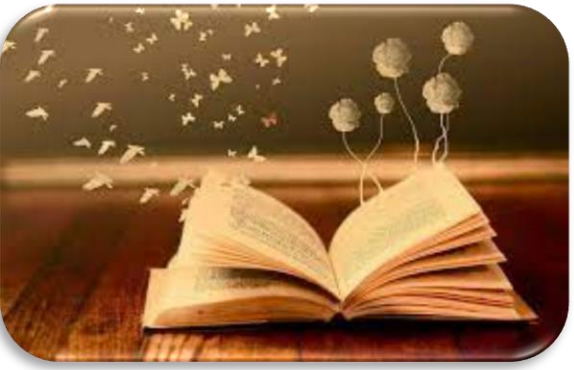
LEAD SCORING CASE STUDY

BY
DORIA SHARMA
&
VISHNU MENON
DSC-43

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

A higher score would mean that the lead is hot, i.e. most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted

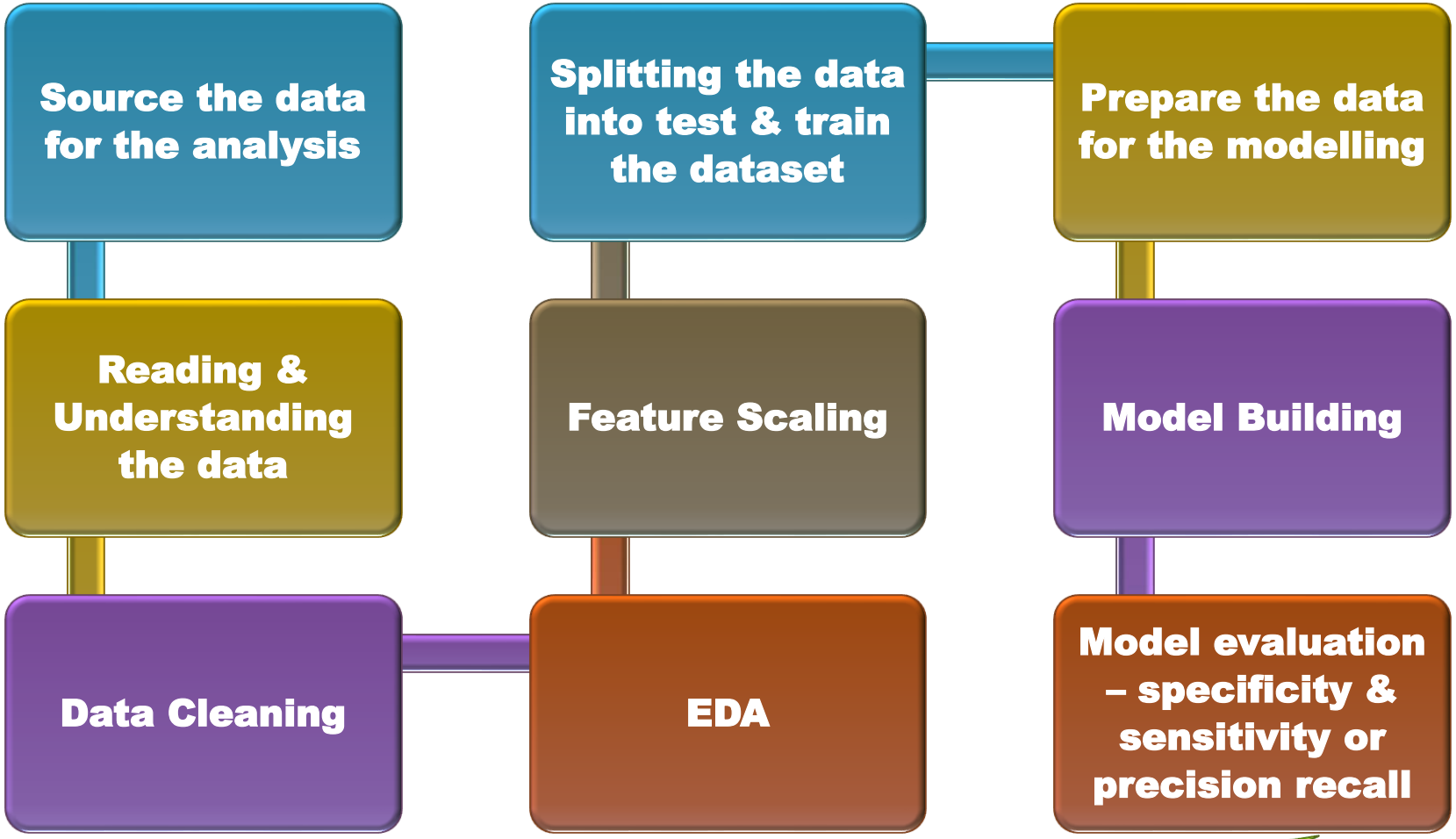
PROBLEM STATEMENT



An education company named X Education sells online courses to industry professionals

