Lead Score Case Study

Logistic Regression Model

PRESENTED BY:

Amit Kumar Chaudhary
Rishabh Kumar Sharma
Lipika Kandari

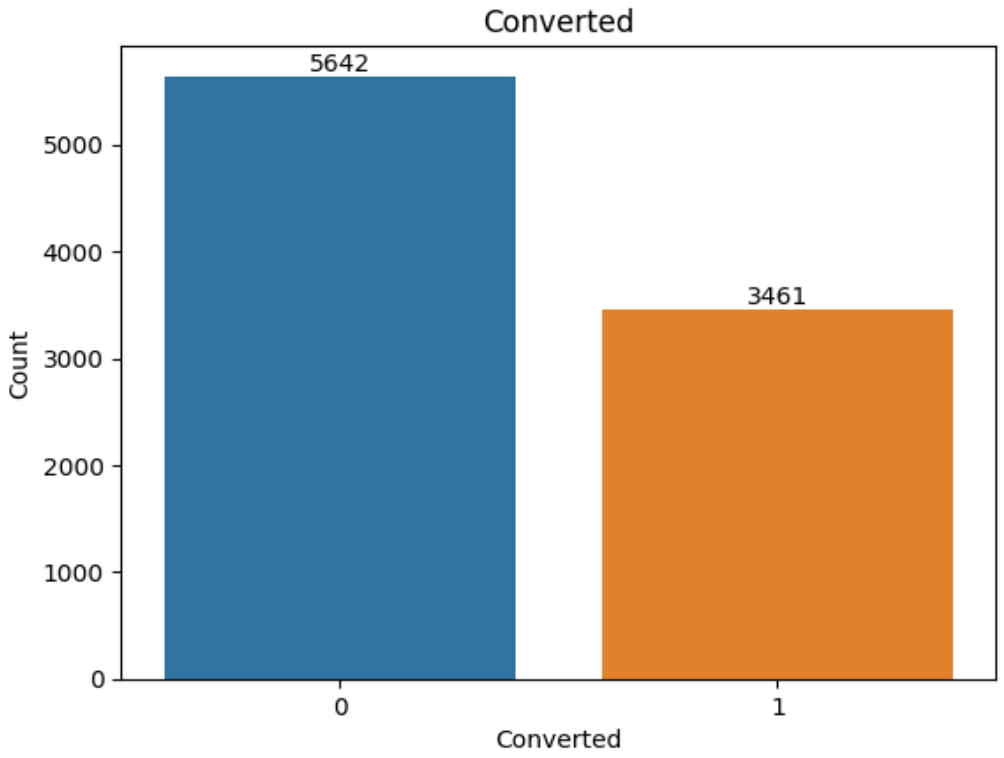
PROBLEM STATEMENT

Industry professionals can purchase online courses from X Education, a company that provides education. On numerous websites and search engines like Google, the firm advertises its courses. Upon arriving at the website, these visitors may browse the courses, submit a form for the course, or watch some videos. These folks are categorised as leads when they fill out a form with their phone number or email address. Additionally, the business receives leads from earlier recommendations. Once these leads are obtained, sales team members begin calling, sending emails, etc. Some leads are converted during this procedure, but most are not. At X Education, the normal lead conversion rate is roughly 30%.

