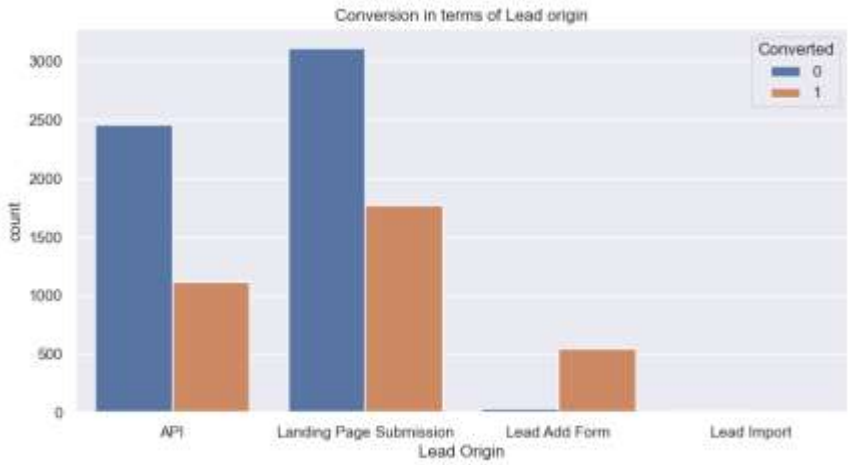
Lead Score Analysis Presentation

- By

Meghana Chinta
Venkatesh Udutha
Gaurav Ojha

Univariate Analysis and Bi-variate Analysis

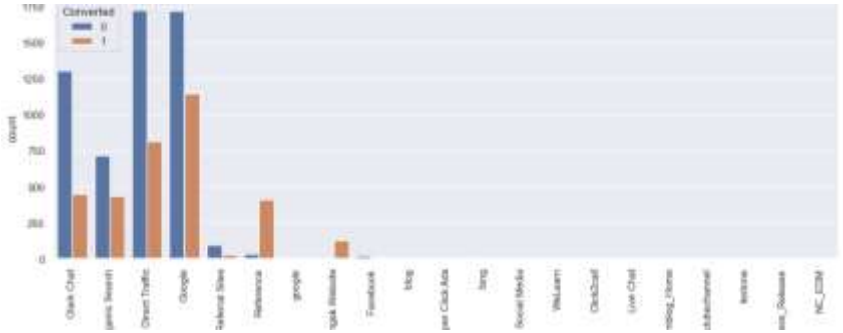


Lead Origin

Inferences

We may deduce the following from the plot description and Lead origin conversion shown above:

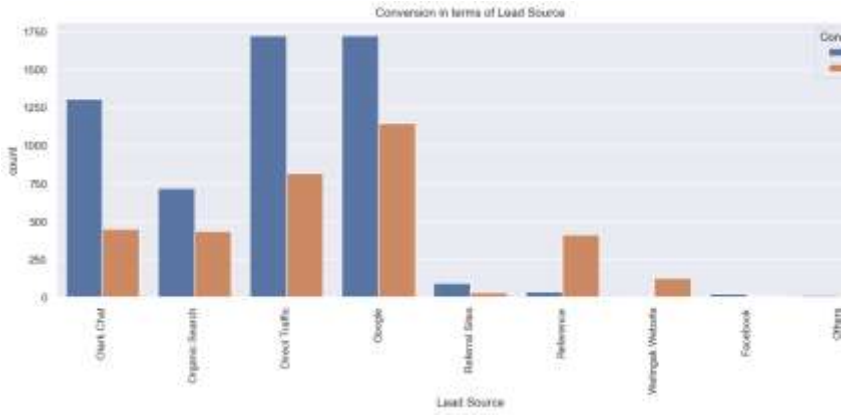
- The greatest conversion rate is 94% for Lead Add Form.
- Although they produce the most leads, API and Landing Page Submission have conversion rates of 31% and 36%, respectively.
- The least number of conversions and leads are generated through Lead Import.
- The lead conversion rate of API and Landing Page Submission should be improved in order to increase the total lead conversion rate. Additionally, create more leads with the Lead Add form as it has a high conversion rate.



Lead Source

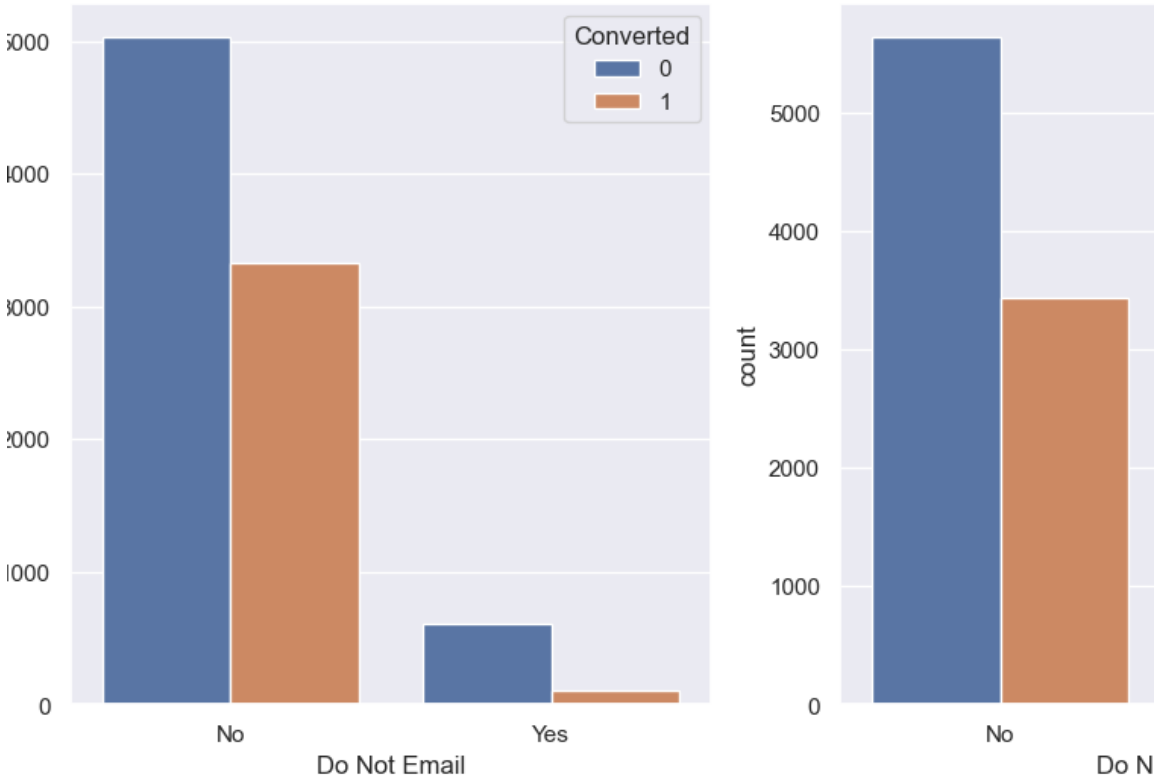
Inferences

- The number of Lead sources is relatively low. Therefore, we may group them together under the category "Others."
- Additionally, there are two categories that share the name Google. So, to have a single category, replace google with Google.



