

---

# LEAD SCORE CASE STUDY

+

o

•

Prepared By : Akhil Sinha

Yazat Sharma

Sandeep K

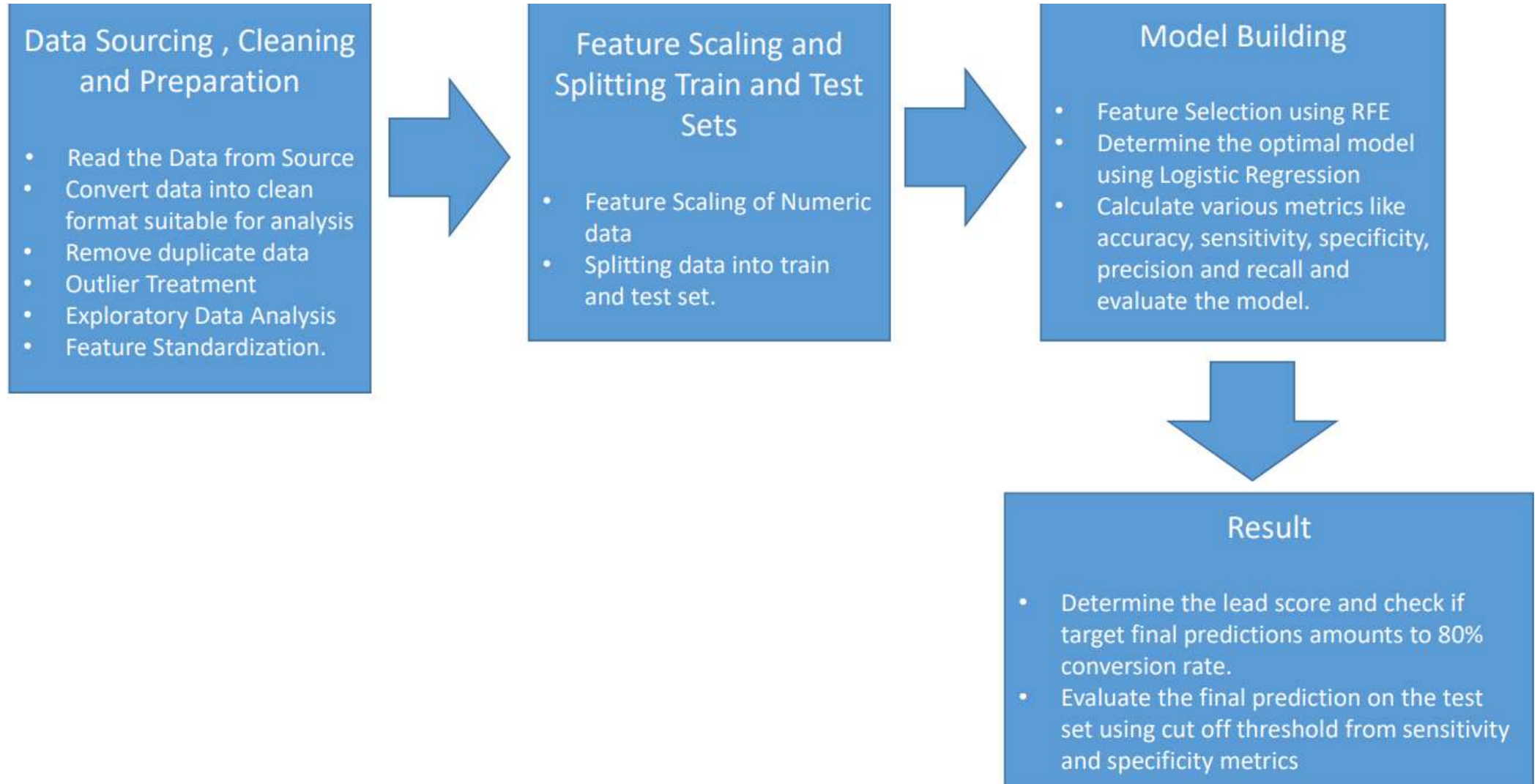
# Problem Statement

- X Education sells online courses to industry professionals. 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
- X Education sells company needs the information of those leads who have high probability of enrolling or buying the course referred as hot leads.

