LEAD SCORING CASE STUDY

Contents :

- Problem Statement
- Business Objective
- Problem Approach
- EDA
- Correlations
- Model Evaluation
- Observations
- Conclusion

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.
- 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, it means that, 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 are successfully identifies this set of leads, the leads 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.

Business Objective:

- The CEO wants to achieve a lead conversion rate of 80%
- They want the model to be able to handle future constraints as well the Peak time action required, how to utilize full man power after achieving target what should be the approaches.
- Lead X wants us to build a model to give every lead as a lead score between 0 – 100, so that they can identify the HOT LEADS and increase their conversion rate as well.

Problem Approach :

- Importing the data and Inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test – Train split
- Feature scaling
- Correlation
- Model building
- Model evaluation
- Making predictions on test set

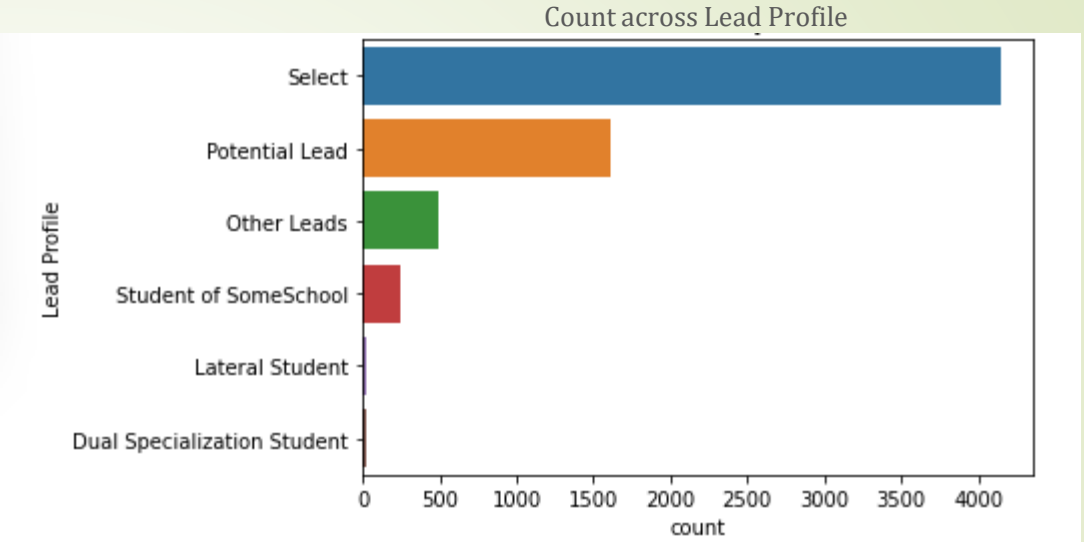
EDA

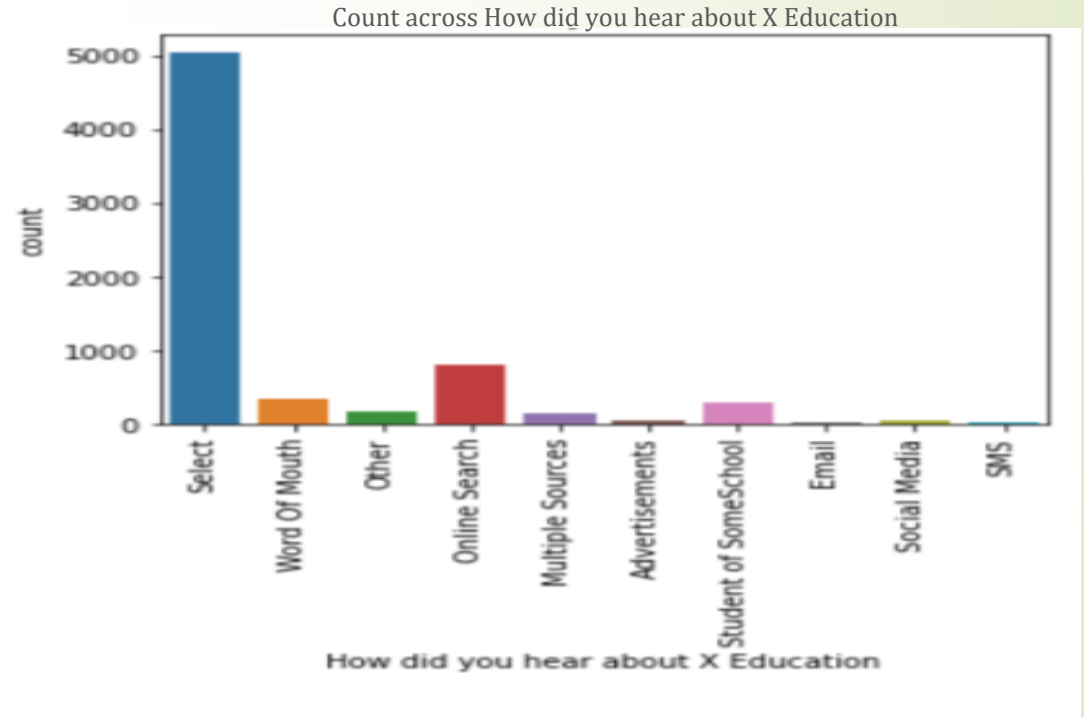
DATA CLEANING :