Inferences

- We may deduce the following from the plot description and Lead origin conversion shown above:
- The most leads are generated by Google and direct traffic, but their conversion rates are 40% and 32%, respectively.
- Although the Welingak website and References have the best conversion rates, at 98% and 93%, respectively, they provide less leads.
- Although they create a sizable amount of leads, Olark chat and organic search only convert about 26% and 38% of those leads, respectively.
- Lead sources in the 'others' category, such as 'Click2call', 'Live Chat', 'NC_EDM', 'Pay per Click Ads', 'Press Release', 'Social Media', 'WeLearn', ‘Bing', 'blog', 'testone', 'welearnblog_Home', and ‘YouTube channel', produce relatively few leads.
- The lead conversion of Google lead source, organic search, direct traffic, and Olark chat should be improved in order to increase total lead conversion rate. Additionally, since reference and Welingak websites have excellent conversion rates, create additional leads from them.



Do Not Email & Do Not Call

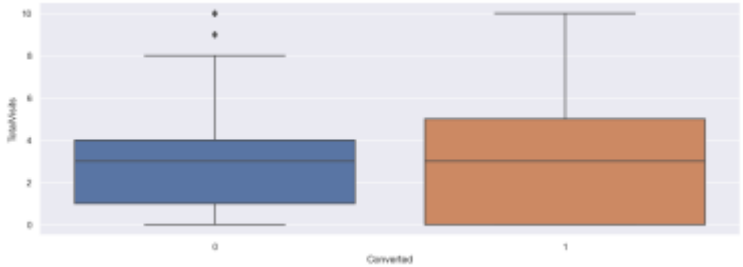
Inferences

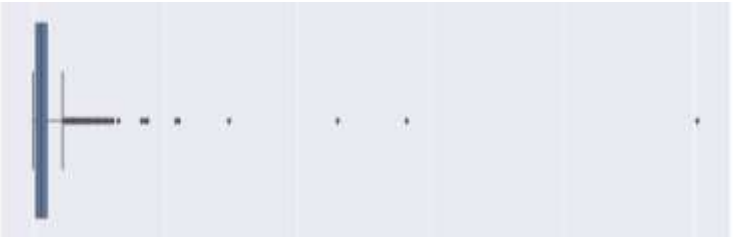
- We may deduce the following from the plot and conversion description above:
- Approximately 99% of consumers dislike getting calls or emails regarding the course.

Total Visits

Inferences

- A few outliers may be seen in the Total Visits column. The outliers will be capped at 95%.

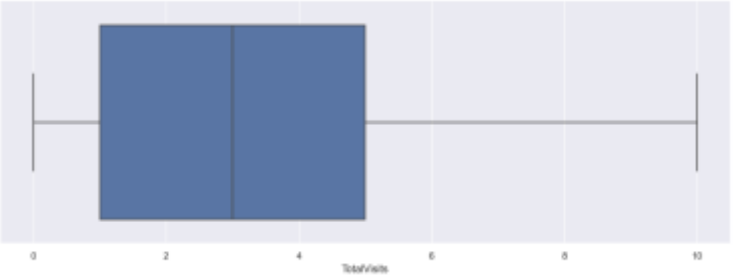




Inferences

From the boxplot shown above, it is clear that:

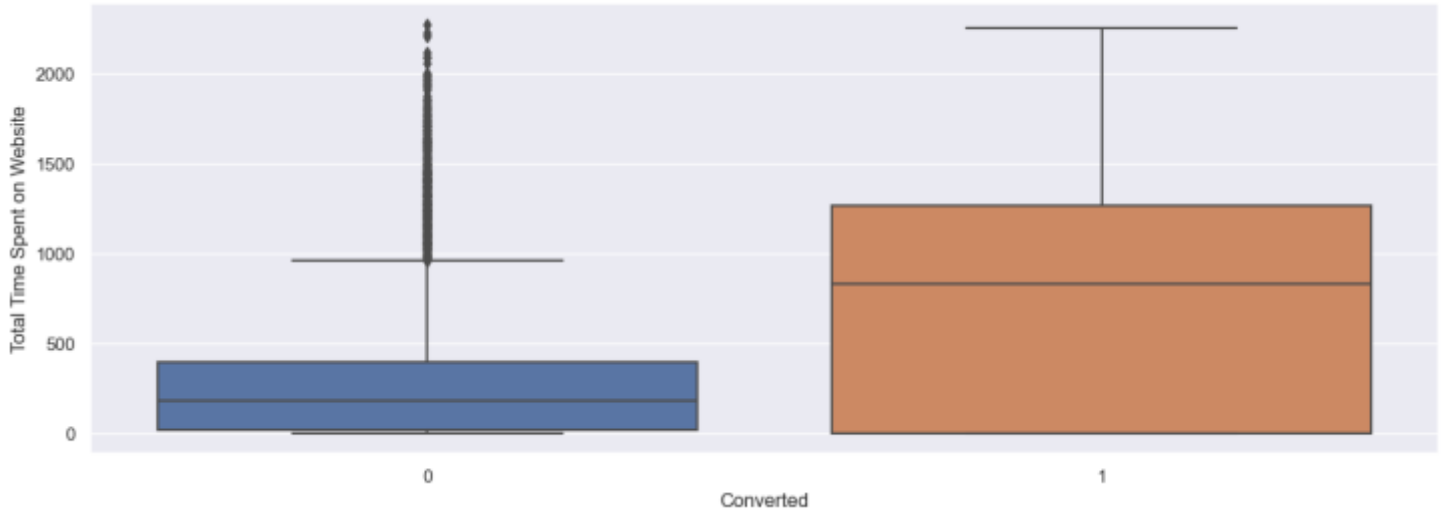
- The median is the same for both converted and non-converted leads.
- Everyone who accesses the site has a 50/50 probability of applying for the course or not.



