

Lead Scoring Case Study

Machine Learning Using Logistic Regression

Prepared and Submitted by:

Dharmvir Singh

Sweta Sinha

Snehal R

Lead Scoring Case Study

Content

1. Problem statement
2. Problem approach
3. EDA
4. Correlations
5. Model Evaluation
6. Observations
7. Conclusion

Lead Scoring Case Study

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. They have process of form filling on their website after which the company that individual as a lead.
- 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, this means 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

Lead Scoring Case Study

Business Objective

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

Lead Scoring Case Study

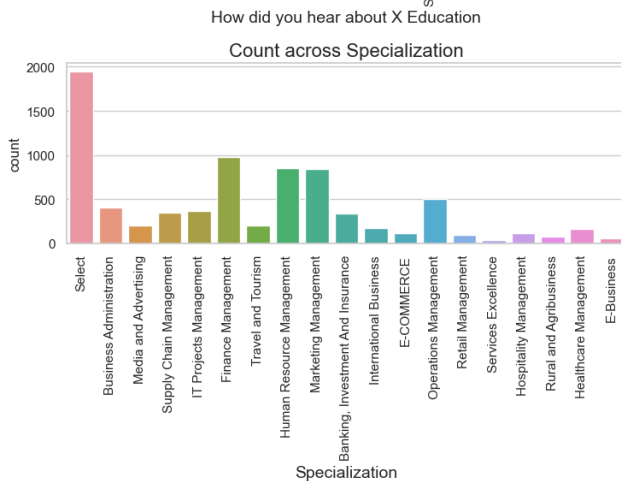
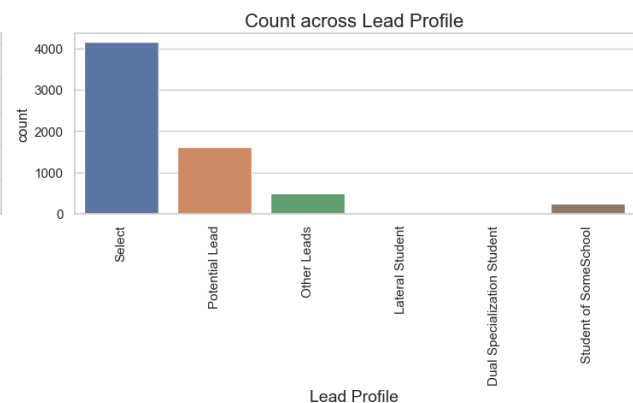
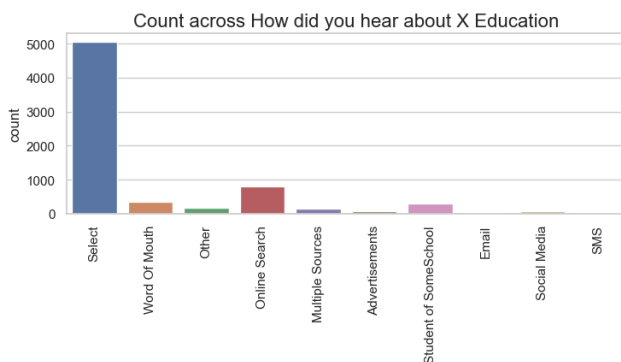
Problem Approach

- Importing the data and inspecting the data frame
- Data preparation
- EDA
- Dummy variable creation
- Test-Train split
- Feature scaling
- Correlations
- Model Building (RFE R-squared VIF and p- values)
- Model Evaluation
- Making predictions on test set

