Lead Score Case Study

By Balakrishna

Archana &

Ananth

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.

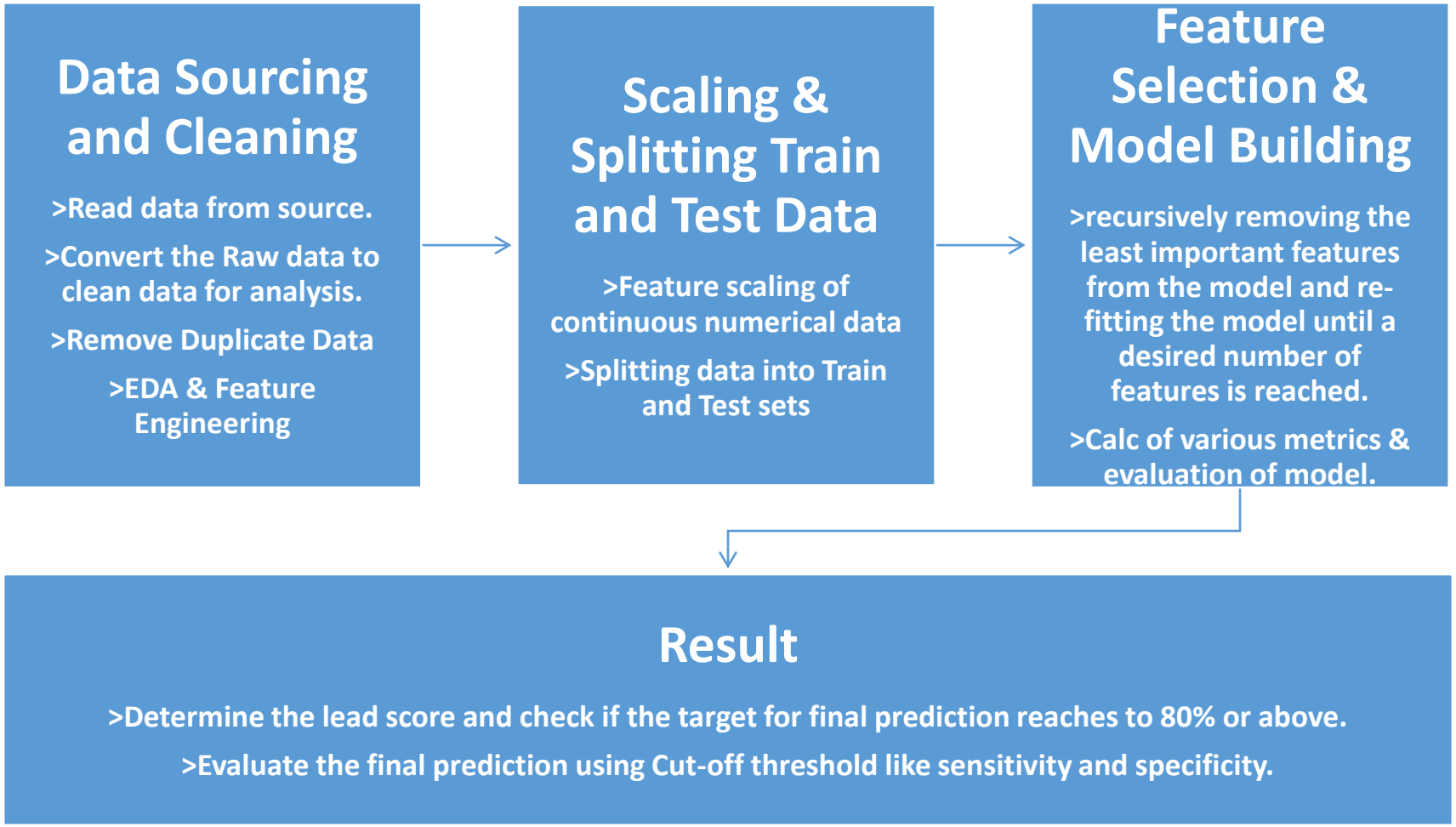
Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

Business Goal:

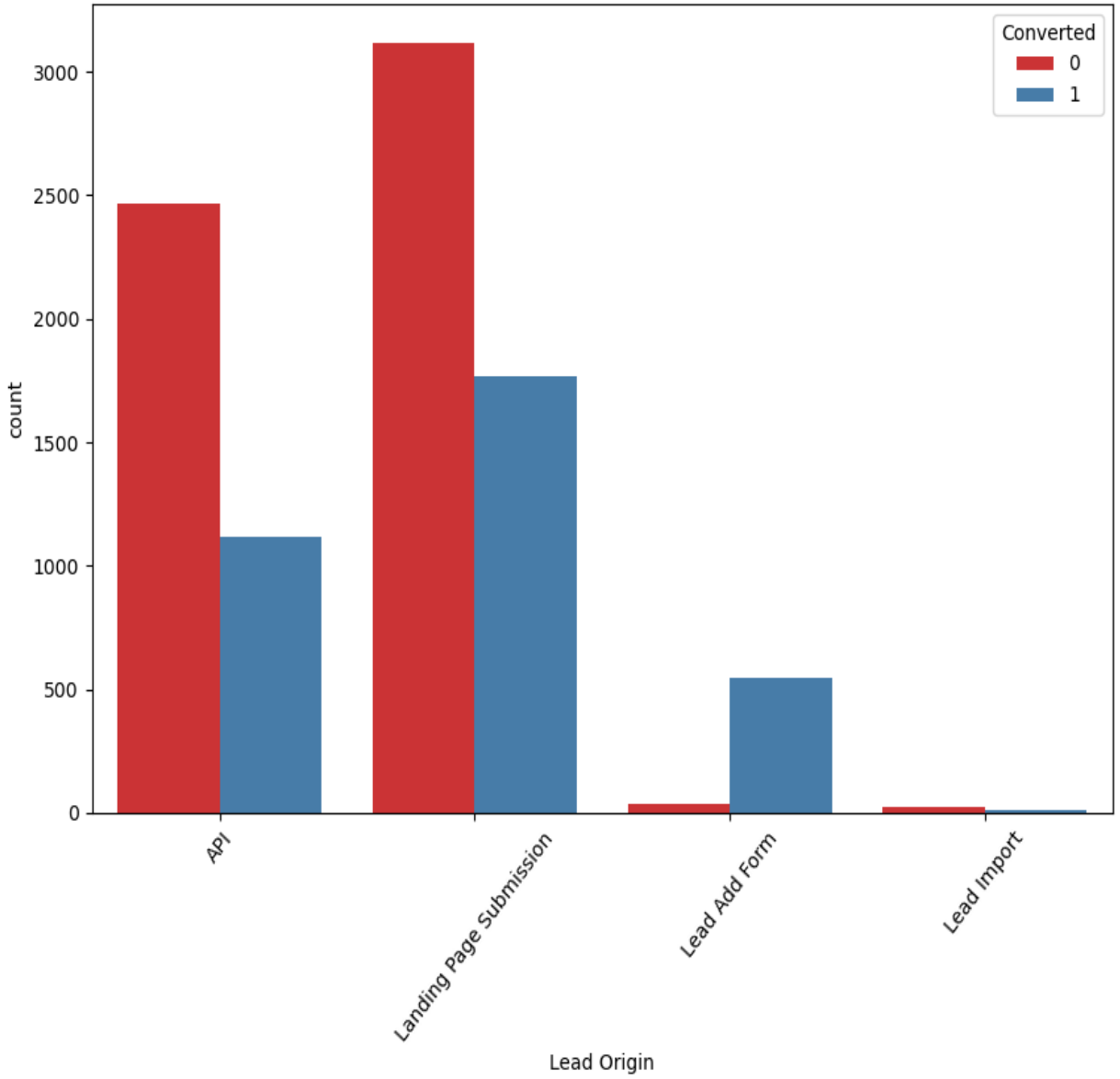
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

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%.