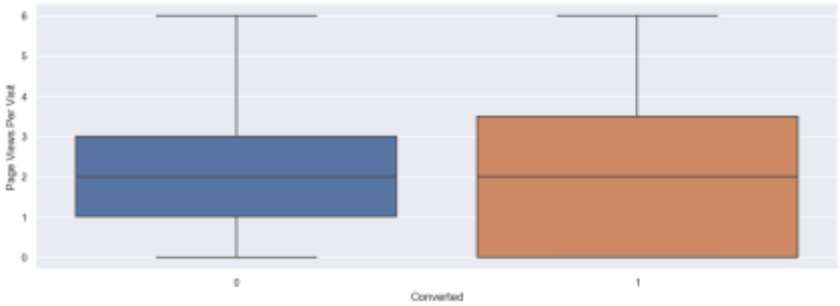
Total Time Spent On Website

Inferences

- From the boxplot shown above, it is clear that:
- People who spend more time on websites are more likely to enroll in a course.
- Less frequent website users didn't sign up for any courses.

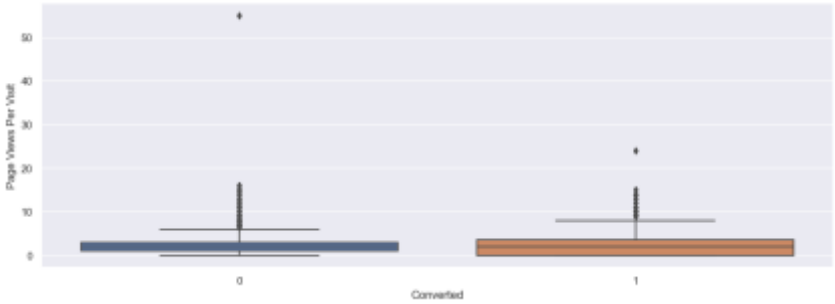


Page Views per visit



Inferences

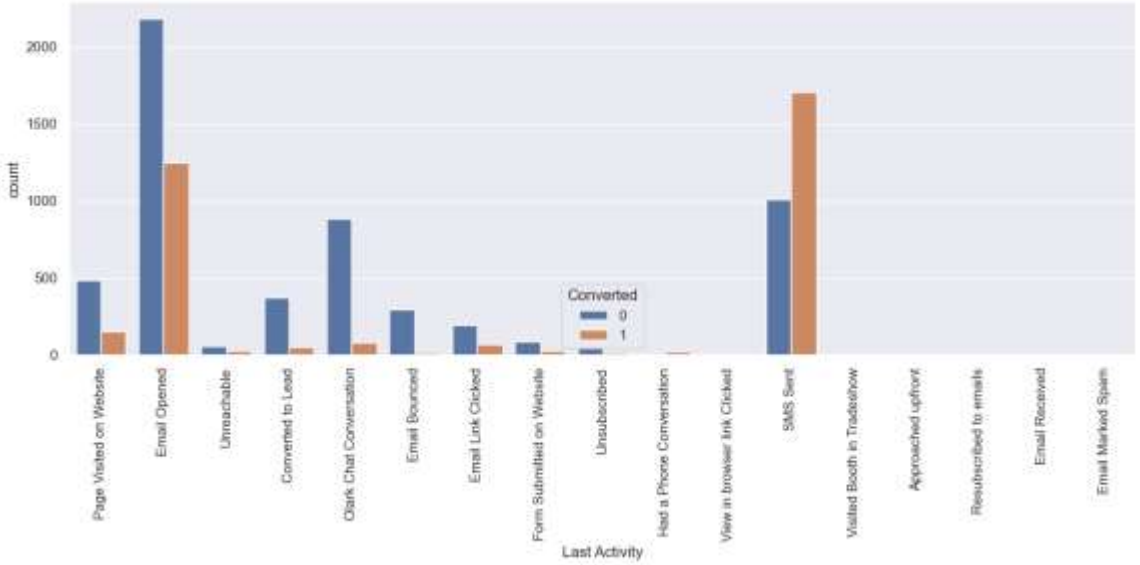
- In the column for Page Views Per Visit, there are several outliers.The outliers will be capped at 95%.



