

Lead Score Case Study

Submitted By:-

Karthik Hulikunte

Devendra Sirsath

Yuvaraj Bhinge

Lead Score Case Study for X Education

Problem Statement:

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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 getting most promising leads i.e. the leads that are most likely to convert into paying customers.

Company requires model wherein you need to 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 CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

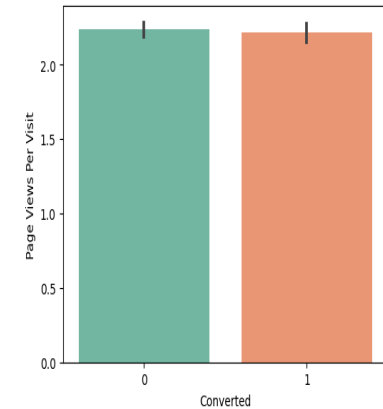
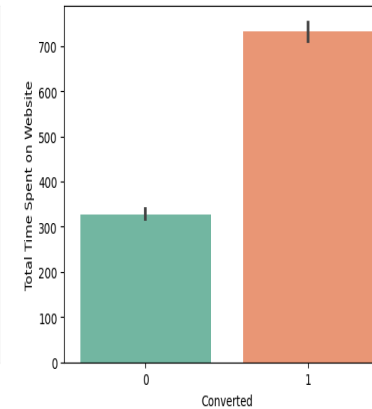
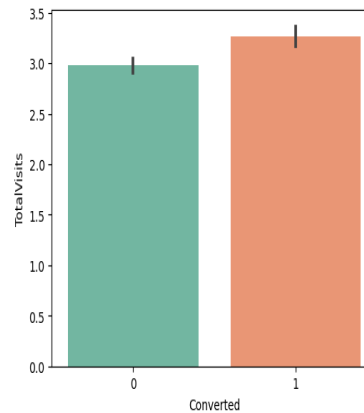
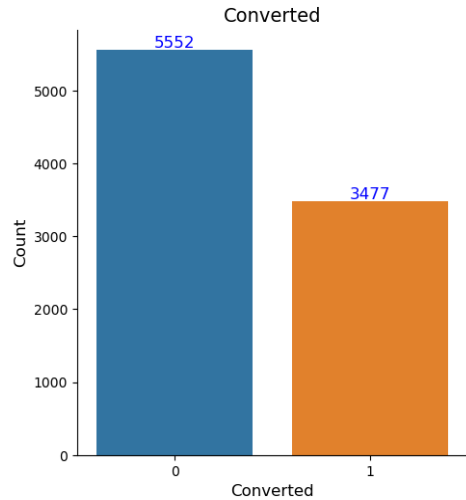
Action Plan:

- Reading, understanding and visualizing the data.
- Cleaning the Data
- Data Analysis
- Data Preparation
- Test-Train split
- Rescaling the features with MinMaxScaler
- Model Building
- Automatic Feature selection using RFE
- Plotting the ROC Curve
- Finding Optimal Cutoff Point
- Making prediction on Test set

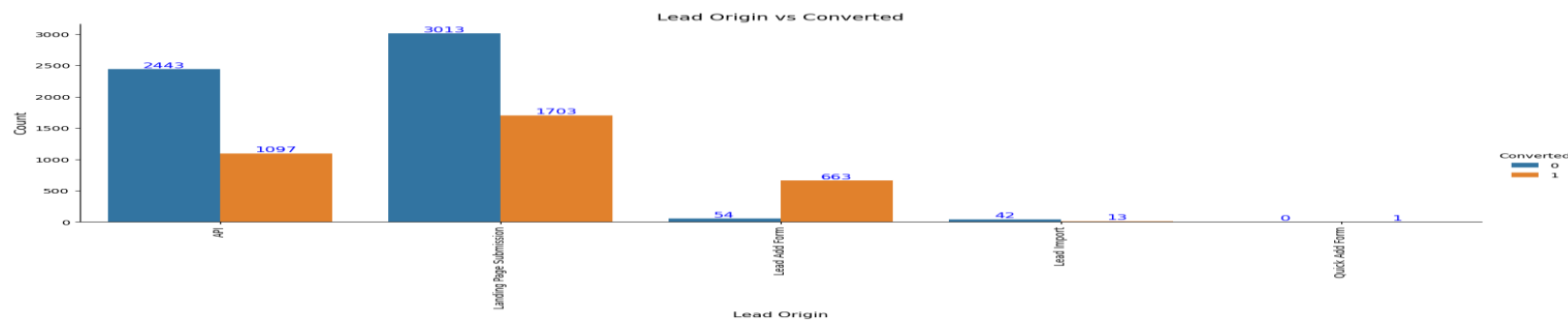
Exploratory Data Analysis

We have around 38.5% conversion rate in total.