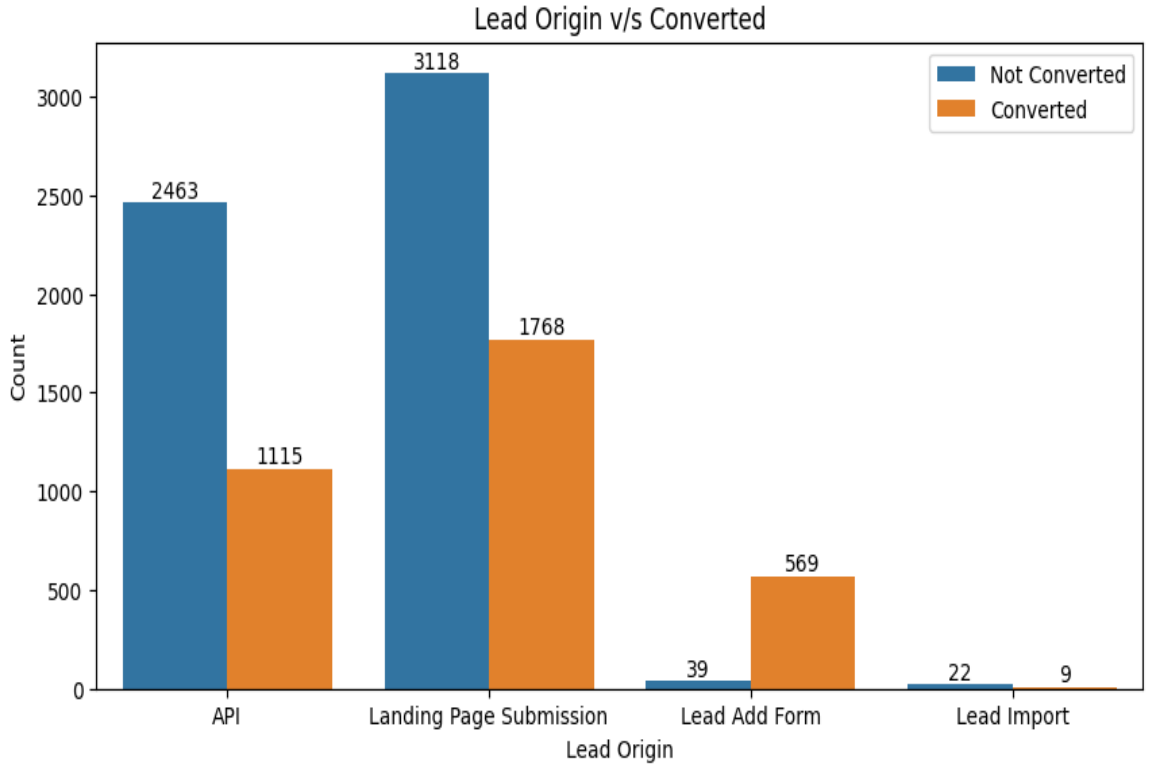
BUSINESS GOAL

You have been asked by X Education to assist them in determining which leads have the best chance of turning into paying clients. The business wants you to create a model where you give each lead a lead score so that leads with higher lead scores have a higher chance of converting than leads with lower lead scores. The CEO has calculated that a lead conversion rate of around 80% would be optimum.

