

Lead Score Case Study

Presented by : Dhanraj Babbar,
Sanjeeta and Bidyut Kumar Roy

Lead Score Case Study for X Education

Problem :

One online X Education sells online courses to industry professionals. The company markets courses on various websites and search engines like Google, Bing etc

When people land on the website, they might browse the courses or fill up a form for that course or may be watch 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 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%.

Strategies

- Importing the packages and dataset
- Clean junk data and prepare data
- Analyzing the Variables.
- Train and Test dataset.
- Building a logistic Regression model and calculate Lead Score.
- Evaluating the model by using different metrics – Accuracy, Sensitivity and specificity.
- Applying the best model in Test data based on the Sensitivity and Specificity Metrics.

Problem solving methodology

Data Sourcing , Cleaning and Preparation

Reading Data from Source
Converting data to clean format which is suitable for analysis
Remove duplicate data
Outlier
Analyzing the variables
Feature Standardization.



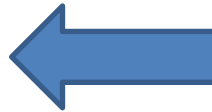
Feature Scaling and Splitting Train and Test Sets

Feature Scaling of Numeric data
Splitting data into train and test set.



Model Building

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

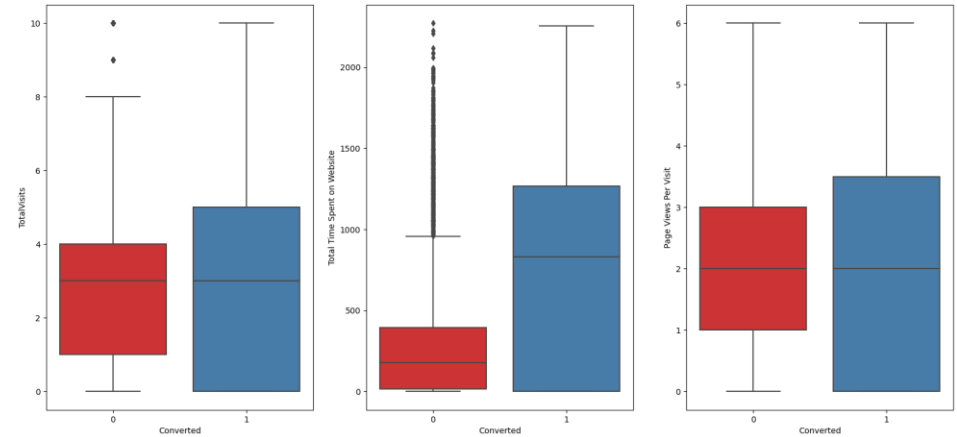


Result

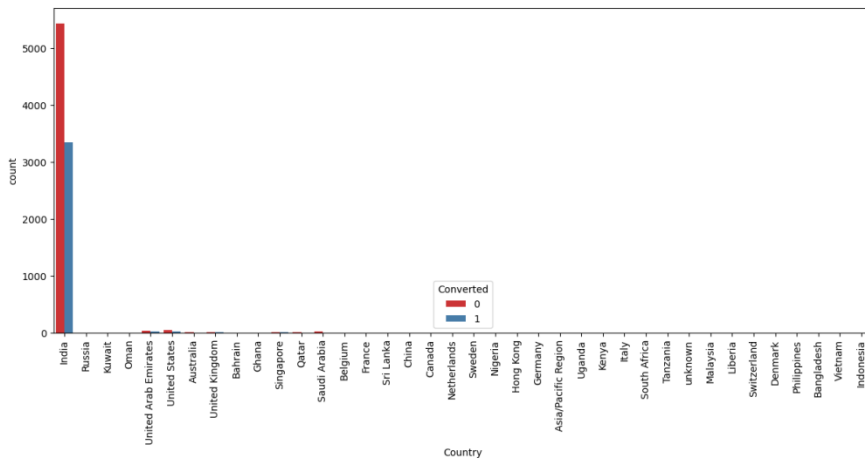
Determine the lead score and check if target final predictions amounts to 80% conversion rate.
Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

Exploratory Data Analysis

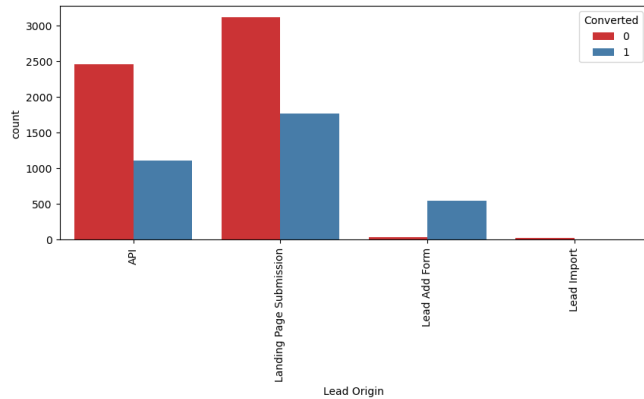
The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit



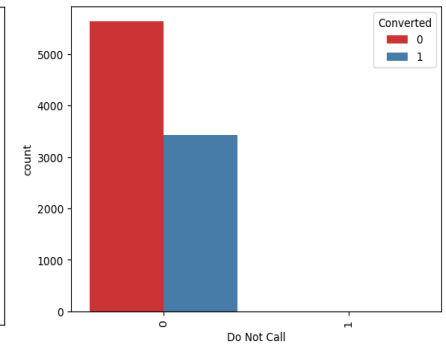
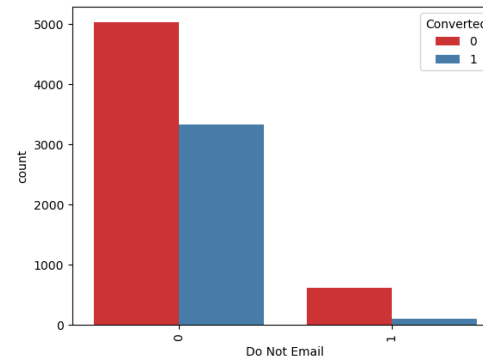
Conversion Chart showing below



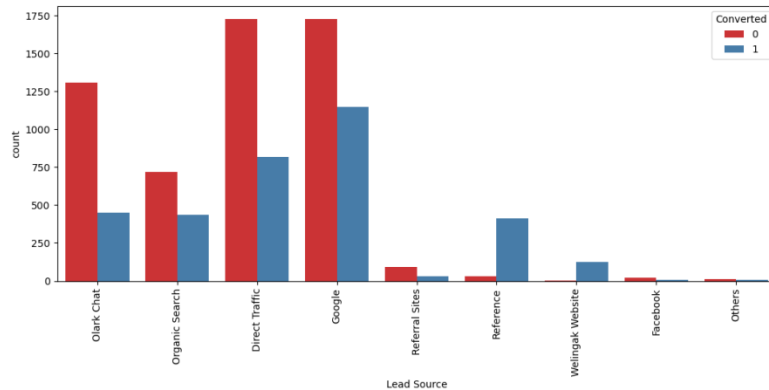
In Lead Origin, maximum conversion happened from Landing Page Submission



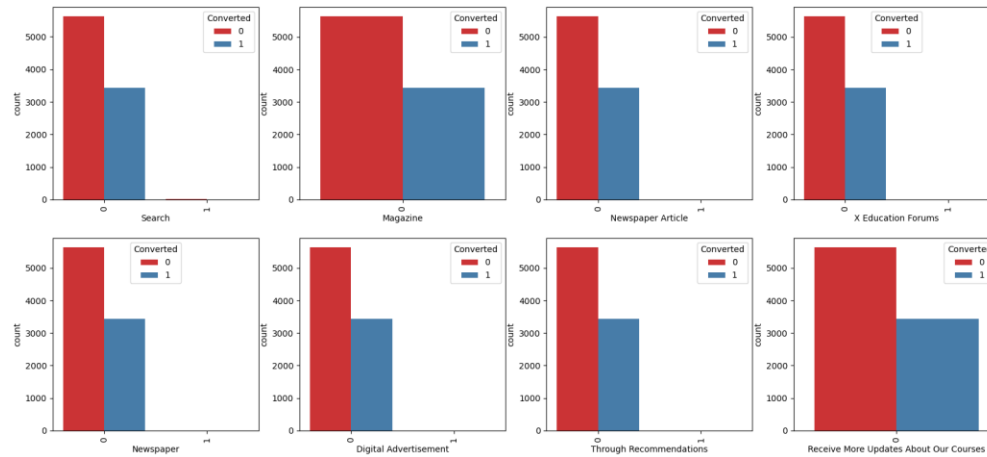
Major conversion has happened from Emails sent and Calls made



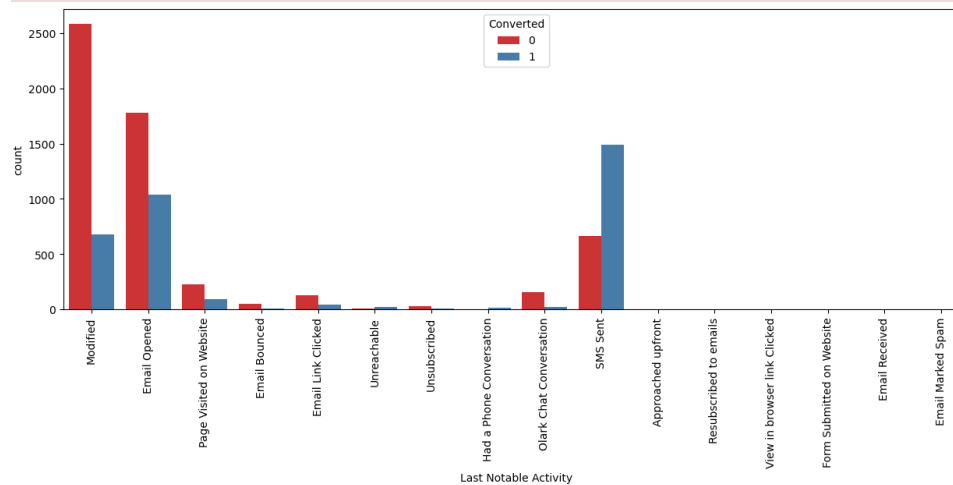
Major conversion in the lead source is from Google



