

Lead Scoring Case Study using logistic regression

SUBMITTED BY : RAJESH BAIRANGULA

Contents

- Problem statement
- Problem approach
- EDA
- Model Evaluation
- Observations
- Conclusion

Problem Statement

- An education company named X Education sells online courses to industry professionals.
- On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around 30%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Business Objective

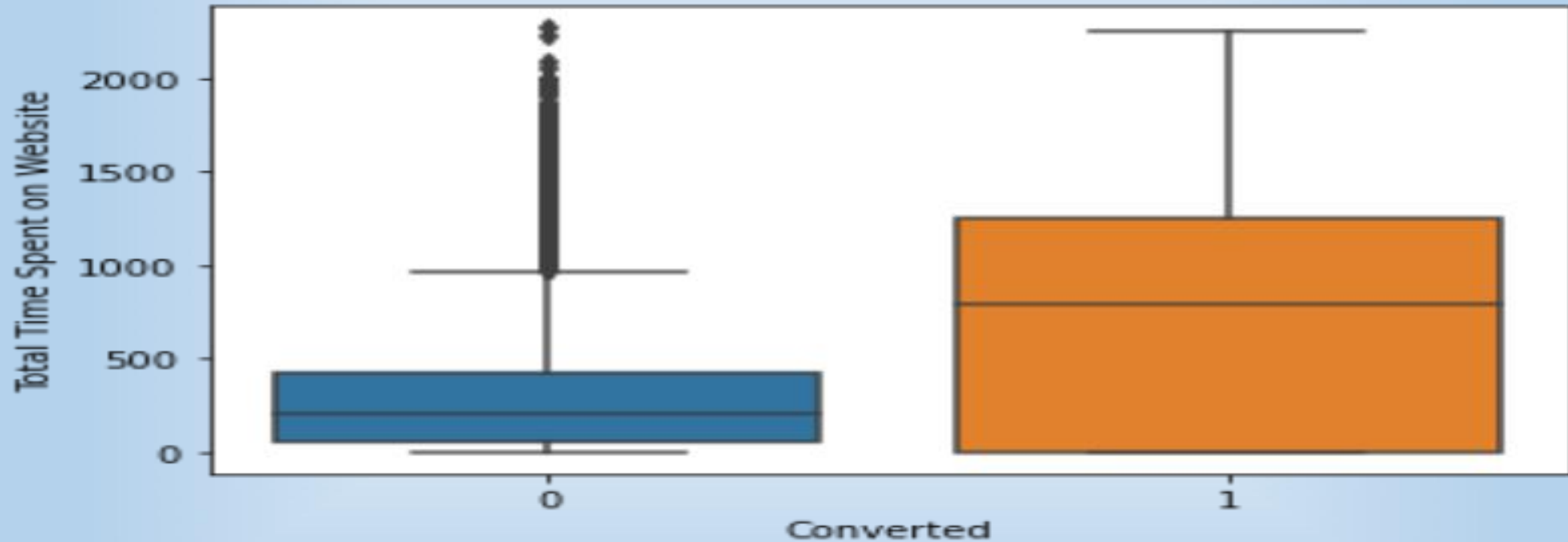
- Lead X wants us to build a model to give every lead a lead score between 0 -100 .
- So that they can identify the Hot leads and increase their conversion rate as well.
- The CEO want to achieve a lead conversion rate of 80%.
- They want the model to be able to handle future constraints as well like Peak time
- actions required, how to utilize full man power and after achieving target what
- should be the approaches

