



Leads Scoring Case Study

By Manish and Ratnesh

Problem Statement

- 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%.

Business Goal

- X Education wants help to know the most promising leads that are most likely to convert into paying customers.
- The company requires help to build a model wherein we 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%.

Solution Methodology

Data Gathering
& Cleaning



EDA
Data
Visualization



Feature Scaling
Model Building



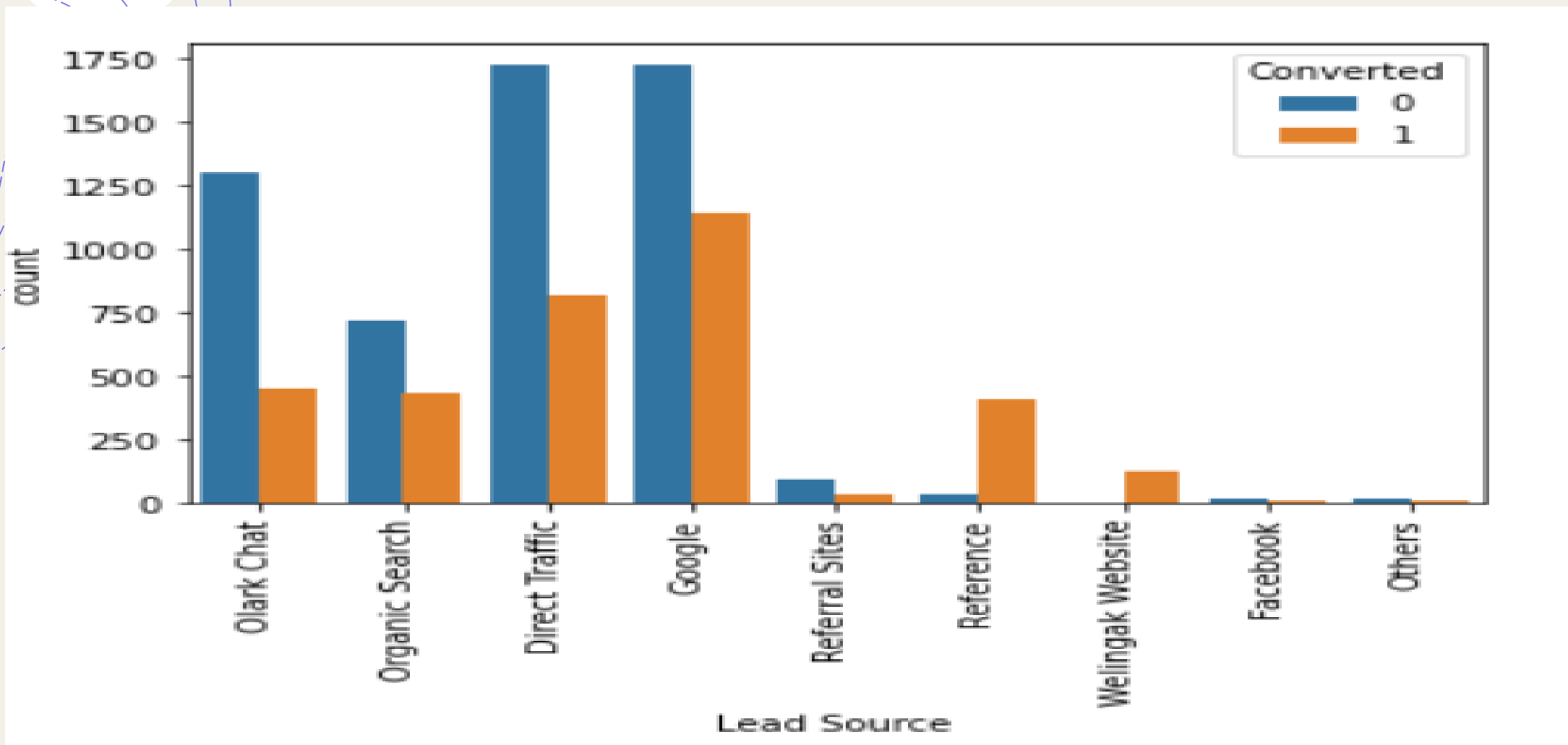
Final Model

- Store and Read source data.
- Remove duplicates, null and redundant columns.
- Perform EDA for numerical and categorical columns
- Perform RFE and Logistic Regression,
- Feature scaling & Dummy variable and encoding of the data
- Model building and tuning using RFE.
- Final Model analysis and test on it.