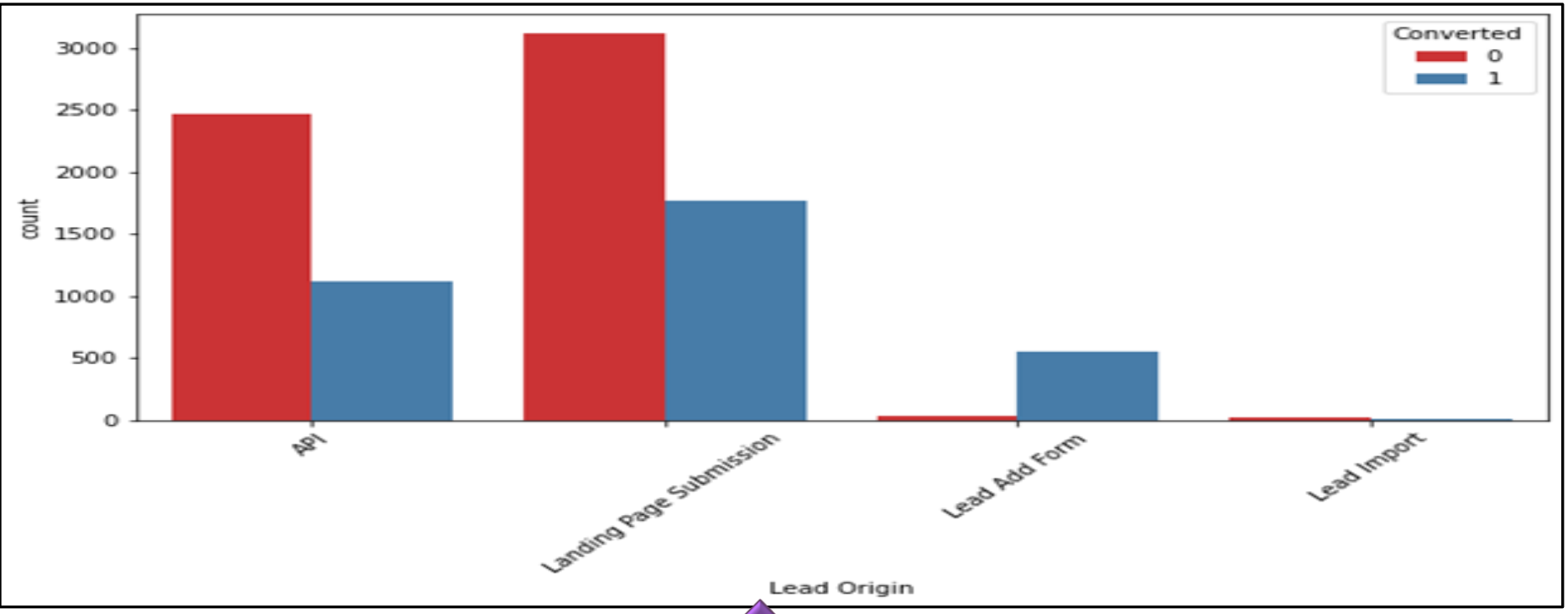
Now although the X Education gets lots of leads, its lead conversion rate is very poor of about 30%