- Lead Profile: A lead level assigned to each customer based on their profile.

- How did you hear about X Education: The source from which the customer heard about X Education.

A large portion of values are **"SELECT"**, which is equivalent to a null. There is no single lead profile that stands out as a clear value for null imputation.



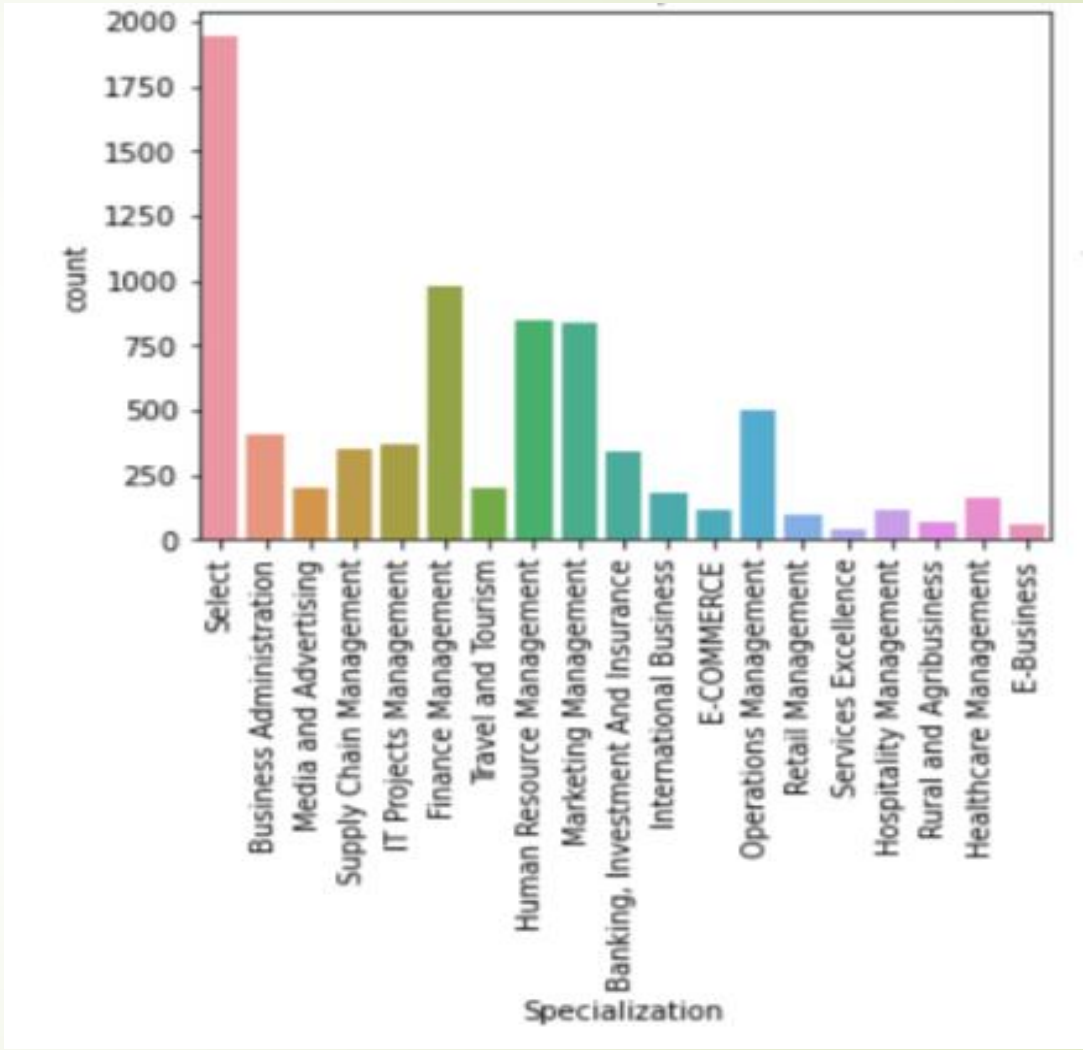


SPECIALIZATION :

The industry domain in which the customer worked before. Includes the level 'Select Specialization' which means the customer had not selected this option while filling the form.

