Lead Scoring Case Study

By-Muskan Ahuja

Problem Statement

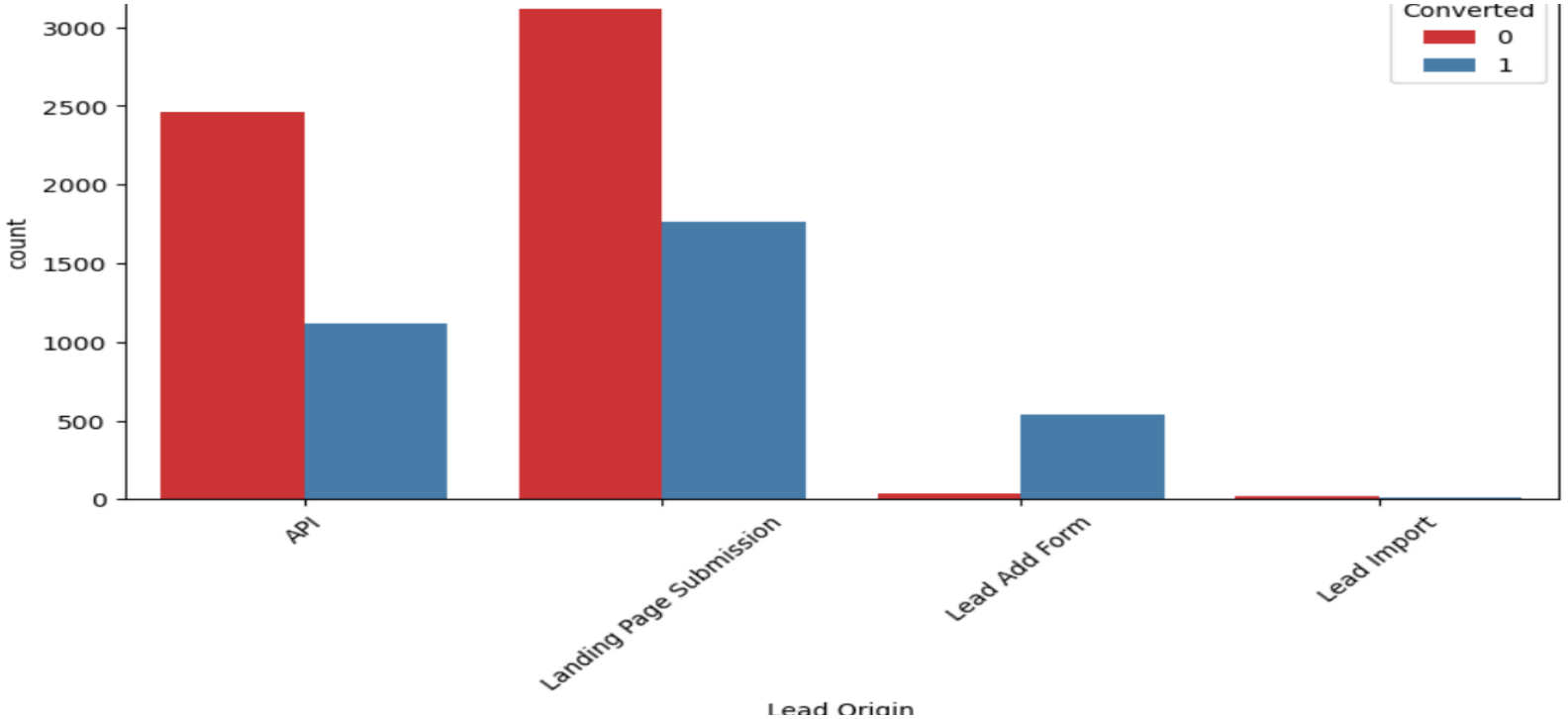
- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
-
- The company markets its courses on several websites and search engines like Google. Once these 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%.
-
- Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone. A typical lead conversion process can be represented using the following funnel:
- Lead Conversion Process - Demonstrated as a funnel
- As you can see, there are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. 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 a higher lead score have a higher conversion chance and the customers with a 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%.