APPROACH

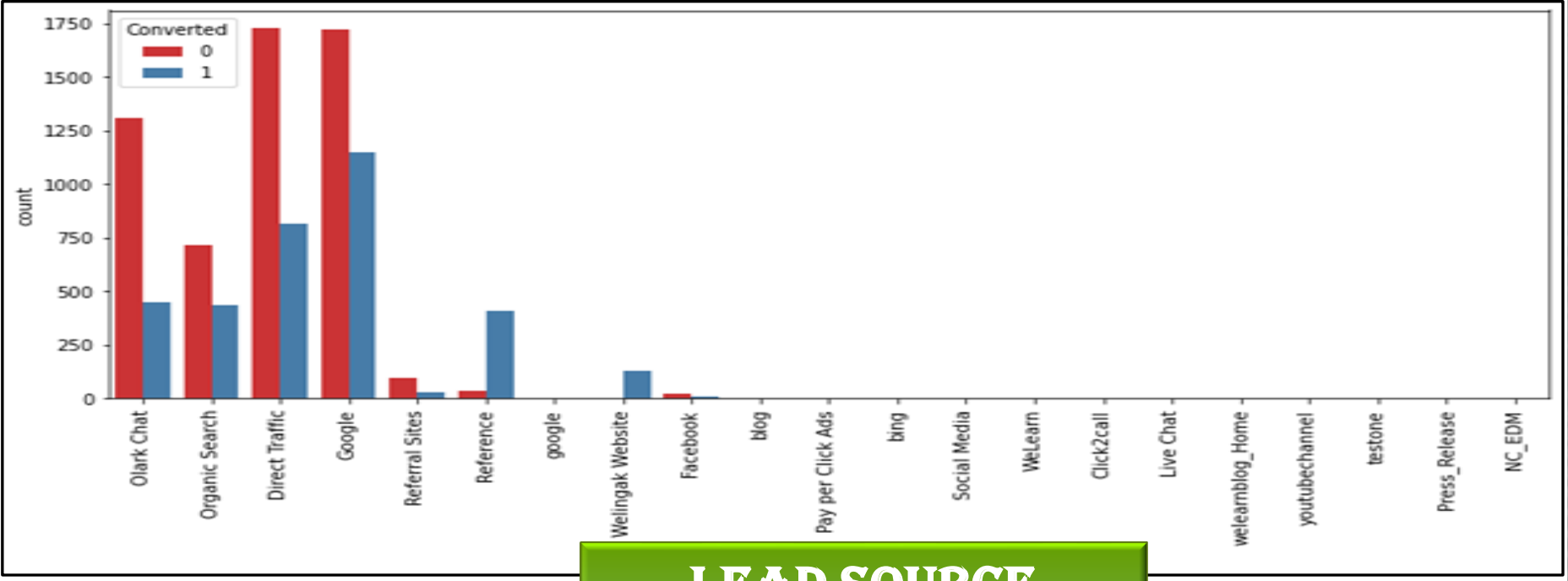

Source the data for the analysis
Splitting the data into test & train the dataset
Prepare the data for the modelling
Reading & Understanding the data
Feature Scaling
Model Building
Data Cleaning
EDA
Model evaluation – specificity & sensitivity or precision recall
Making predictions on the test sets.

Exploratory Data Analysis

LEAD ORIGIN

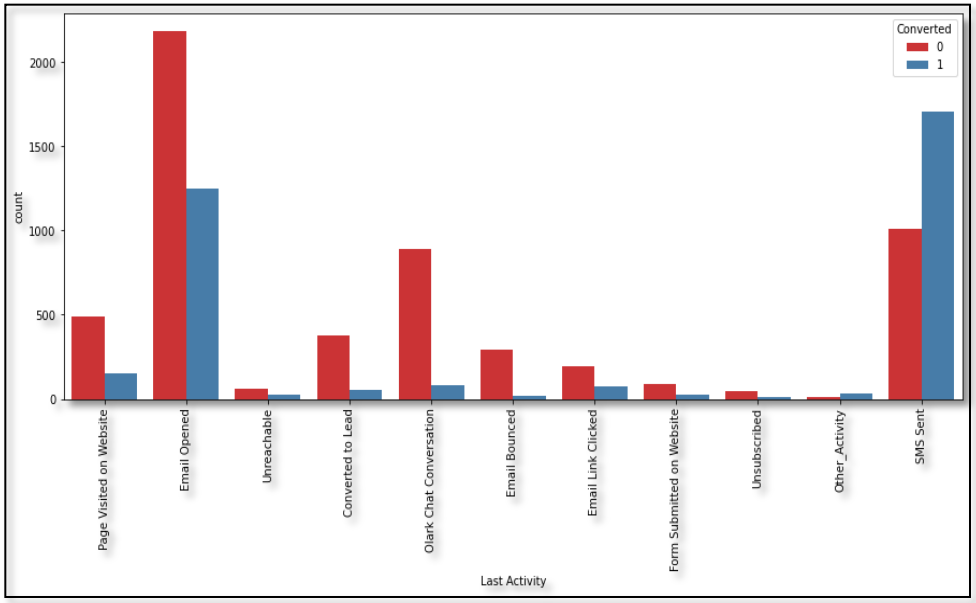


To improve overall lead conversion rate, we need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



