LEAD SCORING CASE STUDY

By Geeta Choudhary

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.

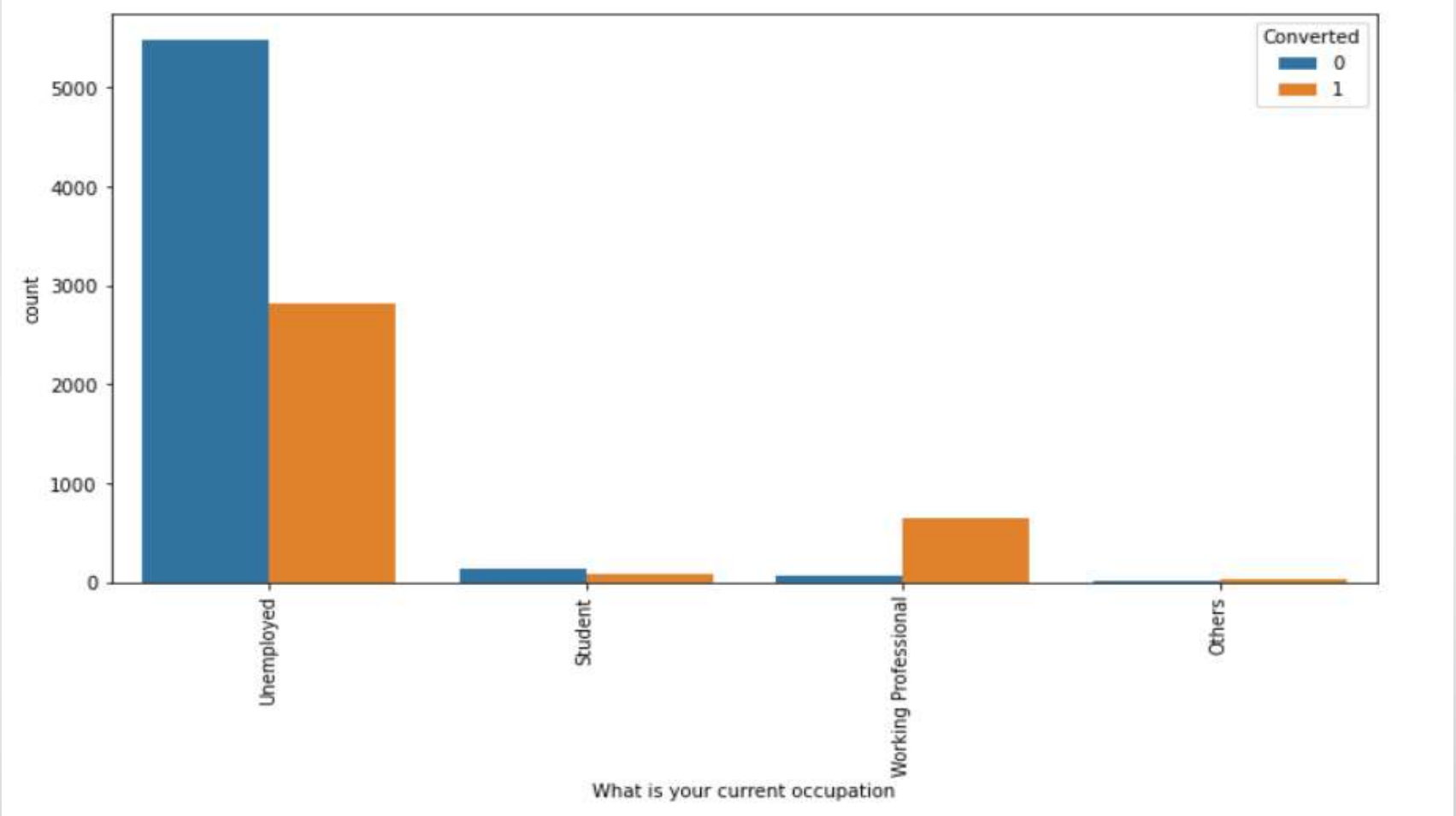
The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified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

There are a lot of leads generated in the initial stage, but only a few of them come out as paying customers. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.

