X Education - Lead Scoring Case study

X Education

X Education is an online learning portal which concentrates on selling online courses to industrial professional. They have considerable amount of website footfall on any given day. They also engage third party ad marketing players like google to promote their courses. As a part of marketing campaign, they also collect the visitor contact information and their interest in the courses offered by them.

The collected information are used by the sales team to contact the leads and promote their paid courses to increase their revenue.

Business Goal

Improving the efficiency in Lead conversion probability by identifying the potential customers for sales team.

A Machine learning model to identify the potential leads and assign lead score to each of the leads. The highest score for high probable customers interested in the paid courses and lowest score for low probable customer with less chance of conversion

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

Analysis Flow

Data Cleaning

EDA

Data Preparation

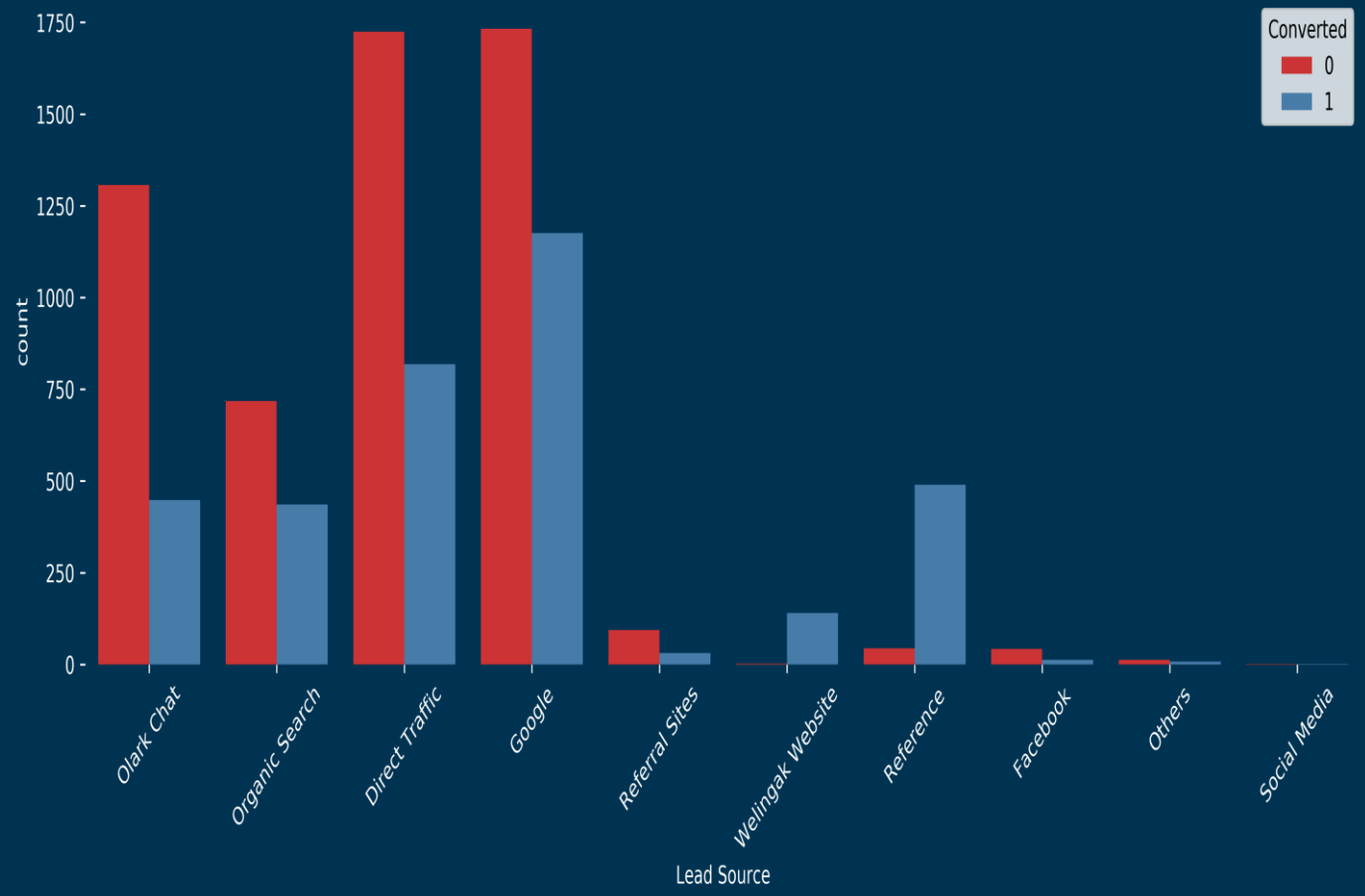
Model Building

Model Validation

Predictions

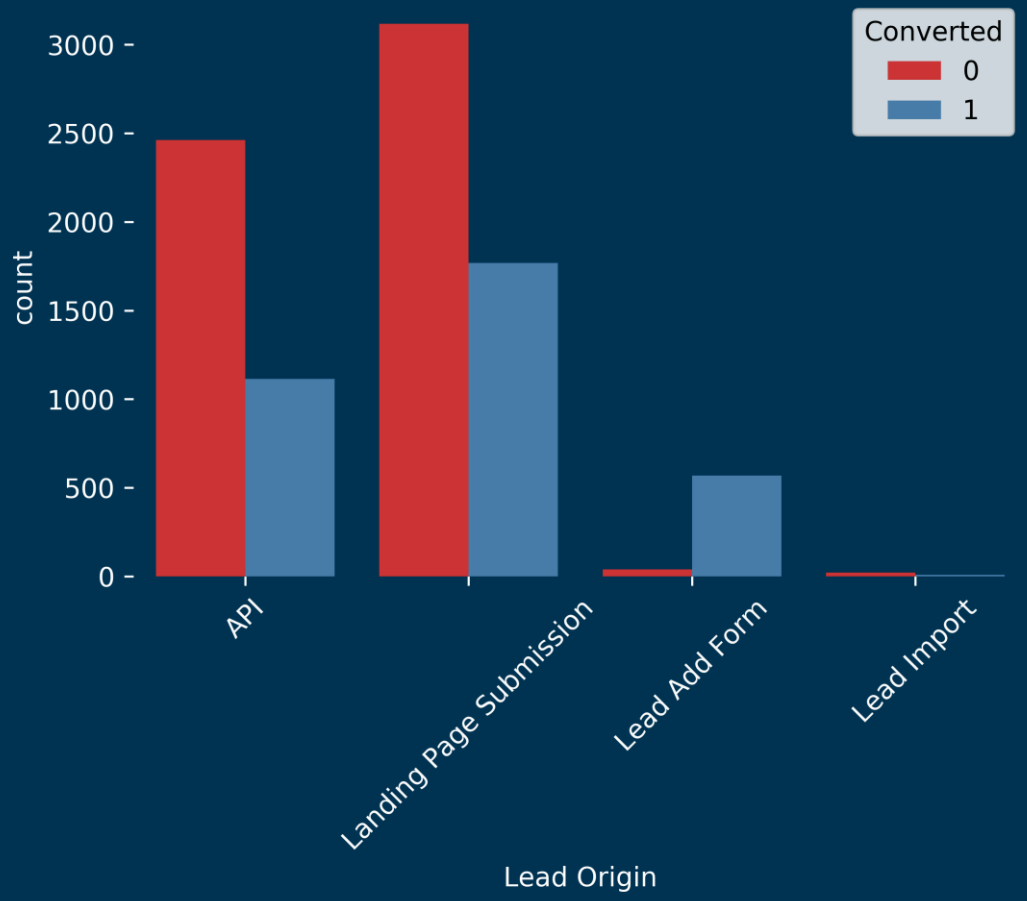
Recommendations

EDA - Lead Source



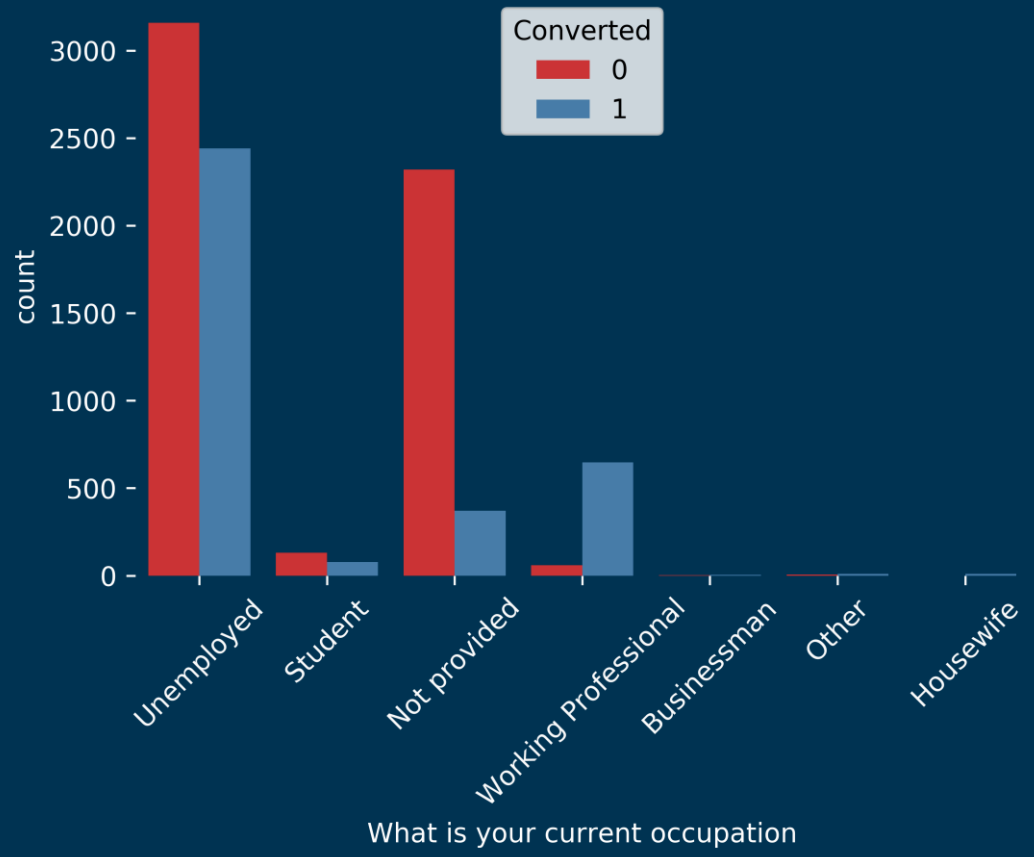
Maximum Leads are generated by Google and Direct Traffic. Conversion rate of Reference leads and Welinkgak Website leads is very high.

