



LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

SUBMITTED BY :

1. Aabisha Nazeer
2. Mohammad Wasim
3. Danvi nanya

CONTENTS

- ☐ Problem statement
- ☐ Problem approach
- ☐ EDA
- ☐ Correlations
- ☐ Model Evaluation
- ☐ Observations
- ☐ Conclusion

PROBLEM STATEMENT

- An education company, X Education, specializes in selling online courses to industry professionals. Prospective students frequently visit the website, where they engage in a form-filling process, transforming them into leads for the company.
- Once leads are obtained, the sales team initiates communication through calls and emails. Despite their efforts, only a portion of the leads undergo conversion. The typical lead conversion rate at X Education stands at approximately 30%. This implies that out of 100 acquired leads, only around 30 are successfully converted.
- To enhance efficiency, the company aims to identify the most promising leads, referred to as Hot Leads. The successful identification of this subset is expected to increase the lead conversion rate, as the sales team can concentrate efforts on communicating with potential leads rather than reaching out to everyone indiscriminately.

BUSINESS OBJECTIVE

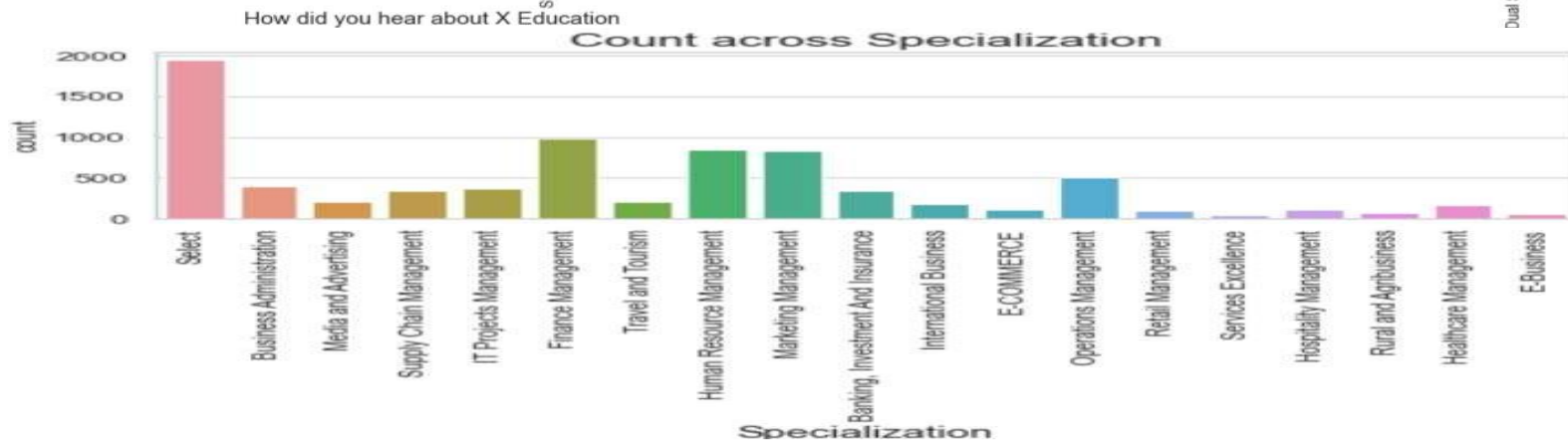
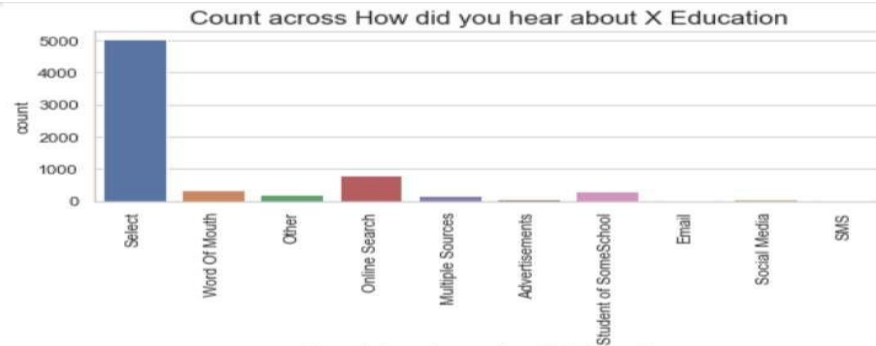
- ☐ Lead X wants us to build a model to give every lead a lead score between 0 -100 .So that they can identify the Hot leads and increase their conversion rate as well.
- ☐ The CEO want to achieve a lead conversion rate of 80%.
- ☐ They want the model to be able to handle future constraints as well like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