Last Activity

Inferences

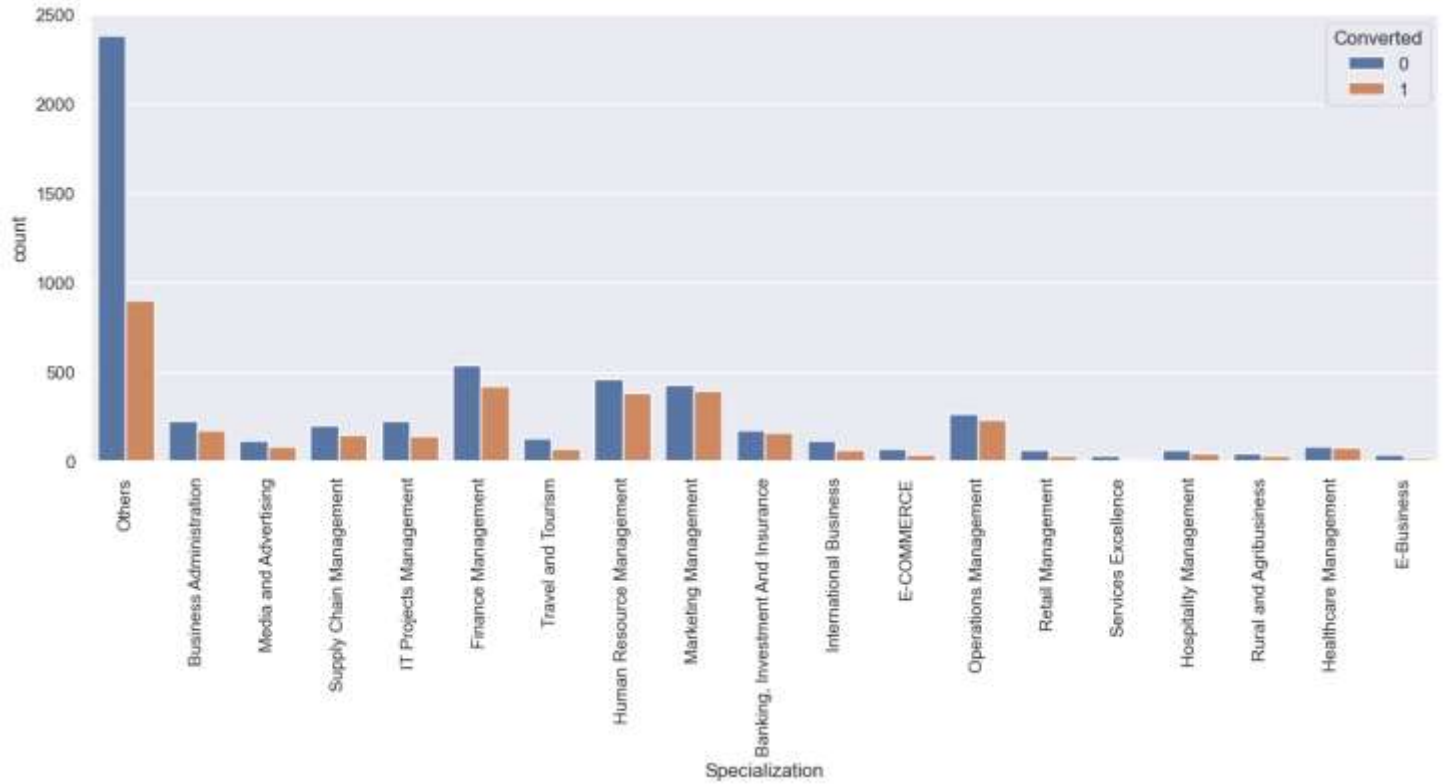
- From the boxplot shown above, it is clear that:
- The median is the same for both converted and non-converted leads.
- People who browse an average of 1 to 3 pages on the website have a 50/50 probability of applying or not for the course.
- The likelihood of conversion is higher for those who never view any pages.



Specialization

Inferences

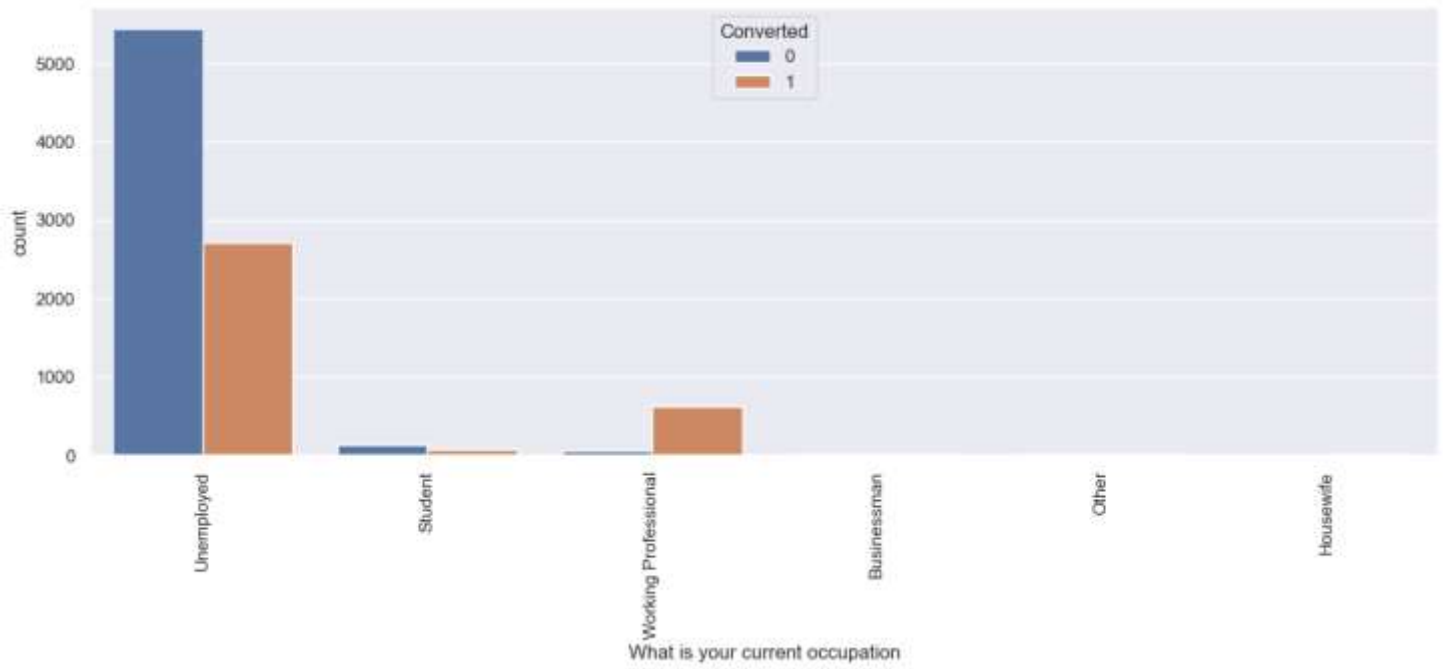
- We may deduce the following from the plot and conversion description above:
- Most specializations have a conversion rate of between 40 and 50 percent.