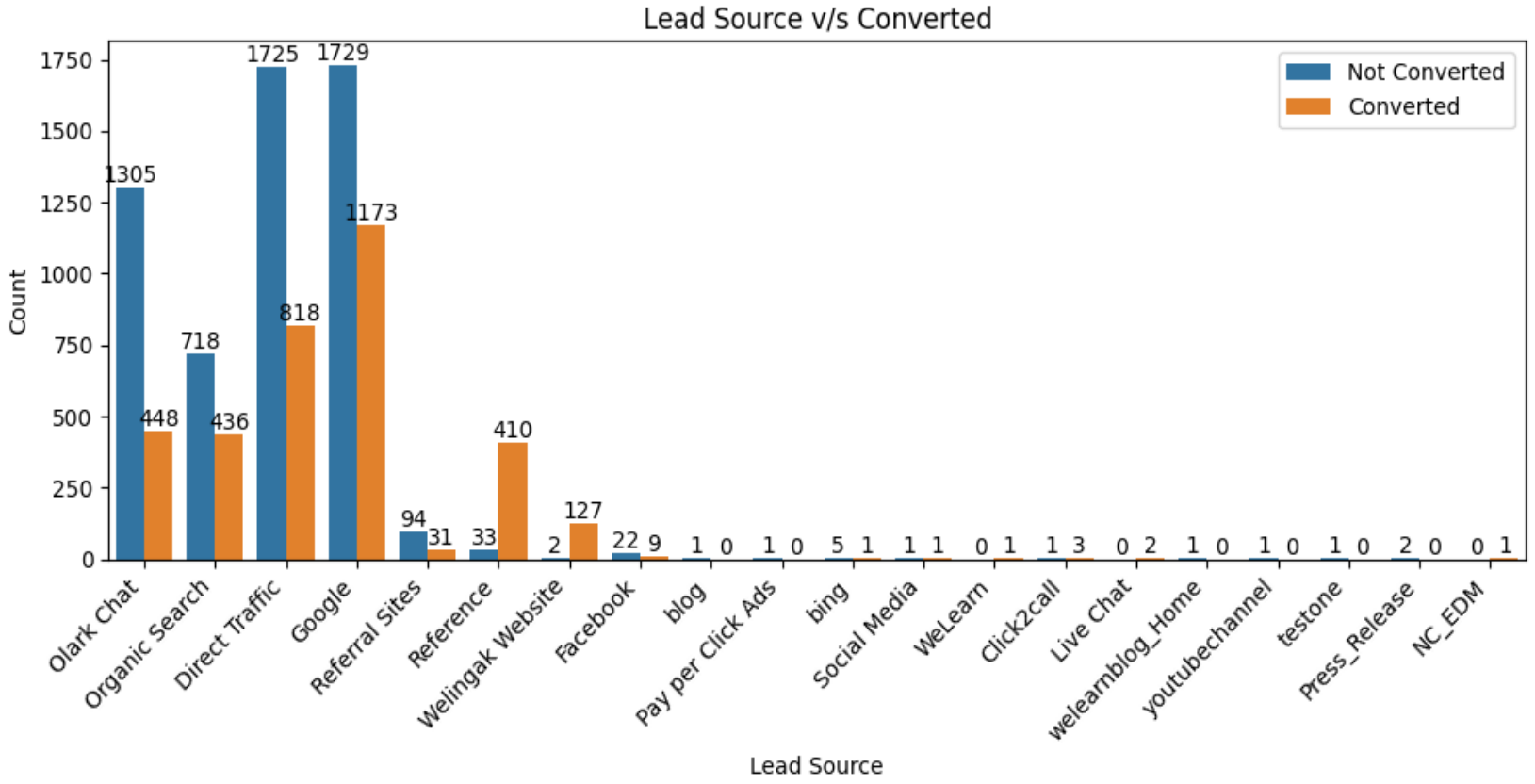
EXPLORATORY DATA ANALYSIS



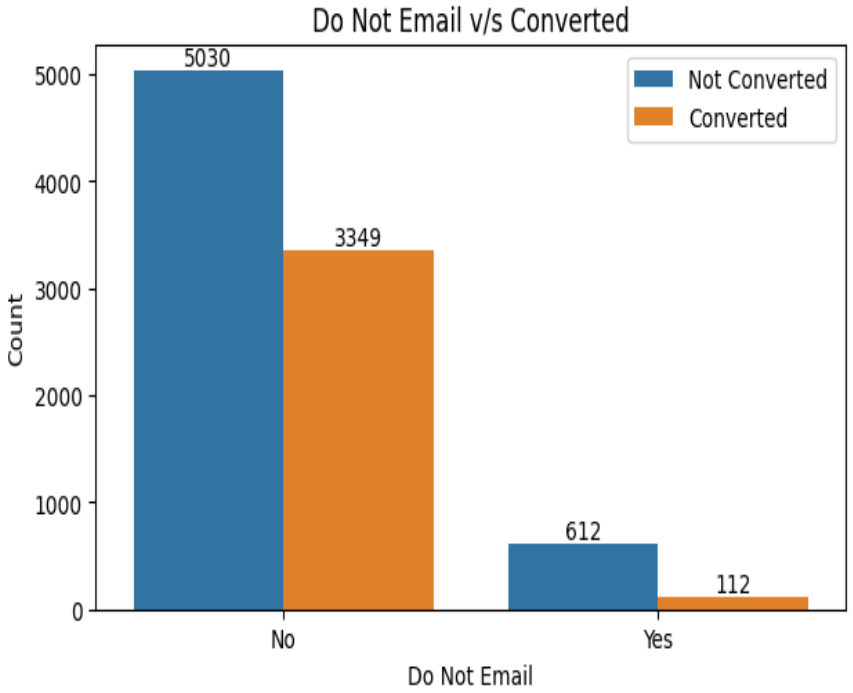
The overall conversion rate for the dataset is 38%.



