LEAD SCORE CASE STUDY

- Simrah Suhail khan
- Siddhant Naik
- Sindhu balaga

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%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor. For example, 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. A typical lead conversion process can be represented using the following funnel:

BUSINESS OBJECTIVE...

1. Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

2. A higher score would mean that the lead is hot, i.e. is most likely to convert know as ‘Hot Leads’

3. A lower score would mean that the lead is cold and will mostly not get converted.

4. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

STEPS INVOLVED .....

1. IMPORTING AND INSPECTING DATASET
2. DATA CLEANING
3. EDA AND DATA PREPARATION
4. MODEL BUILDING
5. MODEL EVALUATION
6. CONCLUSION

DATA CLEANING ....

- We removed the unnecessary columns from the data set eliminating the data set with null Values higher than 40 %

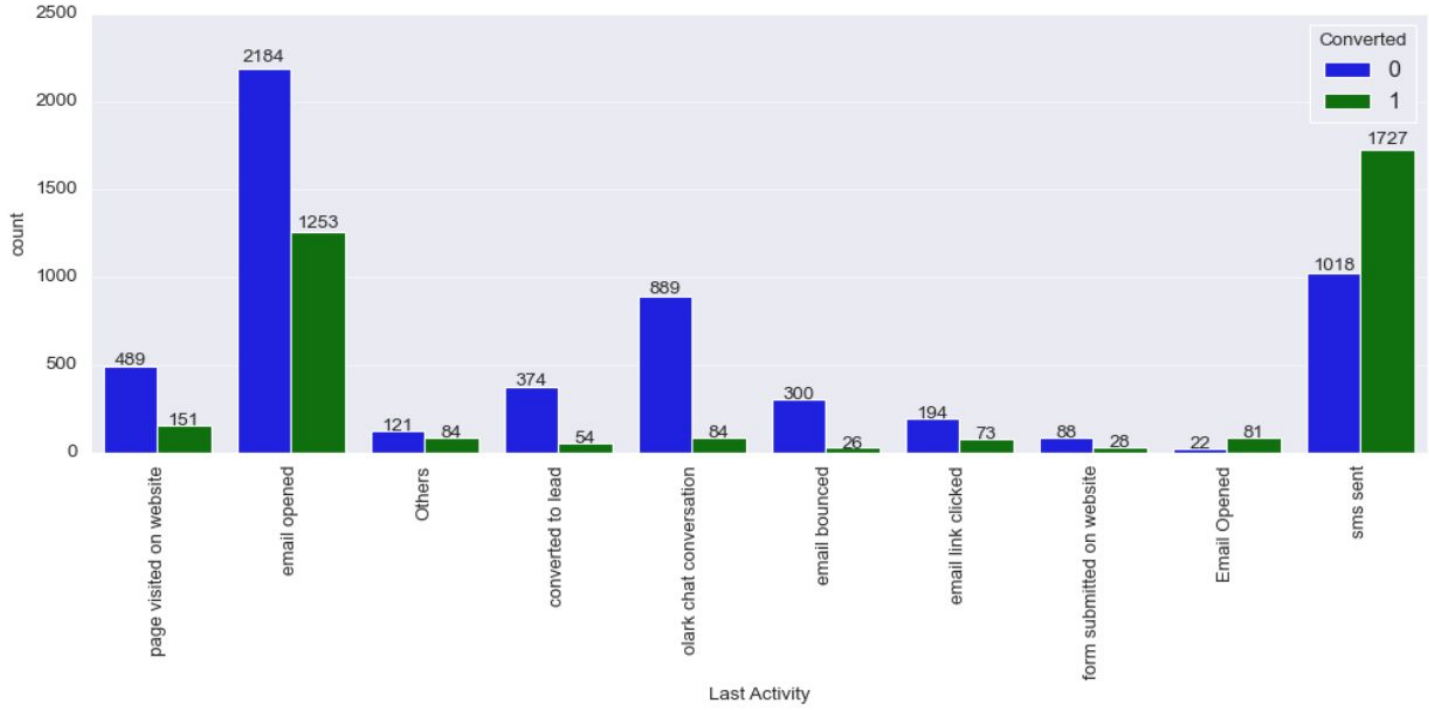
- The columns with skewed values were also removed

