

Lead Scoring Case Study

By

Seema Sarda
Amit Kapoor
Dipanwita Dey



Problem Statement

An education company named X Education sells online courses to industry professionals.

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

X education wants
to know the most
promising leads.

01

For that they want
to build a Model
which identifies
the hot leads.

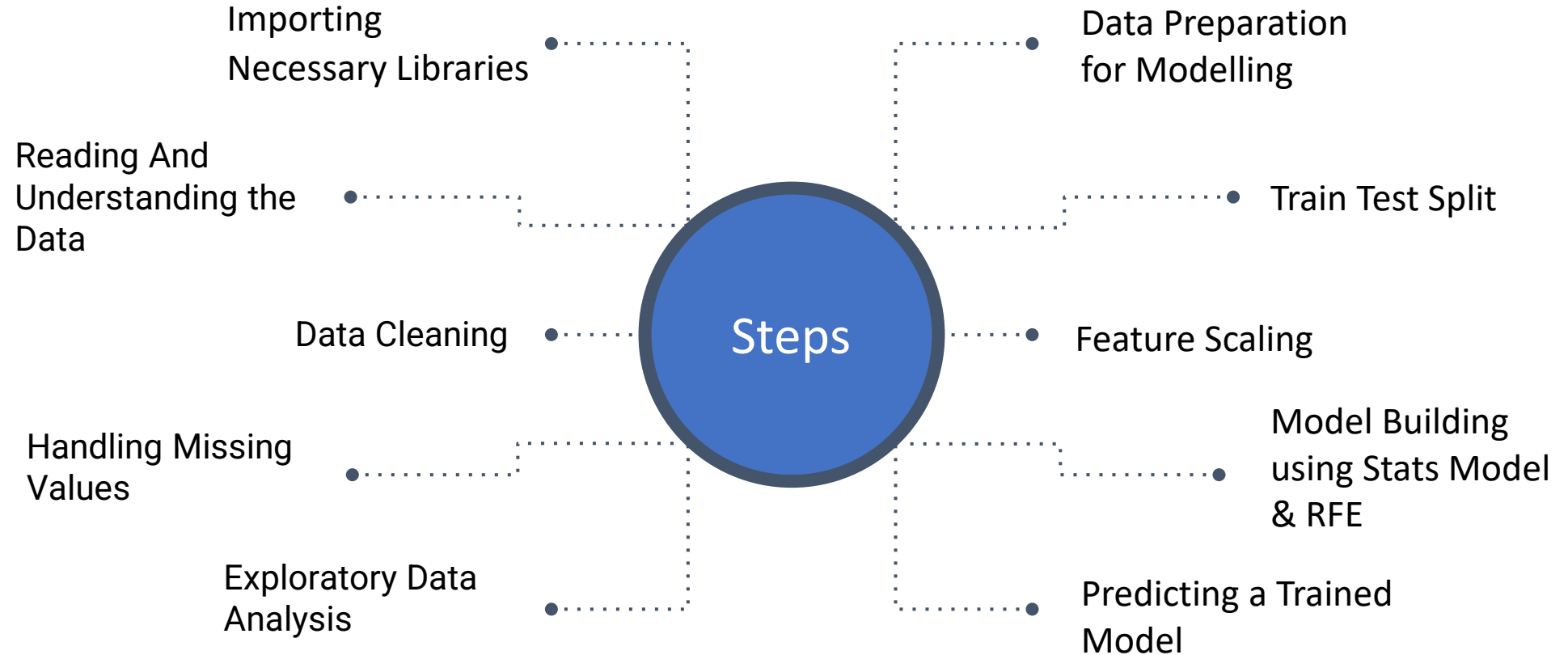
02

Deployment of
the model for the
future use.