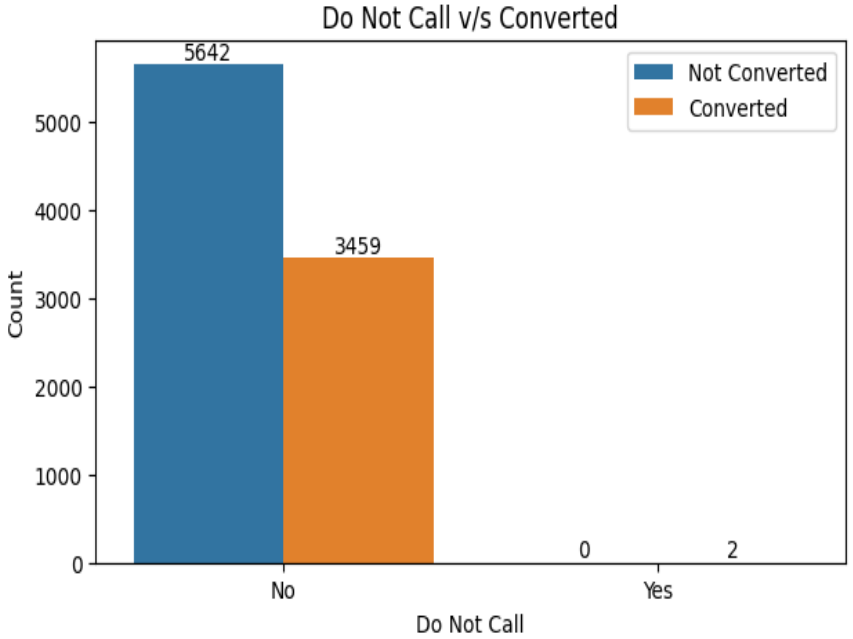
- Maximum leads are originated from Landing Page Submission.
- Maximum conversion also happened from Landing Page Submission with conversion rate around 36.20%.
- Lead Add Form also originates total 608 with the conversion rate around 98.60%.