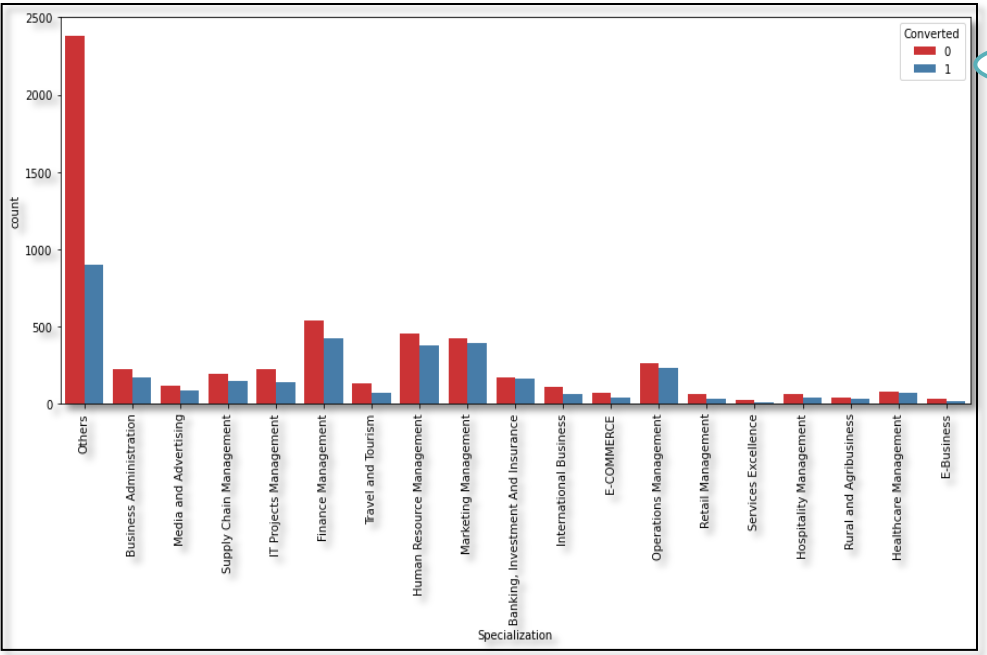
LEAD SOURCE

To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.



LAST ACTIVITY

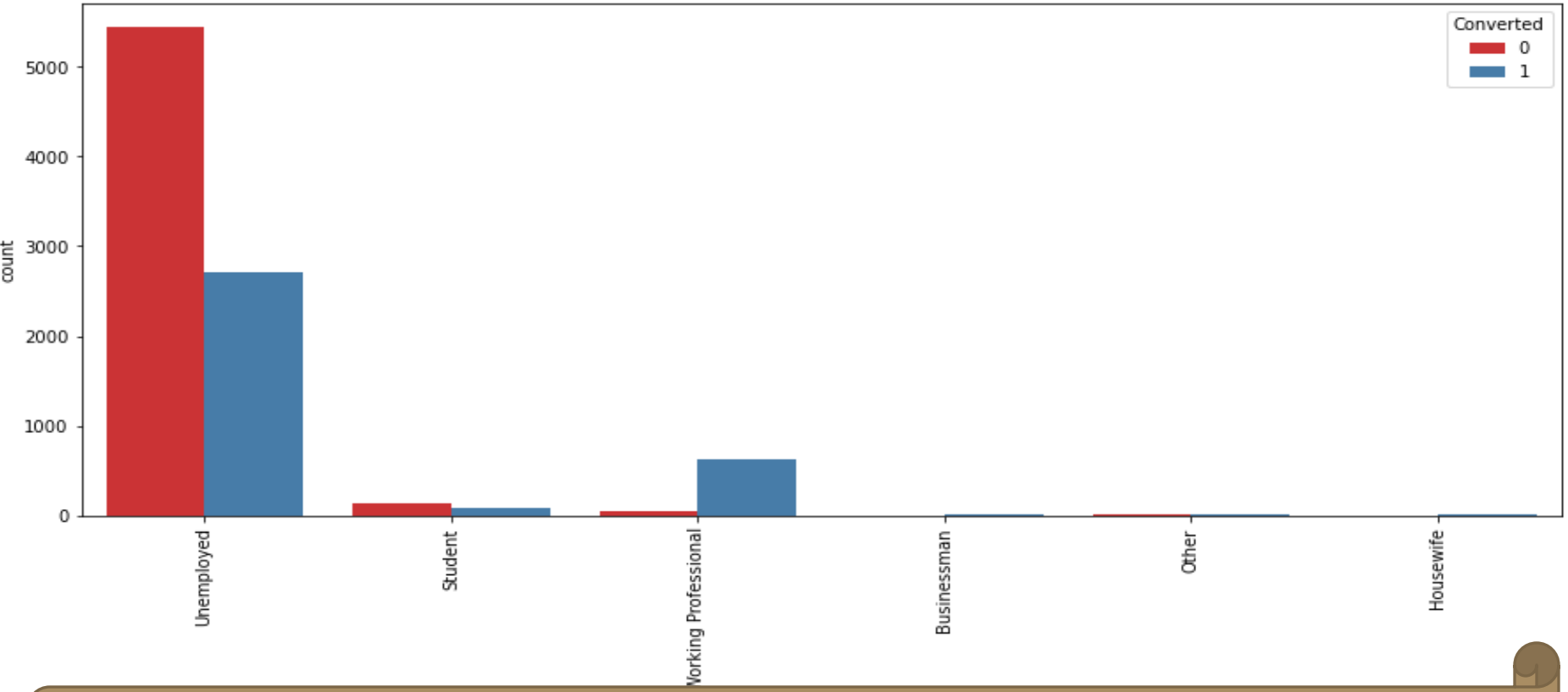
Most of the lead have their Email opened as their last activity. Conversion rate for leads with last activity as SMS Sent is almost 60%.