Problem Approach

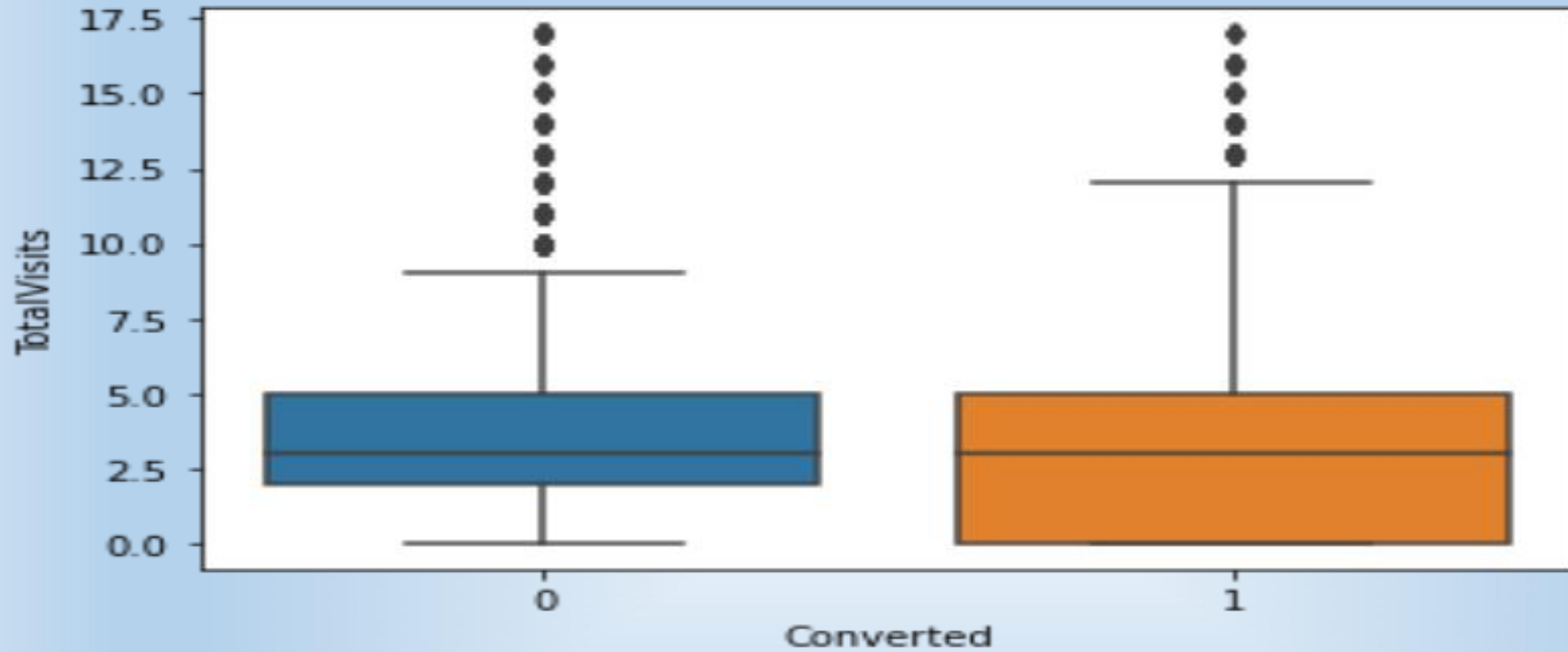
- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE Rsquared VIF and pvalues)
- Model Evaluation
- Making predictions on test set