Lead Scoring Case Study

EDA – Data Cleaning

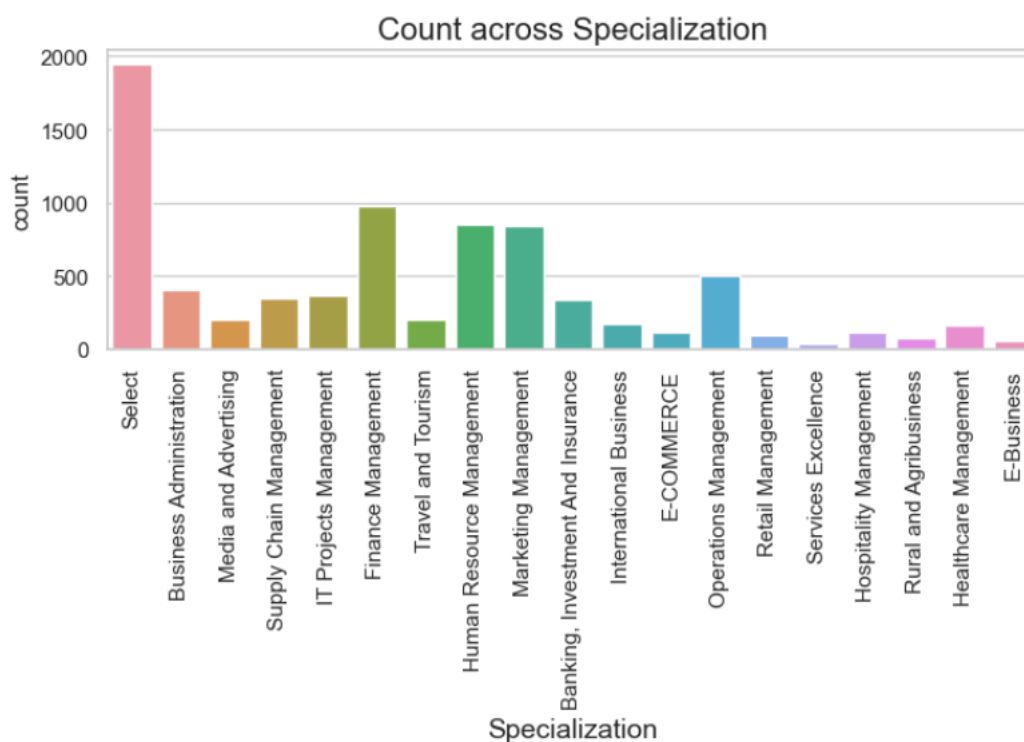
- There are a few columns in which there is a level called 'Select' which is taking care



