

X Education - Lead Scoring Case Study

Increasing lead conversion rate for X education

Background

- ❖ An education company named X-education sells online courses to industry professionals
- ❖ Many interested professionals land on their website
- ❖ The company markets its courses on several websites 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
- ❖ 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%

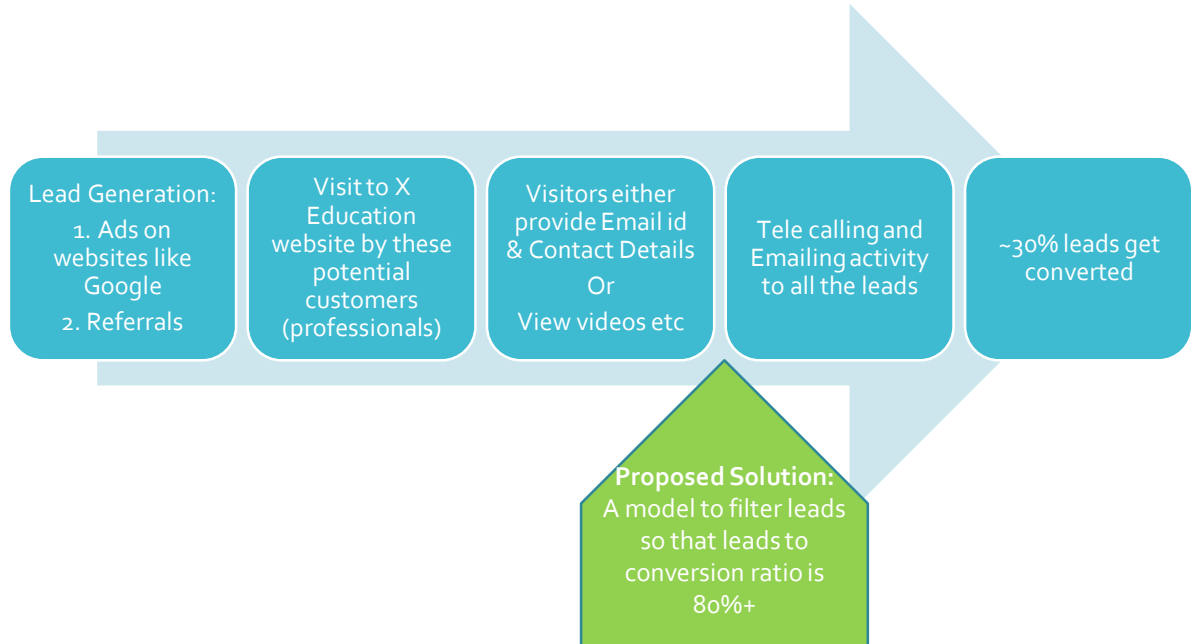
Problem Statement

X Education Company's Problem

- ❖ X Education gets a lot of leads but its lead conversion rate is very poor
- ❖ To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'
- ❖ If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone
- ❖ We will help them to select the most promising leads, i.e. the leads that are most likely to convert into paying customers.
- ❖ We are required to build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance
- ❖ The CEO, in particular, has given a ballpark of the target lead conversion rate to be 80%.

Lead – Conversion Process

Lead to Conversion process



Solution

Selection of Hot Leads

A. Data cleaning and data manipulation:

1. Check and handle duplicate data.
2. Check and handle NA values and missing values.
3. Drop columns, if it contains large amount of missing values and not useful for the analysis.
4. Imputation of the values, if necessary.
5. Check and handle outliers in data.

B. EDA:

1. Univariate data analysis: value count, distribution of variable etc.
2. Bivariate data analysis: correlation coefficients and pattern between the variables etc.