Conversion rates were high for TotalVisits, Total time spend on website and page views per visit

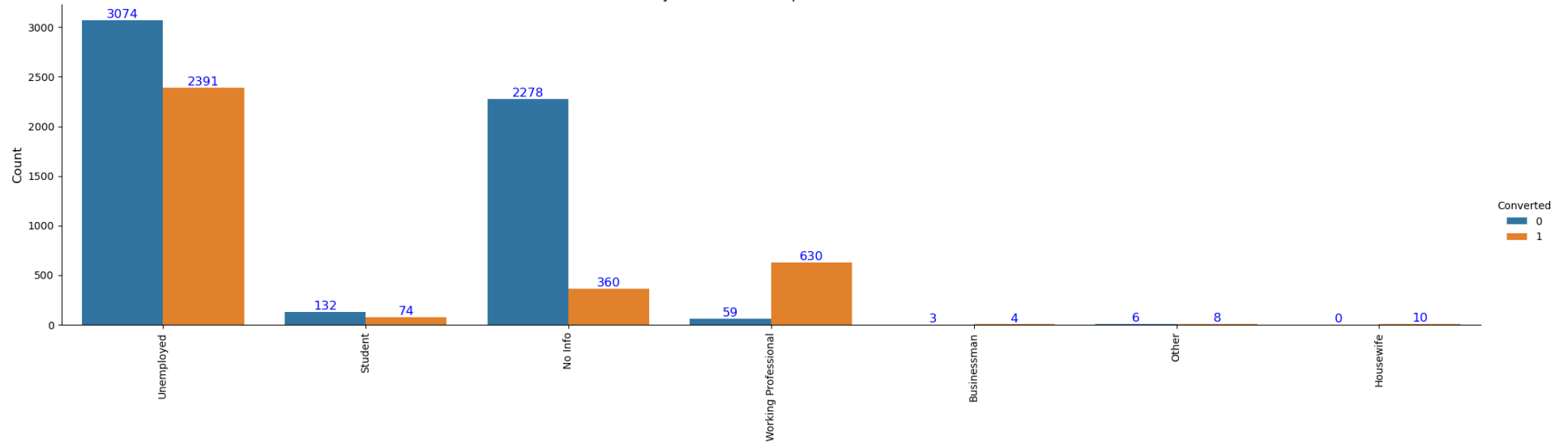


Lead Origin - Maximum Conversion happened from Landing Page Submission, Also one request from Quick Add Form and which is converted.



Major conversion is from unemployed category

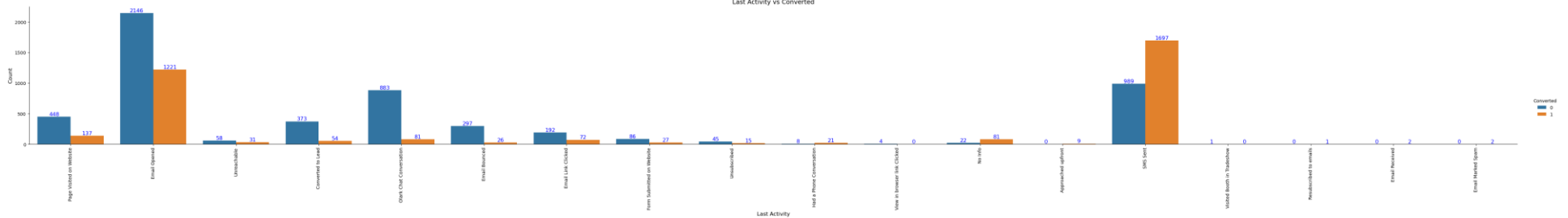
What is your current occupation vs Converted



What is your current occupation

Most conversion are through SMS send.