The overall approach of the analysis



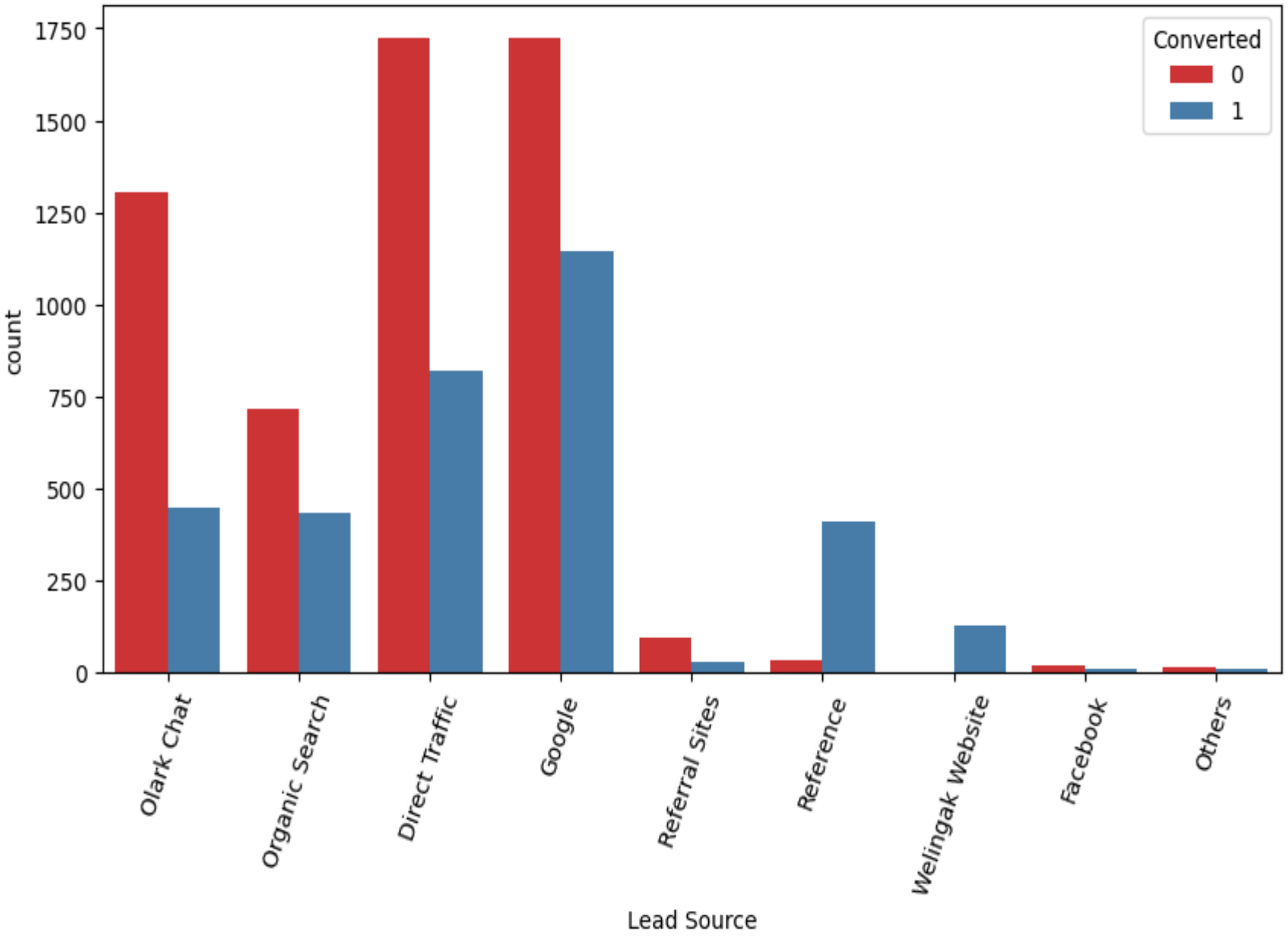
Exploratory Data Analysis



Lead Origin

- API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.

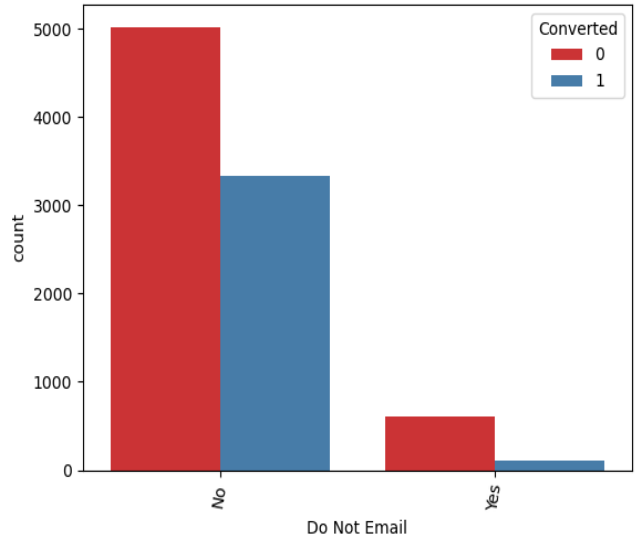
We need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.

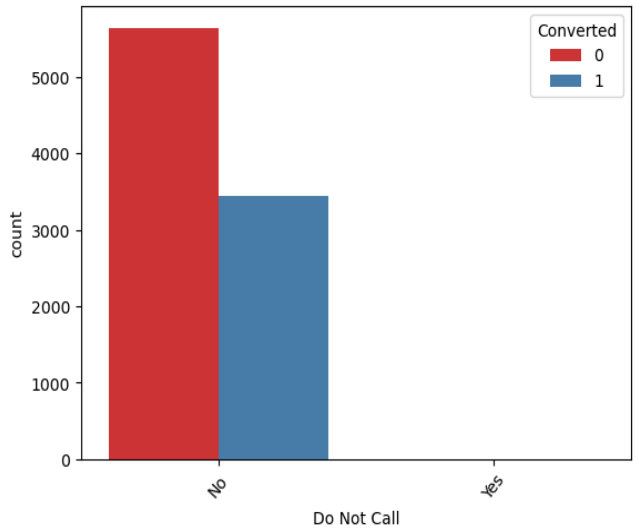


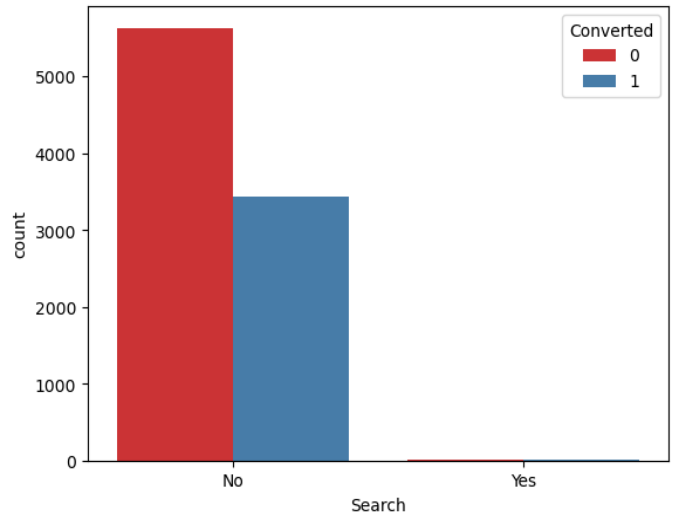
Lead Source:

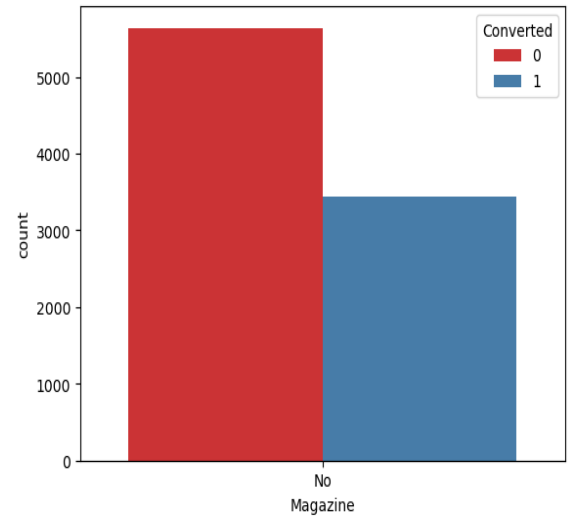
- Google and Direct traffic generates maximum number of leads.
- Conversion Rate of reference leads and leads through welingak website is high.

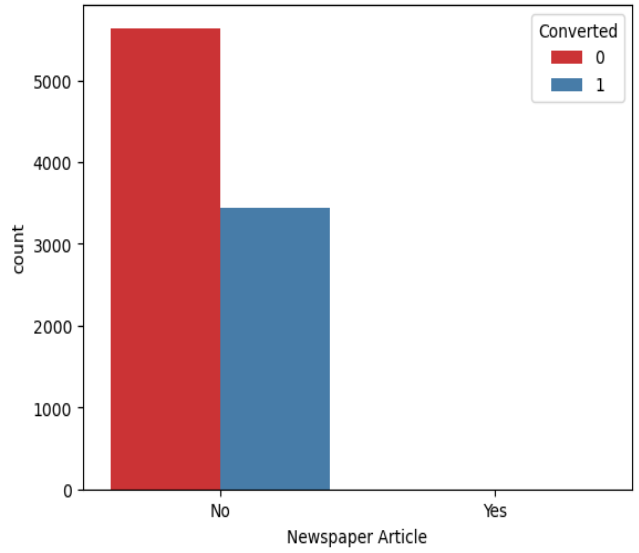
Here, focus should be more on Olark Chat, Organic Search, Google, and lead are very high from Reference and Welingak Website