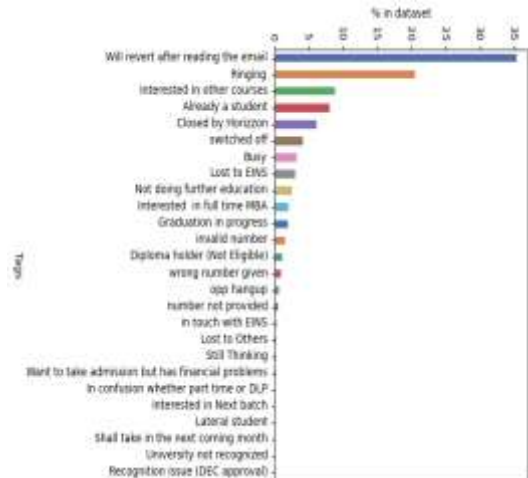
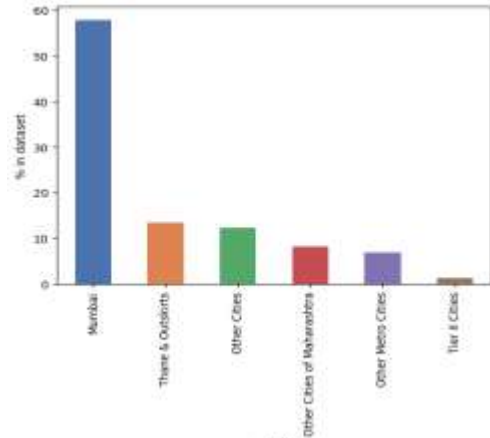
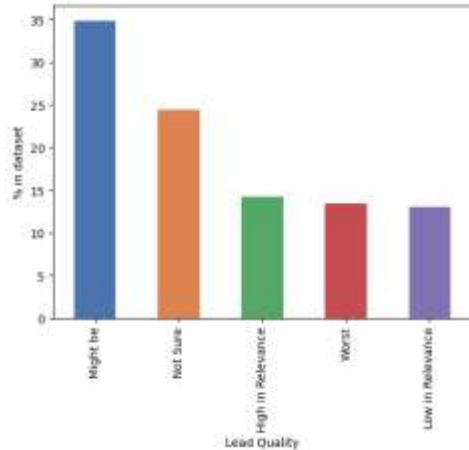
C. Feature Scaling & Dummy Variables and encoding of the data.

D. Classification technique: logistic regression used for the model making and prediction.

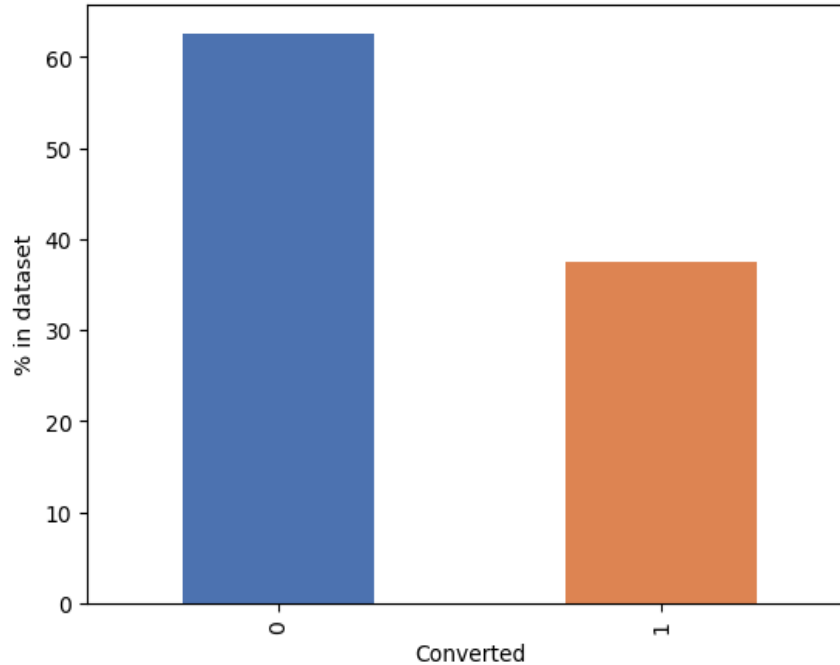
E. Model presentation.

