Lead Scoring Case Study

Problem Statement

- An education company named X Education sells online courses to industry professionals. The company advertises through Google and other portals through which they get leads.
- Sales professionals of X Education contact leads and try to convert them to paying customers
- Here the lead conversion rate is around 30% which needs to be enhanced.
- The company wants to spend time on potential leads rather than every lead.

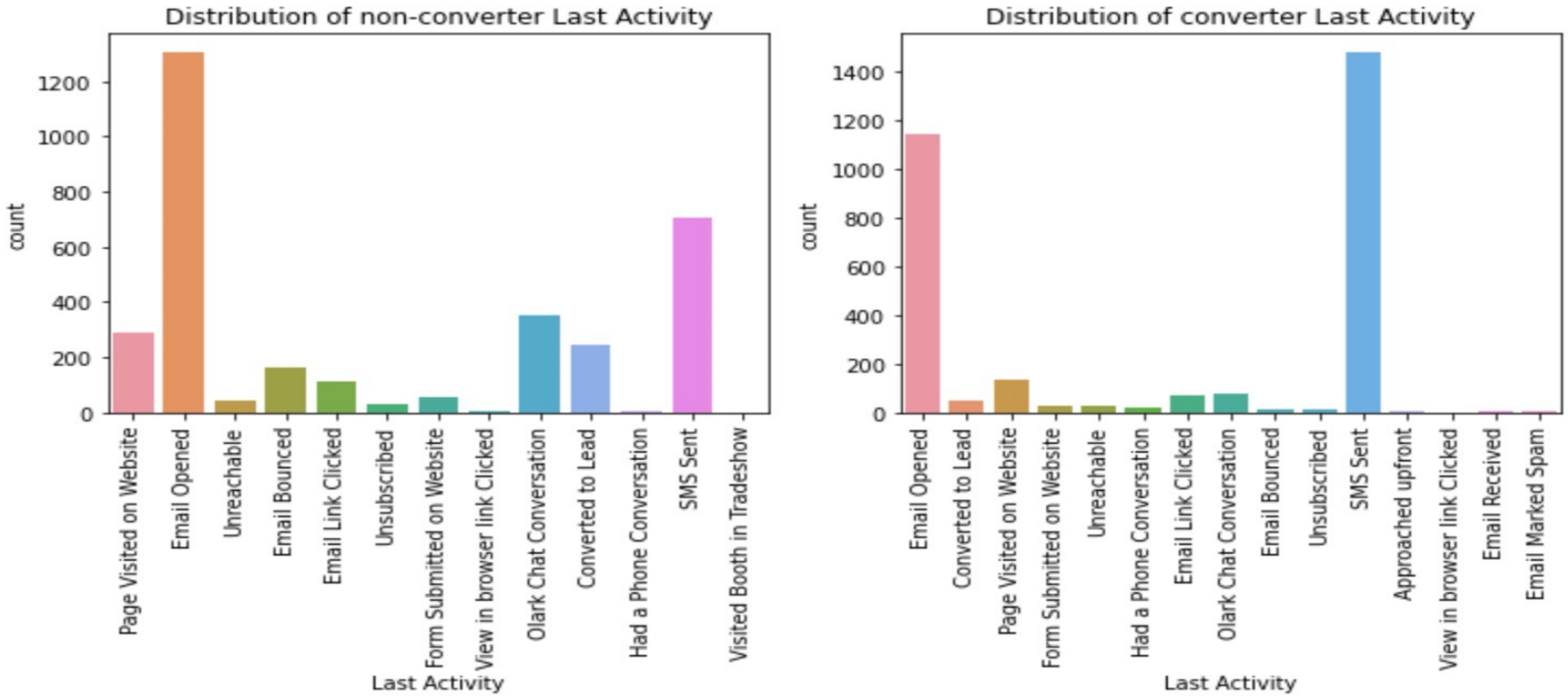
Business Objective

- The company needs to know potential leads.
- They want a model which can predict potential leads for them.
- The model should be able to assign lead scores to each lead.

