Lead Score Case Study

Submitted by:
Abhiram
Abhishek
Rajeshwari

Lead Score Case Study for X Education

Problem Statement :

- X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.
- When people visit the website, they might browse the courses, fill out a form for the course, or watch some videos. When these people provide their email address or phone number, they are considered 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 choosing the most promising leads, i.e. the leads that are most likely to become paying customers.
- The company requires a model where each lead is handed a lead score, and leads with higher lead scores have a better chance of converting, while leads with lower lead scores have a lower chance of converting.
- The desired lead conversion rate was determined by the CEO to be in the range of 80%.

Lead Score Case Study for X Education

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 in Test data based on the Sensitivity and Specificity Metrics.

Problem solving methodology

Data Soring, Cleaning and Preparation

- Understand the Data from source
- Convert Data into suitable format for analysis.
- Duplicate data must terminated
- Outlier treatment
- Data analysis Exploratory
- Standardization Feature

Model Building

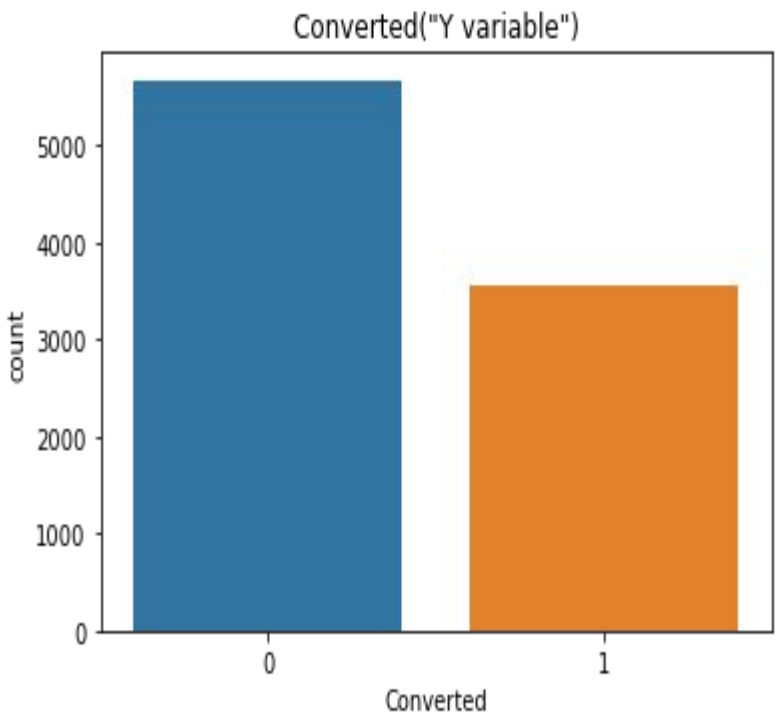
- Feature Selection using RFE
- Determine Optimal Model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, precision and evaluate the model

Result

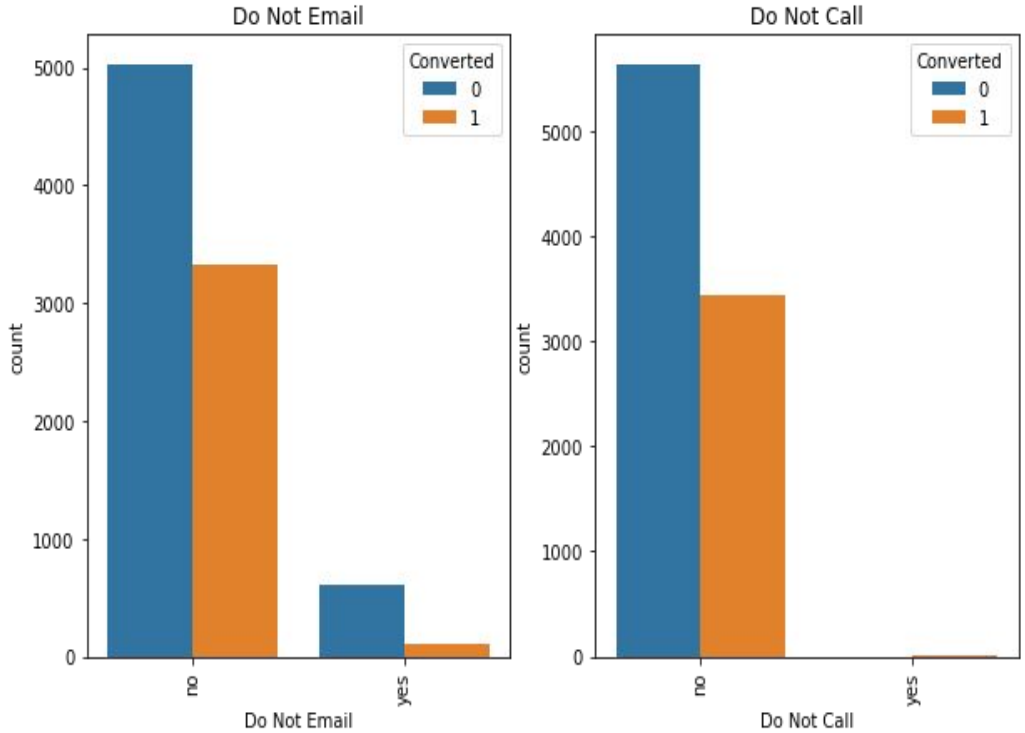
- Validate the desired final projections amount to 80% conversion rate and the lead score.
- Using the cutoff threshold from the sensitivity and specificity measures, assess the final prediction on the test set.