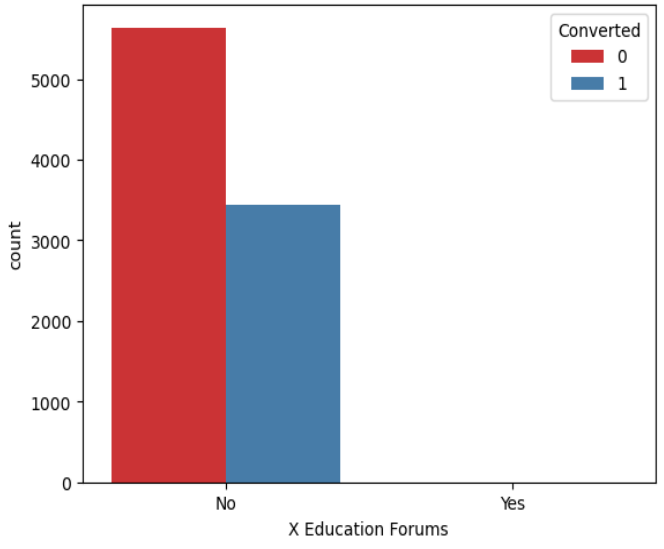


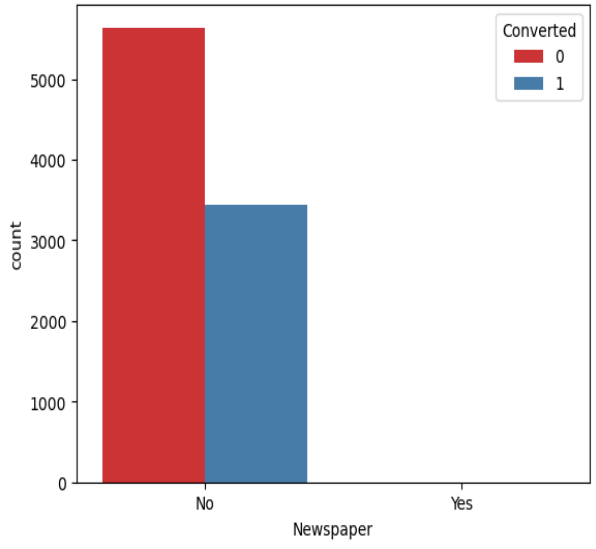


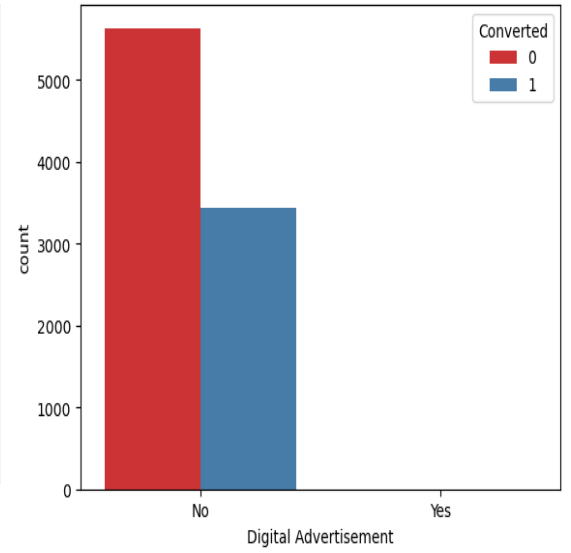




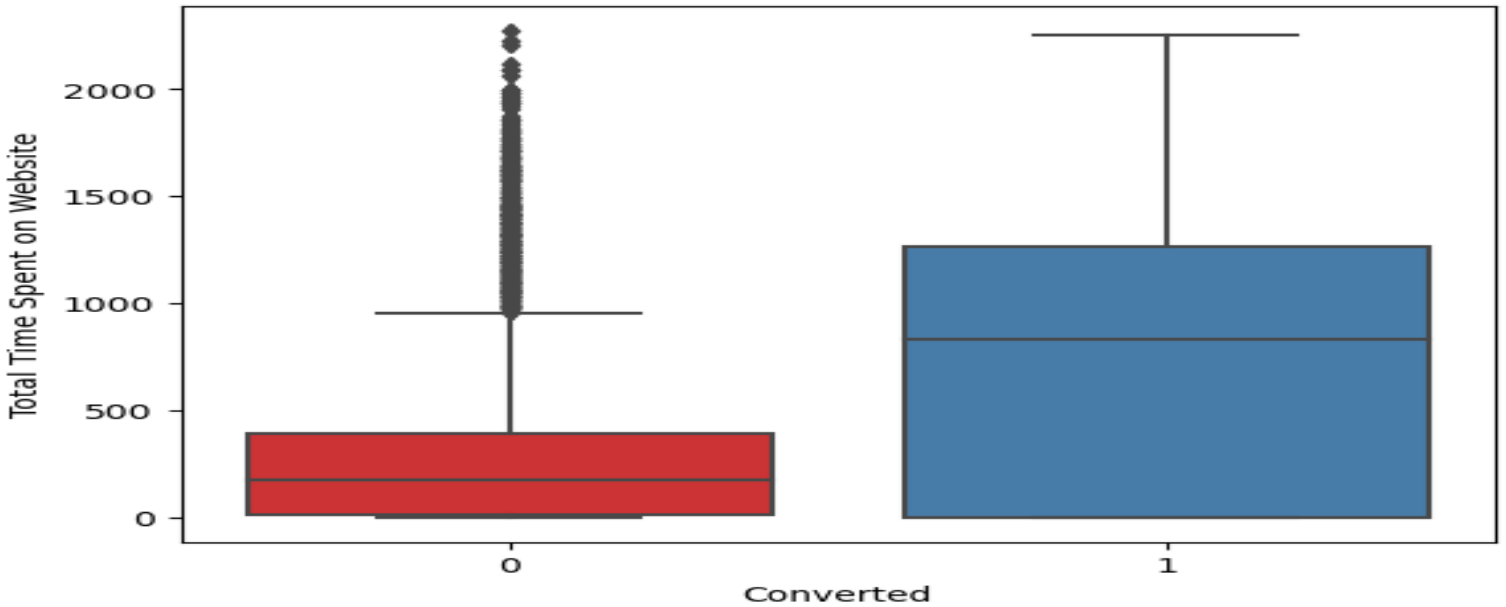




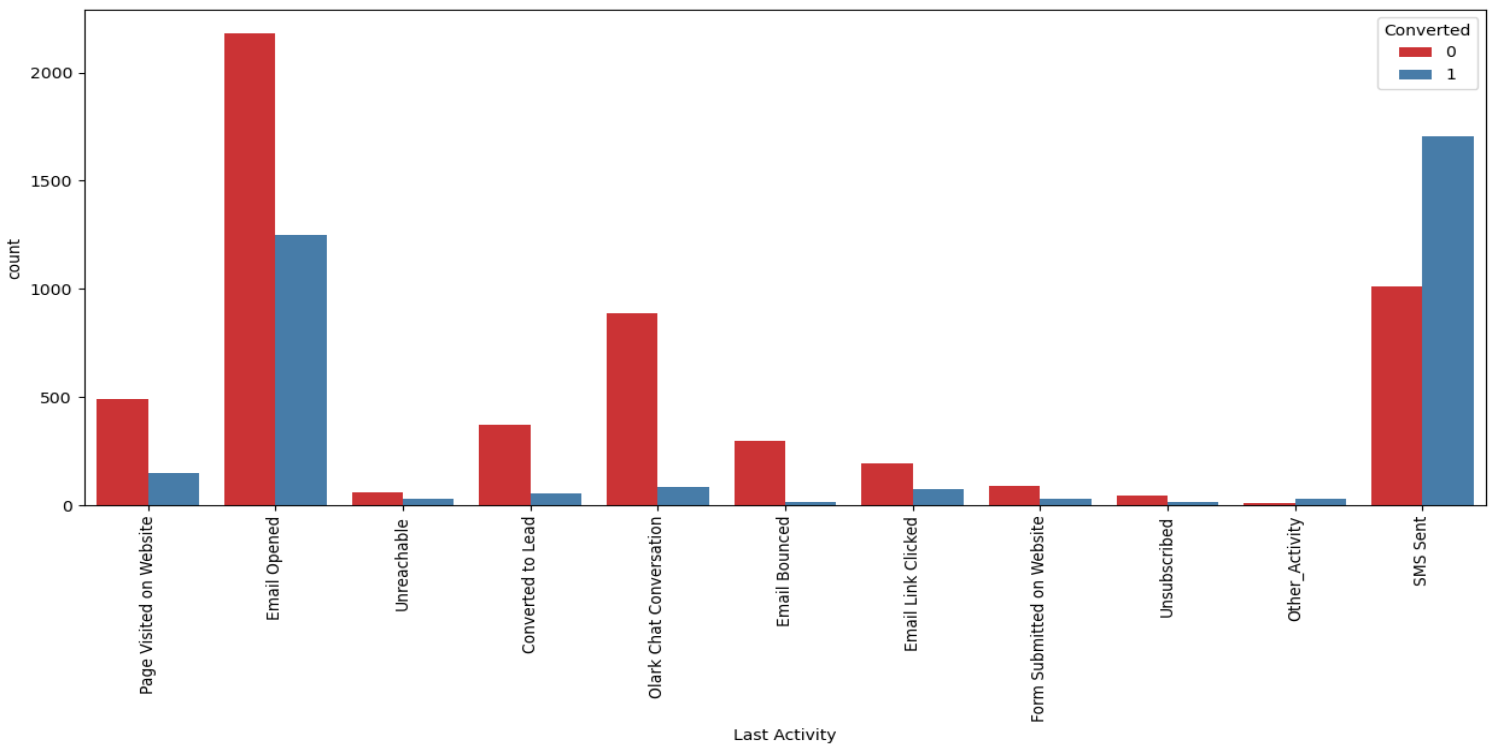




