LEAD SCORE CASE STUDY

-By Mansi

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.

▶ X Education gets a lot of leads, its lead conversion rate is very poor. For example, 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 Objective

▶ The company wants us to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance.

▶ The CEO, wants to achieve a ballpark of the target lead conversion rate to be around 80%.

▶ They want the model to be able to handle future constraints as well like Peak time actions required, how to utilize full man power and after achieving target what should be the approaches

Problem Approach

▶ Importing and inspecting the data.

▶ Data Preparation.

▶ EDA.

▶ Dummy Variable Creation.

▶ Feature Scaling.

▶ Test-Train-Split.

▶ Model Building(RFE Rsquared,VIF,p-values).

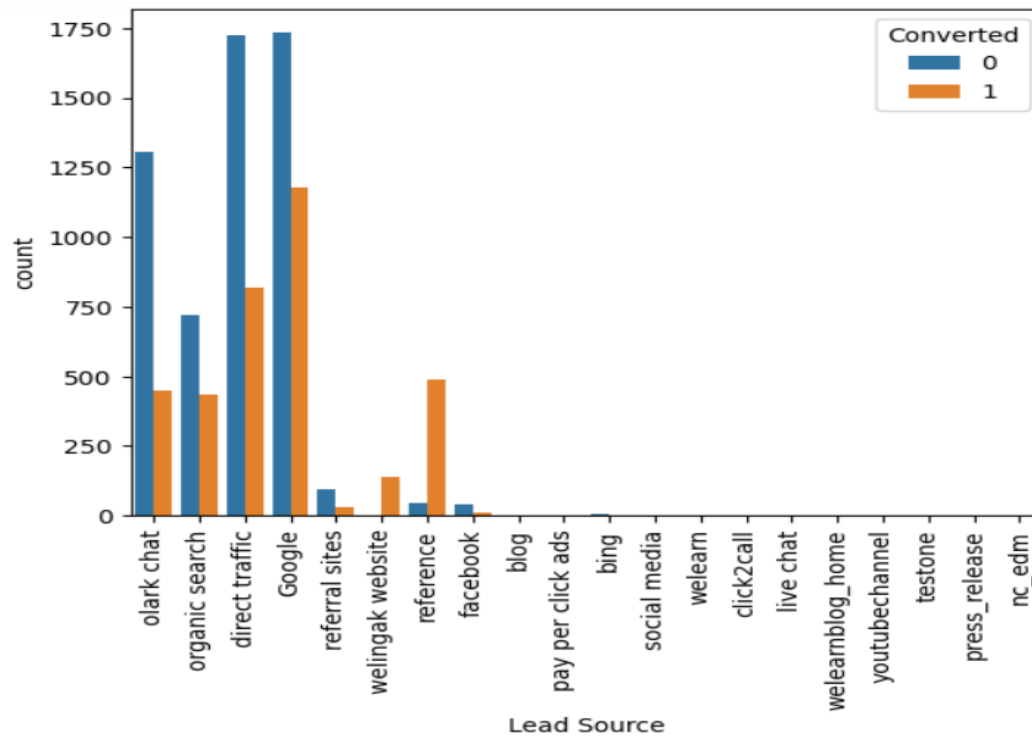▶ Model Evaluation.

▶ Making Predictions on test set.