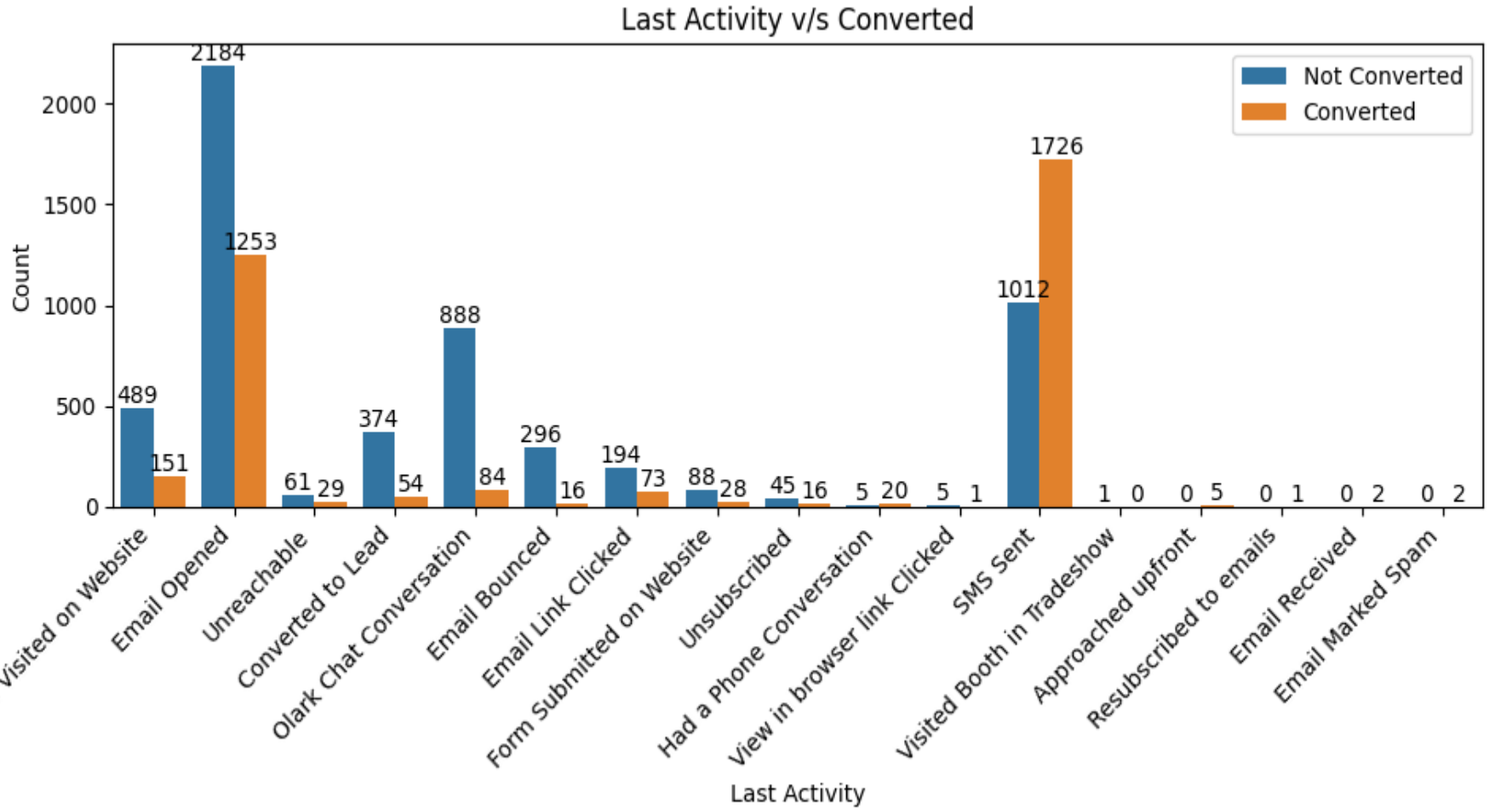
- Google is the source which gives the maximum leads with the conversion rate around 40.42%
- Leads which are coming from the Welingak Website source has the highest rate of conversion around 98.4%
- Leads which are coming from the Refrence Source also has a huge rate of conversion around 92%



