# Lead Score Case Study

**Submitted by :**

Radhay Anand

Onkar Gaikwad

Yashwanth Pallapu

# Lead Score Case Study for X Education

## Problem Statement :

X Education sells online courses to industry professionals. The company promotes its courses through various websites and search engines such as Google.

When visitors arrive at the website, they can browse the courses, fill out a form for a course, or watch videos. If visitors provide their email address or phone number by filling out a form, they are classified as leads. Additionally, the company acquires leads through past referrals.

Once leads are generated, the sales team initiates contact through calls, emails, and other methods. Some of these leads are converted into customers, while the majority are not. The average lead conversion rate at X Education is approximately 30%.

## Business Goal:

X Education requires assistance in identifying the most promising leads—those most likely to convert into paying customers.

The company seeks to implement a model that assigns a lead score to each prospect, ensuring that leads with higher scores have a greater likelihood of conversion, while those with lower scores have a lesser likelihood.

The CEO has set a target for the lead conversion rate to be approximately 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 in Test data based on the Sensitivity and Specificity Metrics.

# Problem solving methodology

## Data Sourcing , Cleaning and Preparation

- 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 and Splitting Train and Test Sets

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

## Model Building

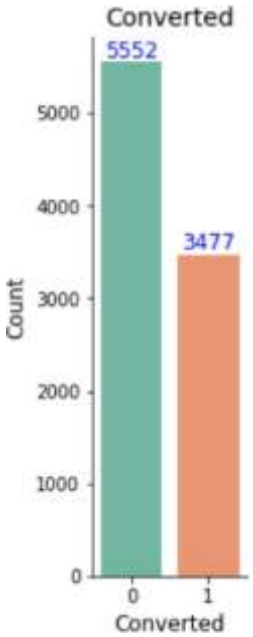
- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model.

## Result

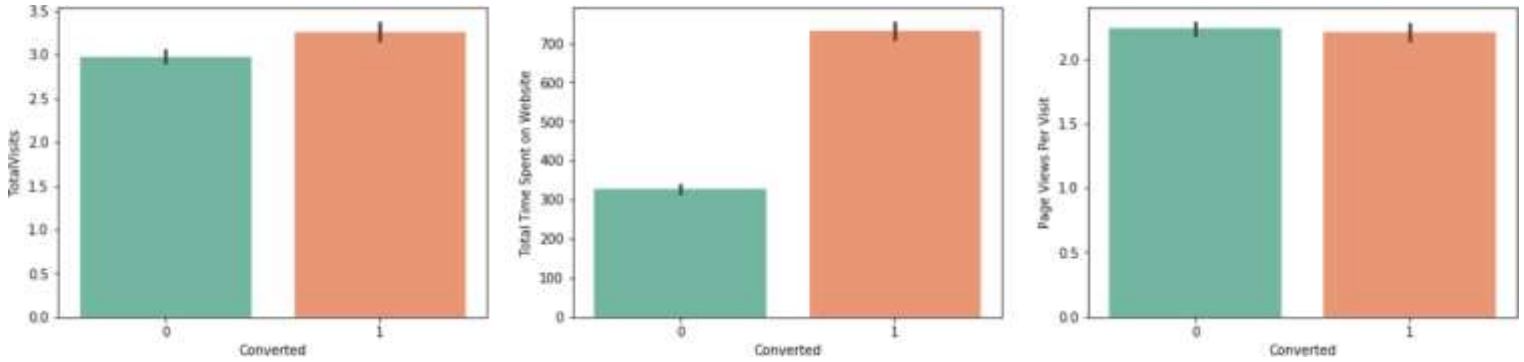
- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics

# Exploratory Data Analysis

We have around 39% Conversion rate in Total

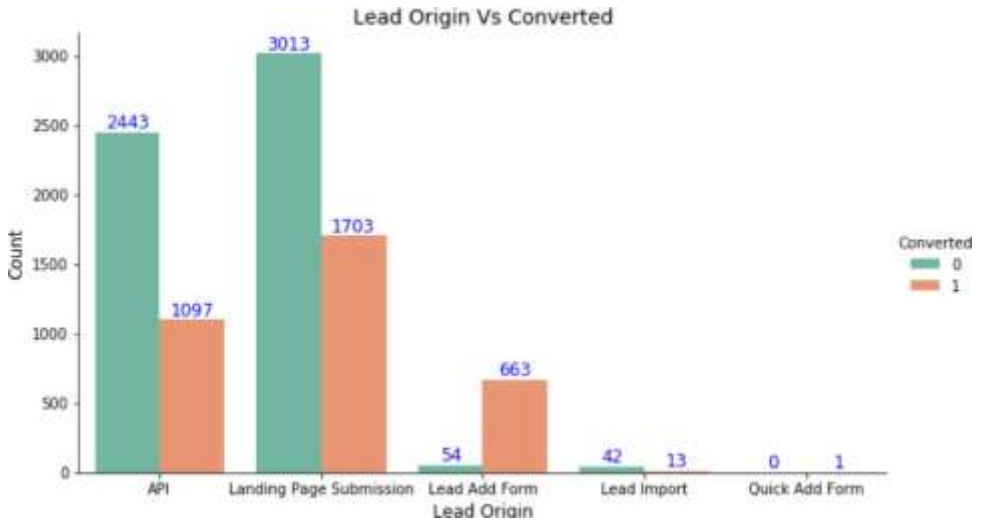



The conversion rates were high for Total Visits, Total Time Spent on Website and Page Views Per Visit

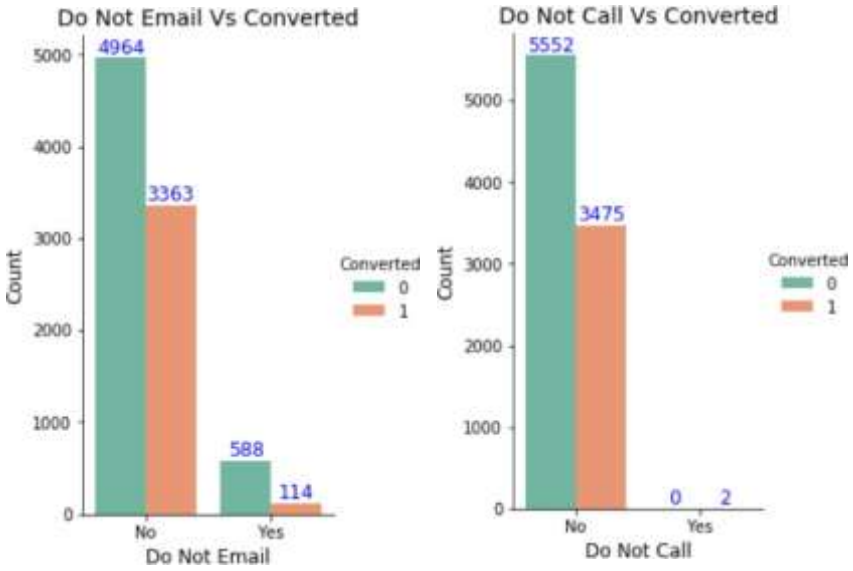



.

