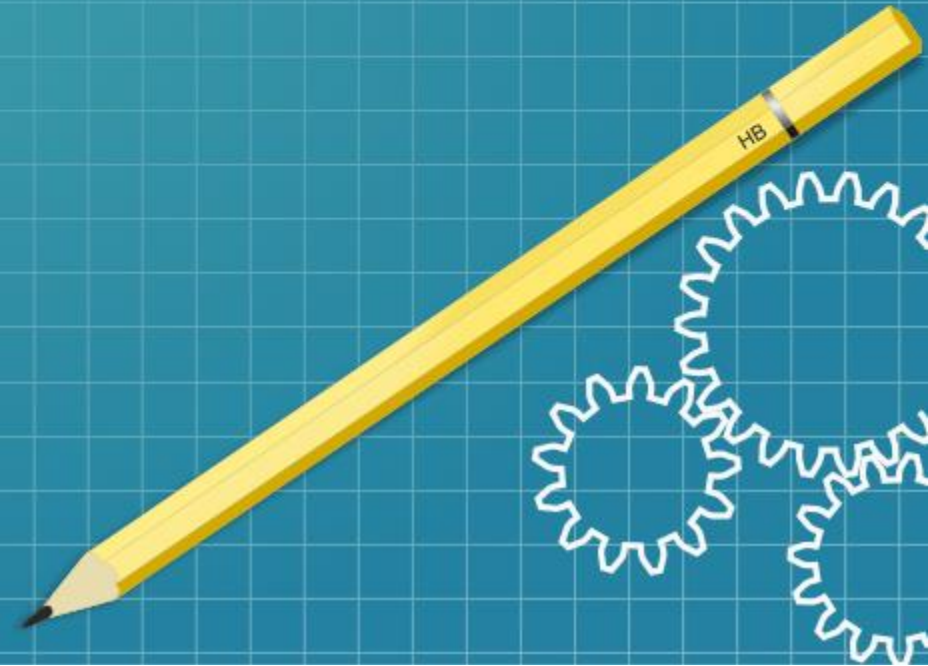# Lead Scoring Case Study Using Logistic Regression

SUBMITTE BY:
ABHISHEK RASTOGI

Content

Problem Statement

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as 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%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. 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

Business Goal/Objective

- Lead X wants us to build a model to give every lead a lead score between 0 -100 . So that they can identify the Hot leads and increase their conversion rate as well.

- The CEO want to achieve a lead conversion rate of 80%.

- They want the model to be able to handle test data that will be fed efficiently.

- Also the actions required to perform, how to utilize full man power and after achieving target what should be the approaches followed to achieve the goal.

Approach Followed

- Importing the required libraries and the dataset

- Data Preparation

- Exploratory Data Analysis EDA

- Univariate/ Bivariate Analysis

- Dummy variable creation

- Train – Test split of the dataset

- Feature scaling

- Correlations

- Model Building (RFE Rsquared VIF and p- values)
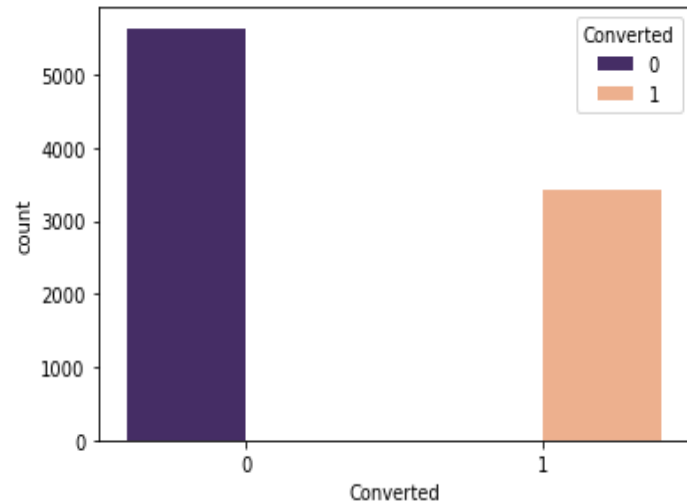
- Model Evaluation

