LEAD SCORE CASE STUDY

Submitted By:
Tanzeela Ali
Neha
Bharat Ramchire
Karthik Ravikanti

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. Moreover, the company also gets leads through past referrals. 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%.

Business Goal

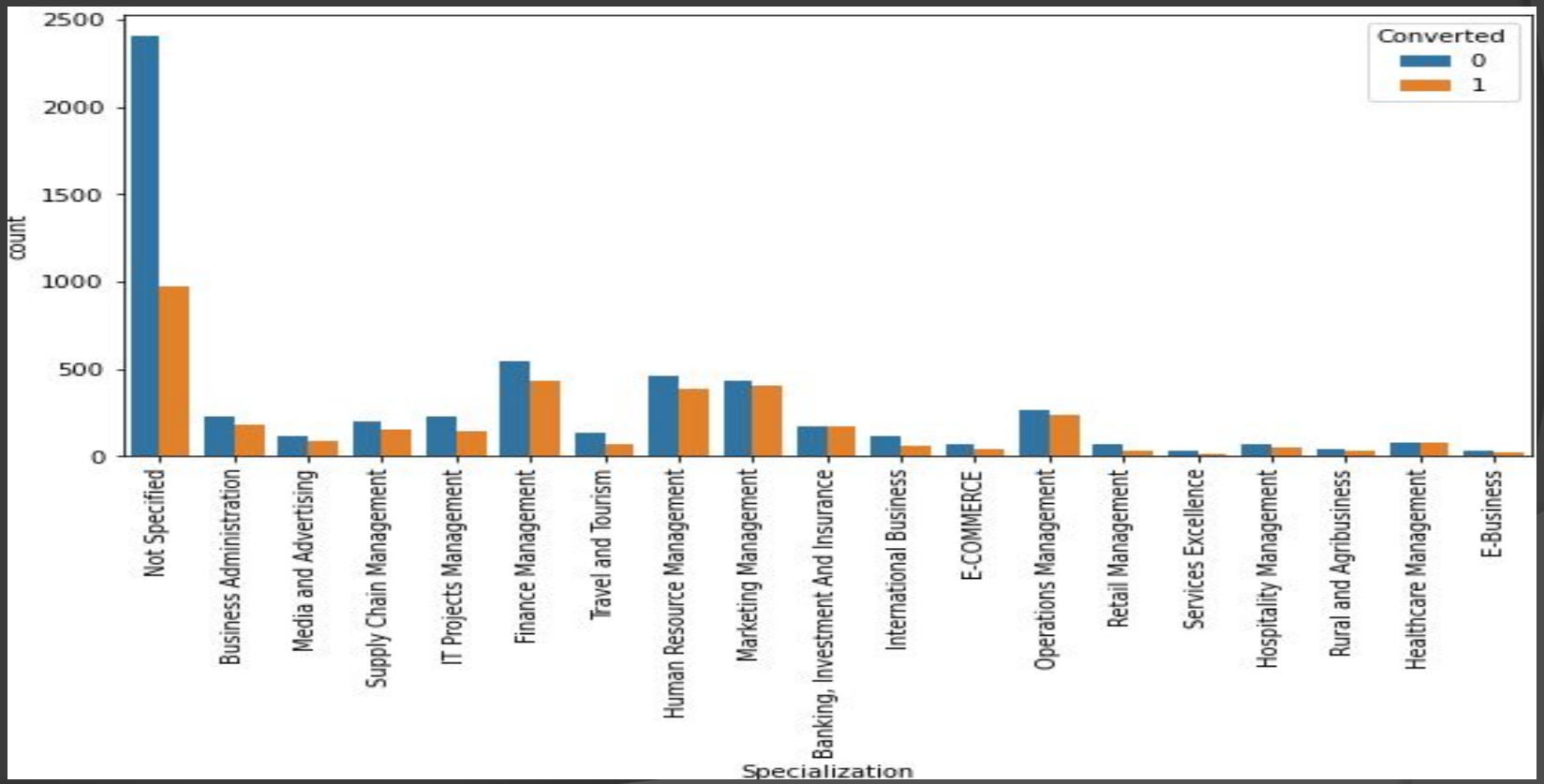
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Steps to Follow