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



Last Activity value of SMS Sent' had more conversion.



Variables Impacting the Conversion Rate

Do Not Email

Total Visits

Total Time Spent On Website

Lead Origin – Lead Page Submission

Lead Origin – Lead Add Form

Lead Source - Olark Chat

Last Source – Welingak Website

Last Activity – Email Bounced

Last Activity – Not Sure

Last Activity – Olark Chat Conversation

Last Activity – SMS Sent

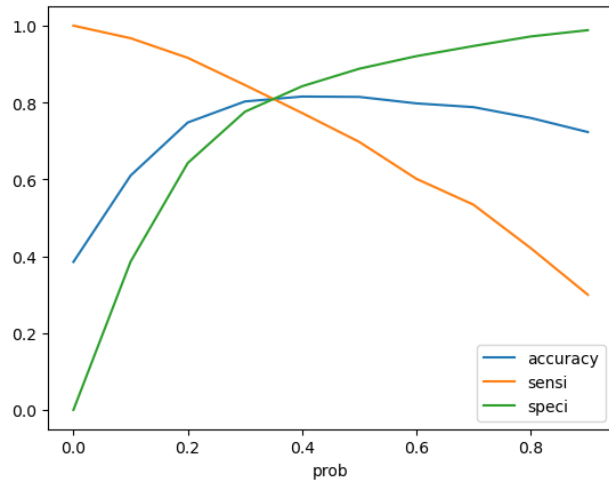
Current Occupation – No Information

Current Occupation – Working Professional

Last Notable Activity – Had a Phone
Conversation

Last Notable Activity - Unreachable

Model Evaluation - Sensitivity and Specificity on Train Data Set



Confusion Matrix

3467

438

740

1706

Accuracy - 81%

Sensitivity – 69 %

Specificity – 88 %

False Positive Rate - 11 %

Positive Predictive Value - 79 %

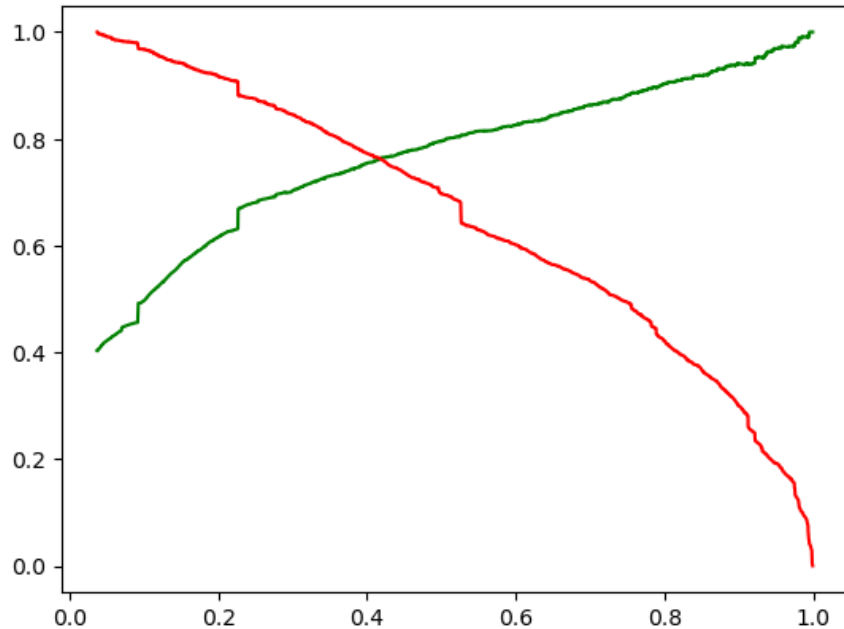
Positive Predictive Value – 82%

Precision - 79 %

Recall - 69 %

Model Evaluation- Precision and Recall on Train Dataset

The graph depicts an optimal cut off of 0.42 based on Precision and Recall



Confusion Matrix

1395	339
192	797

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix

1395	339
192	797

- Accuracy – 80.4 %
- Sensitivity – 80.50 %
- Specificity – 80.40 %