- Making predictions on test set
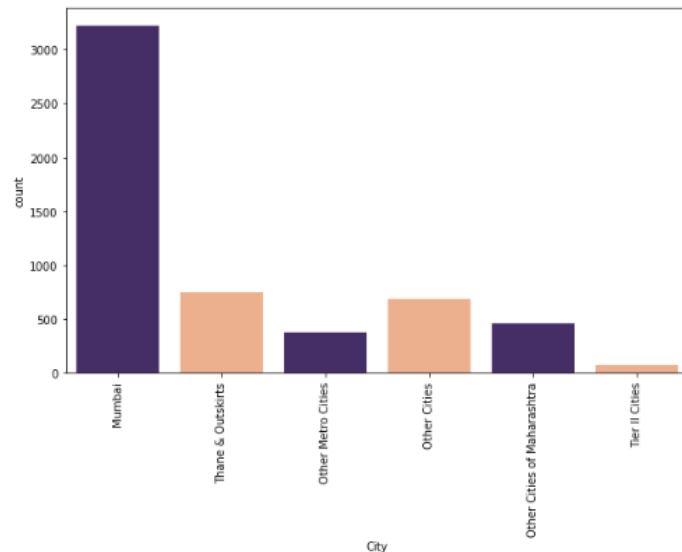
Importing Libraries and dataset

- Importing data cleaning and manipulation libraries like NumPy, Pandas.

- Matplotlib and Seaborn for data visualisation.

- Sklearn for train-test split.

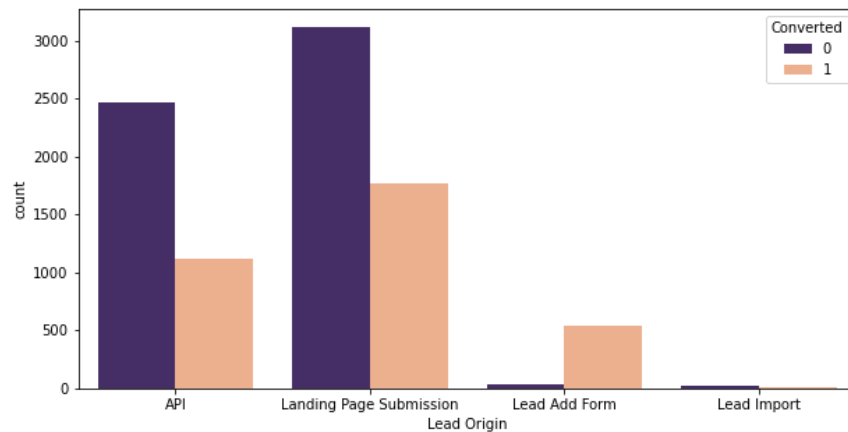- Statsmodels for performing Logisitc Regression

Univariate/Bivariate Analysis

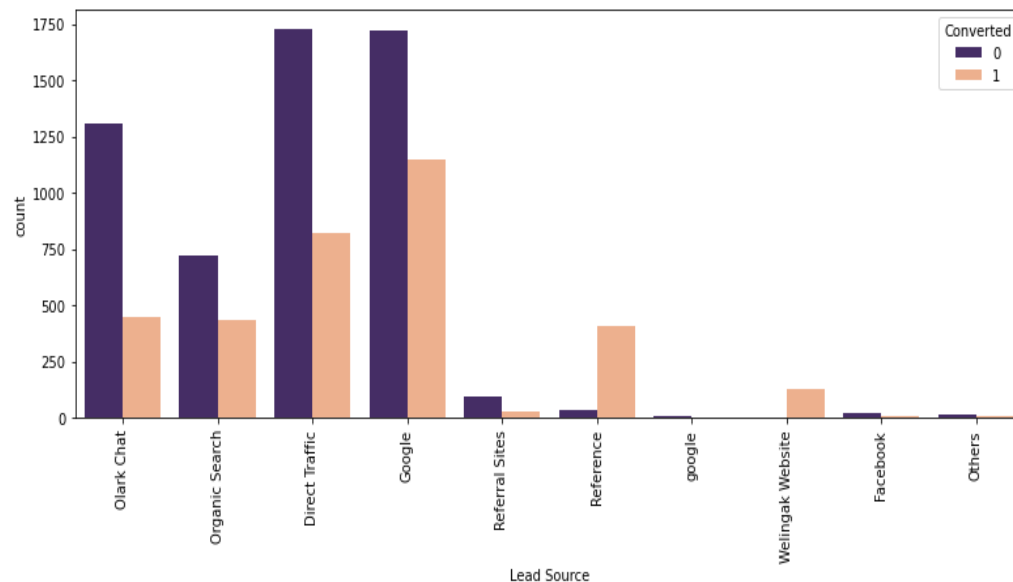The rate of lead conversion is really less as compared to the rate of leads not converted