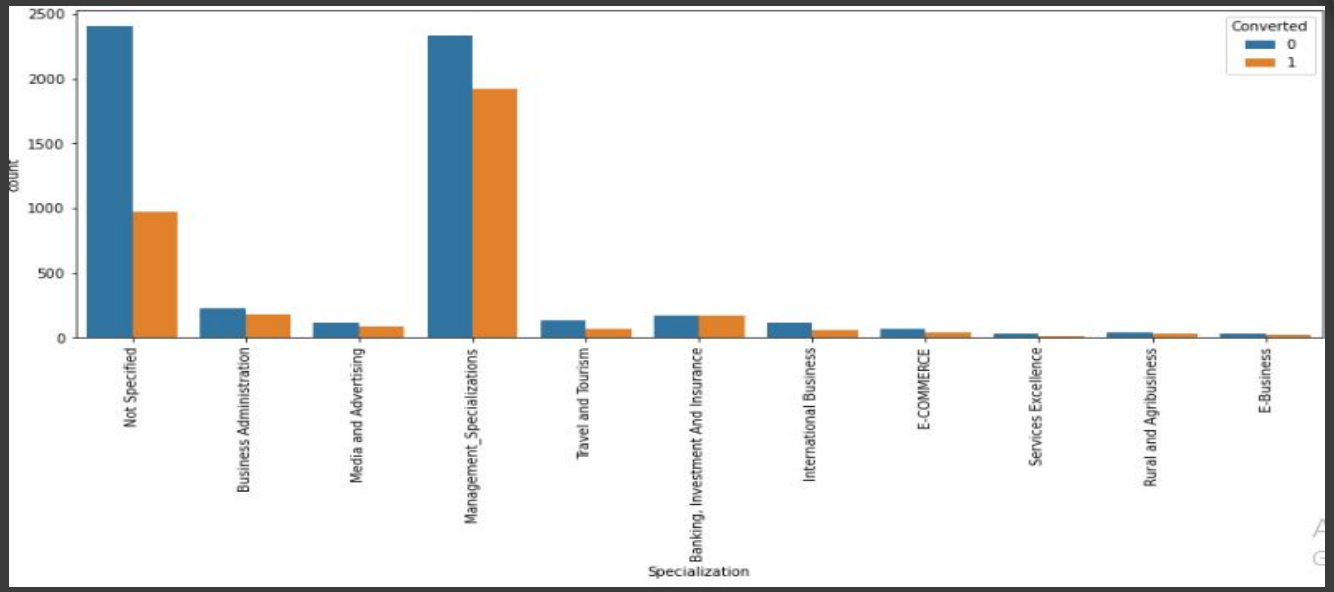
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity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

Exploratory Data Analysis

By the graph We see that specialization with Management are having higher number of leads as well as leads converted. So this is definitely a significant variable