F. Conclusions and recommendations.

EDA

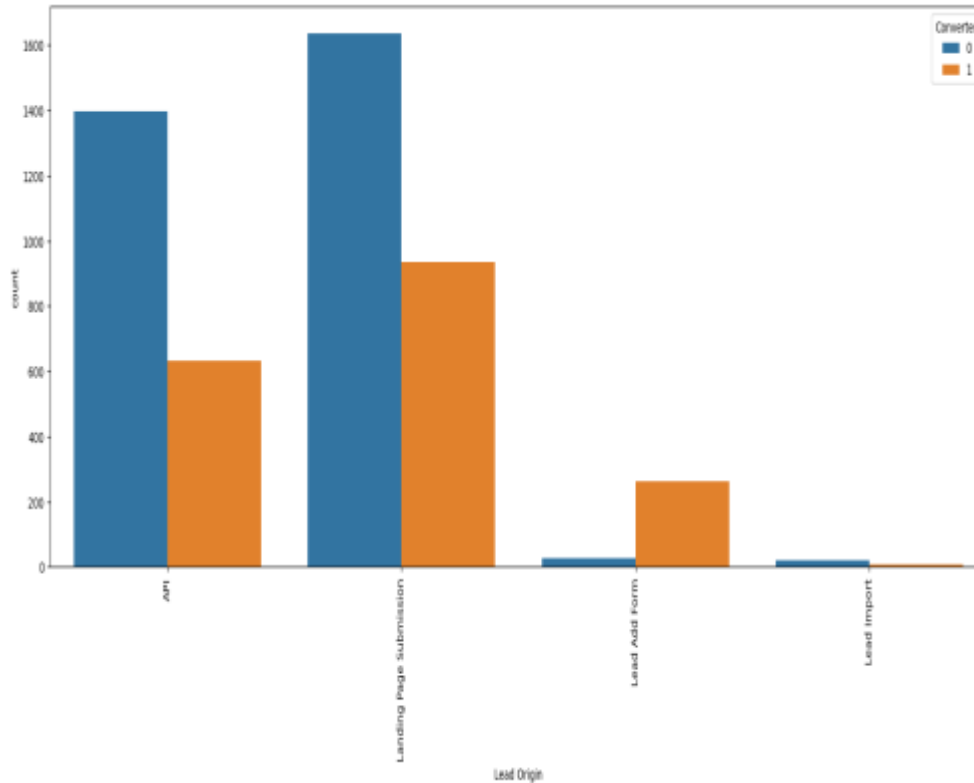


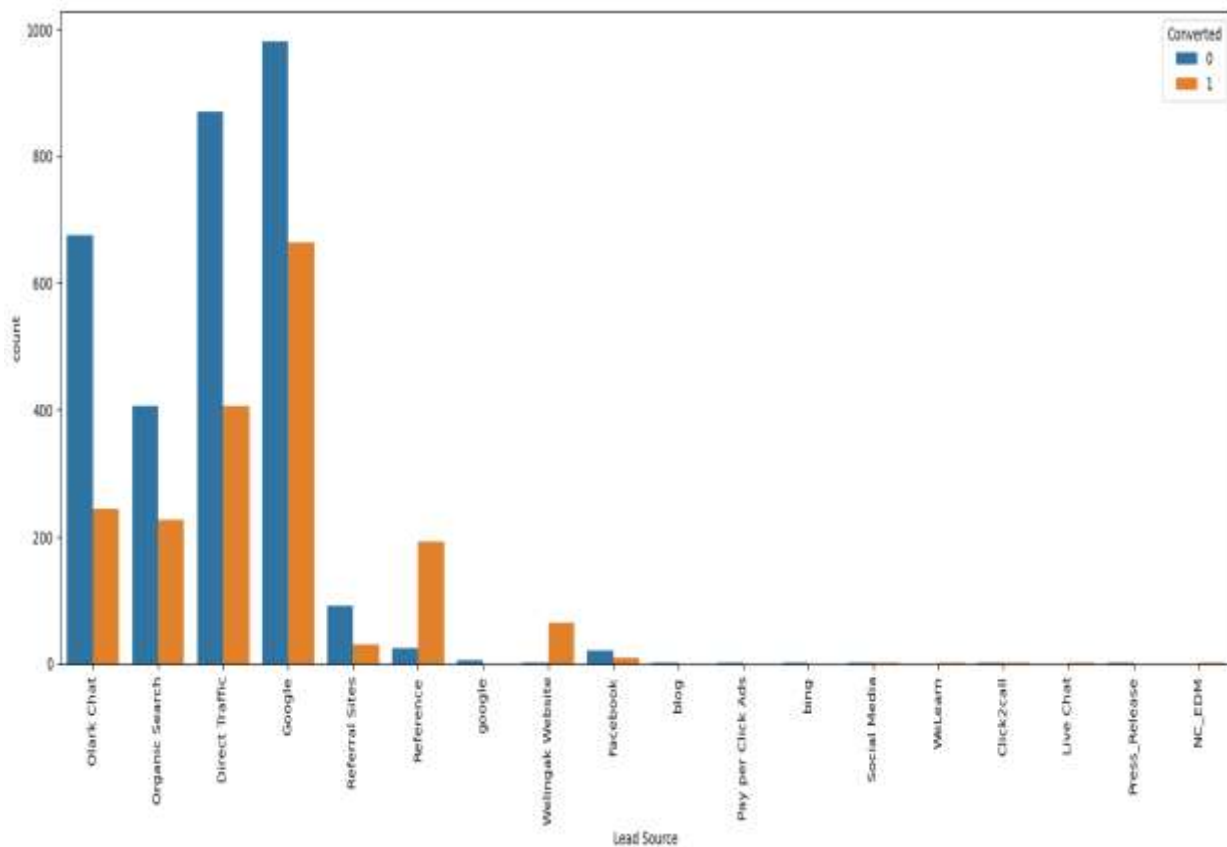