Most of the leads were coming from Mumbai city.
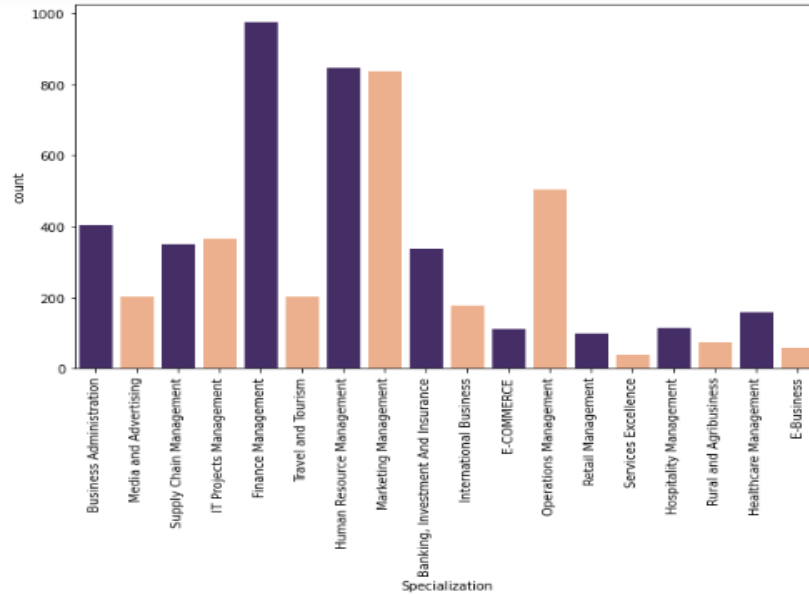
Results from this chart

- 'Landing page submission' is the origin for most of the leads
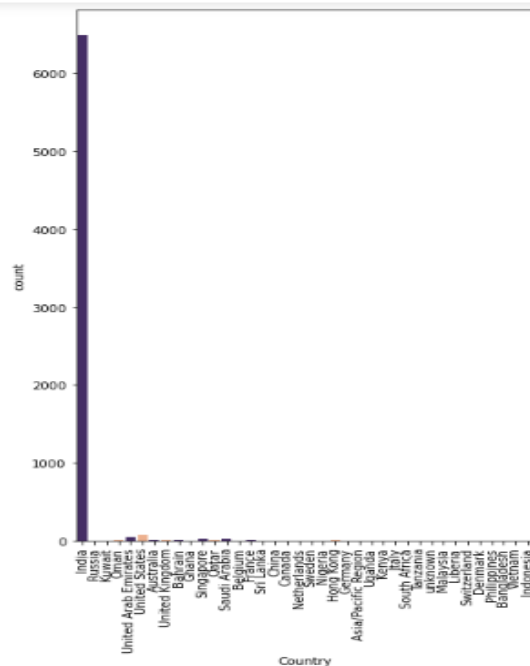- 'Lead Add Form' and 'Lead Import' does not contribute significantly

Leads coming from "Landing Page Submission" has high rate of lead conversion as compared to other types