Exploratory Data Analysis

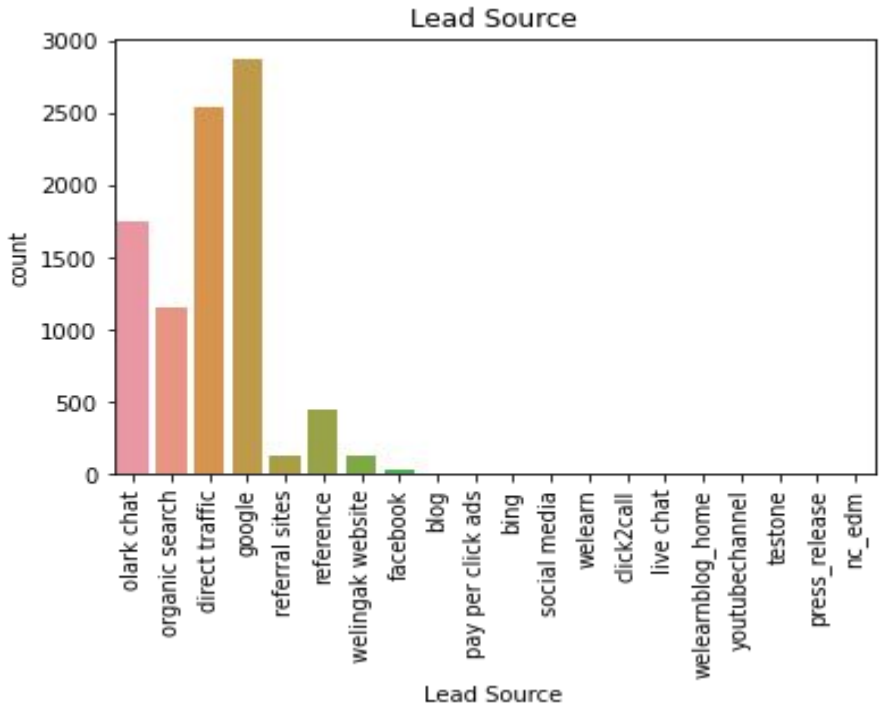
We have around 38% Conversion rate in Total



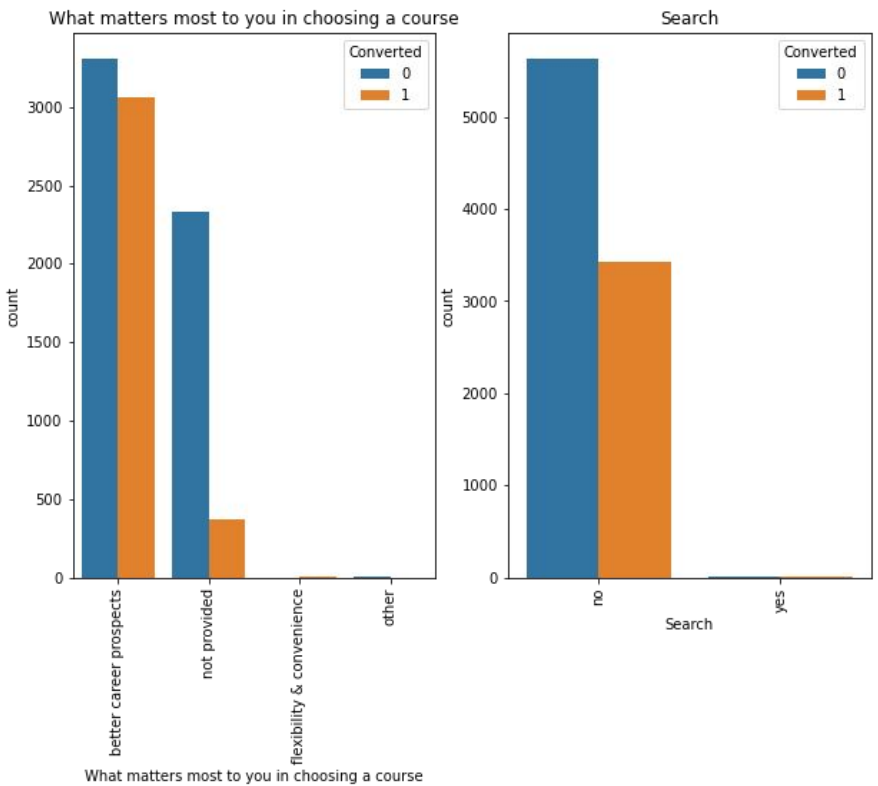
Major conversion has happened from Emails sent and Calls made



