LEAD SCORE CASE STUDY

Ojas Balsotra
Nupur Agarwal
Srijita Chaterjee

Problem Statement

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to assign a lead score 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%.

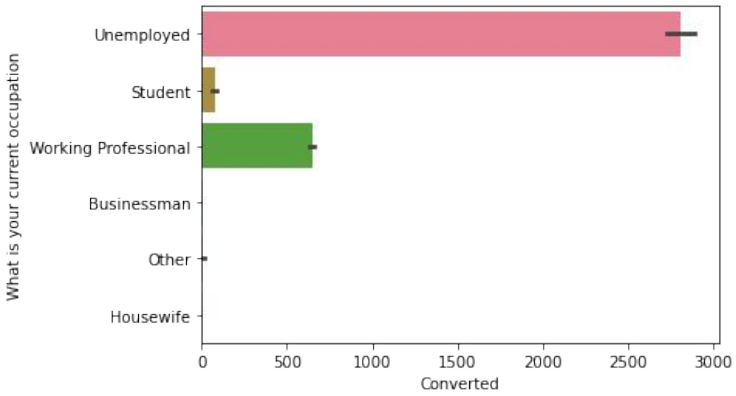
Business Goal

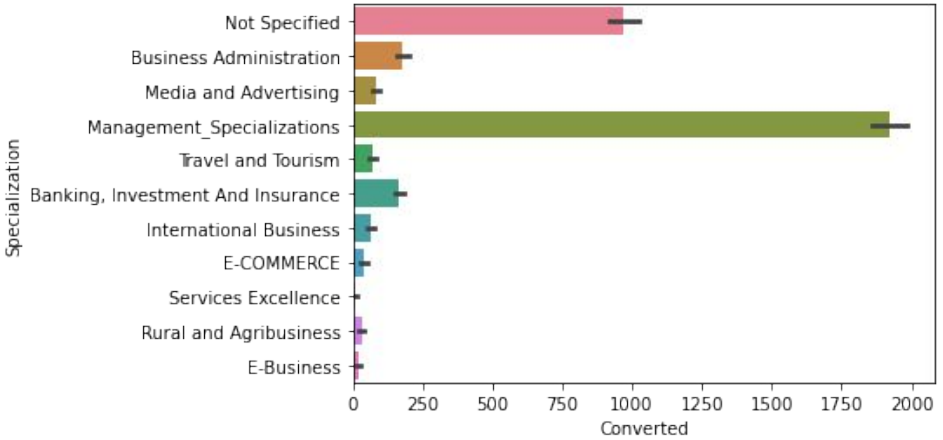
- X Education company wants to identify hot leads.
- Company wishes to identify the most potential leads, also known as “Hot Leads”.
- The company needs a model wherein a lead score is assigned to each of the leads such that the customer with higher lead score have a higher conversion chance and customer with lower lead score have a lower conversion chance.
- The CEO, in particular, has given a ballpark number for the lead conversion rate i.e. 80%.

Solution Approach

1. Importing Data
2. Eda [Exploratory Data Analysis]
   - Data Cleaning & Treatment
3. Feature Scaling And Dummy Variable Creation
4. Train Test Split
5. Logistic Regression Model Building
6. Model Evaluation
7. Precision – Recall
8. Model Evaluation : Specificity , Sensitivity, Precision And Recall

