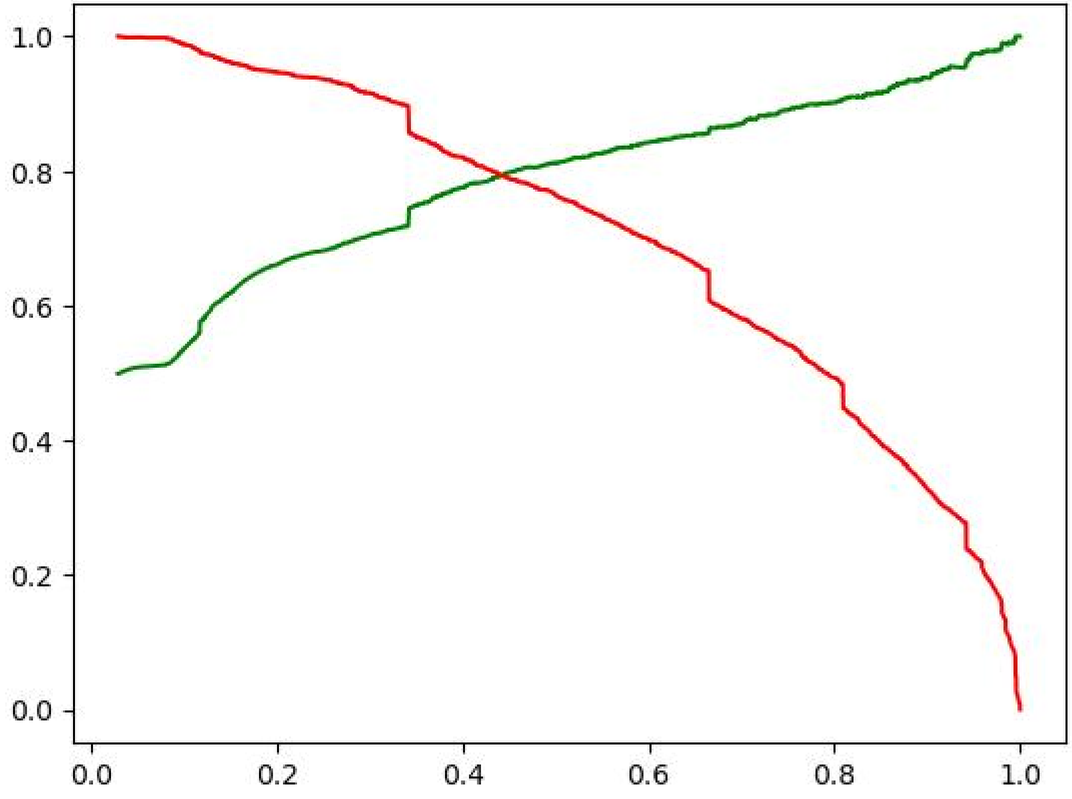
The area under the curve of the ROC is 0.88 which is quite good, so it seems to have a good model.

Optimal Cutoff

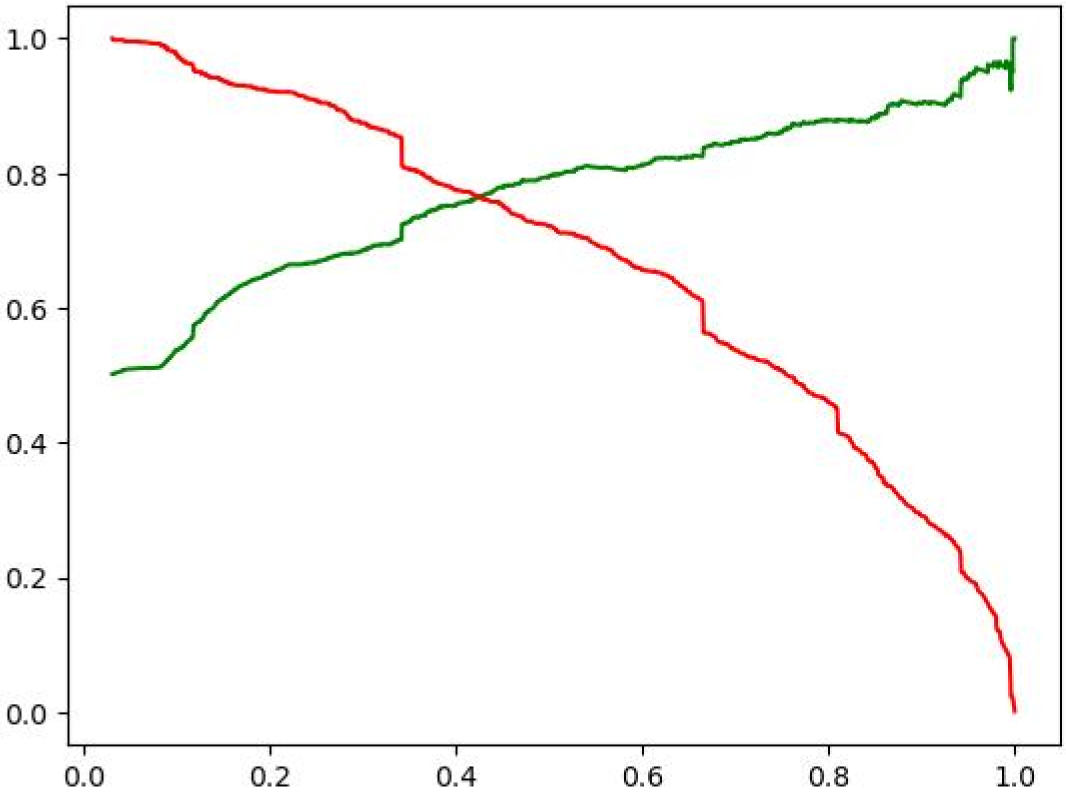


From the above curve, 0.45 is the optimum point to take it as a cutoff probability

Metrics-Precision and Recall



Precision and metrics for the test set



Summary

- Although we have checked both sensitivity-specificity as well as precision and recall metrics well, we have considered the:
- Optimal cutoff based on sensitivity and specificity for calculating the final prediction.
- Accuracy, sensitivity and specificity values of the test set are around 77%, 75% and 79%, which is approximately closer to the respective values calculated using trained set.
- Likewise the lead score calculated in the trained set of data shows the conversion rate on the final prediction model is around 79%.

Hence the overall model seems to be good.