LEAD SCORE CASE STUDY

Submitted by-

Vipul Jadhav

Rohit Venkat

Lead Case Study for X Education

- **Problem Statement**

- X Education sells online courses to industry professionals. The company markets its course on several websites and search engines.

- Once this people land on the websites they might browse a course or fill up the form or watch some videos. When they fill up a form they are classified as lead. Moreover they also get leads from past referrals.

- Once this leads are acquired the sales people call them , writes emails, etc. While doing this some leads gets converted but most of them gets lost. The overall conversion rate is 30%

Lead Case Study for X Education

- **Business Goals**

- X Education needs help in selecting the most promising leads that are most likely to get converted

- The company needs a model where a lead score is assigned to a lead such that a customer with higher lead score is has a higher chance of conversion rate

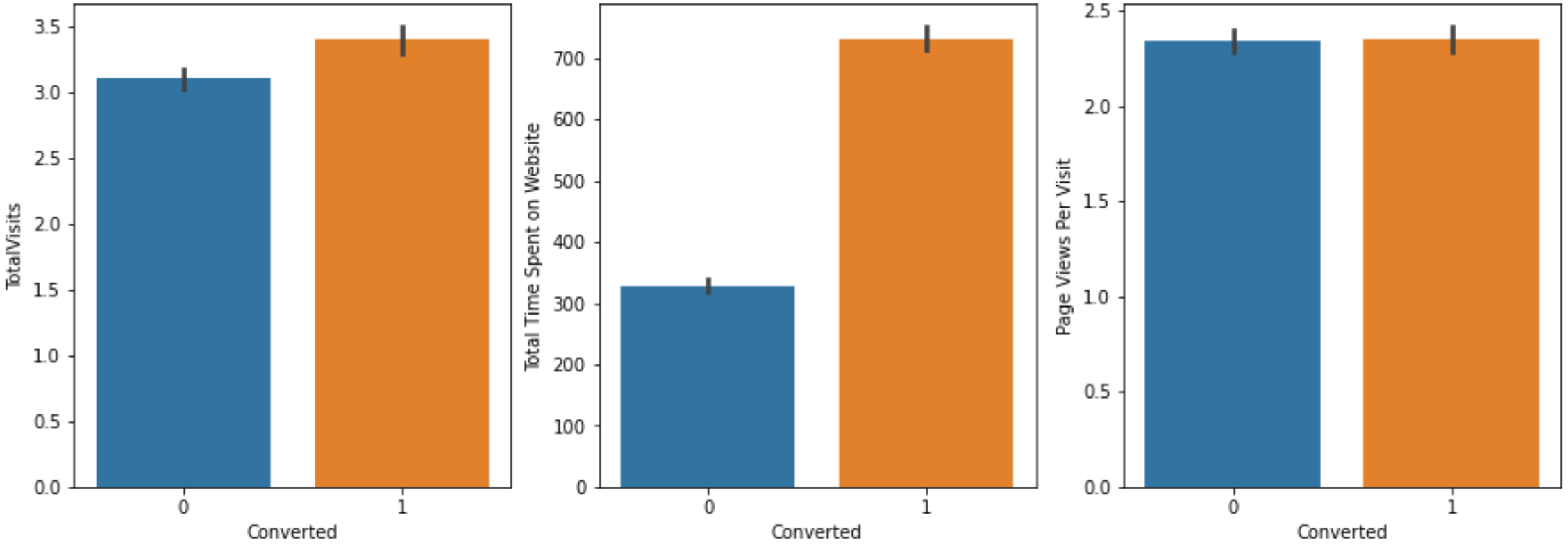
- The lead conversion should be 80%

Plan of Action

- Source the data for Analysis
- Clean and Prepare the Data
- Exploratory Data Analysis
- Feature Scaling
- Splitting the Data into Test and Train dataset
- Building the Logistic Regression Model
- Evaluating the model
- Applying the Best model