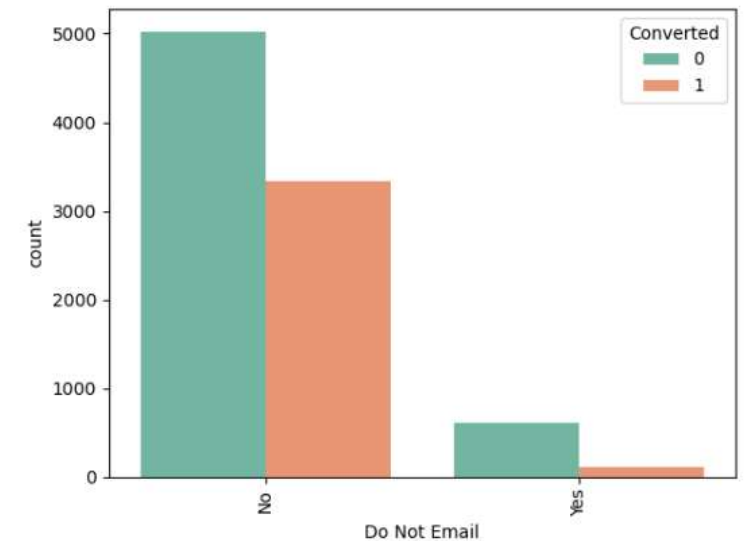
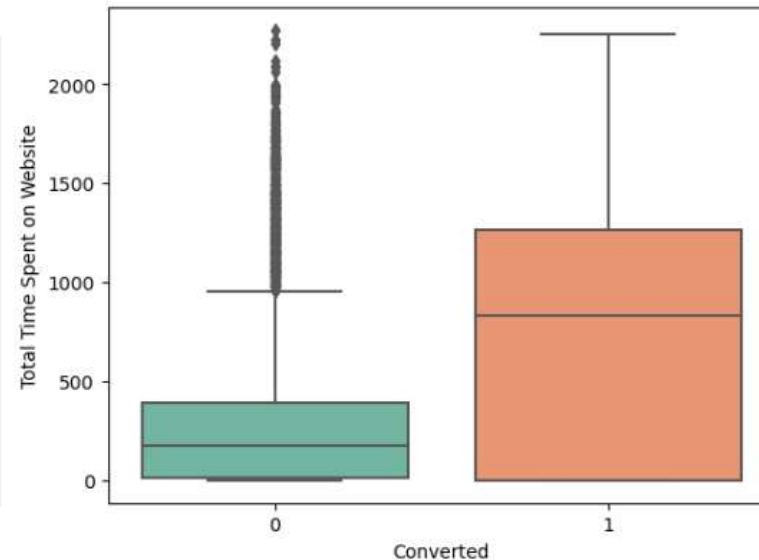
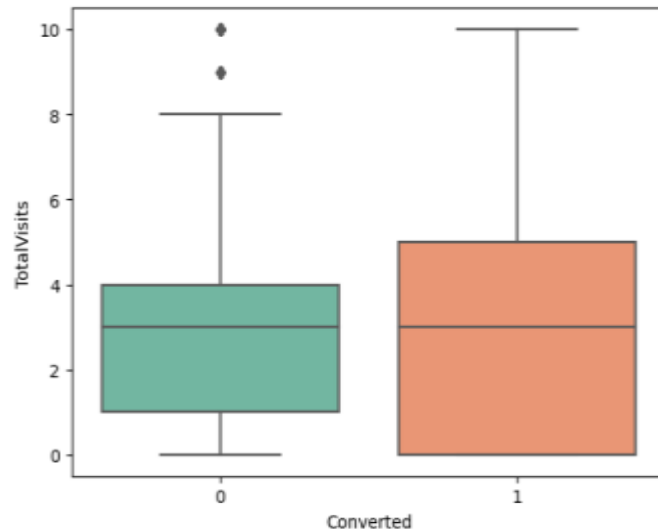
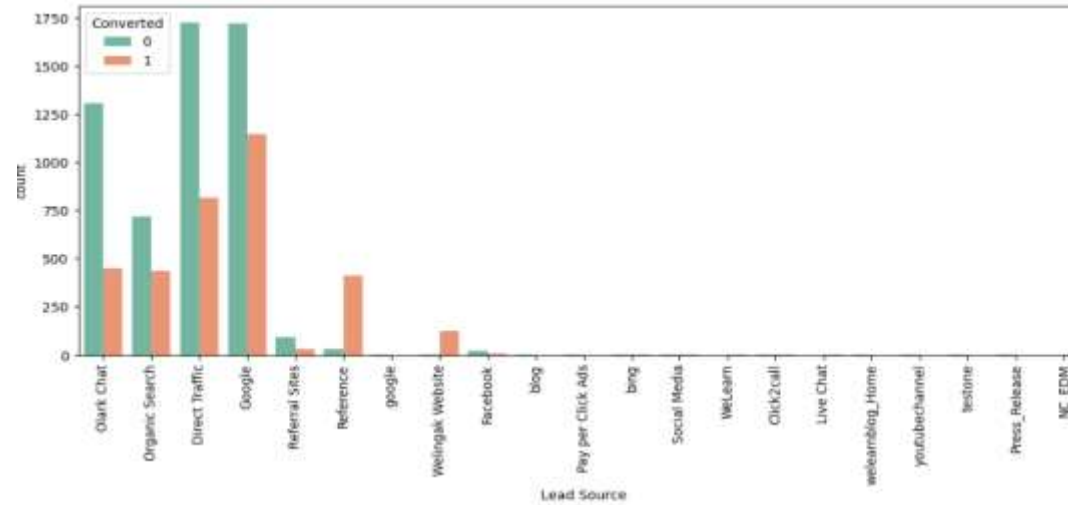
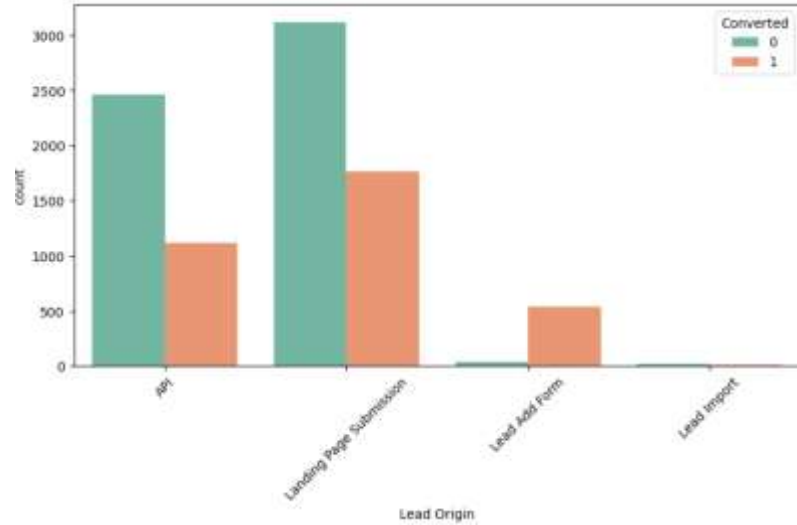
# Strategy Implemented