- After removing we were left with 13 columns to work ahead

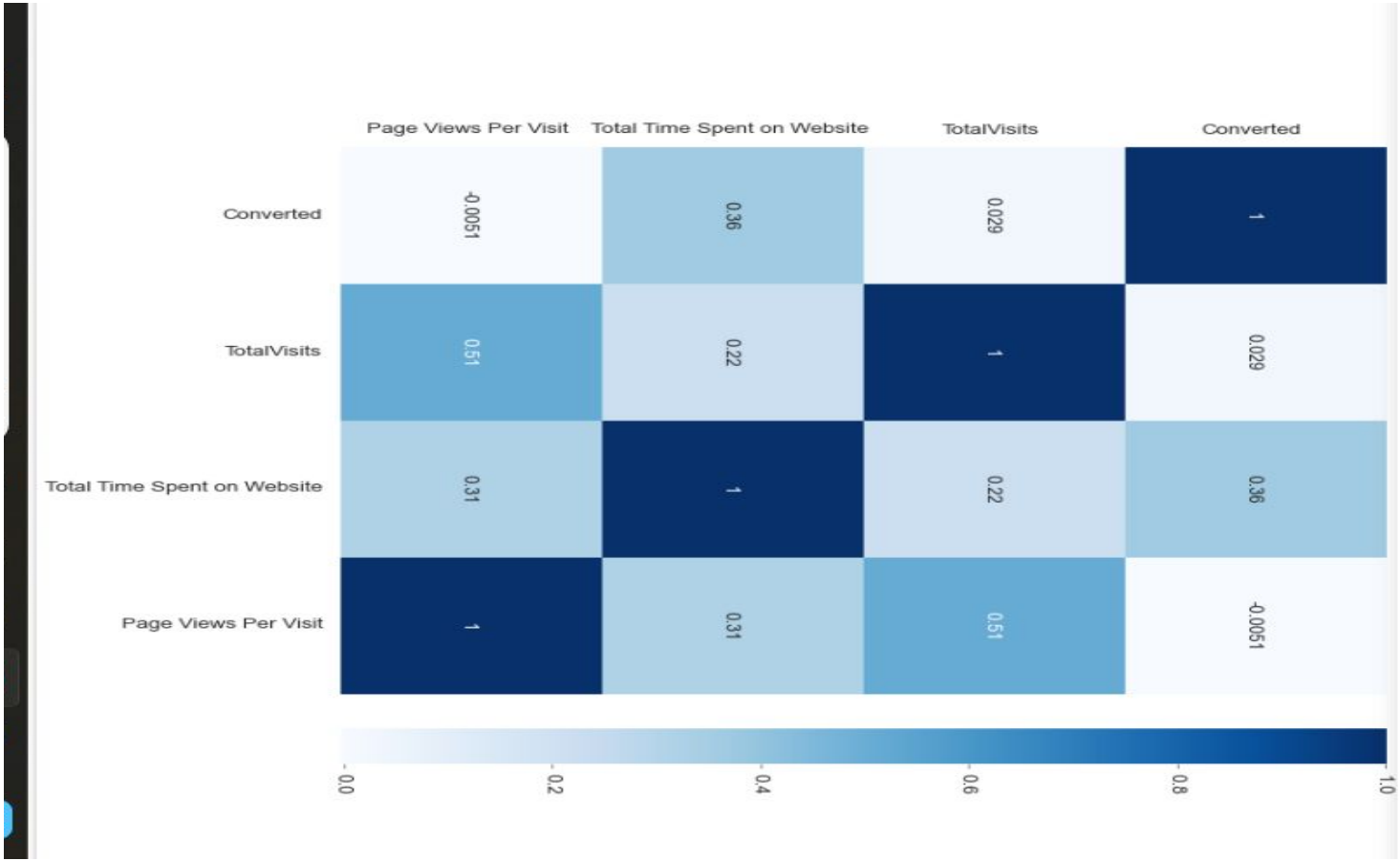
• EXPLORATORY DATA ANALYSIS

