LEAD SCORING CASE STUDY

Submitted by :
Sanjana Dutta
Atul Matvar
Geet Malviya

LEAD SCORE CASE STUDY FOR X EDUCATION

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.

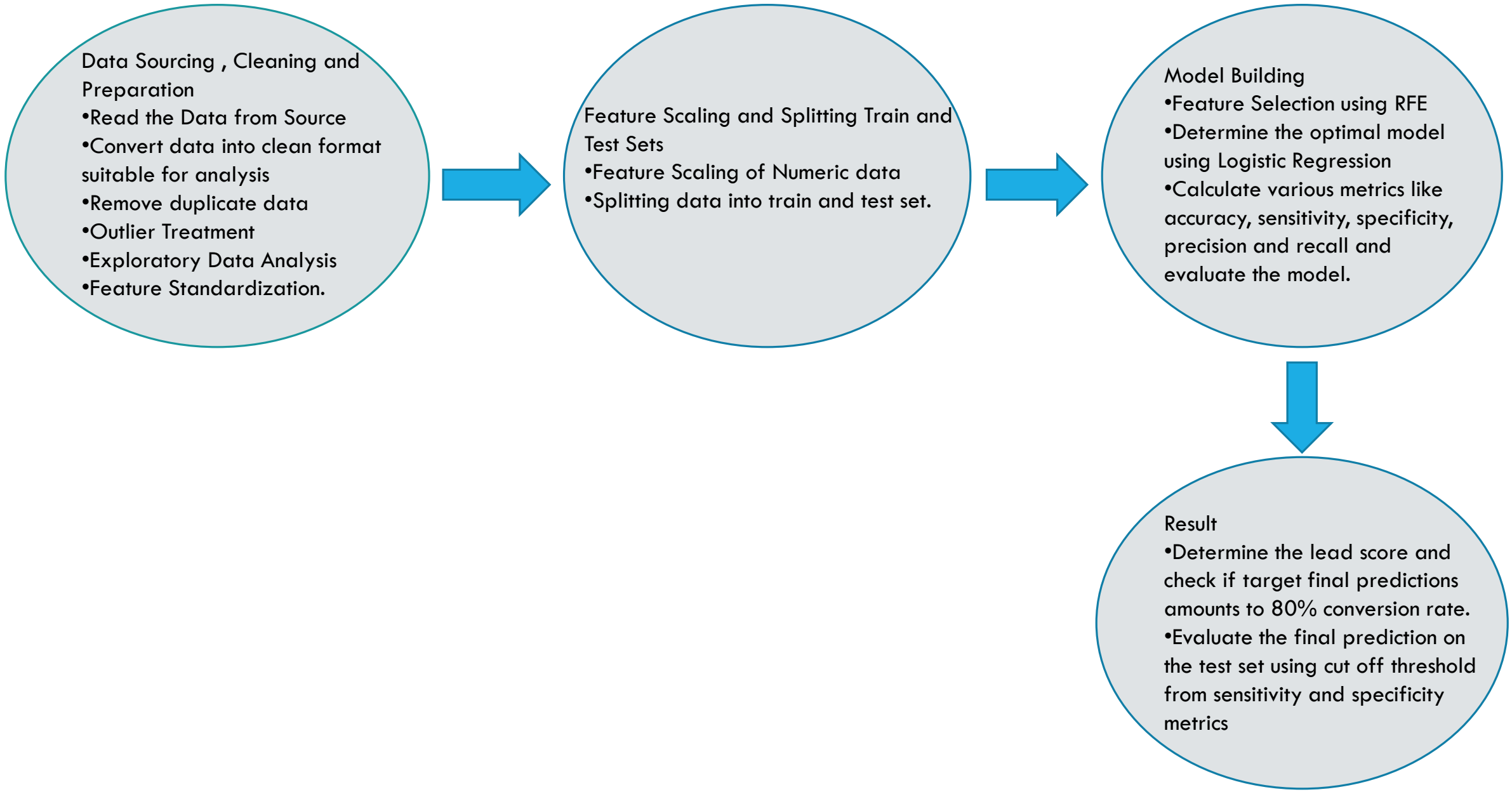
The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance.