03

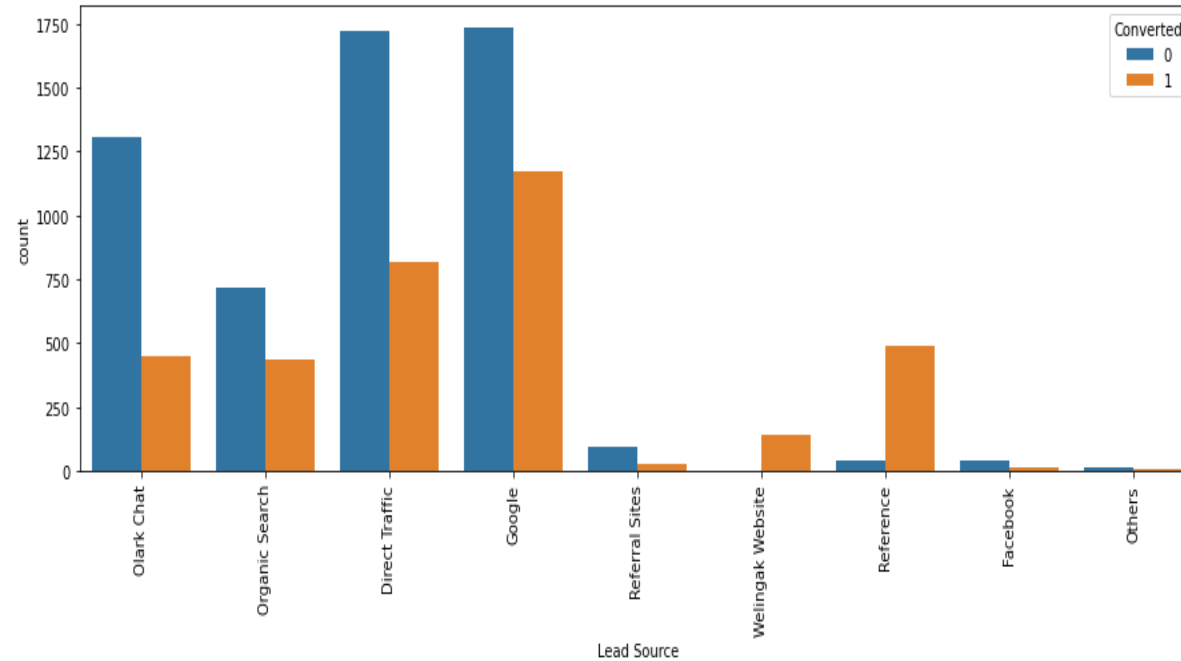


Steps to be Taken



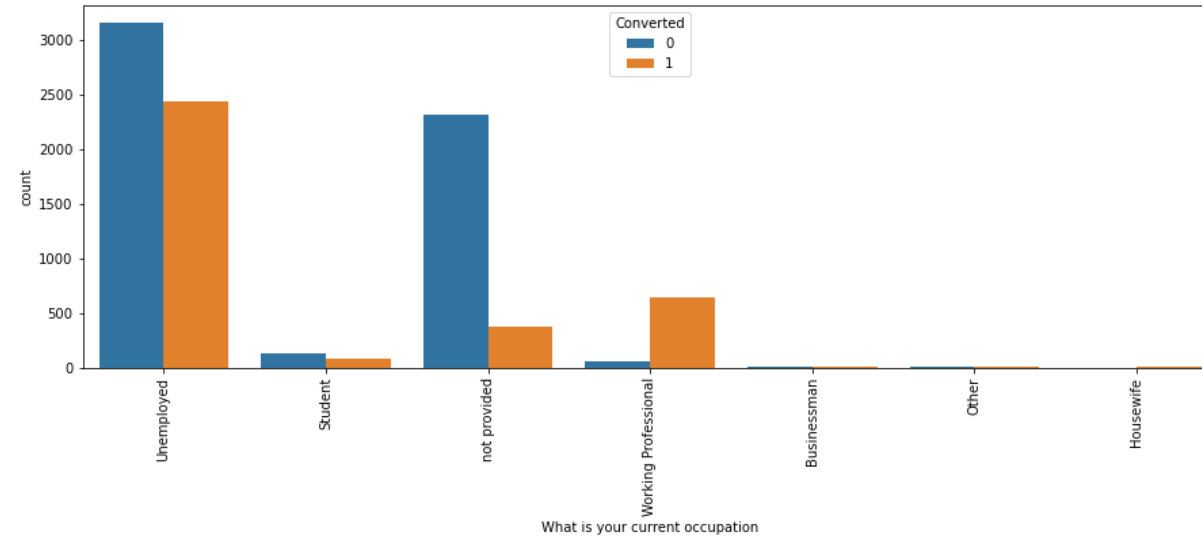
Data Visualization

Lead Source



