Lead Score Case Study Presentation

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 conversion rate at X education is around 30%.

Business Goal:

X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

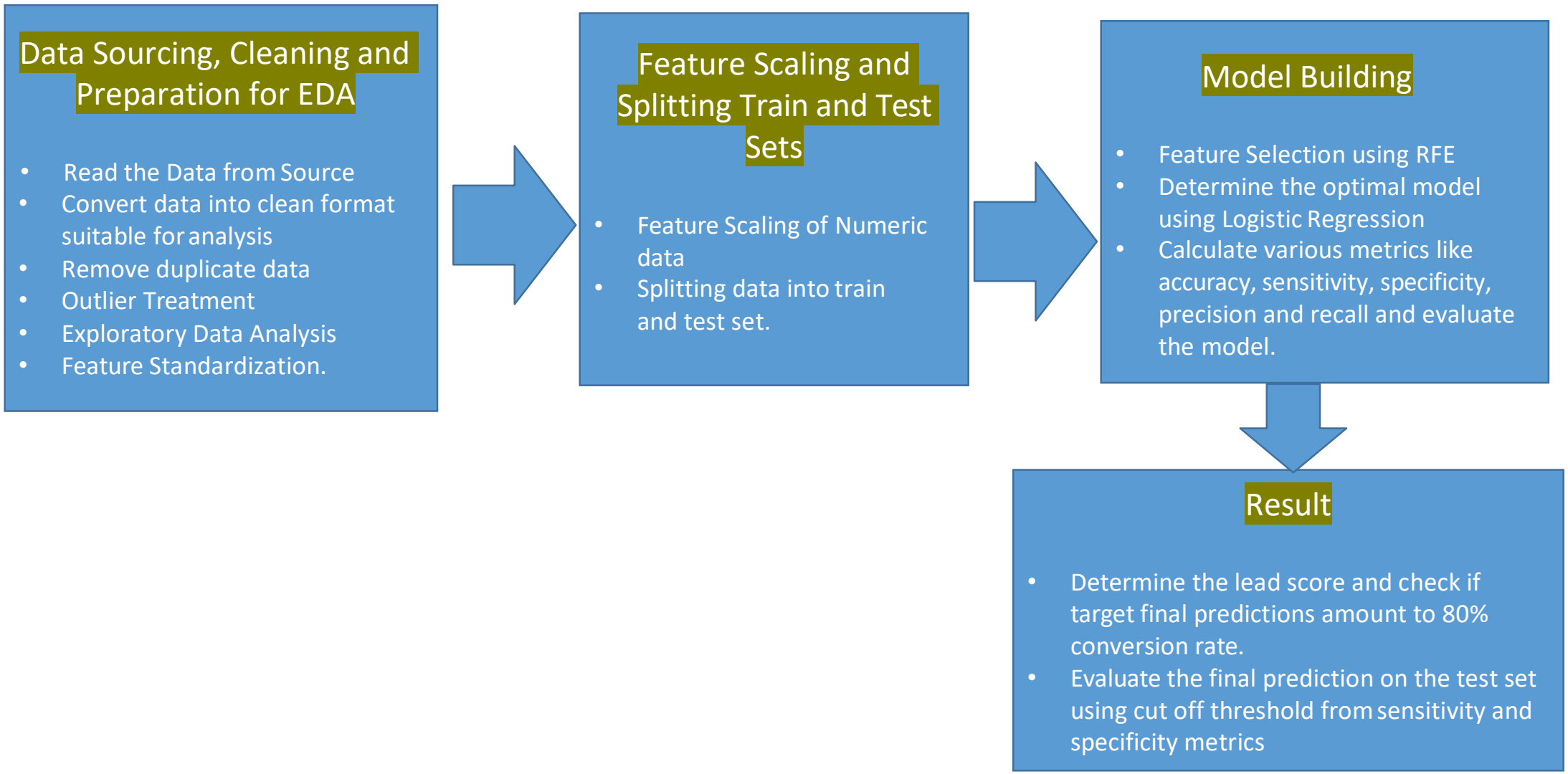
The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