Occupation

Inferences

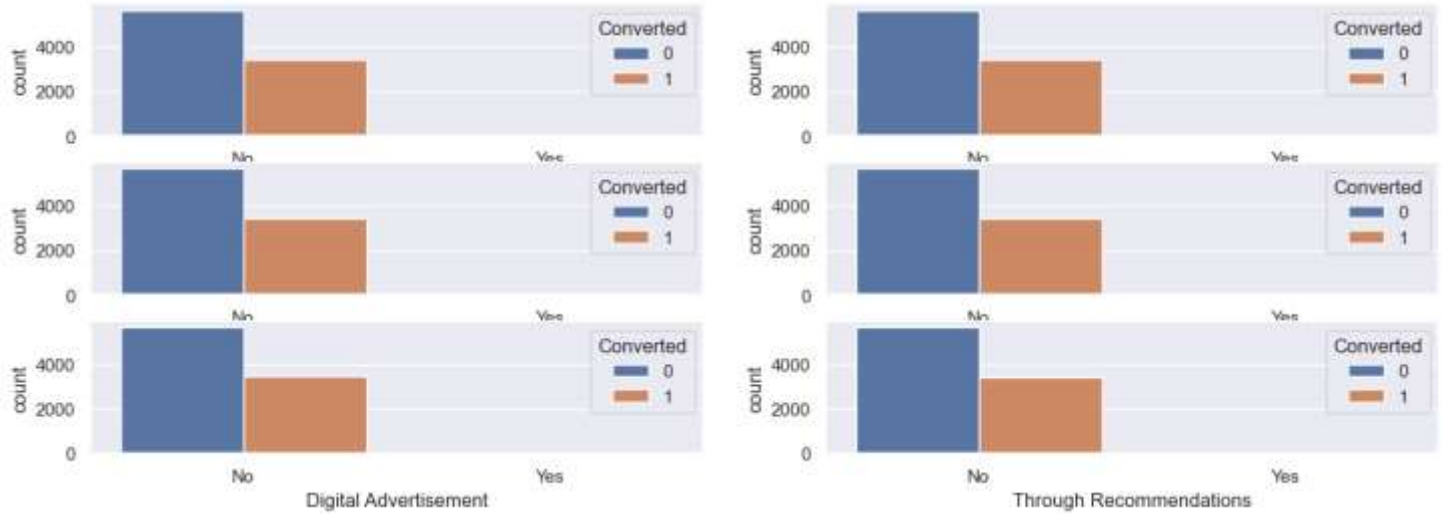
- We may deduce the following from the plot and conversion description above:
- Maximum leads are generated by employed professionals and jobless individuals.
- Professionals in the workforce convert at a high rate of over 92%, whereas unemployed people convert at a rate of about 33%.
- The lead conversion rate of jobless people should be improved in order to increase total lead conversion rate. Increase the number of leads you get from working professionals.



Search, Newspaper article , X Education Forums , Newspaper, Digital Advertisement ,Through Recommendations

Inferences

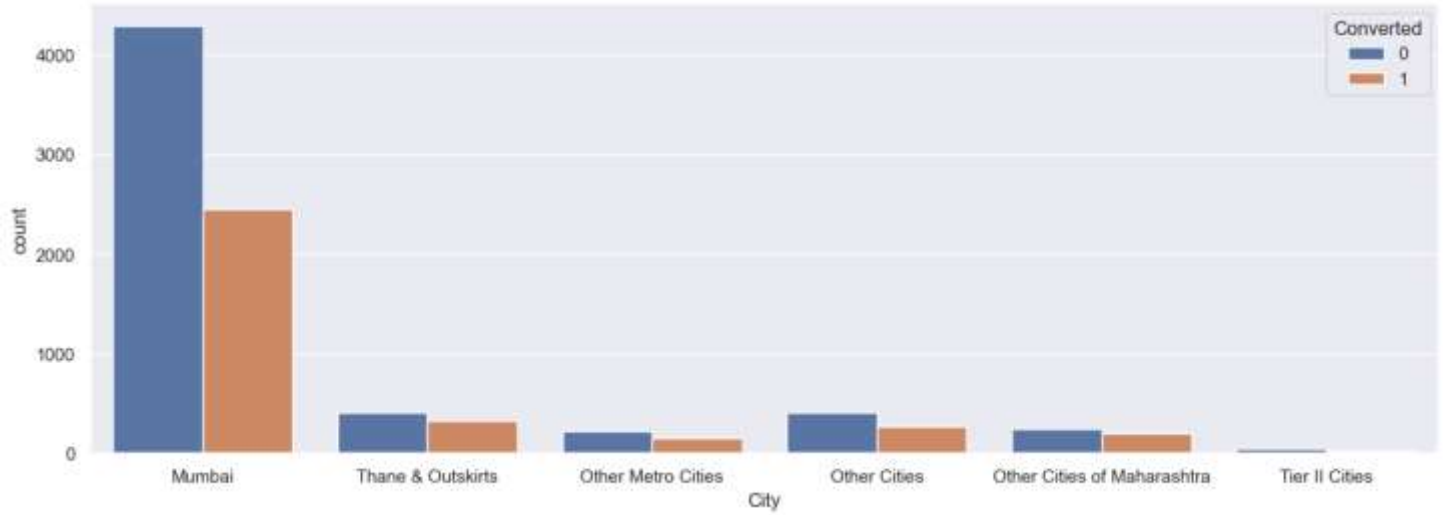
- Nearly 99% of buyers have not come across the X Education advertisement through search, newspaper articles, forums, newspapers, digital advertisements, or recommendations.



City

Inferences

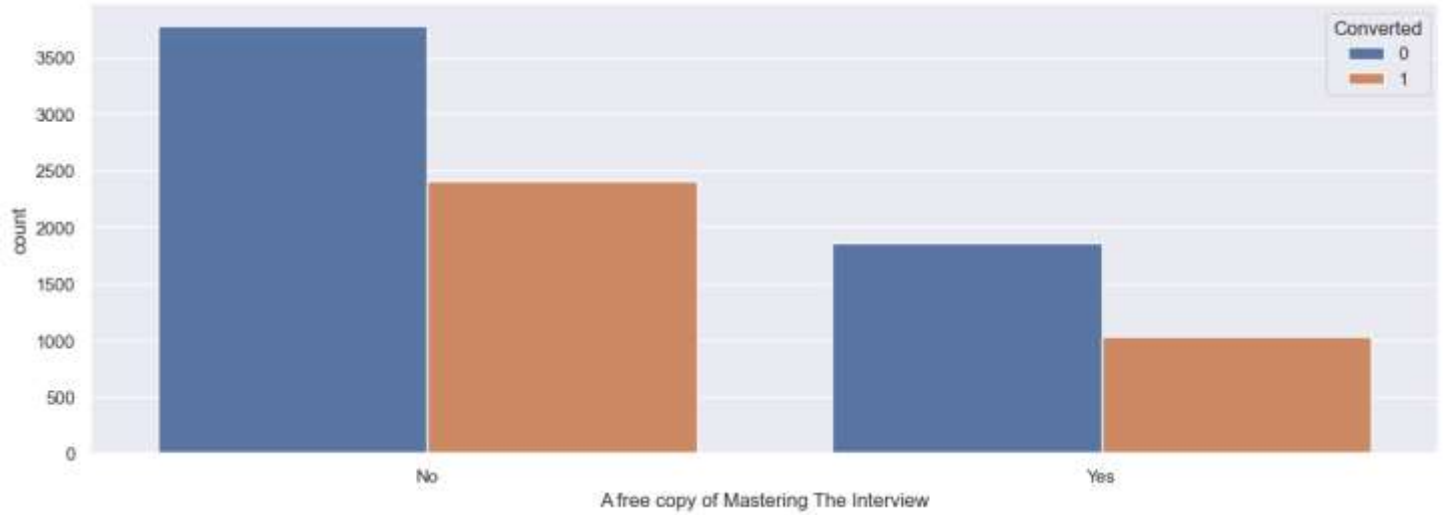
- Mumbai generates the most leads, with a conversion rate of about 36%.Therefore, the focus should be on raising Mumbai's conversion rate.