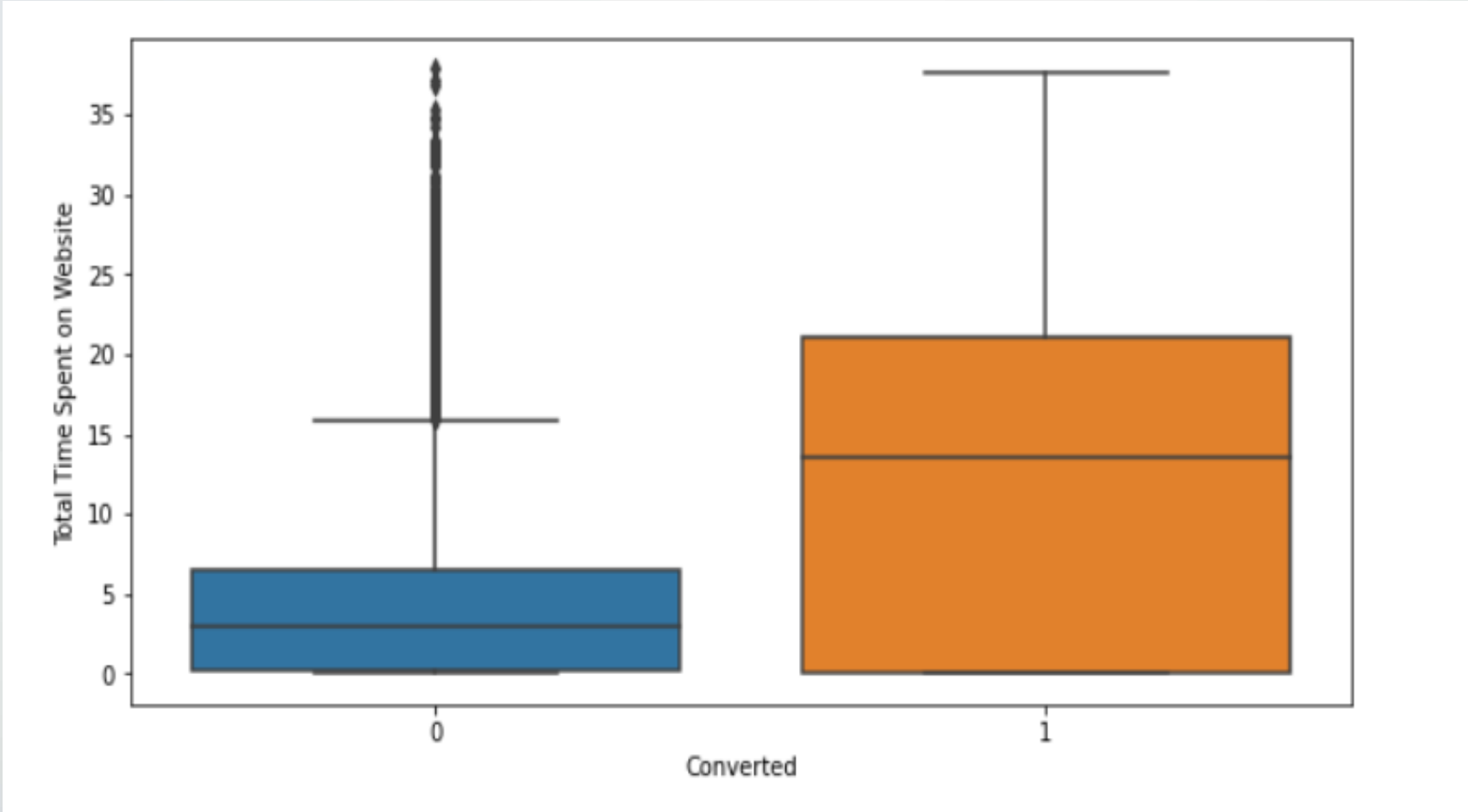
X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have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

STEPS FOLLOWED :

