

# Lead Scoring Case Study using logistic regression

SUBMITTED BY :

1. Om Gaye
2. Sakshi
3. Shayed

# Contents

- **Problem statement**
- **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. They have process of form filling on their website after which

▶ ☐

▶ ☐

▶ ☐

## Business Objective

- ▶ □ Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can

▶ □

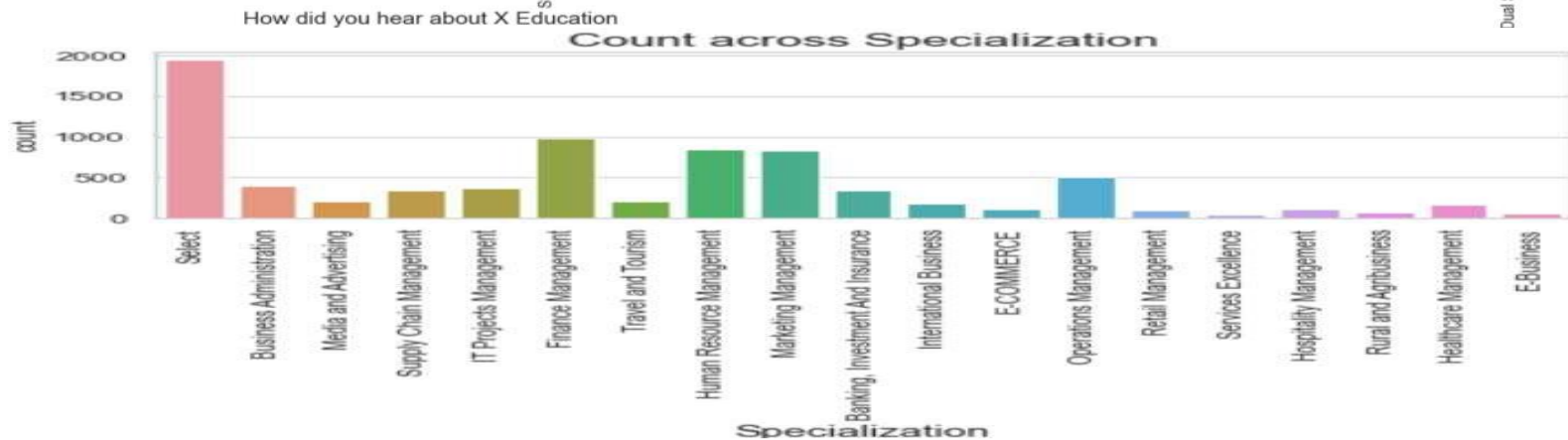
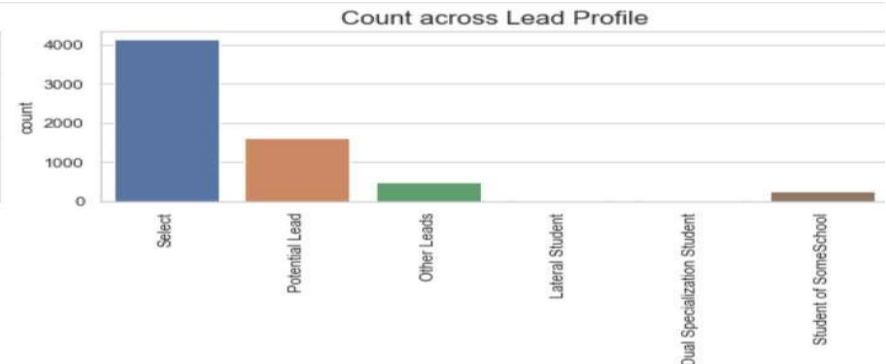
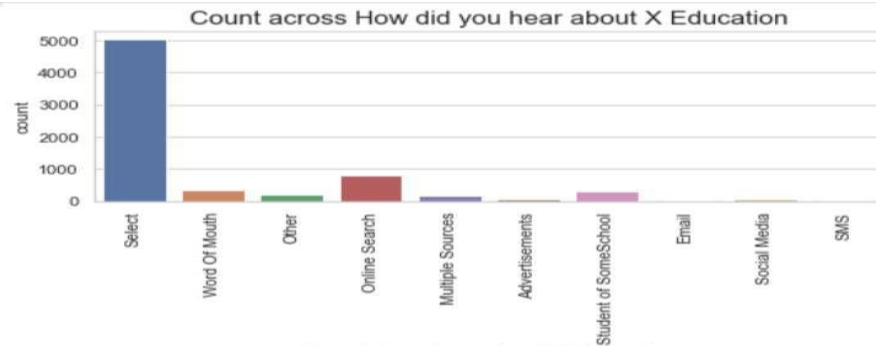
▶ □

