# Lead Score Case Study

SUBMITTED BY :

SWATI SUMAN

SOMIREDDY SRIKANTH REDDY

SIRIKONDA VISHAL KUMAR VERMA

# Lead Score Case Study for X Education

**Problem Statement :**

X Education sells provides online courses to industry professionals.

These courses are advertised on various websites and search engines like Google.

People landing on the website browse the courses and fill up a form providing their email address or phone number.

The company also gets leads through past referrals. After having leads, Sales team contacts them through call or mail, trying to convert them into actual customers.

The typical lead conversion rate at X education is around 30%.

# Lead Score Case Study for X Education

Business Goal :

X Education needs help in selecting the most promising leads who are most likely to convert into paying customers.

The company needs a model where a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

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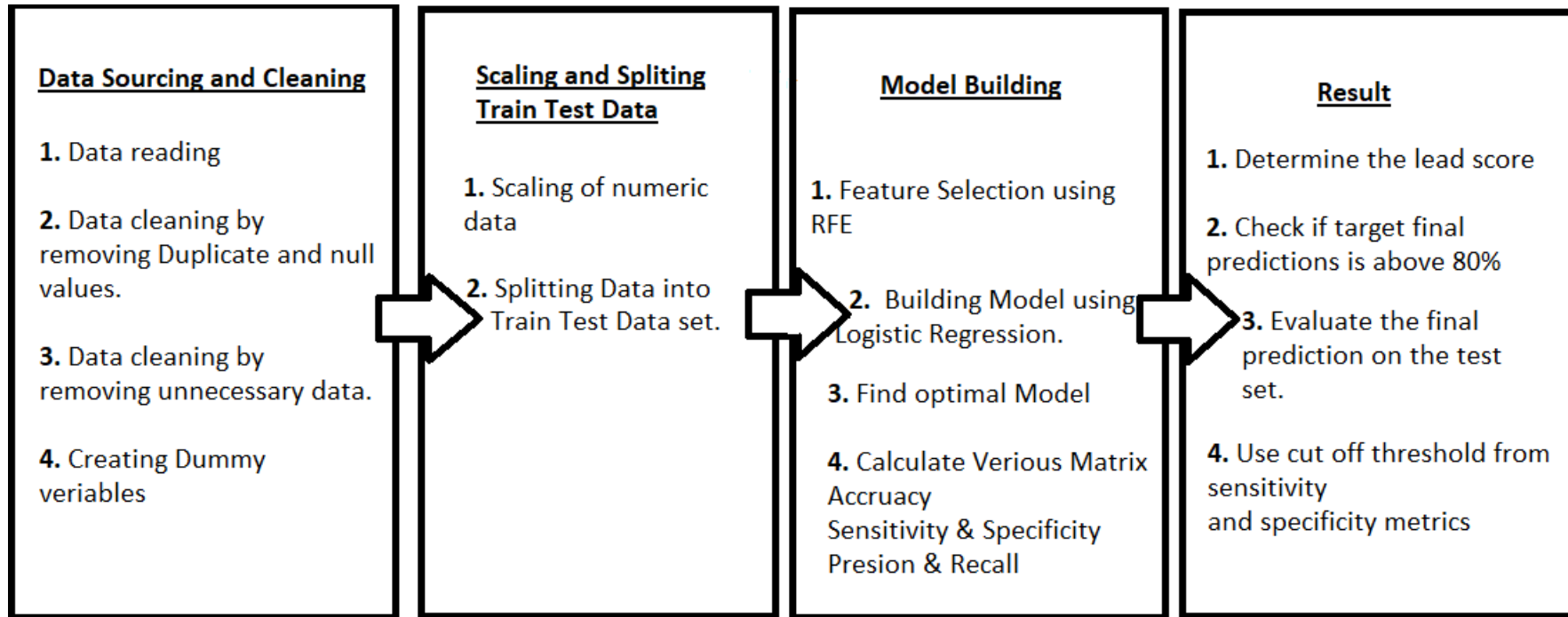
# Model Building Plan

- Source the data for analysis

- Data cleaning and preparation

- Data Analysis using EDA

- Data Scaling

- Test and Train dataset split

- Building a logistic Regression model and calculate Lead Score.