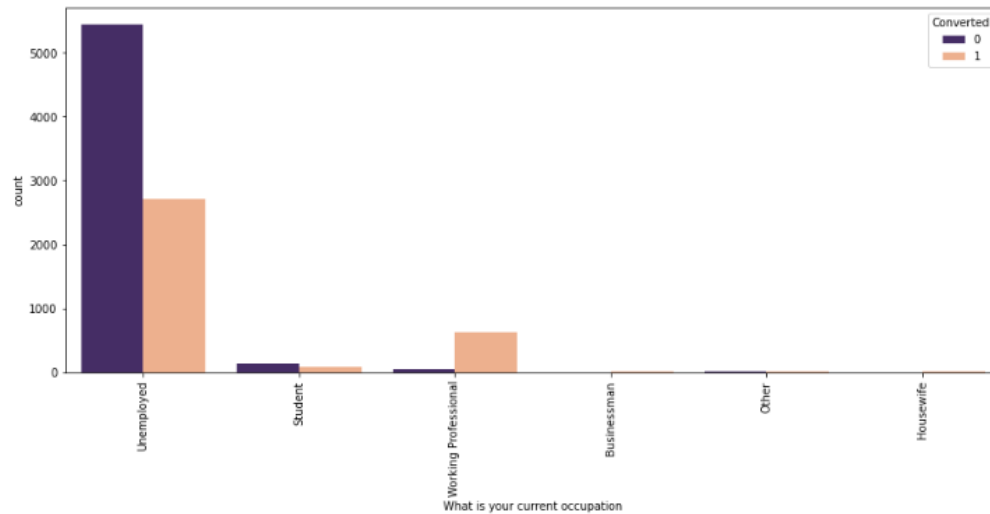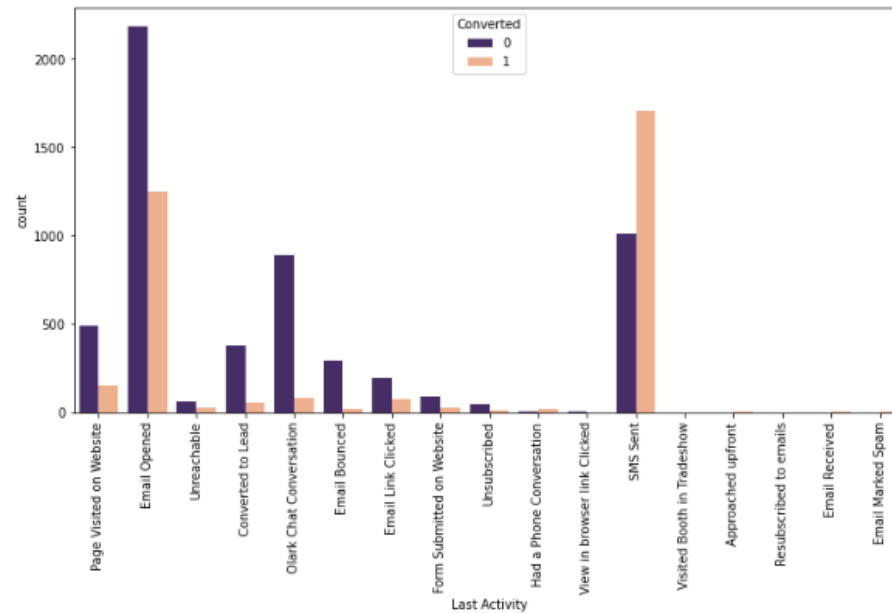Google and Direct Traffic were 2 major sources of bringing potential leads

Most of the leads coming had specialization in "Marketing Management" or "Finance Management

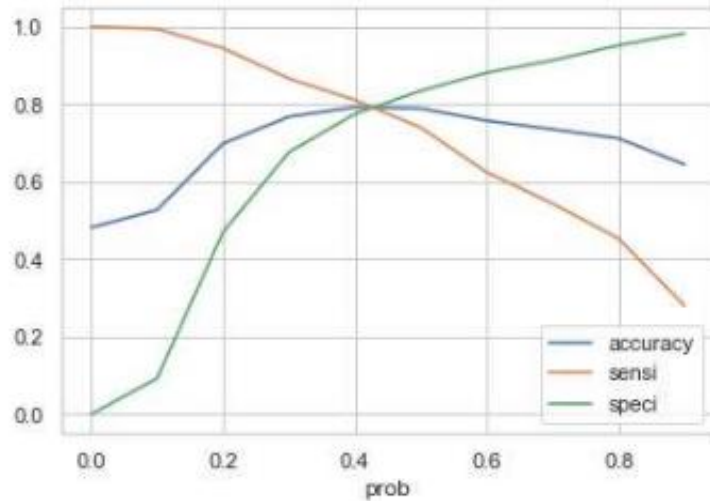Almost all the leads were coming from "India"

Most of the clients were "Unemployed" which is a major finding.