• UNIVARIATE ANALYSIS

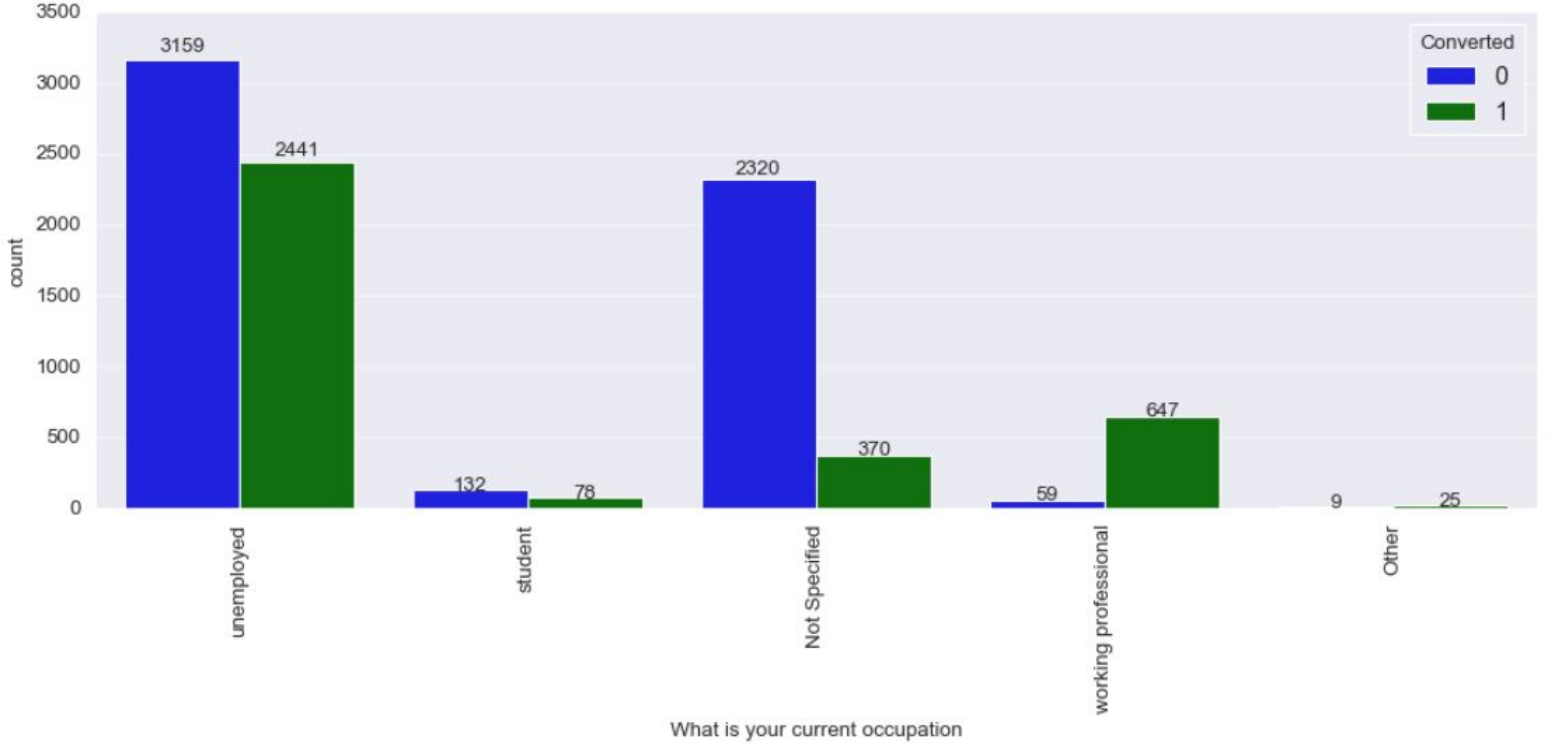
```
varcate('Last Activity')
```



correlation between numerical variables

