Lead Score – Case Study

Submitted by:

- Souvik Haldar

- Kaushtabh Pasalkar

- Greeshma Eliyan

Problem statement

An X Education company needs 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%.

Case study approach

- Data Loading & Cleaning

- Exploratory Data Analysis

- Creating Dummy Variables

- Train – Test Splitting of Data

- Building a Logistic Regression Model

- Evaluating the Model by different Measures and Metrices – Accuracy, Sensitivity, Specificity, Precision and Recall

- Checking the Model Performance over Test Data by Confusion Metrix, Sensitivity, and Specificity
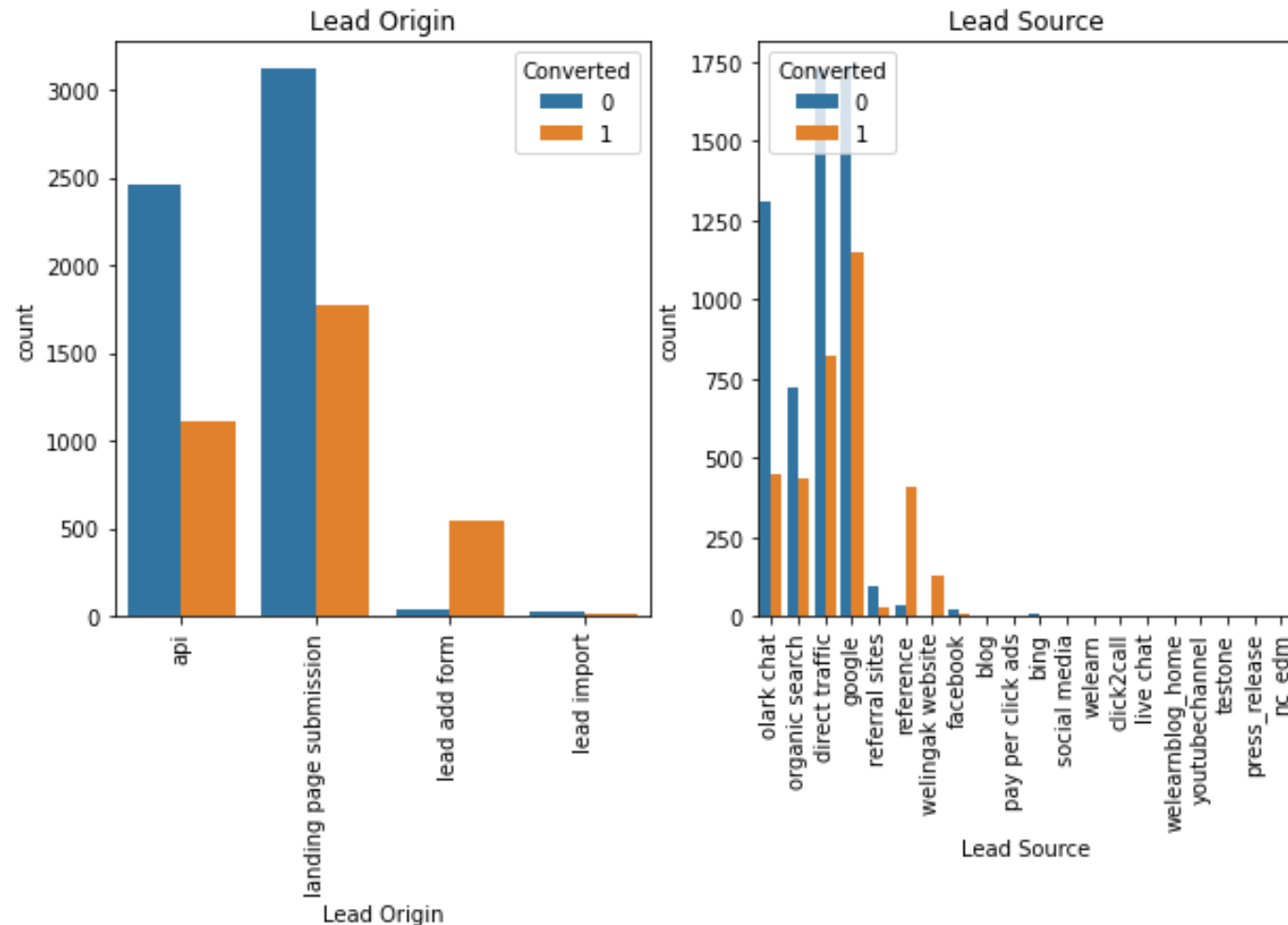
Exploratory Data Analysis