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

Strategy:

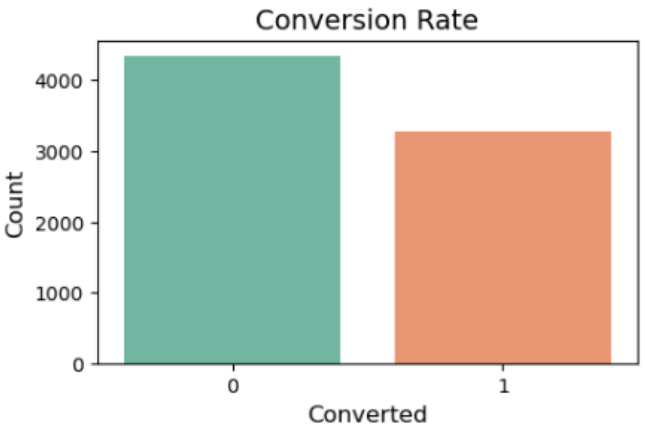
1. Source the data for analysis
2. Clean and prepare the data
3. Exploratory Data Analysis.
4. Feature Scaling
5. Splitting the data into Test and Train dataset.
6. Building a logistic Regression model and calculate Lead Score.
7. Evaluating the model by using different metrics - Specificity and Sensitivity or Precision and Recall.
8. Applying the best model in Test data based on the Sensitivity and Specificity Metrics

Problem solving methodology:



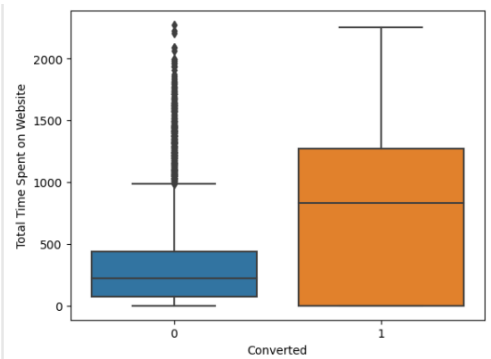
Exploratory Data Analysis:

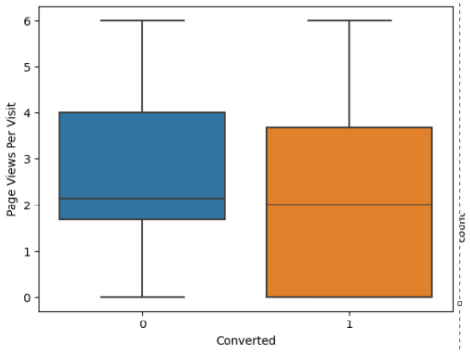
- We have around 43% Conversion rate in Total.

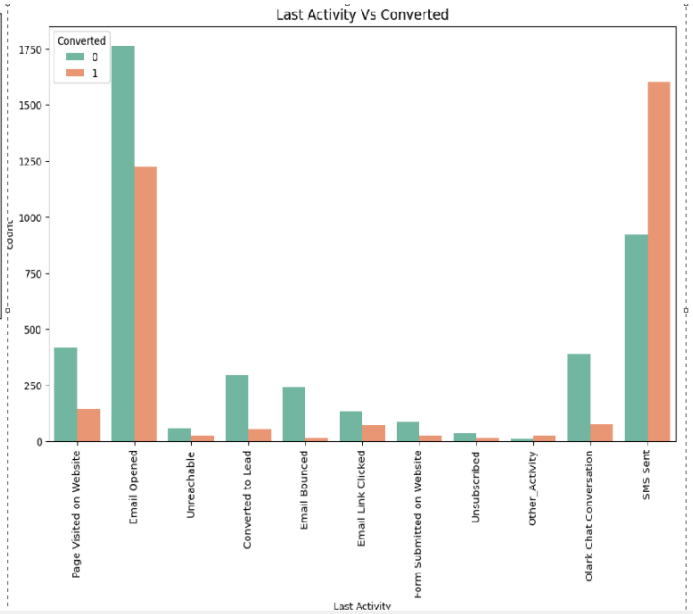


.