# 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

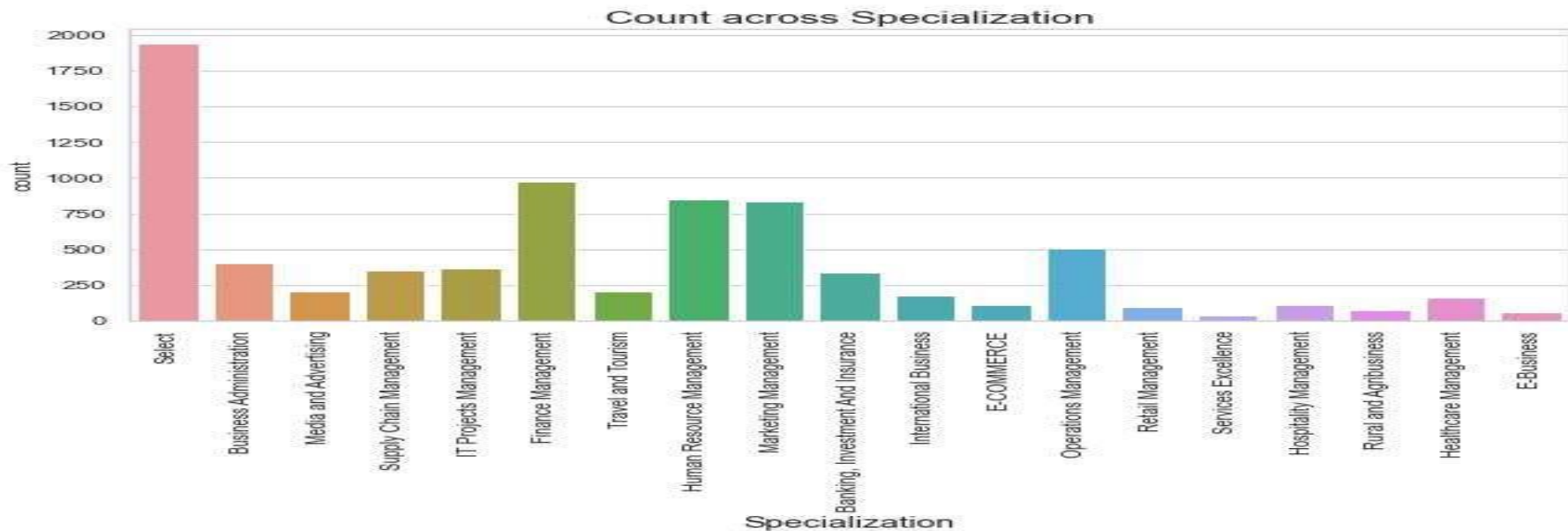
# EDA – Data Cleaning

