LEAD SCORING TO IDENTIFY POTENTIAL LEADS AND INCREASE LEAD CONVERSION RATE

LEAD SCORING PROBLEM STATEMENT

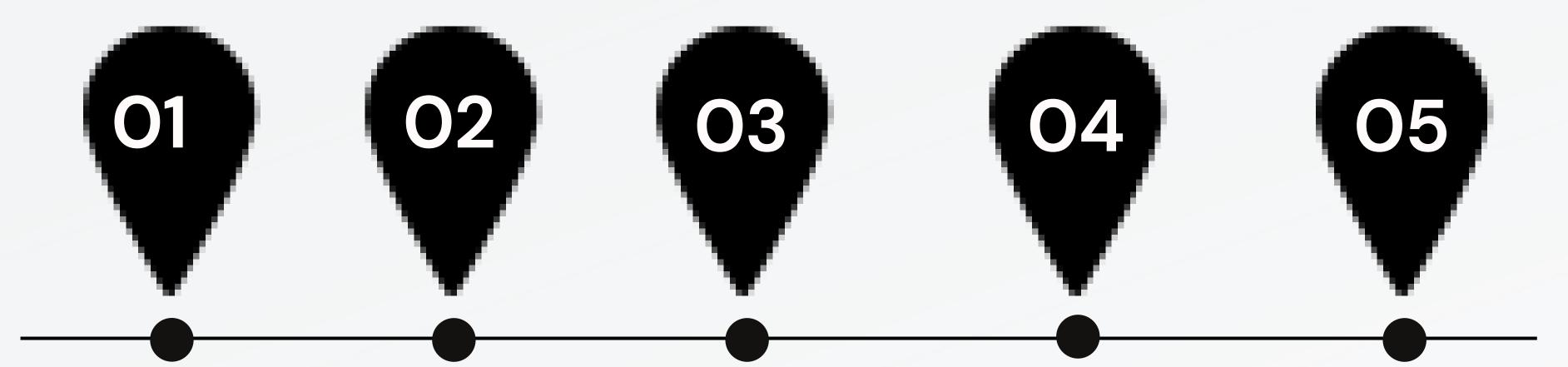


We have received leads dataset from the past with around 9000 data points from X education, consisting of various features such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity etc, the Target variable is 'Converted' feature in the dataset which tells whether a past lead was converted or not.



Business Objective: X Education wants to sort the most promising leads, i.e. the leads that are most likely to convert into paying customers. We need to build a model wherein we will assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The lead conversion rate to be increased to around 80%.

PROBLEM SOLVING STRATEGY



Data importing

Data Inspection

Data Cleaning:

Missing Value
Treatment

EDA
Understanding the
features and
relationships
among them and
Target feature.