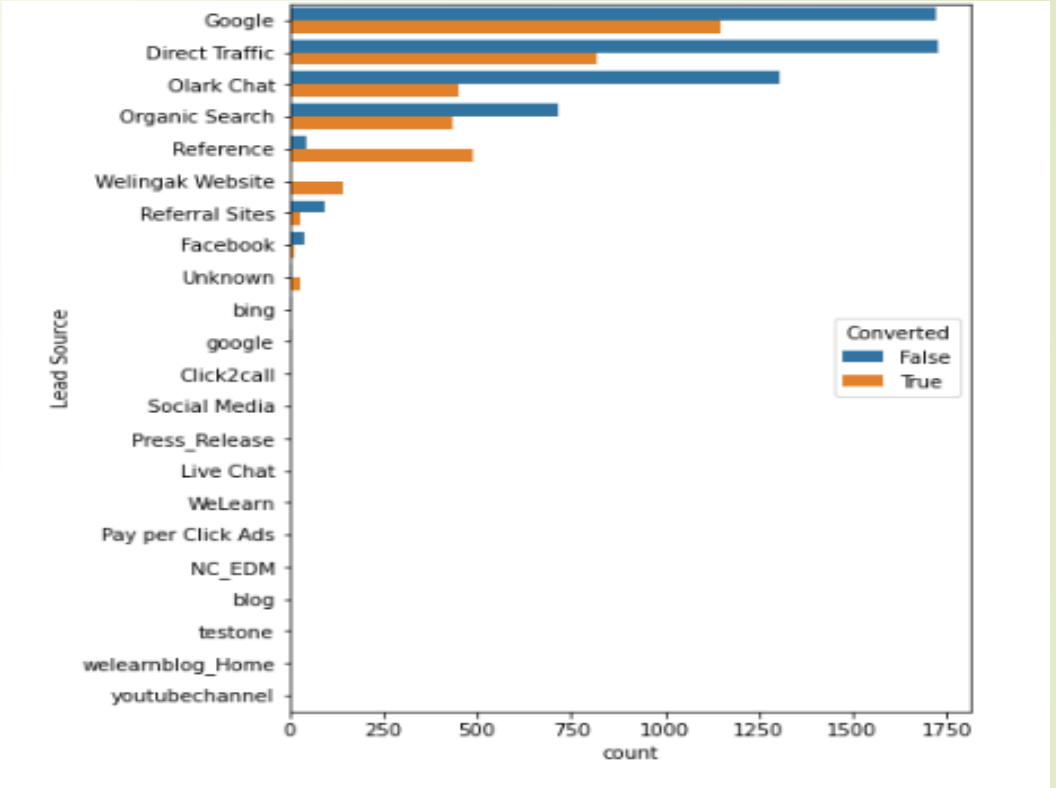
Leads from HR, Finance and marketing management specialization are high probability to convert.

