Lead Score Case Study for X Education

- Narendra Chinchalapu
- Shubham Kokate

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

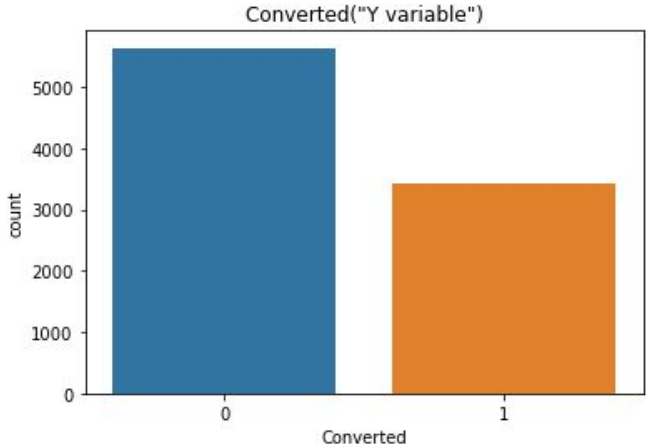
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the 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

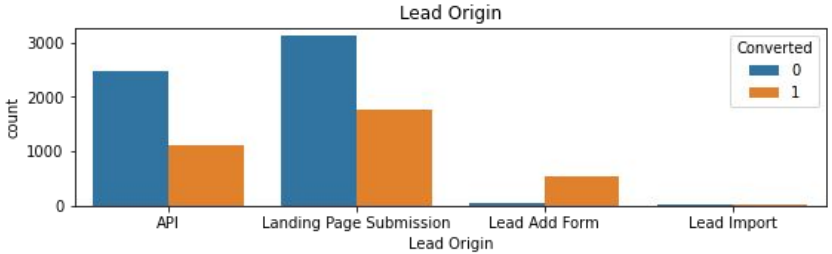
- Source the data for analysis
- Clean and prepare the data
- Exploratory Data Analysis.
- Feature Scaling
- Splitting the data into Test and Train dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics - Specificity and Sensitivity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

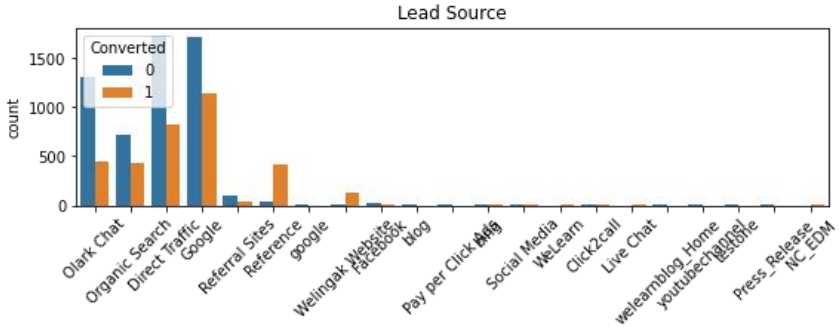
EDA

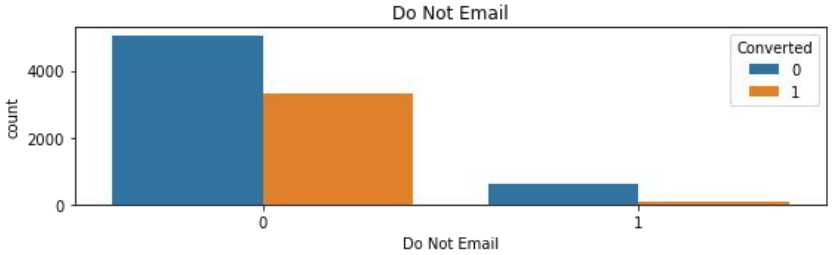
We can see that about 37% of the leads get converted.

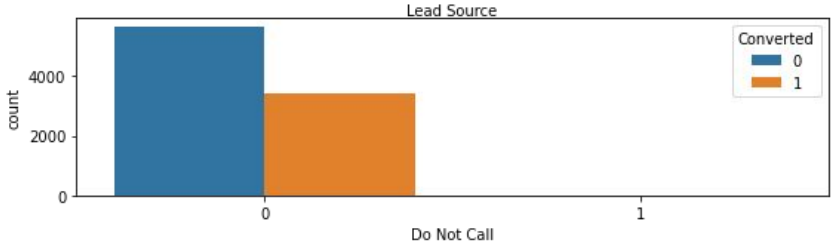


In Lead Origin maximum conversion happened from Landing Page Submission where as Google topped from Lead Source.Major conversion are happened from emails sent and calls made.