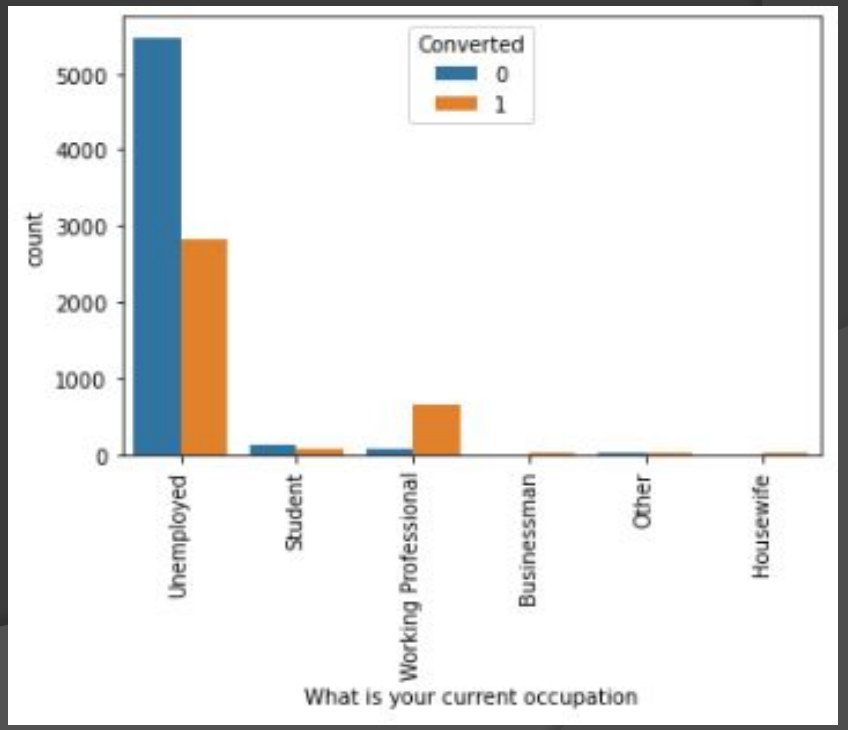


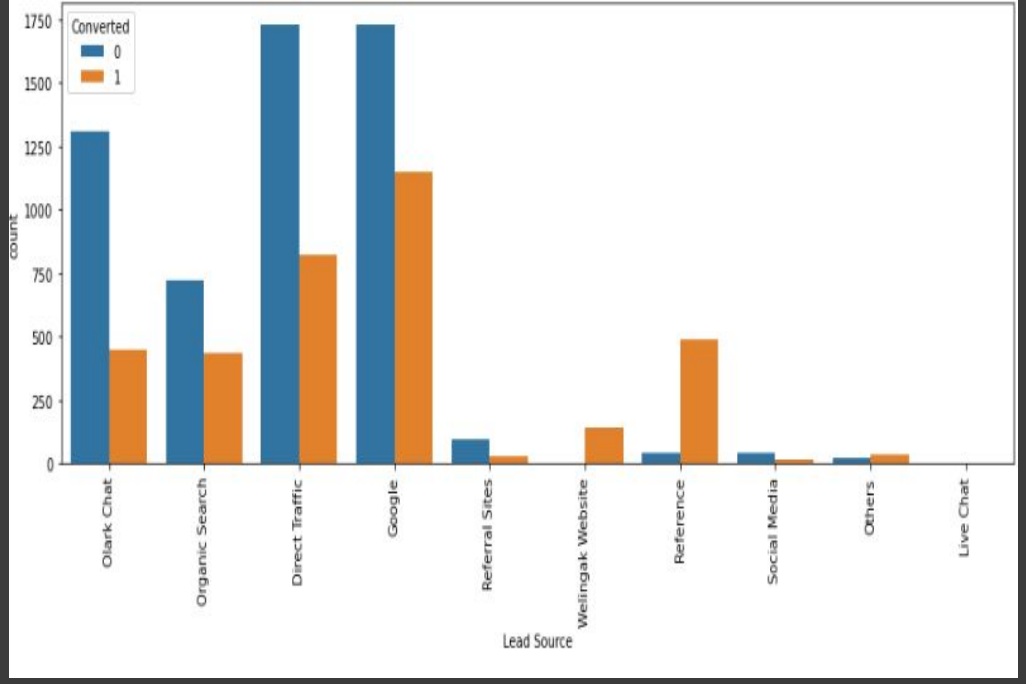
Count of Variable based on Converted value

