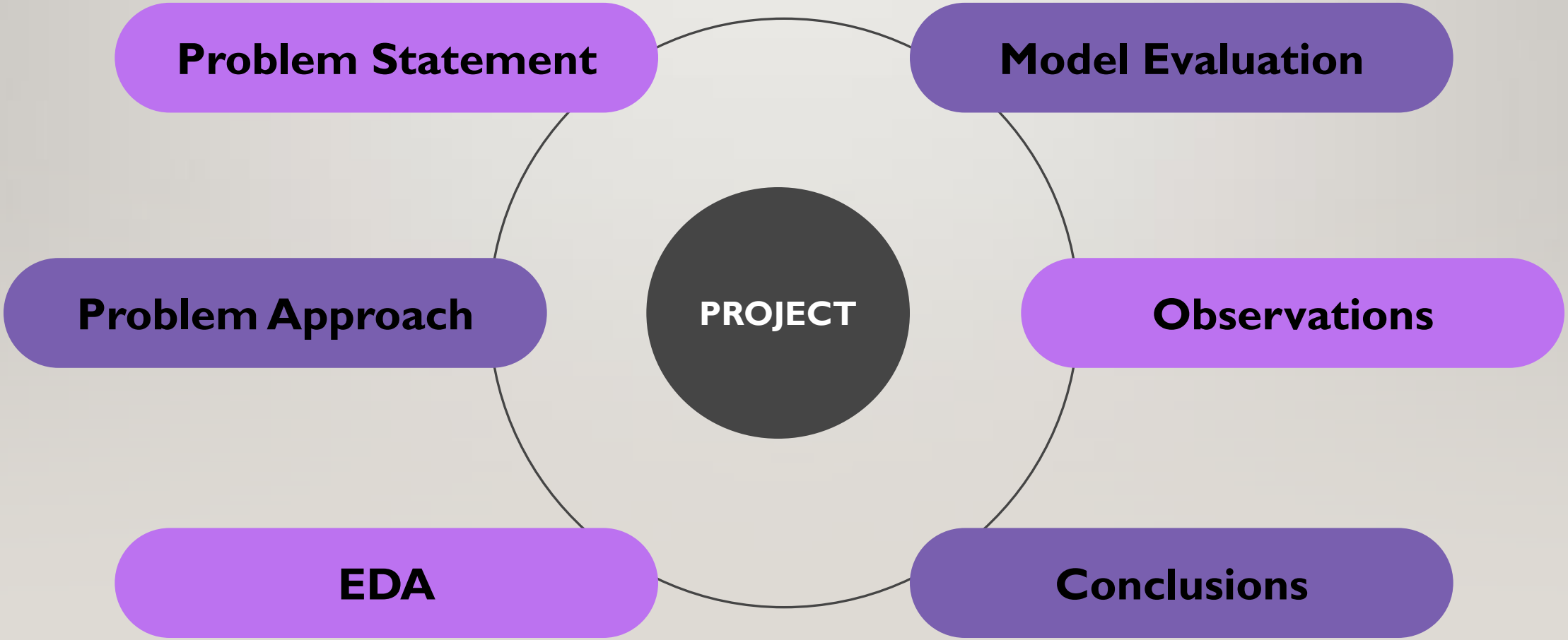
# LEAD SCORING CASE STUDY USING MACHINE LEARNING

PRESENTED BY –

ARGHADEEP CHATTERJEE
APOORVA B PATIL
ANURAG BHARDWAJ

# Contents


Problem Statement
Model Evaluation
Problem Approach
PROJECT
Observations
EDA
Conclusions

# 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%.

# Business Objectives

- We are expected to build a model to give every lead a lead score between 0-100. identify the Hot leads and increase their conversion rate as well.

- The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

- Achieving higher lead conversion.