Analysis Approach

- Import data
- Clean data for further analysis
- Exploratory data analysis to find the important attributes
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evaluation

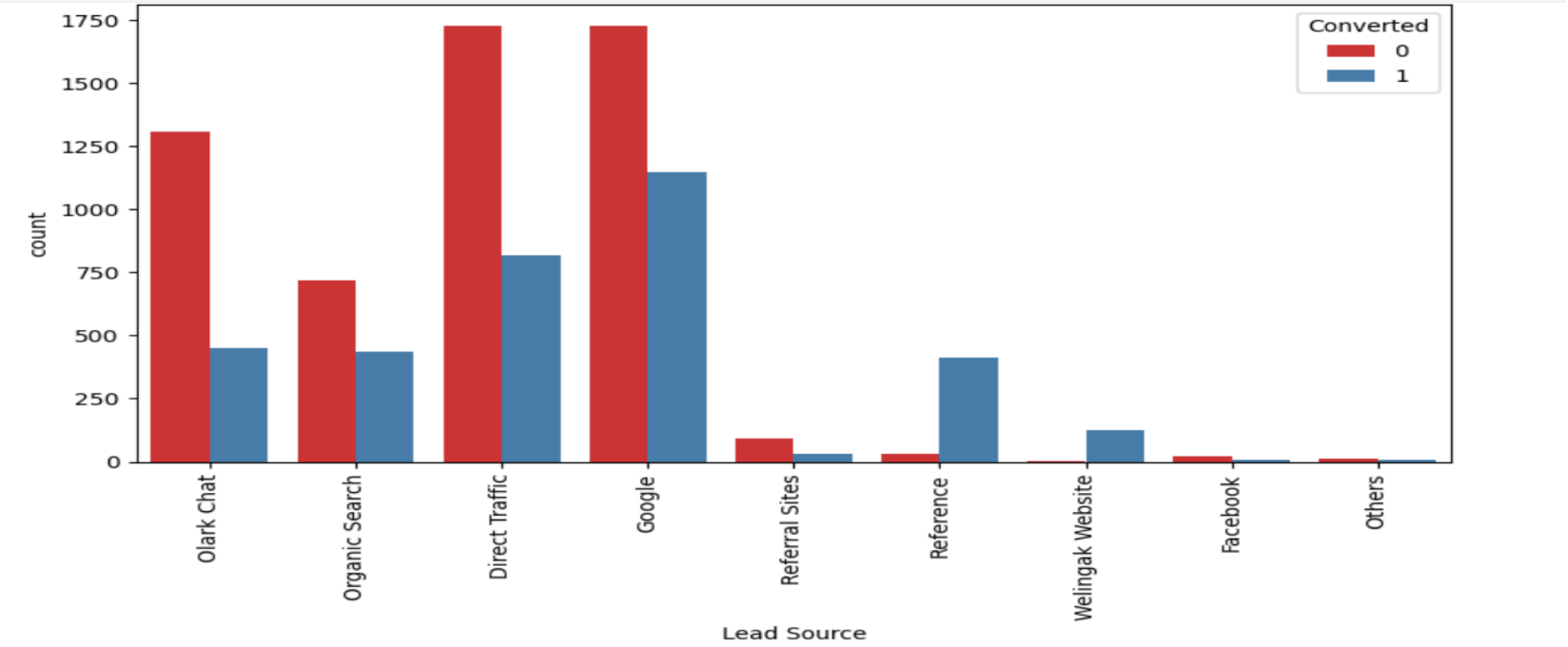
Exploratory Data Analysis

1. Lead Origin



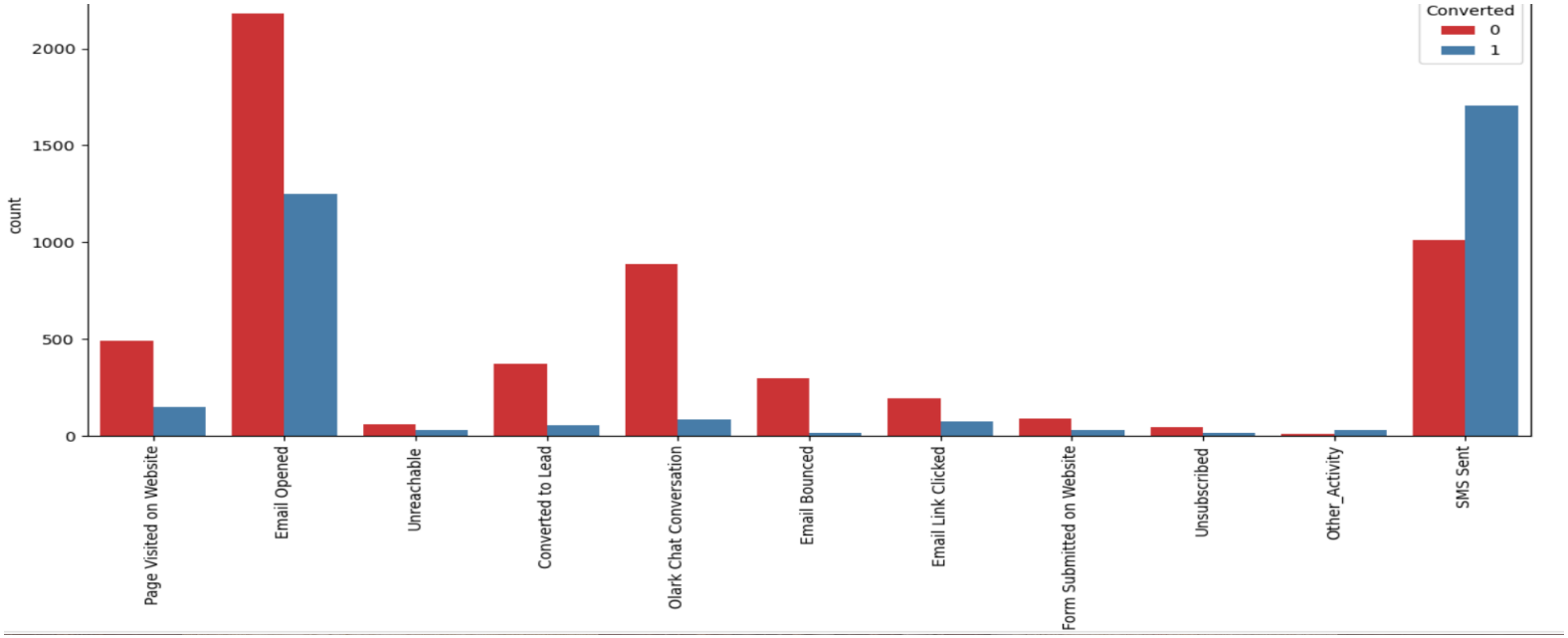
- The conversion rate for API and Landing Page Submission ranges from 30-35%, with a substantial number of leads originating from them.