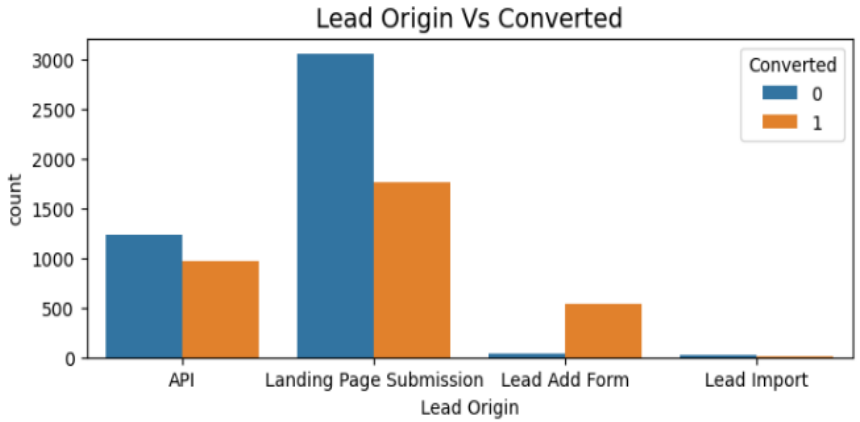
- The conversion rates were high for Page Views per visit, Total Time Spent on Website and Last Activity.