Steps Followed:

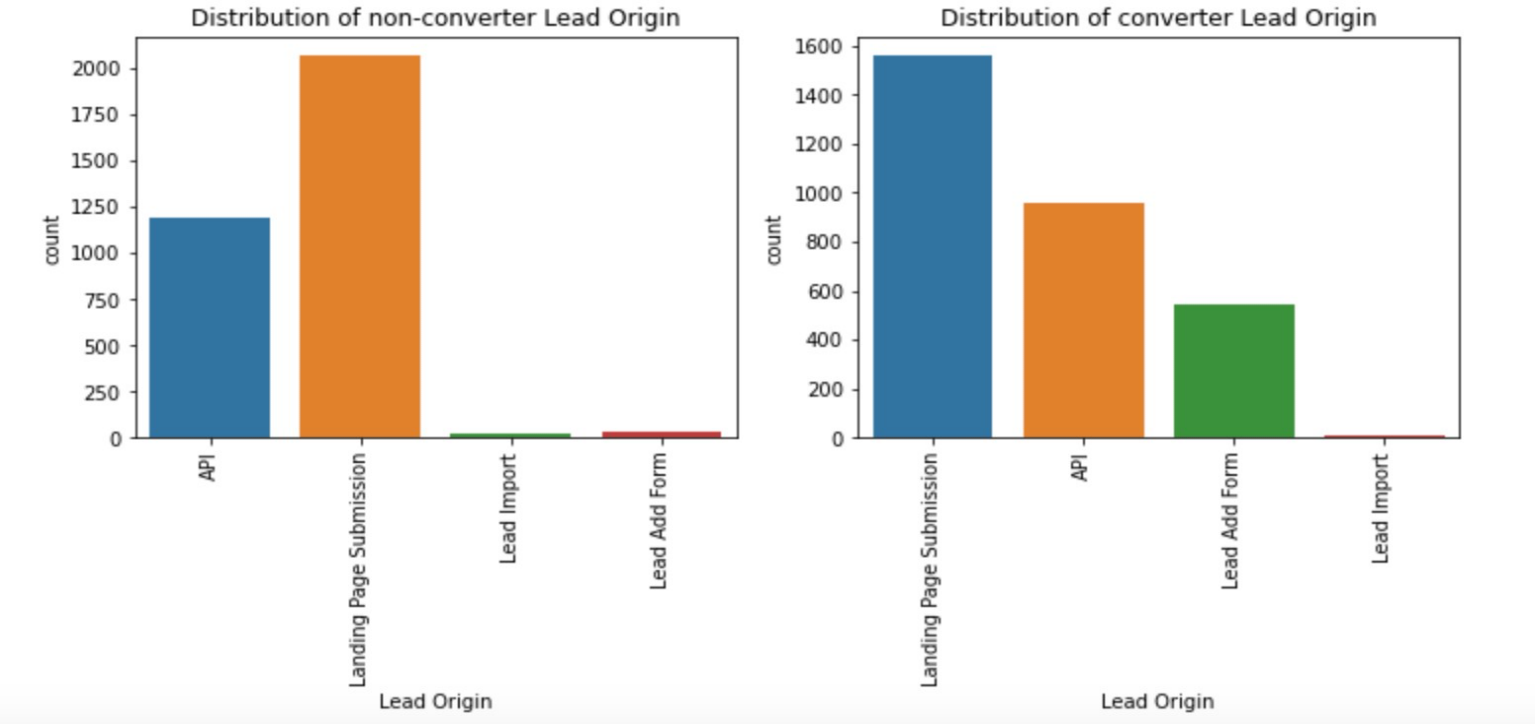
1. Data Understanding
2. Data Cleaning
3. Data Preparation
4. Exploratory Data Analysis
5. Model Building
6. Model Evaluation
7. Predictions on the Test Dataset