Scaling Features

Data Preparation

Building Logistic
Regression model

Assign Lead Score

Test The model on Train Set

Evaluate Model using different metrics

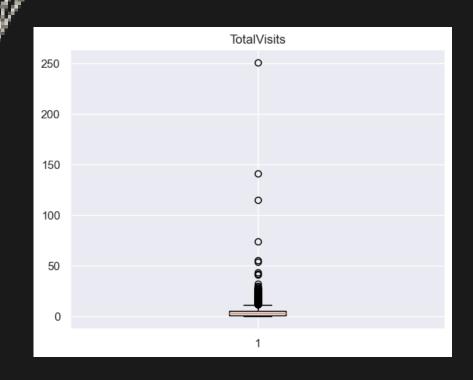
Test The model on Test Set

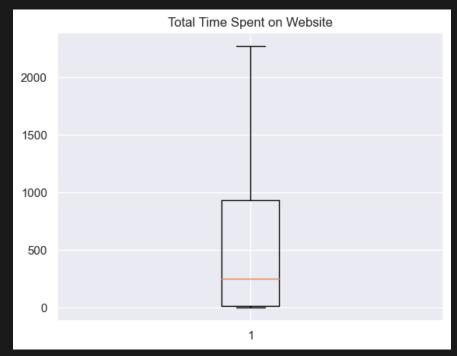
Evaluate Model using different metrics

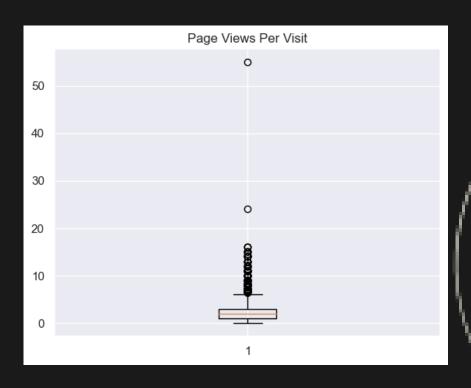
Suggestions and Recommendations

UNIVARIATE ANALYSIS FOR CONTINUOUS COLUMNS

By plotting boxplots, we see there are a few outliers for the 'TotalVisits', 'Page Views Per Visit' columns, although its better to get rid of these outliers, but we see among these points around 40% of leads converted. So we kept the records as is and also removed 'TotalVisits' due to strong relation with other numerical features to have minimum collinearity.

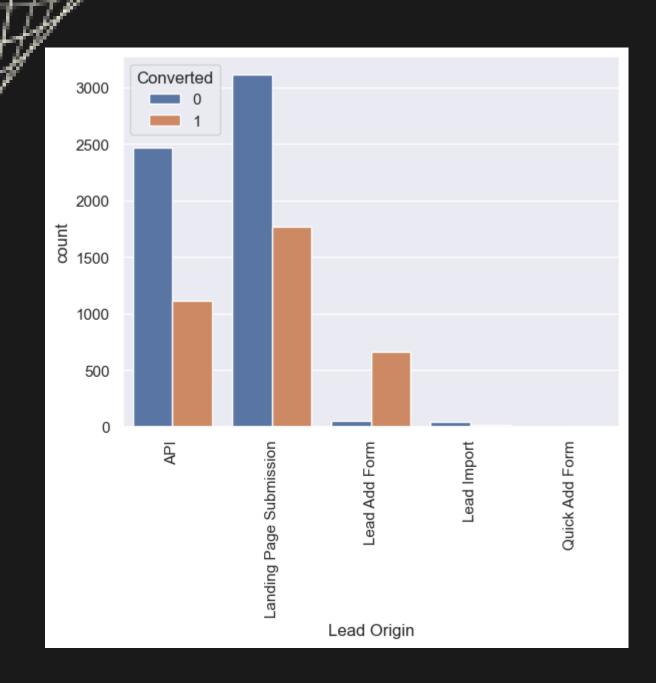


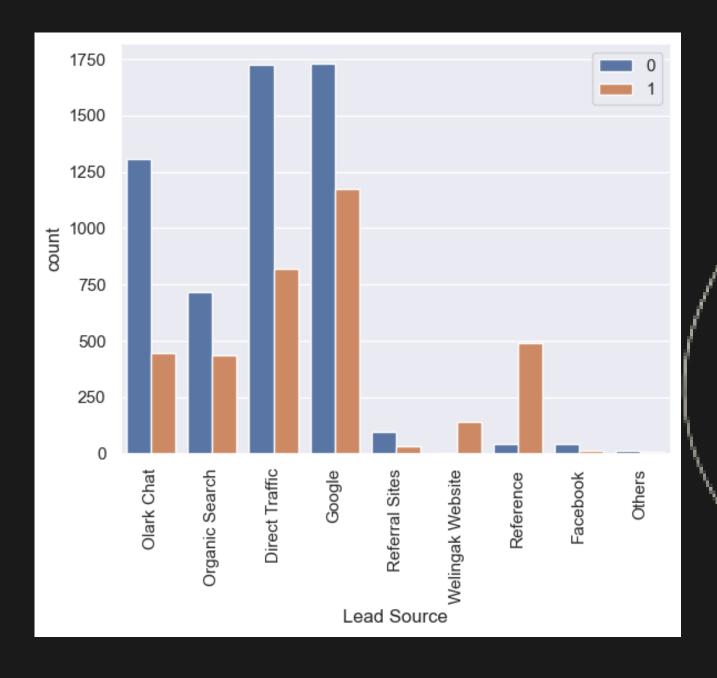




SEGMENTED UNIVARIATE ANALYSIS FOR CATEGORICAL COLUMNS

For categorical columns we have used count plots to understand the frequencies of different categories. Here we have plotted 'Lead Source' and 'Lead Origin'





Understanding Lead Source and Lead Origin Better

Lead Origin

Its clear that most of the converts are coming thru API and Landing Page Submissions.