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 Rsquared VIF and p-values)
- ❑ Model Evaluation
- ❑ Making predictions on test set

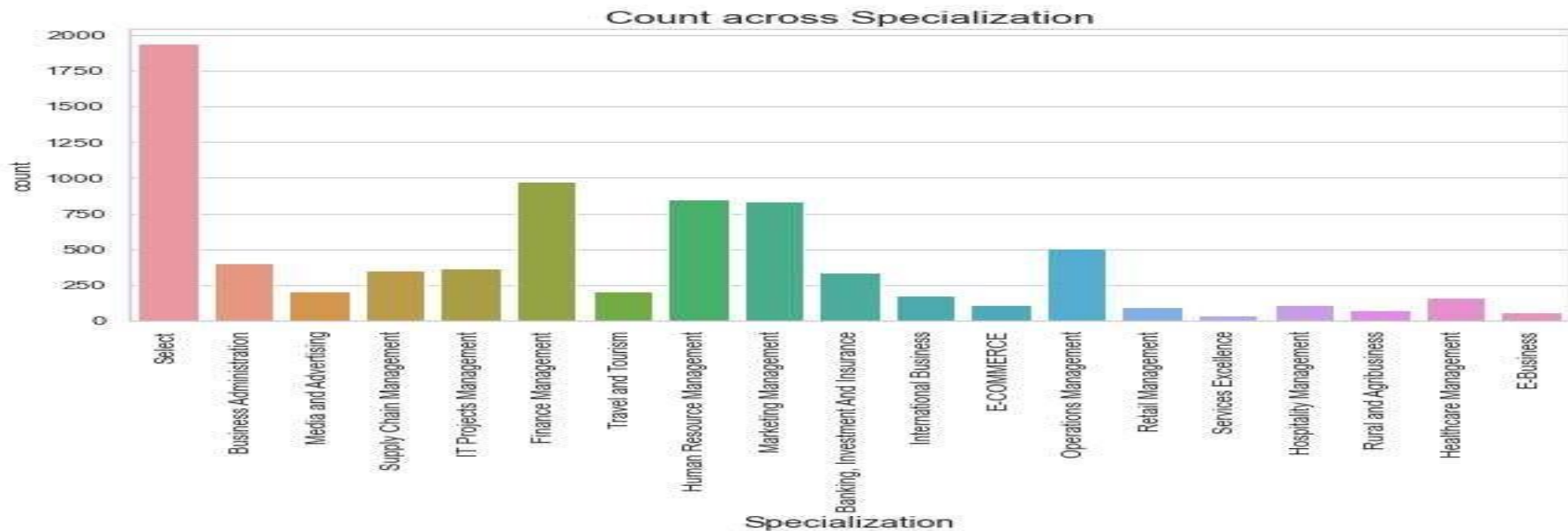
EDA – Data Cleaning