DATA VISUALIZATION - Converted

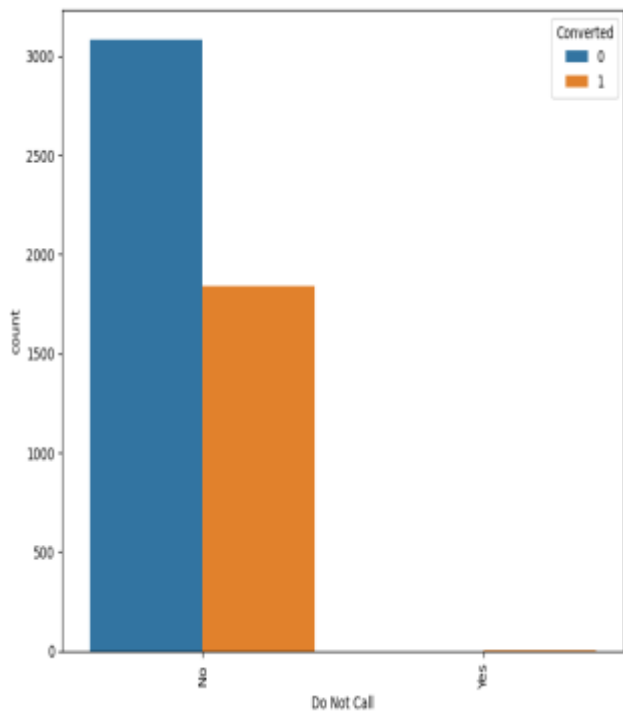
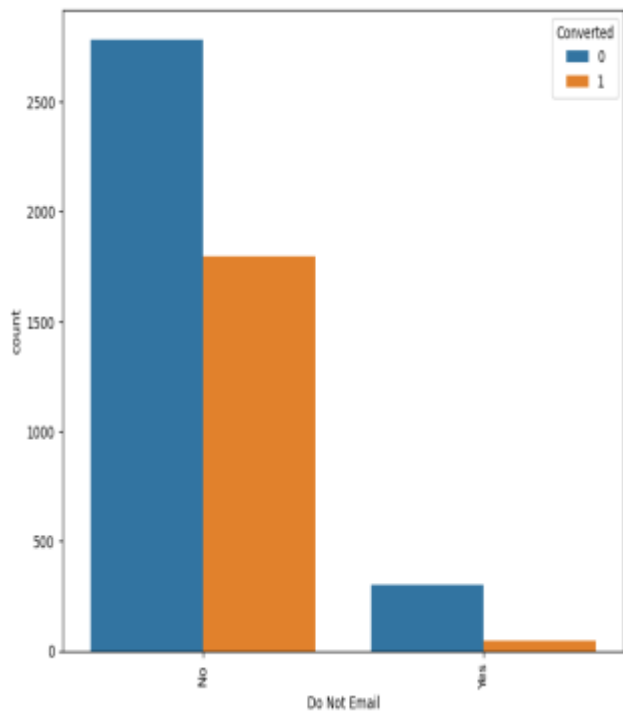