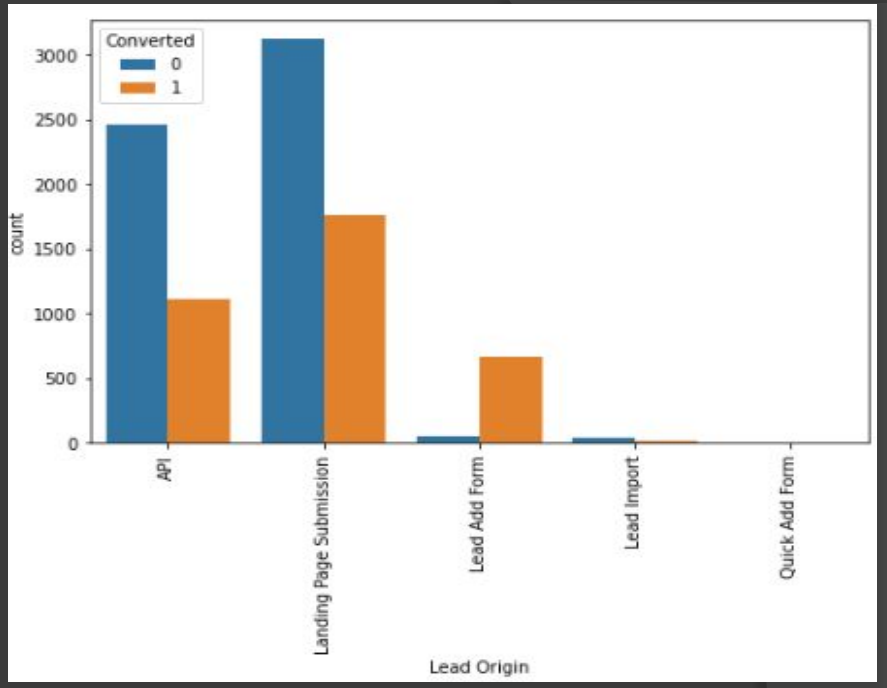


OBSERVATION

Working Professionals going for the course have high chances of joining it. Unemployed leads are the most in terms of Absolute numbers.