The conversion rate is comparatively higher for Lead add form category.

So, the company should ensure to increase efficacy by assisting leads filling up the form in case of queries such that the conversion rate increases.

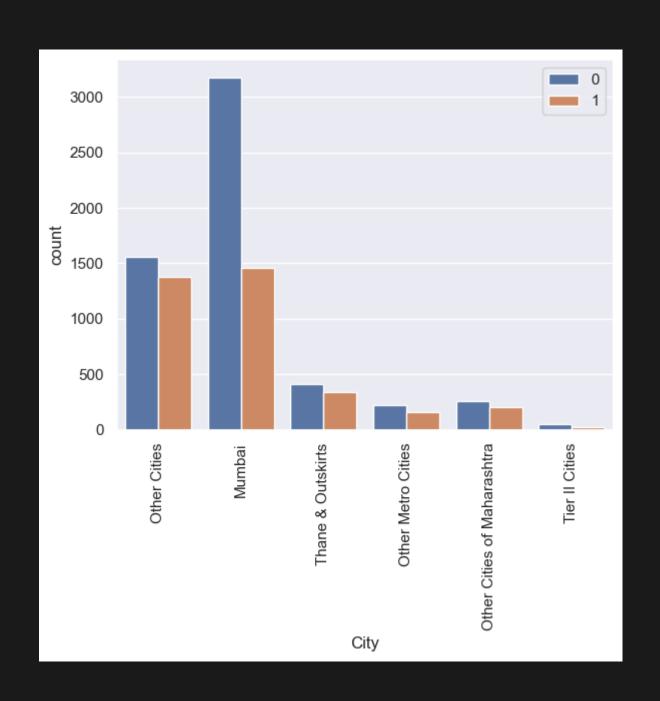
Lead Source

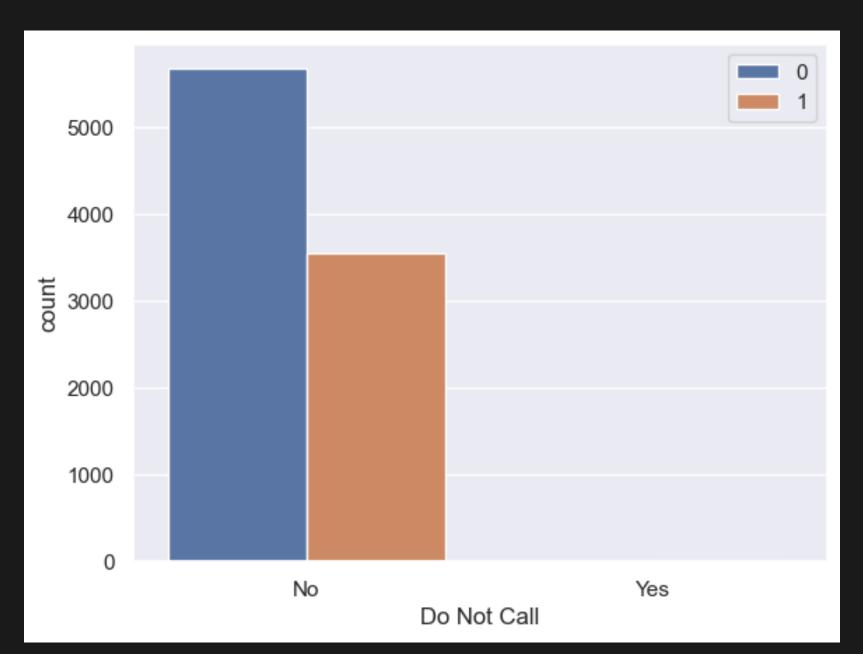
lead conversion rate thru Reference and Welingak Website are pretty high, so marketing for these channels should be increased.