DATA VISUALIZATION

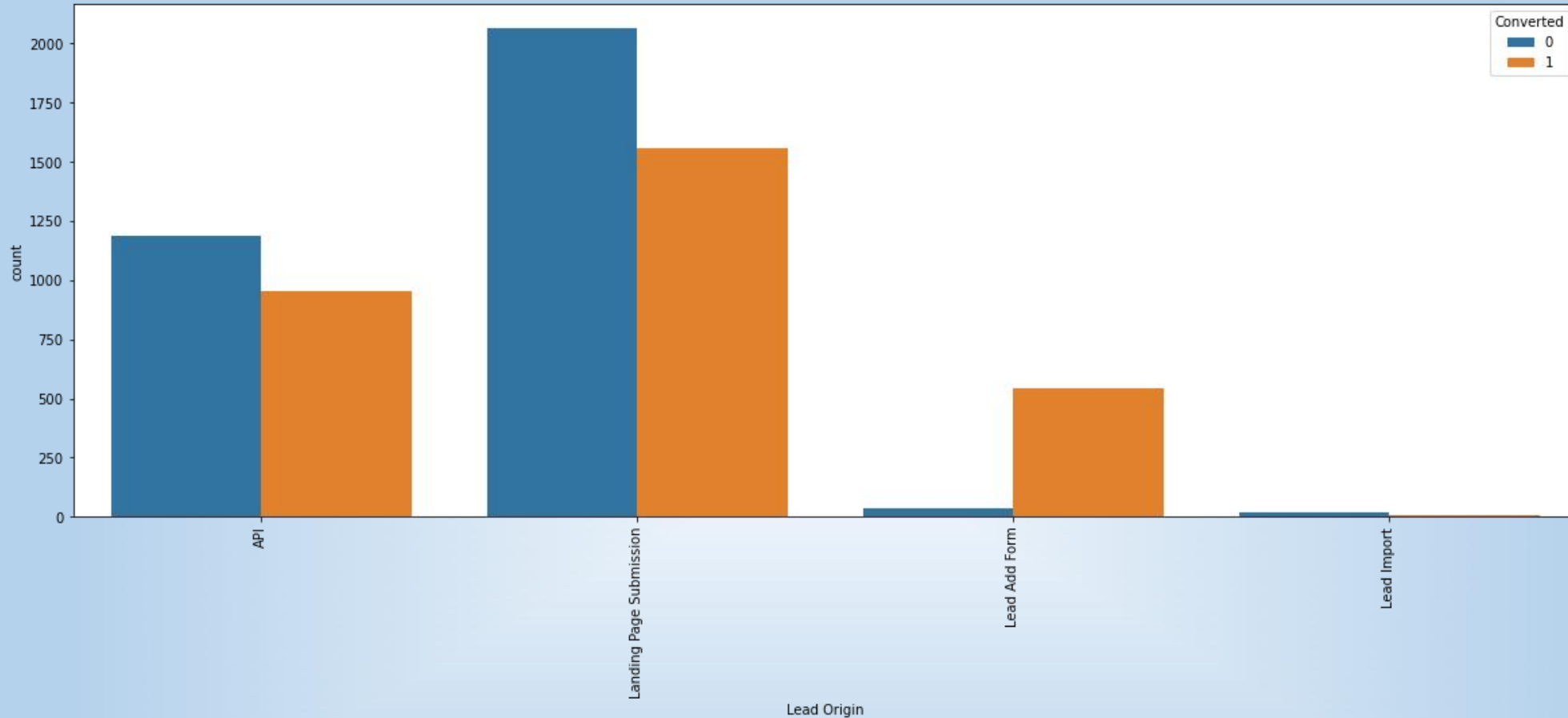
- To identify important features
- To get insights



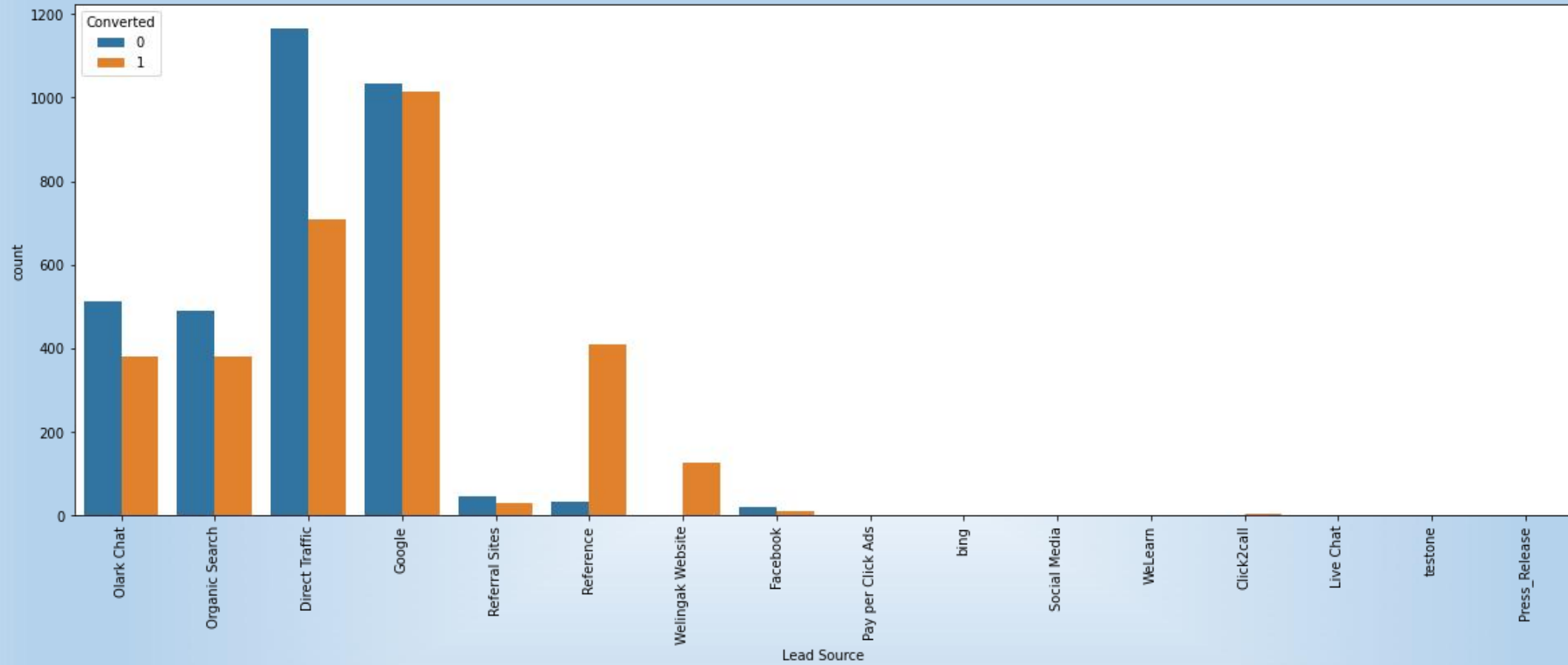
Leads spending more time on the website are more likely to be converted.
Website should be made more engaging to make leads spend more time.