There is more than 35% Conversion rate in Total
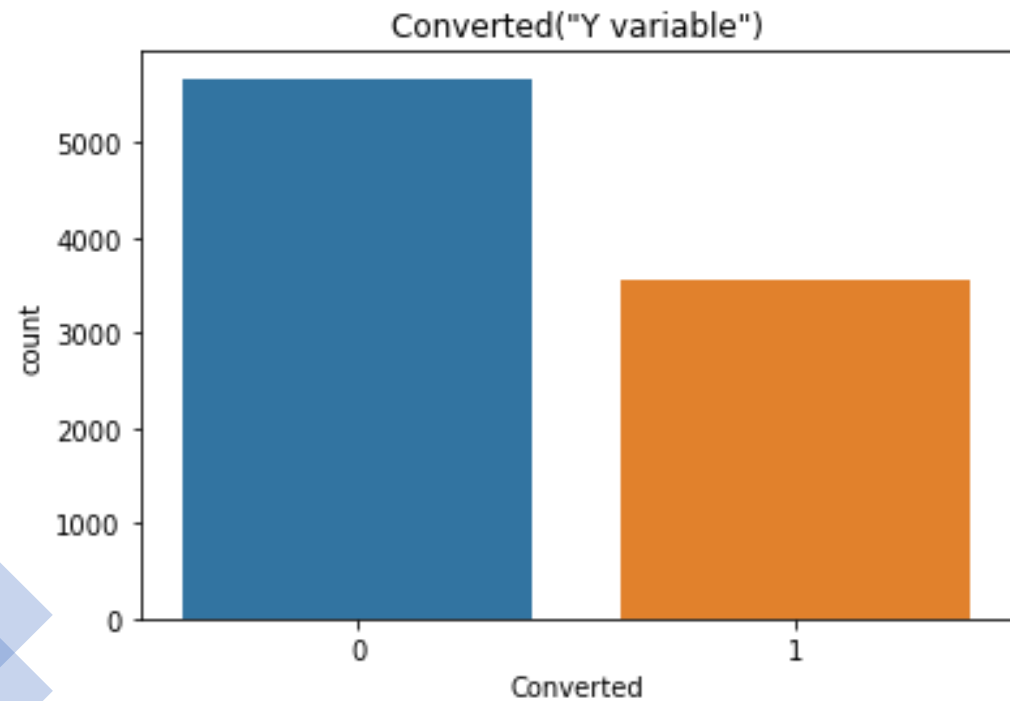
- More users converted when the last notable activity was 'sms sent'.