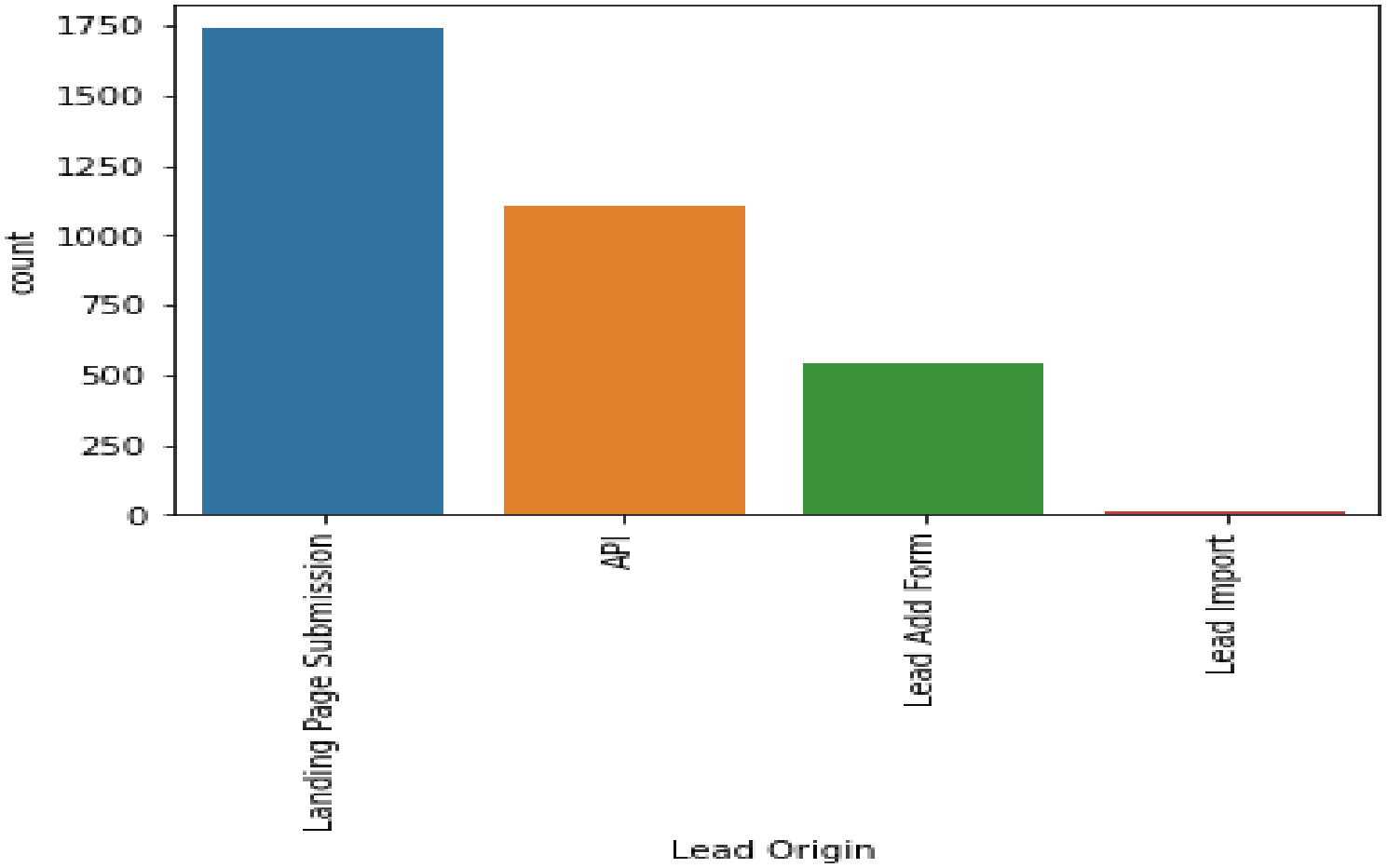
Exploratory Data Analysis

- Conversion from Total Visits, Total Time spent on Website, Total Views per Site (0-Not Converted, 1 – Converted)