Last Activity with respect to Conversions



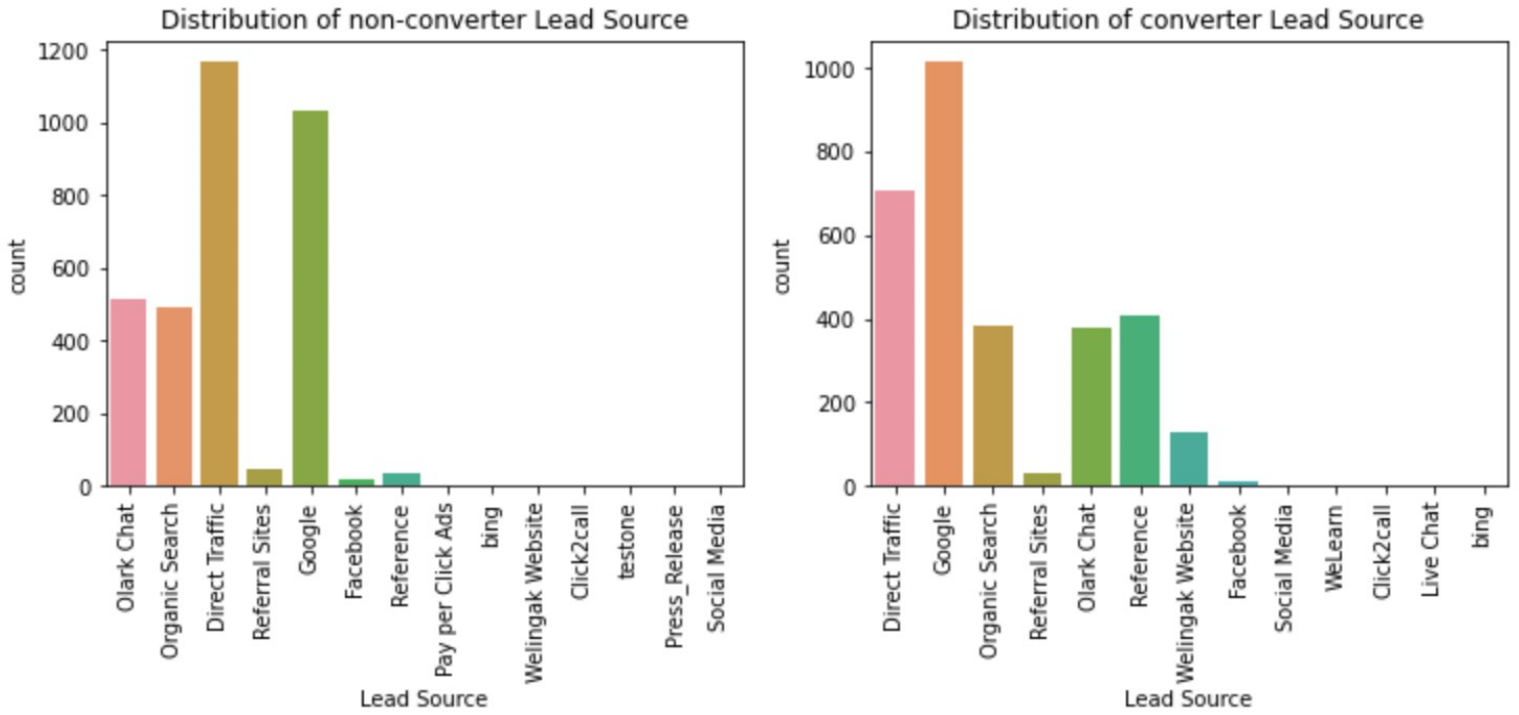
- Customers whose last activity is SMS sent are more likely to be converted.
- These customers are potential leads for X Education team.