Count across Specialization

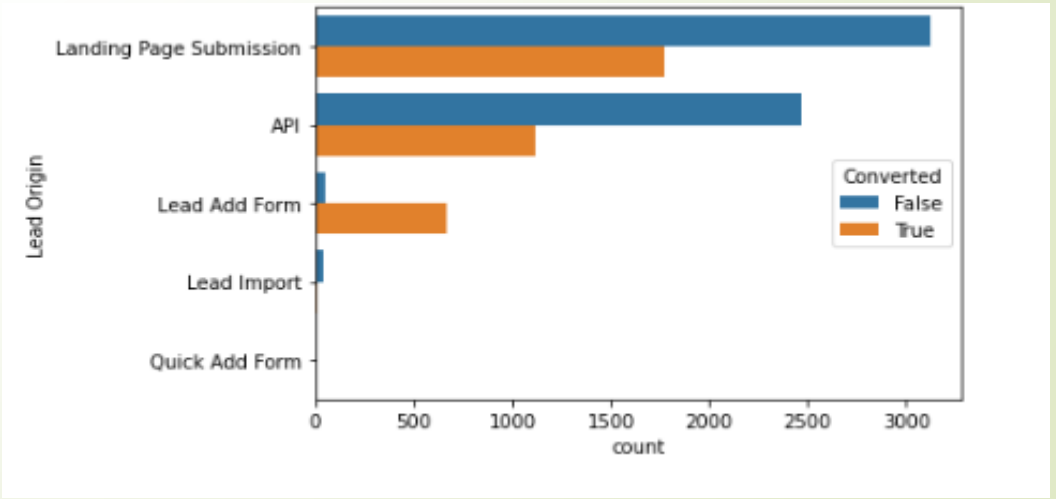


LEAD SOURCE AND LEAD ORIGIN :

- Google and Direct Traffic are the source for a similar number of leads, the conversion rate appears higher among leads that come from Google.



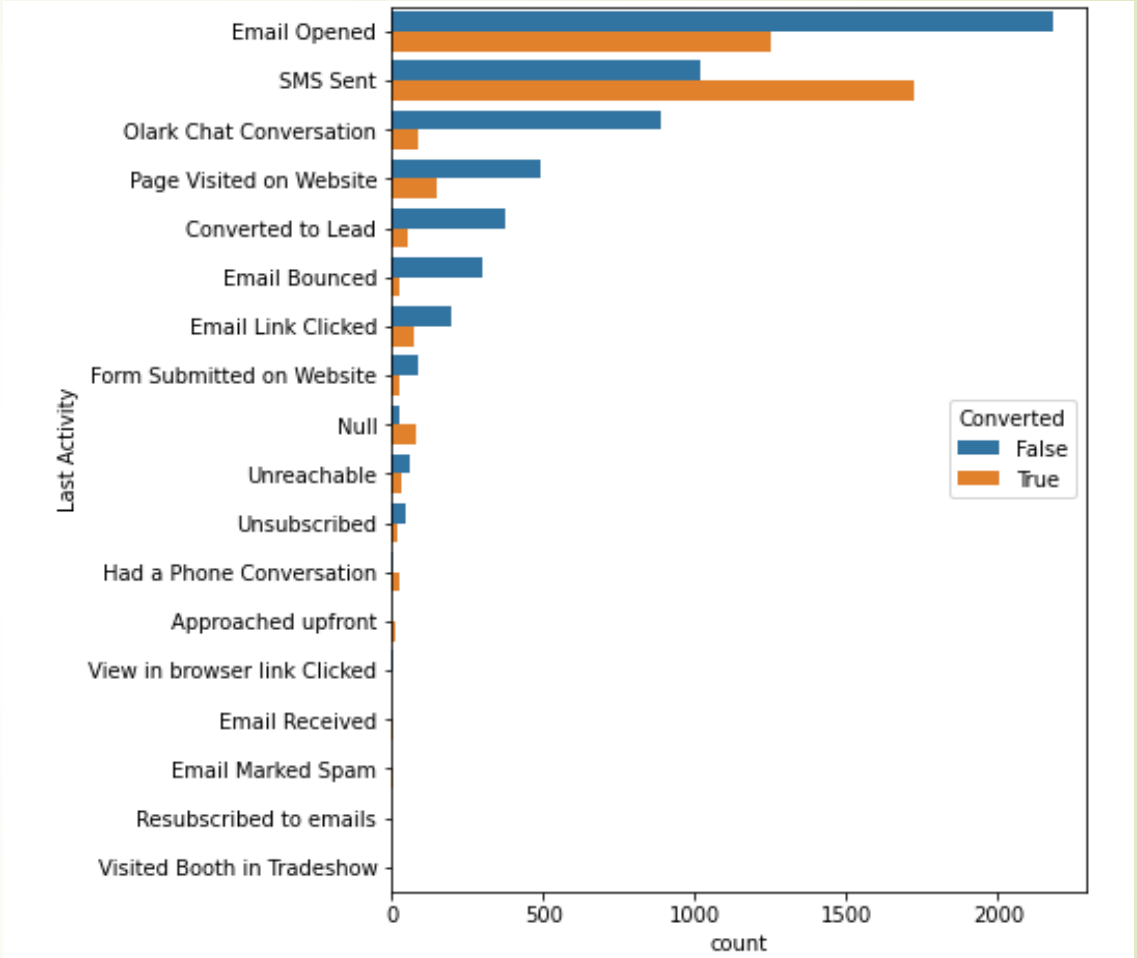
- The lead conversion rate appears slightly higher among leads originating from "Landing Page Submission" than from "API".