EDA

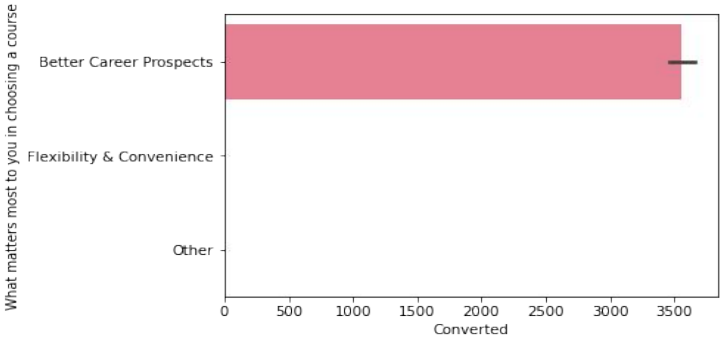


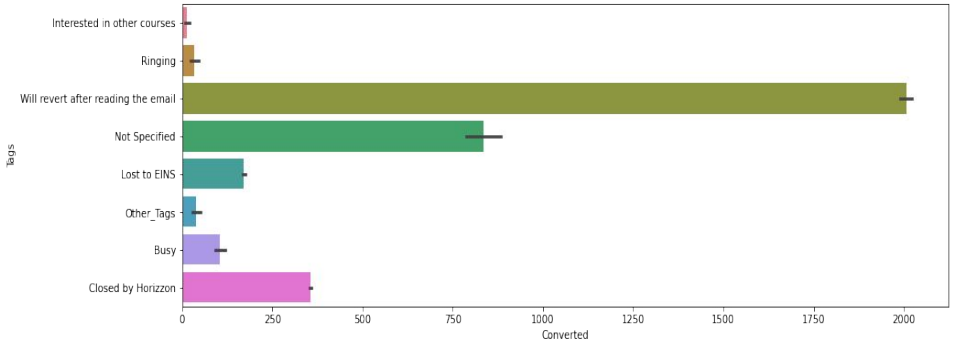


Graphical Analysis Of “What Is Your Current Occupation” Depicts The Unemployed People Are Getting Converted The Most.

Graphical Analysis Of “Specialisation” Depicts Management Specializations Converted The Most.

EDA

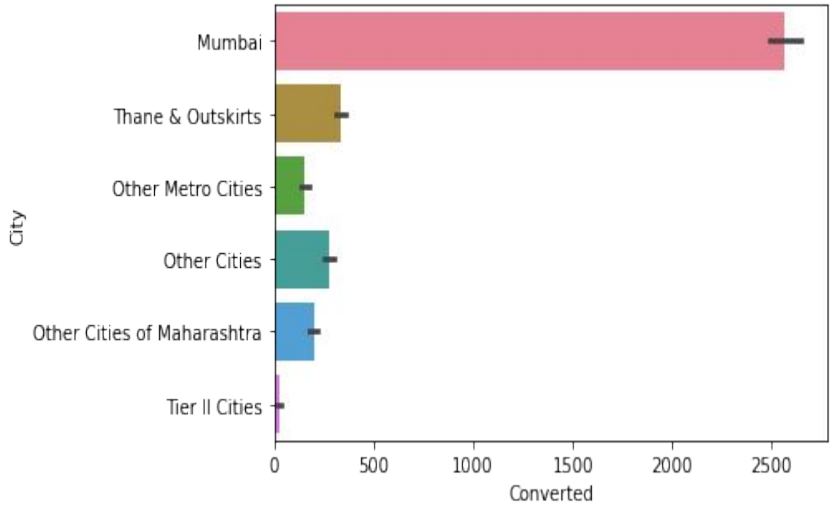




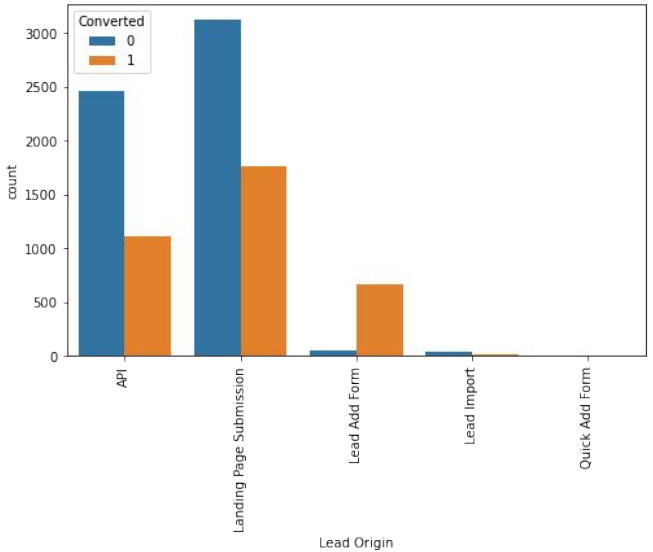
Graphical Analysis Of “What Matters most to you in Choosing a course” Depicts Better Career Prospects with most conversions.

Graphical Analysis Of “Tags” Depicts Will revert after reading the email converted the Most.

EDA

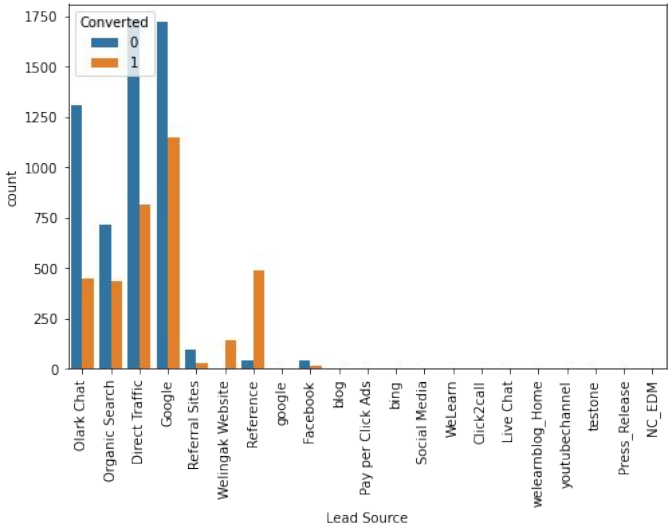


Graphical Analysis Of “City” Depicts Mumbai with most conversions.