- Evaluating the model by using different metrics – (Specificity and Sensitivity) and (Precision and Recall).

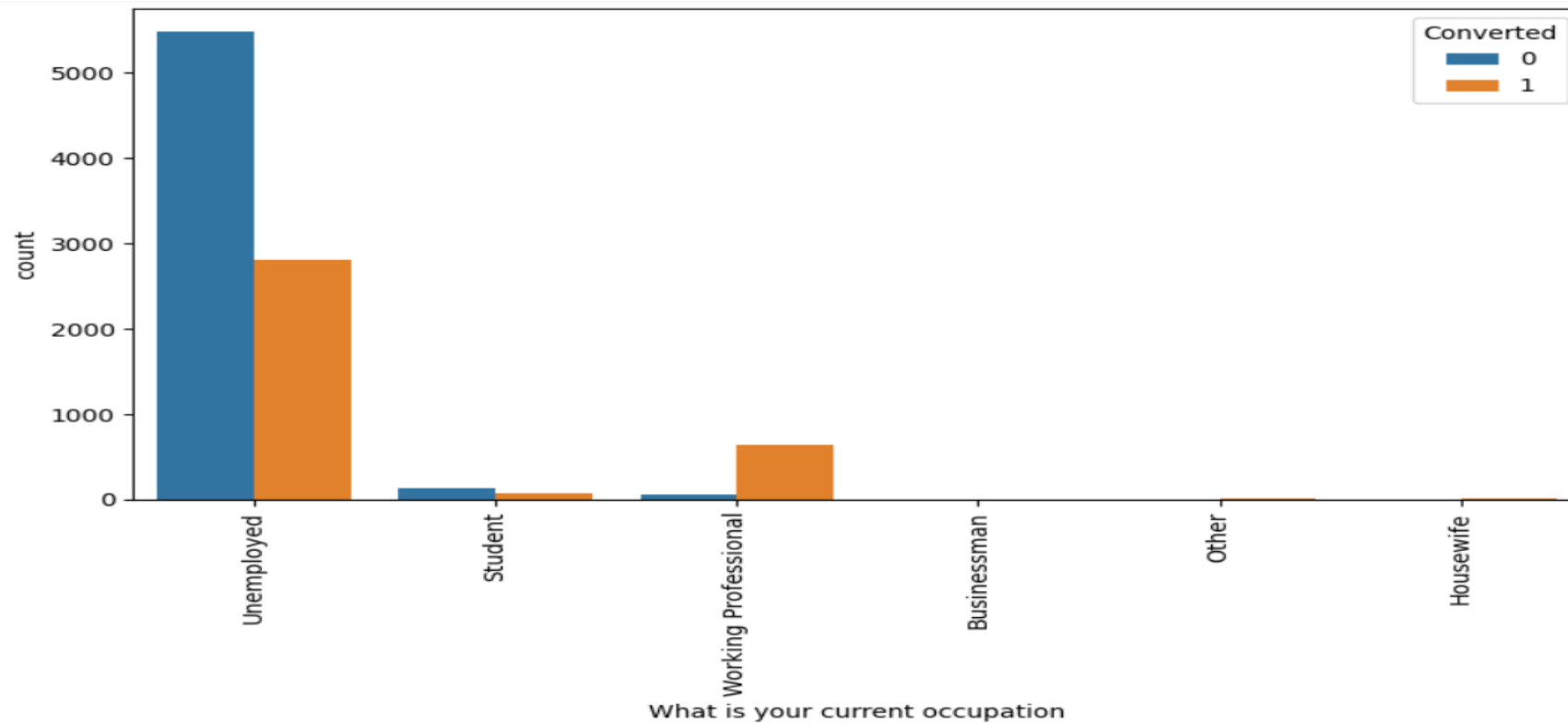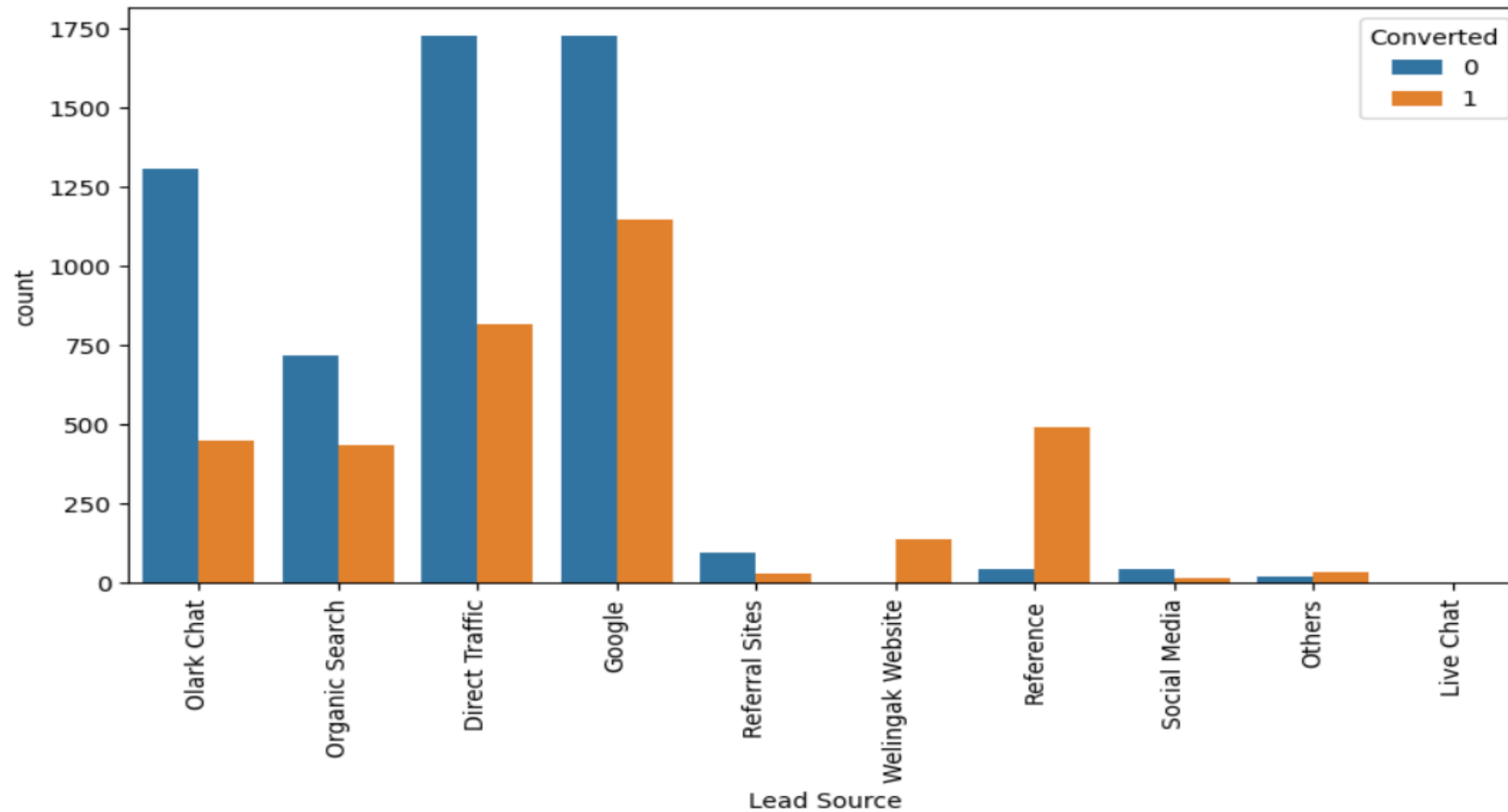- Applying the best model on Test data based.