EDA:
Lead Origin

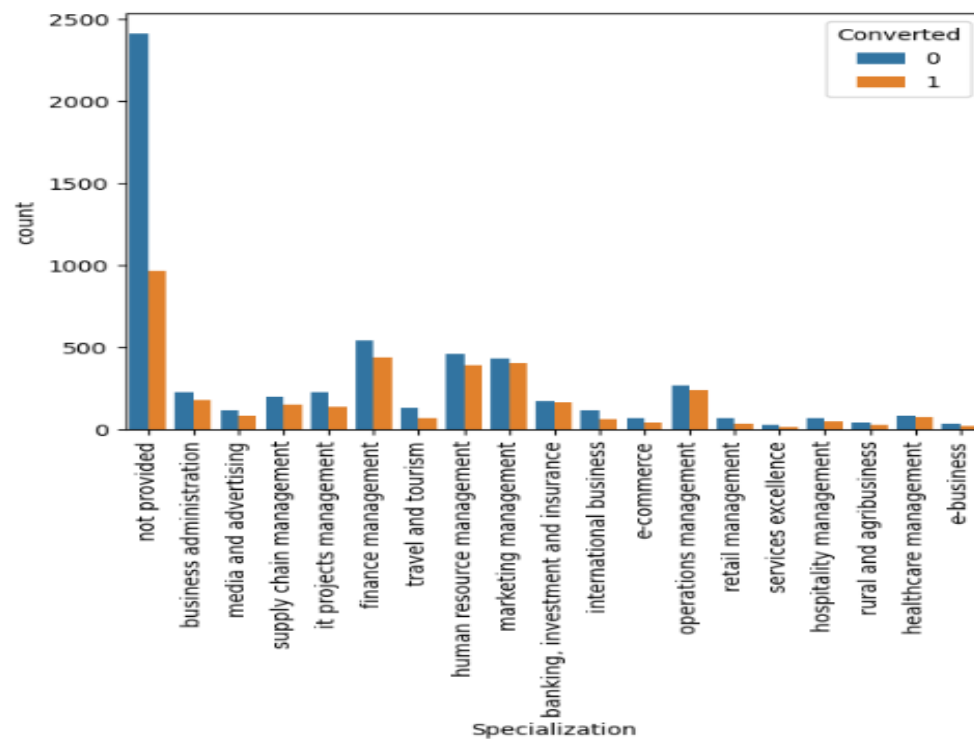▶ In Lead Origin most number of leads are landing on page submission.
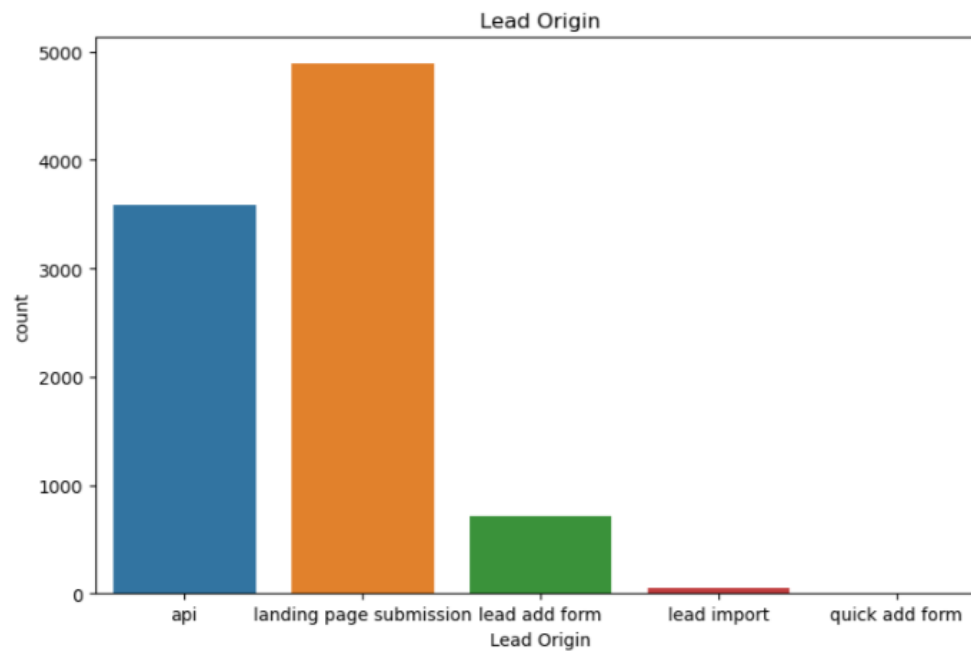
Lead Source