Leads are also getting generated from Google/direct traffic as well, company should look into increasing conversion rate by managing the channels properly.

Olark chart has lesser conversion rate. Facebook and others are not contributing much, so resource and effort can be shifted to more promising channels.

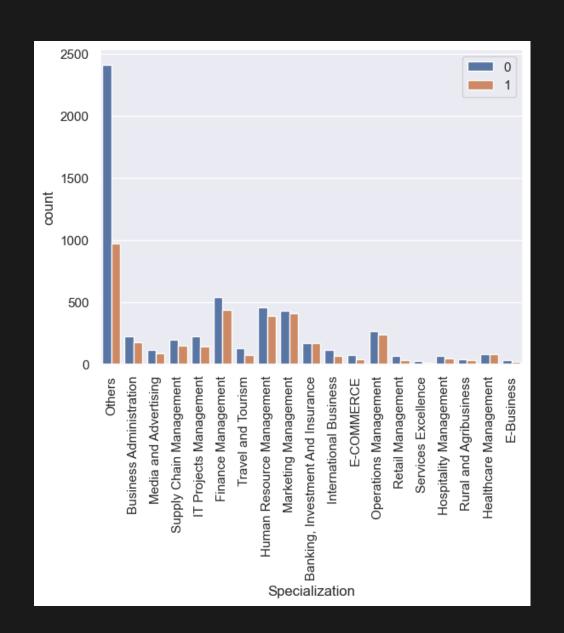
SEGMENTED UNIVARIATE ANALYSIS FOR CATEGORICAL COLUMNS (Contd.)

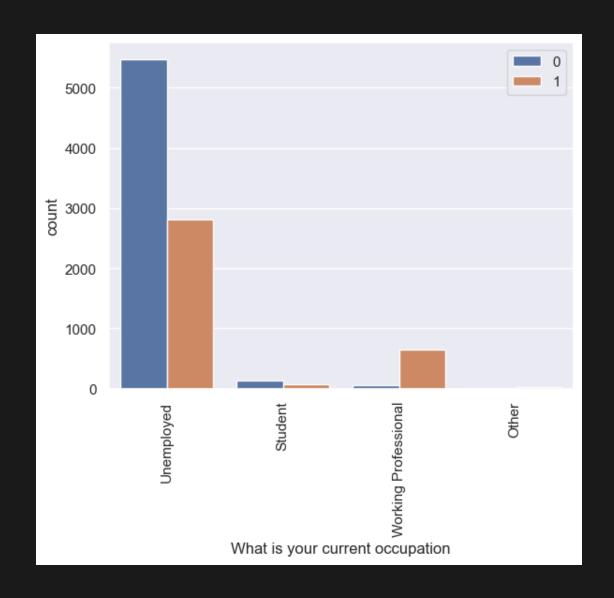




its evident that most leads don't prefer to be called or emailed. Company should look into why it is like that. Are the staffs are not putting enough attention to the customer queries?

SEGMENTED UNIVARIATE ANALYSIS FOR CATEGORICAL COLUMNS (Contd.)