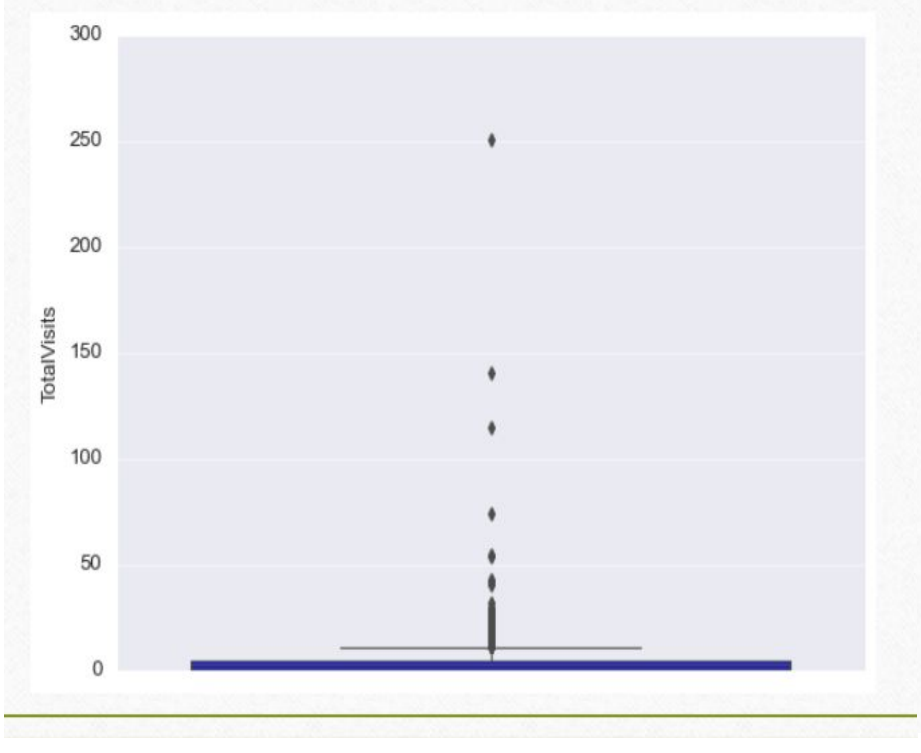


UNIVARIATE ANALYSIS (OCCUPATION)

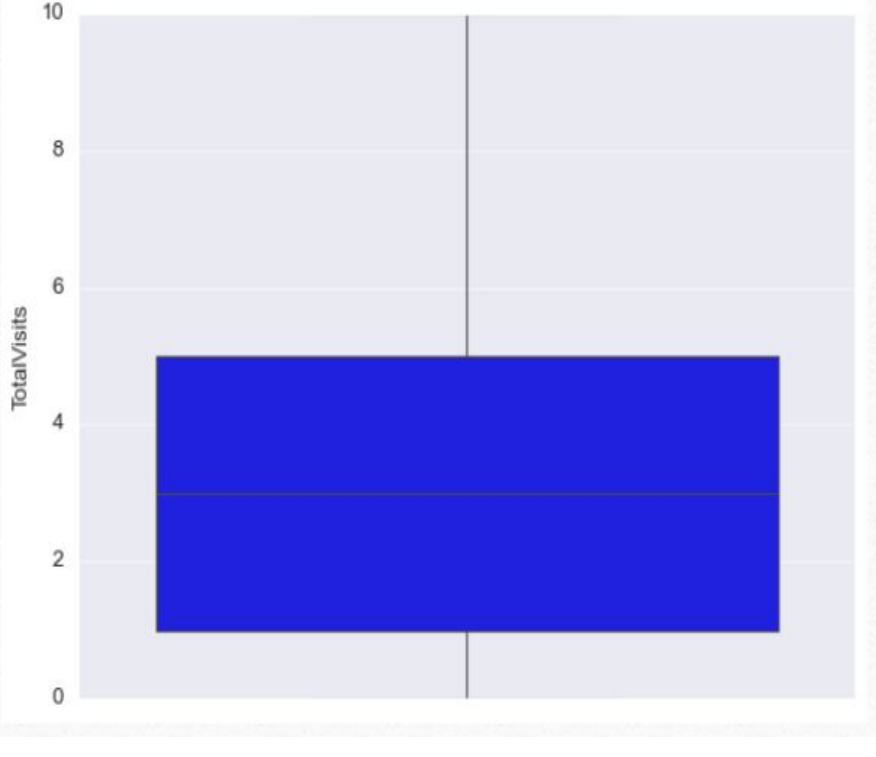


OUTLIER TREATMENT (TOTAL VISITS)