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



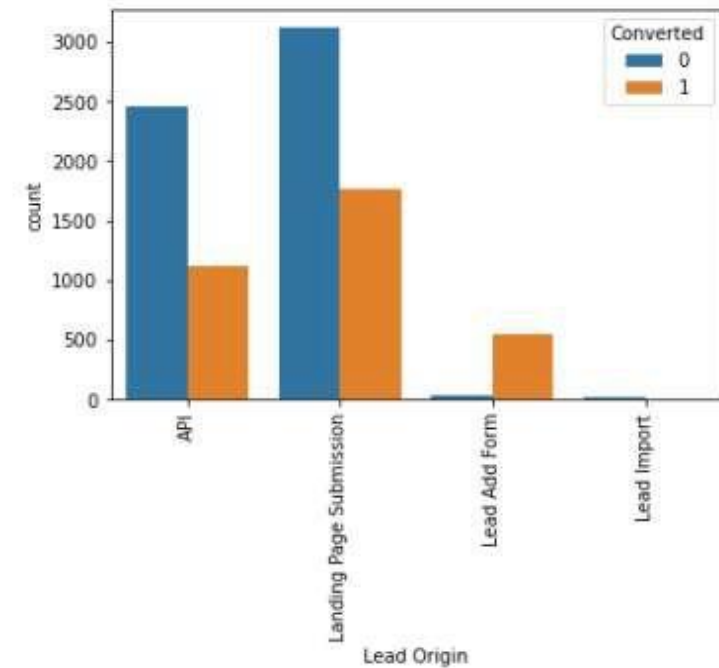
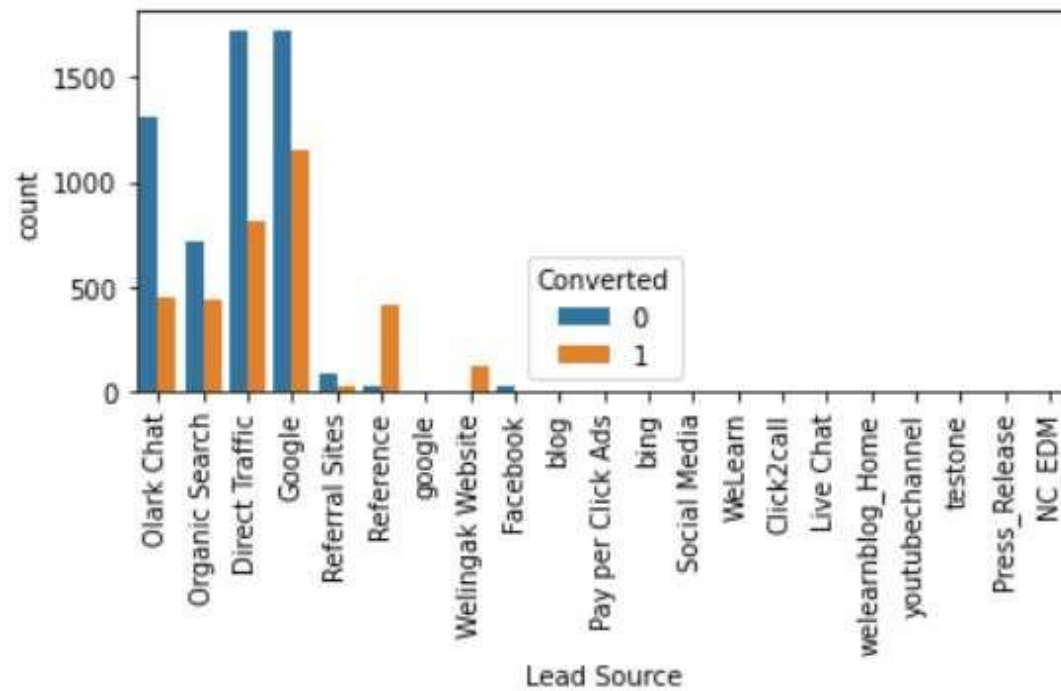
# Specialization