Lead Scoring Case Study

Specialisation

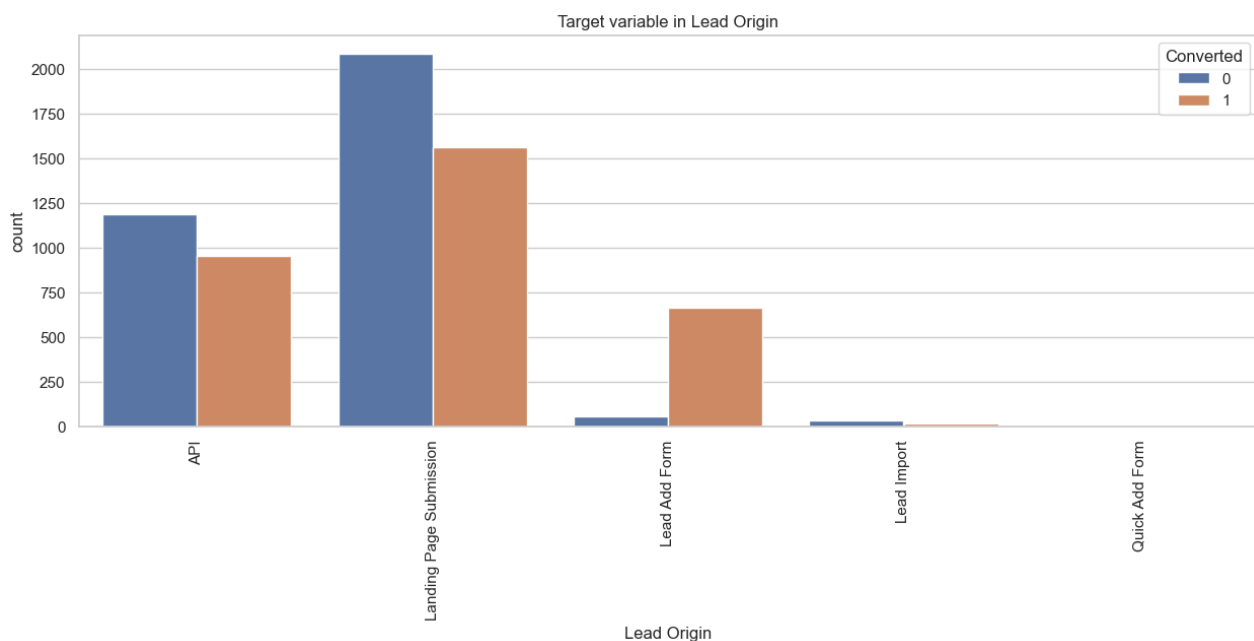
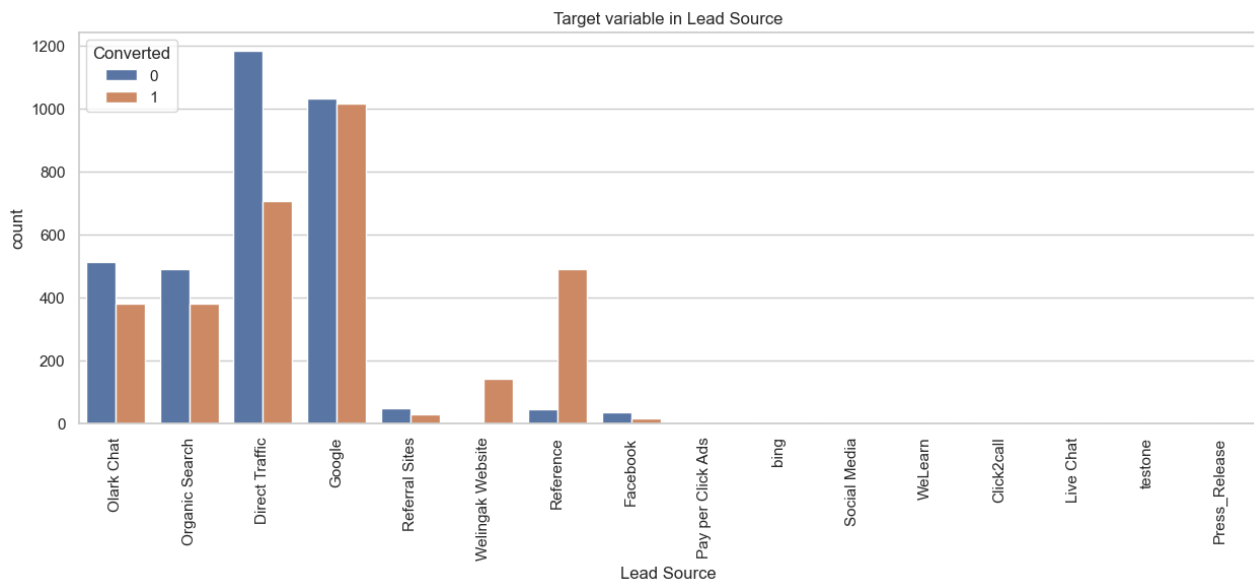
- ▶ Leads from HR, Finance & Marketing management specialisation are high probability to convert.



Lead Scoring Case Study

Lead Source & Lead origin

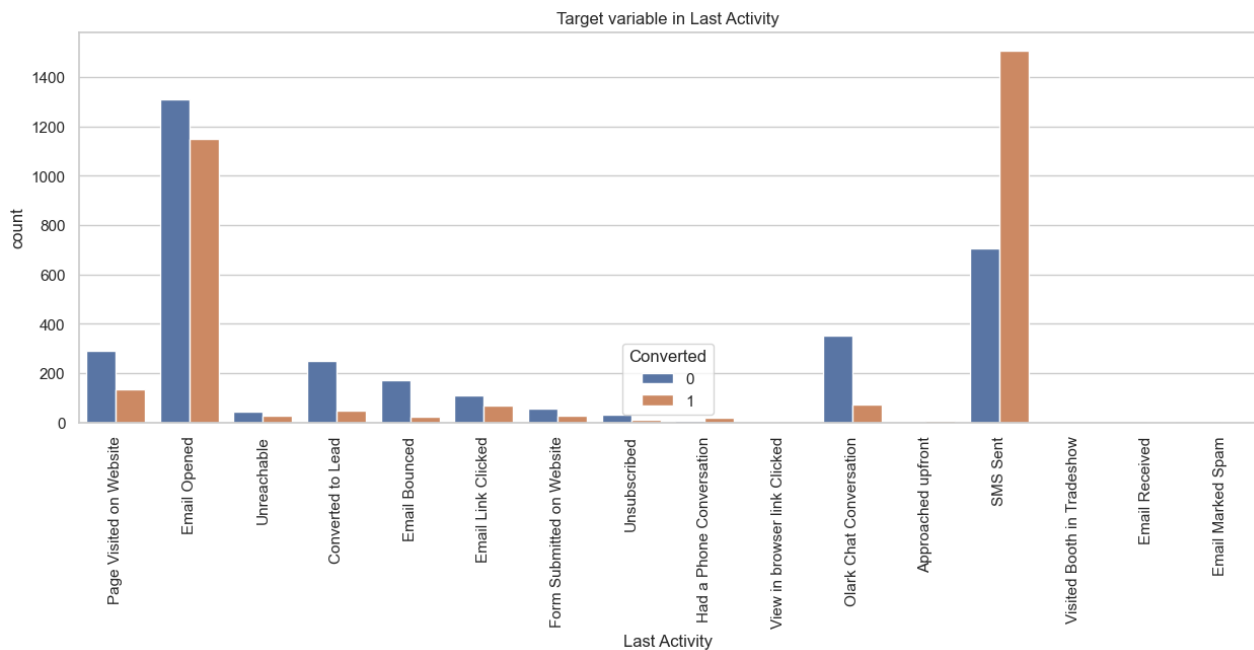
- ▶ In lead source the leads through google & direct traffic high probability to convert
- Whereas in Lead origin most number of leads are landing on submission