# Problem Solving

**Data Sourcing and Cleaning**

**1.** Data reading

**2.** Data cleaning by removing Duplicate and null values.

**3.** Data cleaning by removing unnecessary data.

**4.** Creating Dummy veriables

**Scaling and Spliting Train Test Data**

**1.** Scaling of numeric data

**2.** Splitting Data into Train Test Data set.

**Model Building**

**1.** Feature Selection using RFE

**2.** Building Model using Logistic Regression.

**3.** Find optimal Model

**4.** Calculate Verious Matrix
Accruacy
Sensitivity & Specificity
Presion & Recall

**Result**

**1.** Determine the lead score

**2.** Check if target final predictions is above 80%

**3.** Evaluate the final prediction on the test set.

**4.** Use cut off threshold from sensitivity and specificity metrics

# EDA

Conversion for 'What is your current occupation'

# Conversion From 'Lead Score'

# Collinearity between veriables
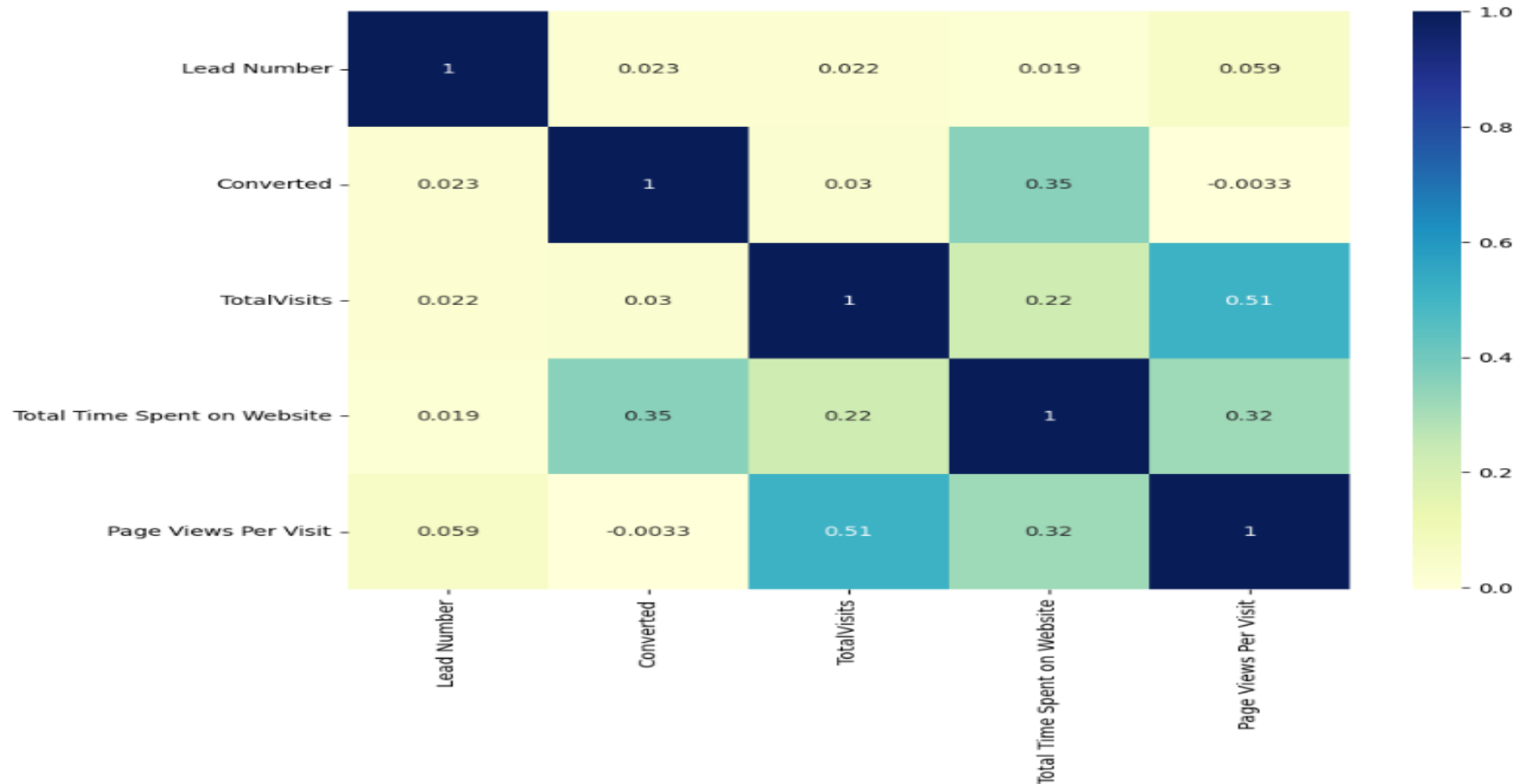
# Feature affecting Conversion Rate

|  | coef |
|---|---|
| const | -0.0323 |
| Total Time Spent on Website | 1.0906 |
| Lead Origin_Lead Add Form | 3.3562 |
| Current_Occupation_Working Professional | 2.7815 |
| Lead Source_Direct Traffic | -1.4611 |
| Lead Source_Google | -1.0043 |
| Lead Source_Organic Search | -1.1360 |
| Lead Source_Referral Sites | -1.3556 |
| Last Activity_Email Bounced | -1.3929 |
| Last Activity_Olark Chat Conversation | -0.9400 |
| Last Activity_SMS Sent | 1.1949 |
| Last Notable Activity_Modified | -0.8610 |
| Last Notable Activity_Olark Chat Conversation | -1.1410 |