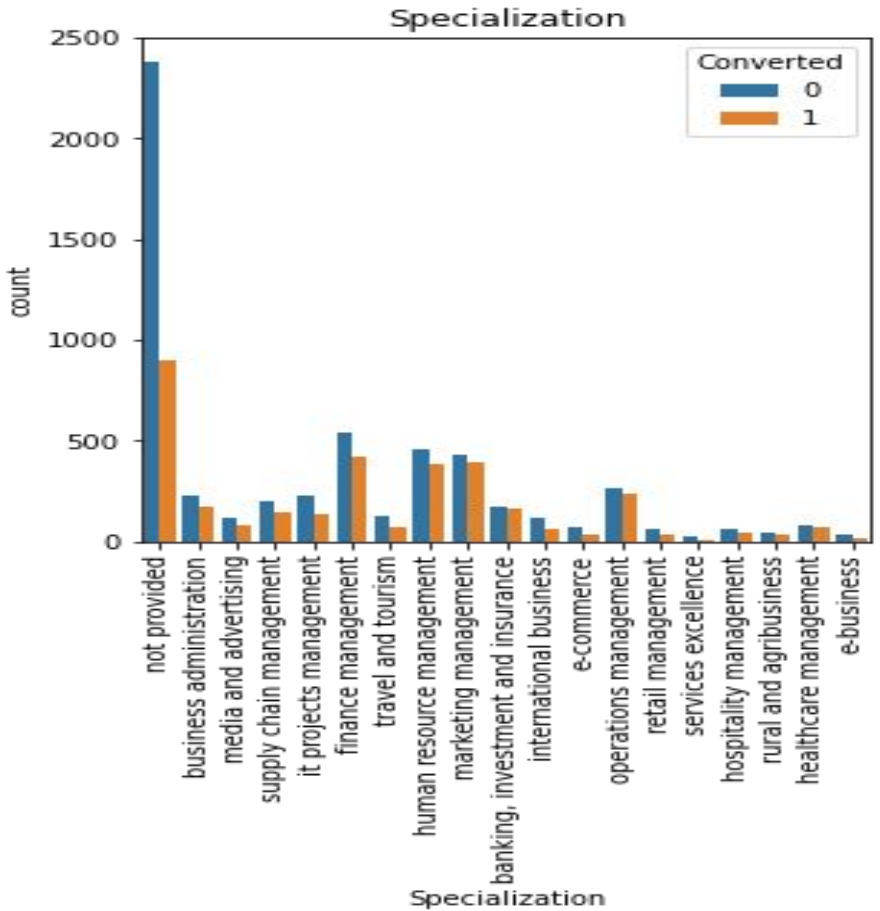
Major conversion in the lead source is from Google