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



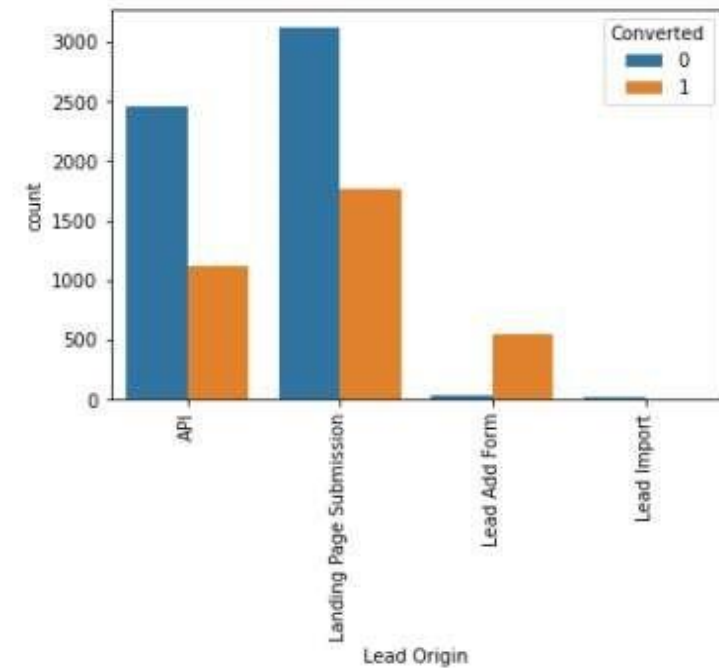
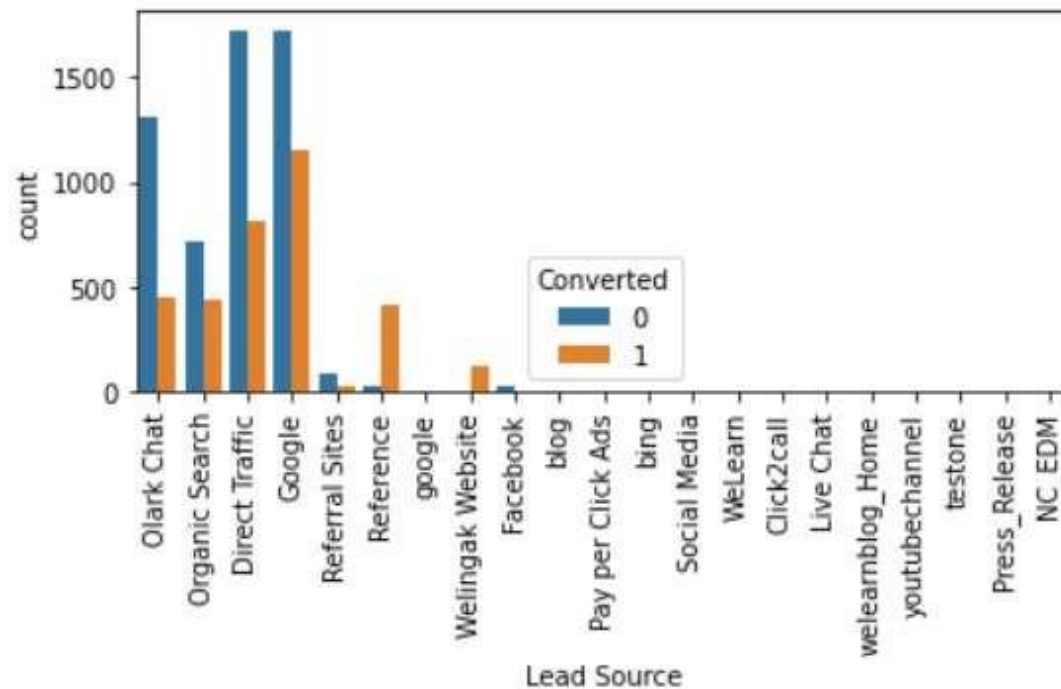
SPECIALIZATION

Leads from HR, Finance & Marketing management specializations are high probability to convert