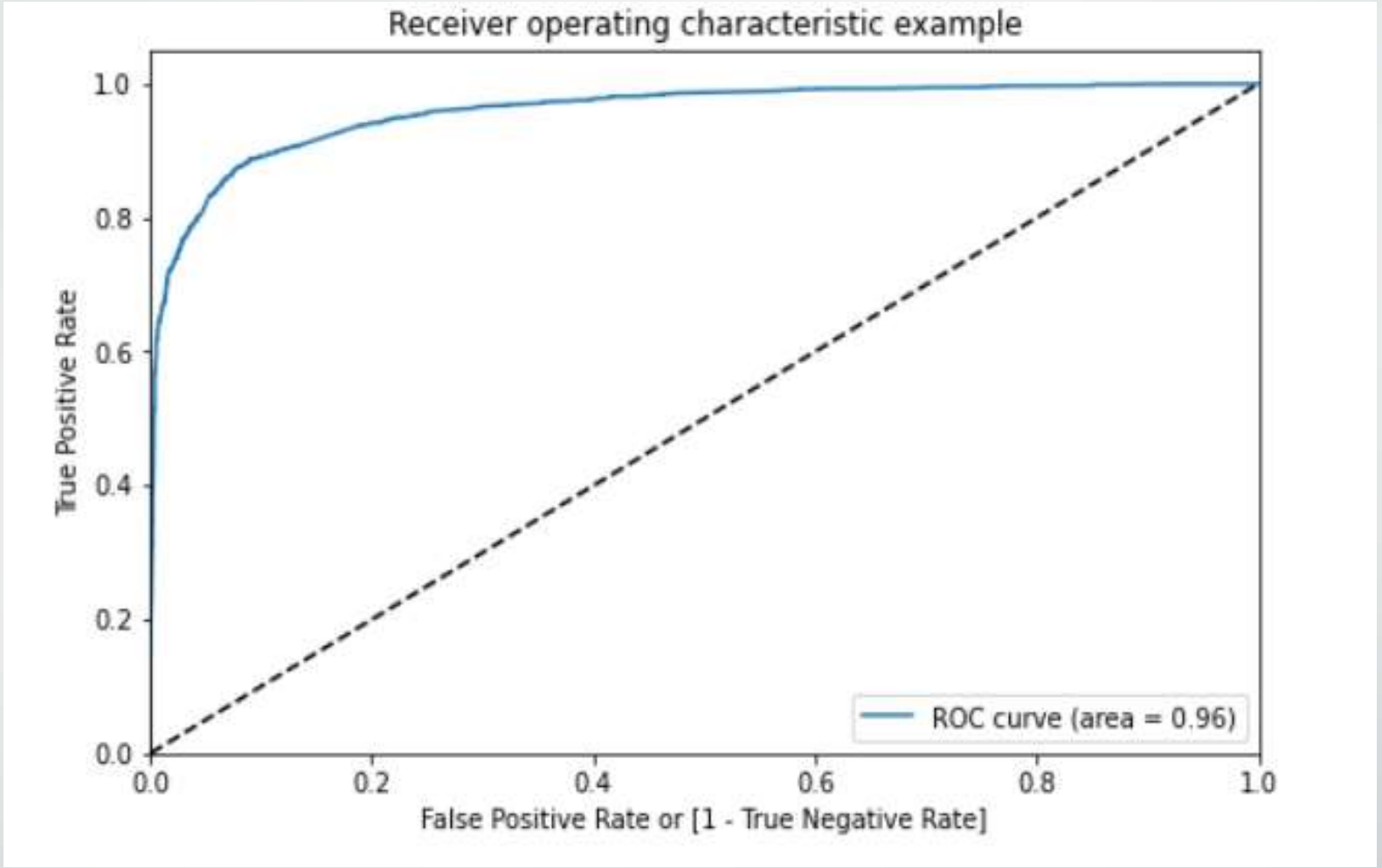
- **Understand the domain/variables.**
- **Import/load the data.**
- **Clean and Prepare the data.**
- **Exploratory Data Analysis.**
- **Create dummy variables.**
- **Splitting the data into Train and Test sets.**
- **Feature Scaling.**
- **Building Logistic Regression Model.**
- **Model Evaluation.**