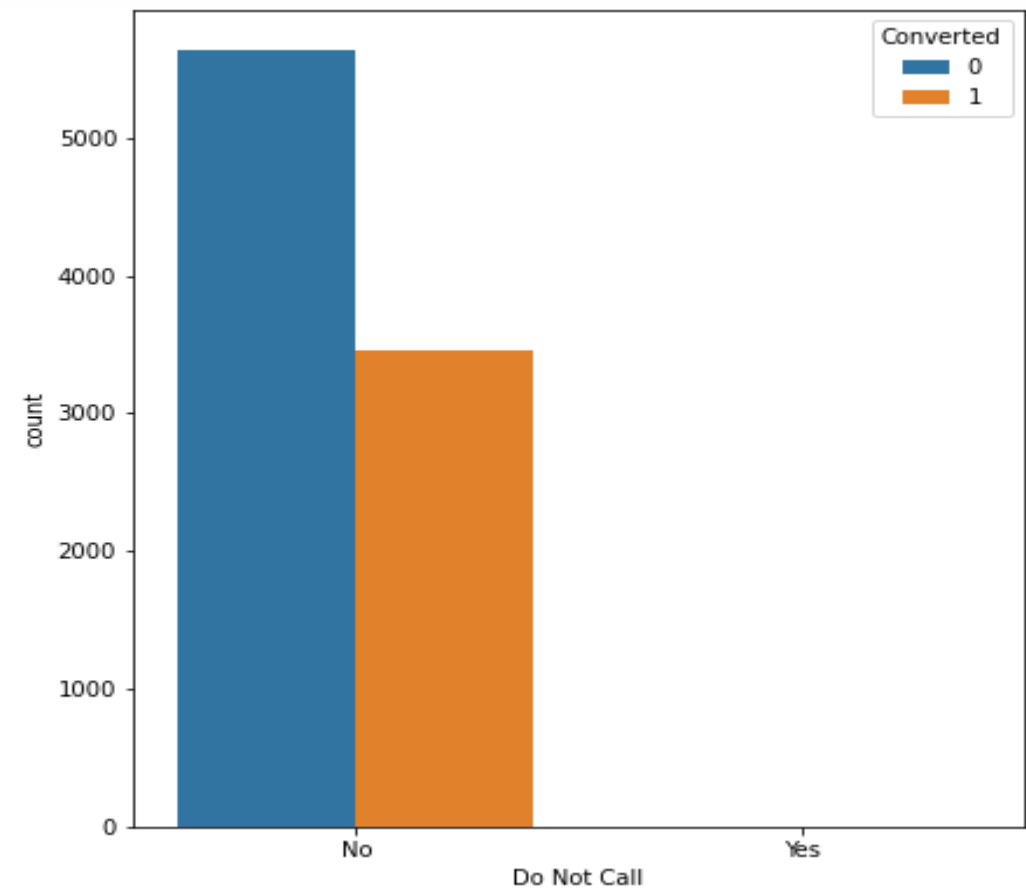
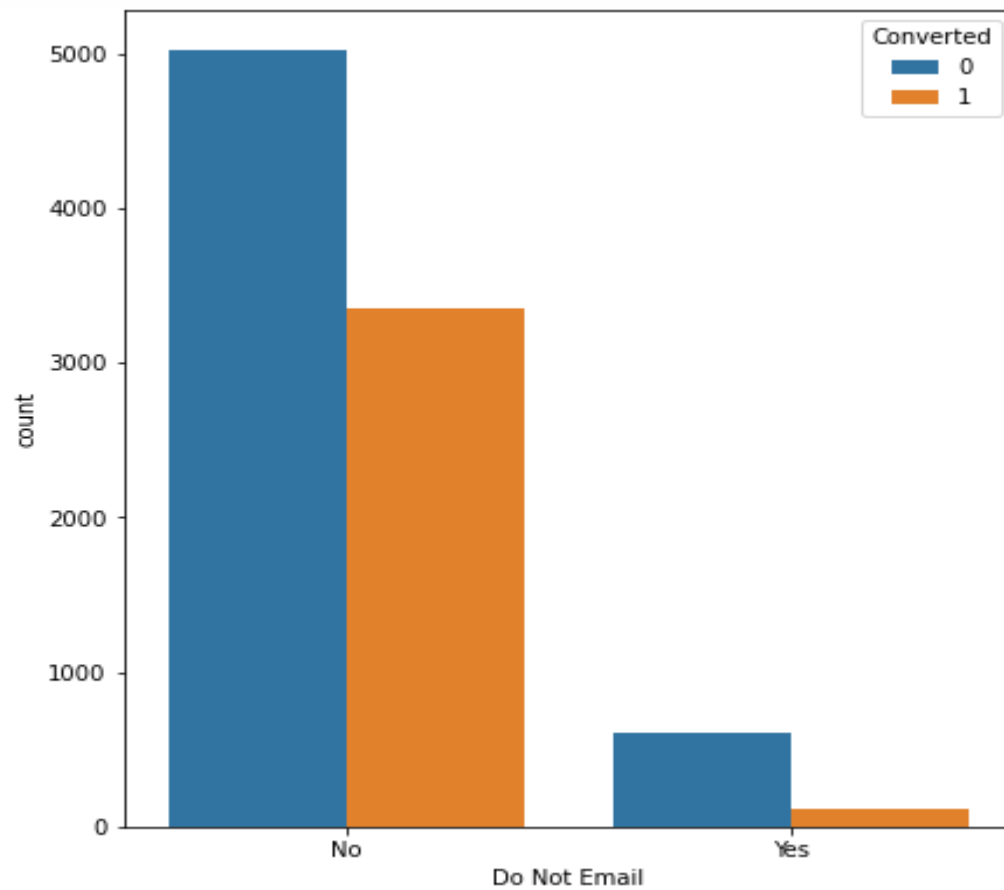