# Lead Score Trend
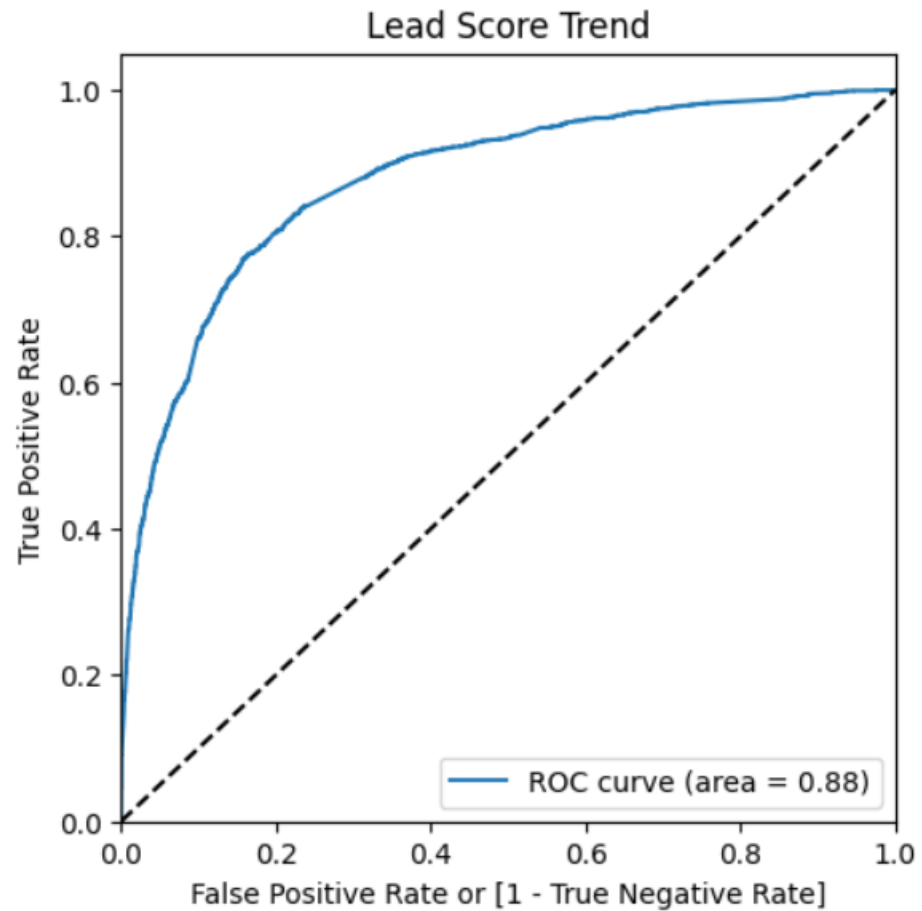
# Sensitivity and Specificity Matrix