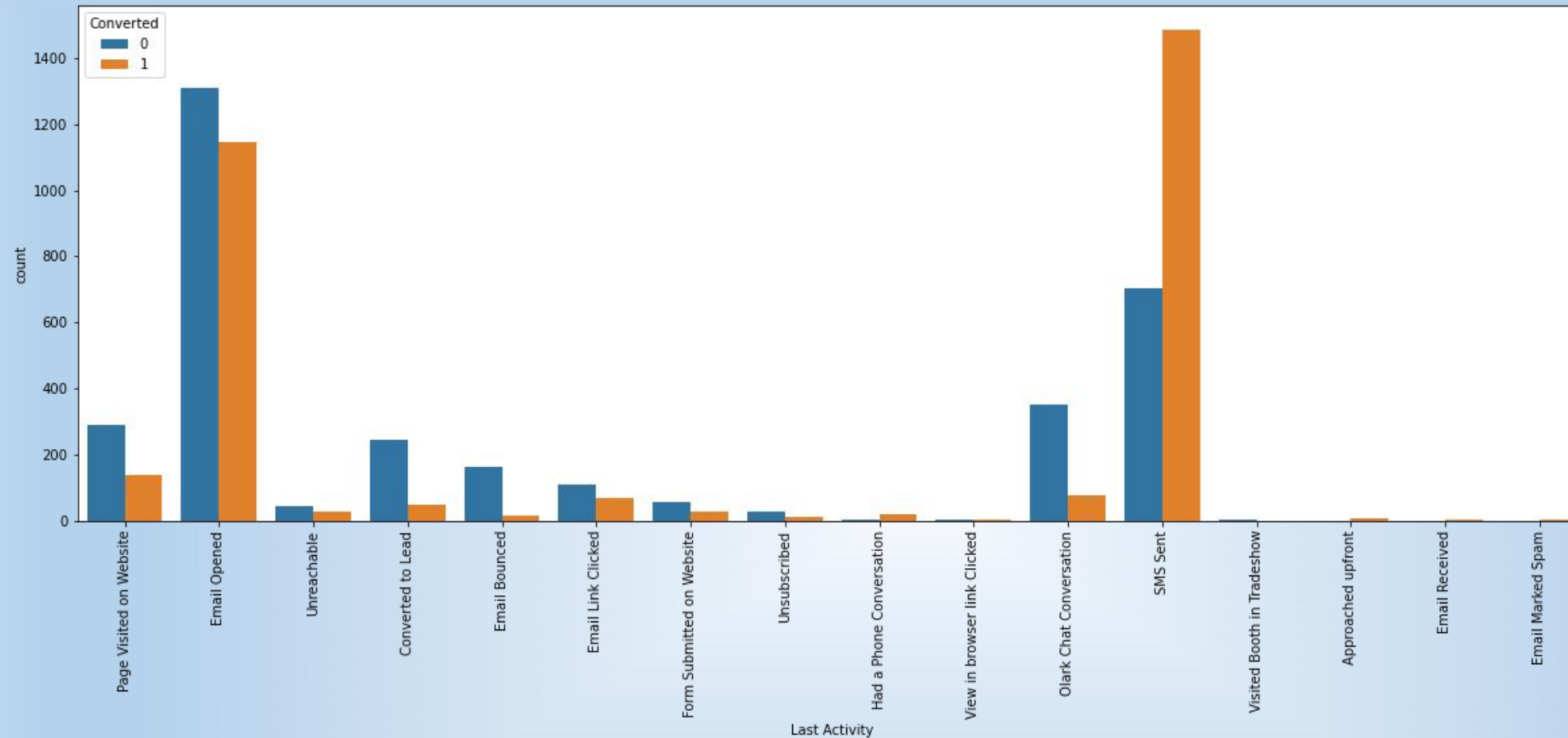
Median for converted and not converted leads are the close.
Nothing conclusive can be said on the basis of Total Visits