- The Lead Add Form boasts a conversion rate exceeding 90%, although the overall lead count is relatively modest.
- Lead Imports have a notably low count.

2. Lead Service



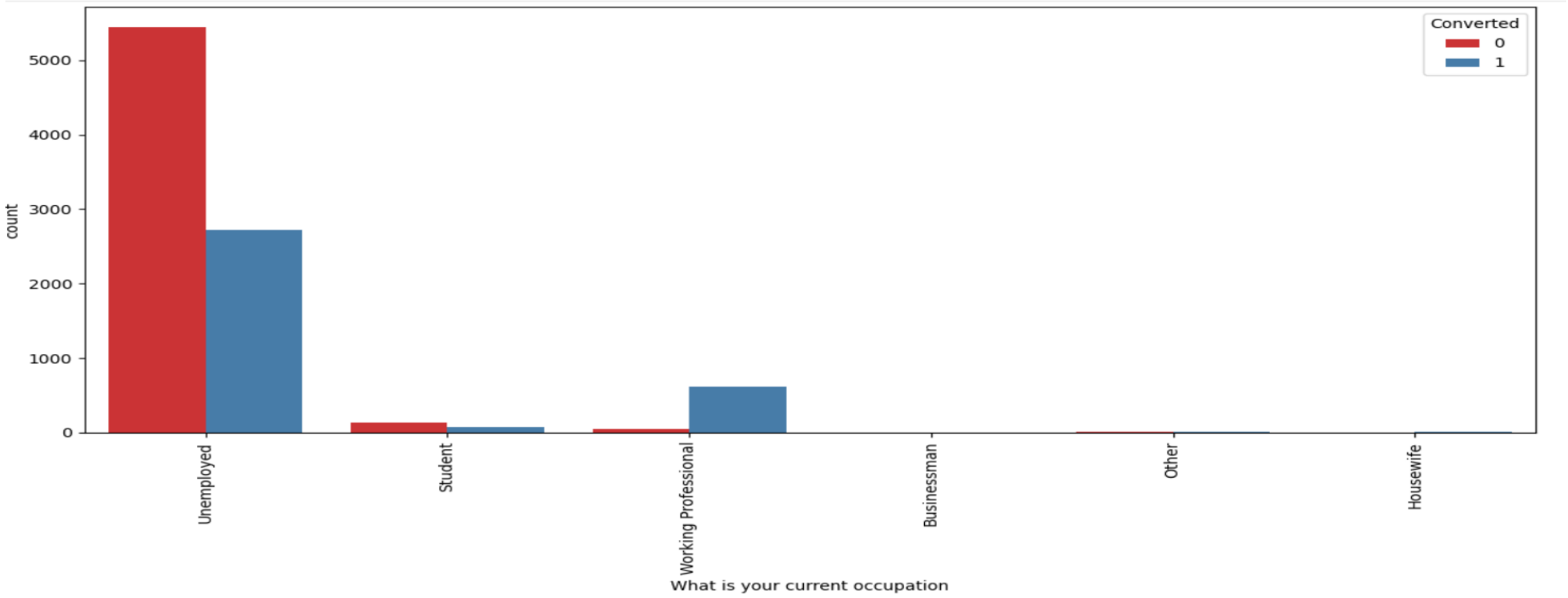
Google and Direct traffic yield the highest number of leads, while the conversion rates for reference leads and leads originating from the Welingak website are notably high.