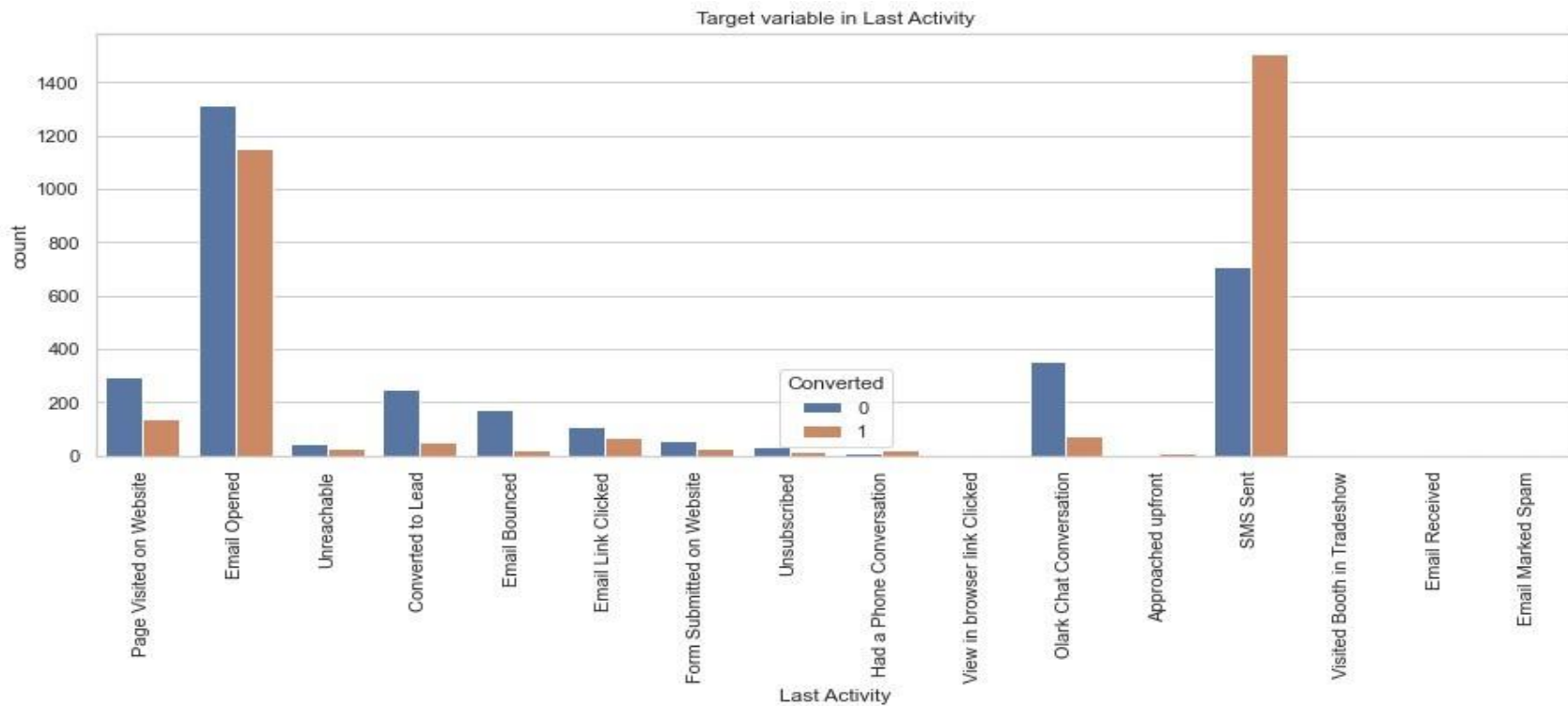
LEAD SOURCE & LEAD ORIGIN

In lead source the leads through google & direct traffic high probability to convert