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



# Lead Source & Lead origin

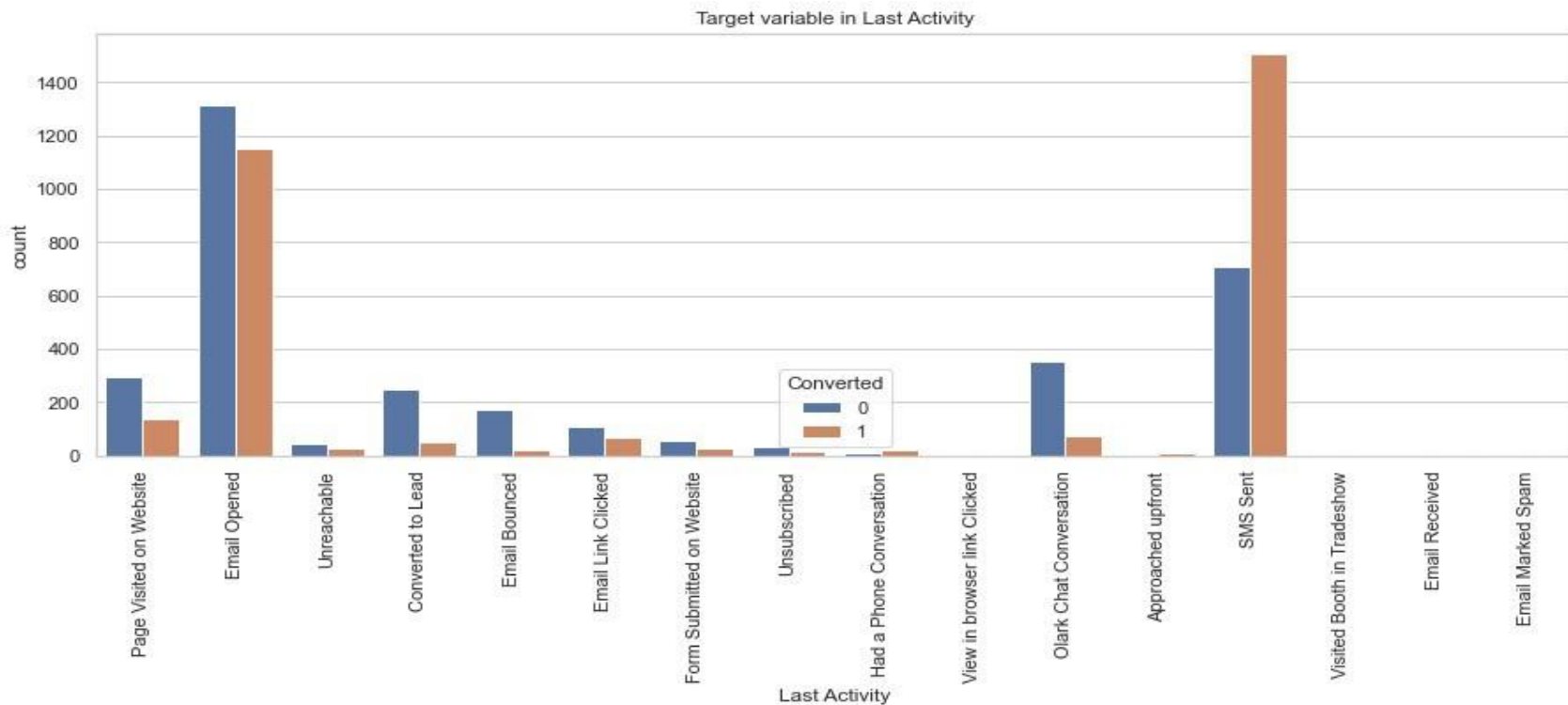
In lead source the leads through google & direct traffic high probability to convert