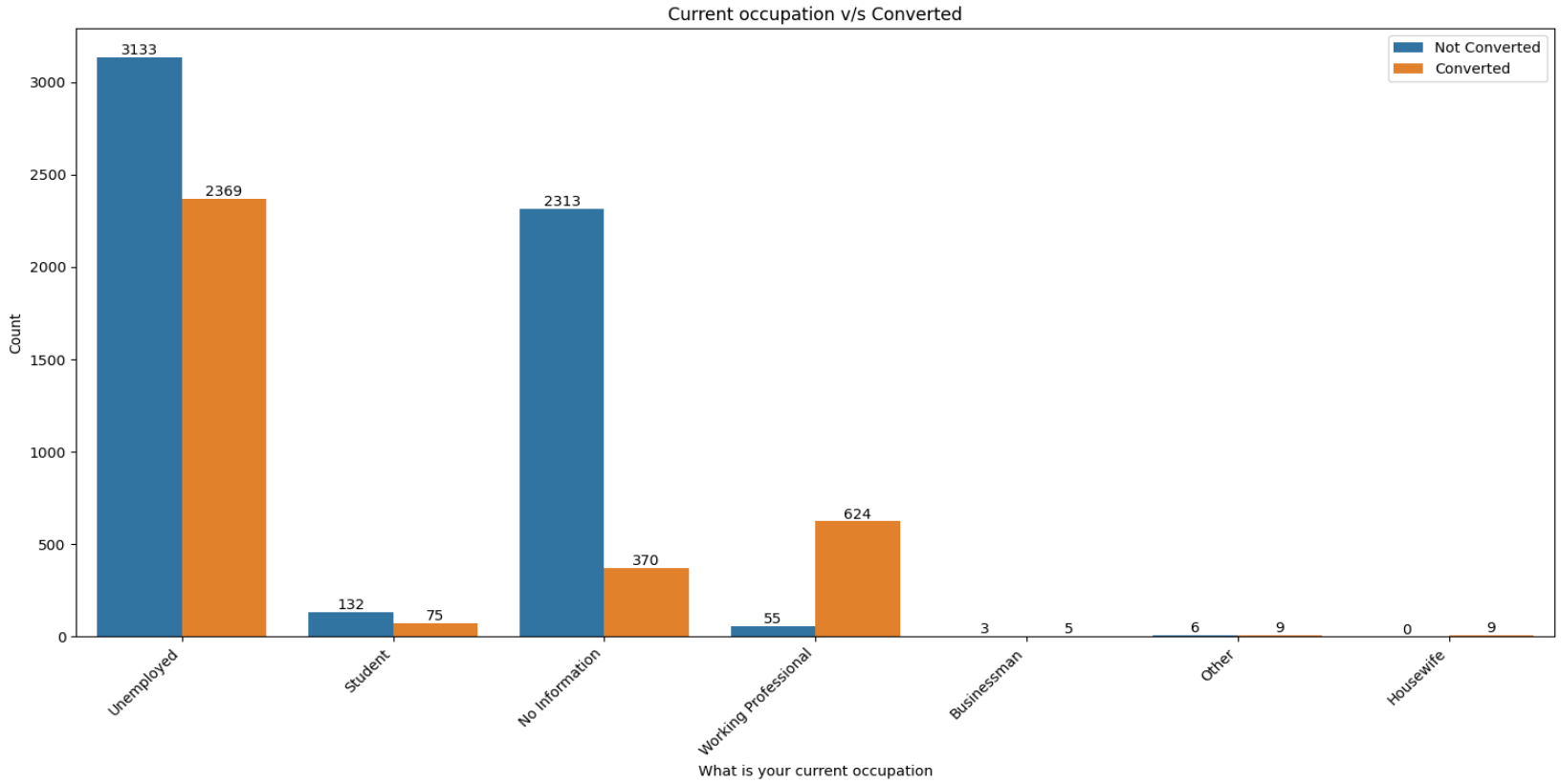
- Most of the conversion has happend from the emails that have been sent



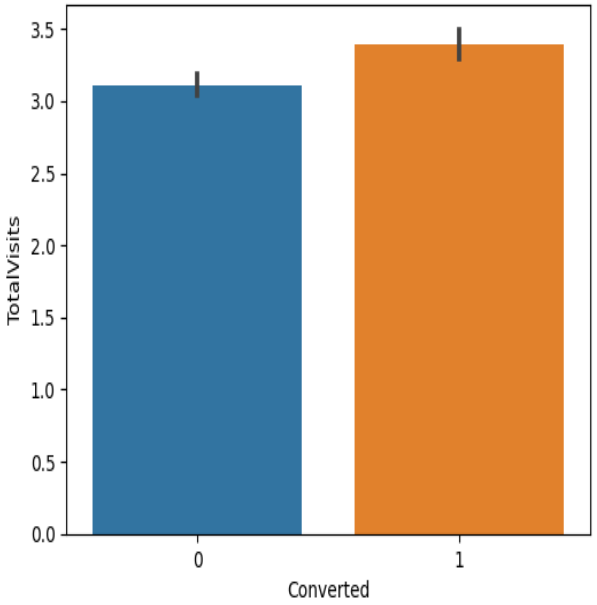
- Mostly the conversions happened when calls were made. However, it can also be seen that 2 leads opted for "Do Not Call", but they still got converted.