EDA - Lead Origin



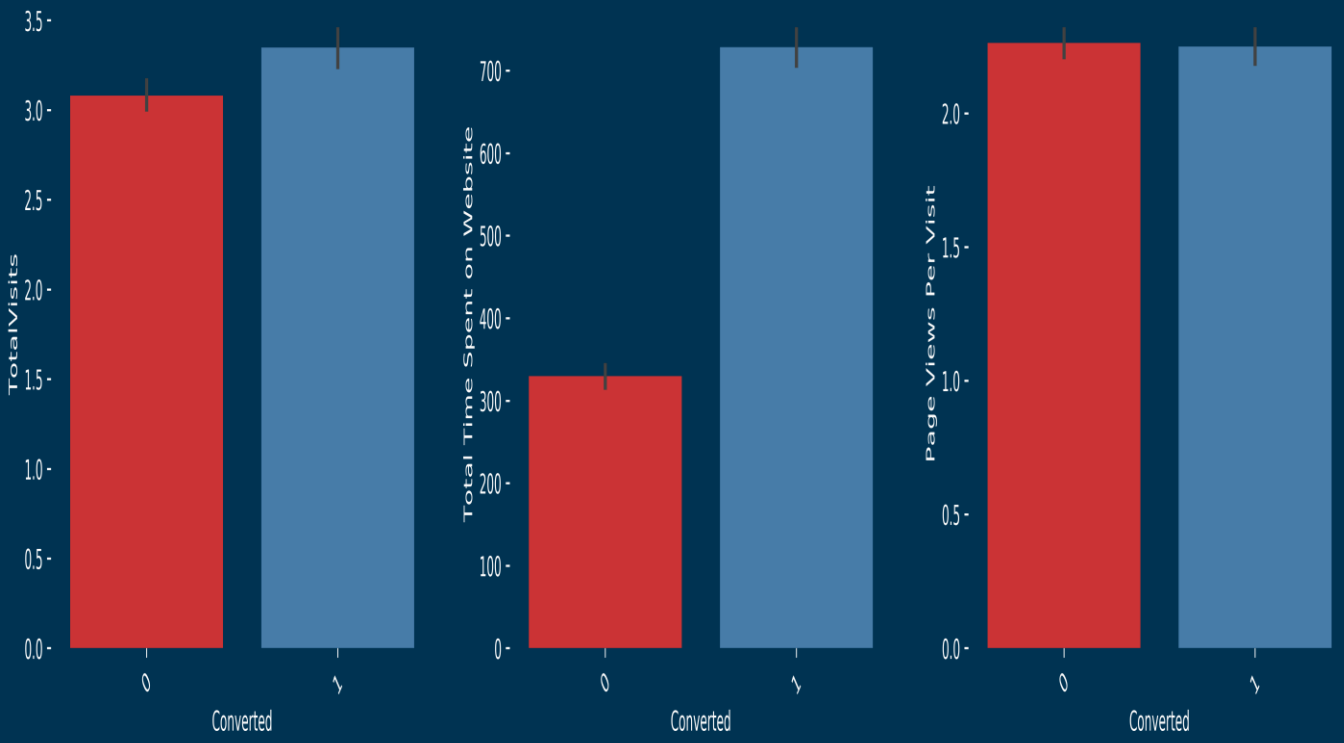
Majority of Leads are coming from API and Landing Page submission, However the Lead Add form has the highest conversion efficiency.

EDA - Lead Occupation



Maximum leads generated are unemployed and their conversion rate is more than 50%. Conversion rate of working professionals is very high.

EDA - Numerical Data



The conversion rate is high for Total Visits, Total Time Spent on Website and Page Views Per Visit.

Model Building

Features selected for building Model based on EDA

1. Lead Source
2. Do Not Email
3. Converted
4. TotalVisits
5. Total Time Spent on Website
6. Page Views Per Visit
7. What is your current occupation

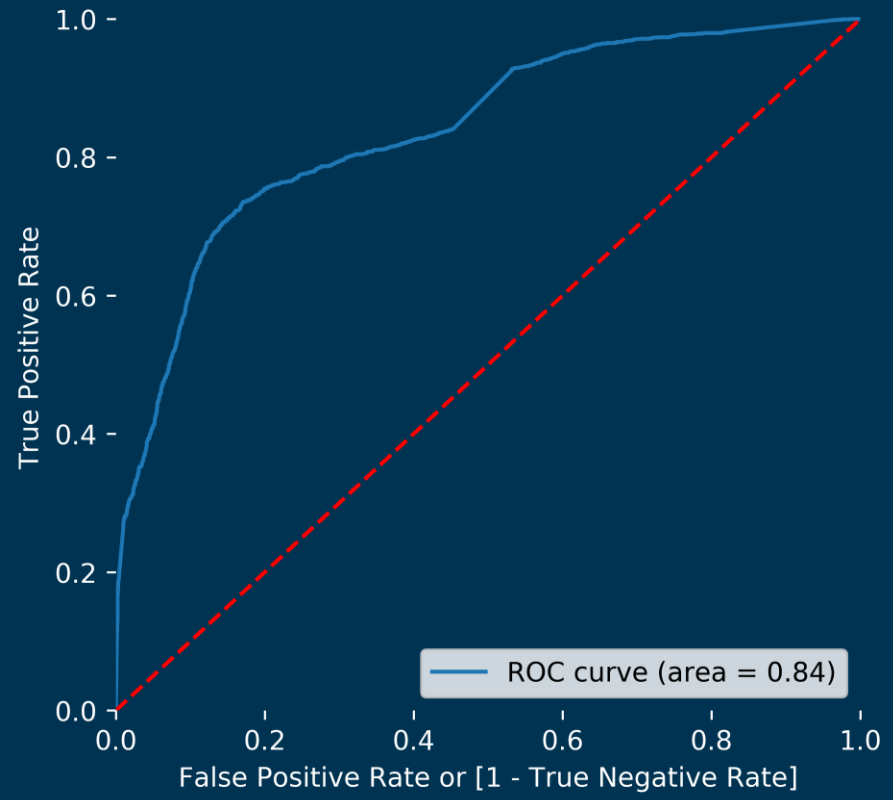
Manual Feature Reduction process was used to build models by dropping variables based on p – value