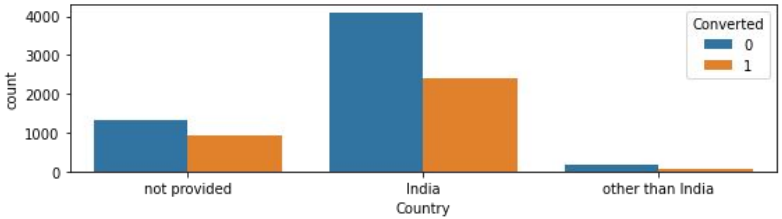


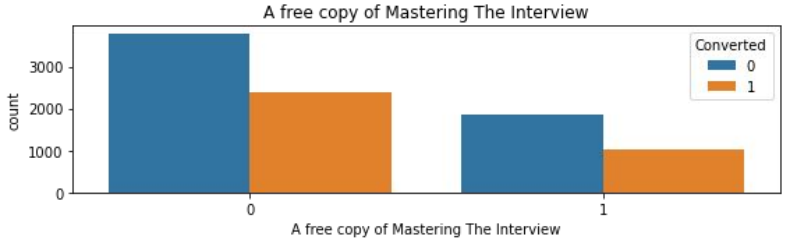


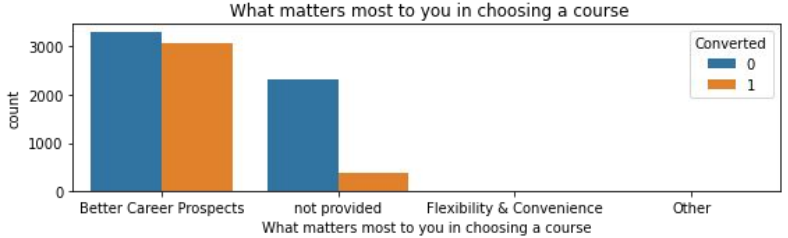




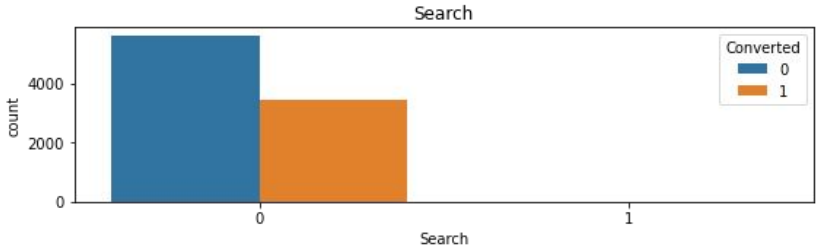
We can observe that most of the people from India are converted .Major conversion can be observed from the people who are unemployed.

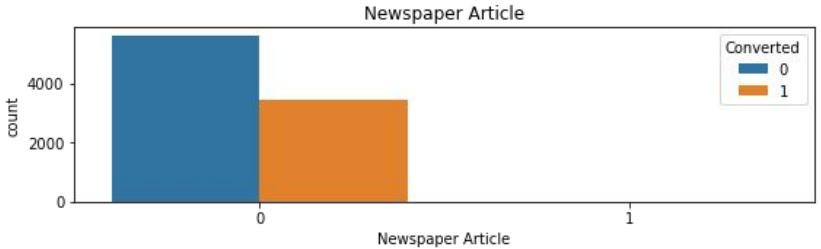


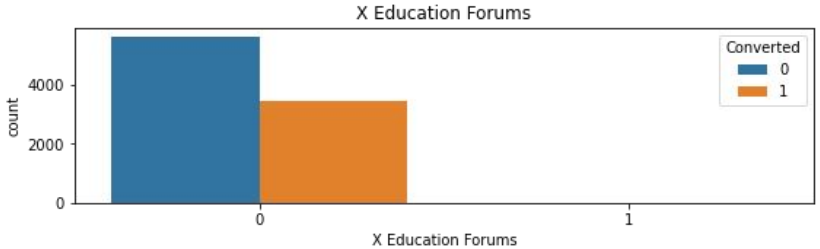


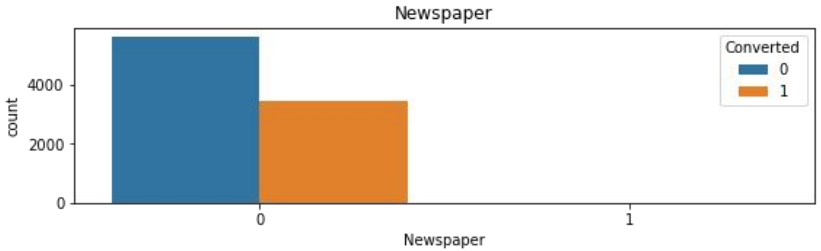


Not much impact on conversion rates through search,X Education forms , Newspaper Article and News Paper.