SPECIALIZATION

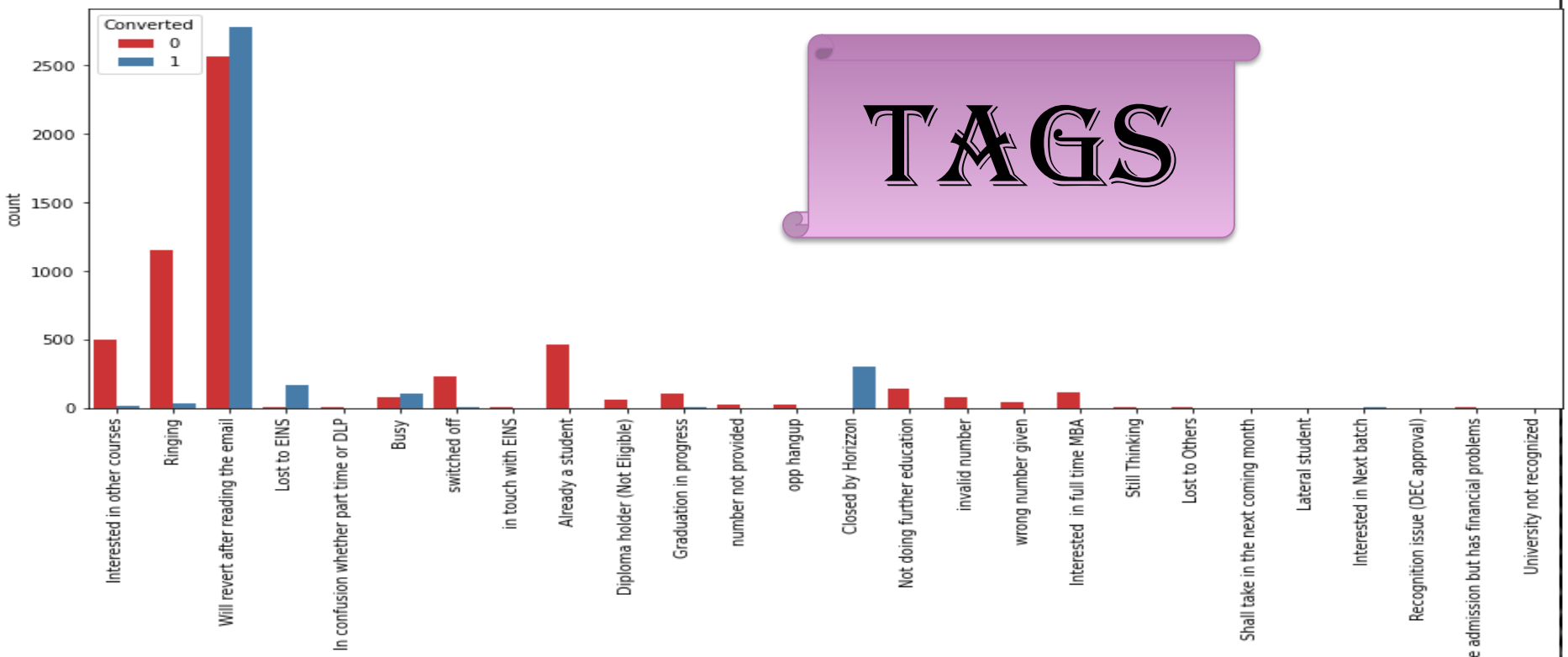
Focus should be more on the Specialization with high conversion rate