LAST LEAD ACTIVITY:

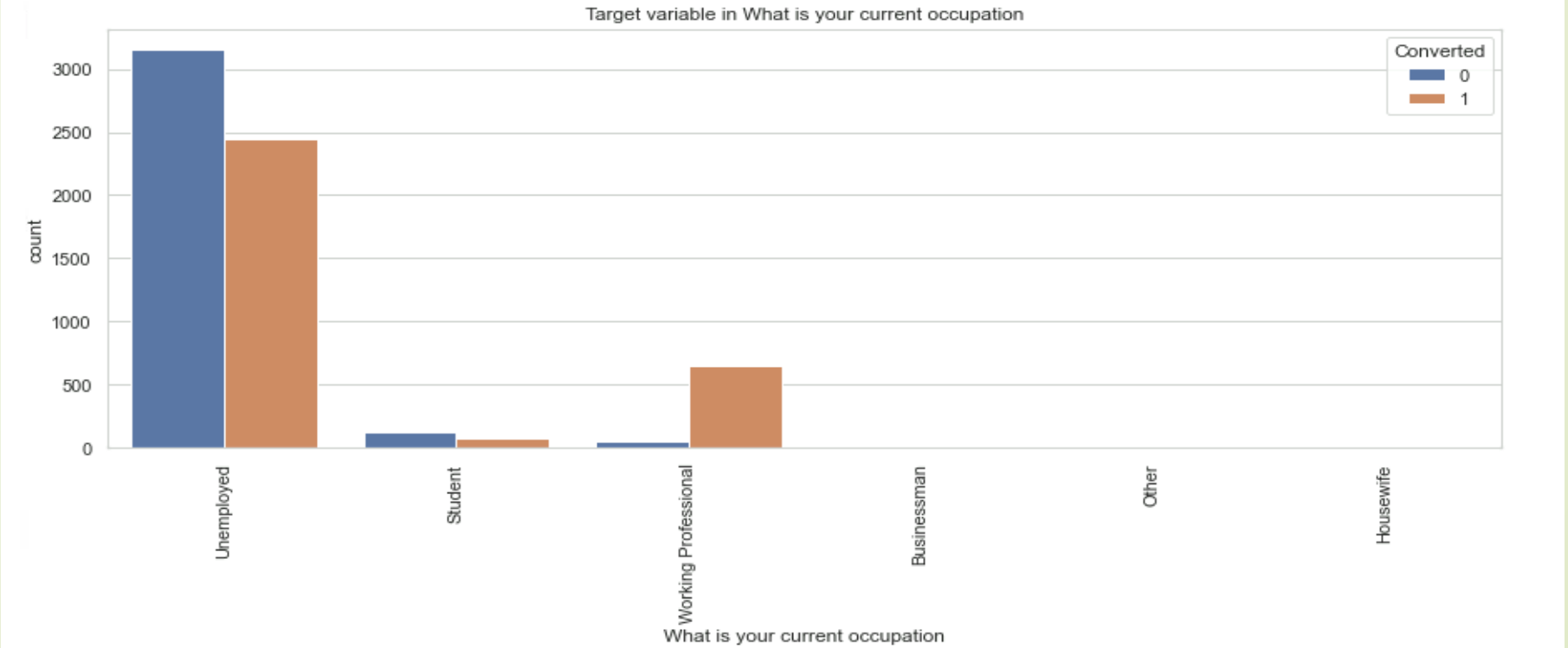
The Last Activity with he highest conversion rate is by far SMS Sent. Same as opening email will also benefit.

Target variable in the Last Activity



LAST WHAT IS YOUR OCCUPATION :

Leads which are unemployed are more interested to join the course than others.



