Lead Score Case Study

Group members :

Yadnesh Chowkekar
Raghuram kuchibhotla
Ujwal Chaudhari

Lead Score Case Study

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:

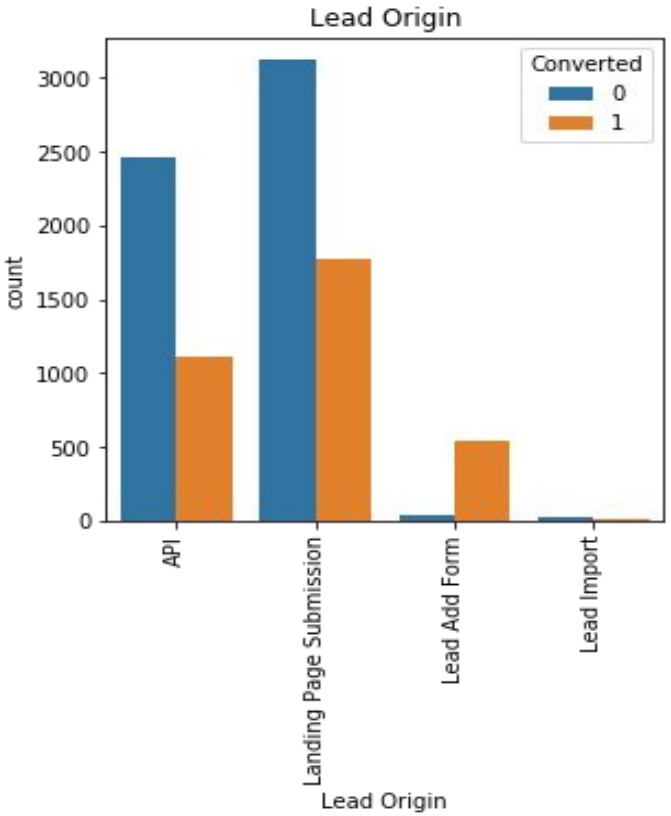
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
The company needs a model wherein you a lead score is assigned to each of the leads such that the customers with higher lead score have a higher conversion chance and the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

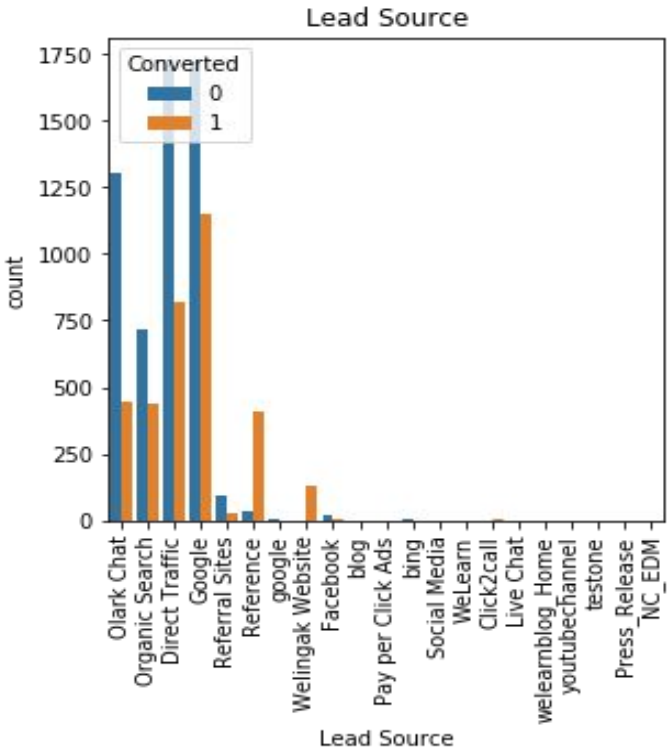
The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Solution workflow