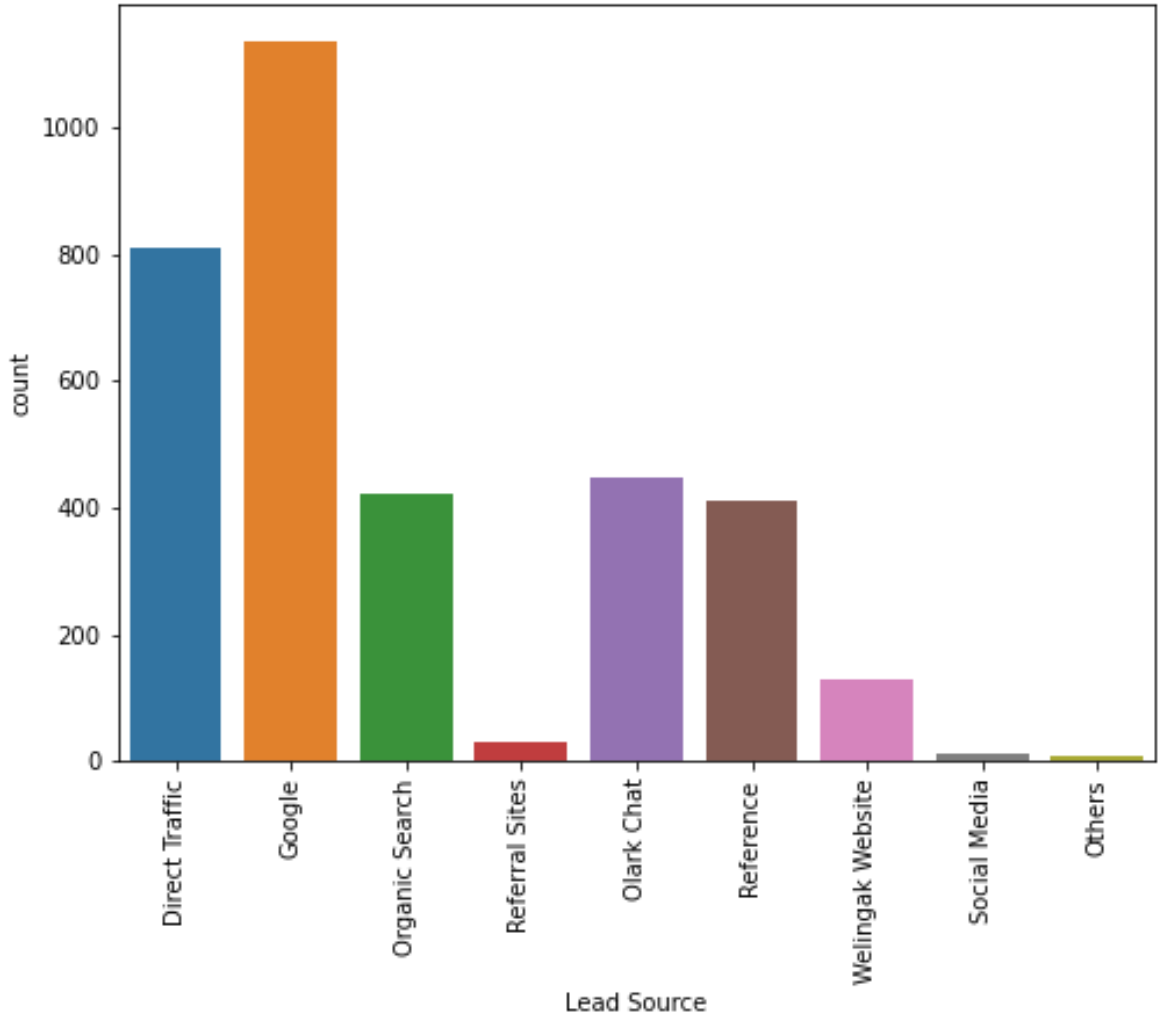
Exploratory Data Analysis

- For Lead Origin maximum Converted happened in landing Page Submission



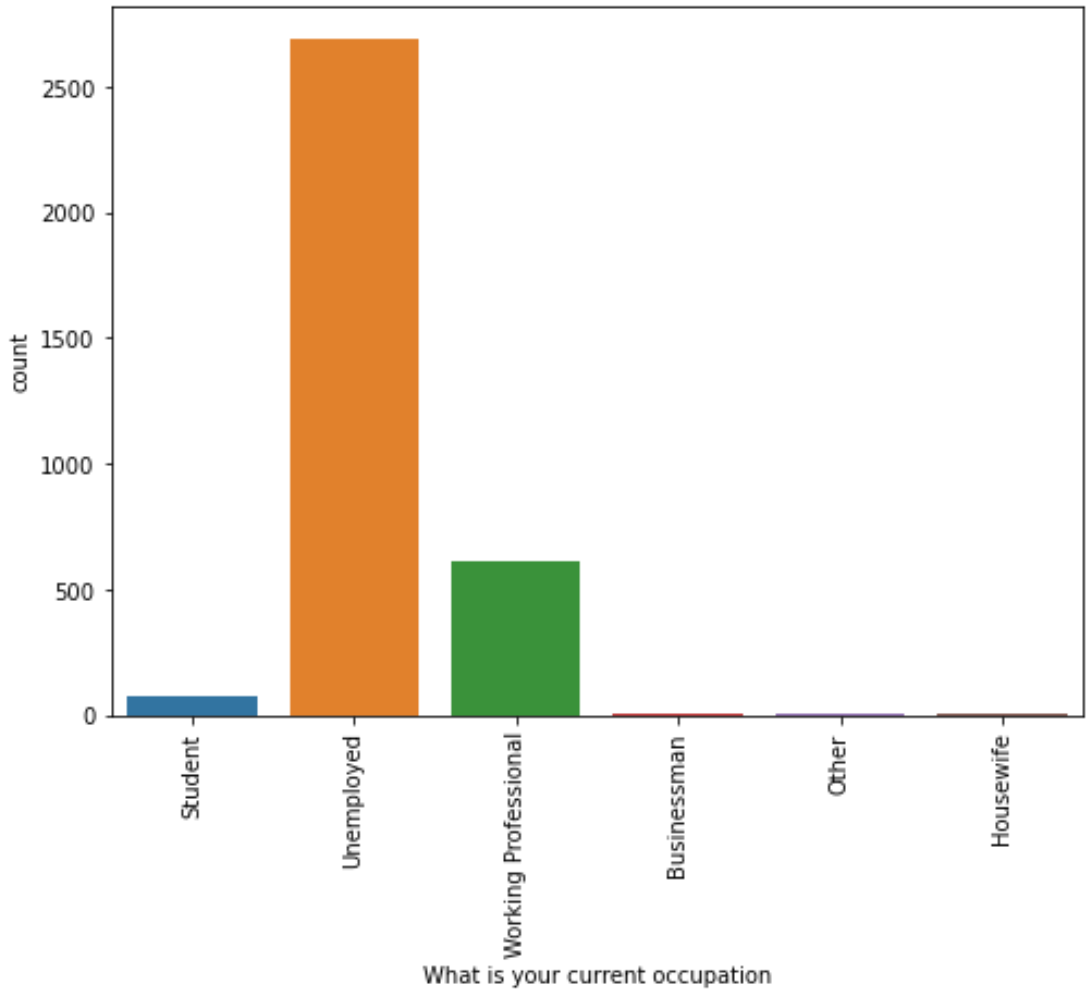
Exploratory Data Analysis

- Major Conversion in Lead Source is from Google



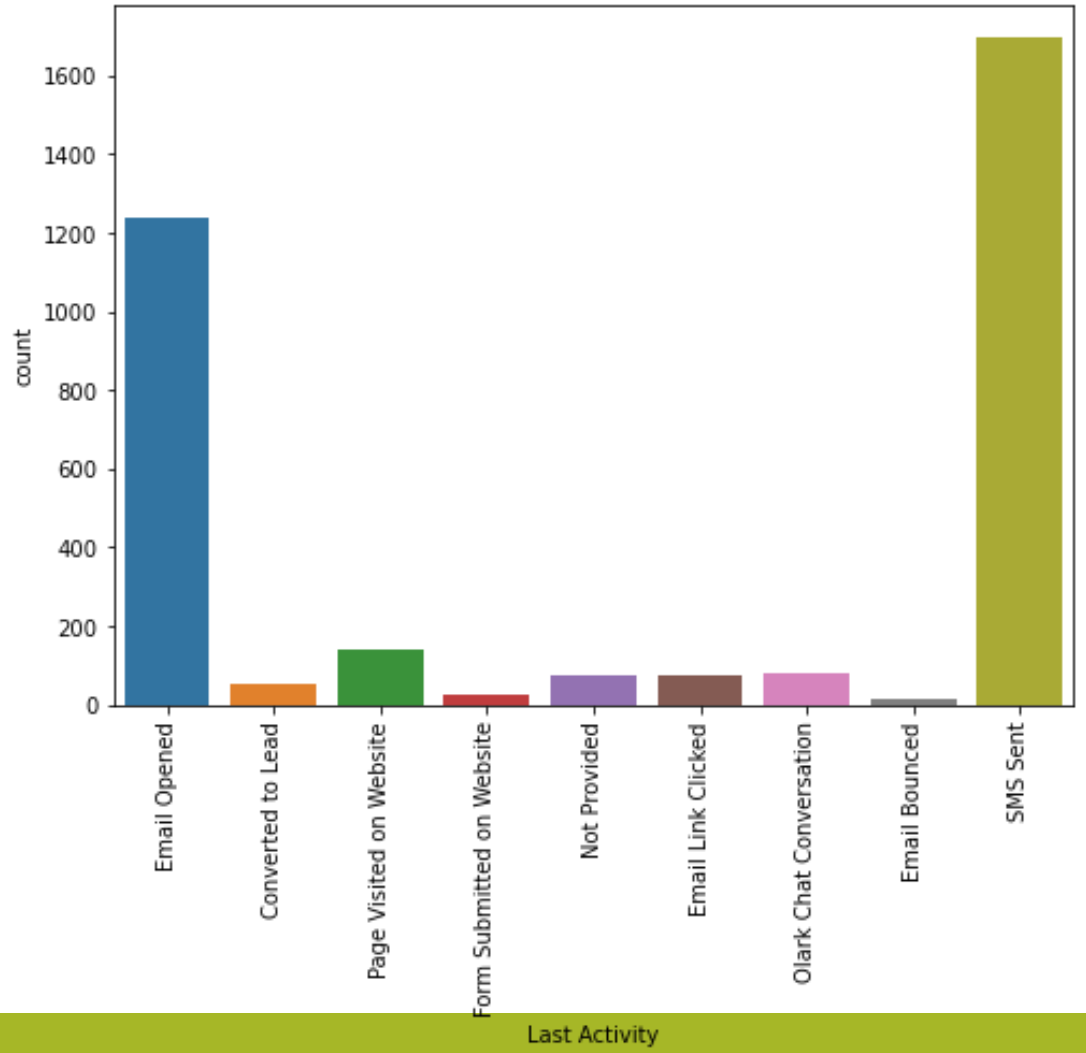
Exploratory Data Analysis