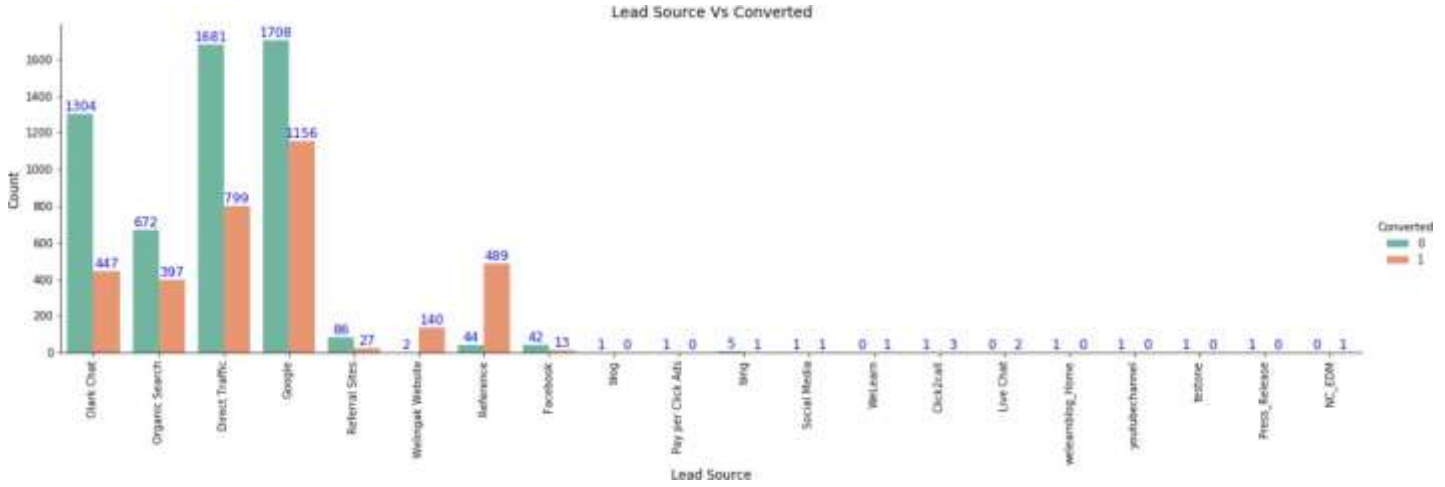
In Lead Origin, maximum conversion happened from Landing Page Submission