Data Visualization

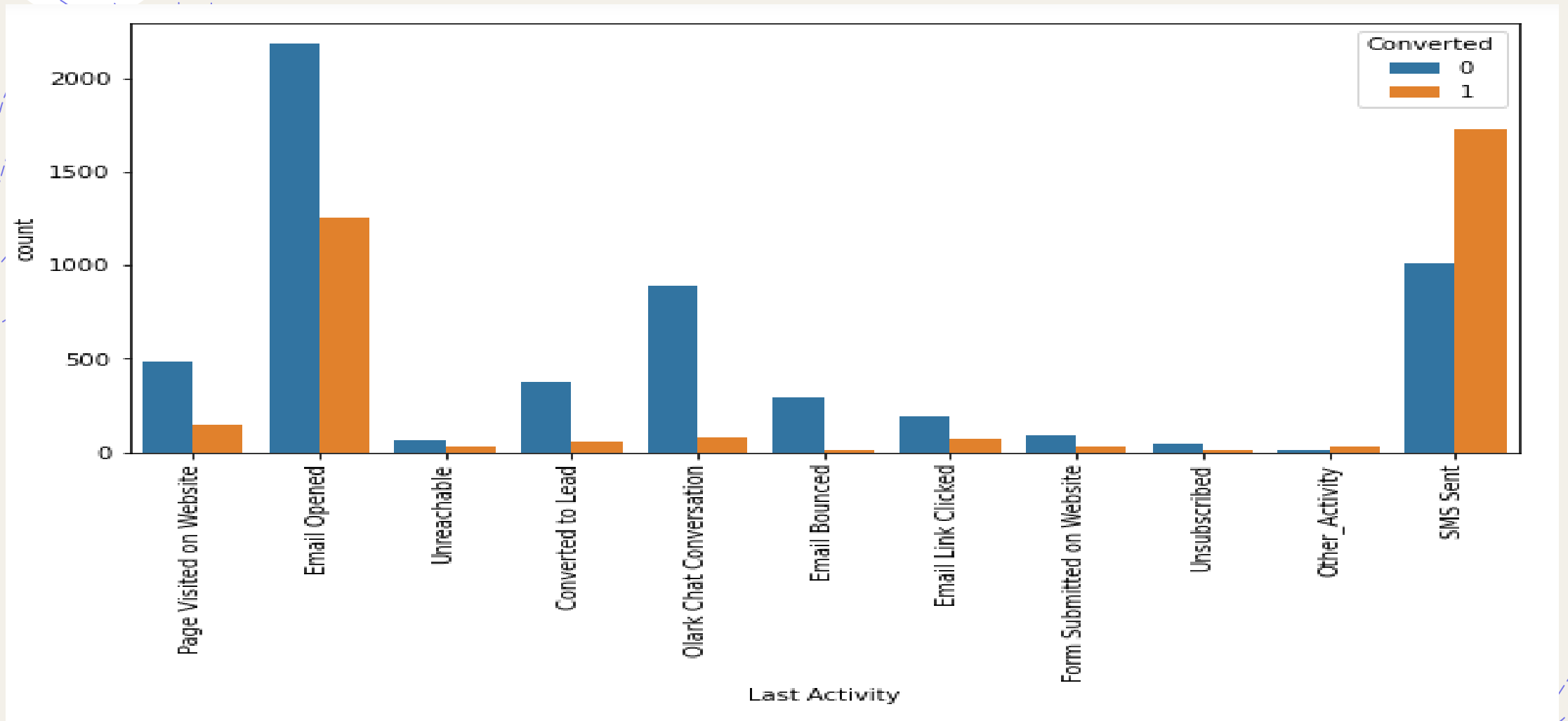
EDA



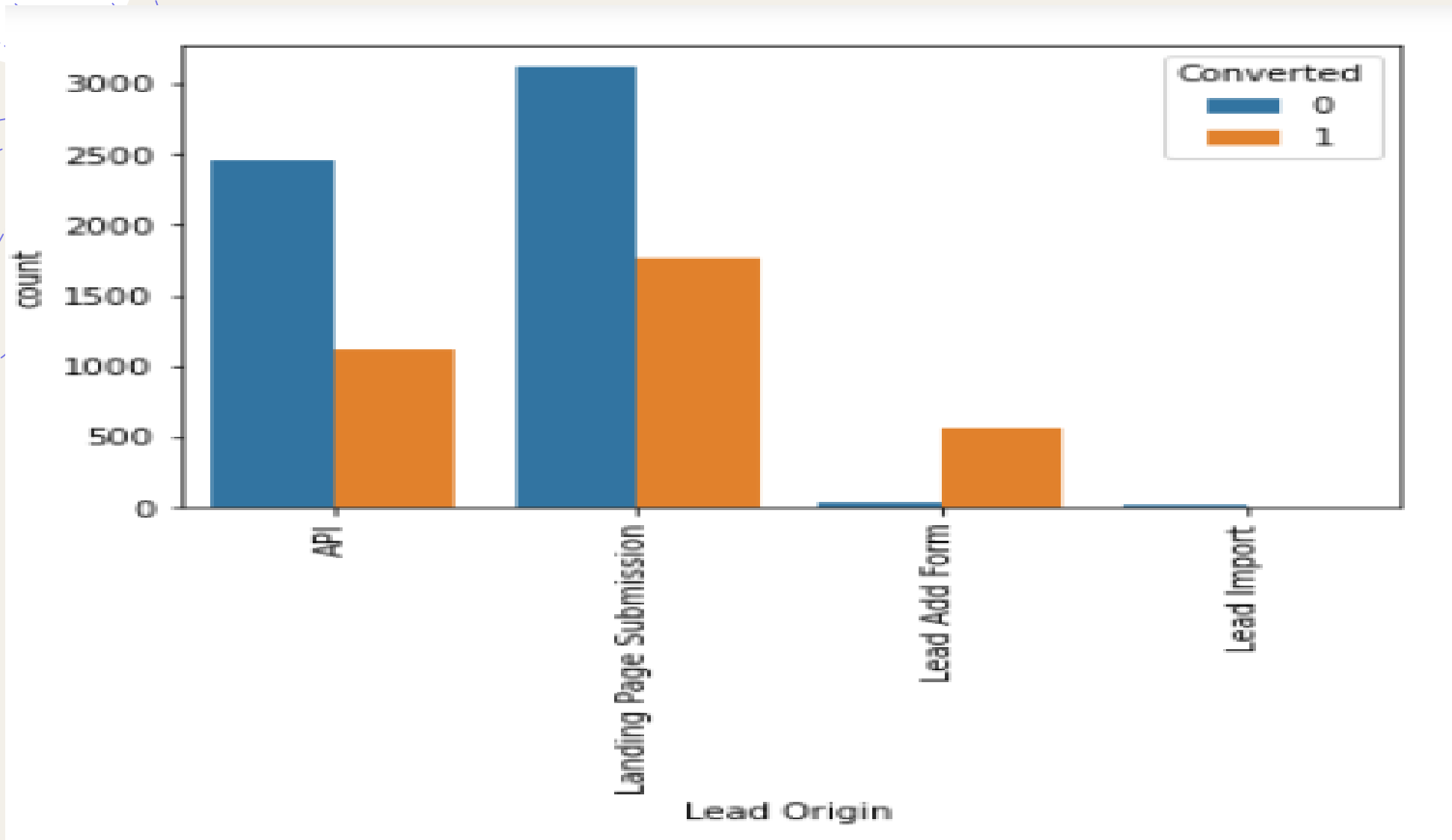
Major conversion in Lead score from **Google**.