- Major Conversion happened with people who are unemployed



Exploratory Data Analysis

- Last Activity as SMS sent has more conversion



Variables Impacting the Conversion Rate

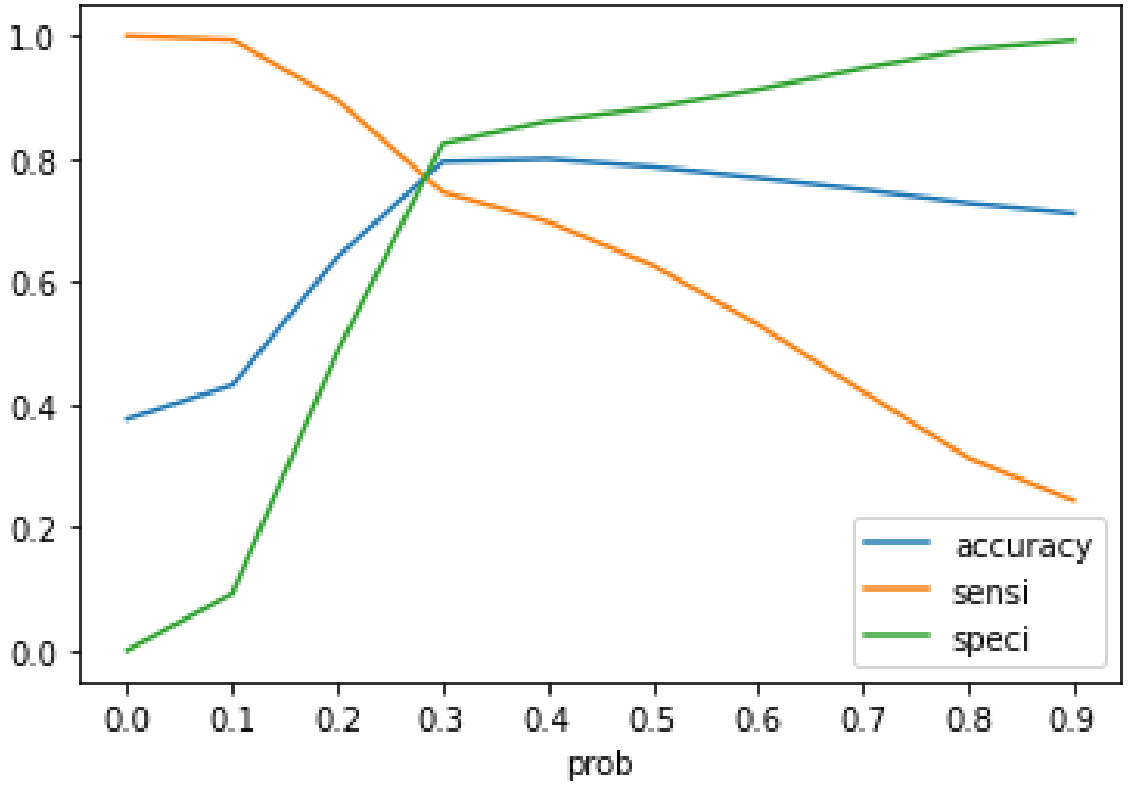
- Lead Origin_Lead Add Form
- Lead Source_Reference
- Lead Source_Welingak Website
- Country_India
- What is your current occupation_Working Professional
- Total Time spent on Website
- Do Not Email_Yes
- Specialization_Banking, Investment And Insurance
- Specialization_Travel and Tourism
- Specialization_Media and Marketing
- Country_Outside India
- Specialization_Ecommerce
- Specialization_Rural and Agribusiness
- What is your current occupation_Housewife

Variables Selected for the Model

- Lead Source_Reference
- Country_India
- What is your current occupation_Working Professional
- Total Time spent on Website
- Do Not Email_Yes
- Specialization_Banking, Investment And Insurance

Model Evaluation – Sensitivity and Specificity on Train Data Set

- The Graph depicts an optimal cut off of 0.30 based on Accuracy, Sensitivity and Specificity