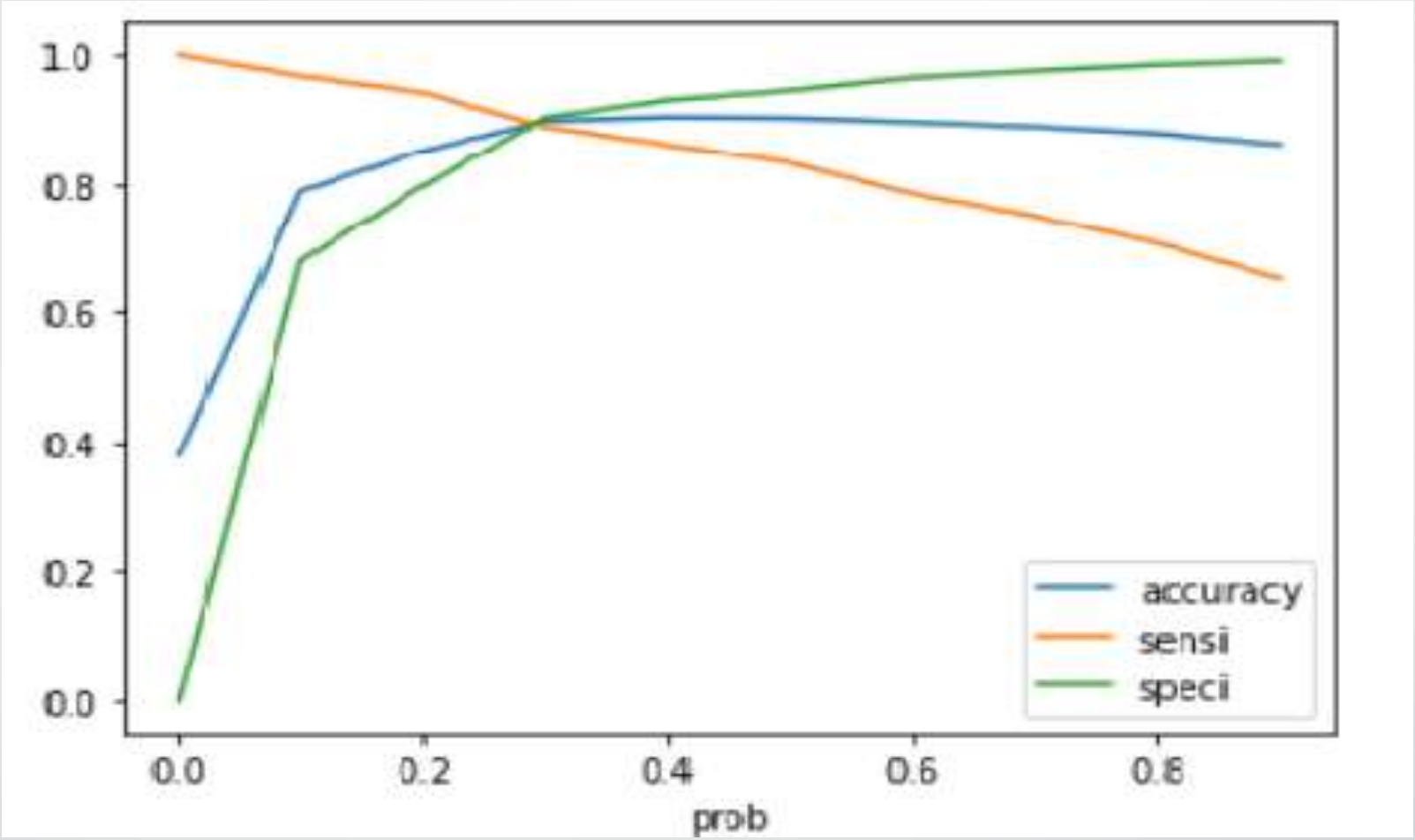
- **Working Professionals going for the course have high chances of joining it.**
- **Unemployed leads are the most in terms of Absolute numbers.**