- Source the data for analysis.
- Data Cleaning and Preparation.
- Exploratory Data Analysis.
- Feature Scaling – using RFE(Recursive Feature Elimination)
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

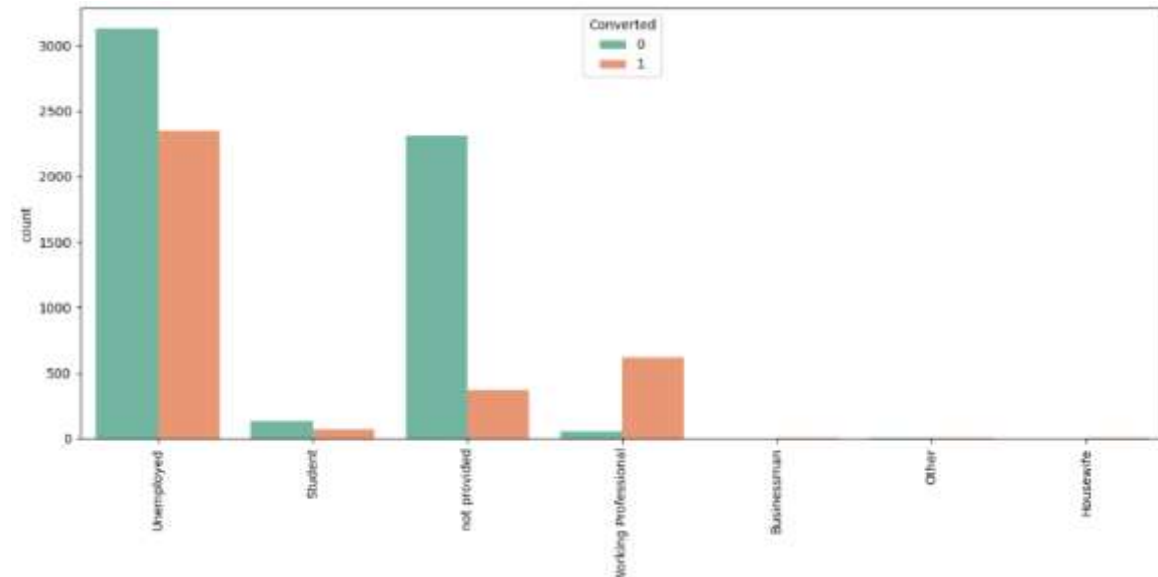
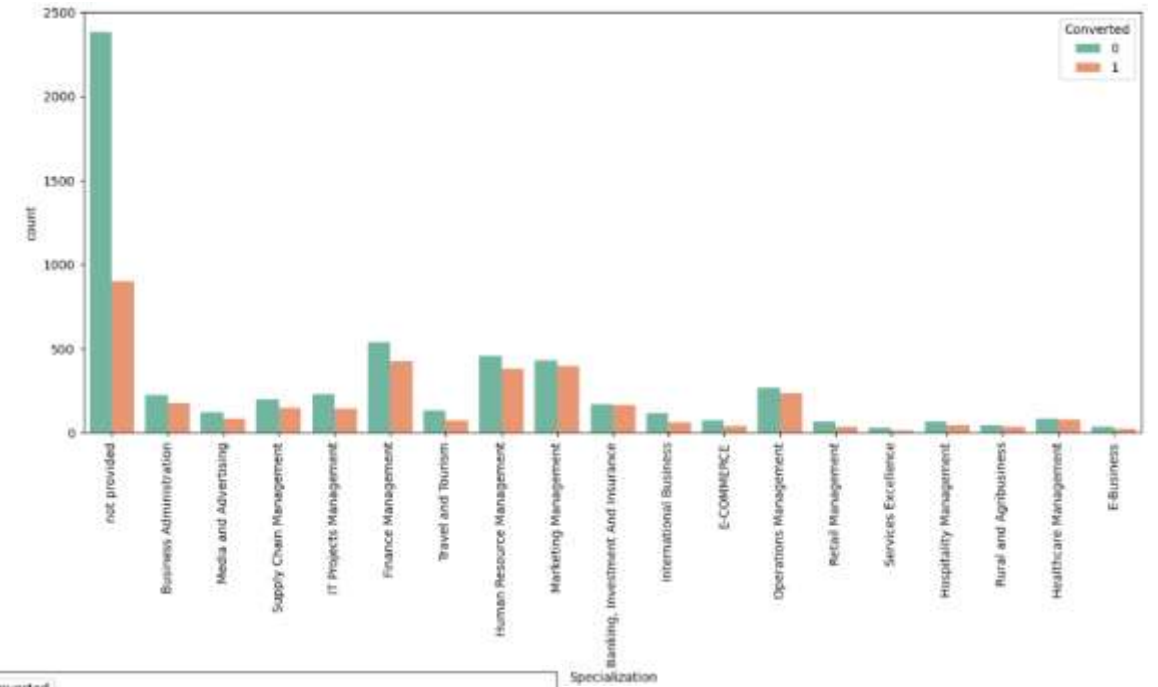
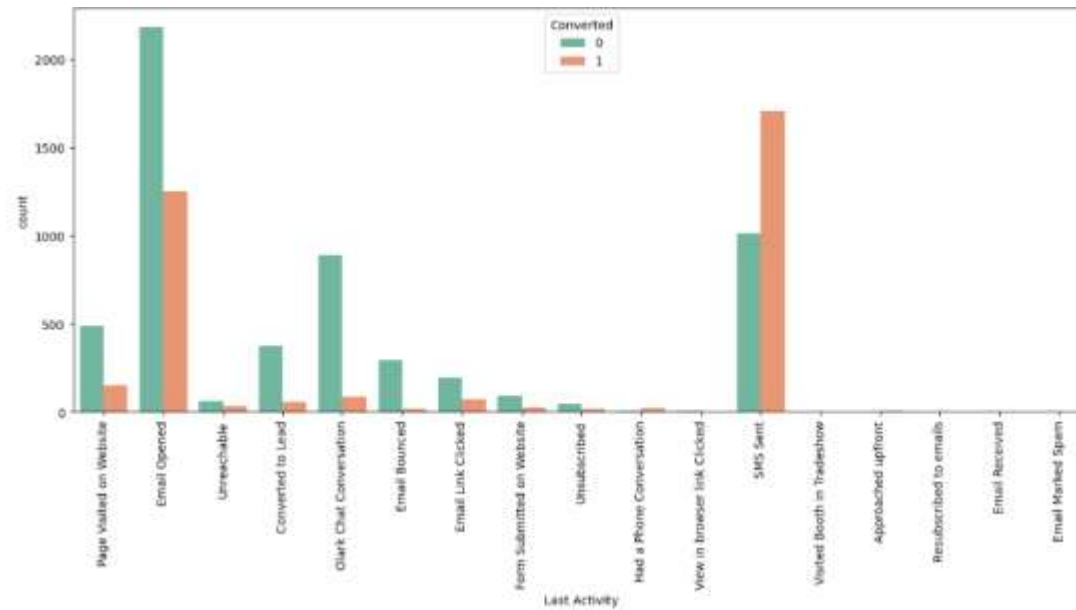
# Methodology