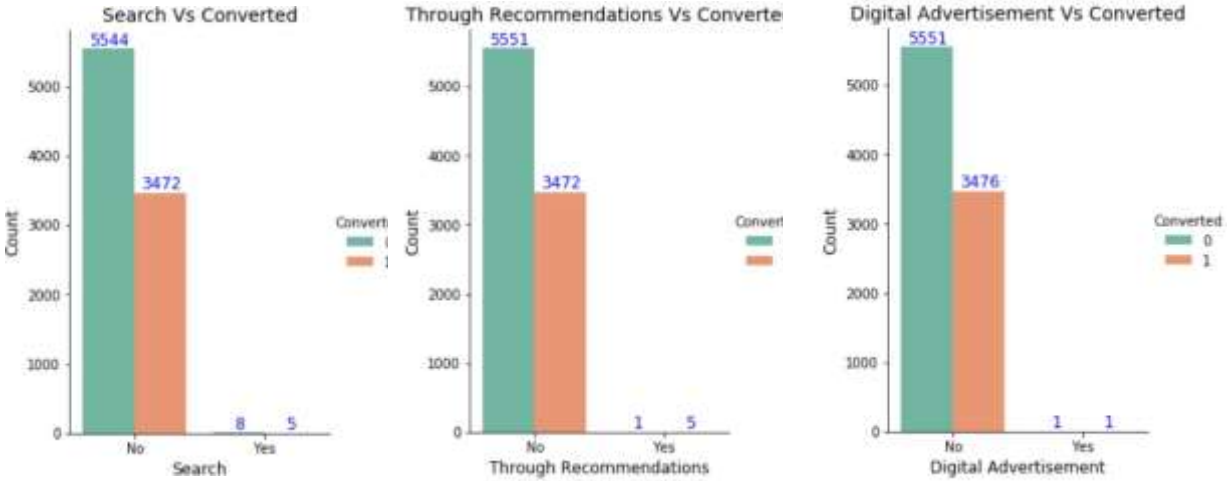

Major conversion has happened from Emails sent and Calls made



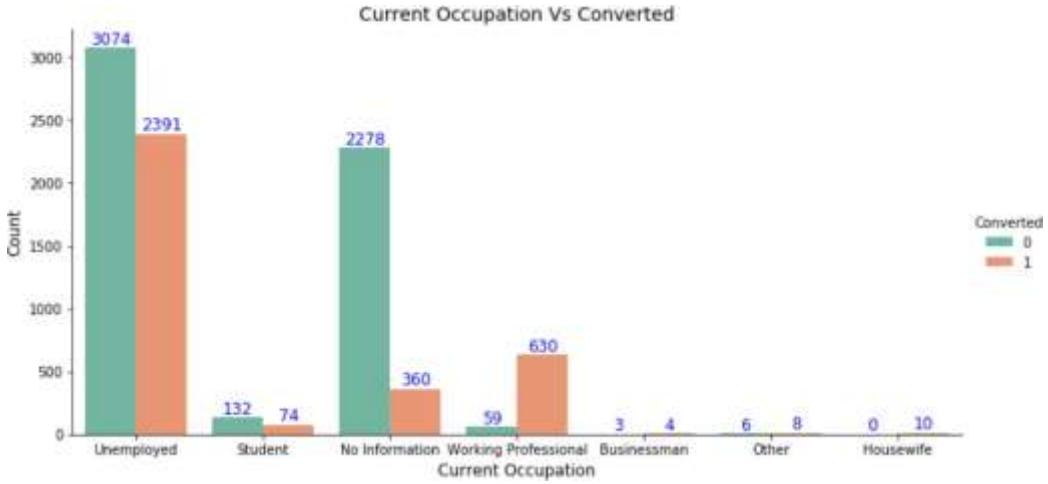

Major conversion in the lead source is from Google



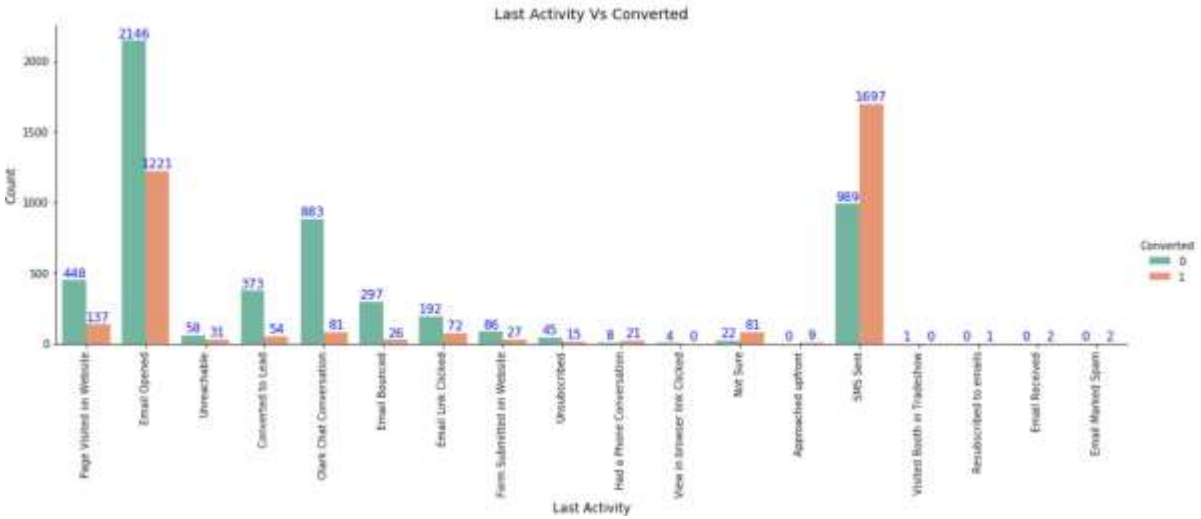
Not much impact on conversion rates through Search, digital advertisements and through recommendations




More conversion happened with people who are unemployed




Last Activity value of SMS Sent' had more conversion.