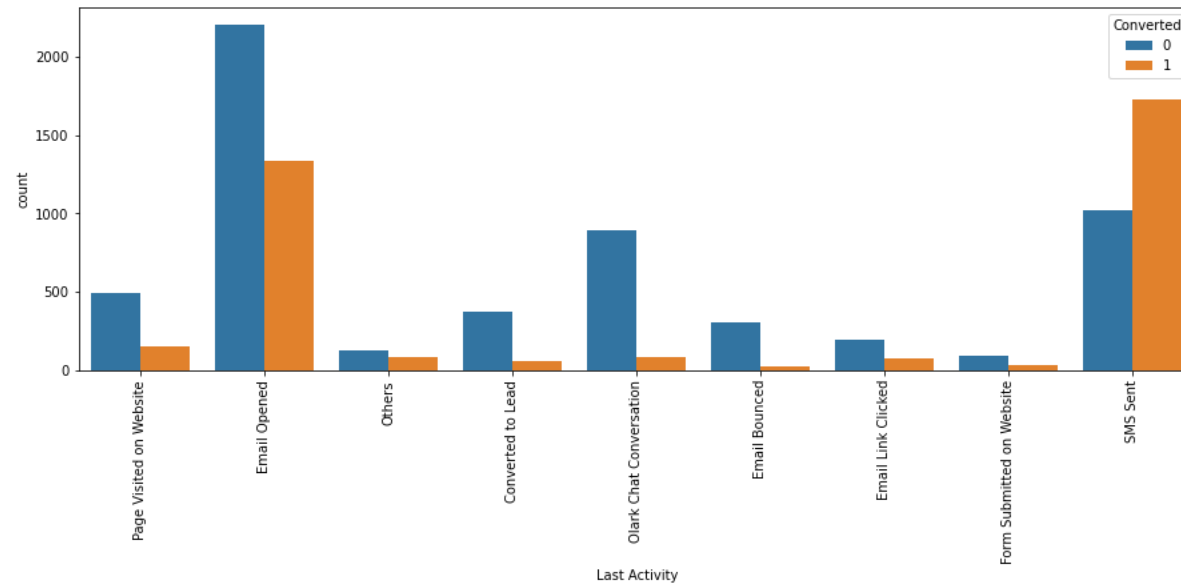
Maximum Leads are generated by Google and Direct Traffic. The conversion rate of Reference leads and Welinkgak Website leads is very high.