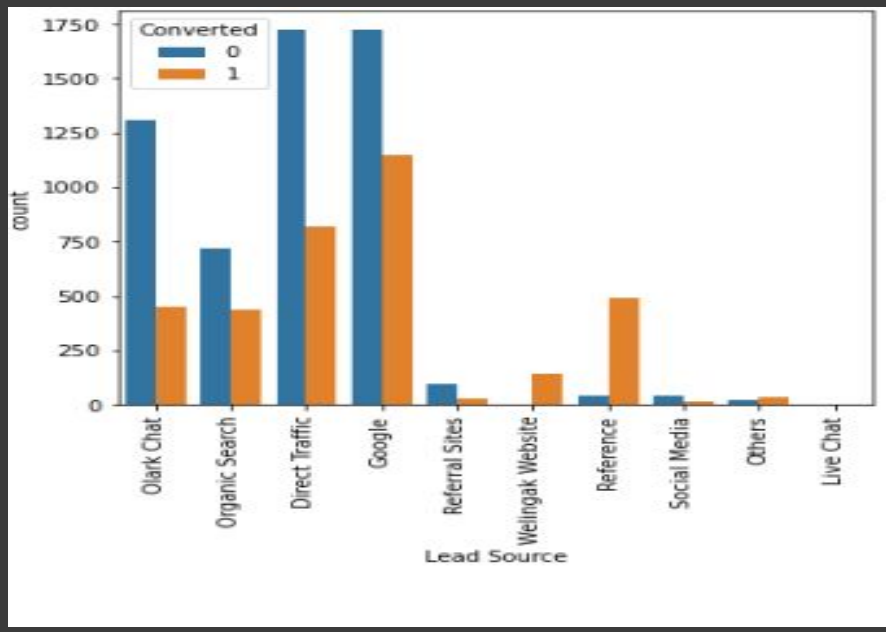


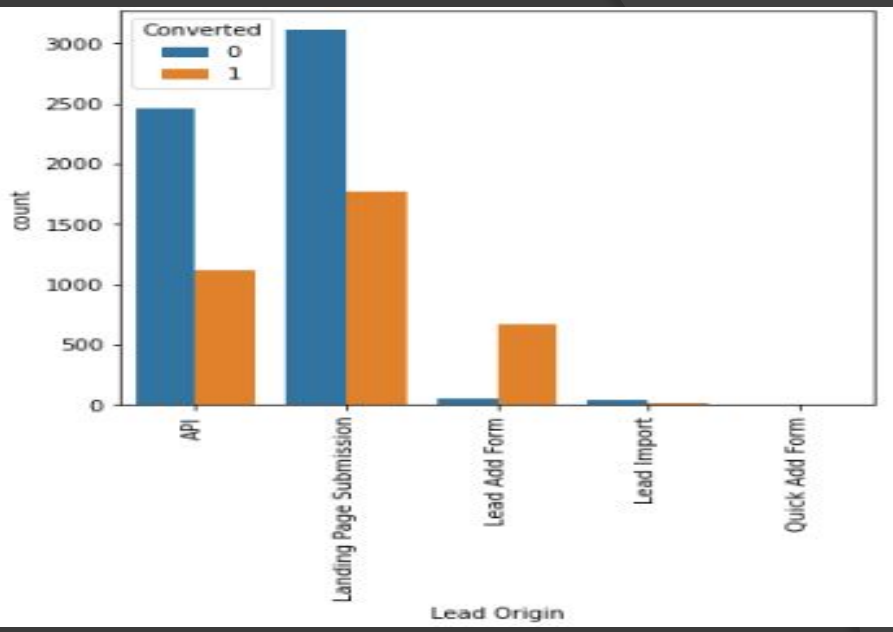
OBSERVATION

- Maximum number of leads are generated by Google and Direct traffic.
- Conversion Rate of reference leads and leads through welingak website is high.
- To improve overall lead conversion rate, focus should be on improving lead converion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.

OBSERVATION

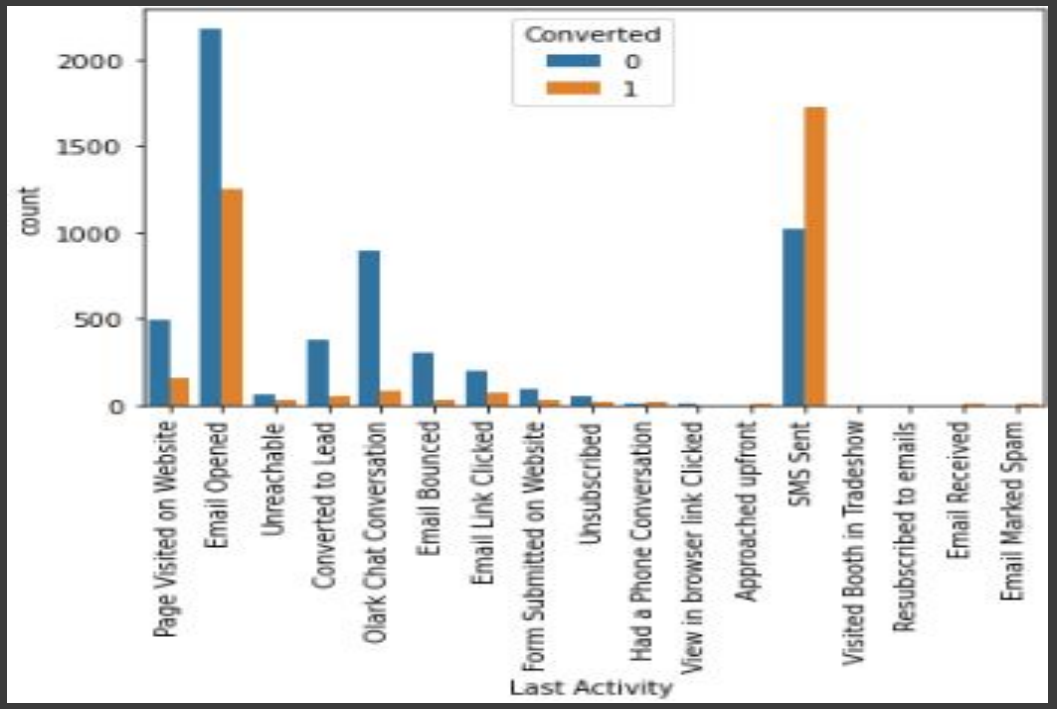
- API and Landing Page Submission bring higher number of leads as well as conversion.
- Lead Add Form has a very high conversion rate but count of leads are not very high.
- Lead Import and Quick Add Form get very few leads.
- In order to improve overall lead conversion rate, we have to improve lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form