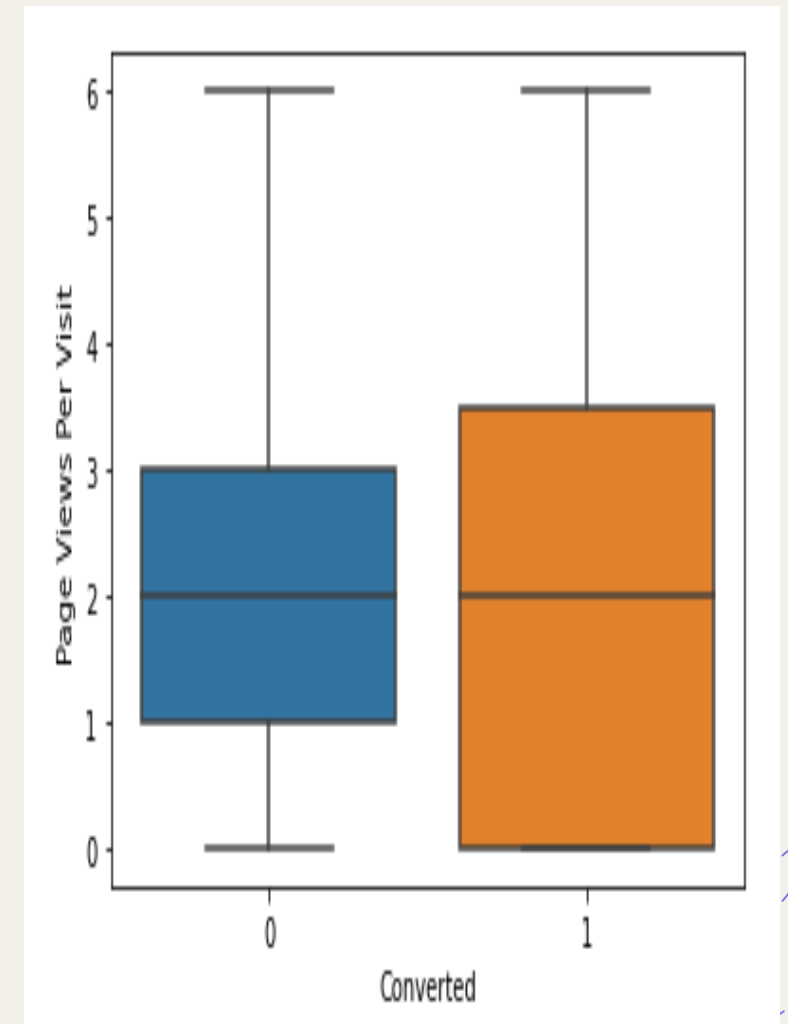
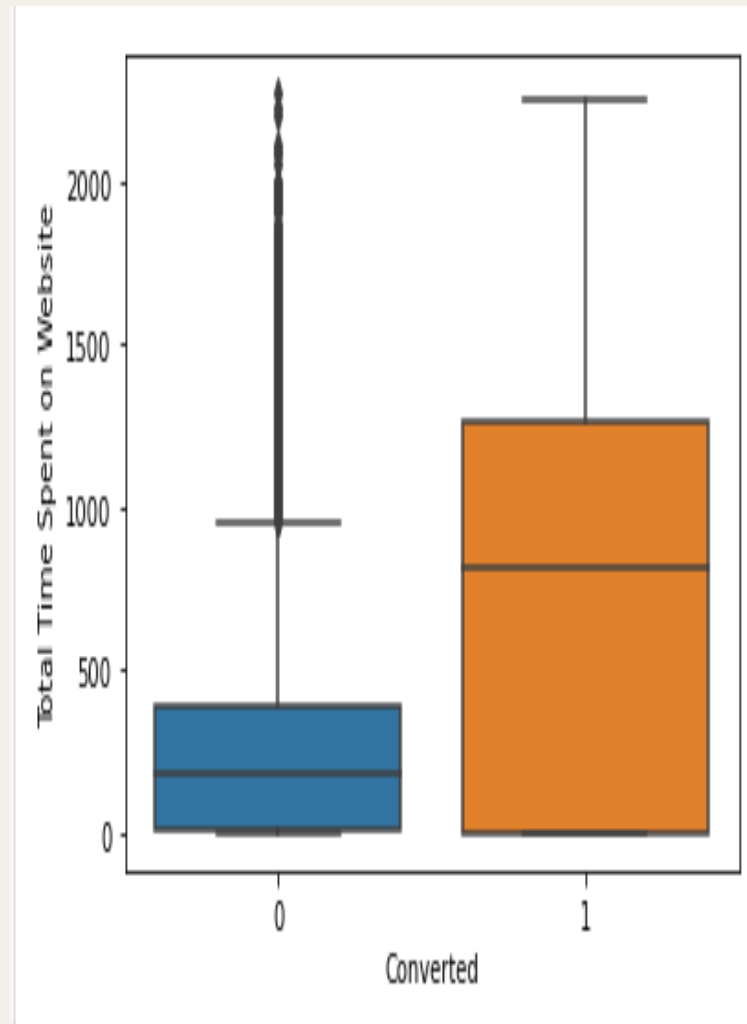
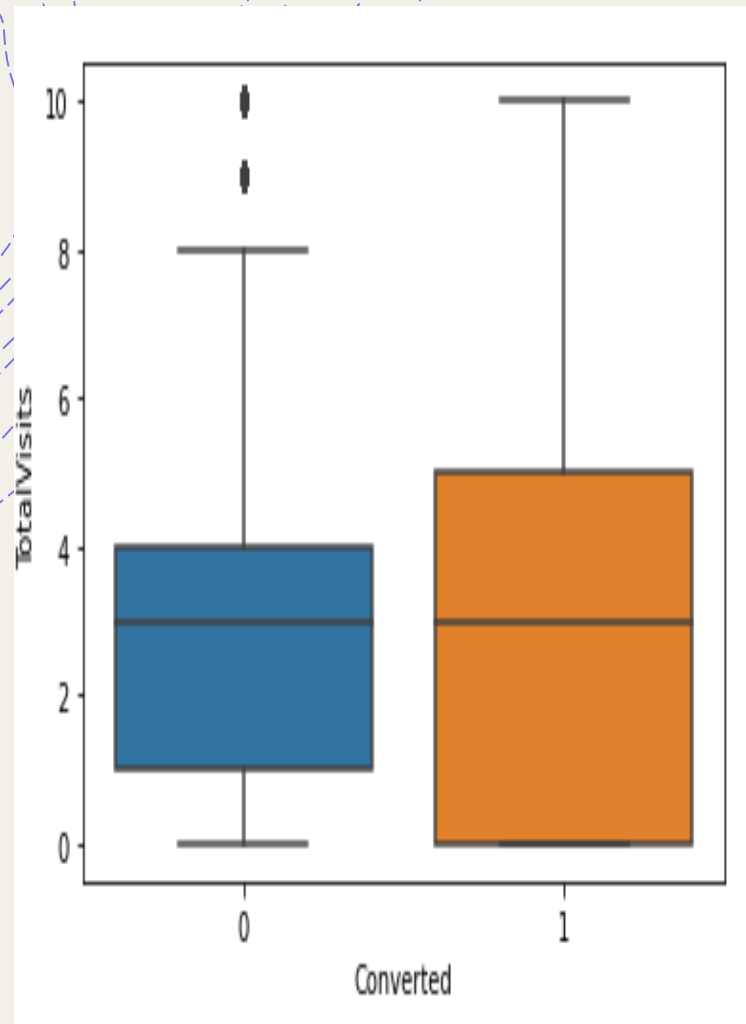
Major conversion has been performed from '**Email sent**' and '**calls made**'