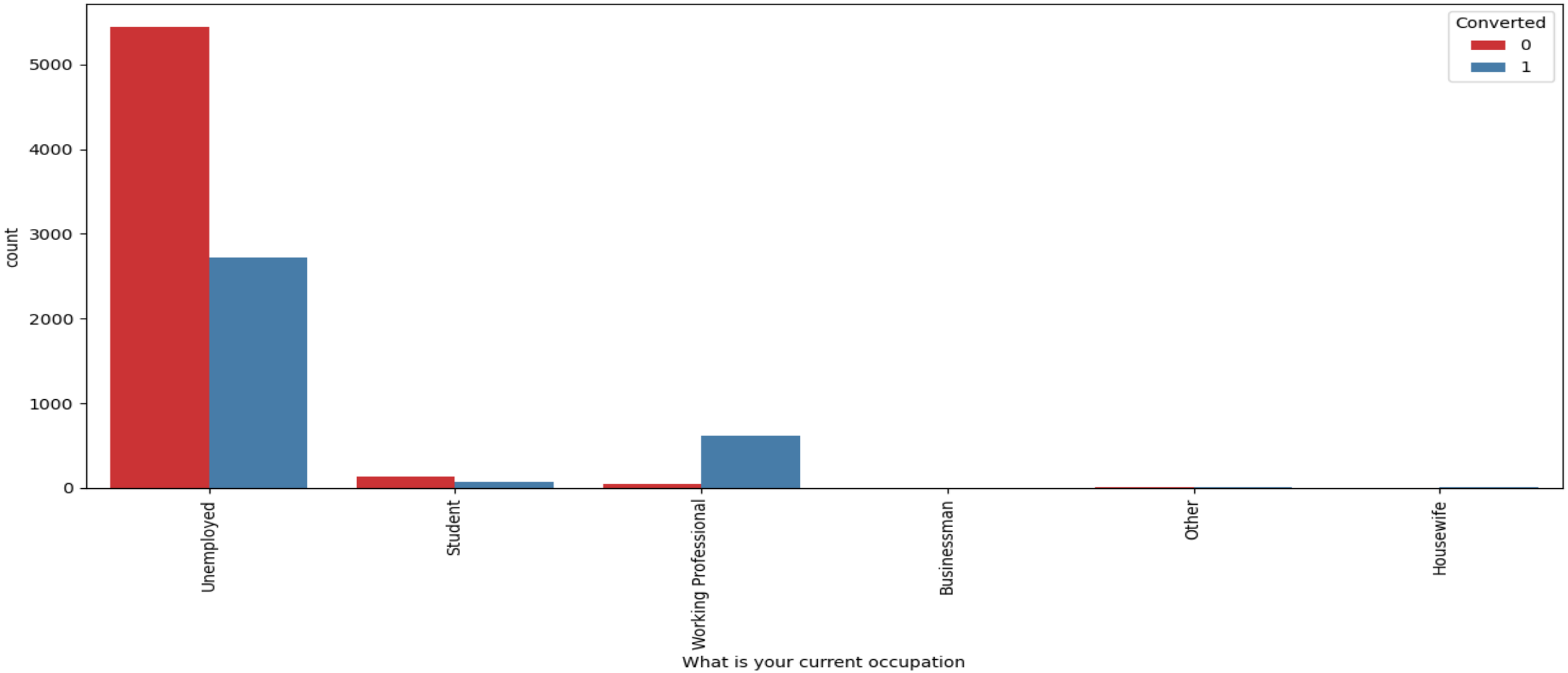
- Inference cannot be drawn from all of the above features including few more features with Through Recommendation, Receive More Updates About Our Courses, Update on my supply chain content, Get