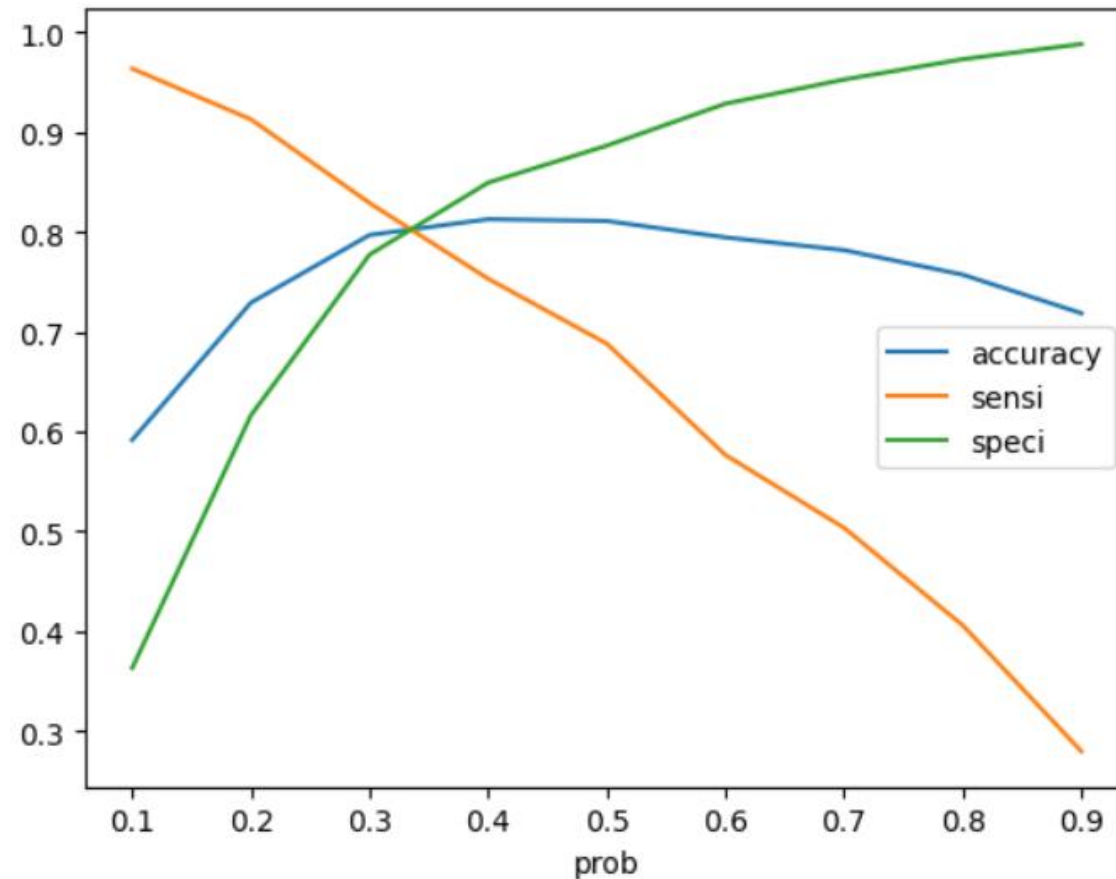
# Observation

So as we can see above the model seems to be performing well.