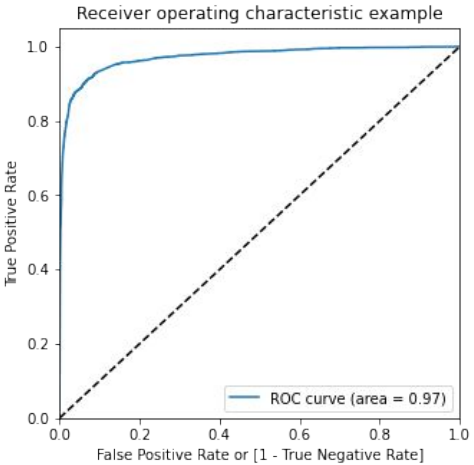
Graphical Analysis Of API and Landing Page Submission bring higher number of leads as well as conversion.

EDA

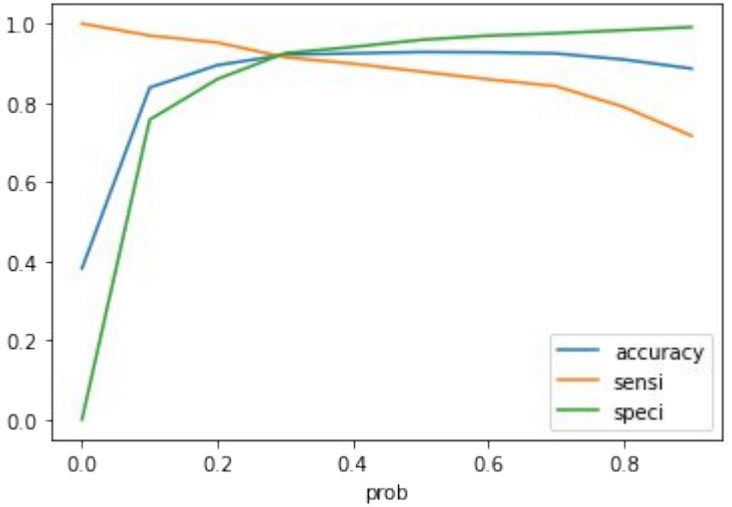


Graphical Analysis Of “Count”
Depicts When the lead source was:
a. Google
b. Direct traffic
c. Organic search
Has the conversion and leads with higher number

MODEL BUILDING



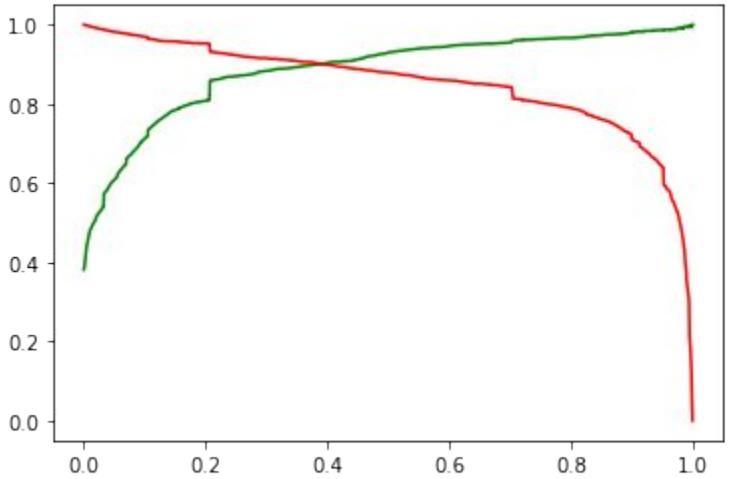
ROC CURVE - 0.97



OPTIMAL CUT-OFF -> 0.3

The ROC curve has a value of 0.97, which is very good.

MODEL BUILDING



PRECISION RECALL CURVE

— MODEL EVALUATION

TRAINING DATA	
Accuracy	92.14%
Sensitivity	91.49%
Specificity	91.49%

TEST DATA	
Accuracy	92.57%
Sensitivity	91.18%
Specificity	93.46%

CONCLUSION

- The logistic regression model is used to predict the probabillty of conversion of a customer.
- In Business terms, this model has capability to adjust with the company's requirements in coming future.
- The variables that were most important to potential buyers are
- The total time which was spent on the Website.
- Total number of visits.
- When the lead source was : - Working Professional and Unemployed.