Model 5 looks stable after five iteration with p-values within the threshold (p-values < 0.05) and No sign of multicollinearity with VIFs less than 5

The model 5 will be the final model, and it will be used for Model Evaluation and prediction

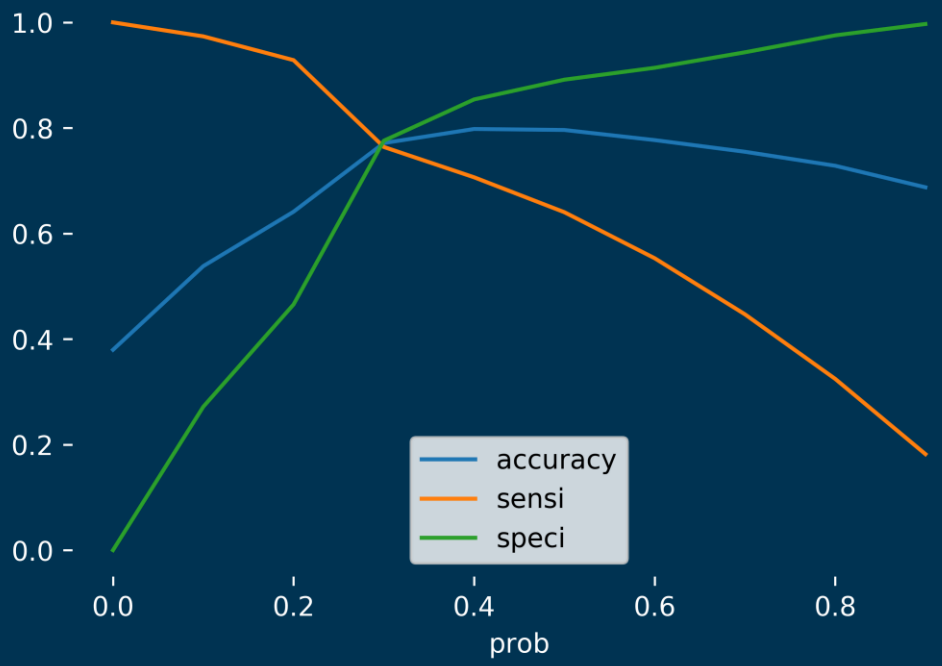
Model Evaluation

Receiver operating characteristic example



The ROC curve with value of 1 indicates the good predictability of the model

Model Evaluation



From the plots, 0.3 can be taken as optimum point to take it as cutoff probability

Recommendation

Based on the test data with Accuracy : 77.52% Sensitivity :76.39% Specificity : 77.58%, the recommendation to improve the efficiency of Lead conversion are as follows

1. Sales team should focus on Leads which originates from the Add Form
2. Working professionals are the key leads who can be converted to subscribe for the paid courses
3. People spending more time on the website are more interested to understand the offerings and can be converted as potential lead