A free copy of Mastering The Interview

Inferences

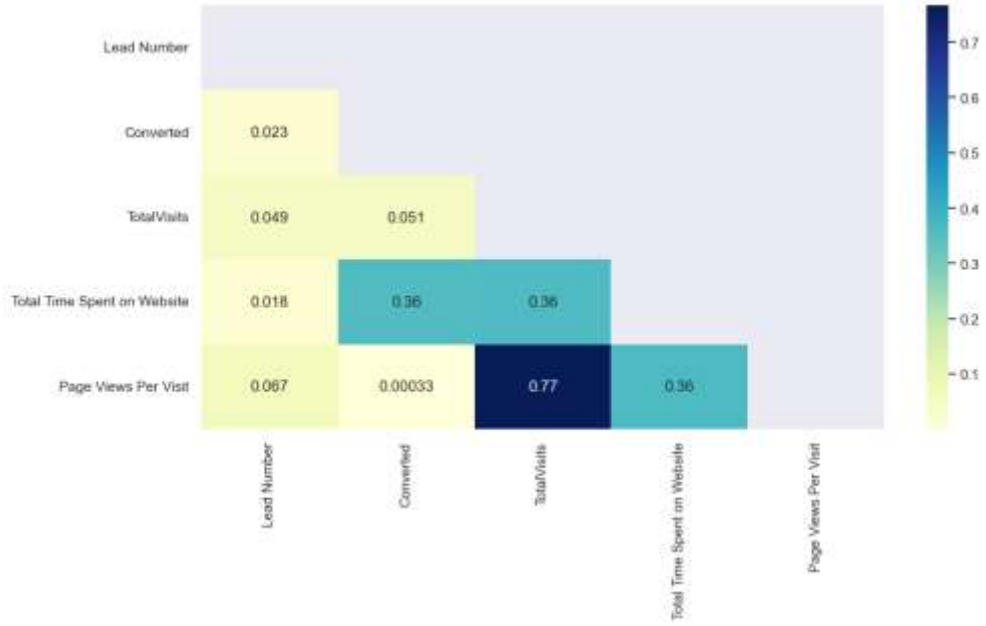
- Most buyers weren't interested in the complimentary copy of Mastering The Interview.
- Customers who chose the free copy had a 36% conversion rate, while those who did not choose had a 39% conversion rate.



Last Notable Activity

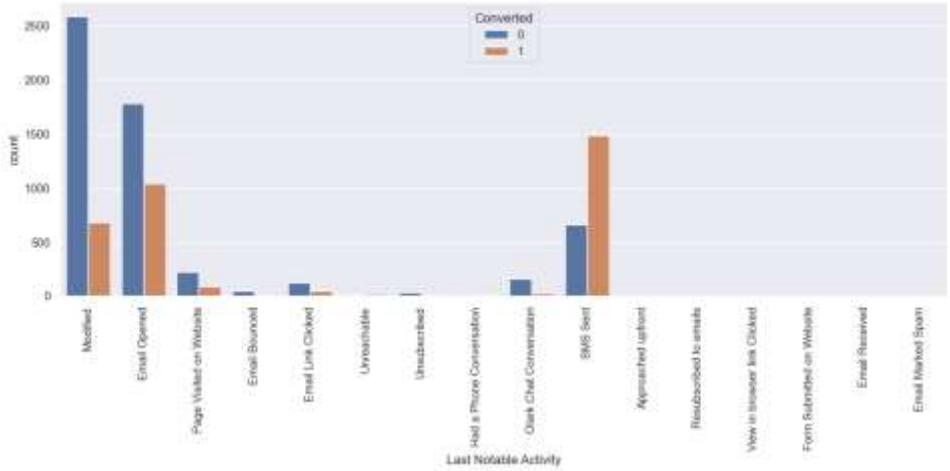
Inferences

- The columns for overall visits and page views per visit are connected.
- Thus, to prevent multi-collinearity, we should include any of these columns in our model.