Lead Origin with respect to Conversions



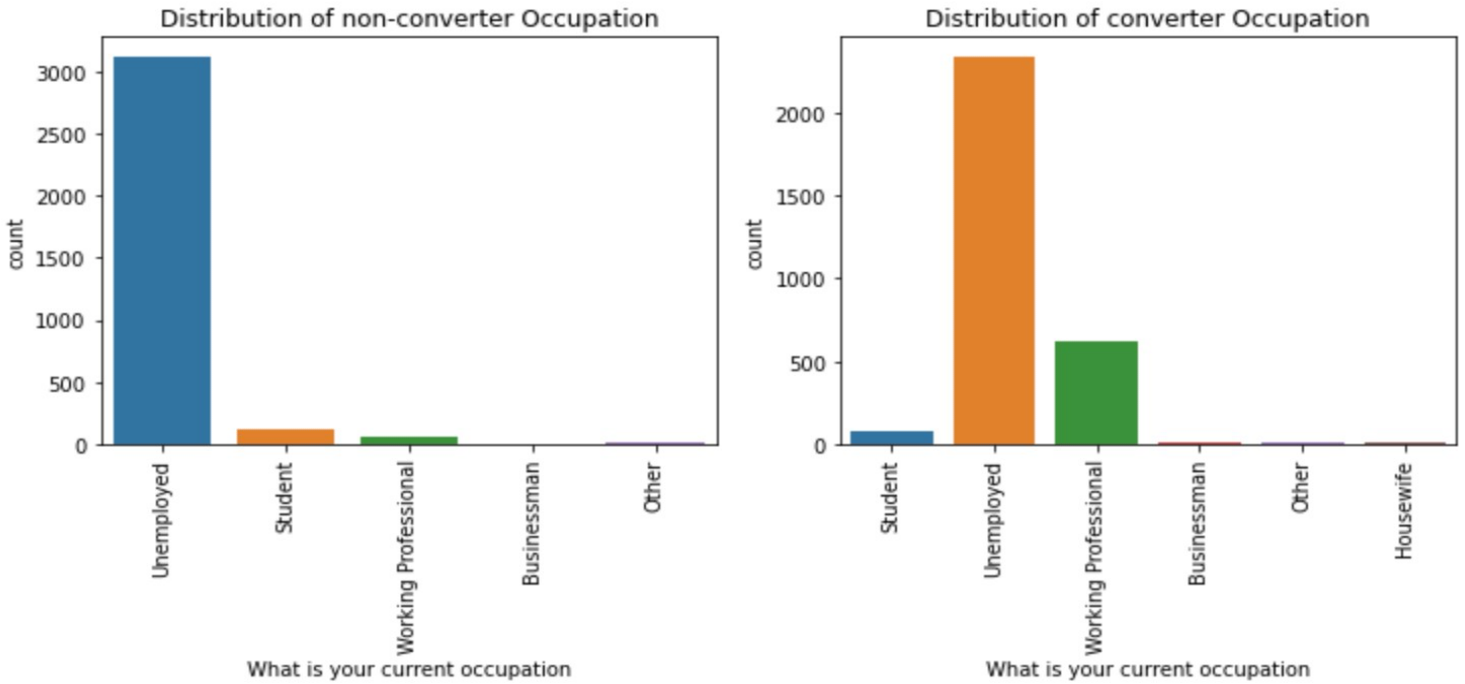
- Leads which are originated from Lead Add Form are most likely to get converted.

Lead Source with respect to Conversions



- If the Lead source is coming from Reference or Welingak Website, they are potential leads.

Occupation with respect to Conversions



• Working professionals are spending more time in enhancing their skills and hence more likely to be converted to potential lead.