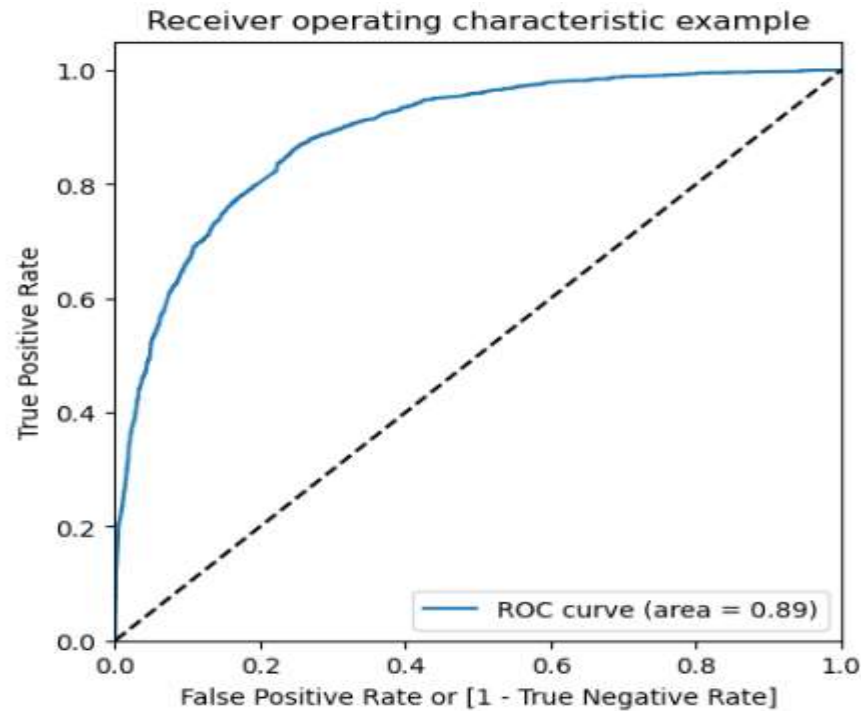
# Exploratory Data Analysis



# Exploratory Data Analysis



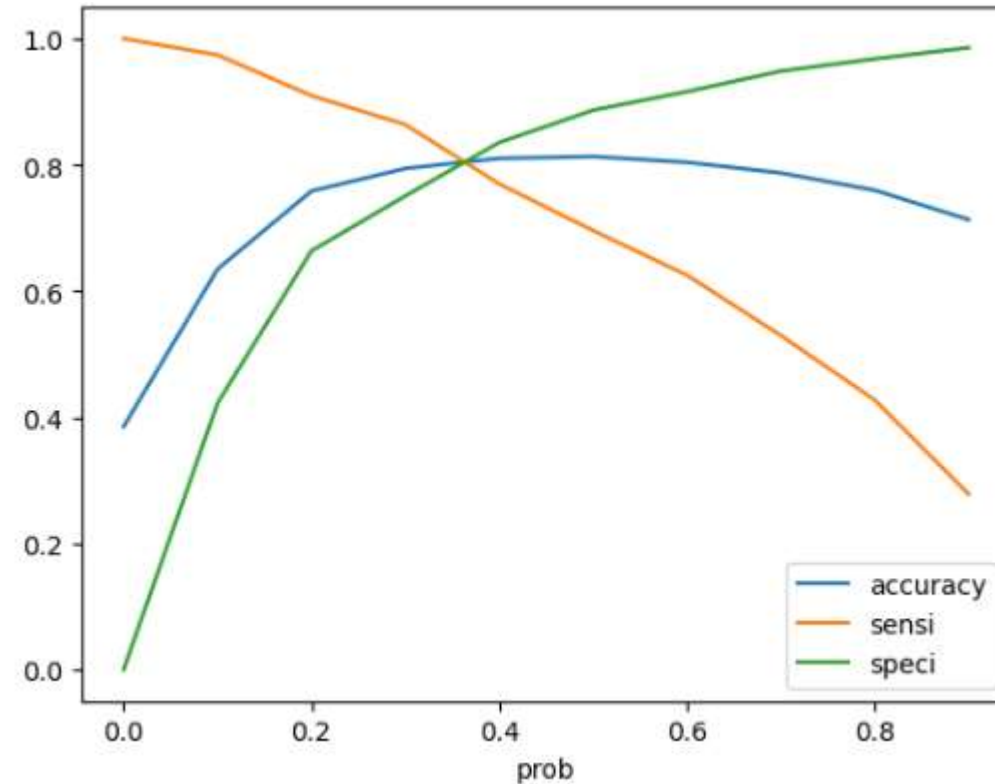
# Model Evaluation



ROC Curve

```
array([[1359, 375],  
       [ 173, 816]],  
      )
```

Confusion  
matrix



Train Data:

1.Accuracy : 79.7 %

2.Sensitivity : 82.2 %

3.Specificity : 78.1 %

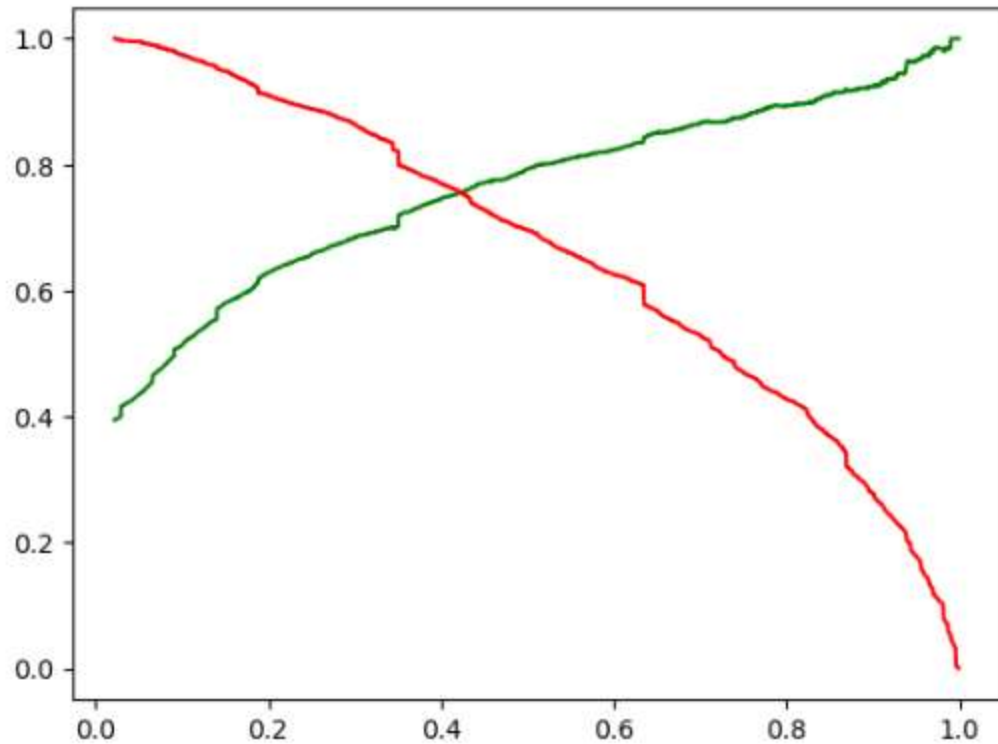
Test Data:

1.Accuracy : 79.8 %

2.Sensitivity : 82.5 %

3.Specificity : 78.4 %

# Model Evaluation – Precision & Recall





# Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 79.7%, 82.2% and 78.1% which are approximately closer to the respective values calculated using trained set
- Accuracy, Sensitivity and Specificity values of test set are around 79.8%, 82.5% and 78.4% which are approximately closer to the respective values calculated using test set.
- Overall the model is a fit model.
- The top 3 variables that contribute for lead getting converted in the model are :

## Important Features from the final model:

```
res.params.sort_values(ascending=False)
```

Lead Origin_Lead Add Form	4.835271
Total Time Spent on Website	4.566509
What is your current occupation_Working Professional	2.409864
Lead Source_Olark Chat	1.413681
Last Activity_SMS Sent	1.167273