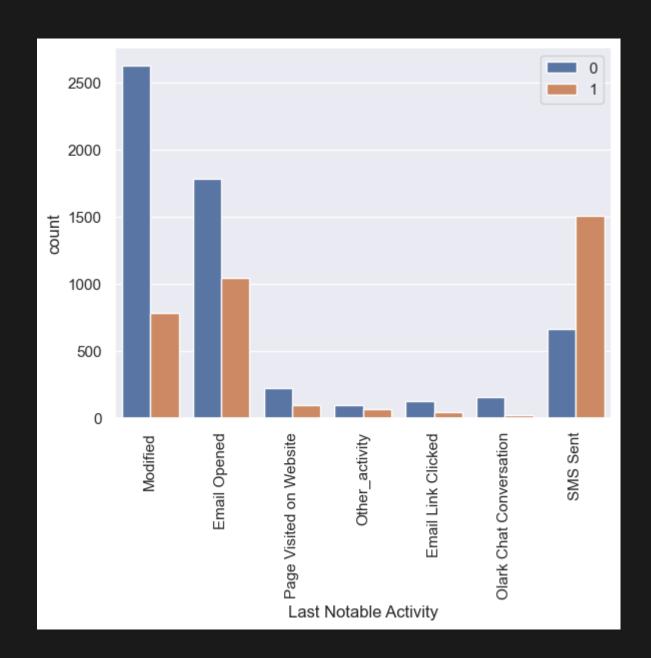


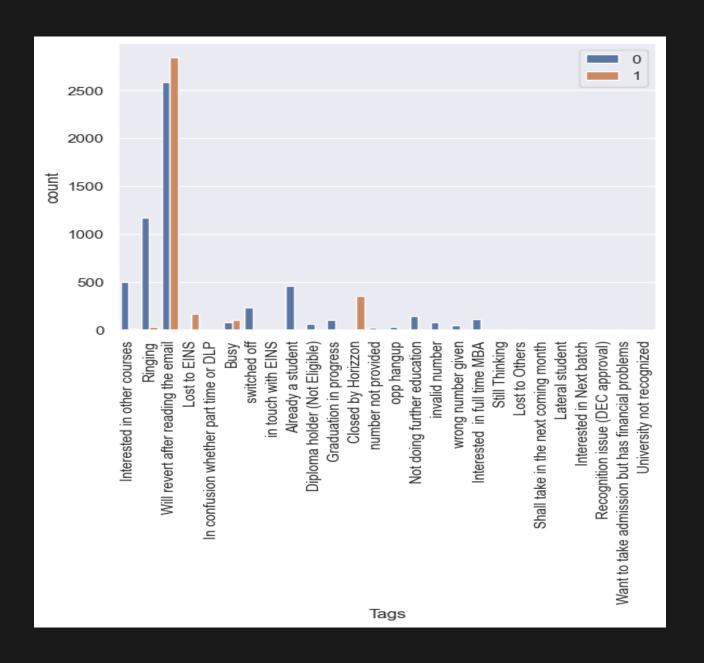


here, we see, people with specialization in Finance, HR, Banking, OM are having good conversion rates, so company should target these groups.

For current occupation, unemployed is by far the most lead generating section. Working professionals having very high conversion rate. We can put remaining in others category.

SEGMENTED UNIVARIATE ANALYSIS FOR CATEGORICAL COLUMNS (Contd.)





'Email Opened' and 'SMS sent' are among the highest contributors for this category in terms of conversion rate.

For 'Tags' feature, a lot of sub categories are ambiguous, so we decided to drop this feature as well.