Inferences

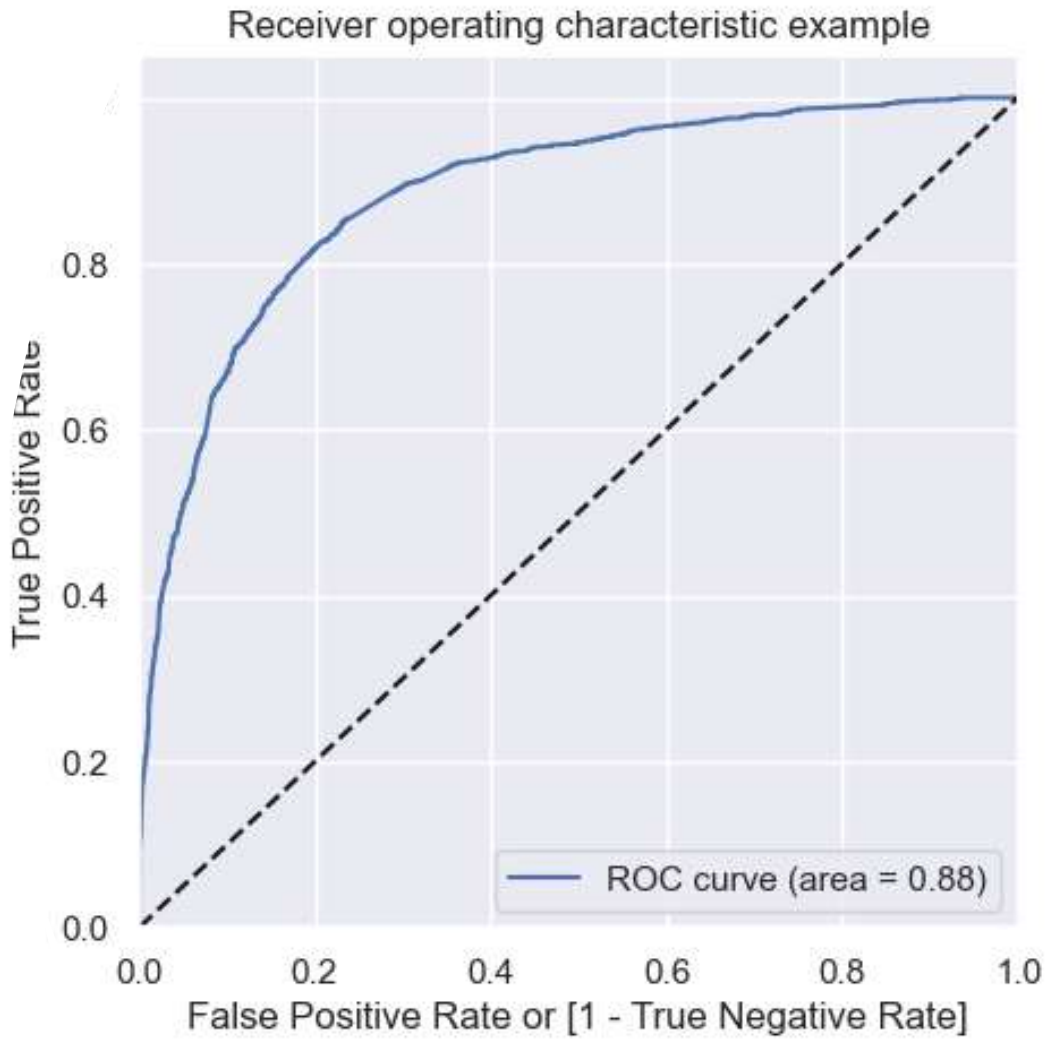
- This column and the Last Activity column are extremely similar.



ROC Curve

Inferences

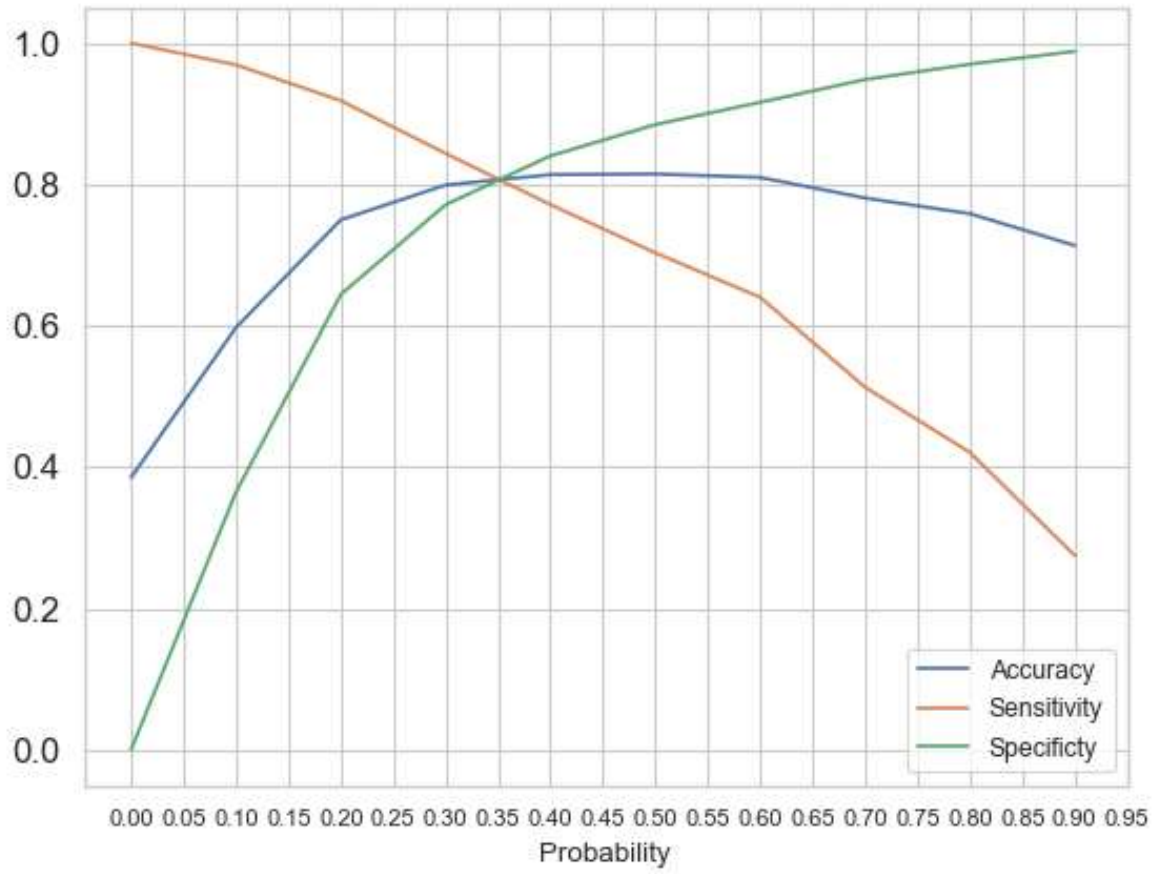
- The ideal value for the ROC Curve is close to 1. A result of 0.88 is returned, which denotes a strong prediction model.