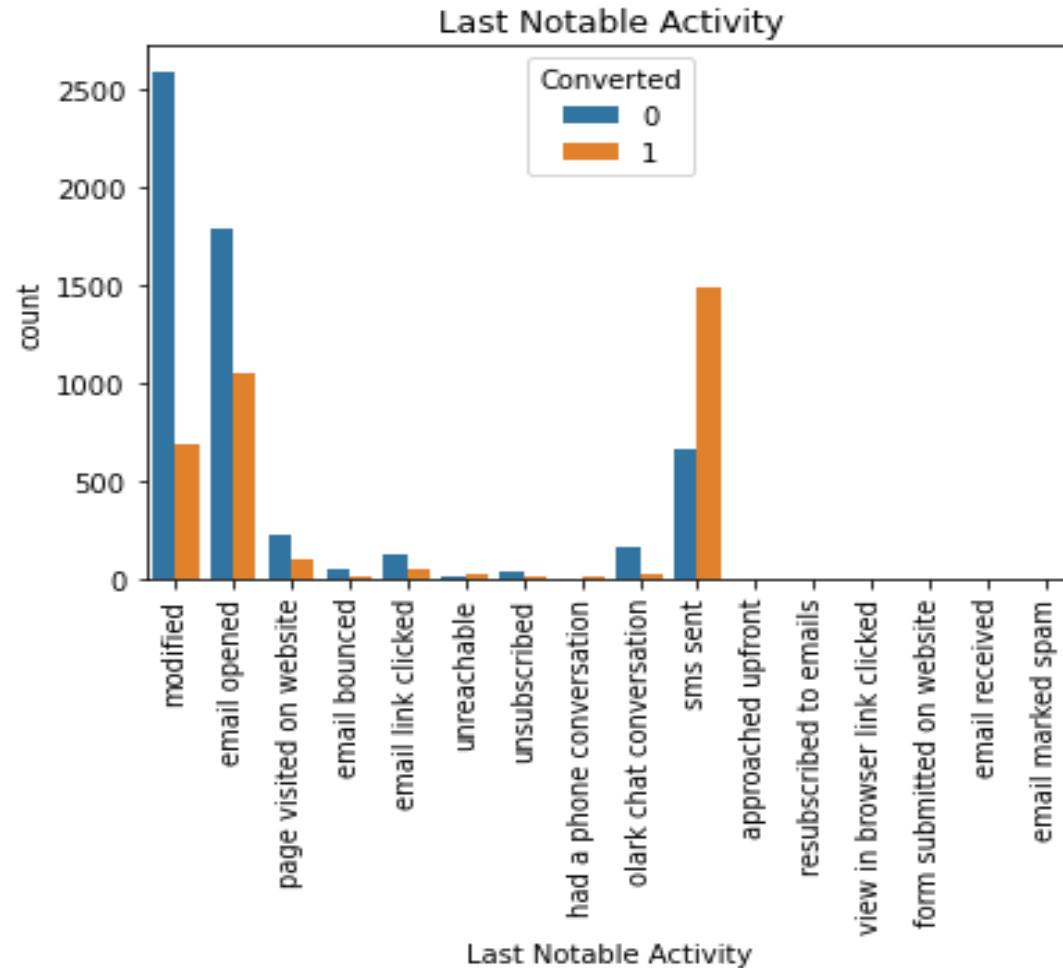
- More conversion happened when Lead source was Google, Direct Traffic, Organic Search or Welingak website.