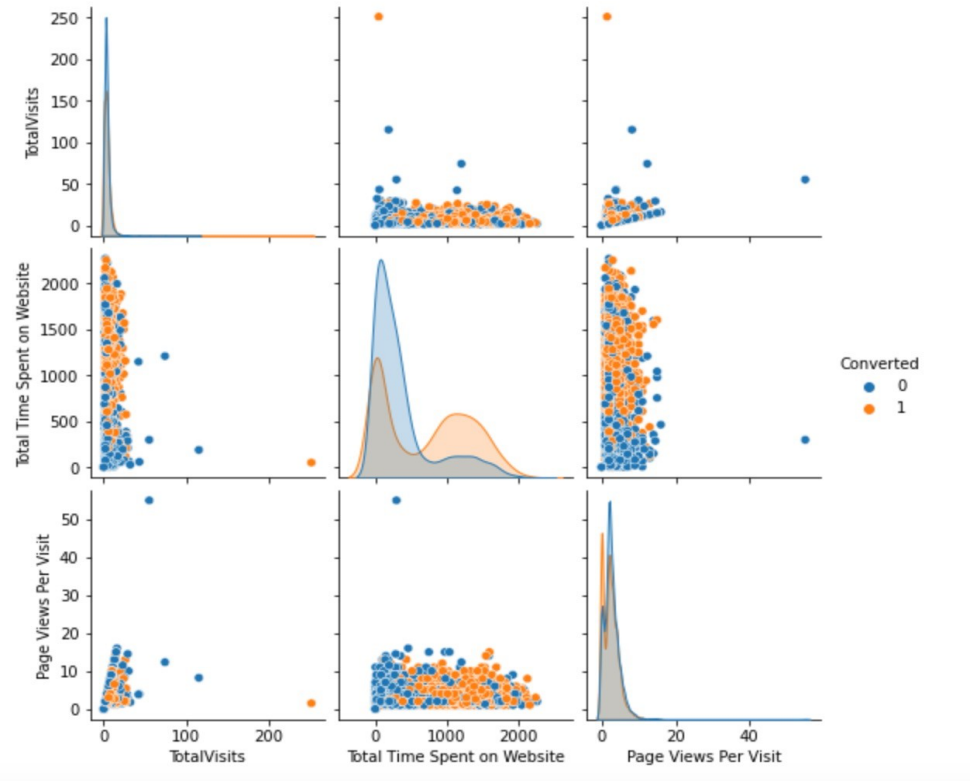
Top Features Determining High Conversions

TotalVisits
Total Time Spent on Website
Lead Origin_Lead Add Form
Lead Source_Olark Chat
Lead Source_Welingak Website
Do Not Email_Yes
Last Activity_Had a Phone Conversation
Last Activity_SMS Sent
What is your current occupation_Student
What is your current occupation_Unemployed
Last Notable Activity_Unreachable

These features play significant role in enhancement of conversion rate.

Total Visits, Total Time Spent on Website and Lead Origin_Lead Add form are the top 3 variables which positively impact the Lead Conversions. So these 3 variables should be given more importance.

Significant features with respect to Conversions

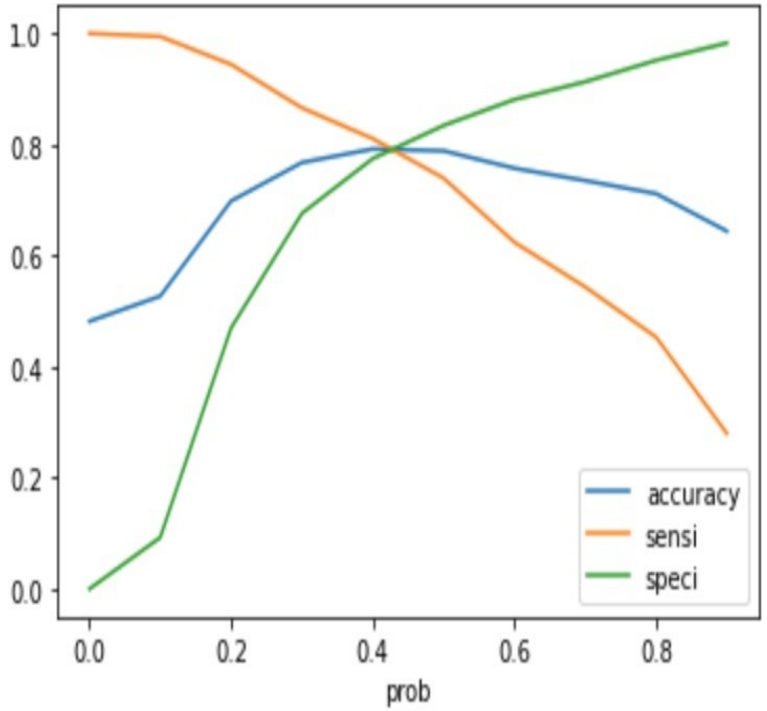


Total Visits, Total Time Spent on Website and Page views per Visit are highly impacting conversion.