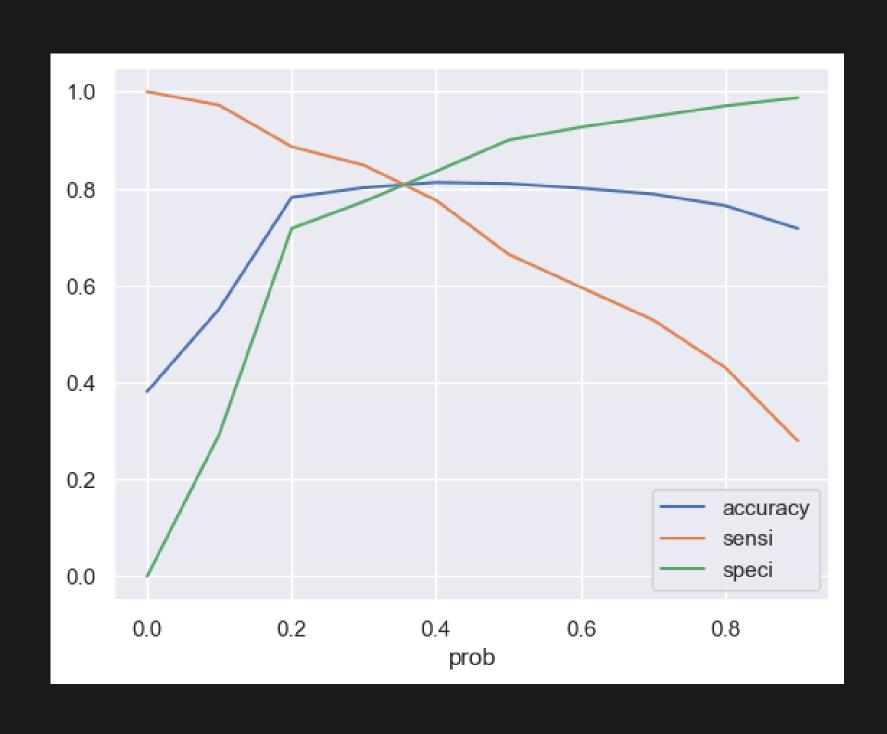
INFERENECES FROM FINAL LOGREG MODEL

Below are the observations from the final logistic regression model:

- 1. Total time spent on website is very influential attribute for hot leads.
- 2. When lead origin is thru lead add form / landing page submission, leads are likely to be converted.
- 3. When customer source is olark chat or welingak website, they are more likely to buy courses.
- 4. Customers having hospitality management as specialization are more likely to enroll.
- 5. Working professionals have more inclination to upskill themselves thru such courses than other categories.
- 6. When customers are involved in olark chat conversation or SMS conversations as last notable activity, they found to be more likely to buy courses from X education.

MODEL EVALUATION (TRAIN)

ACCURACY, SENSITIVITY AND SPECIFICITY



TN FP

3216 786

FN TP

463 2003

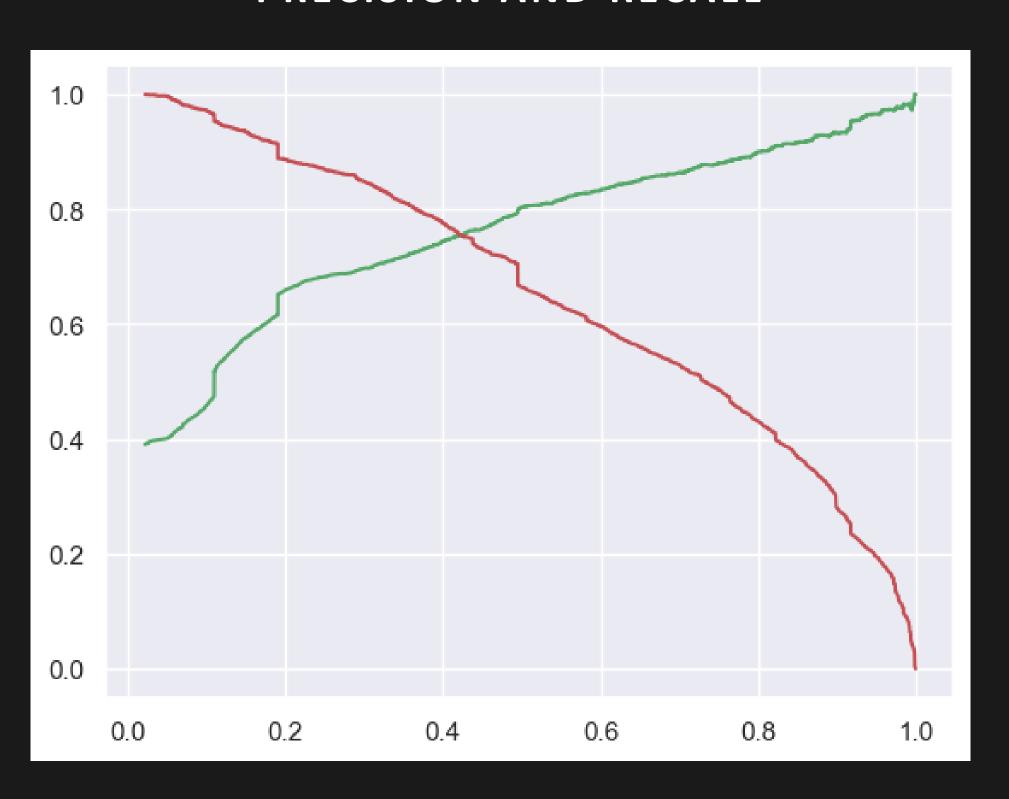
Accuracy: 81%

Sensitivity: 81%

Specificity: 80%

MODEL EVALUATION (TRAIN)