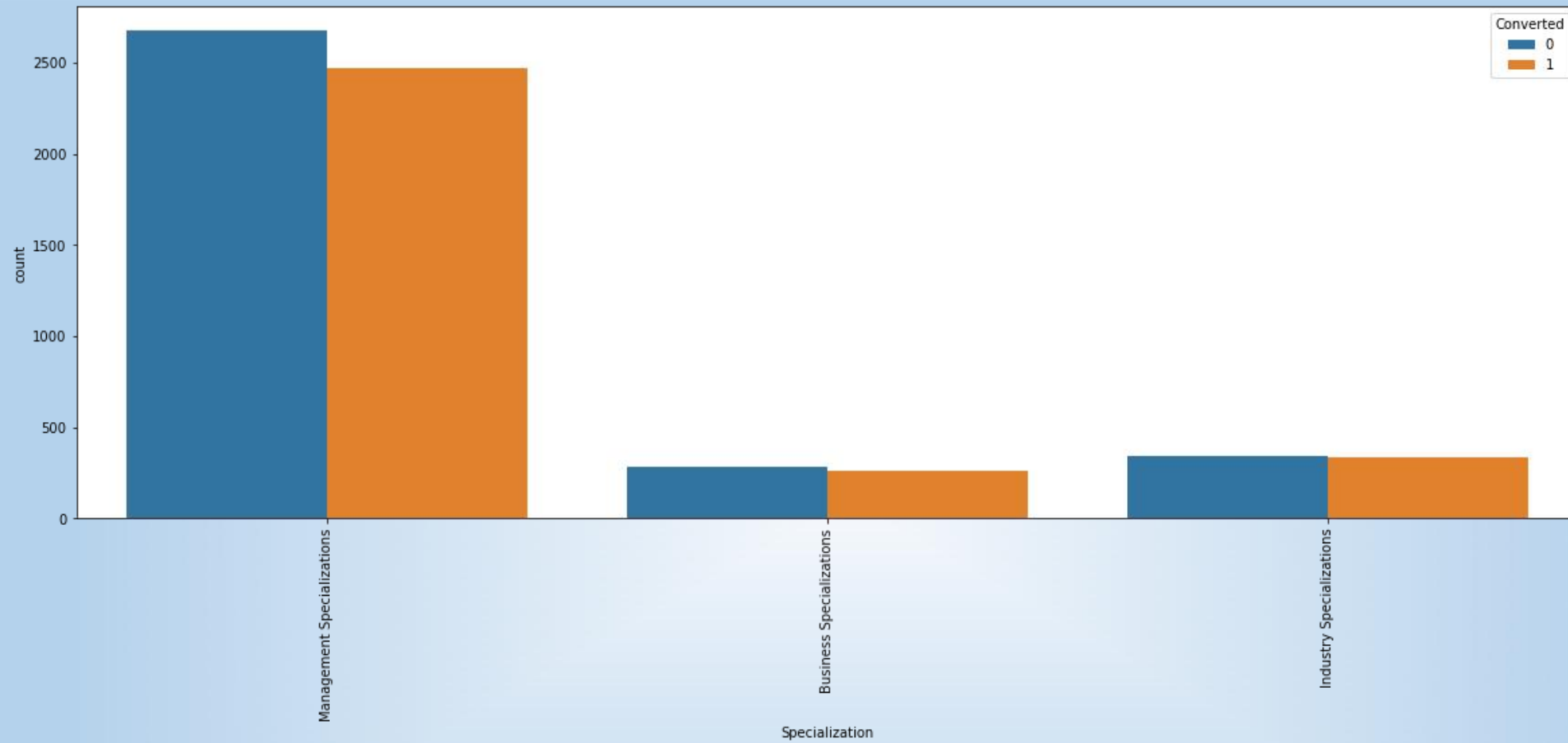
'API' and 'Landing Page Submission' generate the most leads but have less conversion rates of around 30%. Whereas, 'Lead Add Form' generates less leads but conversion rate is great. We should try to increase conversion rate for 'API' and 'Landing Page Submission', and increase leads generation using 'Lead Add Form'. 'Lead Import' does not seem very significant.