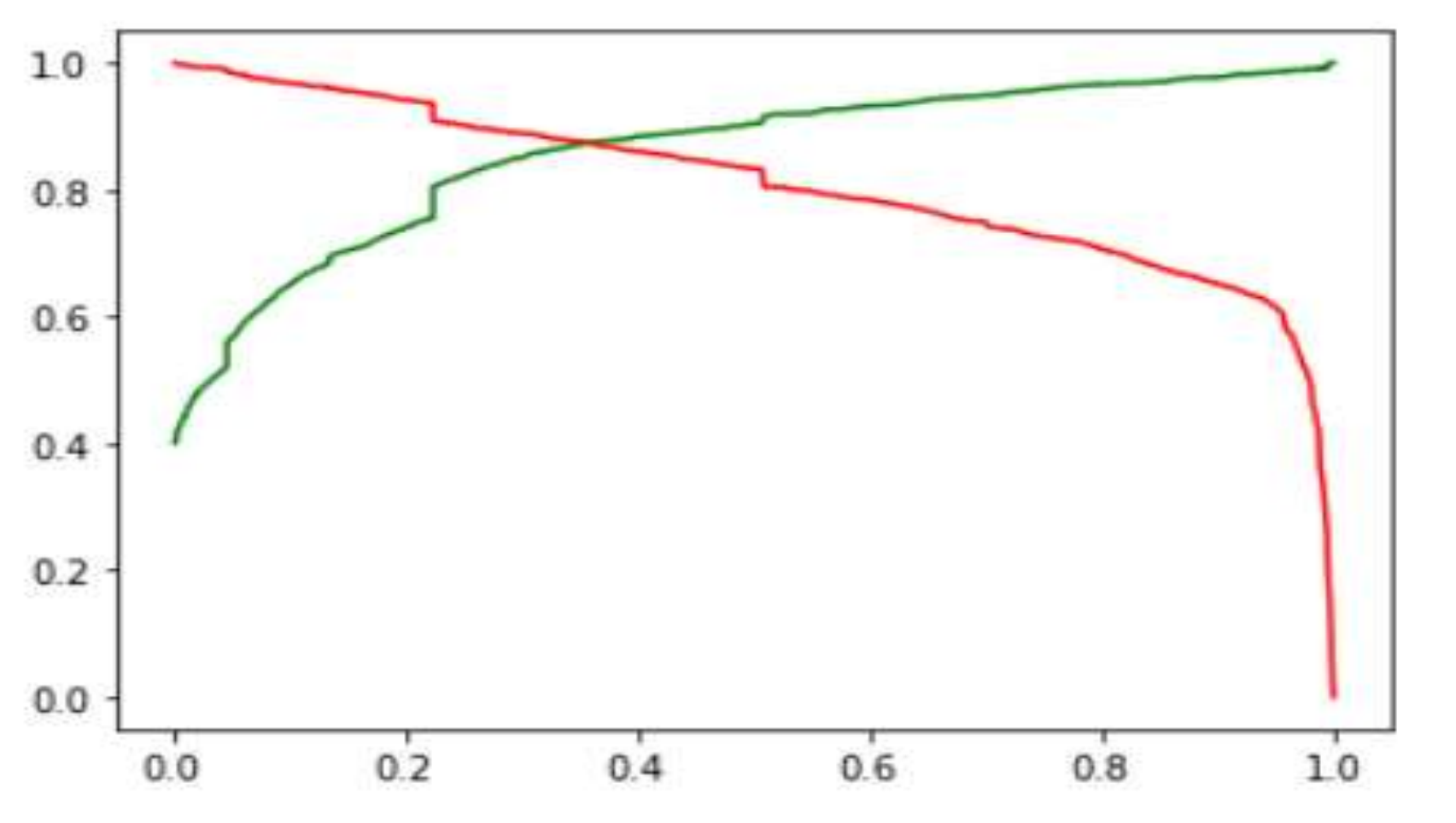
- **leads who are spending time around 10 to 15 minutes on website are mostly conversing**



- **Area under cover is 96% which is indicating a good predictive model as ROC curve is towards TPR side and also having value close to 1.**



- **From the curve above, 0.3 is the optimum point to take it as cutoff probability**



- **Above curve of precision and recall is also give optimal cut off probability of 0.3 .**

IMPORTANT VARIABLES IMPACTING THE CONVERSION RATE

- Total Time Spent on Website
- What is your current occupation Working Professional
- Lead Origin Landing Page Submission
- Lead Origin Lead Add Form
- Last Activity Converted to Lead
- Last Activity Email Bounced
- Last Activity Email Opened
- Last Activity Olark Chat Conversation
- Last Activity SMS Sent
- Lead Source Olark Chat
- Lead Source Welingak Website
- Tags Already a student
- Tags Closed by Horizon
- Tags Interested in other courses
- Tags Will revert after reading the email

CONCLUSION

- Conversion rate are high for reference and welingak website.
- Focused should be more on improving lead conversion of Olark chat, organic search, direct traffic, and google leads as most number of leads are approaching website from these platform
- Accuracy of both train and test data is above 90% which indicates that model is good.
- As our focus was on predicting conversion rate correctly, the Model seems to predict the Conversion Rate very well as sensitivity is 90.38% on train set and 90.01% on test set So we should be able to give the CEO confidence in making good calls based on this model
- Lead Origin Landing Page Submission, Last Activity Converted to Lead, Last Activity Email Bounced, Last Activity Olark Chat Conversation, Tags Already a student, Tags Interested in other courses, Tags Ringing are the variables which are negatively correlated with target variable ‘ Converted ’