Lead Scoring Case Study

Last lead Activity

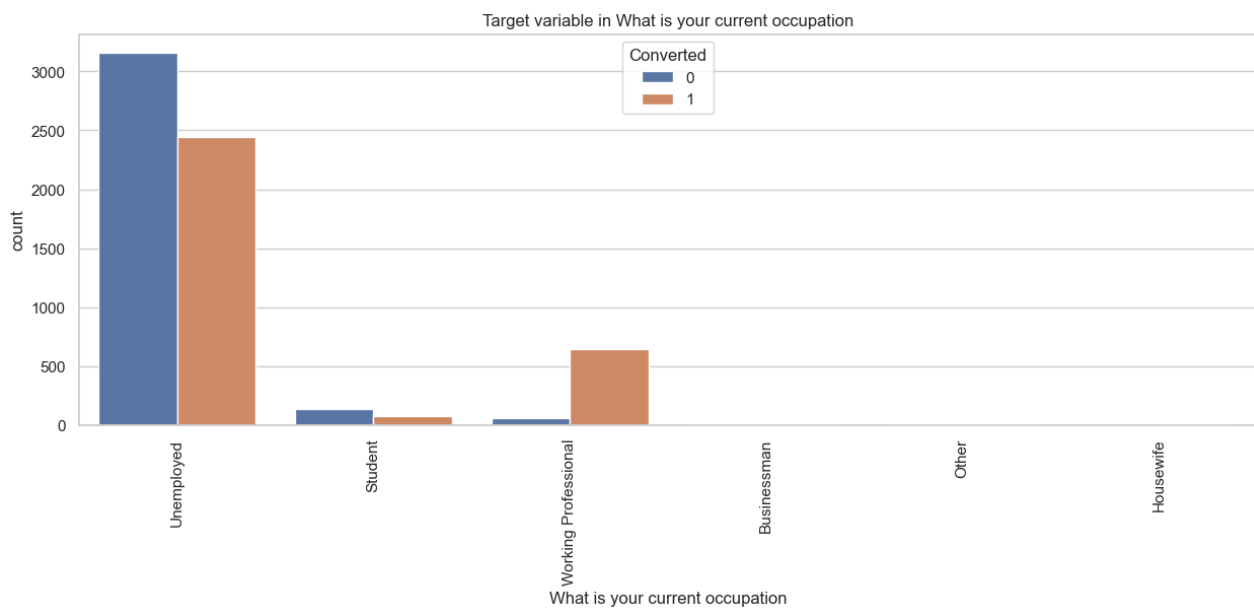
Leads which are opening email have high probability to convert, Same as Sending SMS will also benefit.