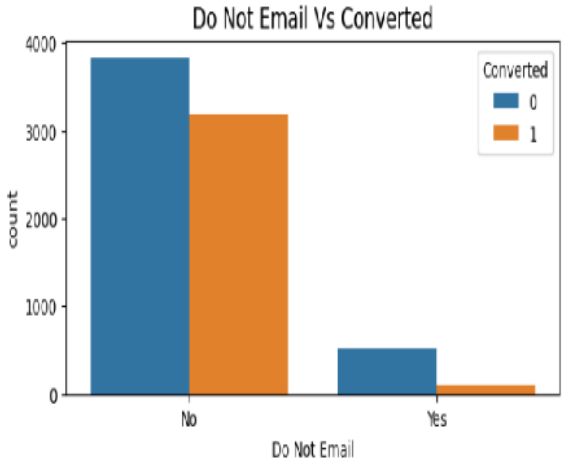


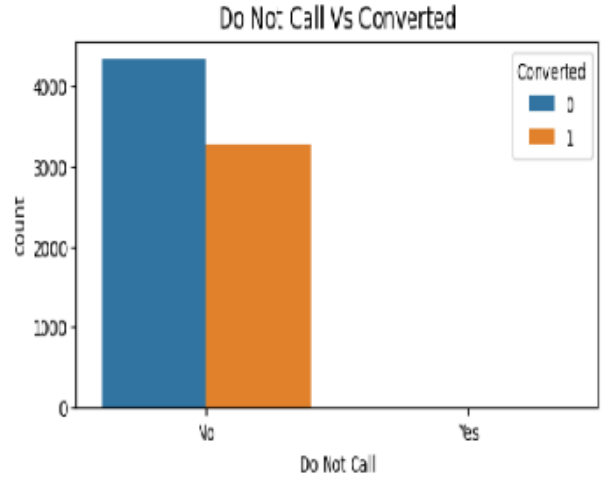


- In Lead Origin, maximum conversion happened from Landing Page Submission

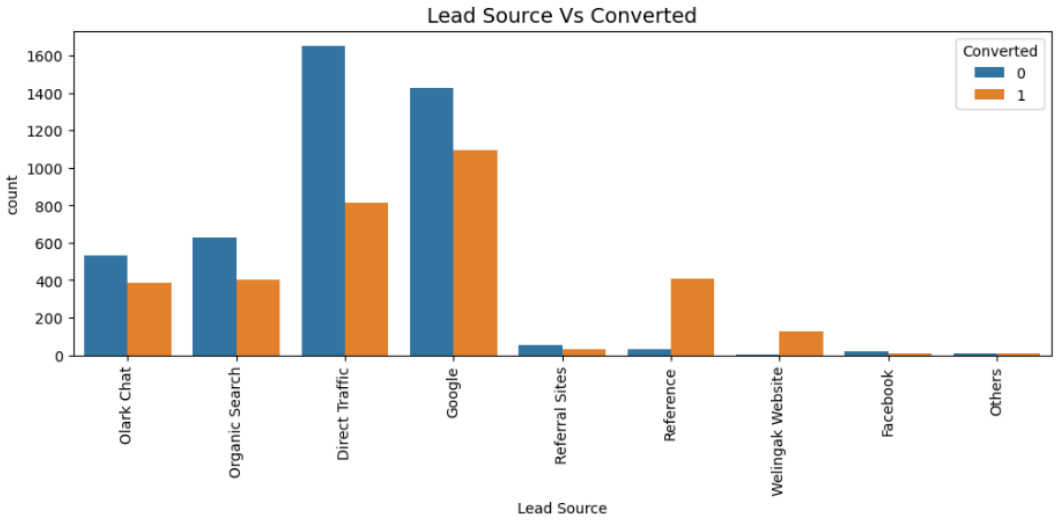


- Major conversion has happened from without Email sent and Without Calls made

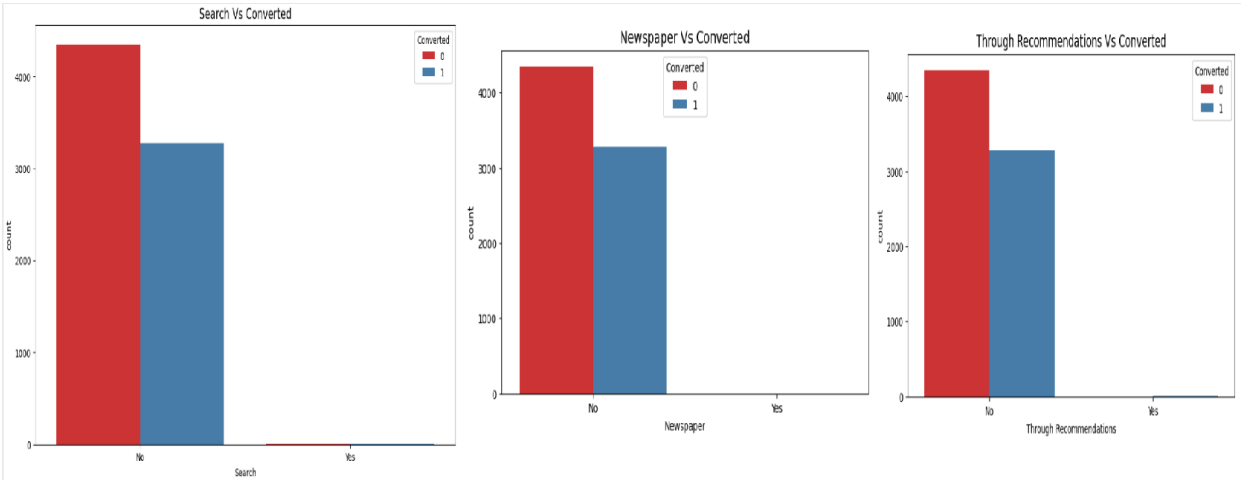




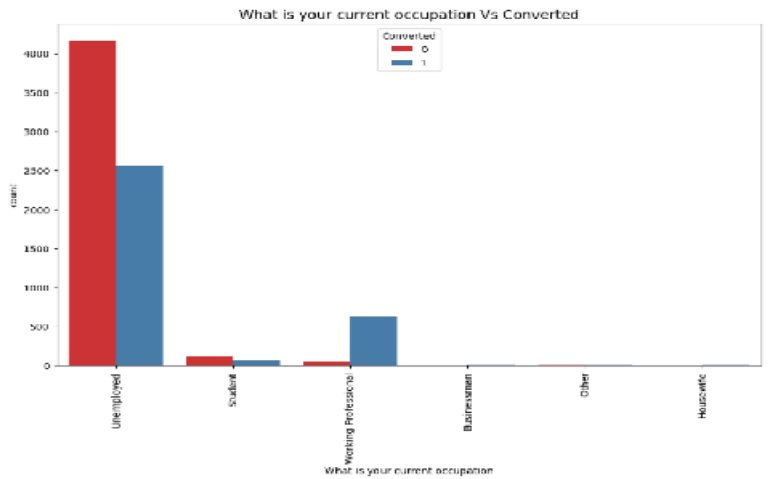