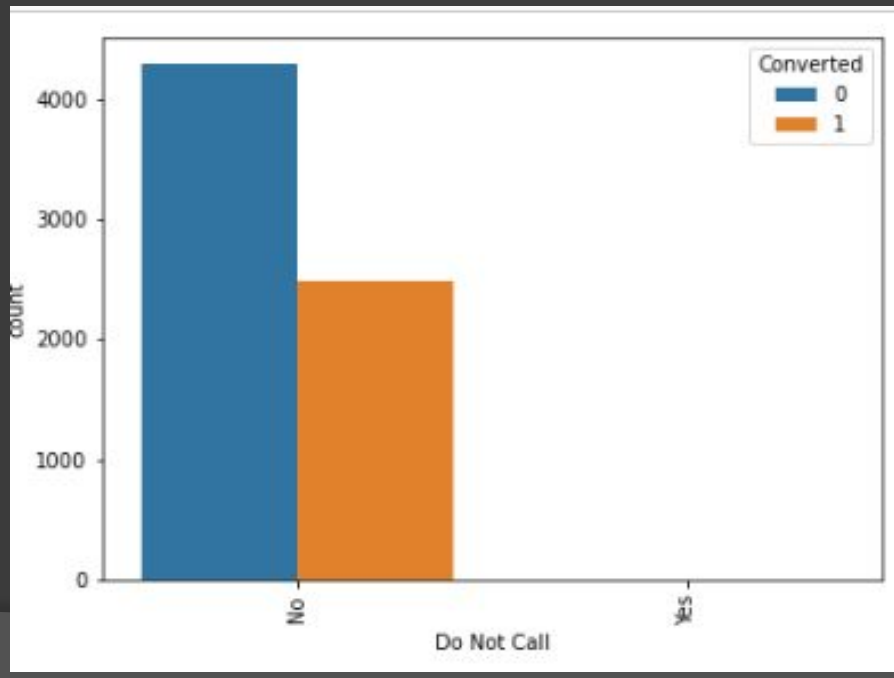
OBSERVATION

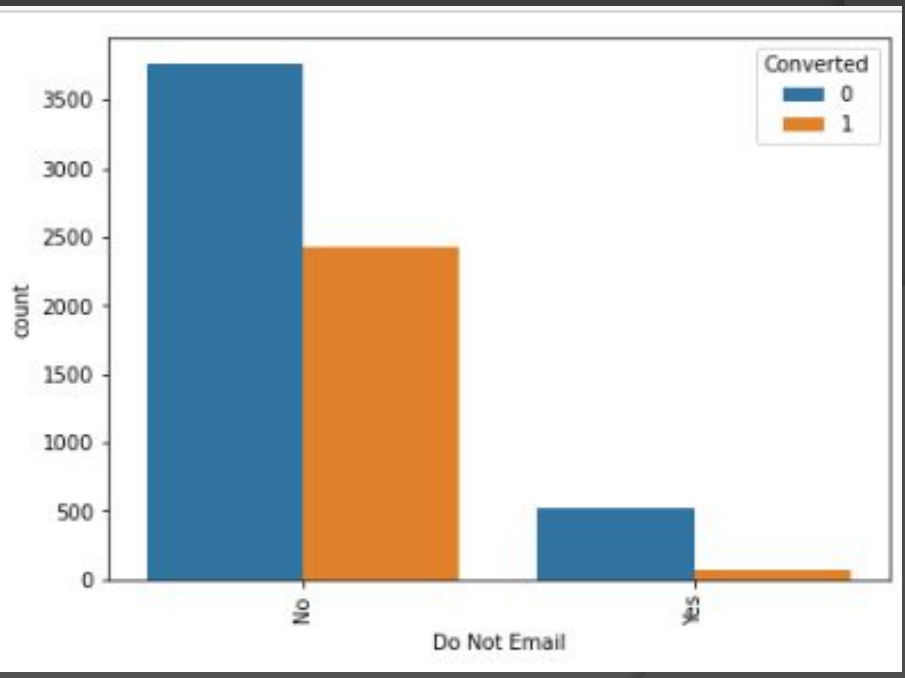
- Despite having a relatively lower conversion rate of approximately 30%, both API and Landing Page Submission generate a substantial number of leads.
- Conversely, the Lead Add Form generates a significantly lower count of leads, yet boasts a notably high conversion rate.
- Lead Import contributes negligibly to both lead count and conversion rate and can be disregarded.
- To enhance the overall lead conversion rate, efforts should be directed towards improving the conversion rates of API and Landing Page Submission, while simultaneously increasing lead generation.



OBSERVATION

- The highest count among last activities is recorded for "Email Opened".
- The maximum conversion rate is observed for the last activity being "SMS Sent".