Optimal Cutoff Point

Inferences

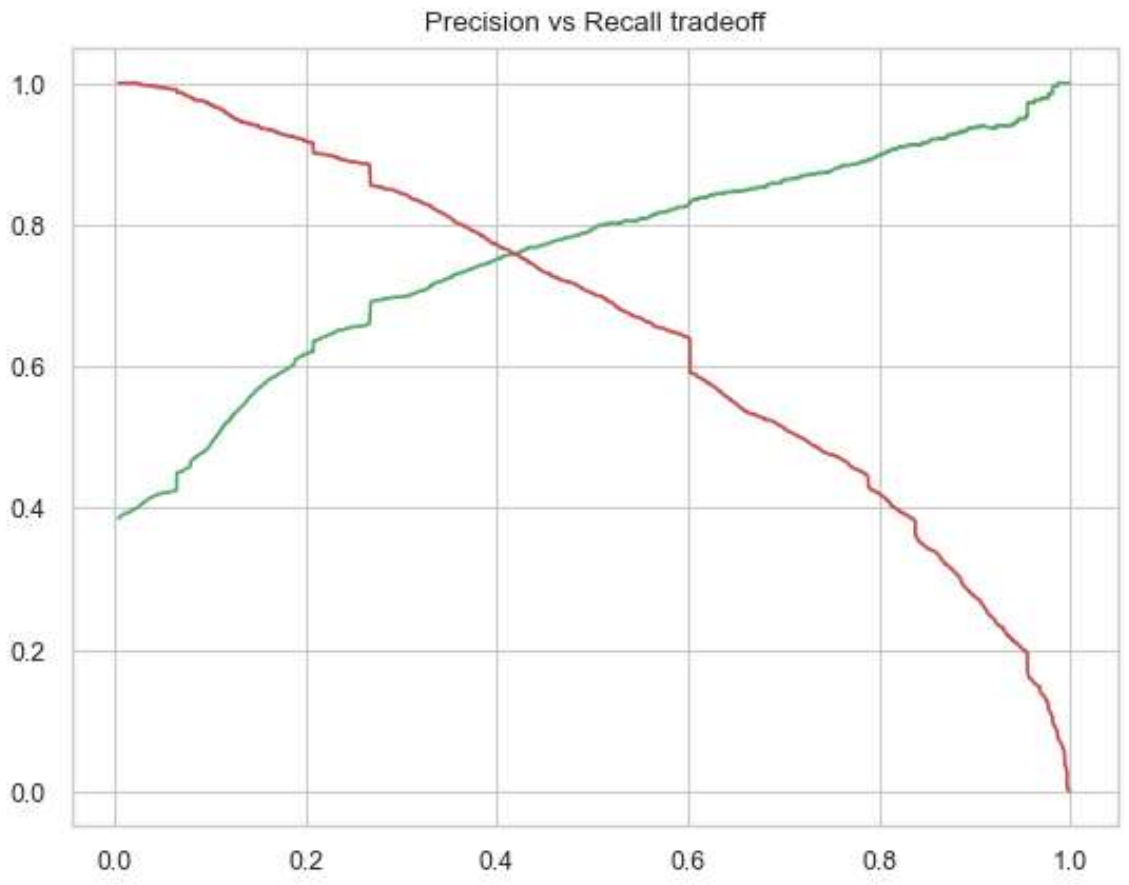
- The ideal cutoff is at 0.35, as seen by the graph above. At this moment, accuracy, sensitivity, and specificity are all equally balanced.



Metrics - Precision and Recall

Inferences

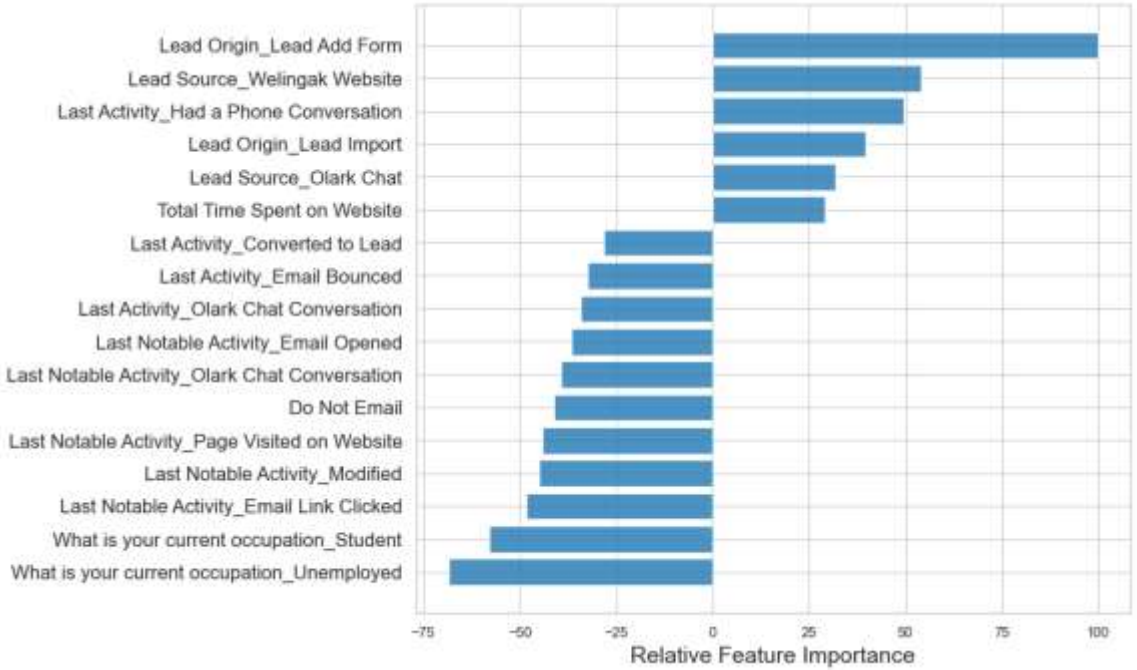
- The recall % is very important for our company purpose since we don't want to miss any hot leads who are ready to convert.
- Thus, recall of 81% indicates a good model.
- Precision and recall have an inverse relationship, so if one improves, the other will actually decline.



The feature variables based on their relative coefficient values

Inferences

- On performance measures comparing train and test data, we found a 1% difference. This suggests that our final model is working well and did not overfit the training set.
- High specificity will guarantee that leads who are on the cusp of being converted or not are not picked, whilst high sensitivity will ensure that practically all leads who are likely to convert are accurately predicted.
- In accordance with the needs of the company, we may raise or lower the probability threshold value, which will in turn affect the model's sensitivity and specificity.



Thank You