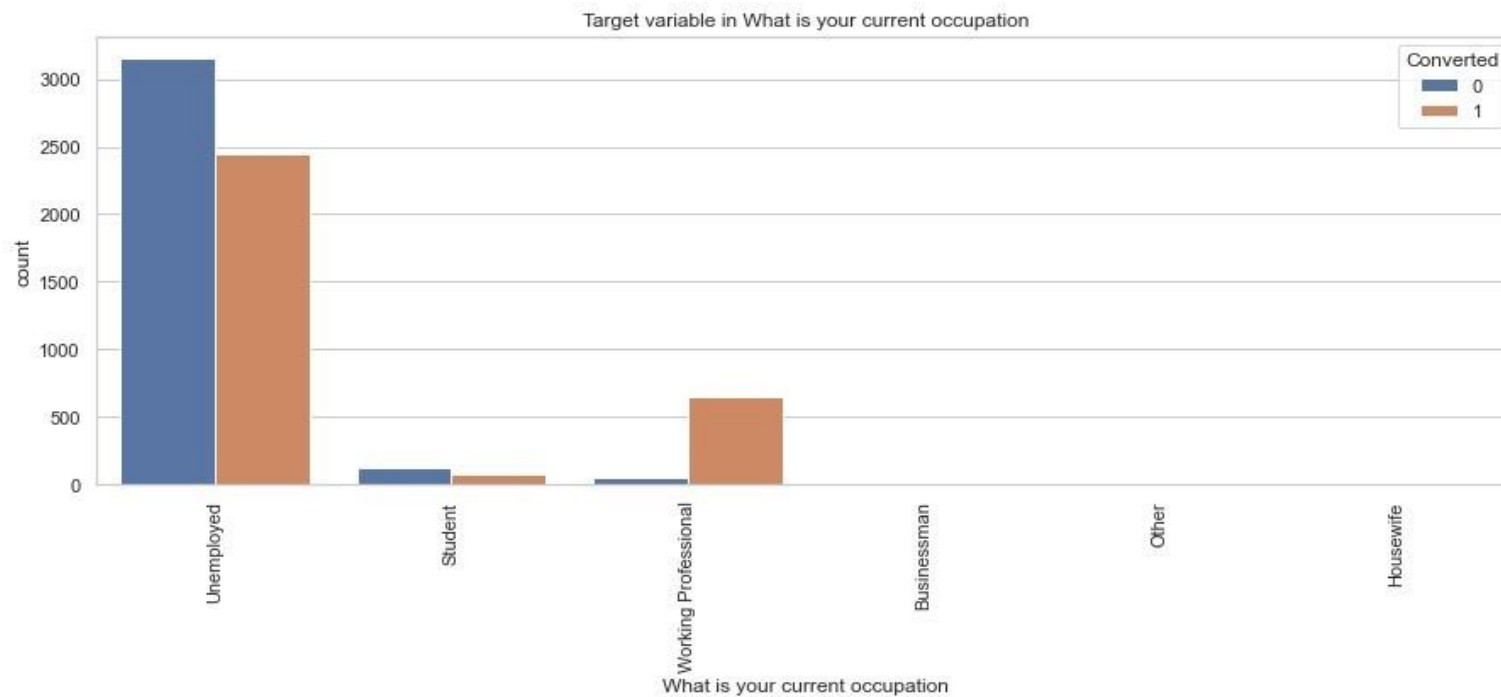
LAST LEAD ACTIVITY

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



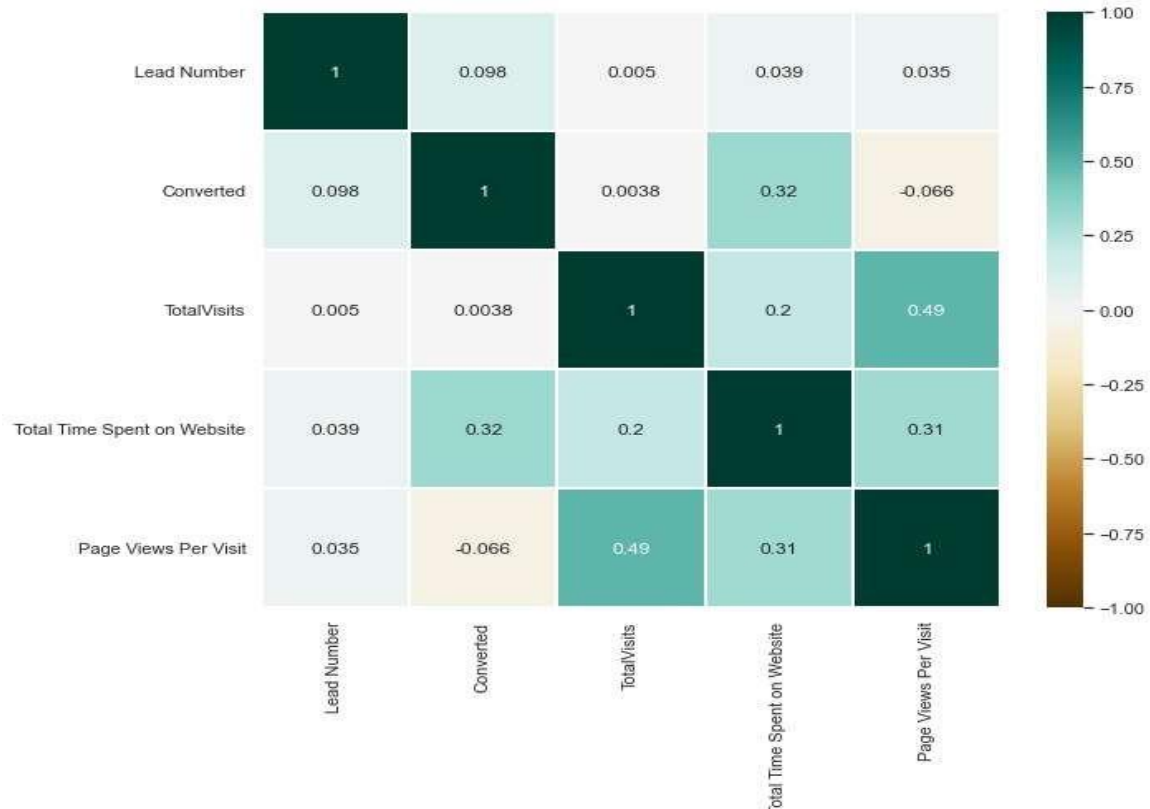
LAST WHAT IS YOUR OCCUPATION

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



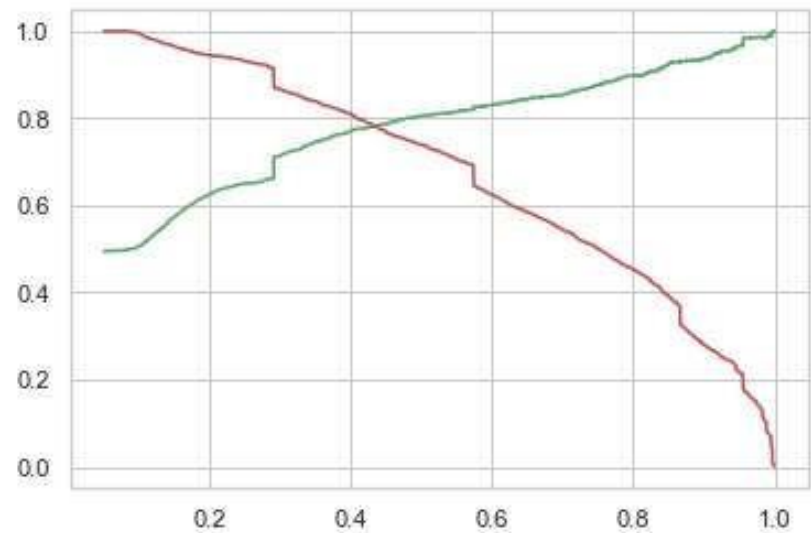