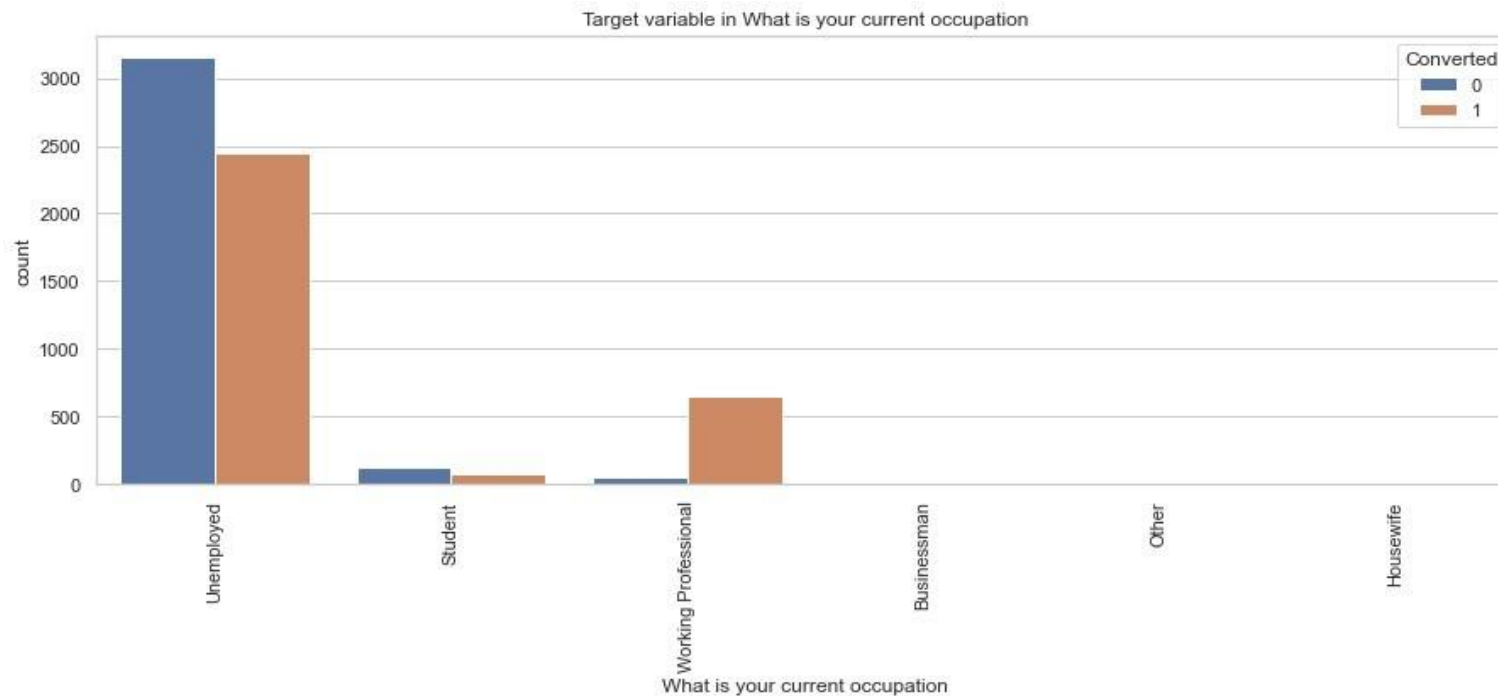
# Last lead Activity

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



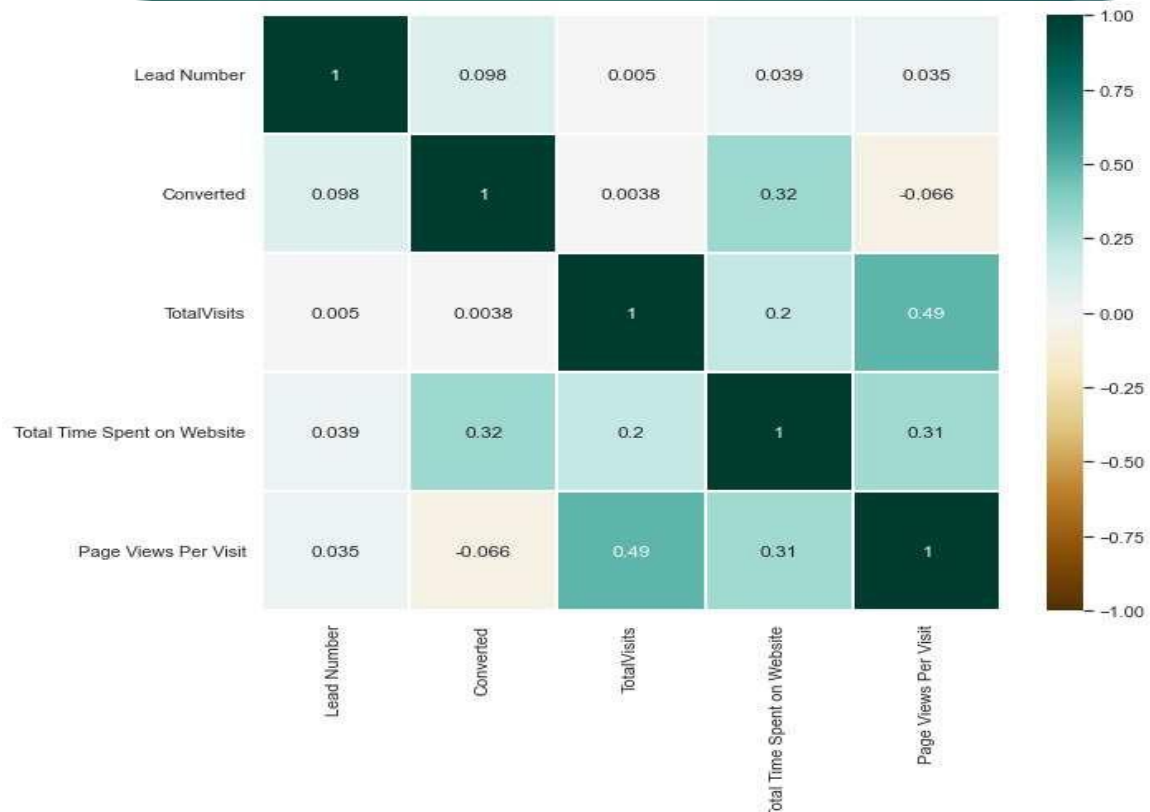
# Last What is Your Occupation

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