WHAT IS YOUR CURRENT OCCUPATION



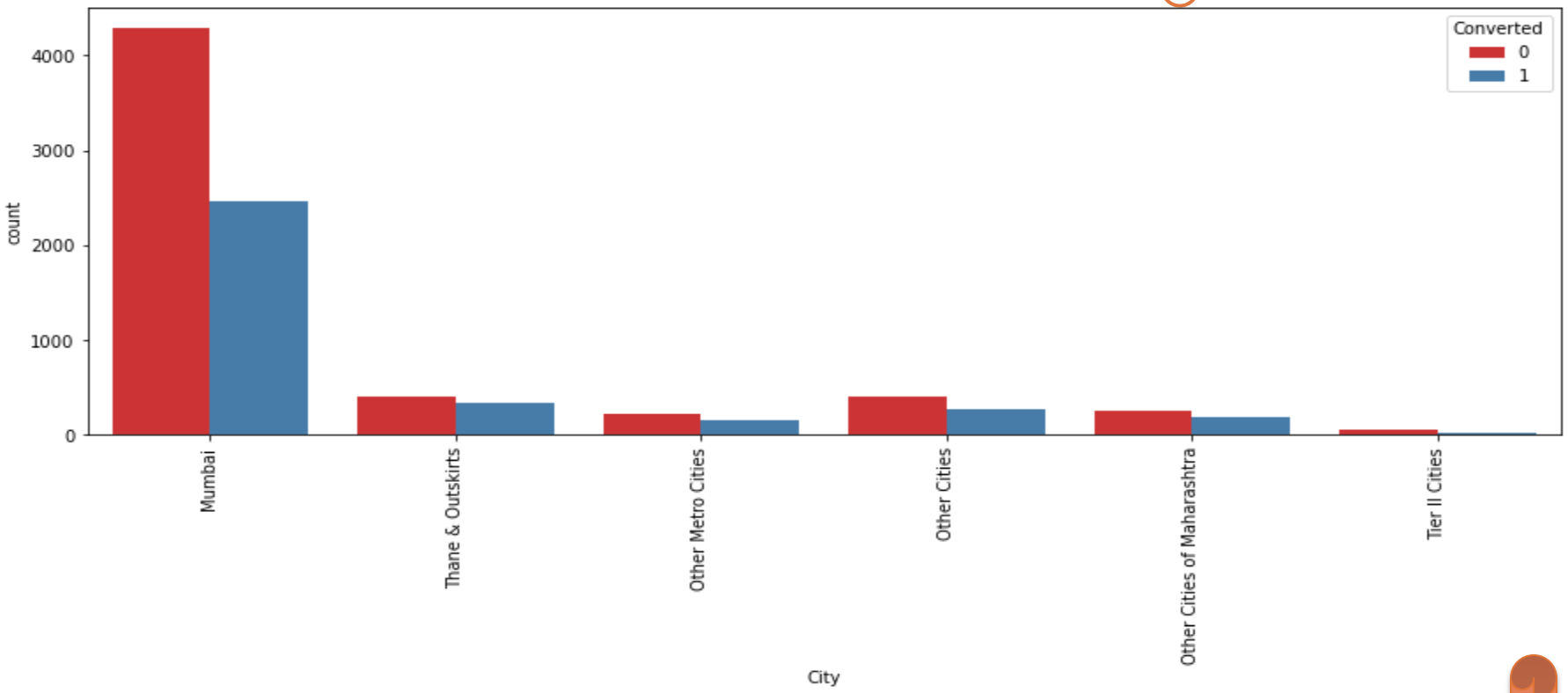
Working Professionals going for the course have high chances of joining it.
Unemployed leads are the most in numbers but has around 30-35% conversion rate.