Updates on DM Content, I agree to pay the amount through cheque, A free copy of Mastering the Interview,



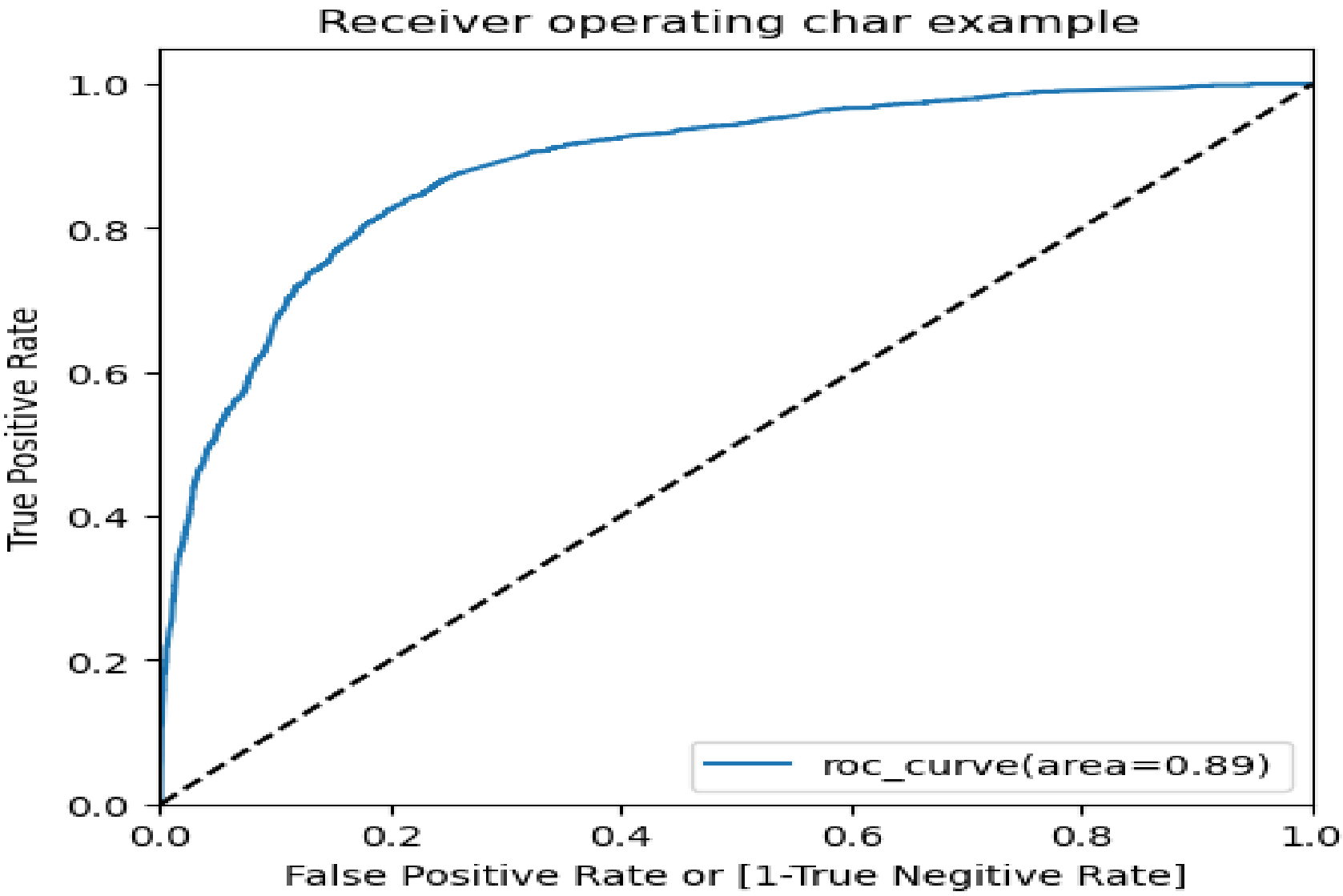
- Those who spent more time on website will more likely to be converted.