# Problem Approach

The below mentioned steps have been taken for the problem approach

- Importing Dataset
- Data Preparation
- EDA
- Dummy Variable Creation
- Train-Test Split of Data
- Scaling
- Model Building
- Model evaluation
- Test-set Predictions using Model

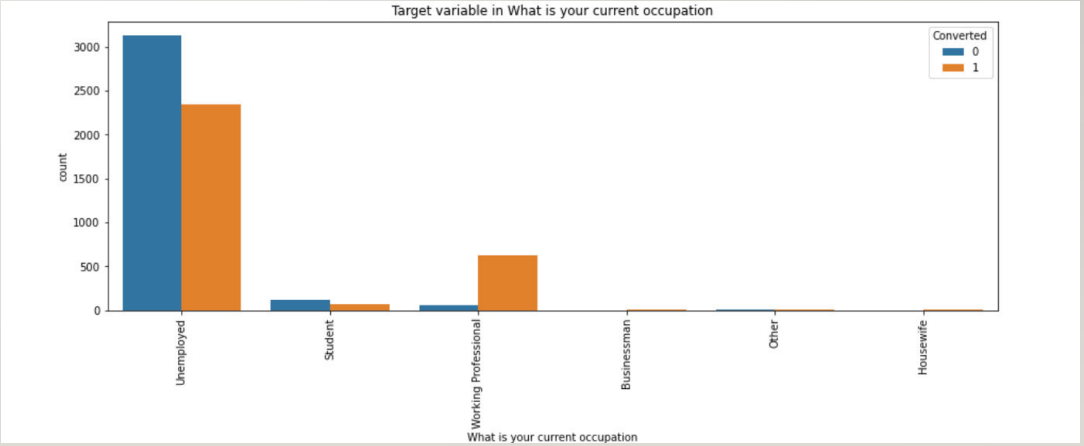
# EDA

1. Data Cleaning:

   Columns like (How did you hear about X Education', 'What matters most to you in choosing a course', 'Lead Profile) have been removed which are not required for analysis.

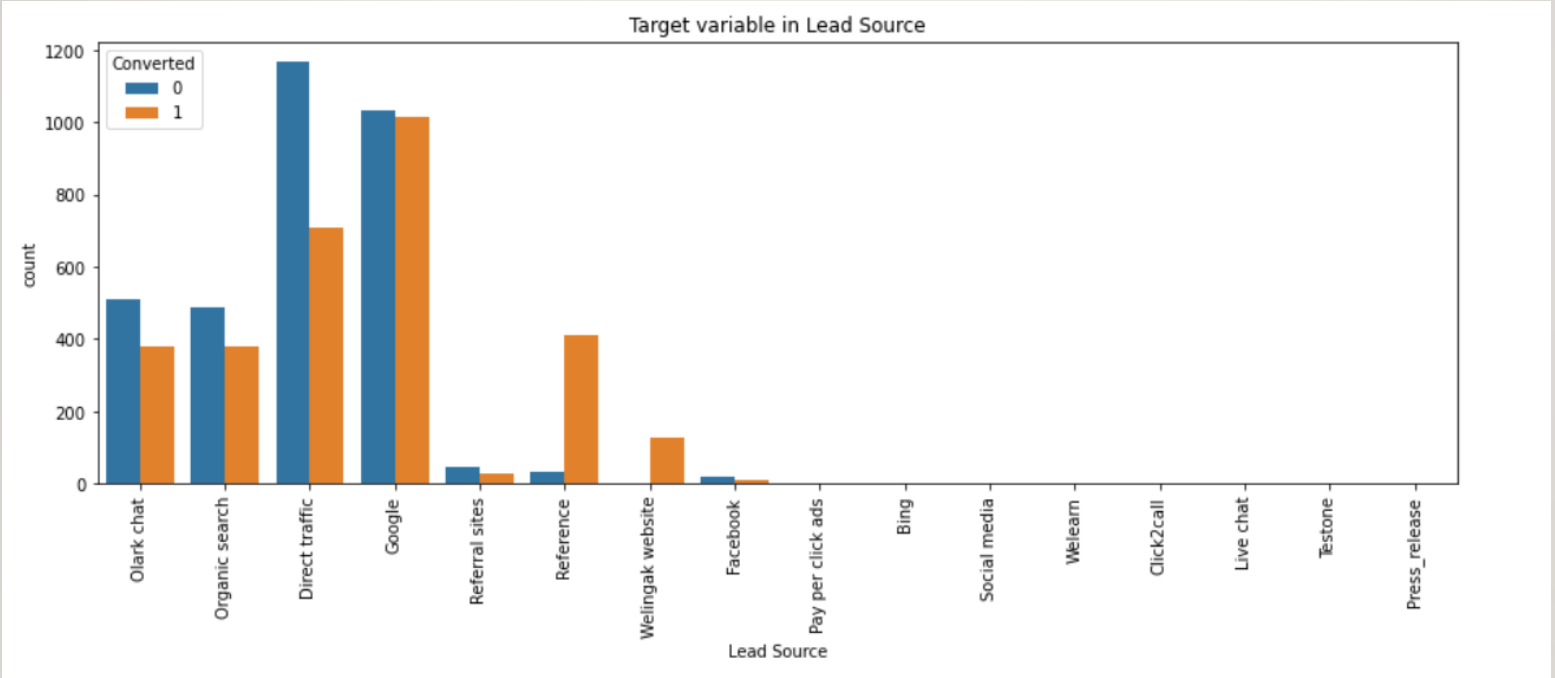
## 2. Bi-variate Analysis:

- **<u>Current Occupation vs converted bar plot:</u>**



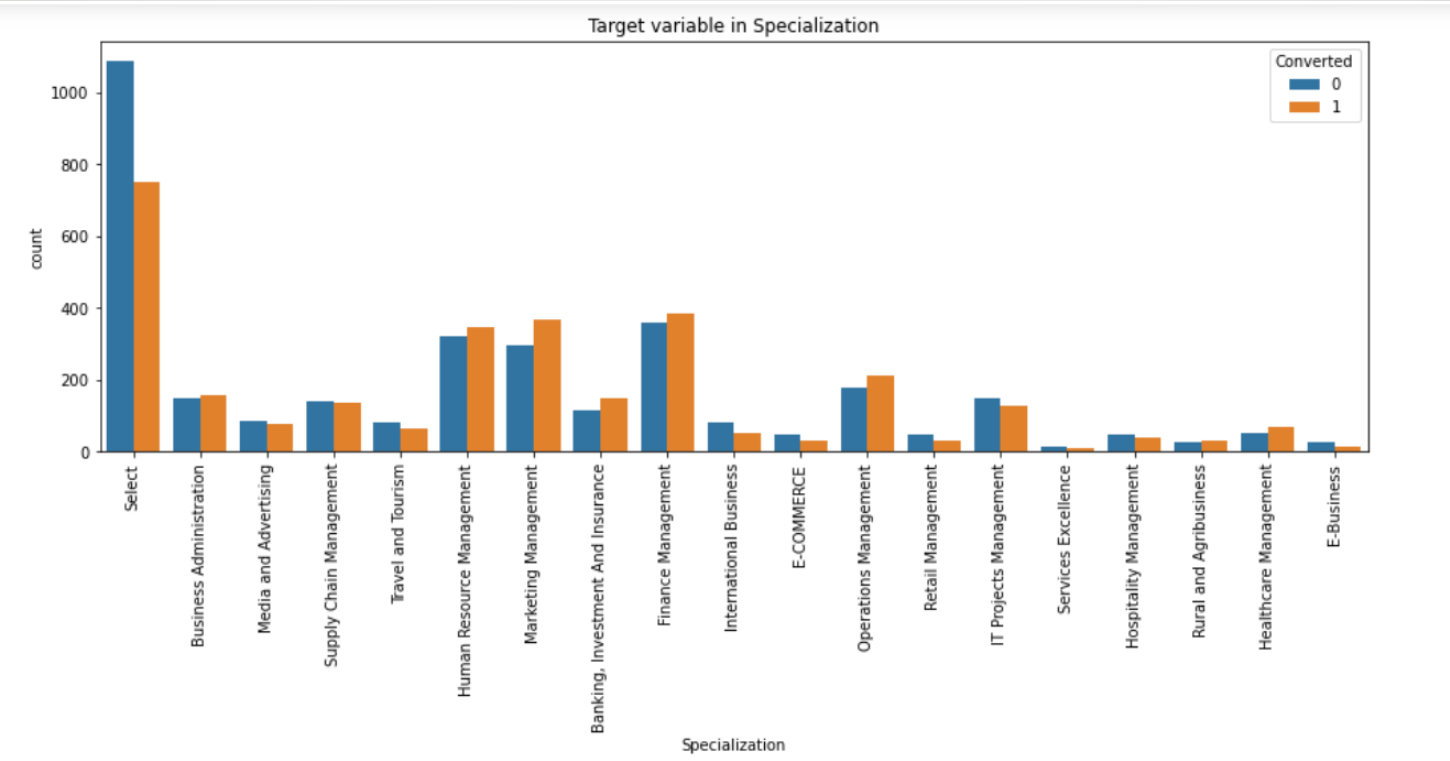

- From the above plot we can observe people who are unemployed are higher chance of taking course