The ROC curve has a value of 0.88, which is very good. We have the following values for the Train Data:
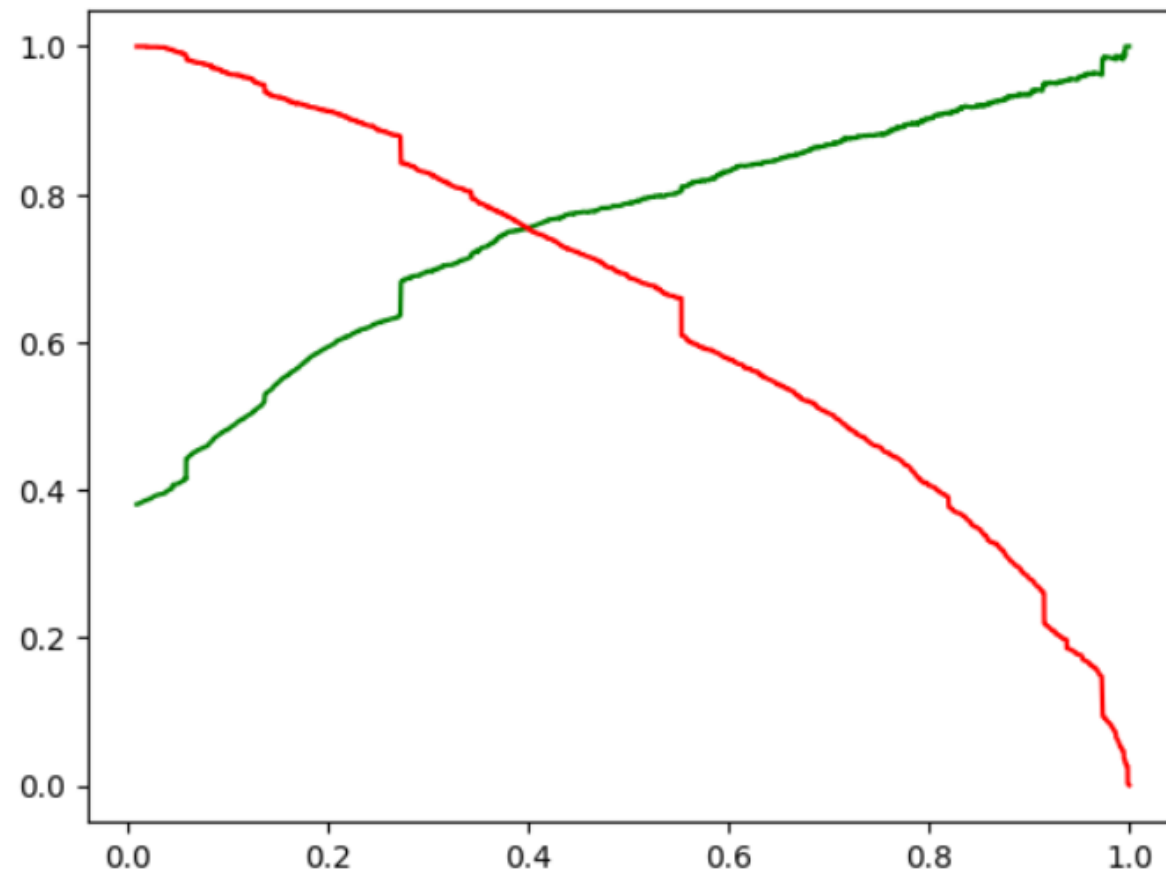
Accuracy : 80%

Sensitivity : 80.83%

Specificity : 79.5%

# Precision and Recall Metrix

# Final Observation

Train Data:

Accuracy : 80%

Sensitivity : 80.83%

Specificity : 79.5%

Test Data:

Accuracy : 80.2%

Sensitivity : 81%

Specificity : 79.7%

# Conclusion

❑ We have considered the optimal cut off of 0.33 based on Sensitivity and Specificity for calculating the final prediction.

❑ Accuracy, Sensitivity and Specificity values of test set are around 80.2%, 81% and 79.7% which are approximately closer to the respective values calculated using trained set.

❑ Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set

❑ The top 3 variables that contribute for lead getting converted in the model are

• Lead Add Form (Lead Origin )

• Working Professional (Current_Occupation)

• SMS Sent (Last Activity)


❑ Overall model prediction score is 80%.