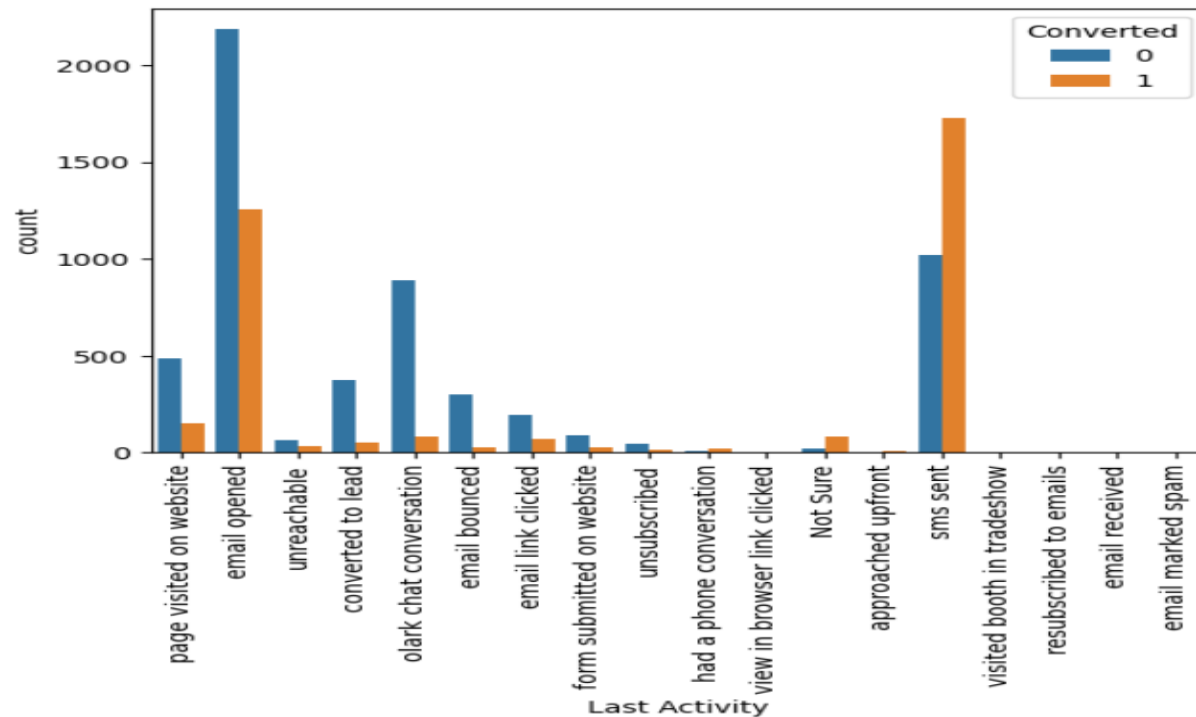▶ In Lead Source the leads through Google and Direct Traffic have high probability to convert.

Specilization

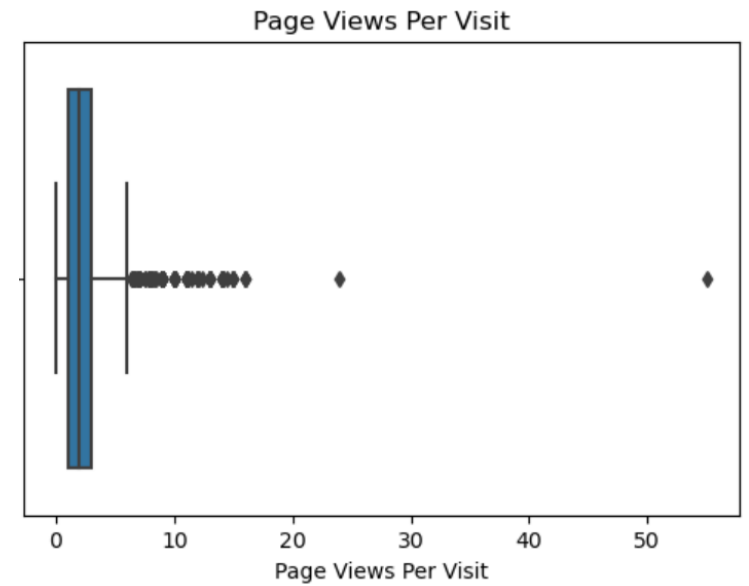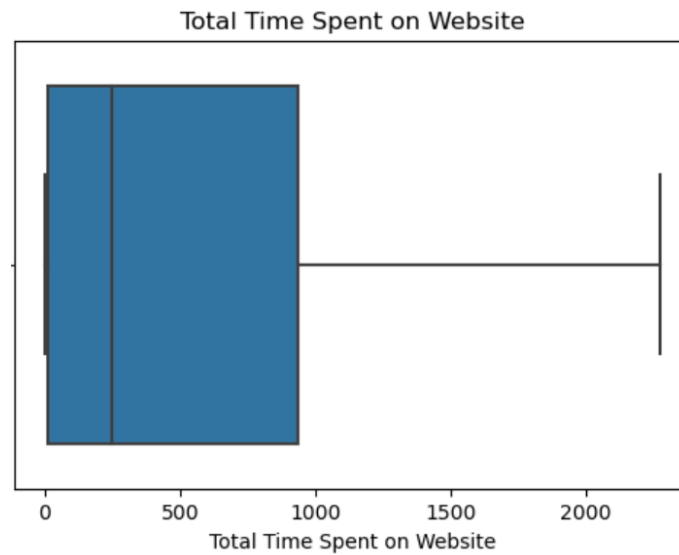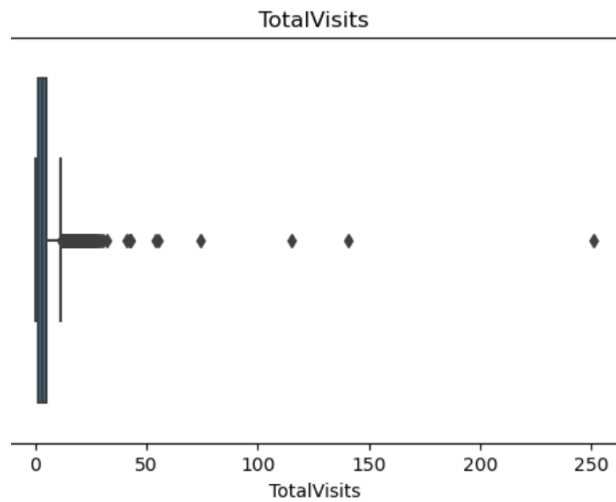▶ Leads from HR,Finance and marketing management have high probability to convert.