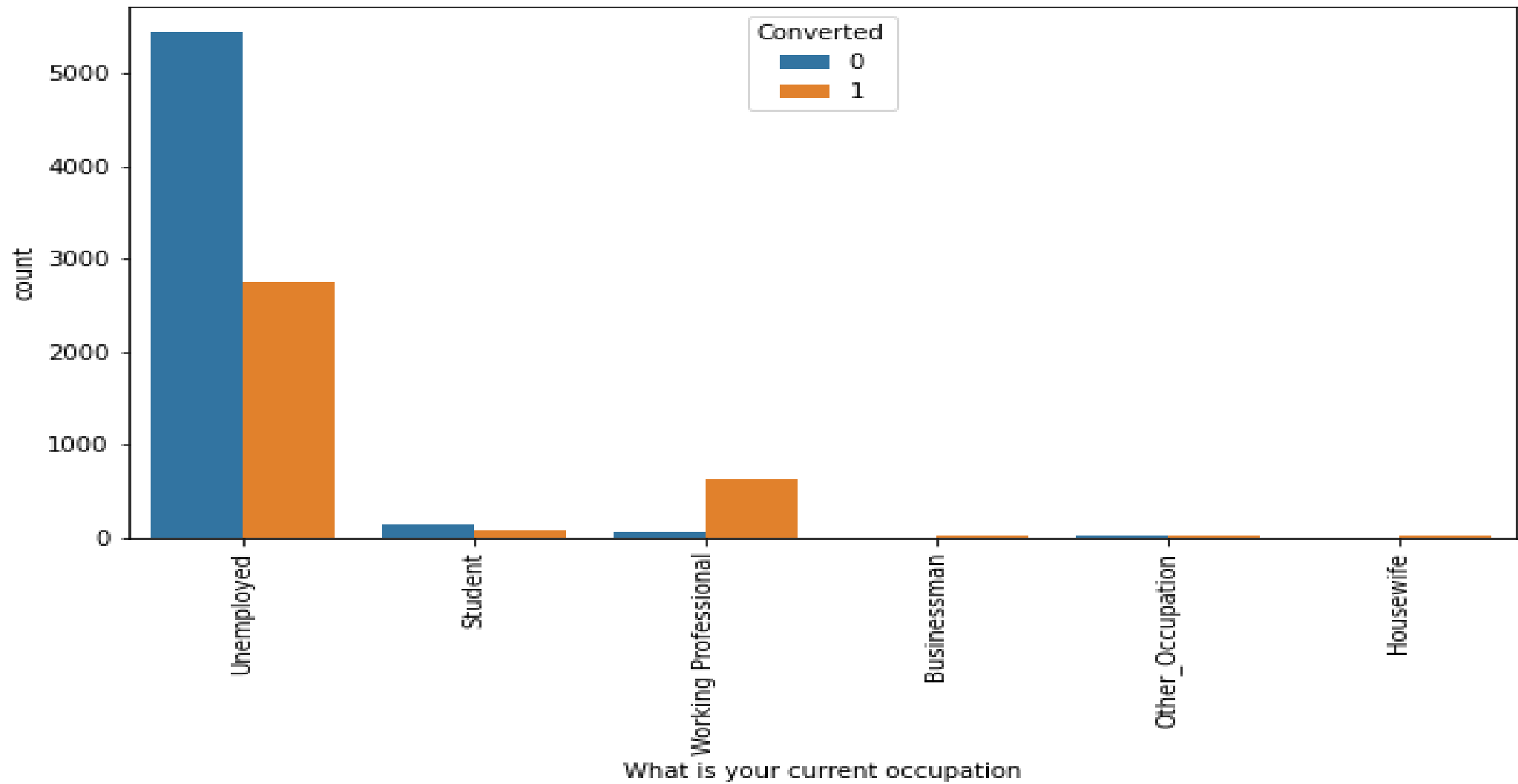
The count of Last Activity for '**SMS sent**' is higher than others.