Current Occupation



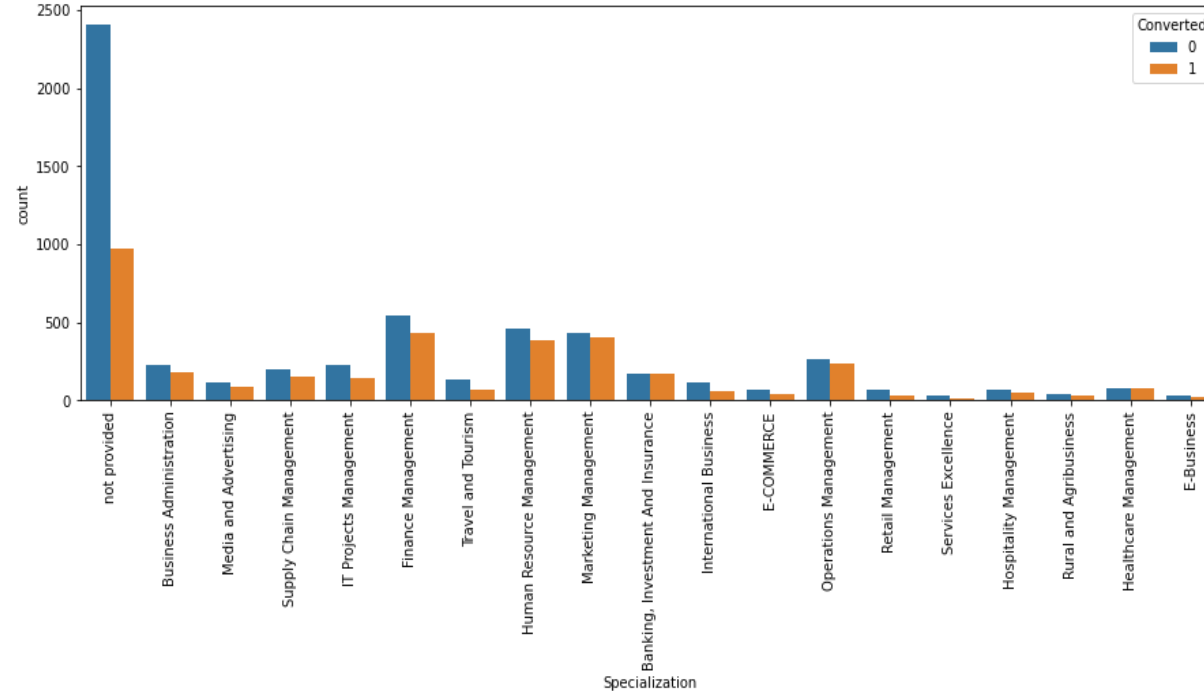
Maximum leads generated are unemployed and their conversion rate is more than 50%. Conversion rate of working professionals is very high.

Last Activity



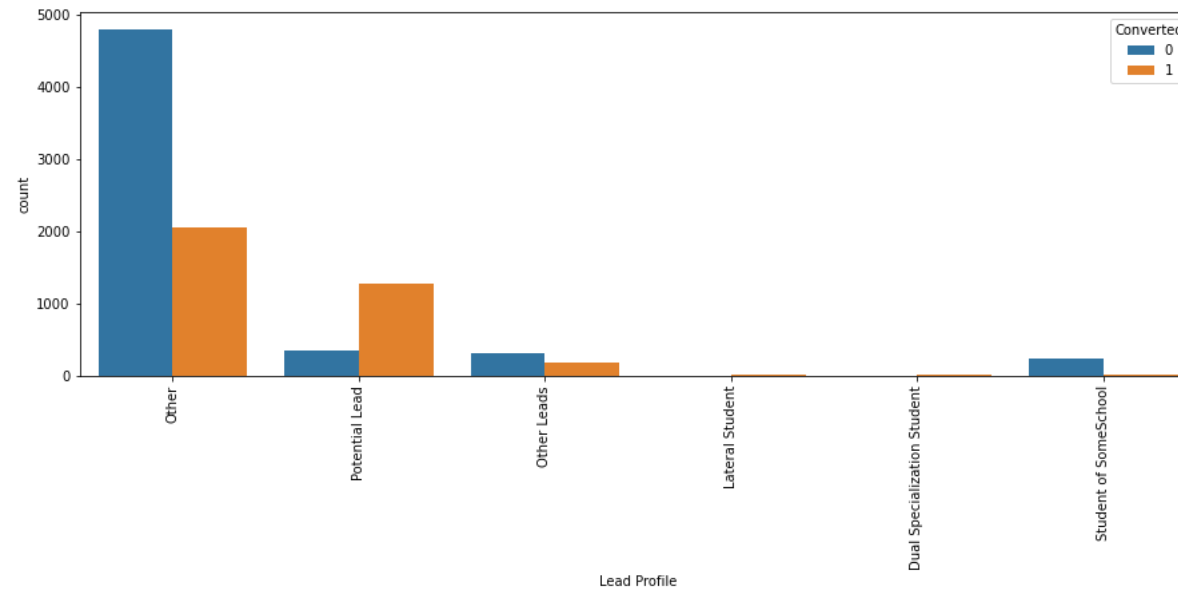
Maximum leads are generated having last activity as Email opened but conversion rate is not too good. SMS sent as last activity has high conversion rate.