3.Last Activity



1. The majority of leads exhibit their last activity as having their email opened.
2. Leads with their last activity recorded as SMS Sent demonstrate an almost 60% conversion rate.

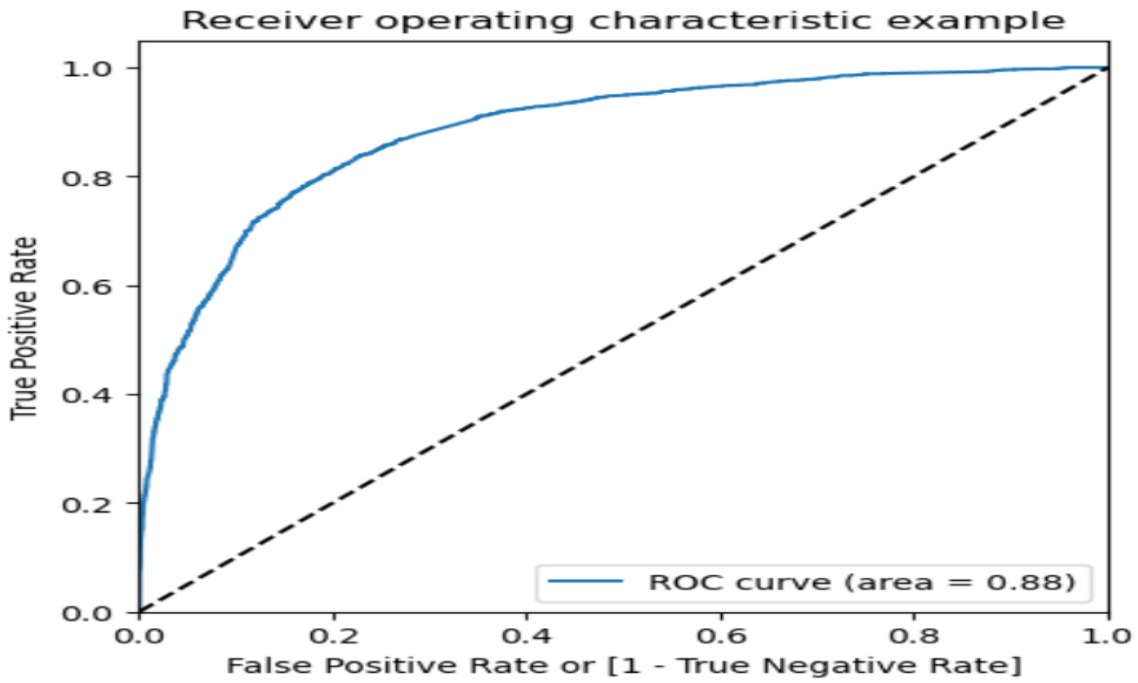
4. What is your current Occupation?