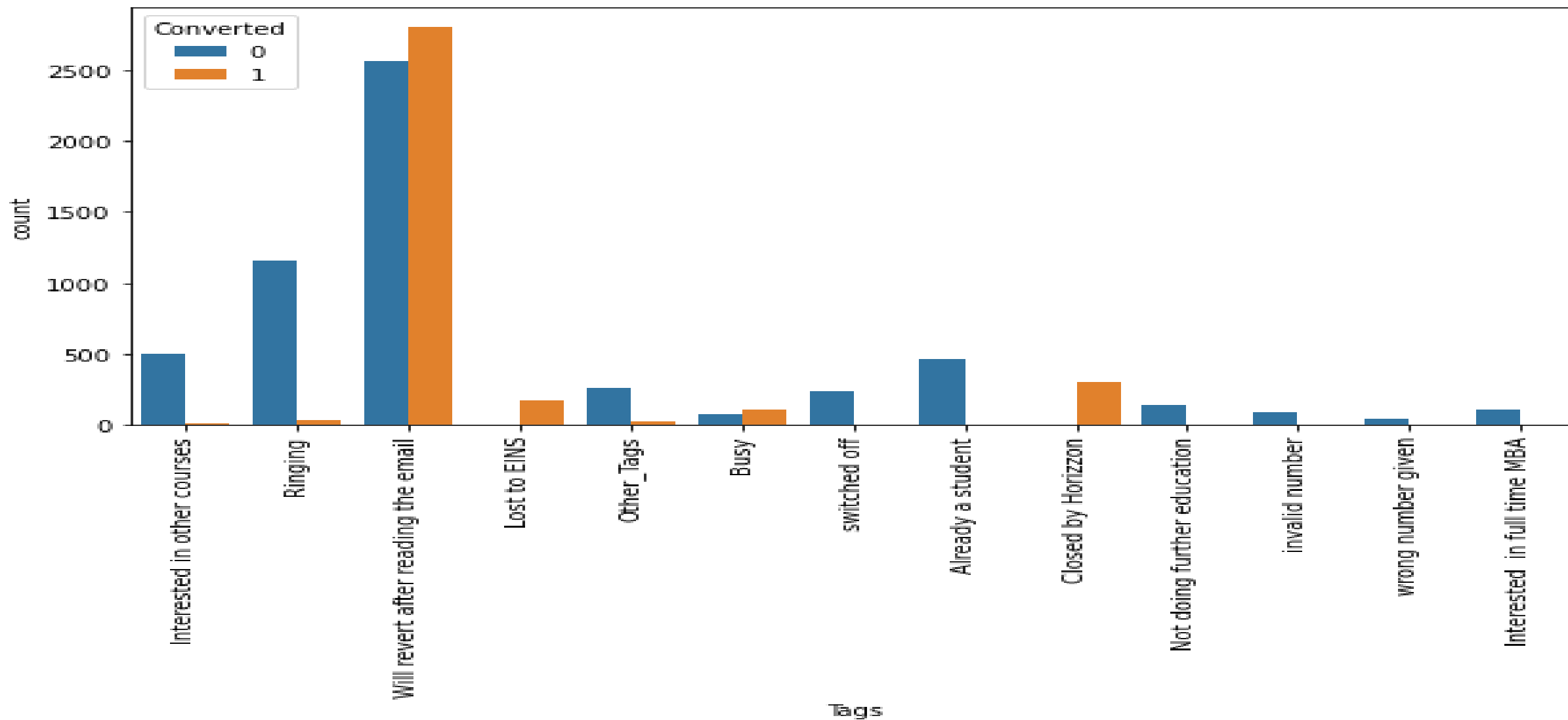
'Landing Page Submission' count is higher than others..



It describes the variation of numerical columns whoever converted or not.