Specialization



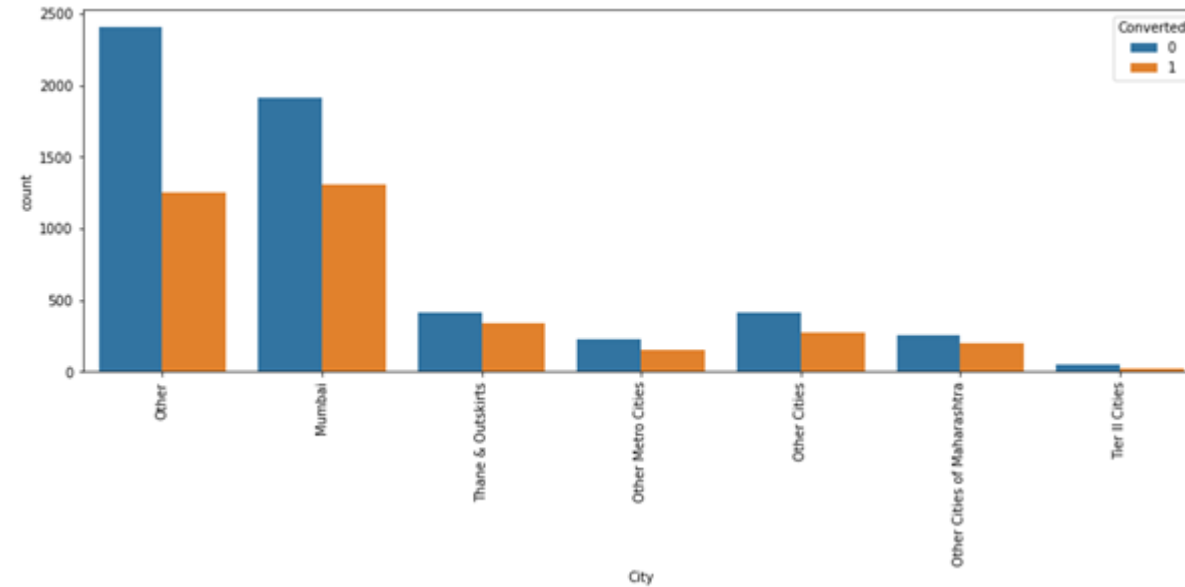
There is a good conversion rate of the leads from almost all specialization except which are not provided, so this feature may be appended in the list of columns to drop.

Lead Profit



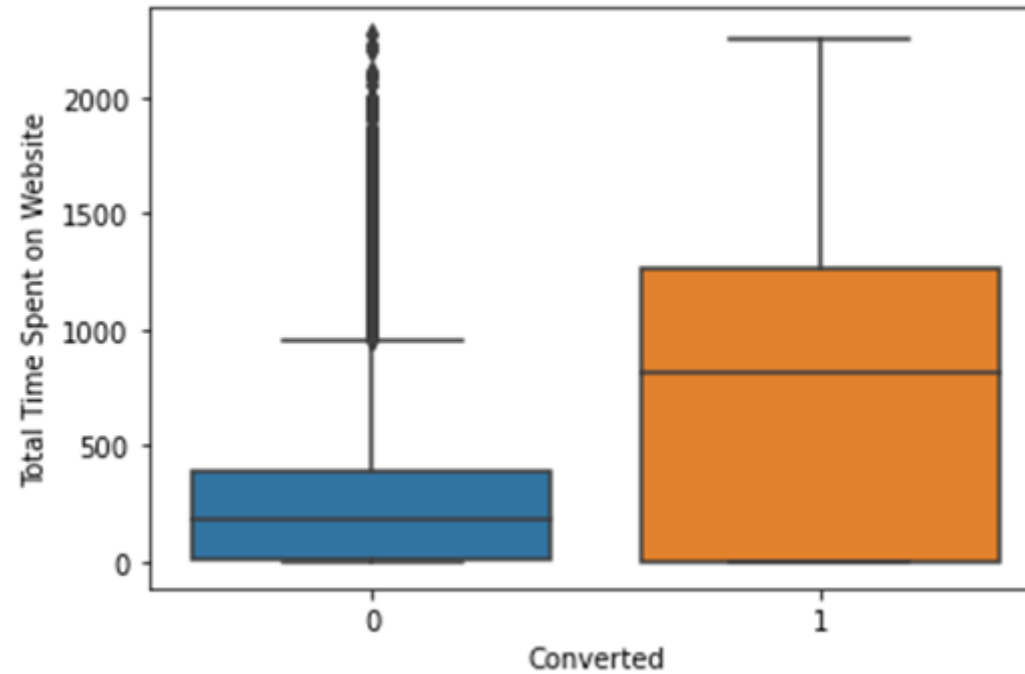
This feature also has the same problem as the above features like How did you hear about X Education, hence can be appended to the list of col to drop