These features are directly proportional with Target Conversion Variable.

Model Evaluation(Train):

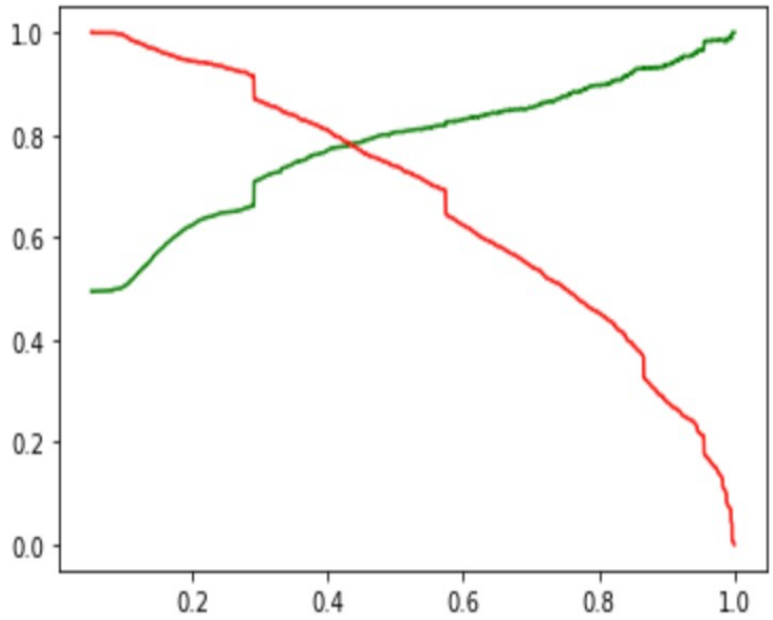
ROC Curve



- The ROC curve shows, at around 0.42, we get the optimal values of the three metrics i.e., accuracy, sensitivity and specificity.
- Accuracy : 79.08%
- Sensitivity : 79.33%
- Specificity : 78.84%

Model Evaluation(Train) Continued..

ROC Curve



ROC curve based on Precision Recall matrix.

- Keeping the threshold 0.42
- Precision : 77.71%
- Recall : 79.33%

Model Prediction(Test)

- Accuracy : 78.45%
- Sensitivity: 77.94%
- Specificity: 78.91%
- Precision: 77.27%
- Recall : 77.94%

Lead Score Prediction

	Converted	Conversion_Prob	final_predicted	Lead_Score
0	1	0.996296	1	100.0
1	0	0.129992	0	13.0
2	0	0.703937	1	70.0
3	1	0.299564	0	30.0
4	1	0.720796	1	72.0
5	1	0.792250	1	79.0
6	0	0.704038	1	70.0
7	1	0.464521	1	46.0
8	0	0.282978	0	28.0
9	1	0.786460	1	79.0

- The model has assigned lead score between 0 and 100 to each of the leads which can be used by the company to determine potential customers.

Findings

- In order to get a higher lead conversion, first, sort out the best prospects from the leads you have generated.
- 'TotalVisits' , 'Total Time Spent on Website' , 'Page Views Per Visit' features contribute most towards the probability of a lead getting converted. So, company needs to focus more on this.
- The model is able to capture correct leads having high probability of conversions also leads having low probability.
- The accuracy of the model is close to 80% with high sensitivity and specificity which will help the company reaching target customers.

THANK YOU