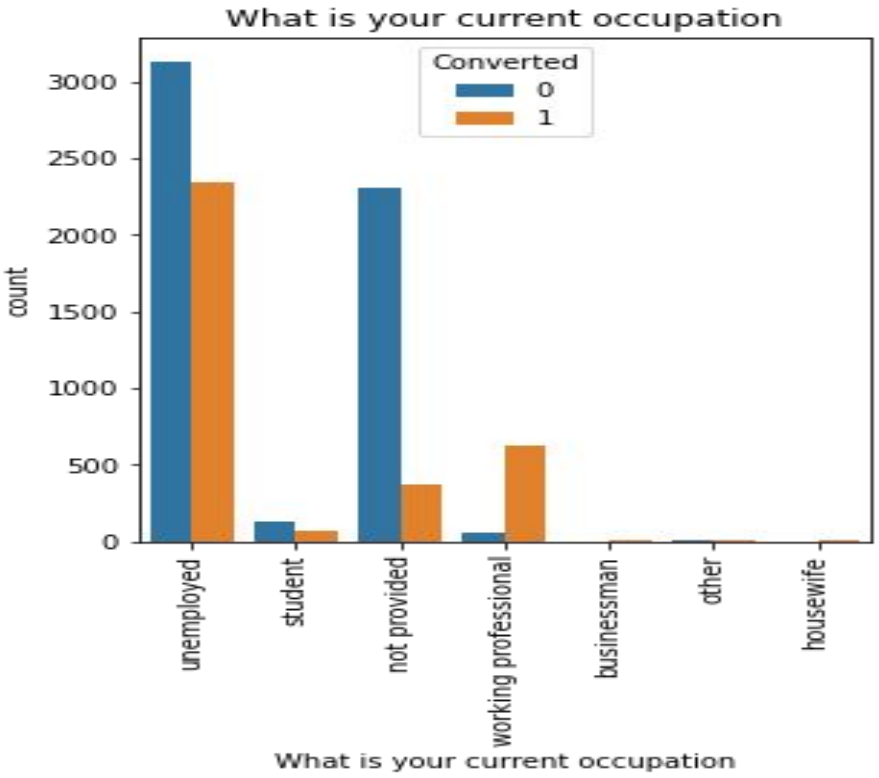


Mainly choosing course for better career Prospects

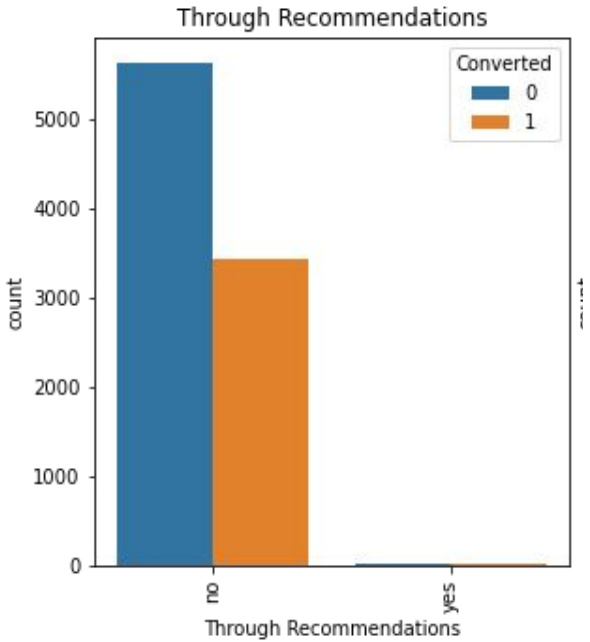


More conversion happened with people who are unemployed

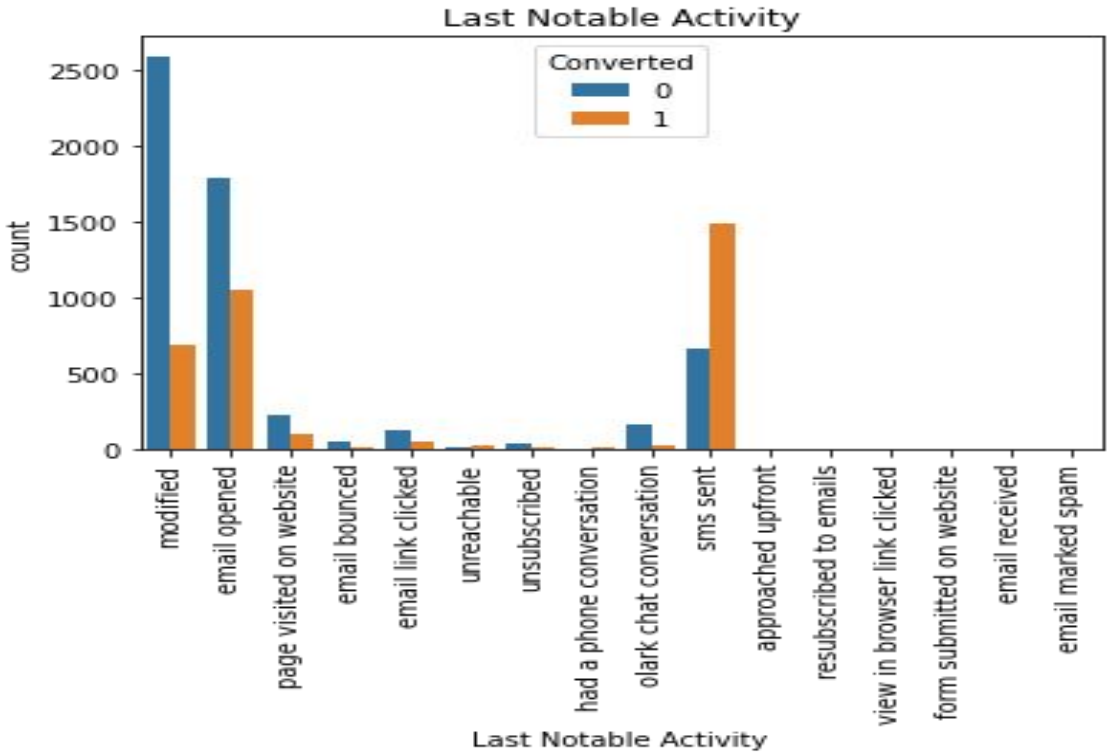




➢ **Not much impact on conversion rates through recommendations**



➢ **Last Activity value of SMS Sent' had more conversion.**

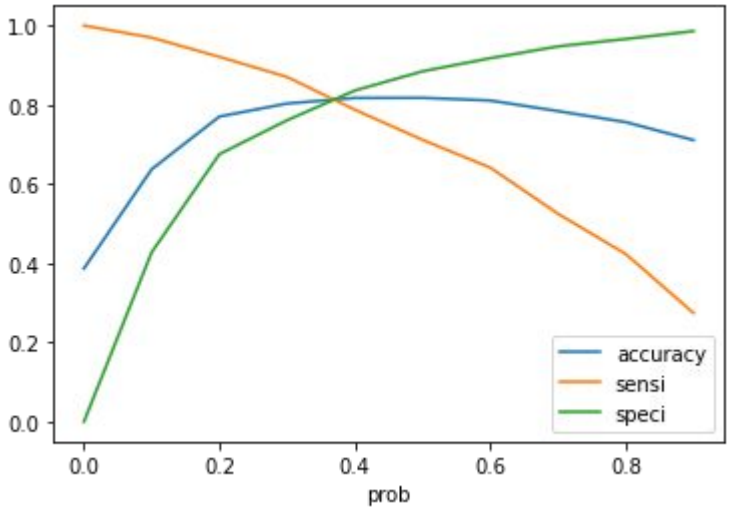


Variables Impacting the Conversion Rate

- 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

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



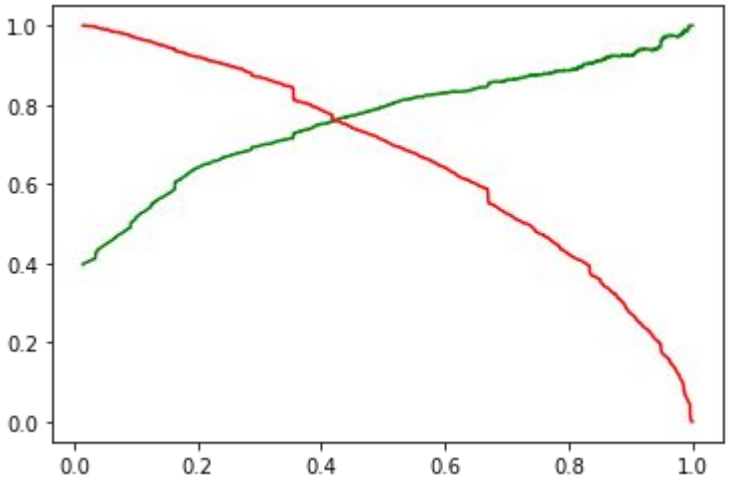
Confusion Matrix

3065	**830**
379	**2077**

- Accuracy - 80%
- Sensitivity - 79 %
- Specificity - 78 %

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

3273	**622**
541	**1915**

- Precision - 74%
- Recall - 76 %

Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix

1474	270
214	765

- Precision - 73%
- Recall - 75%

Conclusion :

- The best cutoff based on sensitivity and specificity was taken into consideration when determining the final prediction, even though we also examined precision and recall metrics and sensitivity-specificity metrics.
- Accuracy, Sensitivity and Specificity values of test set are around 80% which are approximately closer to the respective values calculated using trained set.
- On top of that, the lead score assessment indicates that the end predicted model's conversion rate is approximately 81% (in the train set) and 80% (in the test set).
- In the model, the top three factors that influence a lead's conversion are:
  - Total time spent on the website
  - Lead Add Form from Lead Origin
  - Had a Phone Conversation from Last Notable Activity.
- As a result, this model appears to be **ideal.**