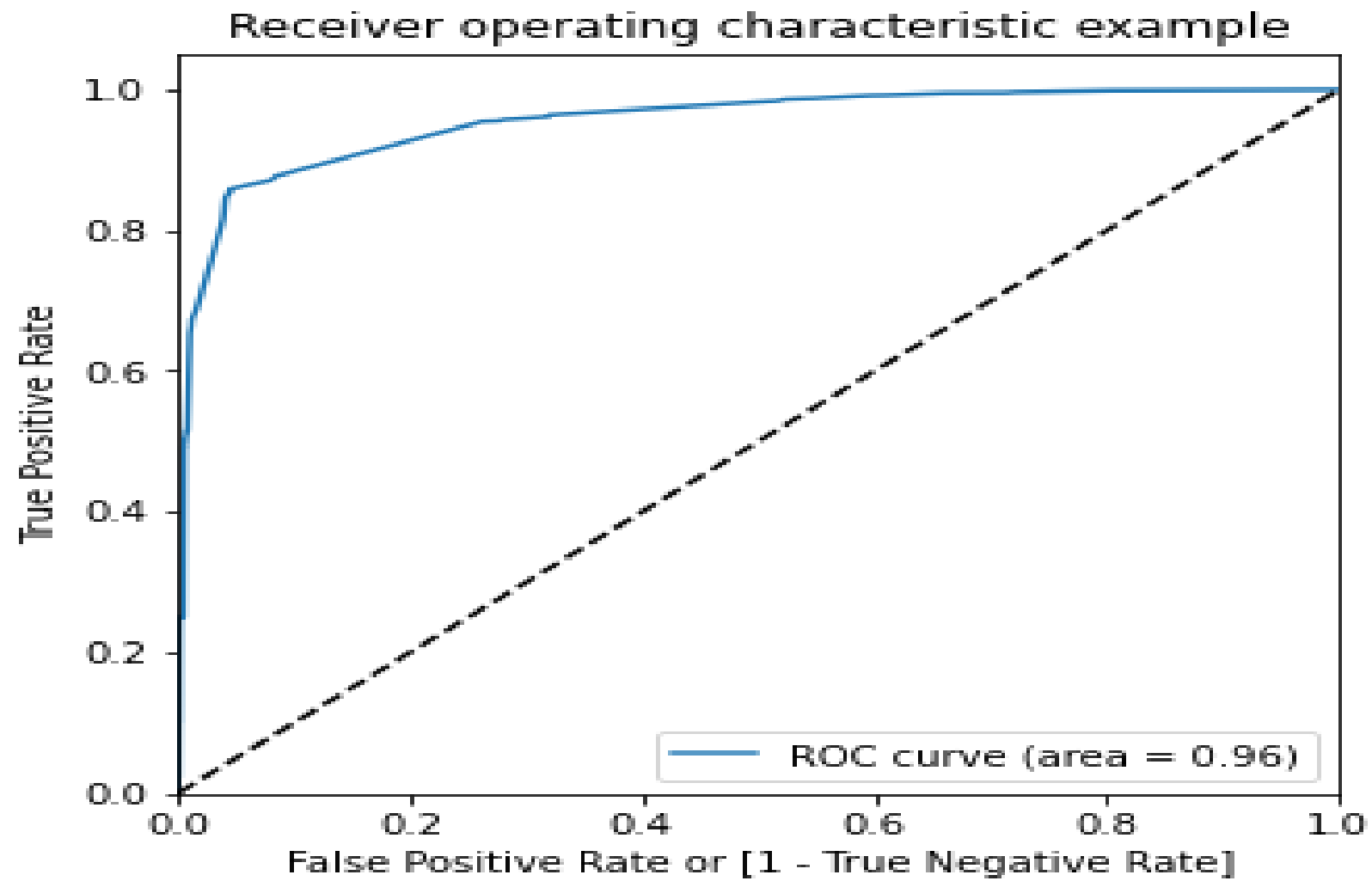
Unemployed people are higher converted than others



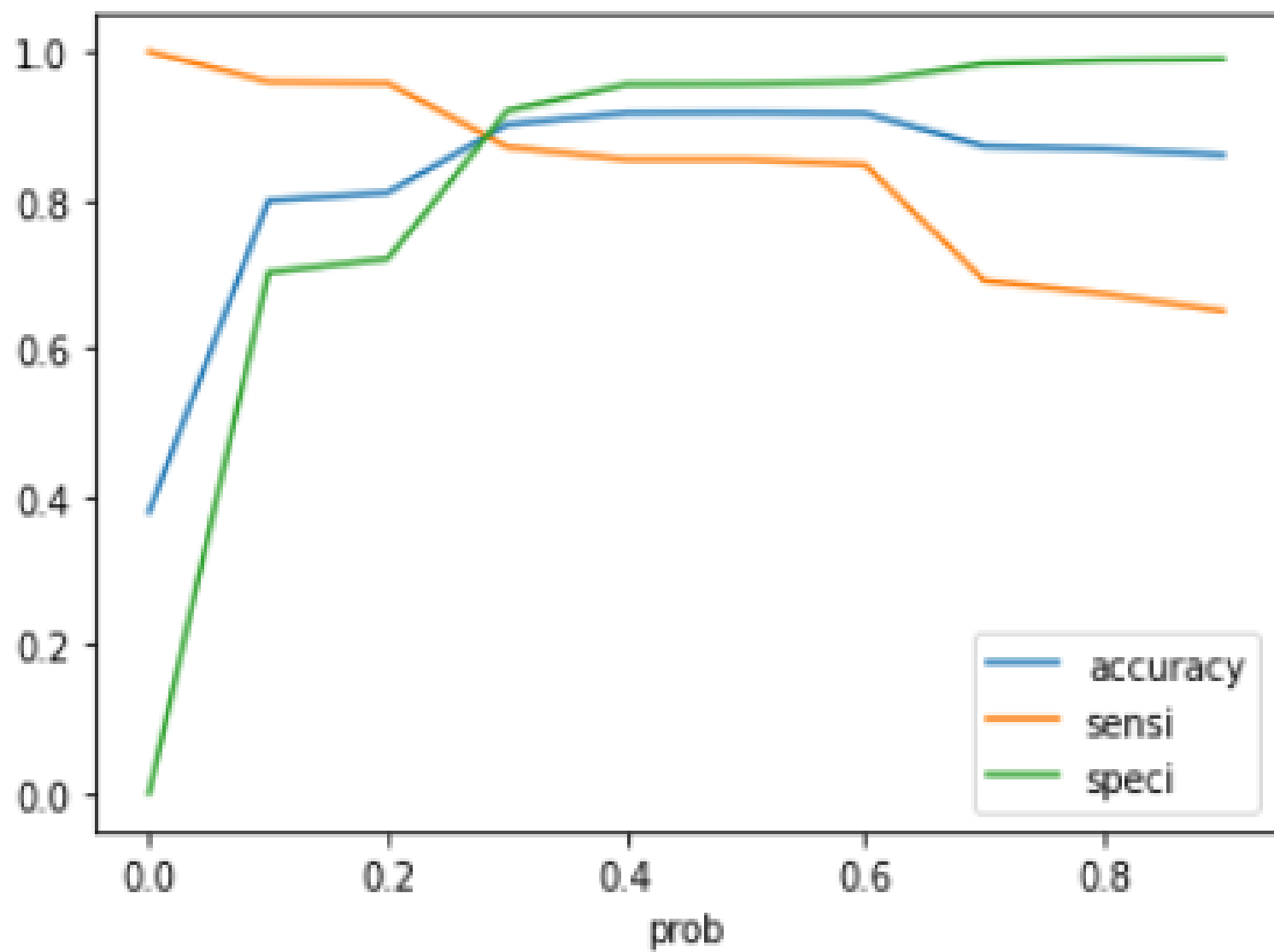
The count of 'tags' for 'Will revert after reading the email' is much higher than others.