PRECISION AND RECALL



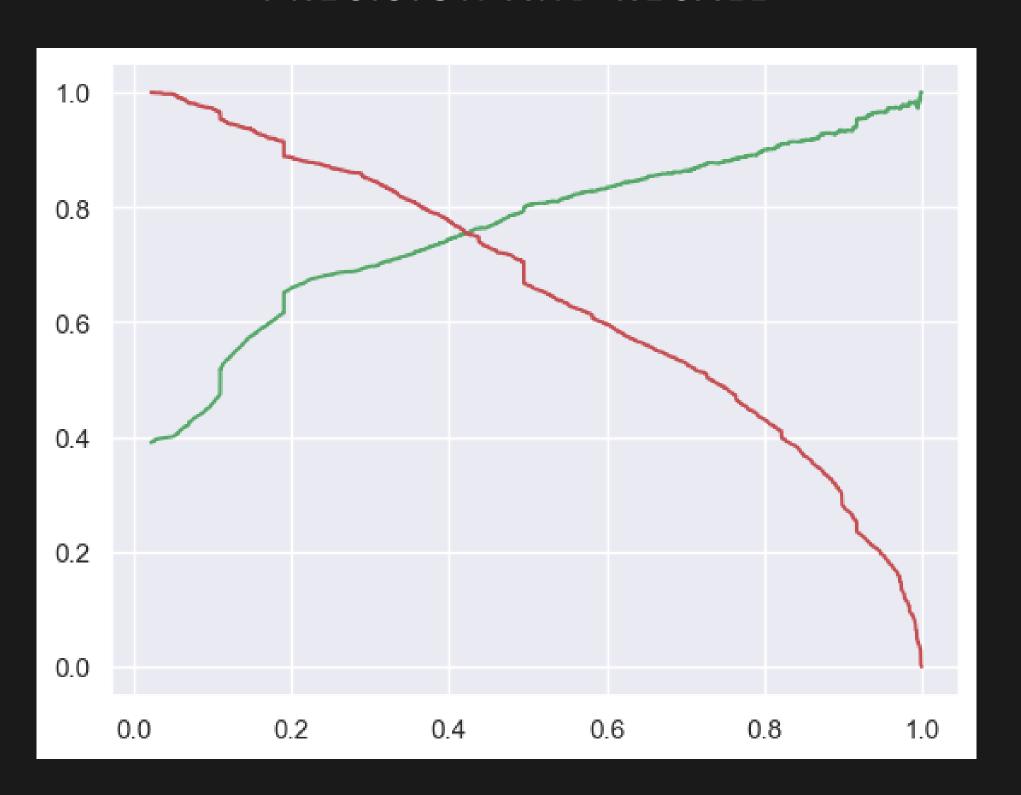
Accuracy: 81%

Precision: 75%

Recall: 75%

MODEL EVALUATION (TEST)

PRECISION AND RECALL



Accuracy: 81%

Precision: 77%

Recall: 75%

Test set threshold has been set as

0.41

SUGGESTIONS AND RECOMMENDATIONS

Below are the observations from EDA on dataset:

- 1. People spending higher than average time are promising leads, so targeting them could help increasing conversions.
- 2. SMS messages can have high impact on lead conversion.
- 3. Customer source such as olark chat or welingak website are effective in lead conversion.
- 4. Marketing management, HR management have high conversion rates, people from these specializations can be promising.
- 5. References and offers for referring a lead can be a good source for higher conversions.
- 6. Landing page submissions can help find out more leads.

Below are the observations from the final logistic regression model:

- 1. Total time spent on website is very influential attribute for hot leads.
- 2. When lead origin is thru lead add form / landing page submission, leads are likely to be converted.
- 3. When customer source is olark chat or welingak website, they are more likely to buy courses.
- 4. Customers having hospitality management as specialization are more likely to enroll.
- 5. Working professionals have more inclination to upskill themselves thru such courses than other categories.
- 6. When customers are involved in olark chat conversation or SMS conversations as last notable activity, they found to be more likely to buy courses from X education.

THANK YOU