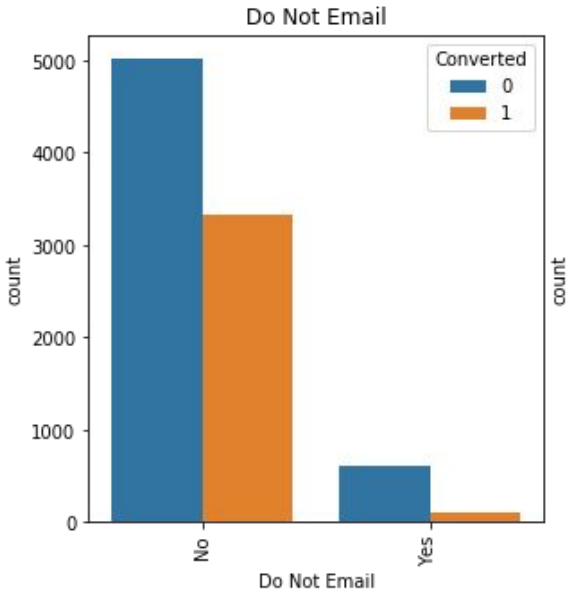
Data Sourcing , Cleaning and Preparation	Feature Scaling and Splitting Train and Test Sets	Model Building	Result
• Read the Data from Source • Convert data into clean format suitable for analysis • Remove duplicate data • Outlier Treatment • Exploratory Data Analysis Feature Standardization	• Feature Scaling of Numeric data • Splitting data into train and test set.	• Feature Selection using RFE • Determine the optimal model using Logistic Regression • Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.	• Determine the lead score and check if target final predictions amounts to 80% conversion rate. • Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

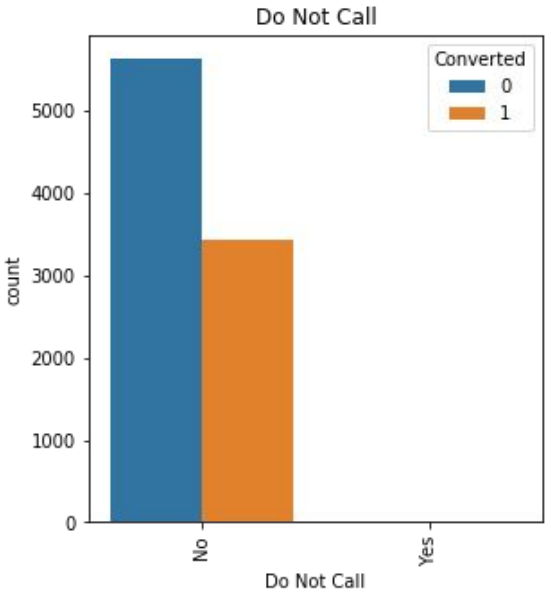
In Lead Origin, maximum conversion happened from Landing Page Submission