Note: 37.8% of the 'Converted' data is 1 ie, 37.8% of the leads are converted. This means we have enough data of converted leads for modelling.

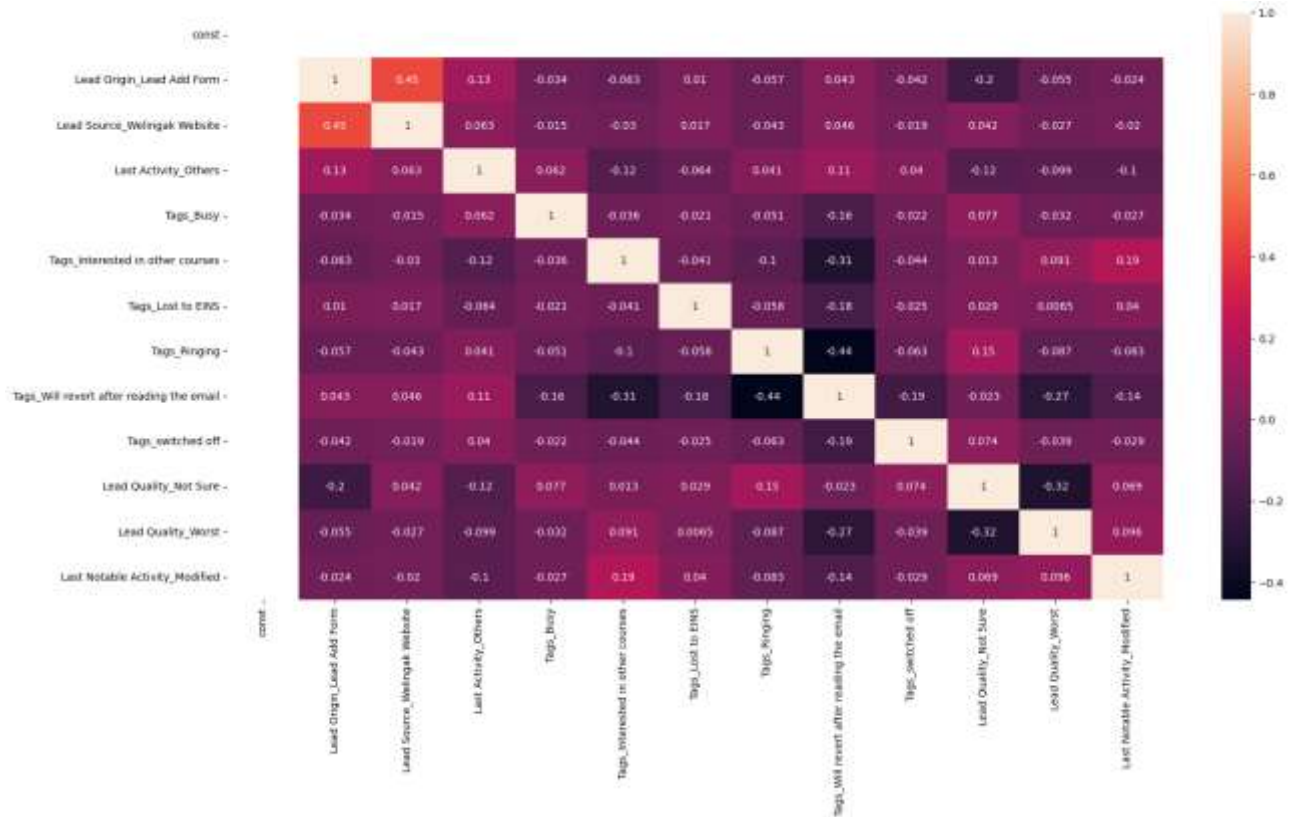
CATEGORICAL VARIABLES







Heatmap

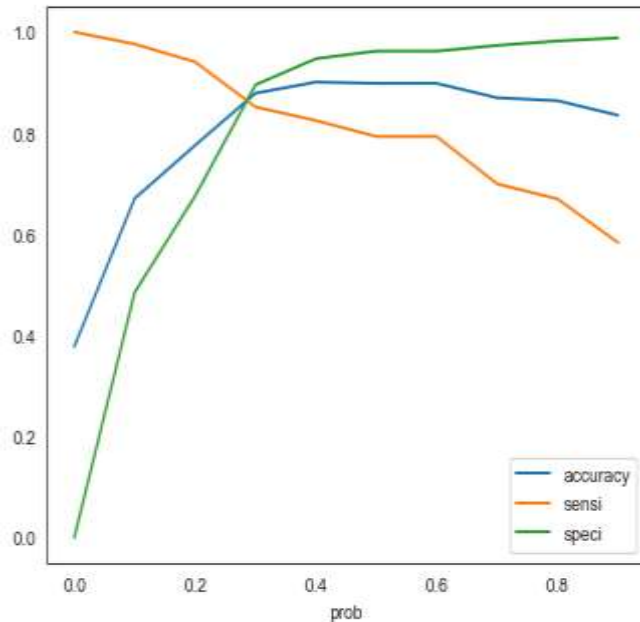
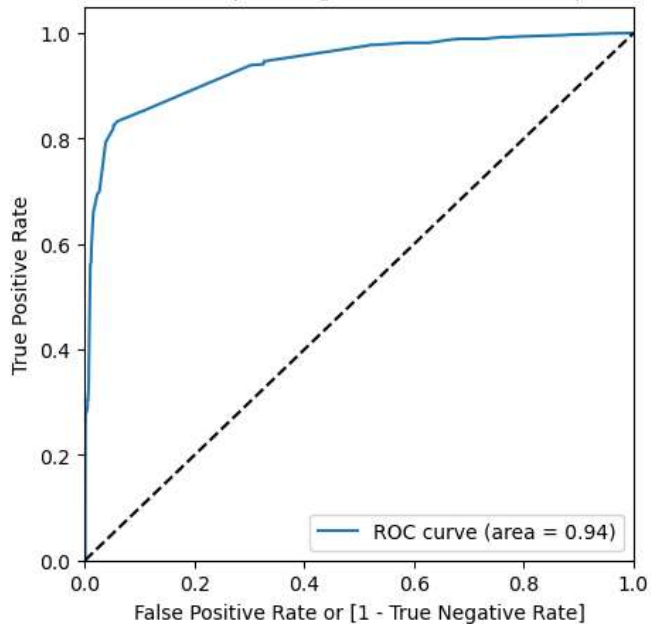


Model Building

- Splitting the Data into Training and Testing Sets.
- The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- Use RFE for Feature Selection.
- Building Model by removing the variable whose p- value is greater than 0.05 and vif value is greater than 5.
- Predictions on test data set.

ROC Curve

Receiver operating characteristic example



Conclusion

It was found that the variables that mattered the most in the potential buyers are:

- a. The total time spend on the Website.
- b. Total number of visits.
- c. When the lead source was: a. Google b. Direct traffic c. Organic search
- d. When the lead origin is Lead add format.
- e. When their current occupation is as a working professional.

The X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

Thank-you