- Lead Source vs Converted:



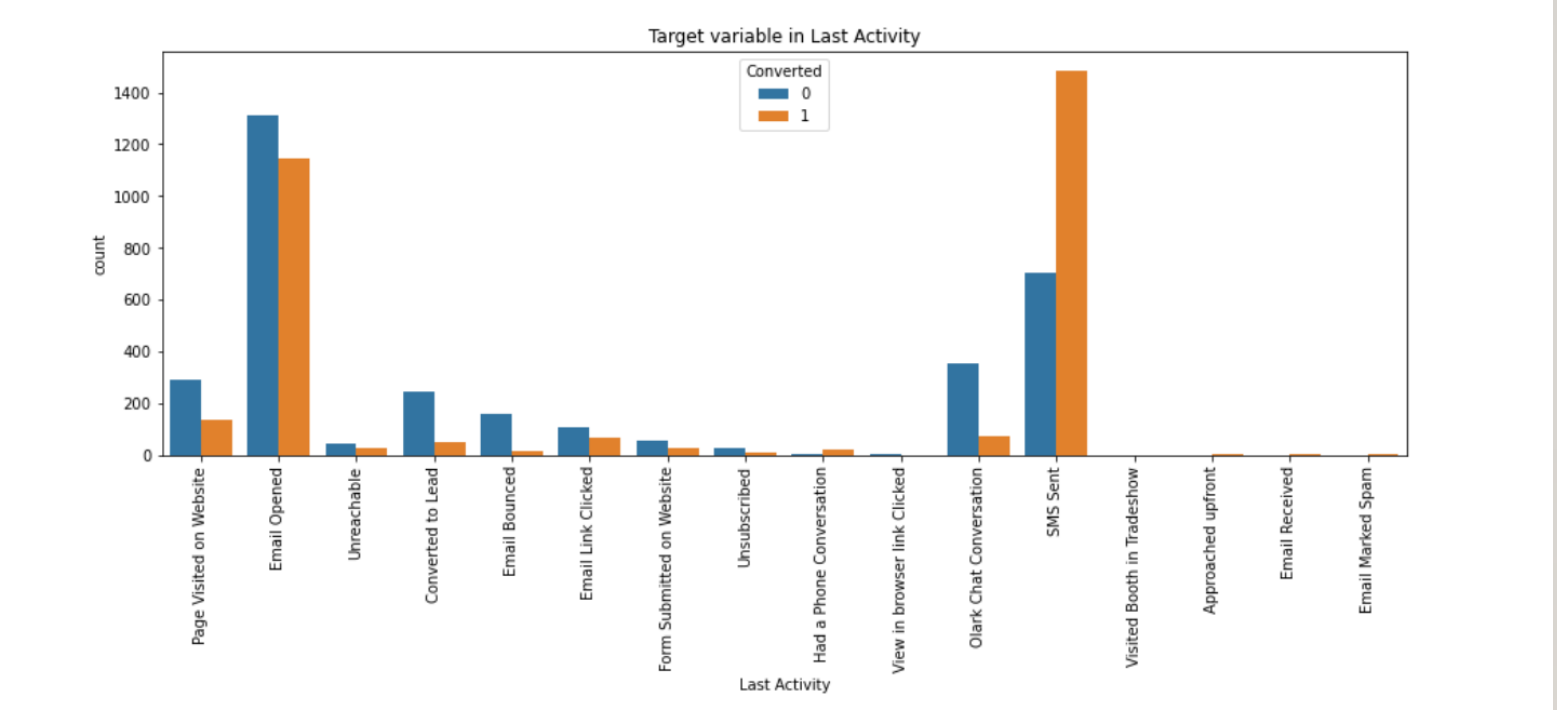

- From the above bar plot we can observe that google and direct traffic have high probability of conversion