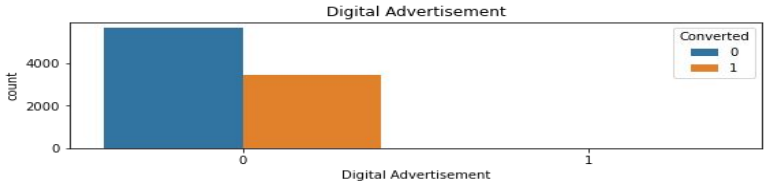


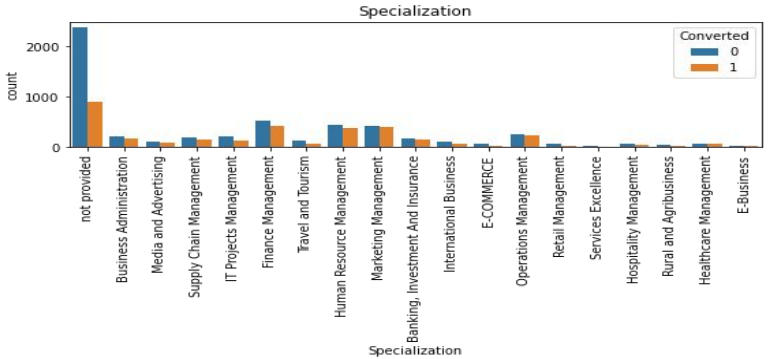


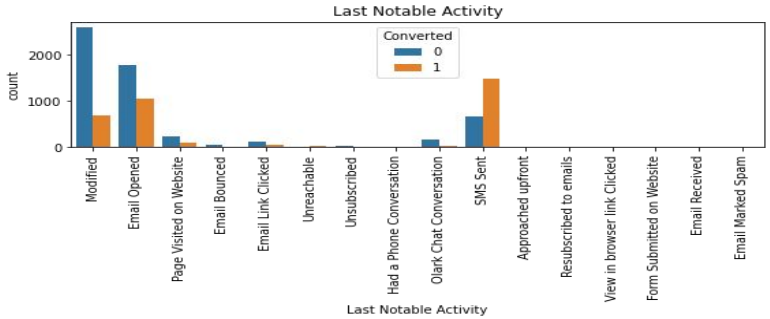




Through Digital Advertisement and Reccomendations there is no improvement in conversion rate.Through SMS we can get better conversion rate.







Variables Impacting on Lead Conversion Rate

- Do Not Email
- Total Time Spent on Website
- Lead Origin_Lead Add Form
- Lead Source_Olark Chat
- Lead Source_Welingak Website
- Last Activity_Had a Phone Conversation
- Last Activity_Olark Chat Conversation
- Last Activity_SMS Sent
- What is your current occupation_Working Professional
- Last Notable Activity_Modified
- Last Notable Activity_Unreachable
- Lead Quality_Low in Relevance
- Lead Quality_Might be
- Lead Quality_Not Sure
- Lead Quality_Worst

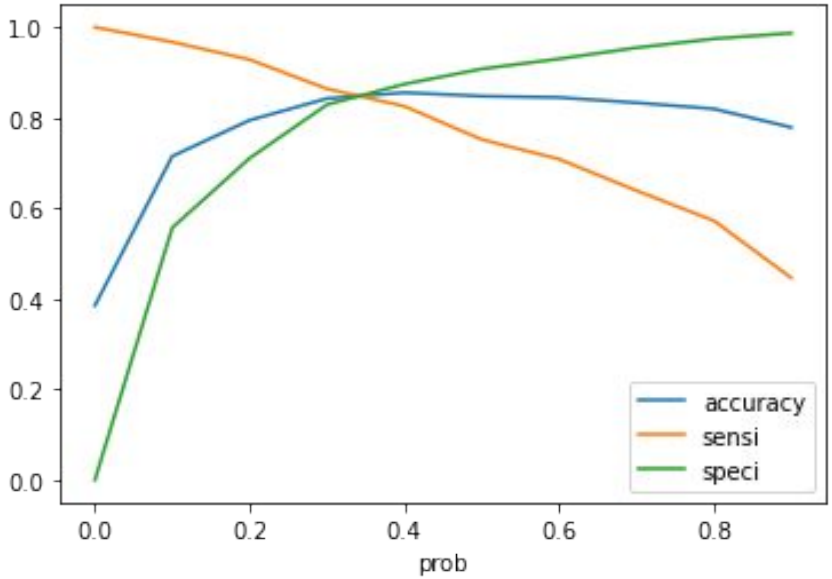
MODEL EVALUATION ON TRAIN DATA SET

This graph shows optimal cut-off 0.35 based on accuracy,sensitivity and specificity.