TAGS



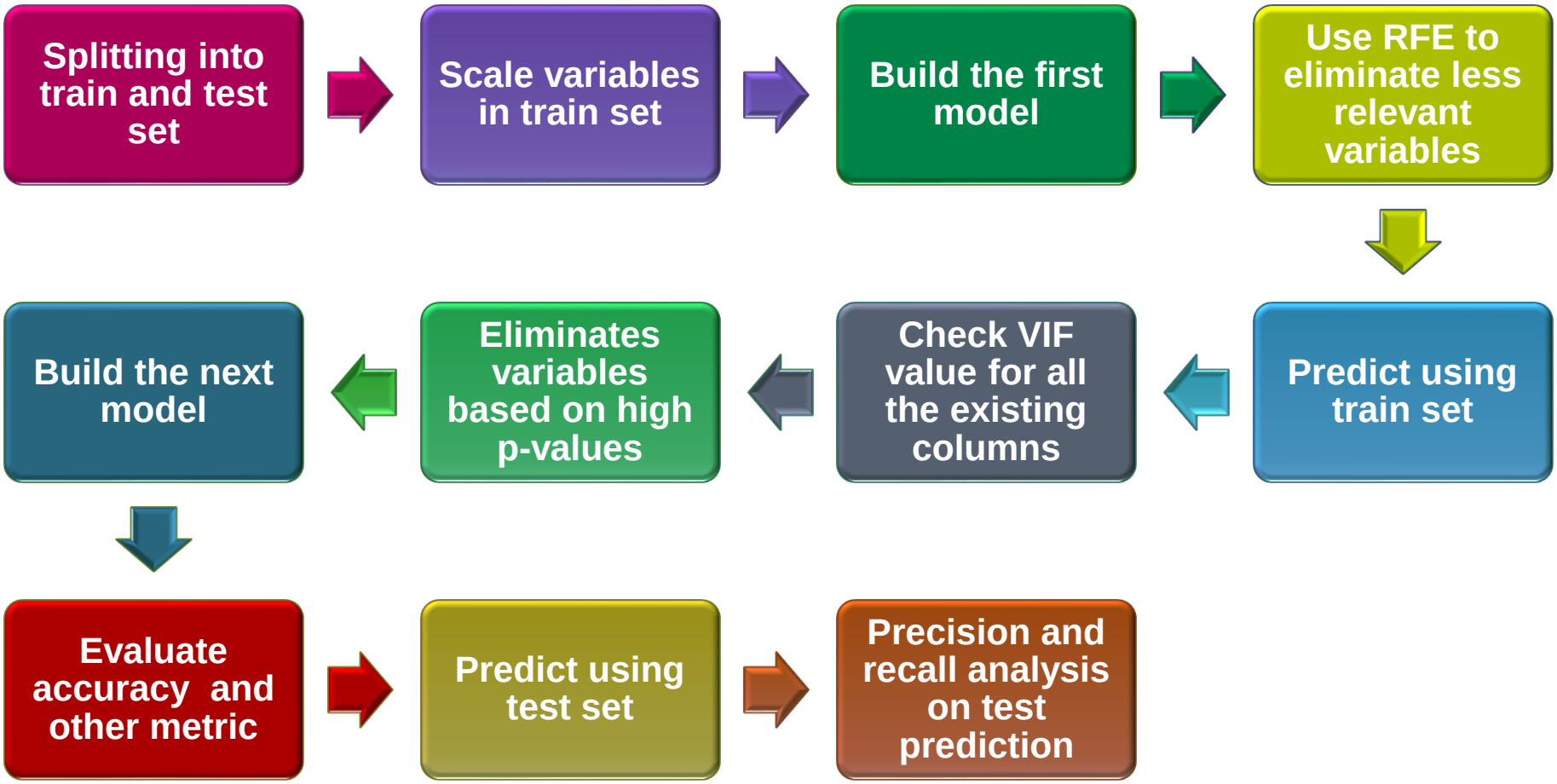
Since this is a column which is generated by the sales team for their analysis , so this is not available for model building . So we will need to remove this column before building the model.

CITY

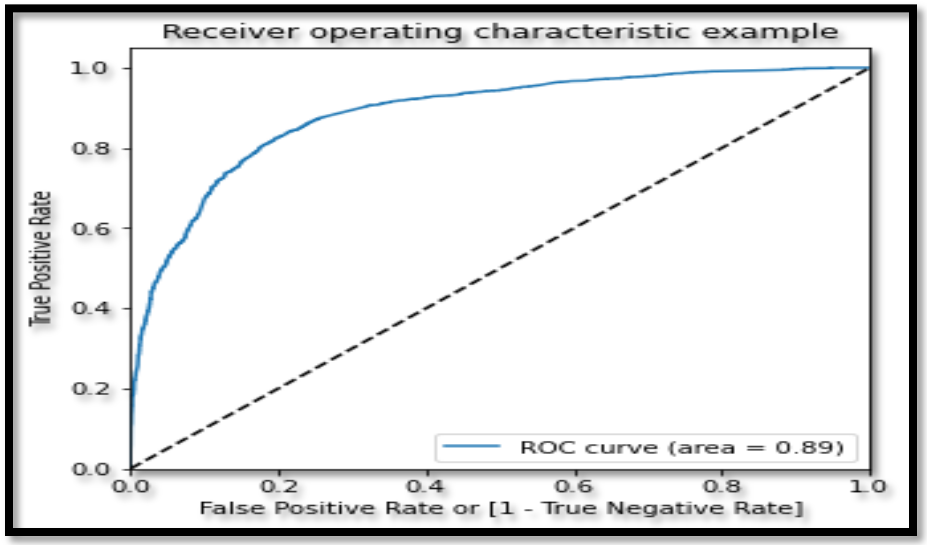


Most leads are from mumbai with around 50% conversion rate.