- Specialization vs Converted Analysis:




- From the above bar plot we can observe that people from HR, Finance, marketing are more probable to conversion

- Last activity vs Converted Analysis:

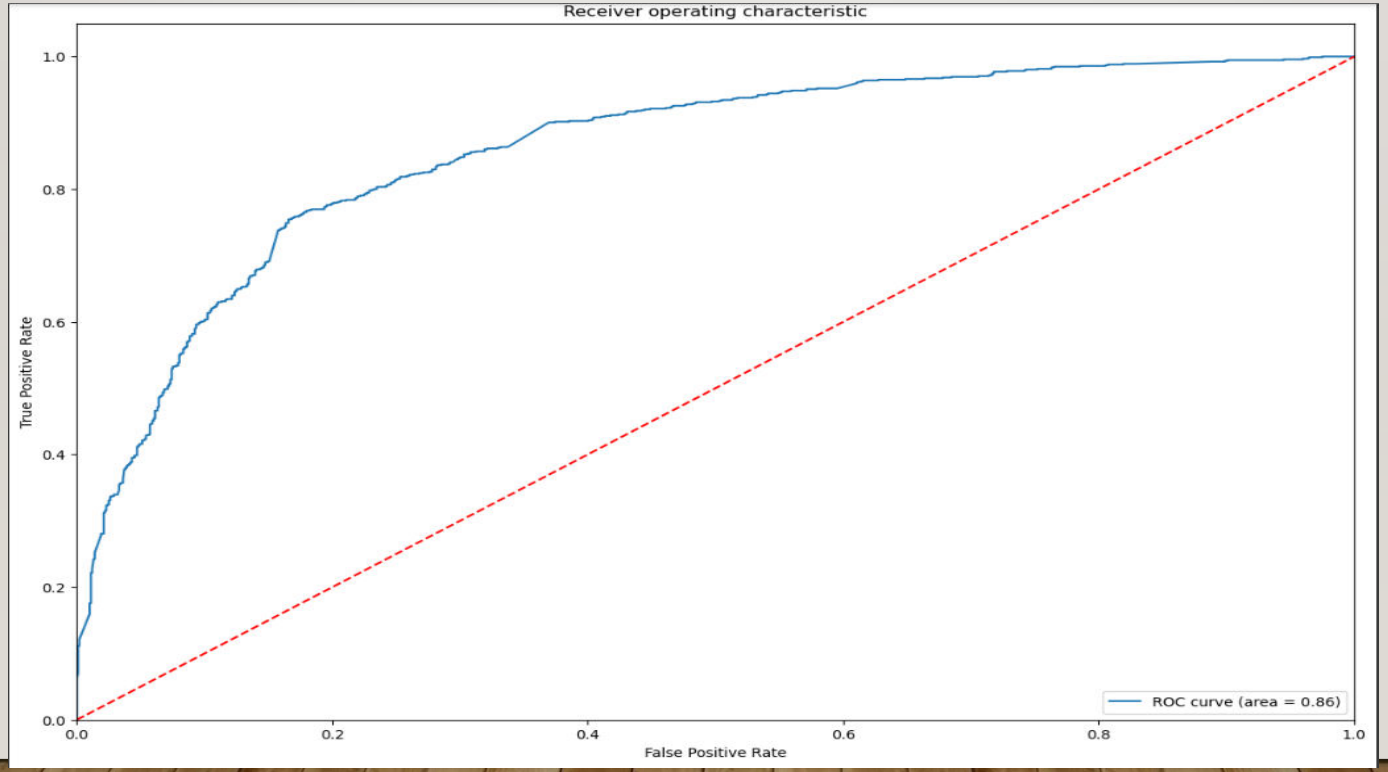



- From the above bar plot we can observe that people to whom SMS was sent and people who opened email have higher chance of conversion