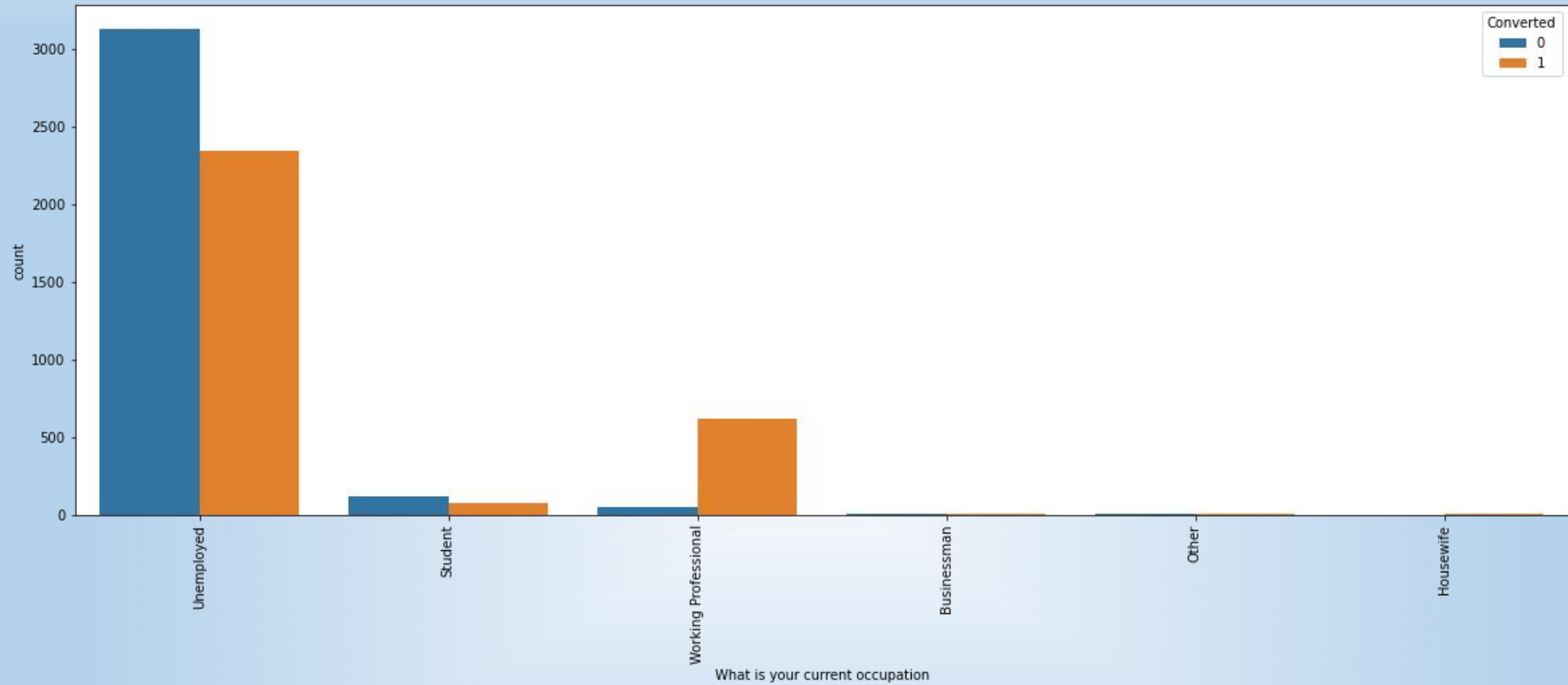
'Direct Traffic' and 'Google' generate maximum number of leads while maximum conversion rate is achieved through 'Reference' and 'Welingak Website'.