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

STRATEGY

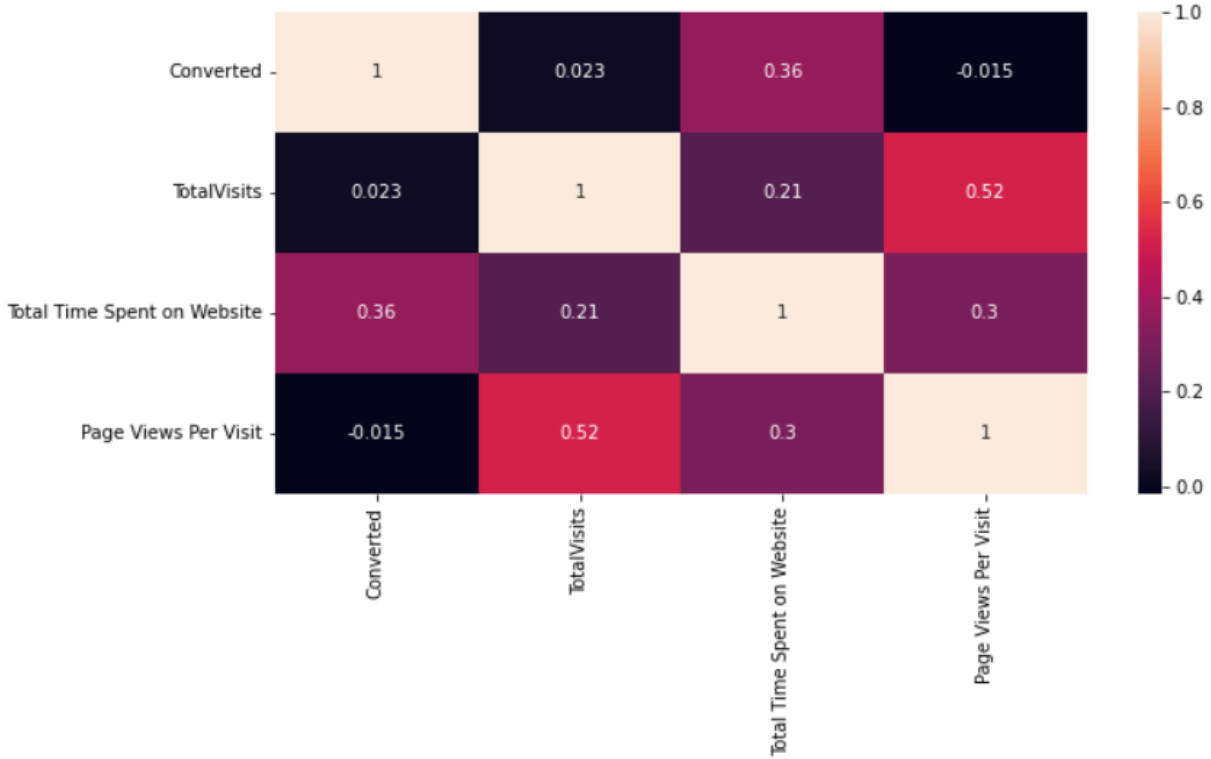
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics -Specificity and Sensitivity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