# Model Evaluation:

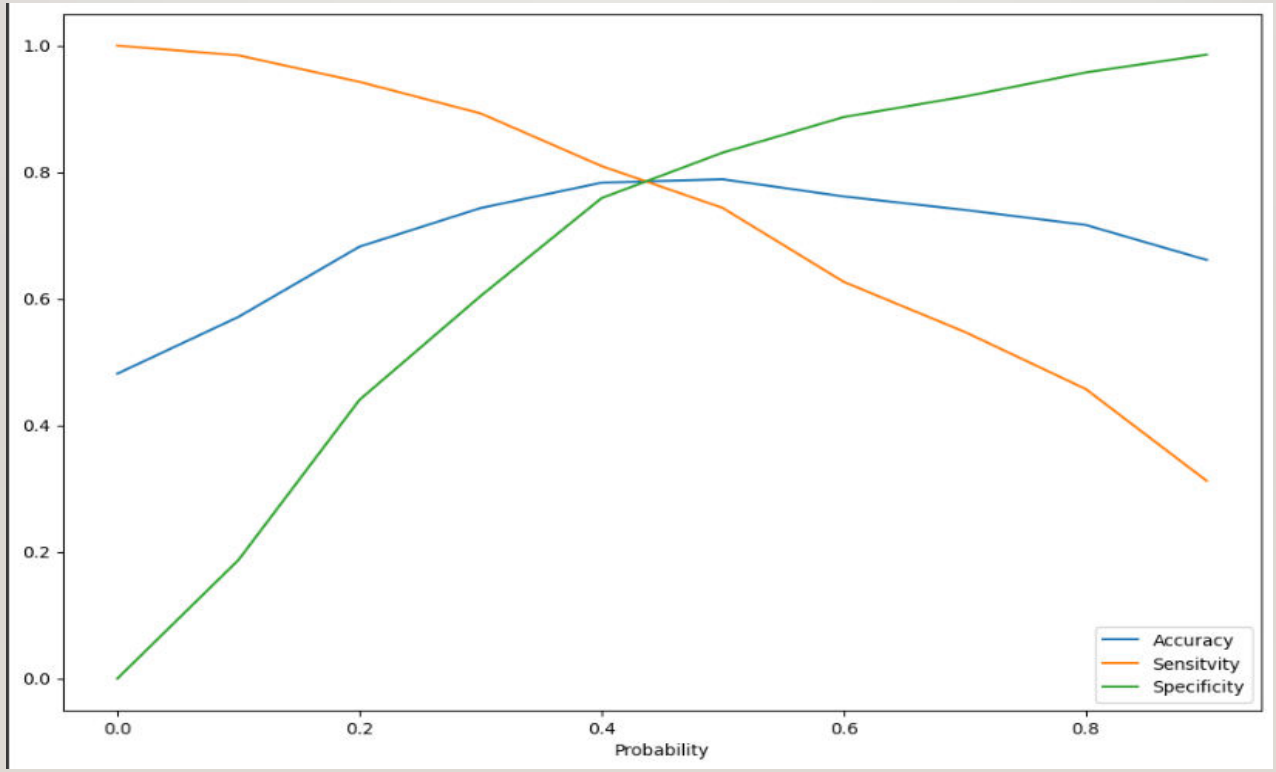
- ROC Curve:

  After building the final model making prediction on it(on train set), we created ROC curve to find the model stability with AUC score(area under the curve) As we can see from the graph plotted on the right side, the area score is 0.86 which is a great score  and our graph is leaned towards the left side of the border which means we have good accuracy