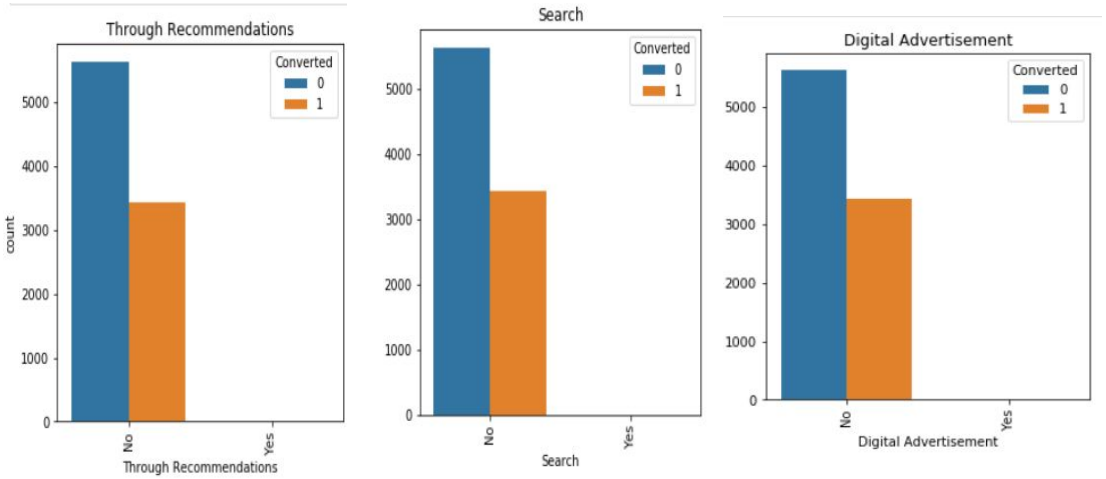


Major conversion has happened from Emails sent and Calls made

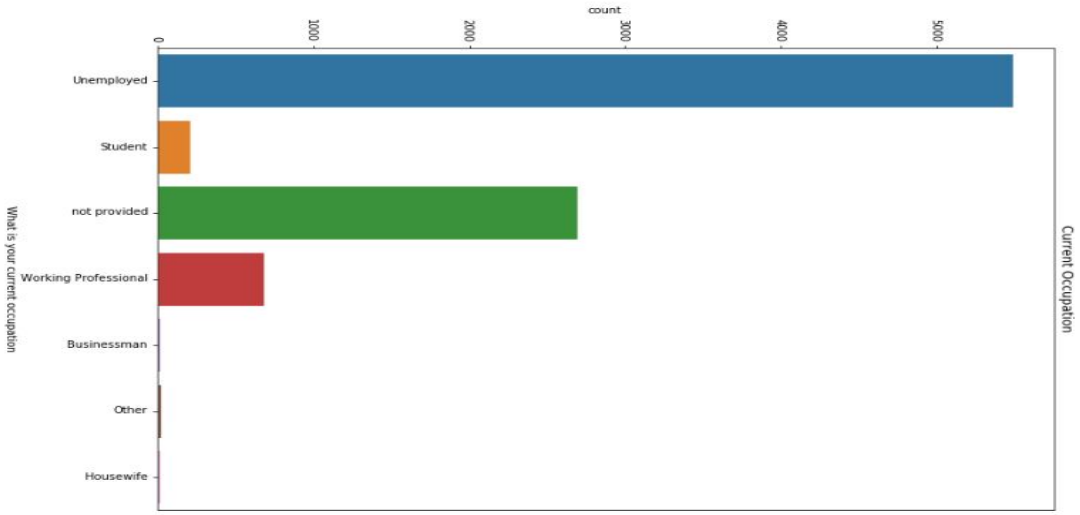


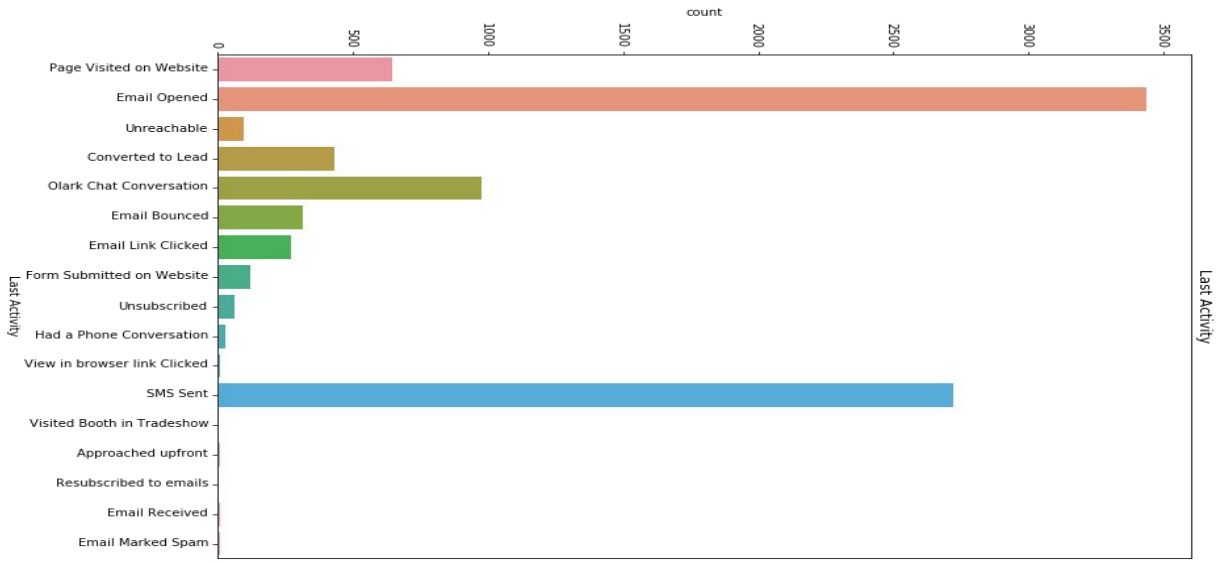


Not much impact on conversion rates through Search, digital advertisements and through recommendations