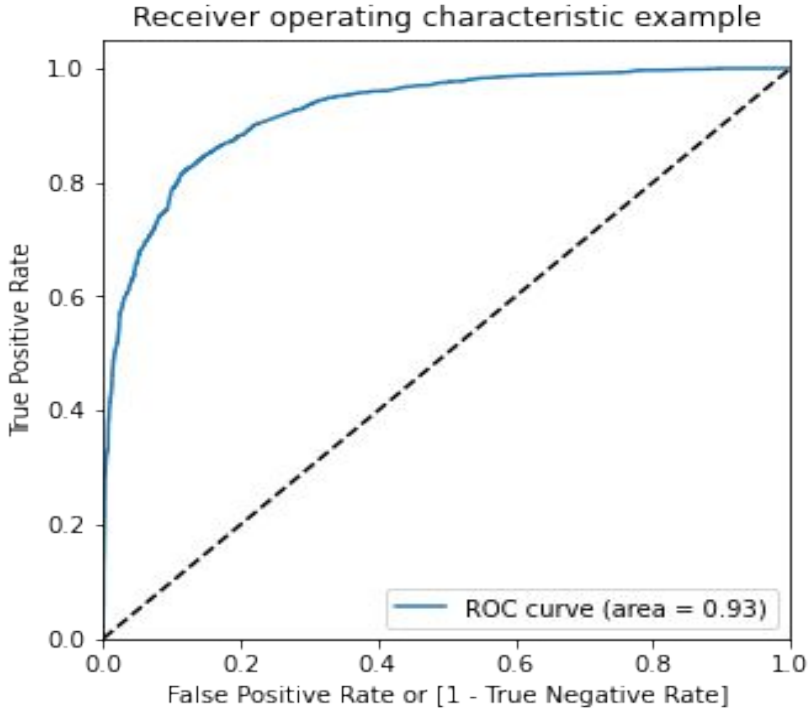
Confusion Matrix :

[[3341 564]
[383 2063]]

Accuracy : 85%
Sensitivity:84%
Specificity :85.5%
False Positive Rate :14%
Positive Predictive Rate :78%
False Predictive Power:89%



RECEIVER CHARACTERISTIC OPERATING CURVE



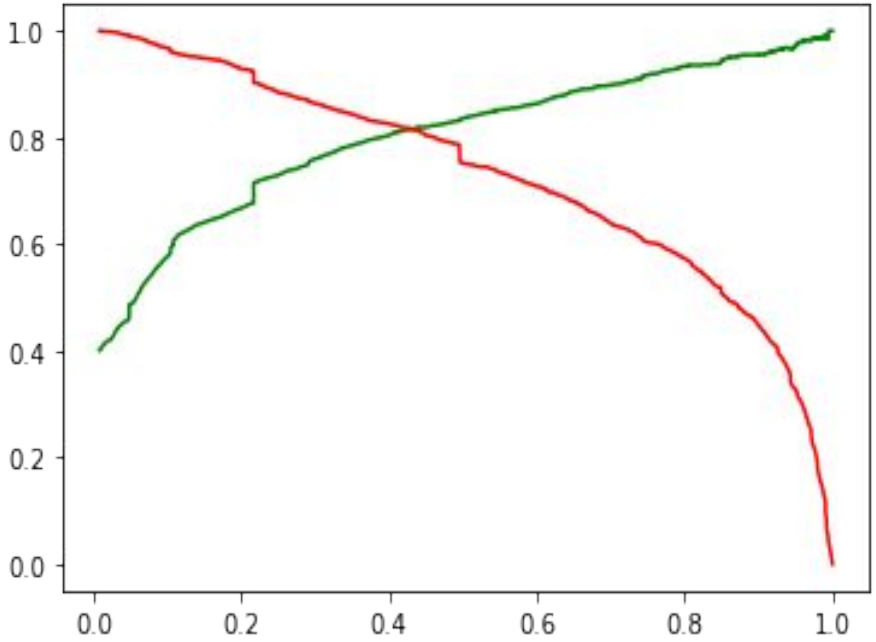
Precision Recall on Train Data set

The Graph shows Optimal cut-off of 0.41
based on Precision and Recall Value.

Precision : 81%
Recall : 82%

Confusion Matrix :

```
[[3434,  471]
[ 434, 2012]]
```



Model Evaluation on Test Data Set :

Confusion Matrix :

[1511, 223]

[197, 792]]

Accuracy : 84.5%

Specificity :87.13%

Precision : 78 %

Recall : 80%

Conclusion

We performed logistic regression to classify if a lead will get converted or not. It was observed that features such as Lead Source, Lead Quality, Occupation and Communication mode matter in classifying if a lead will convert or not.

We built a model with 84.5% accuracy.