City



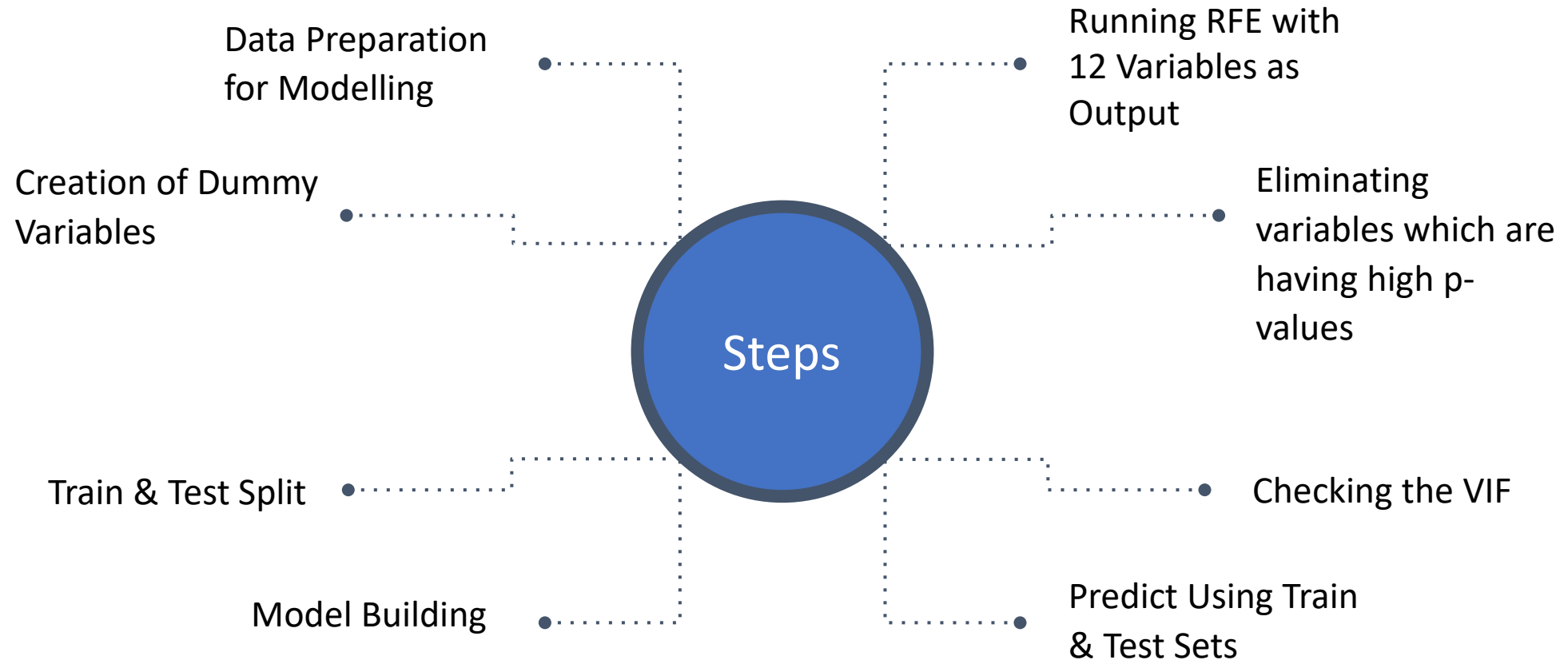
This feature "city" may be dropped too as this also has the same problem like in Lead Profile