More conversion happened with people who are unemployed





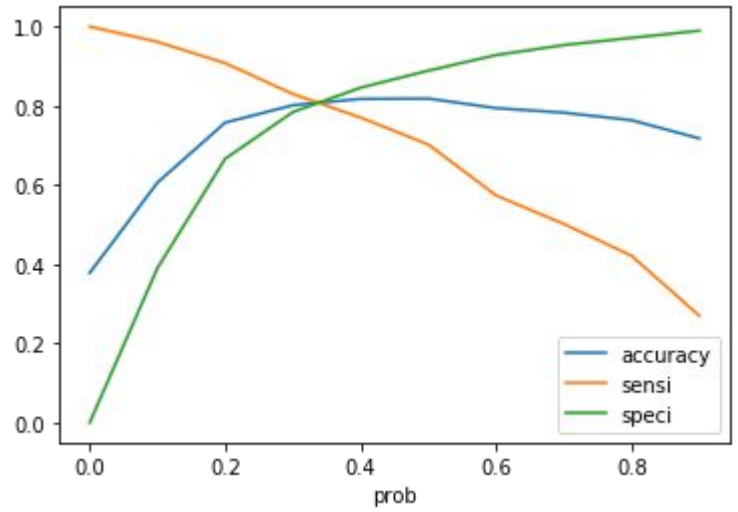
Last Activity value of SMS Sent had more conversion.

Variables Impacting the Conversion Rate

- Total Visits
- The total time spent on the Website.
- Lead Origin_Lead Add Form
- Lead Source_Direct Traffic
- Lead Source_Google
- Lead Source_Welingak Website
- Lead Source_Organic Search
- Lead Source_Referral Sites
- Lead Source_Welingak Website
- Do Not Email_Yes
- Last Activity_Email Bounced
- Last Activity_Olark Chat
- Conversation

Model Evaluation - Sensitivity and Specificity on Train Data Set

The graph depicts an optimal cut off of 0.35 based on Accuracy,
Sensitivity and Specificity



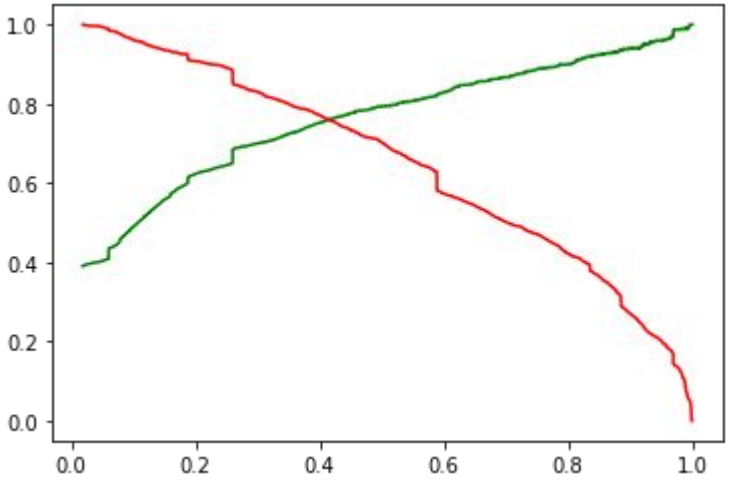
Confusion Matrix

[3191, 724],
[477, 1901]

- Accuracy - 81%
- Sensitivity - 80 %
- Specificity - 82 %
- False Positive Rate - 18 %
- Positive Predictive Value - 74 %
- Positive Predictive Value – 86%

Model Evaluation- Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.41 based on Precision and Recall



Confusion Matrix

[3333, 582],
[563, 1815]

- Precision - 79 %
- Recall - 71 %

Checking overall accuracy

Confusion Matrix

[1426, 251],
[273, 748]

- Accuracy - 80.5 %
- Sensitivity - 79 %
- Specificity - 82 %

With the current cut off as 0.41 we have Precision around 75% , Recall around 73% and accuracy 80.5%.

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

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. –
- With the current cut off as 0.41 we have Precision around 75% , Recall around 73% and accuracy 80.5% calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 80% (in train set) and 79% in test set
- The top 3 variables that contribute for lead getting converted in the model are
  - Total time spent on website
  - Lead Add Form from Lead Origin
  - Last Activity_Olark Chat Conversation
- Hence overall this model seems to be good.