Lead Scoring Case Study

Last What is Your Occupation

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



Lead Scoring Case Study

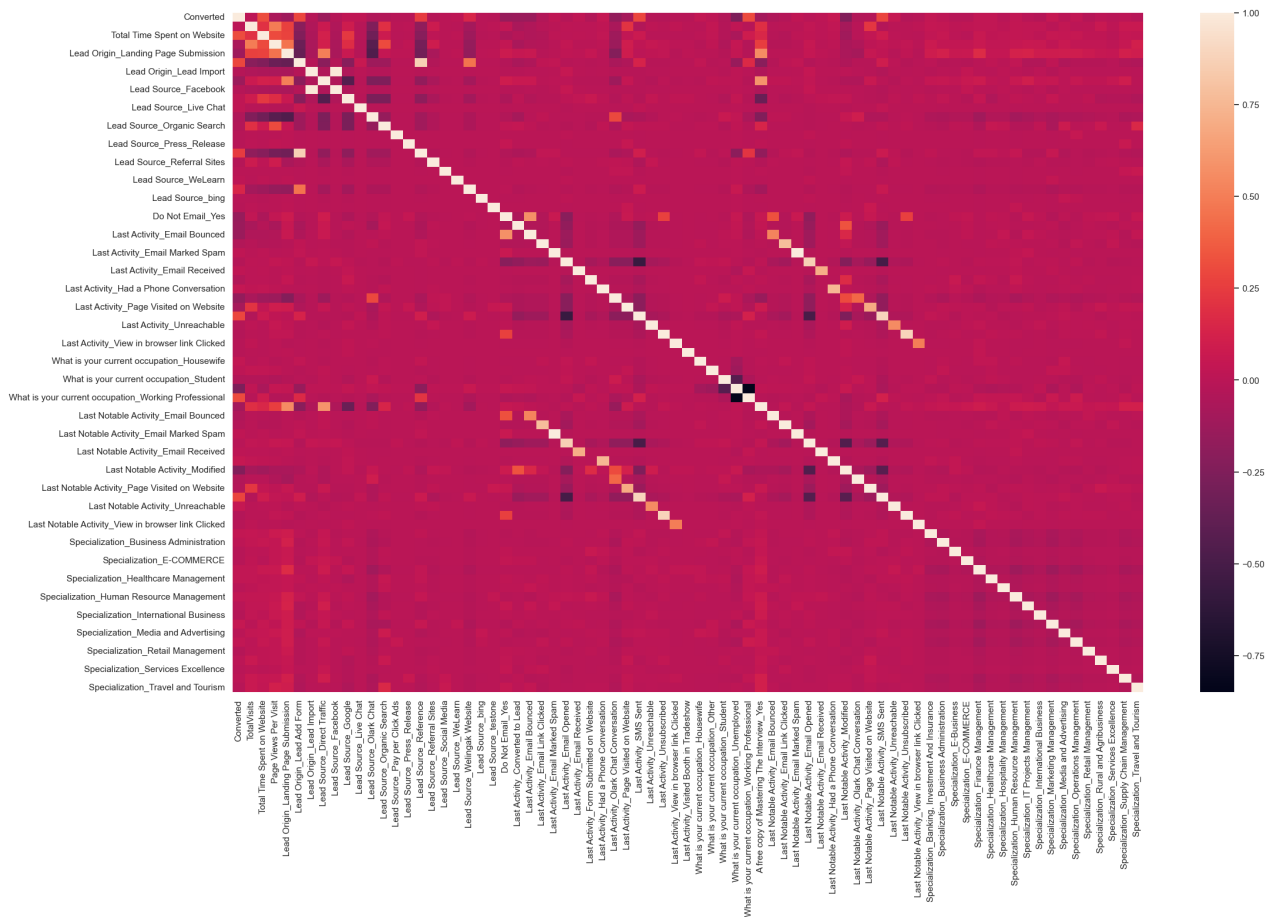
Correlation

There is no correlation between the variables



Lead Scoring Case Study

Looking at the correlations - Heatmap



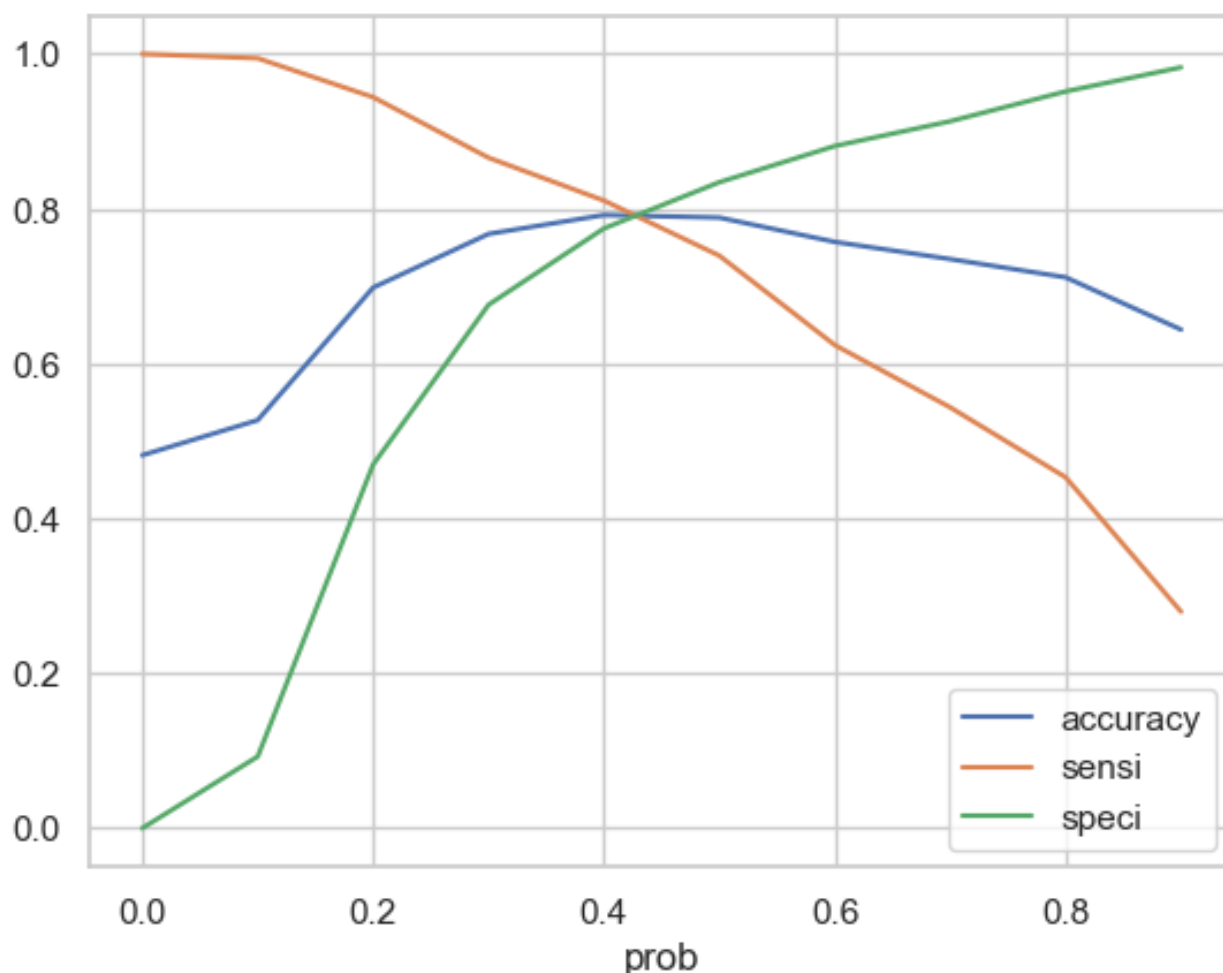
Lead Scoring Case Study

Model Evaluation

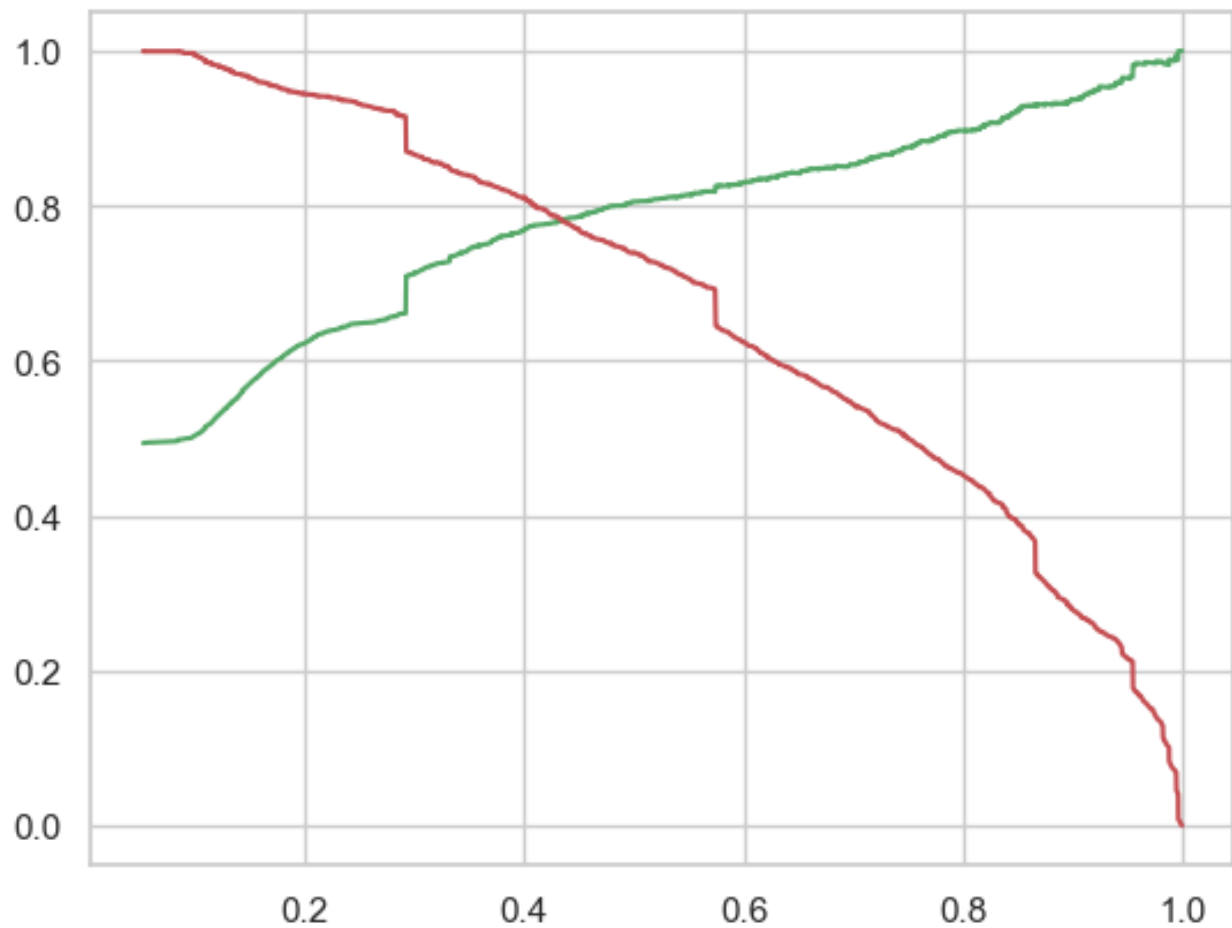
ROC curve

0.42 is the tradeoff between Precision and Recall -

Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 42 % to be a hot Lead



Lead Scoring Case Study



Lead Scoring Case Study

Observations

Train Data: Accuracy : 80% Sensitivity : 77% Specificity : 80%

Test Data:

Accuracy : 80% Sensitivity : 77% Specificity : 80%

Final Features list:

- Lead Source_Olark Chat
- Specialization_Others
- Lead Origin_Lead Add Form
- Lead Source_Welingak Website
- Total Time Spent on Website
- Lead Origin_Landing Page Submission
- What is your current occupation_Working Professionals
- Do Not Email

Lead Scoring Case Study

Conclusion

- Observing the data, we notice that the conversion rate is around 30-35%, which is close to the average, for API and Landing Page submissions. However, it is significantly lower for Lead Add forms and Lead imports. This suggests that we should prioritise our efforts on improving the conversion rates for leads generated through API and Landing Page submissions.
- We observe that the highest number of leads originate from Google and direct traffic sources. The highest conversion ratio is achieved through references and the Welingak website.
- Leads who invest more time on the website have a greater likelihood of conversion.
- The most frequently observed last activity is "email opened," while the activity with the highest conversion rate is "SMS sent." A majority of leads are categorised as unemployed, and the highest

Lead Scoring Case Study

conversion rate is achieved with working professionals.

*****End of Case Study Summary*****