BEFORE TREATMENT



AFTER TREATMENT

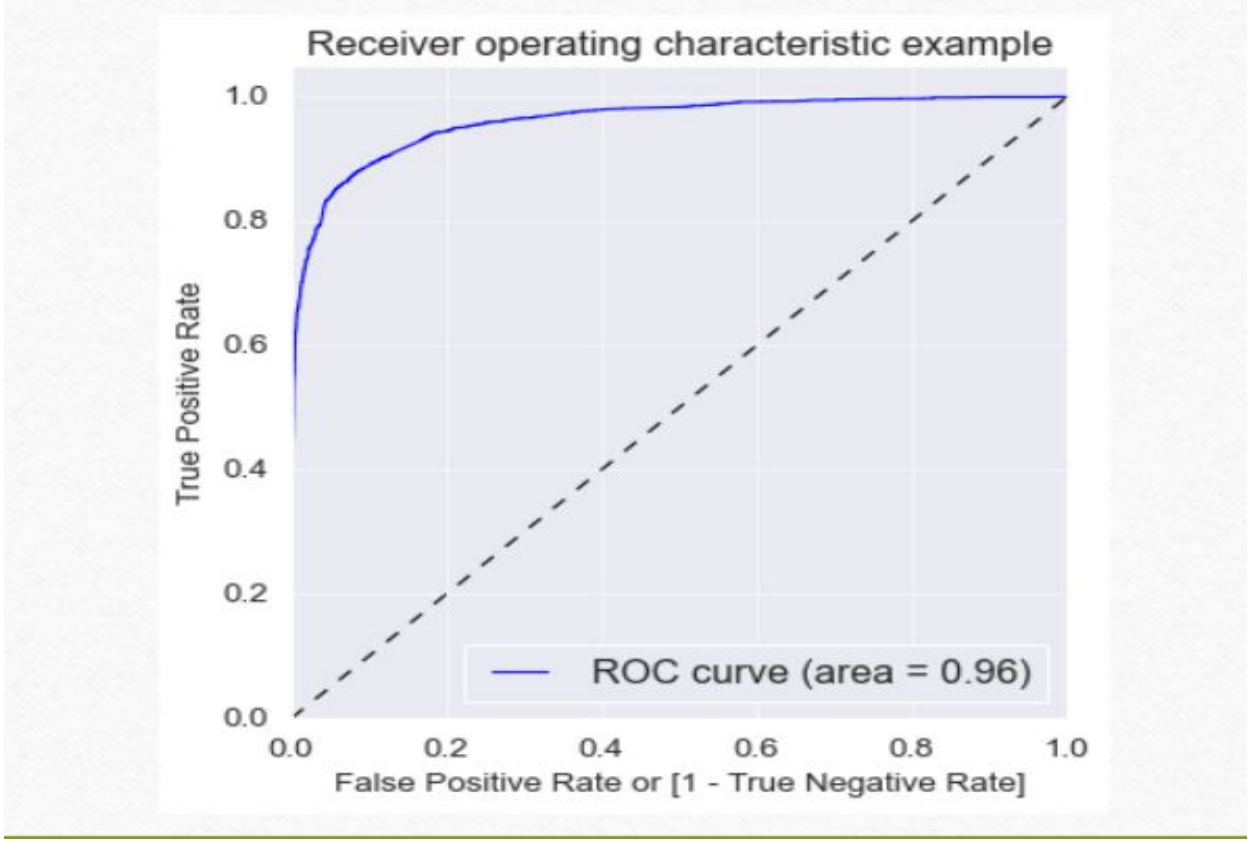


DATA PREPARATION

- Dummy variables were created for the categorical columns
- Scaling of the data was done • The test train split on the data was carried out
- Features selection was done on the data set with the help of RFE
- VIF was also used to check the optimal features
- p-value 0.05 and RFE < 5