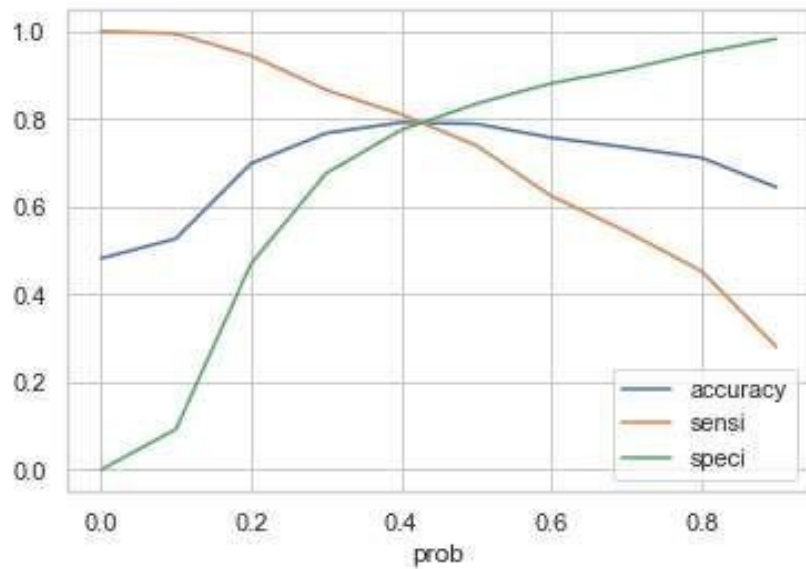
CORRELATION

There is no correlation between the variables



MODEL EVALUATION

ROC curve



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

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.

THANK YOU