Last Activity

- Leads which are opening email and SMS sent have high probability to convert,olark chat conversation will also benefit.
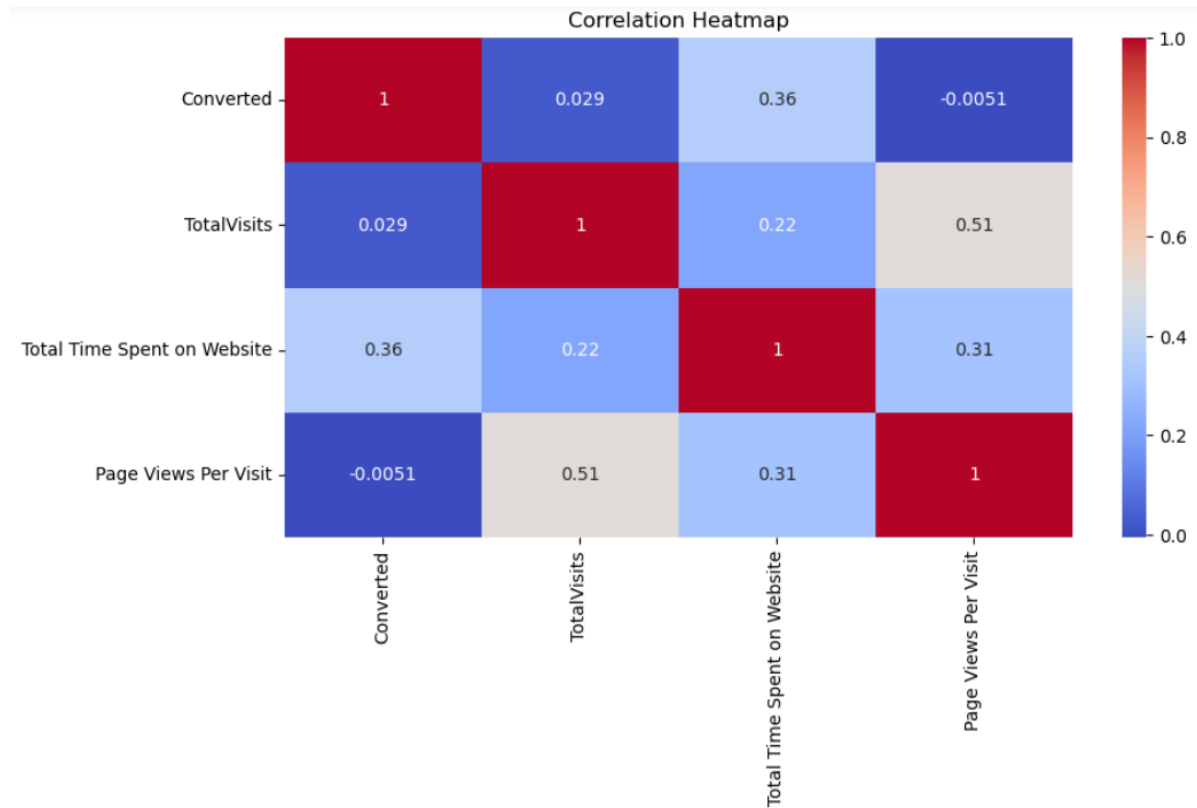
Box Plots
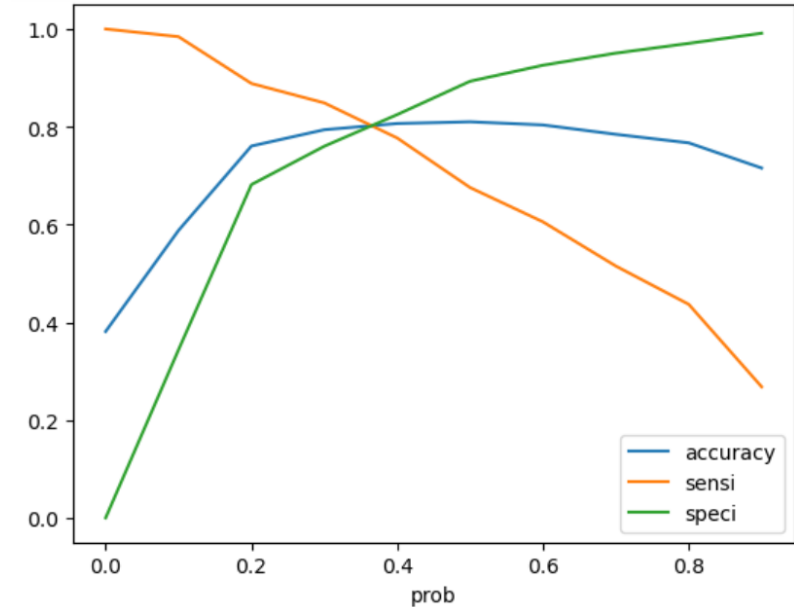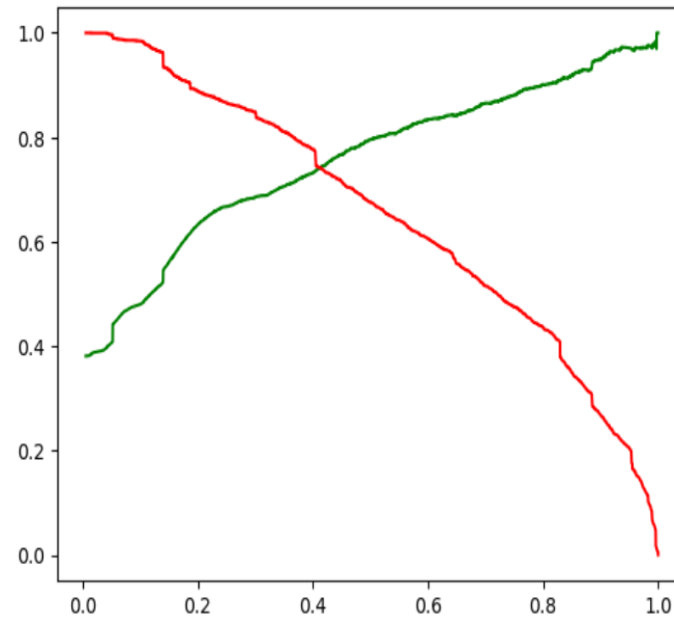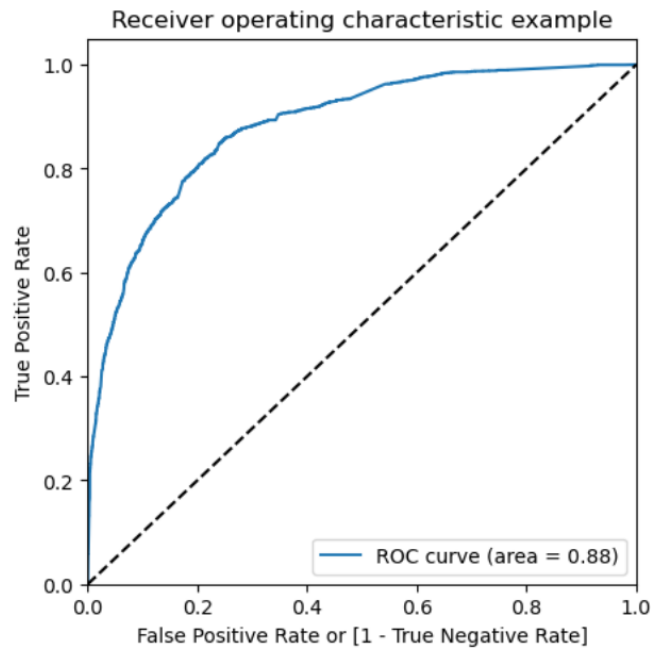
Correlation

Correlation Heatmap

Model Evaluation
ROC Curve

0.4 is the tradeoff between Precision and Recall - Thus we can safely choose to consider any Prospect Lead with Conversion Probability higher than 40 % to be a hot Lead

Observations

- Train Data:

- Accuracy : 79%

- Sensitivity : 73%

- Specificity : 83%

- Test Data:

- Accuracy : 79%

- Sensitivity : 73%

- Specificity : 83%

Conclusions

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

- Total Time Spent on website.

- Total number of visits.

- When the lead source was Google,Direct Traffic,Organic Search.

- When the last activity was SMS,Olark Chat Conversation.

- When lead orgin is lead add format Keeping this in mind the X_Education can flourish as they have high chances to get almost all potetial buyers to change their mind and buy the courses.