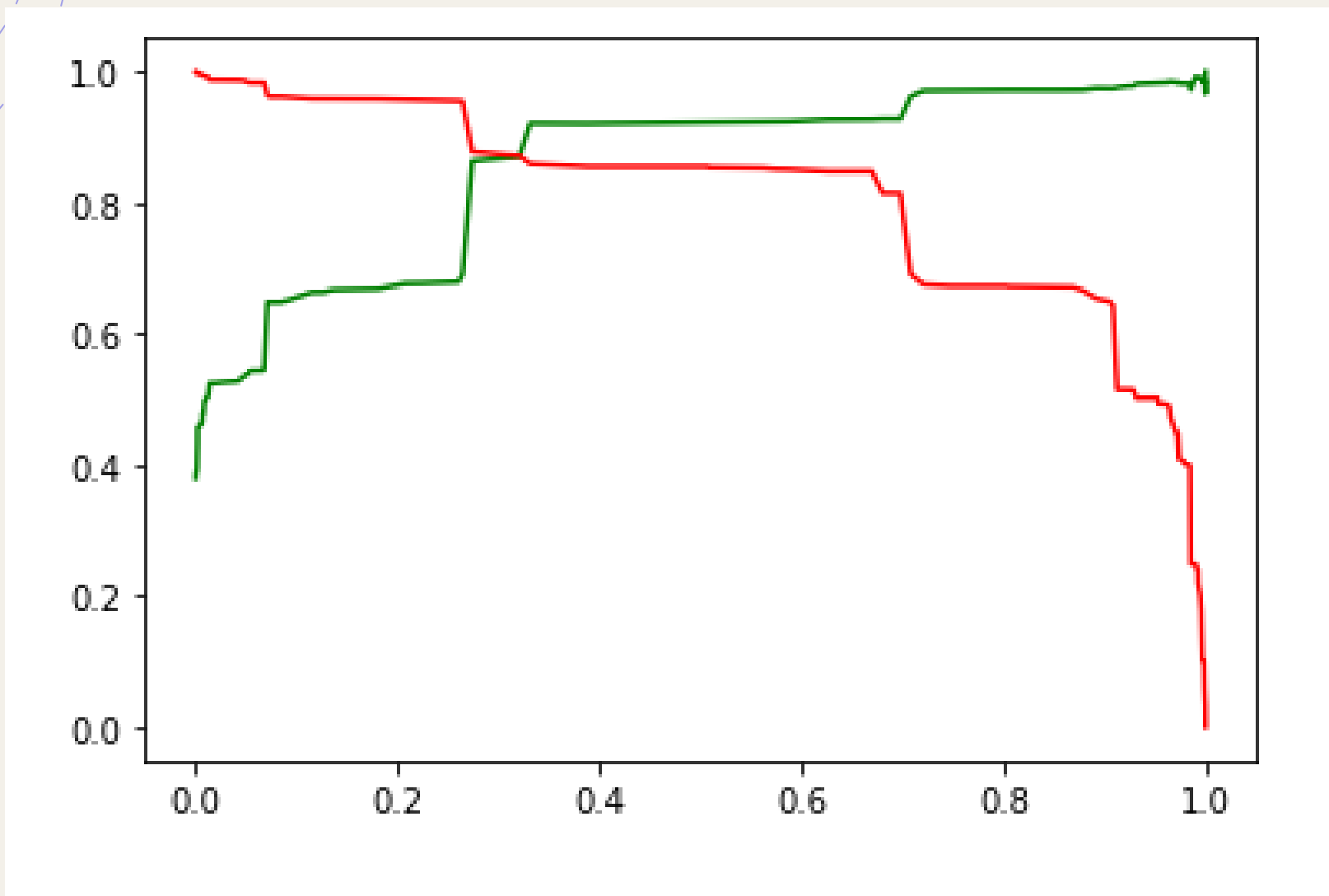
The background is a light beige color. In the top-left corner, there is a white circle partially cut off by the edge, with several blue dashed wavy lines flowing downwards and to the right. In the bottom-right corner, there is another white circle partially cut off, with blue dashed wavy lines flowing upwards and to the left. A solid orange line also flows from the bottom-left towards the bottom-right, curving slightly upwards.

Model Evaluation



Linear Regression: ROC curve = 0.96.





Conclusion

- We have calculated Sensitivity, specificity as well as Precision, and Recall Metrics.
- Sensitivity, Specificity, and accuracy values on test around 95%, 73%, and 81%
- These are the top 3 variables that contribute most towards the probability of a lead getting converted
 - a) Total Visits
 - b) Total Time Spent on Website
 - c) Page Views Per Visit
- We should also focus on increasing sending messages, engagement on website, and advertising.
- Inform about new courses, services, job offers, and future higher studies. Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information, or courses that suits best according to the interest of the leads