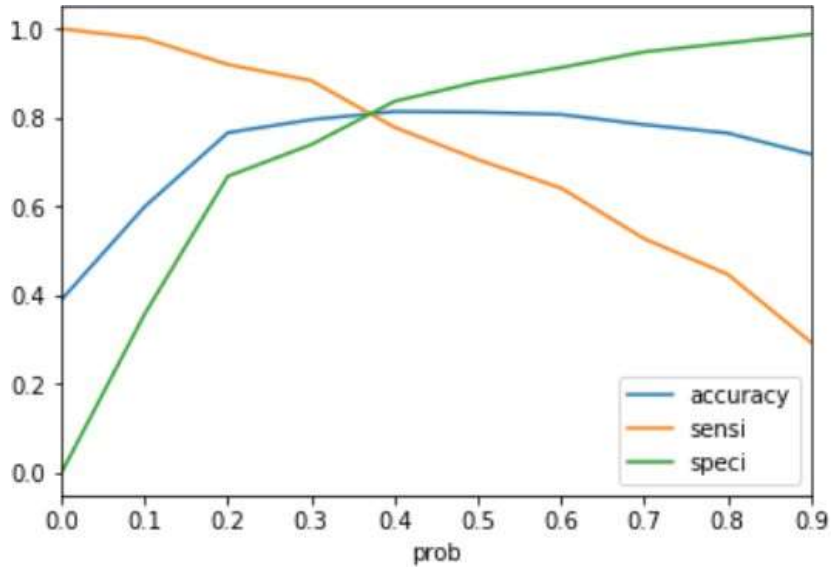


# Variables Impacting the Conversion Rate

- 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 0.37 based on Accuracy, Sensitivity and Specificity



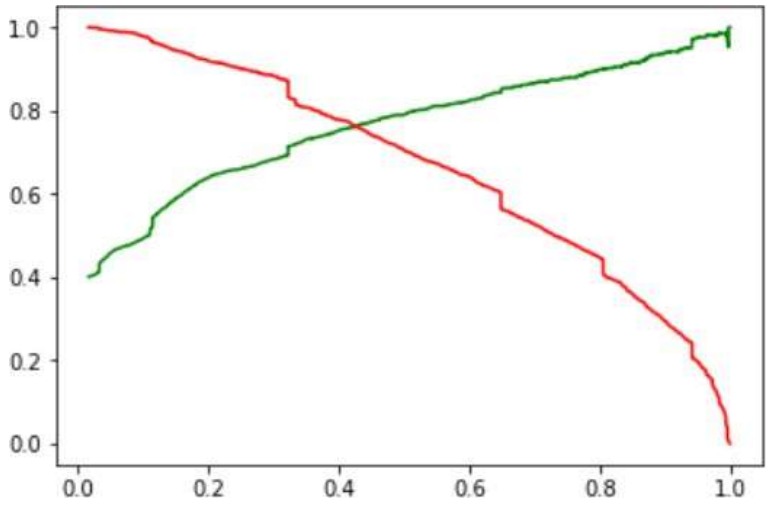

Confusion Matrix

| | |
|---|---|
| 3161 | 697 |
| 974 | 1965 |

- 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 0.42 based on Precision and Recall



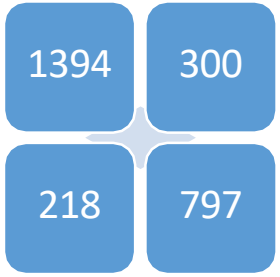

Confusion Matrix

| | |
|---|---|
| 3397 | 461 |
| 725 | 1737 |

- Precision - 79 %
- Recall - 71 %

# Model Evaluation – Sensitivity and Specificity on Test Dataset

Confusion Matrix




- Accuracy - 81 %
- Sensitivity - 79 %
- Specificity - 82 %

# Conclusion

- While we have evaluated both Sensitivity-Specificity and Precision-Recall metrics, we determined the optimal cutoff for the final prediction based on Sensitivity and Specificity.

- The accuracy, sensitivity, and specificity values for the test set are approximately 81%, 79%, and 82%, respectively, which are close to the values obtained from the training set.

- The calculated lead score indicates that the conversion rate for the final predicted model is around 80% on the training set and 79% on the test set.

- The top three variables contributing to lead conversion in the model are:

  - Total time spent on the website
  - Lead Add Form from Lead Origin
  - Had a Phone Conversation from Last Notable Activity

- Overall, this model appears to be effective.