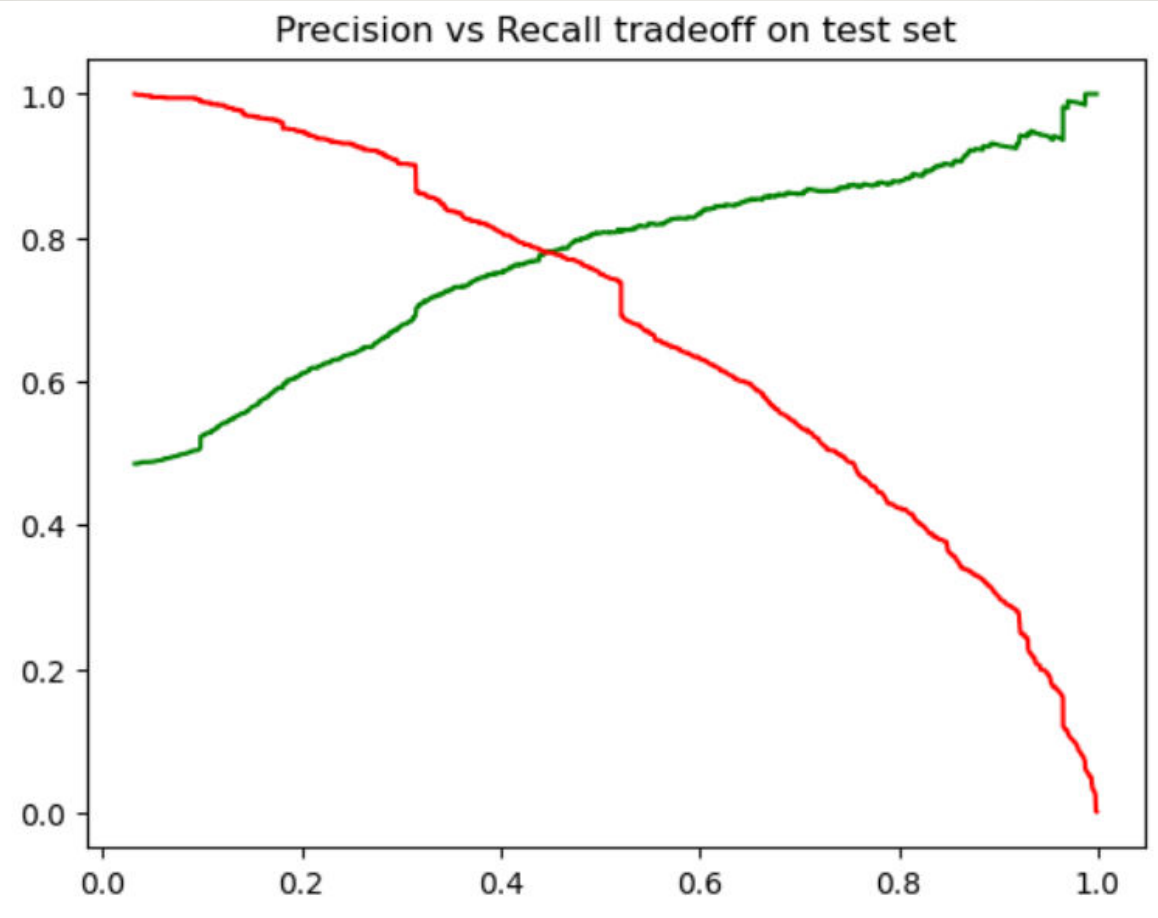
- Finding the optimal cut off point

We found that on 0.45 point all the score of accuracy, sensitivity and specificity are in a close range which is the ideal point to select and hence it was selected.



- PRECISION AND RECALL TRADE OFF POINT

The below graph is showing us the trade off between Precision and recall.• We found that there is a trade off between Precision and Recall and the meeting point is approximately at 0.45



# Observations:

After model evaluation we can observe the sensitivity, specificity, precision, recall and accuracy of both train and test data set.

```
Train Data Set metrics:

Sensitivity: 89.3
Specificity: 60.42
Precision: 67.71
Recall: 89.3
Accuracy: 74.33

Test Data Set metrics:

Sensitivity: 90.28
Specificity: 60.54
Precision: 67.79
Recall: 90.28
Accuracy: 74.79
```

# CONCLUSION

- The ROC curve area is 0.86 which is very acceptable
- The Accuracy, Precision and Recall score we got from the test data are in the acceptable region.
- It has been observed that number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.
- Most common last activity is email opened, and max conversion with working professional.

# THANK YOU