- Major conversion in the Lead Source is from Google



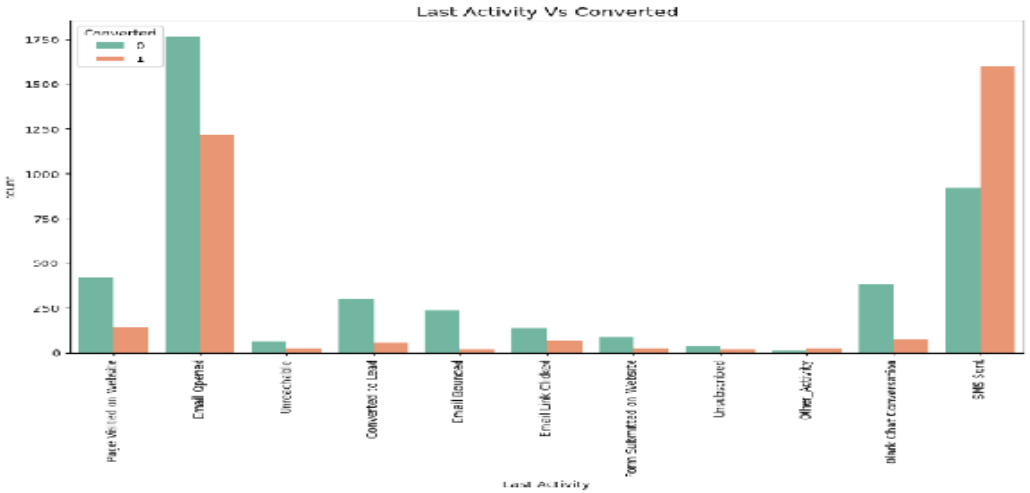
- Not much impact on conversion rates through Search, digital advertisements and through recommendations



- More conversion happened with people who are unemployed



- Last Activity value of SMS Sent had more conversion.



Variables Impacting the Conversion Rate:

Do Not Email

Total Time Spent on Website

Lead Origin_Landing Page Submission

Lead Origin_Lead Import

Lead Source_Olark Chat

Lead Source_Reference

Lead Source_Welingak Website

Last Activity_Olark Chat Conversation

Last Activity_Other_Activity

Last Activity_SMS Sent

Specialization_Hospitality Management

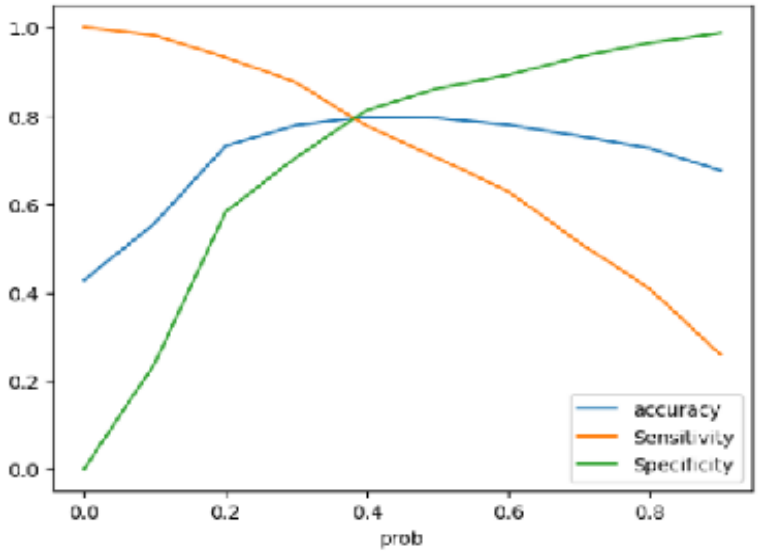
What is your current occupation_Working Professional

Last Notable Activity_Modified

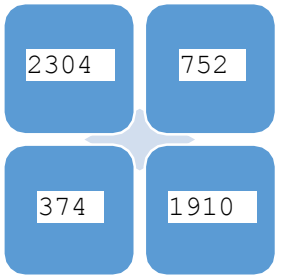
Last Notable Activity_Page Visited on Website

Model Evaluation - Sensitivity and Specificity on Train Data Set:

- The graph depicts an optimal cut off of 0.36 based on Accuracy, Sensitivity and Specificity

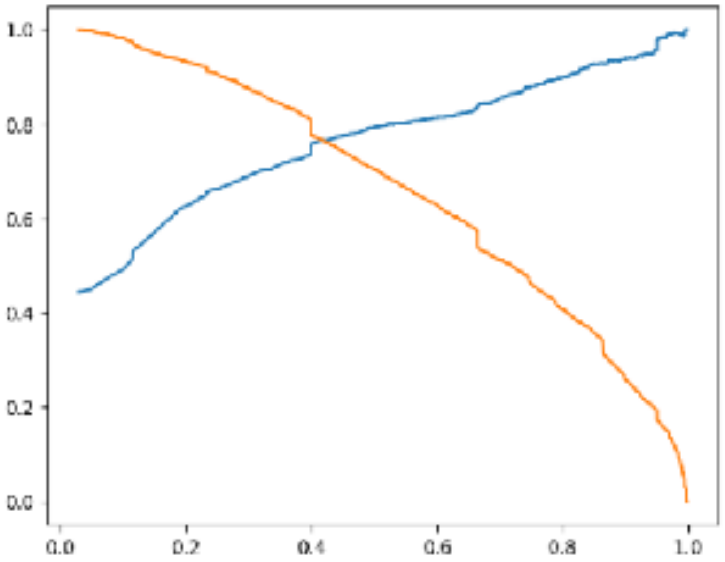


- Confusion Matrix



- Accuracy – 78.9%
- Sensitivity – 83.6 %
- Specificity – 75.3 %
- False Positive Rate – 24.6 %
- Positive Predictive Value – 71.8 %
- Positive Predictive Value – 86.1%

- The graph depicts an optimal cut off of 0.43 based on Precision and Recall



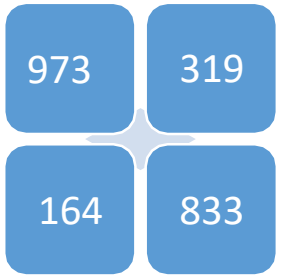
- Confusion Matrix

2633	423
676	1608

- Precision – 79.1 %
- Recall – 70.4 %

Model Evaluation – Sensitivity and Specificity on Test Dataset:

- Confusion Matrix



- Accuracy – 78.8 %
- Sensitivity – 83.5 %
- Specificity – 75.3 %

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 78.8%, 83.5% and 75.3% 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
    1. Lead Source_Welingak Website
    2. Lead Source_Reference
    3. What is your current occupation_Working Professional
- Hence overall this model seems to be good.