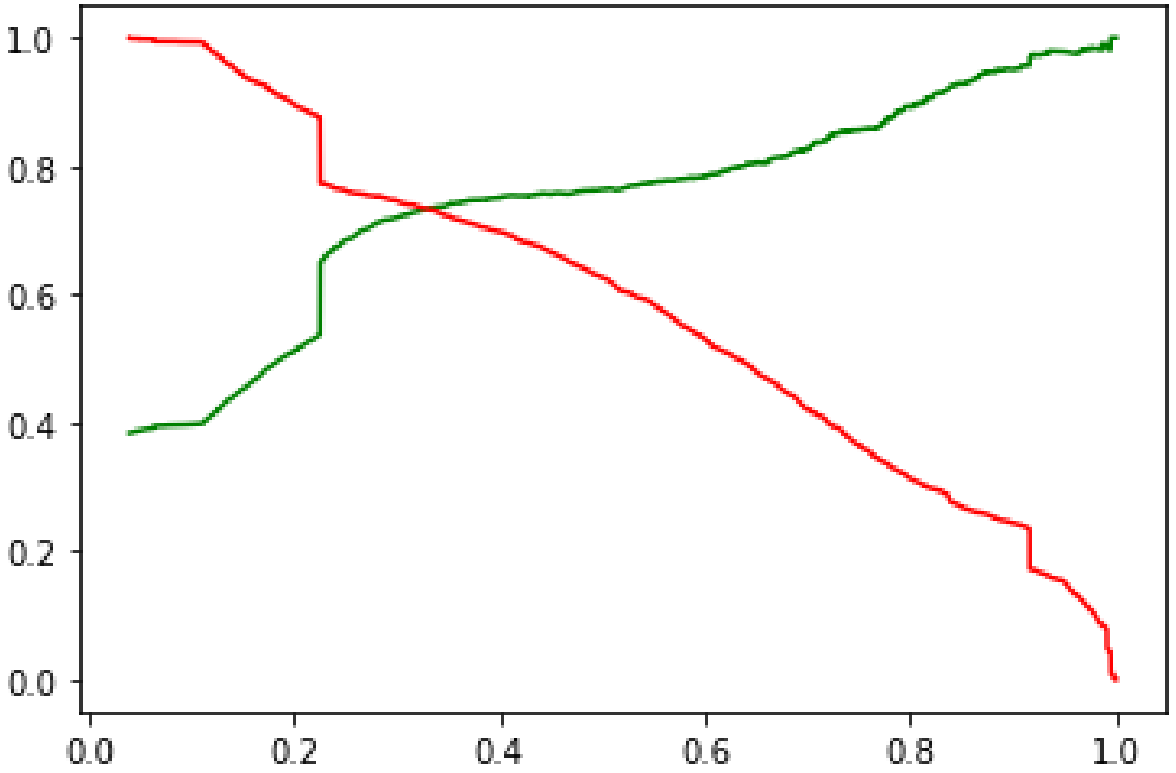


Model Evaluation – Sensitivity and Specificity on Train Data Set

- Accuracy – 80.02%
- Sensitivity –72.15%
- Specificity – 84.76%
- False Positive Rate –15.23%
- Positive Predictive Rate – 74.06%
- Negative Predictive Value – 83.47%

Model Evaluation – Precision and Recall on Train Dataset

- The Graph depicts an optimal cut off (0.3) of based on Precision and Recall



Model Evaluation –Precision and Recall on Train Data Set

- Confusion Matrix

3328	598
659	1708

- Precision – 74.07%
- Recall – 72.16%

Model Evaluation – Sensitivity and Specificity on Test Data Set

- Confusion Matrix

1389	277
265	767

- Accuracy – 79.92%
- Sensitivity –74.32%
- Specificity – 83.38%

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision-Recall , we have considered the optimal cut off

- Accuracy, Sensitivity and Specificity of test set are 80%, 74% and 83% which are closer to train set

- The top 3 variable that contribute for lead getting converted are
  - Lead Source_Reference
  - Country_India
  - What is your current occupation_Working Professional

- The overall model seems to be good