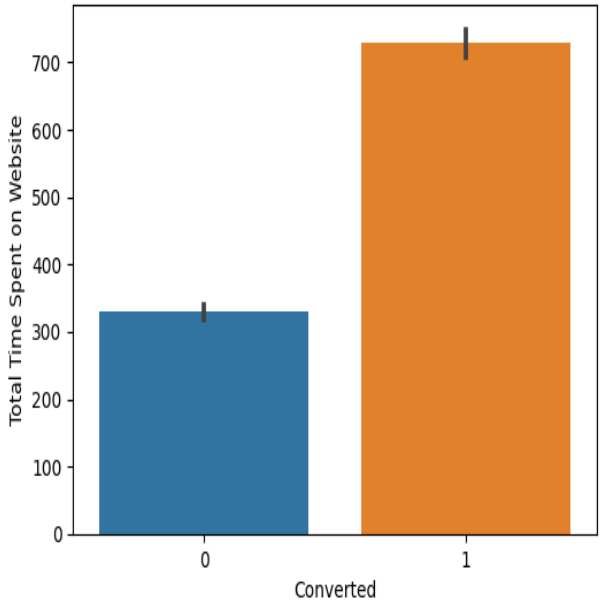


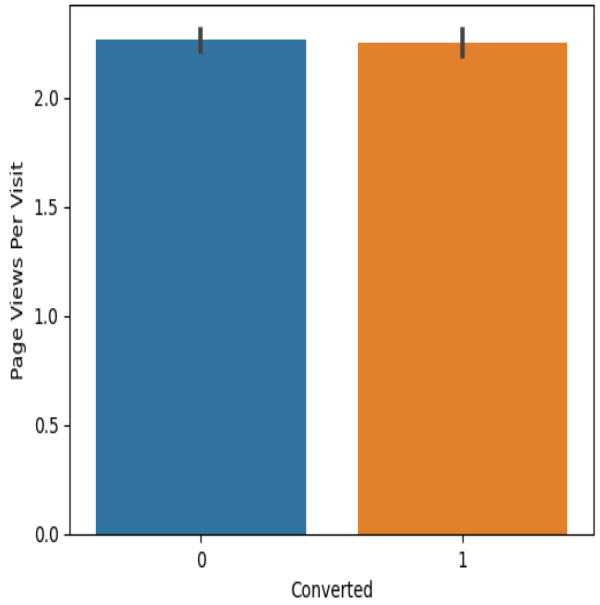
- Last activity value of 'SMS Sent' had maximum conversion with conversion rate around 63%



- Here we can see that maximum coversion takes place from the section of people who are unemployed
- Working Professionals has the huge rate of conversion
- Total 9 out of 9 Housewife got converted
- Out of 8 Businessman 5 got converted and 9 out of 15 got convertd from Other section