CORRELATION :

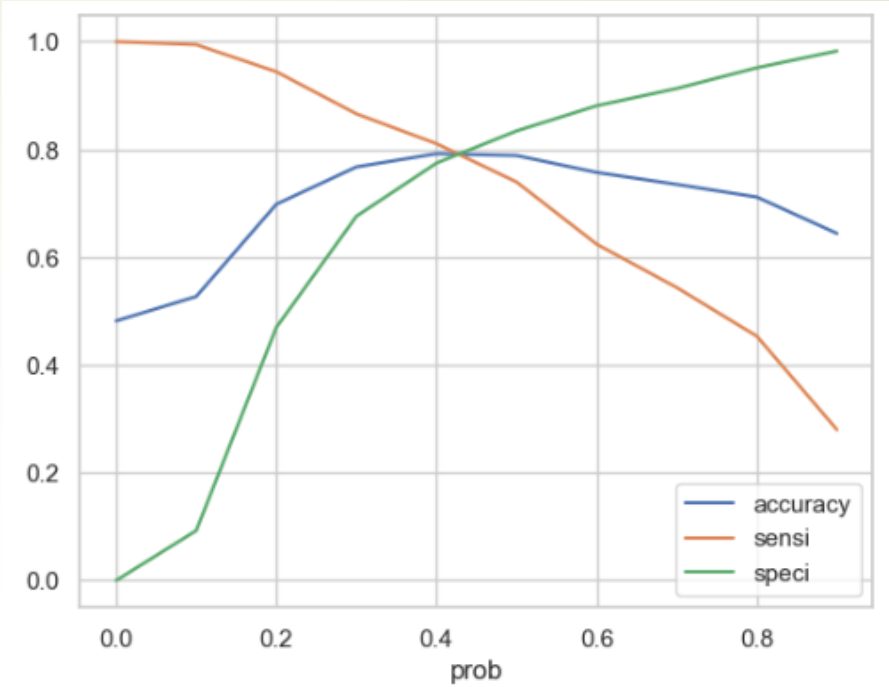
There is no correlation between the variables.

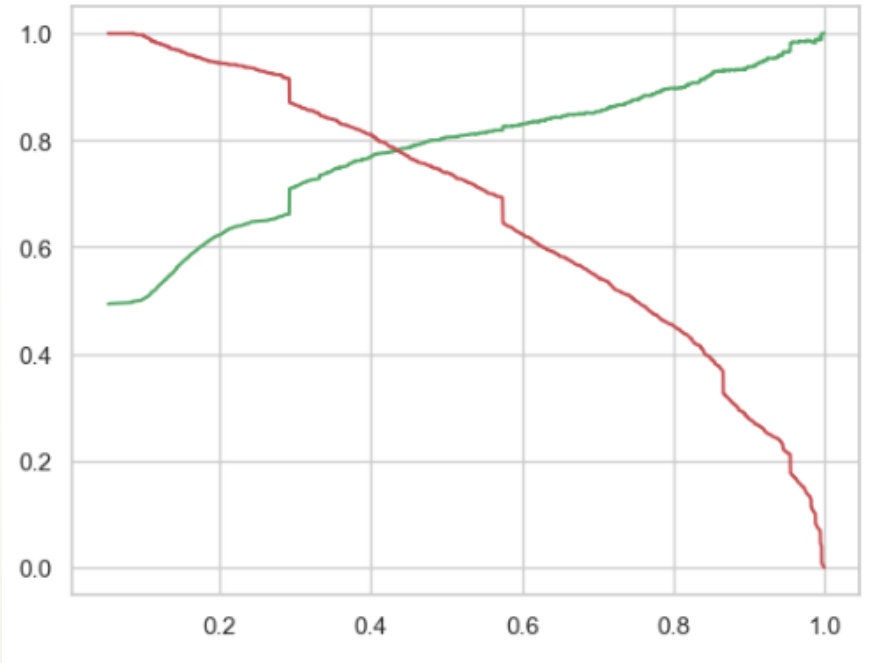
MODEL EVALUATION :

ROC Curve :

0.42 is the trade off between Recall and Precision

Thus we can safely choose to consider any prospect lead with the conversion **probability higher than 42% to be a HOT LEAD**





OBSERVATIONS :

- **Train Data**
  - Accuracy – 79.08%
  - Sensitivity – 79.33%
  - Specificity – 78.84%
- **Test Data**
  - Accuracy – 78.95%
  - Sensitivity – 78.40%
  - Specificity – 77.7%

Final Features List

- Lead Source_Olark Chat
- Specialization_Others
- Lead Source_Welingak Website
- Lead Origin_Lead Add Form
- Total Time Spent on Website
- What is your company occupation_Working Proffesionals
- Do Not Email

CONCLUSION :

- We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore we can intervene that we need to focus more on the leads originated from API and Landing page submission.
- We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.