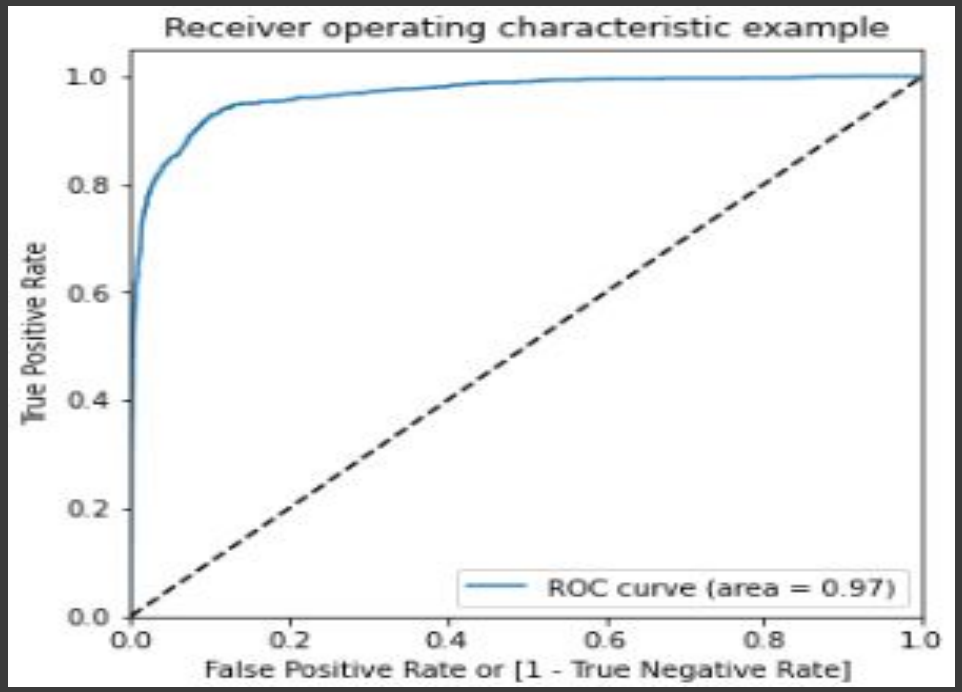


Variables impacting the conversion rate

- Do Not Email
- Total Visits
- Total Time Spent On Website
- Lead Origin – Lead Page Submission
- Lead Origin – Lead Add Form
- Lead Source - Olark Chat
- Last Source – Welingak Website
- Last Activity – Email Bounced
- Last Activity – Not Sure
- Last Activity – Olark Chat Conversation
- Last Activity – SMS Sent • Current Occupation – No Information
- Current Occupation – Working Professional
- Last Notable Activity – Had a Phone Conversation
- Last Notable Activity - Unreachable