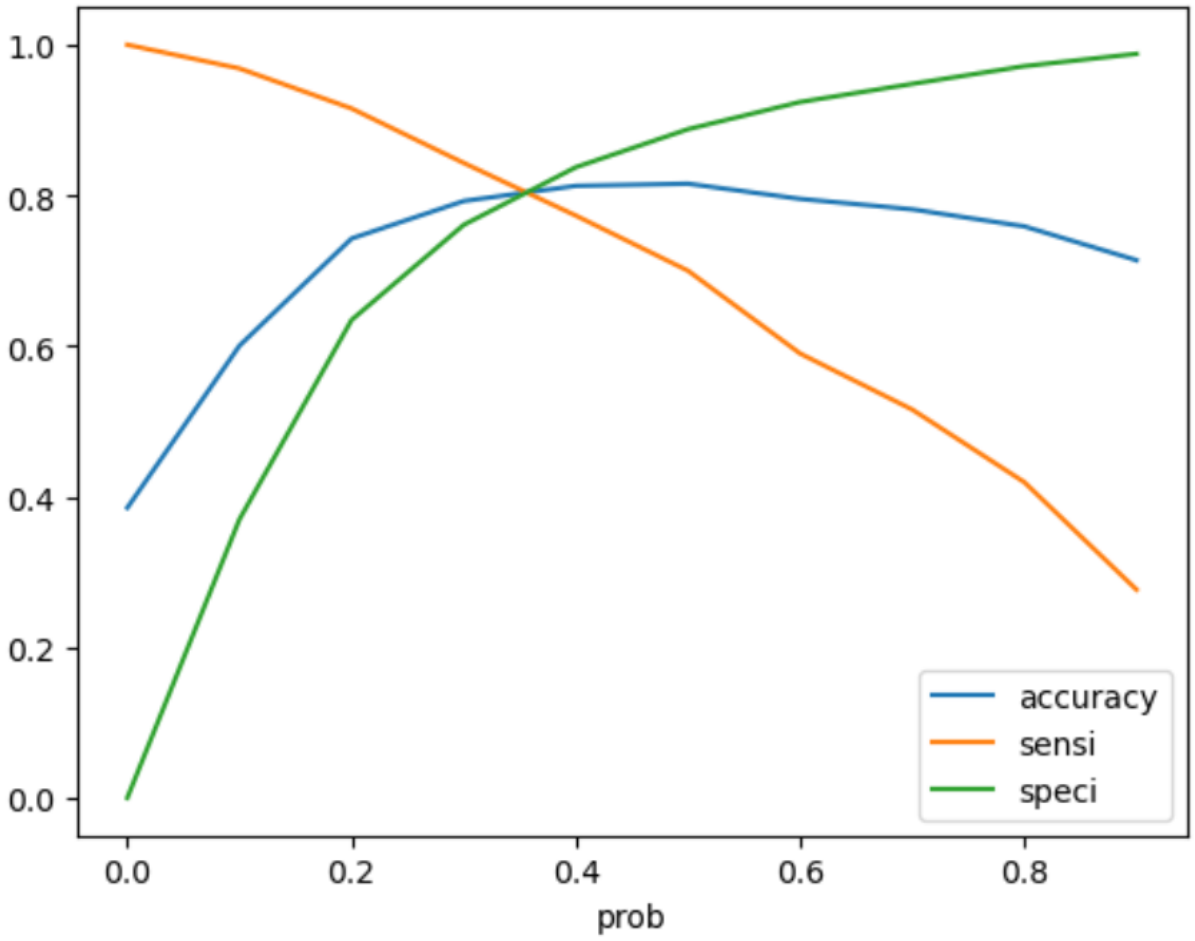
- Individuals employed as working professionals and opting for the course exhibit a high likelihood of joining.
- Although unemployed leads constitute the largest group in numbers, their conversion rate hovers around 30-35%.

Results:

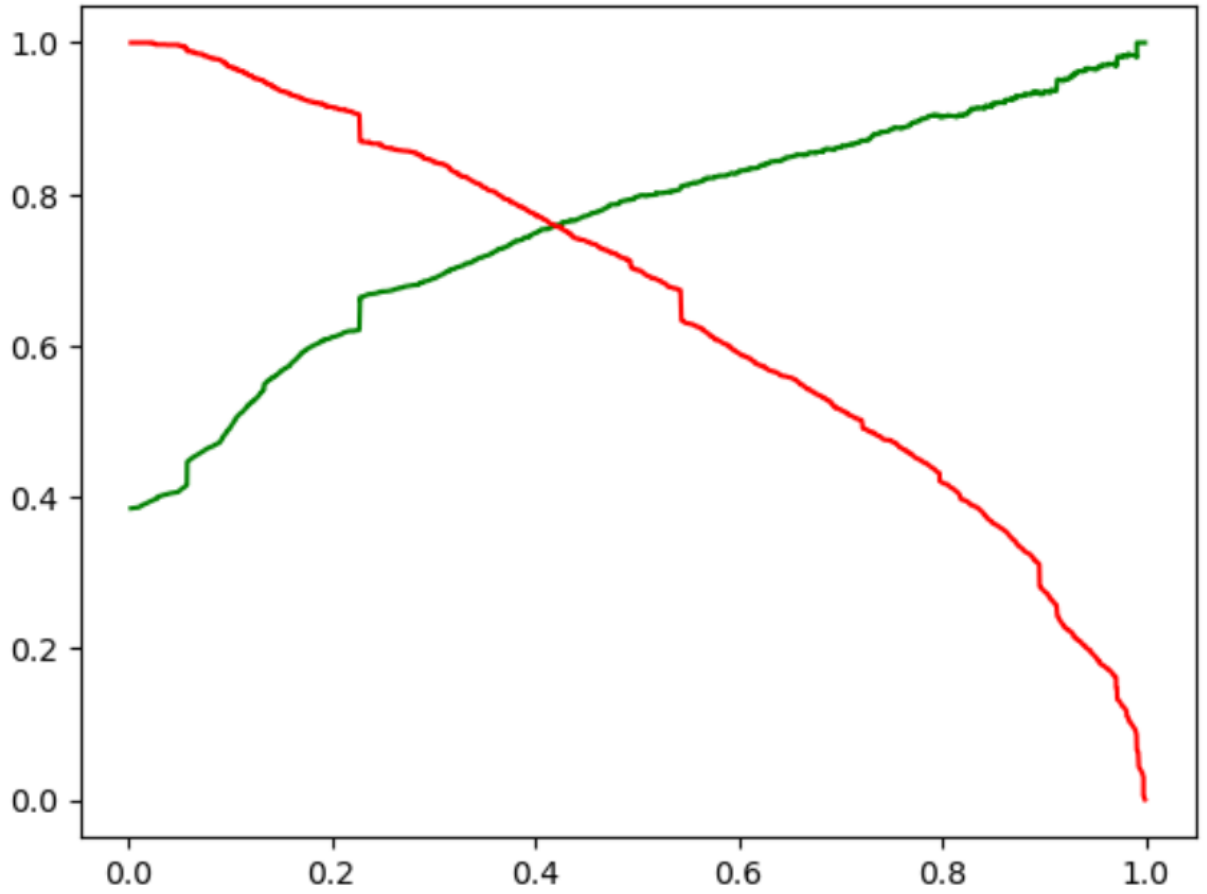
- ROC Curve:



Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.



From the curve above, 0.34 is the optimum point to take it as a cutoff probability.



The above graph shows the trade-off between the Precision and Recall .

Comparing the values obtained for Train & Test:

Train Data:

- Accuracy : 81.0 %
- Sensitivity : 81.7 %
- Specificity : 80.6 %

Test Data:

- Accuracy : 80.4 %
- Sensitivity : 80.4 %
- Specificity : 80.5 %

Thus, we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 80%.

- The customers which should be contacted are the customers whose "Lead Score" is equal to or greater than 85. They can be termed as 'Hot Leads'.
- So there are 368 leads which can be contacted and have a high chance of getting converted.

Important Features from our final model:

```
res.params.sort_values(ascending=False)
```

```
Lead Source_Welingak Website                                  5.811465
Lead Source_Reference                                         3.316598
What is your current occupation_Working Professional          2.608292
Last Activity_Other_Activity                                  2.175096
Last Activity_SMS Sent                                        1.294180
Total Time Spent on Website                                   1.095412
Lead Source_Olark Chat                                        1.081908
const                                                        -0.037565
Last Notable Activity_Modified                               -0.900449
Last Activity_Olark Chat Conversation                        -0.961276
Lead Origin_Landing Page Submission                          -1.193957
Specialization_Others                                        -1.202474
Do Not Email                                                 -1.521825
dtype: float64
```

Summary/Recommendations:

- Prioritize making calls to leads from "Welingak Websites" and "Reference" sources, as they exhibit a higher conversion likelihood.
- Target calls towards "working professionals" among leads, as they demonstrate a higher likelihood of conversion.
- Focus on leads who spend more time on the websites, as they are more likely to convert.
- Prioritize calls to leads from the "Olark Chat" source, given their higher conversion potential.
- Target leads whose last activity was "SMS Sent," as they are more likely to convert.
- Avoid making calls to leads with the last activity as "Olark Chat Conversation," as they show lower conversion likelihood.
- Refrain from calling leads from the "Landing Page Submission" source, as their conversion likelihood is lower.
- Avoid contacting leads with Specialization marked as "Others," as they are less likely to convert.
- Do not make calls to leads who have selected "Do not Email" as "yes," as their conversion likelihood is lower.