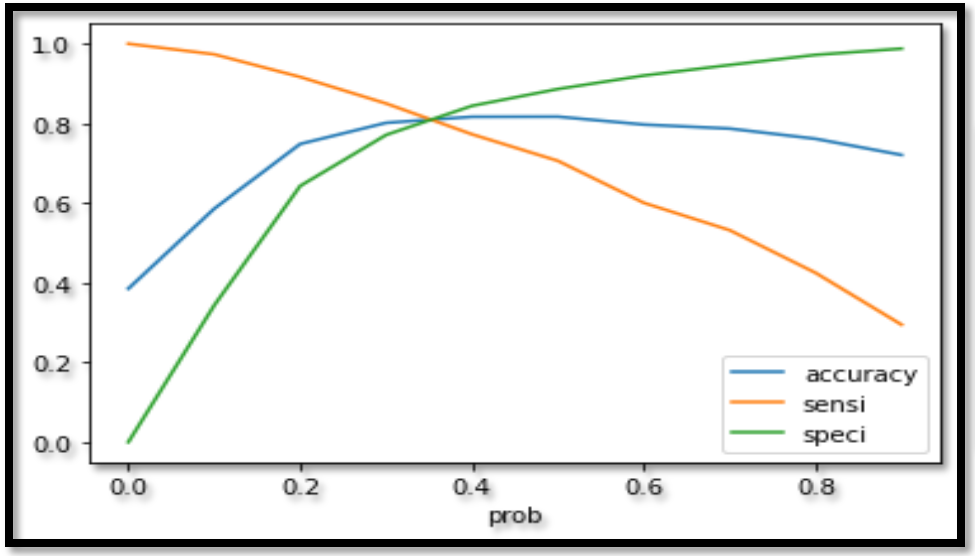
MODEL BUILDING



ROC CURVE & PROBABILITY PREDICTION

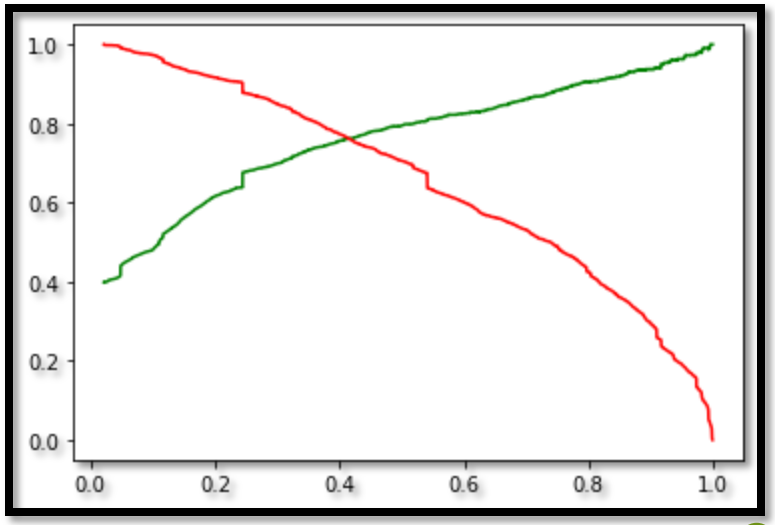


Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.



0.34 is the optimum point to take it as a cutoff probability

MODEL EVALUATION (TRAIN & TEST)



The above graph shows the trade-off between the Precision and Recall

Train Data

Accuracy : 81.0 %
Sensitivity : 81.7 %
Specificity : 80.6 %

Precision – 79.5%
Recall – 70.6%

Test Data

Accuracy : 80.4 %
Sensitivity : 80.4 %
Specificity : 80.5 %

Precision – 70.1%
Recall – 80.4%

Thus we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 80%.

RECOMMENDATION

The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.

The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.

The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.

The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.

The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.

RECOMMENDATION

The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.

The company should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted

The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.

.

.The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.