PROBLEM SOLVING METHODOLOGY

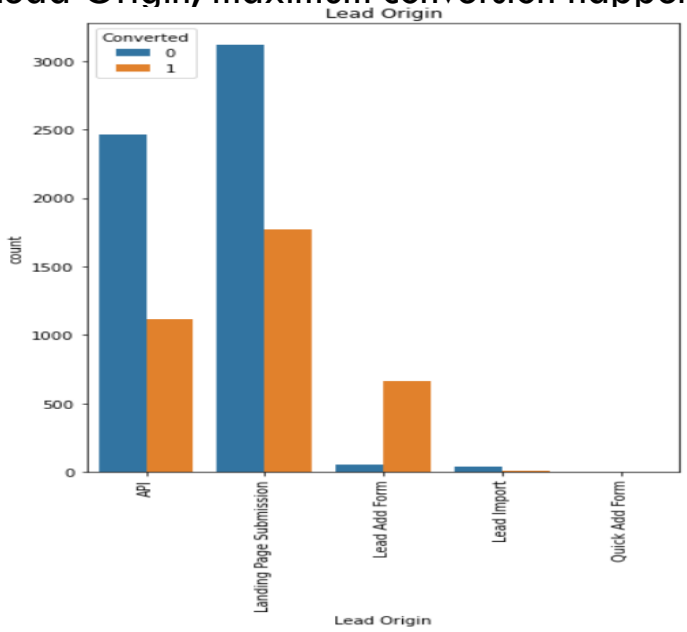


EXPLORATORY DATA ANALYSIS

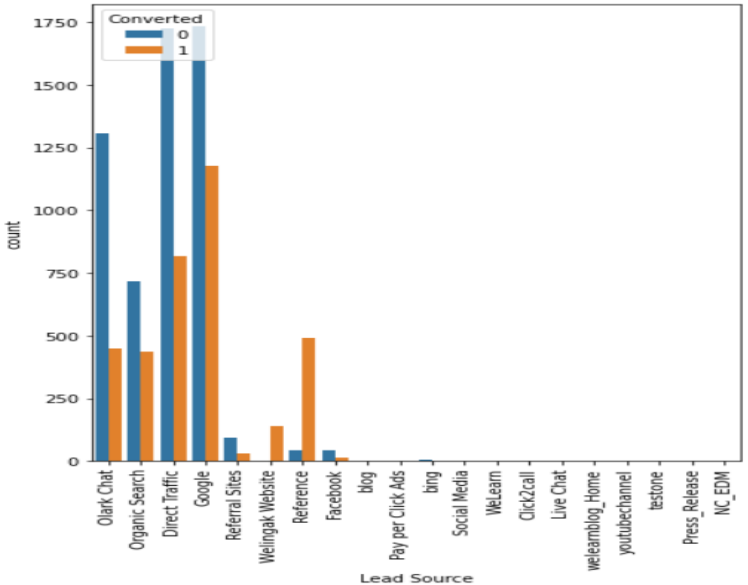
Conversion rate between Total Visits, Total time spent on website and Page view per visit(multivariate analysis heatmap).



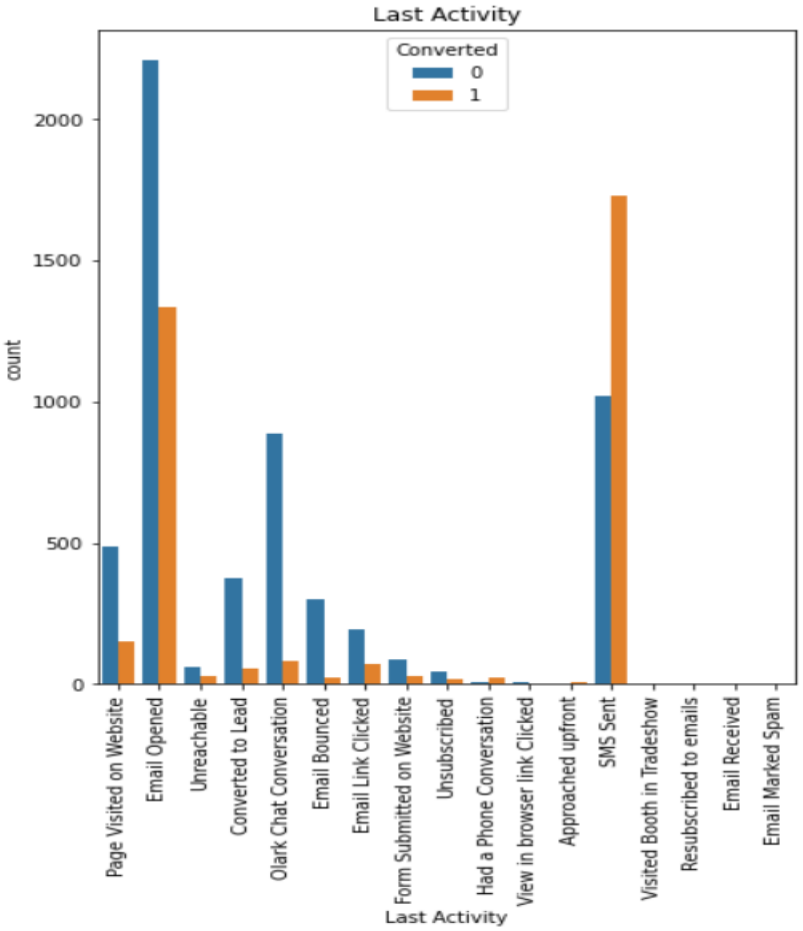
In Lead Origin, maximum conversion happened from Landing Page Submission:



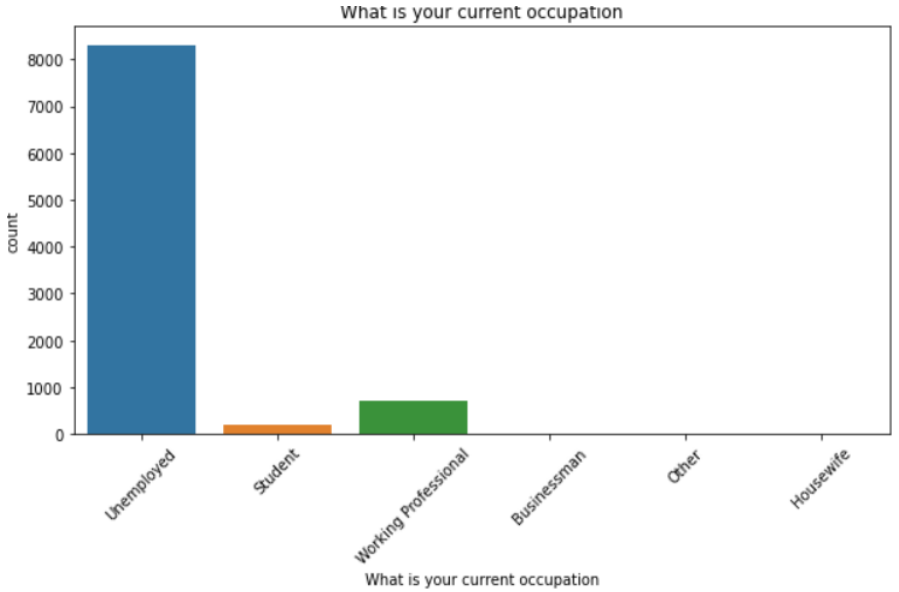
In Lead source maximum conversion happened from Direct Traffic:



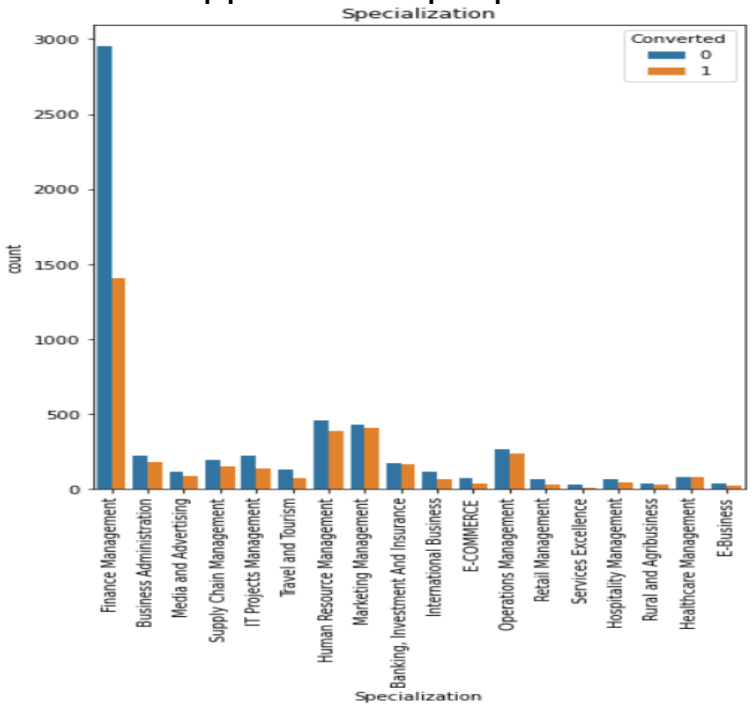
In Last Activity, maximum conversion happened from Email Opened:



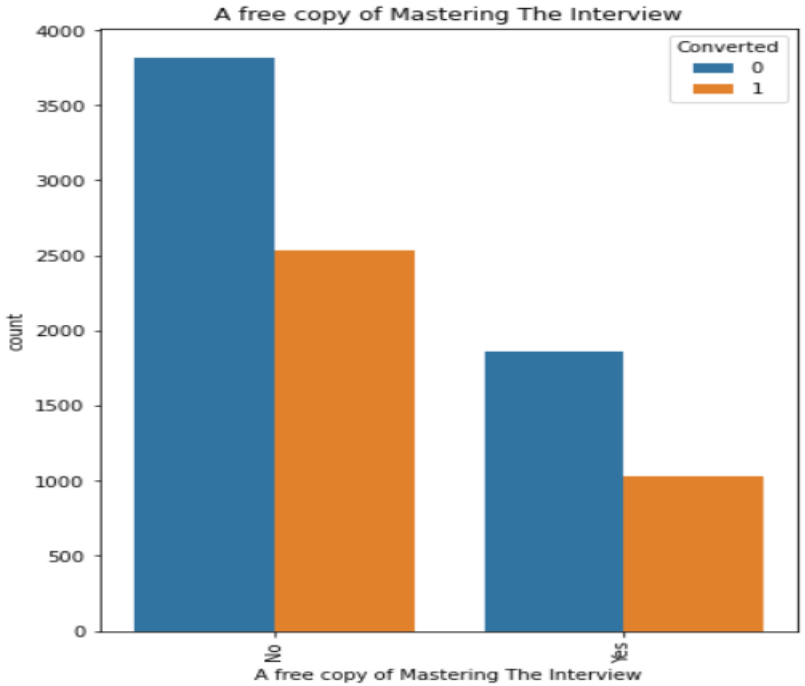
More conversion happened with people who are unemployed:



Most Conversion happened with people from Finance Management:



Converted leads didn't demand for a free copy of Mastering the Interview:



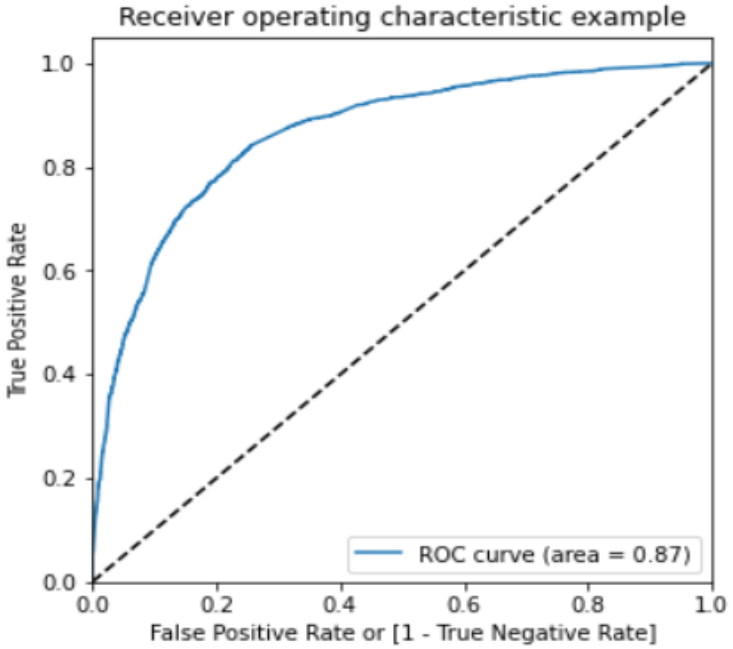
Few inferences derived from the Analysis:

- Landing Page Submission & API has highest count respectively
- Most of the people reached out via Google followed by Direct Traffic source
- Most of peoples choose Don't Email option
- Most of peoples choose Don't Call option
- Most of peoples Open the Email and Send the SMSs
- Leads who are management profession in any fields are more likely convert in actual lead
- Most of the peoples are "Unemployed"
- Most of peoples haven't search about this ads.
- All the peoples haven't seen about this ads on Magazine
- Most of peoples haven't seen any adds on News paper Article.
- Most of peoples haven't seen any adds on X Education Forums
- Most of peoples haven't seen any adds on Digital Advertisement
- Most of peoples haven't come Through Recommendations
- No one receive any update About this Courses
- No one wants any Update me on Supply Chain Content
- No Get updates on DM Content
- No agreed to pay amount through cheque
- Some of peoples wants A free copy of Mastering The Interview from X education
- Counts of activities like Modified, Email Opened and Sms Sent are high compared to other activities.