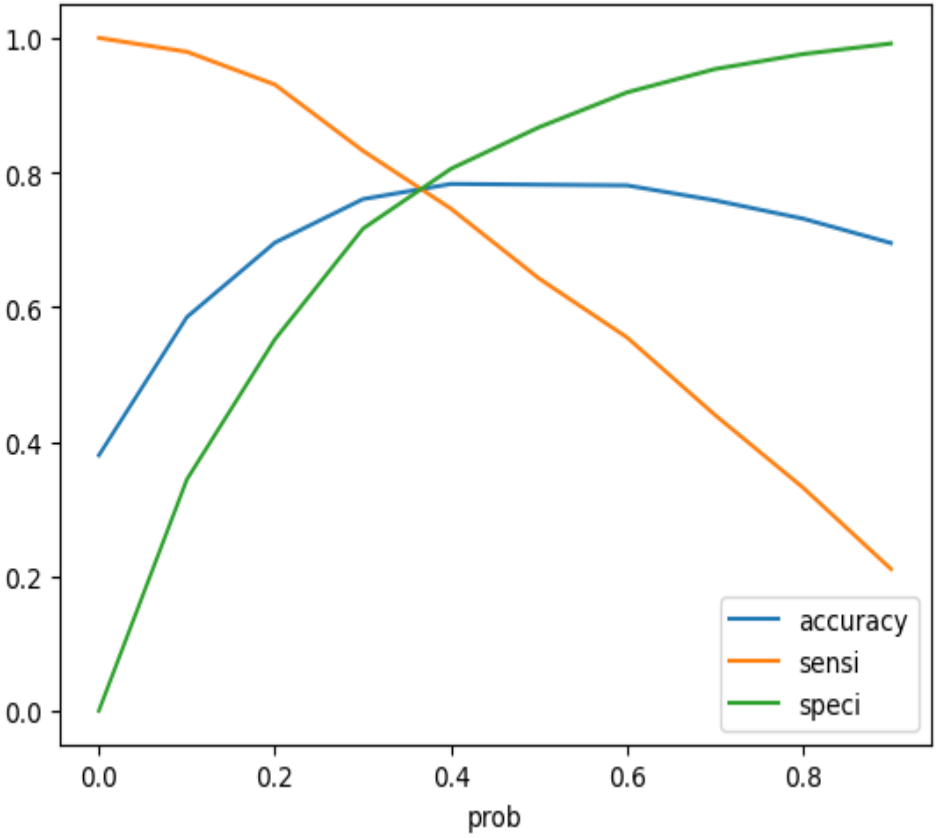




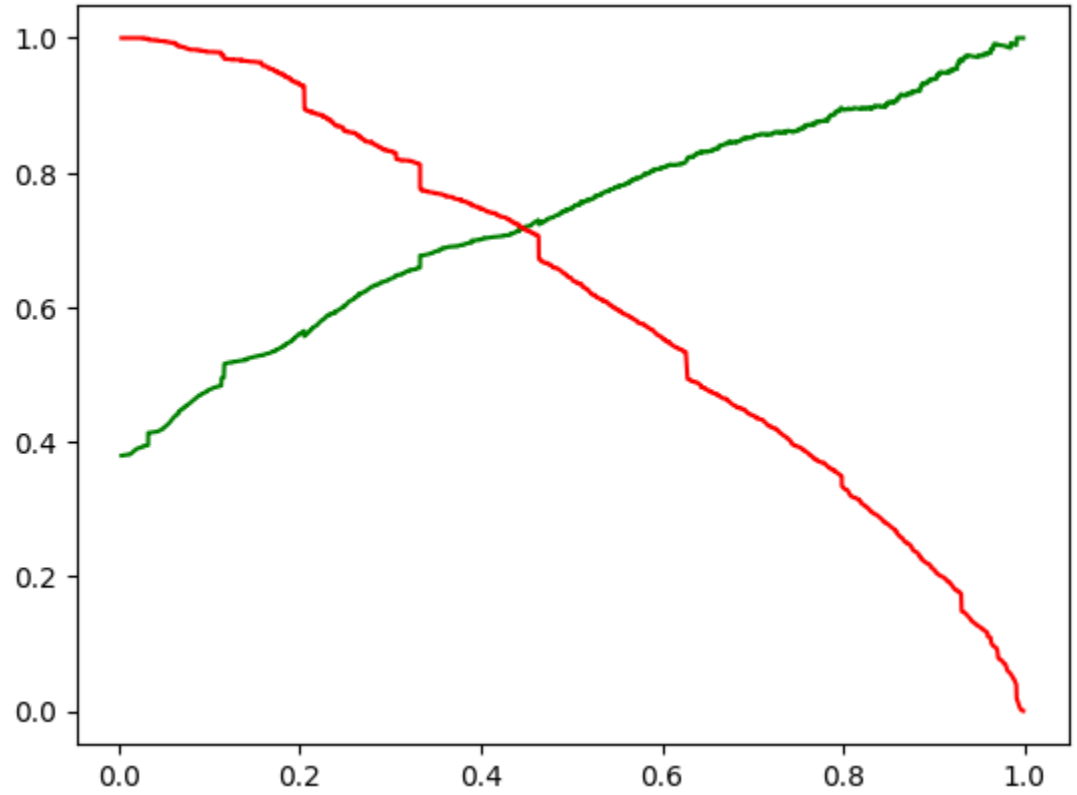


- From graph we can see that, people who spent more time on website has the more chance to get converted.
- The people whose website visit time is more has the more chance to get converted.

MODEL EVALUATION (TEST)



- From the curve above, 0.37 is the optimum point to take it as a cutoff probability.



Based on the precision-recall trade-off curve, the cutoff point seems to be 0.404. We will use this threshold value for test data evaluation.

CONCLUSION

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 79%, 77% and 80% 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 80%
- Hence overall this model seems to be good.

THANK YOU