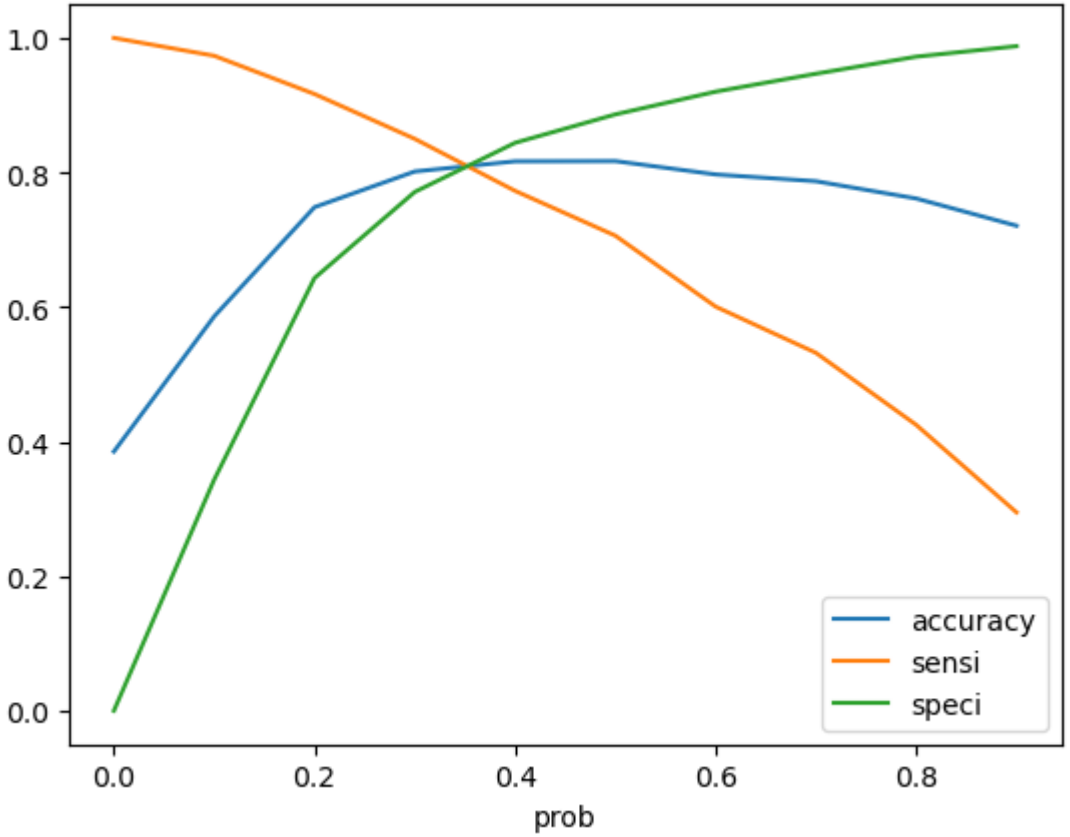
- Most of the lead have their Email opened as their last activity.
- Conversion rate for leads with last activity as SMS Sent is almost 60%.



- Unemployed leads are the most in number, but conversion rate is not so much.
- Working professional has high conversion rate comparing to any other attribute.



- This ROC model suggests it is good one as we have ROC_Curve(0.89).



Train Data:
Accuracy : 81.0 %
Sensitivity : 81.7 %
Specificity : 80.6 %

Test Data:¶
Accuracy : 80.4 %
Sensitivity : 80.4 %
Specificity : 80.5 %

This represents optimal cut-off is 0.37 based on accuracy, Sensitivity, and Specificity.

Conclusion:

- From the both the methods Sensitivity, specificity, and Precision and Recall we have considered the optimal cut-off to calculate the final prediction.
- Accuracy, Sensitivity, and Specificity values of test set are 81%,81% and 80% which are approximately closer to the respective values of trained data.
- Lead score of the final prediction reaches 81% in trained data and 80% in test data.
- Henceforth the model is considered to be good one.