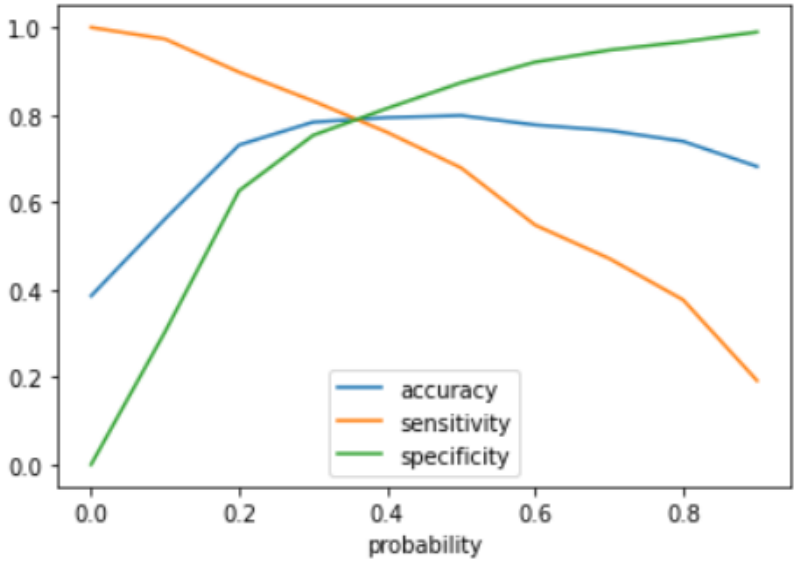
MODEL EVALUATION - SENSITIVITY AND SPECIFICITY ON TRAIN SET

The ROC Curve should be a value close to 1.

We are getting a good value of 0.87 indicating a good predictive model.



The graph depicts an optimal cut off of 0.37 based on Accuracy, Sensitivity and Specificity.



Confusion Matrix:

```
[3700,  841]
[ 684, 2167]
```

- **Observation of Train Set**
- Cutoff = 0.4 Accuracy = 79.37%
- Sensitivity = 76.00% Specificity = 81.48%

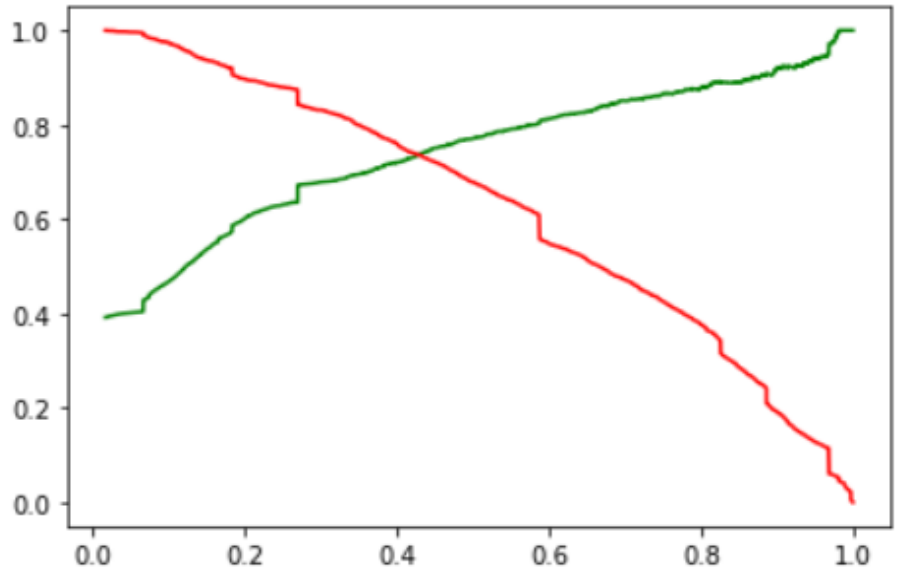
MODEL EVALUATION - SENSITIVITY AND SPECIFICITY ON TEST DATASET

Confusion Matrix:

```
[935, 203]
[184, 526]
```

- Accuracy = 79.06%
- Sensitivity = 74.08%
- Specificity = 82.16%

The graph depicts an optimal cut off of 0.42 based on Precision and Recall



- Precision = 72.15%
- Recall = 74.08%

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%, 74% and 82% 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 76% (in train set) and 74% (in test set)
- The top 3 variables that contribute for lead getting converted in the model are
  - Total Visits
  - Lead Add Form from Lead
  - Had a phone conversation from Last notable activity

Hence overall this model seems to be good.