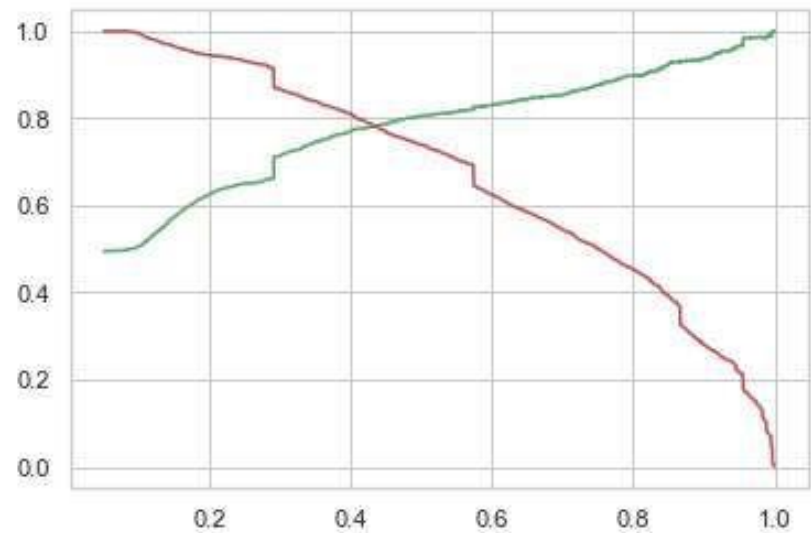
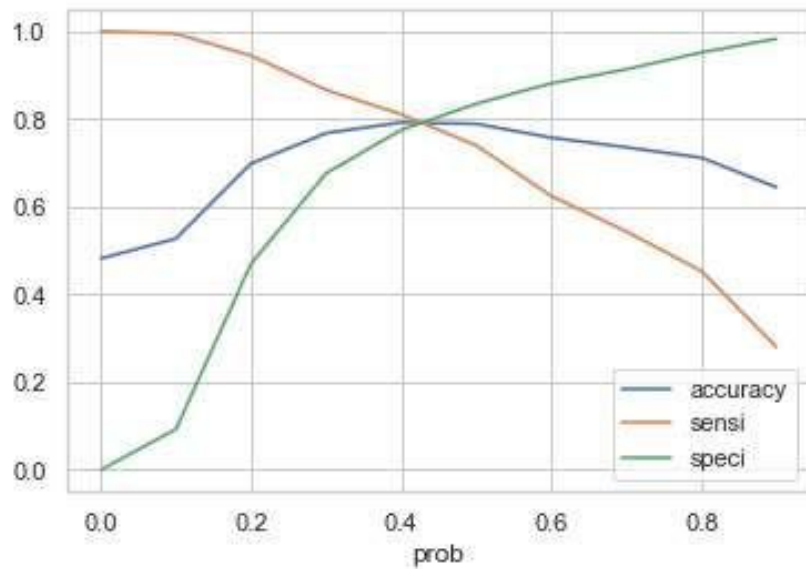
# Correlation

There is no correlation between the variables



# Model Evaluation

## ROC curve



# Observations

- 
- 
- 
- 
- 
- 
- 
-

# 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.