Highest number of lead are generated where the last activity is 'Email Opened' while maximum conversion rate is for the activity of 'SMS Sent'. Its conversion rate is significantly high. Categories after the 'SMS Sent' have almost negligible effect. We can aggregate them all in one single category



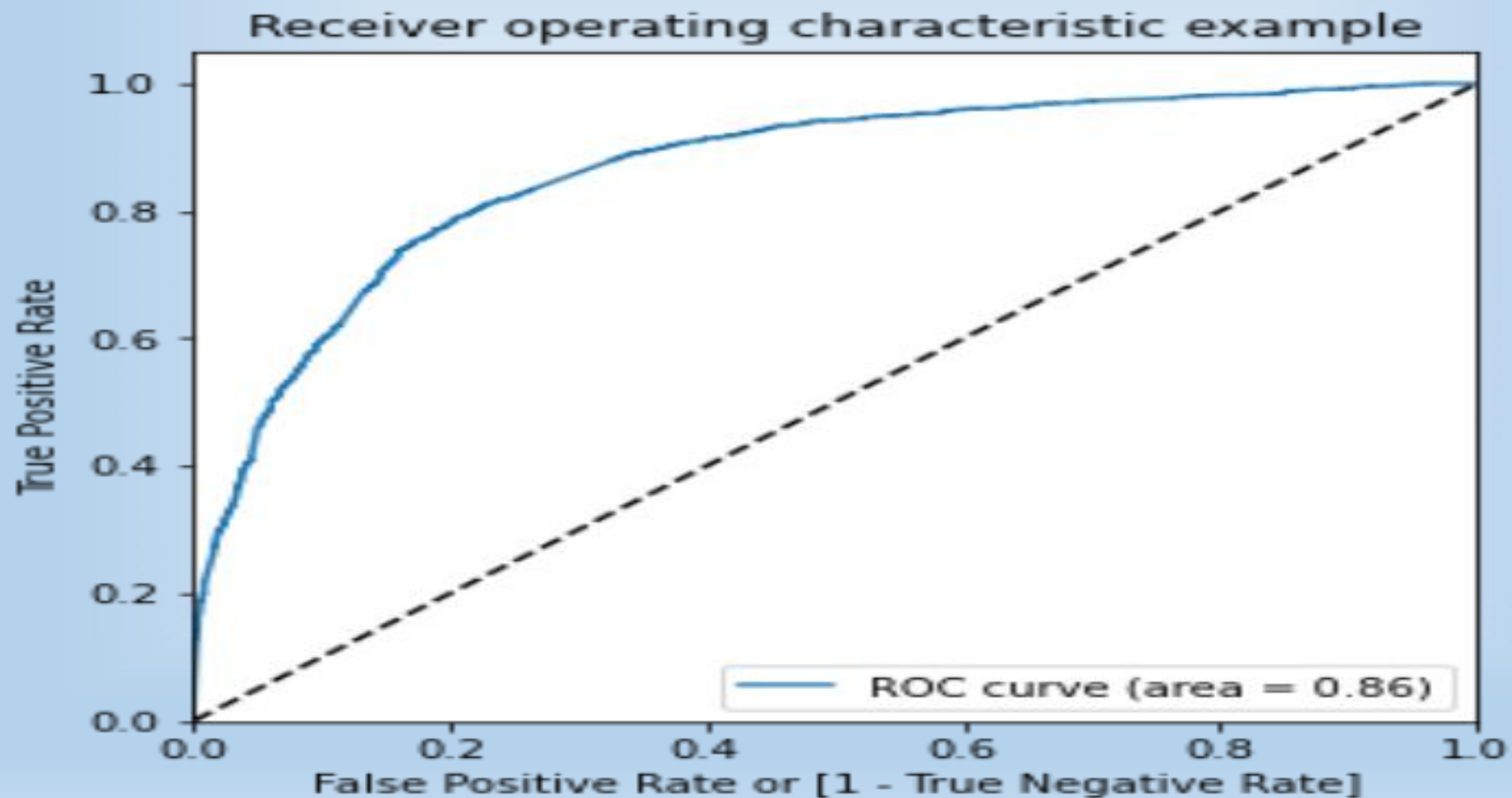
The highest conversion rate is for 'Management Specialization' category.