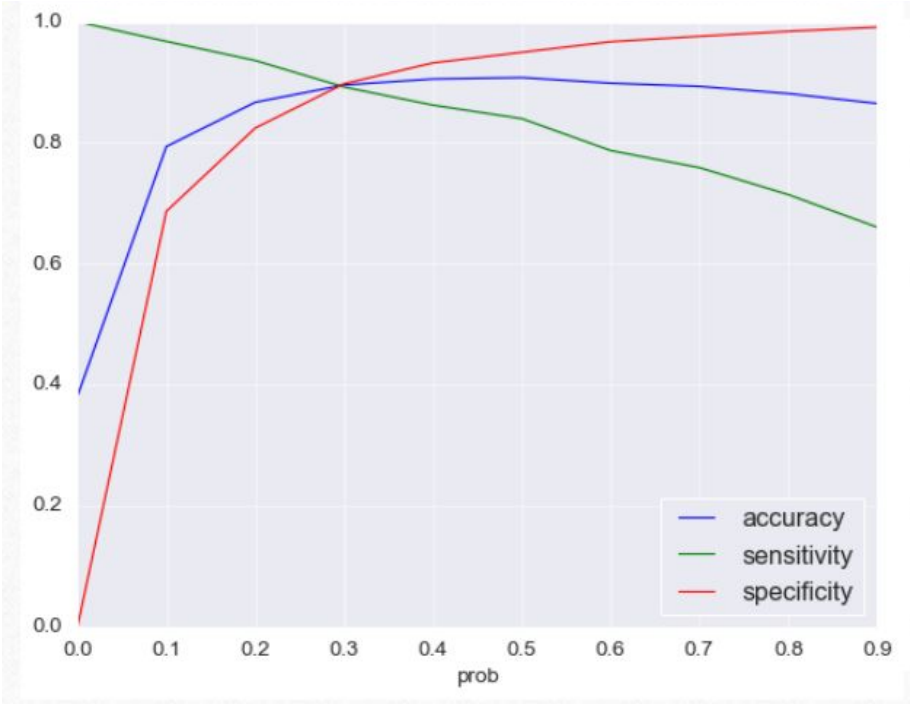
MODEL EVALUATION

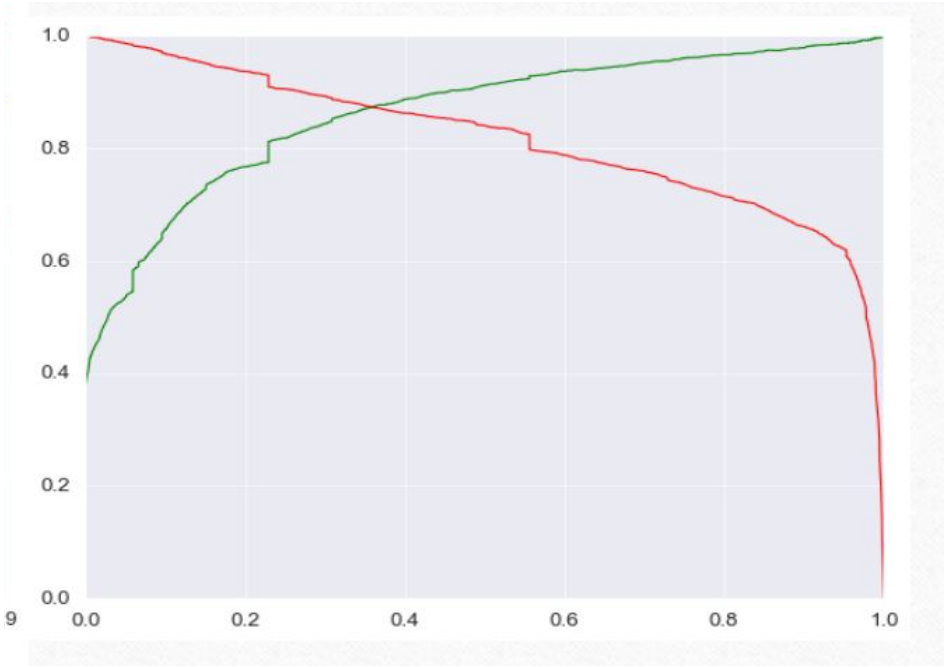
- **ROC Curve**



Model evaluation Train

1. Accuracy – 89.54%
2. Sensitivity-89.21%
3. Specificity – 89.75%
4. Precision-84.29%
5. Recall – 89.21%





MODEL EVALUATION TEST

Accuracy - 89.61%

Sensitivity – 89.68%

Specificity – 89.56%

Precision – 84.87%

Recall – 89.68%

RECOMMENDATIONS

Important features responsible for good conversion rate or the ones' which contributes more towards the probability of a lead getting converted are :

- Tags_closed by horizzon
- Tags_will revert after reading the email
- Lead Source_welingak website

- The evaluation metrics are pretty close to each other so it indicates that the model is performing consistently across different evaluation metrics in both test and train dataset.

- A. The model achieved a sensitivity of 89.21% in the train set and 89.68% in the test set, using a cut-off value of 0.345.

- B. The model also achieved an accuracy of 89.61%, which is in line with the study's objectives.The CEO of X Education had set a target sensitivity of around 80%.