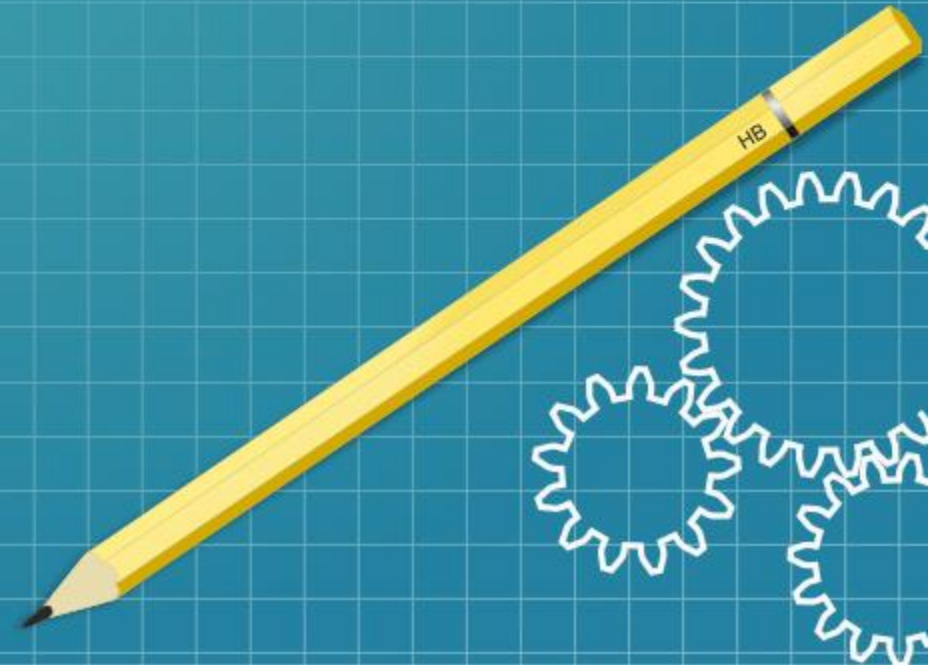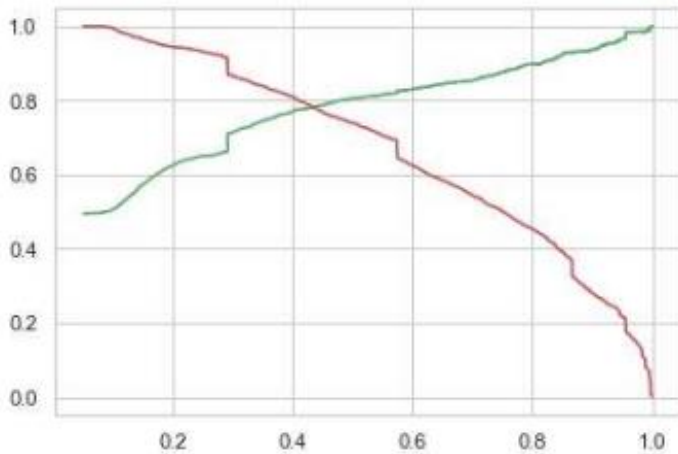Most of the converted leads were being sent a SMS or an email opened.

Model Evaluation

0.42 is the tradeoff between Precision and Recall -
Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than
42 % to be a hot Lead

Important results

Train Data Statistics: Accuracy 80%
- Sensitivity 77%
- Specificity 80%

Test Data Statistics: Accuracy 80%
- Sensitivity 77%
- Specificity 80%

Important attributes and features

- Lead Source_Olark Chat

- Total Time Spent on Website

- Lead Origin_Landing Page Submission

- Specialization_Others

- Lead Source_Google

- Lead Origin_Lead Add Form

- Do Not Email

Things to keep in mind

We see that the conversion rate is 30-35% (close to average) for API and Landing page submission. But very low for Lead Add form and Lead import. Therefore

We can intervene that we need to focus more on the leads originated from API and Landing page submission.

We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.

Leads who spent more time on website, more likely to convert.

Most common last activity is email opened. highest rate = SMS Sent.

- Maximum people are unemployed.

- Maximum conversion is with working professional.

THANK YOU

Submitted By: Abhishek Rastogi