Last Activity vs Converted

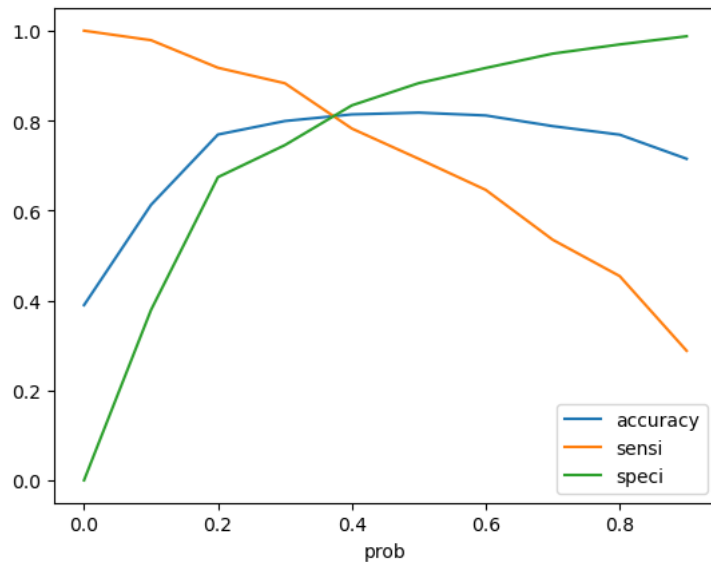


Features impacting the Conversion Rate

- Do Not Email
- TotalVisits
- Total Time Spent on Website
- Page Views Per Visit
- LeadOrigin_Lead Add Form
- LeadSource_Olark Chat
- LeadSource_Welingak Website
- LastActivity_Converted to Lead
- LastActivity_Email Bounced
- LastActivity_No Info
- LastActivity_Olark Chat Conversation
- LastActivity_Unreachable
- CurrentOccupation_No Info
- CurrentOccupation_Working Professional
- NotableActivity_Email Bounced
- NotableActivity_Had a Phone Conversation
- NotableActivity_SMS Sent
- NotableActivity_Unreachable

Model Evaluation – Sensitivity and Specificity on Train Data Set

Graph shows that optimal cut off is 0.37 based on Accuracy, Sensitivity and Specificity



Confusion Matrix

3256	702
478	1984

Accuracy 81.32%

Sensitivity 80.50%

Specificity 81.80%

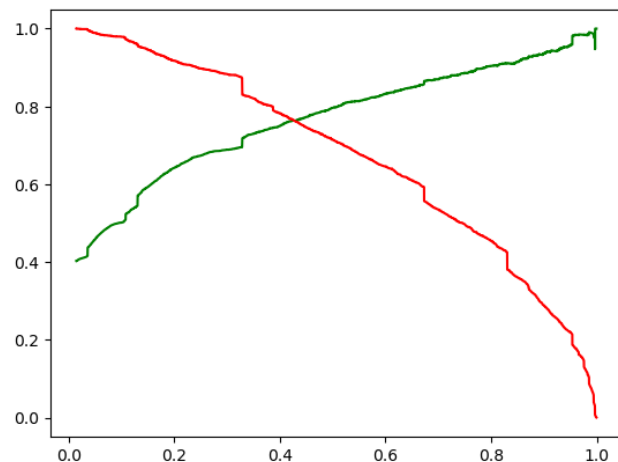
False Positive Rate 18.19%

Negative Predictive value 86%

Positive Predictive value 72%

Model Evaluation – Precision and Recall on Train Data Set

The graph shows optimal cut-off based on precision and recall is 0.42



Confusion Matrix

3408

450

702

1760

Precision 79%

Recall 71%

Model Evaluation – Sensitivity and Specificity on Test Data Set

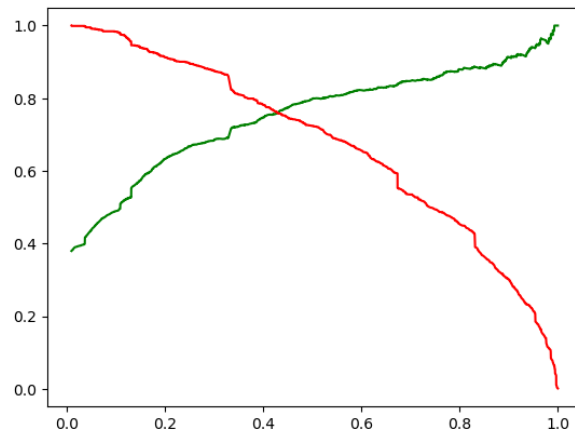
Accuracy 81%
Sensitivity 80%
Specificity 82%

Confusion Matrix

1392	302
204	811

Precision and Recall metrics for the test set

Same 0.42 cut-off from precision recall graph



Precision 73%

Recall 80%

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 81%, 79% and 82% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%
- Hence overall this model seems to be good.