Converted

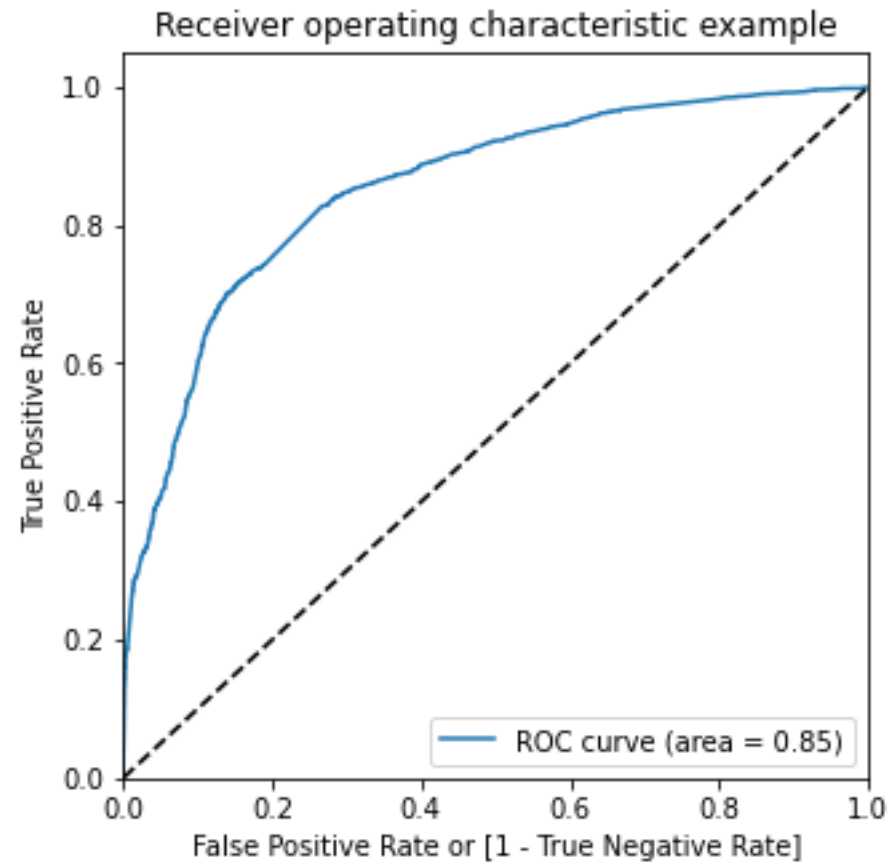


As can be seen, leads spending more time on the website are more likely to convert, thus website should be made more engaging to increase the conversion rate