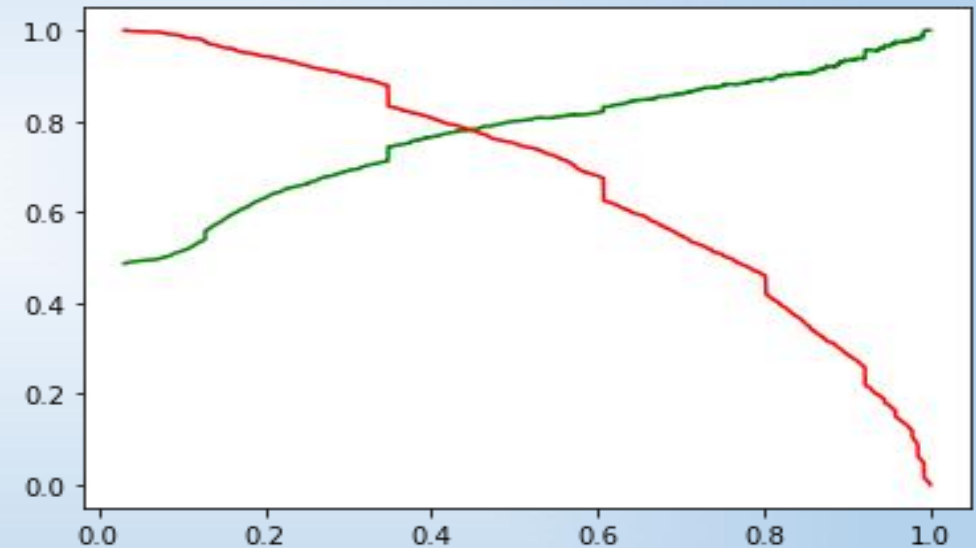
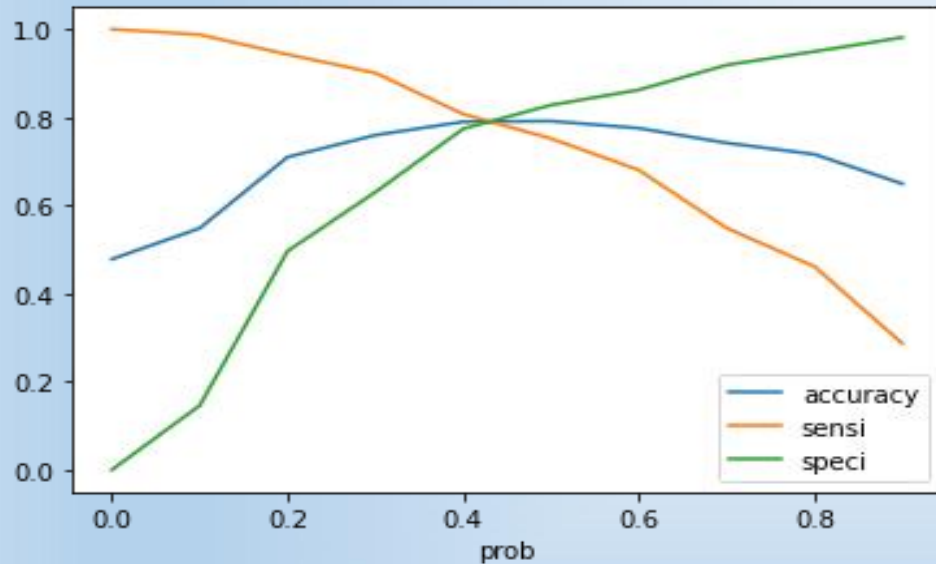
The highest conversion rate is for 'Working Professional'. High number of leads are generated for 'Unemployed' but conversion rate is low.

Model Evaluation

ROC curve



0.42 is the tradeoff between Precision and Recall -
Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 42 % to be a hot Lead



Observations

- **Train Data:**
- **Accuracy : 79%**
- **Sensitivity : 79%**
- **Specificity : 78%**
- **Test Data:**
- **Accuracy : 79%**
- **Sensitivity : 78%**
- **Specificity : 79%**
- **Final Features list:**
- **Lead Source_Welingak Website**
- **Last Activity_Email Bounced**
- **Last Notable Activity_Unreachable**
- **Last Notable Activity_Unsubscribed**
- **Lead Origin_Lead Add Form**
- **What is your current occupation_Working Profes.**
- **Last Activity_SMS Sent**
- **Lead Source_Organic Search**
- **Last Activity_Olark Chat Conversation**
- **Lead Source_Direct Traffic**

Conclusion

- **We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore**
- **we can intervene that we need to focus more on the leads originated from API and Landing page submission.**
- **We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.**
- **Leads who spent more time on website, more likely to convert.**
- **Most common last activity is email opened. highest rate = SMS Sent.**
- **Max are unemployed. Max conversion with working professional.**