Model Building

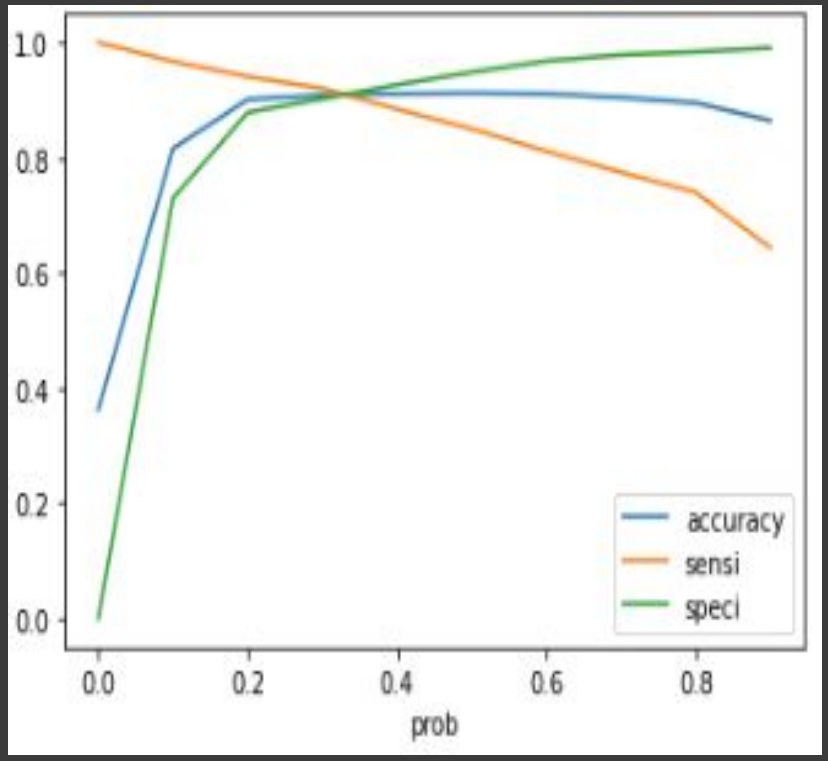
Model Evaluation on Train Data Set



ROC Curve

The ROC Curve should be a value close to 1. We are getting a good value of 0.97 indicating a good predictive model.

Sensitivity and specificity on Train Data Set



OBSERVATION

The ROC curve has a value of 0.97, which is very good. We have the following values for the Train Data:

Accuracy : 90.89%

Sensitivity : 92.05%

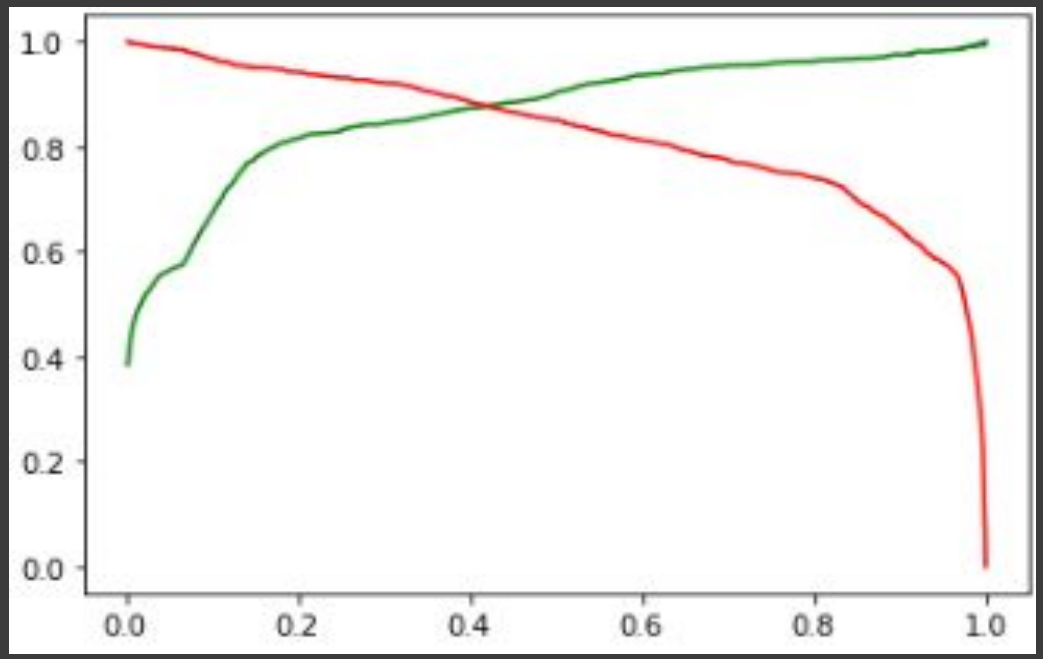
Specificity : 90.23%

False Positive Rate : 9.7%

Positive Predictive Value : 84.31%

Negative Predictive Value : 95.22%

Precision and Recall on Train Data Set



OBSERVATION

Precision : 84.31%

Recall : 92.05%

Model Evaluation on Test Data Set

Sensitivity and specificity on Test Data Set

OBSERVATION

After running the model on the Test Data these are the figures we obtain:

Accuracy : 90.71%

Sensitivity : 91.14%

Specificity : 90.46%

Precision : 84.79%

Recall : 91.14%

Conclusion

- The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model
- The final model has Sensitivity of 0.91, this means the model is able to predict 91% customers out of all the converted customers, (Positive conversion) correctly.
- While we have checked both Sensitivity-Specificity we have considered the for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 90%, 91% and 90% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 90%
- Hence overall this model seems good.