Model Building

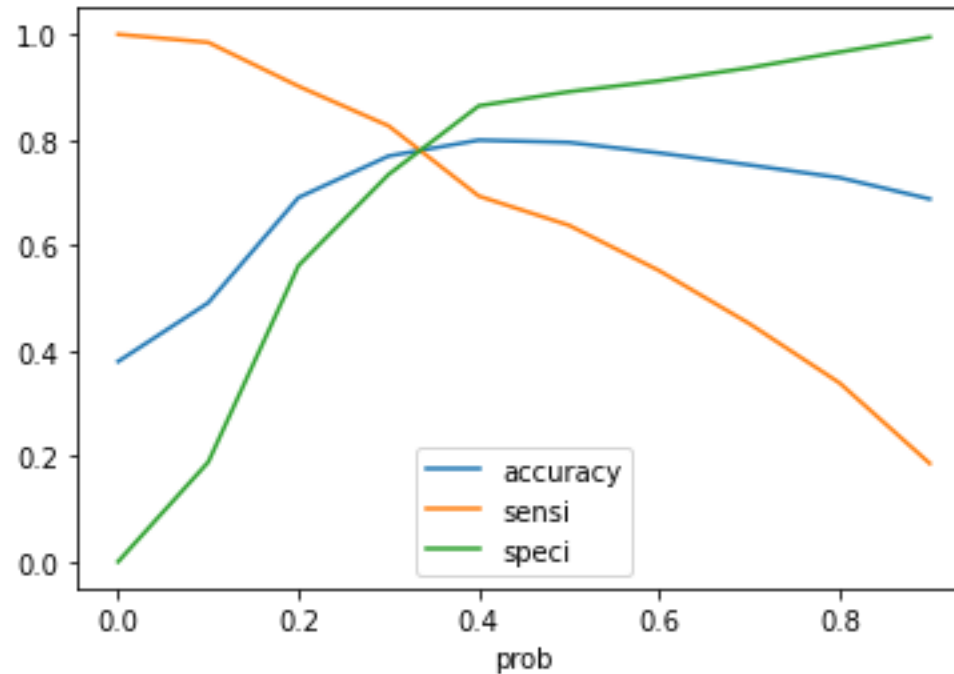


ROC CURVE



The ROC Curve should be a value close to 1. We are getting a good value of 0.85 indicating a good predictive model

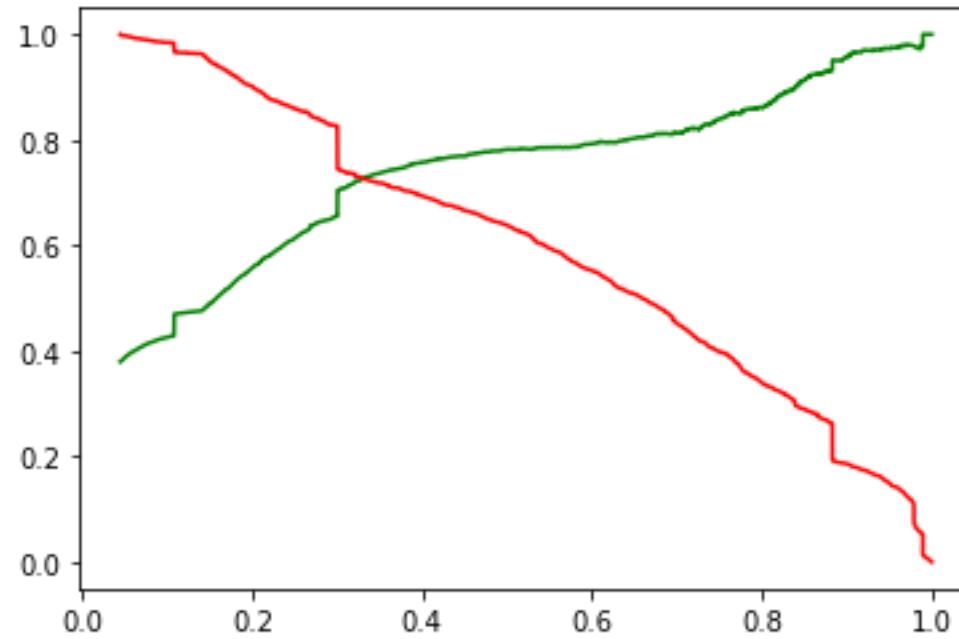
Accuracy, Sensitivity & Specificity



From the curve above, 0.3 is the optimum point to take it as a cut-off probability

- Accuracy => 76.93%
- Sensitivity => 82.63%
- Specificity => 73.43%

Precision and Recall



Precision score => 65.5%

Recall score => 82.6%

Conclusion

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut-off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of the test set are around 78%, 83% and 75% which are approximately closer to the respective values calculated using the trained set.
- Also, the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%
- Hence overall this model seems to be good.

Important features responsible for a good conversion rate or the ones which contribute more towards the probability of a lead getting converted are :

- Lead Origin_Lead Add Form
- What is your current occupation_Working Professional
- Total Time Spent on Website

Thank You