- Also, conversion rate was high when the lead origin is 'lead add format'.

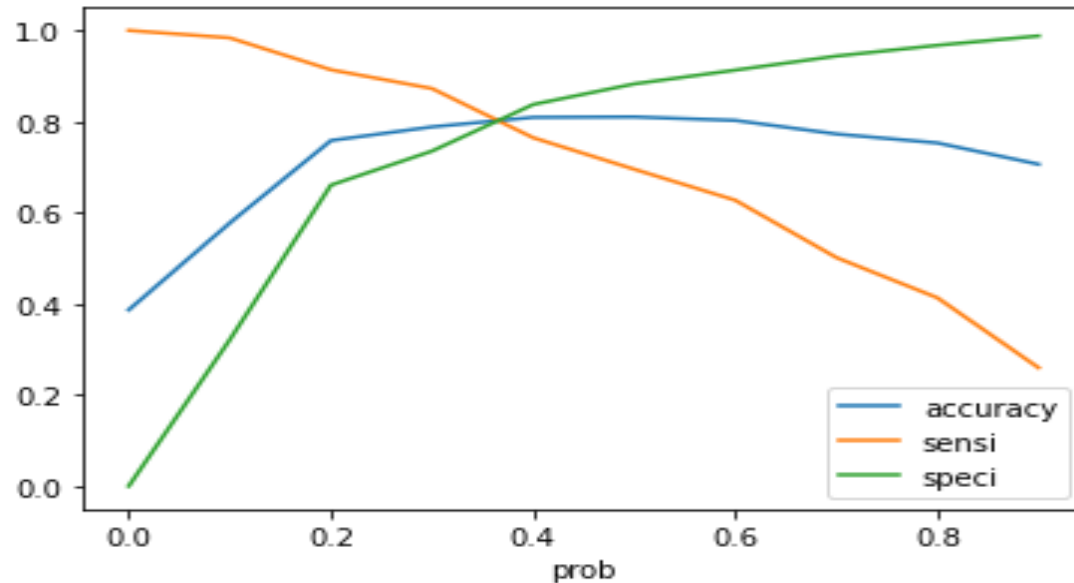- More potential users are working professional. They have more conversion rate.
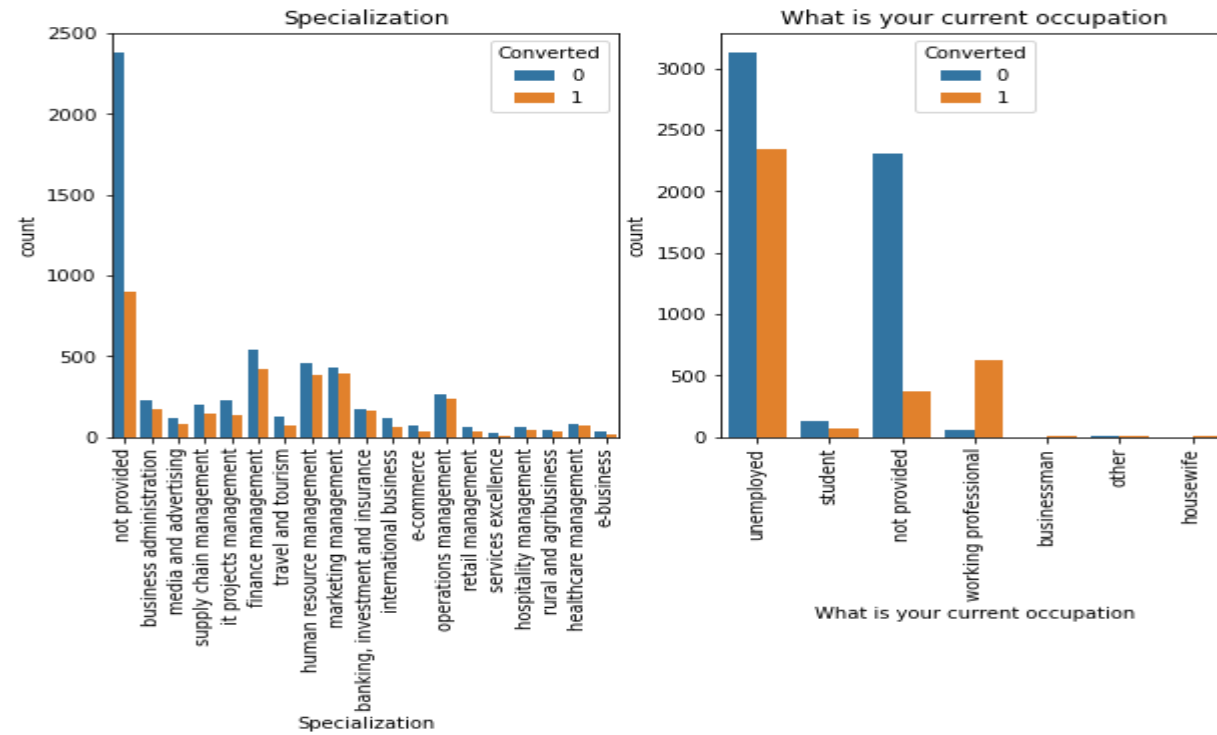
Model Evaluation – ROC Curve

Receiver operating characteristic example
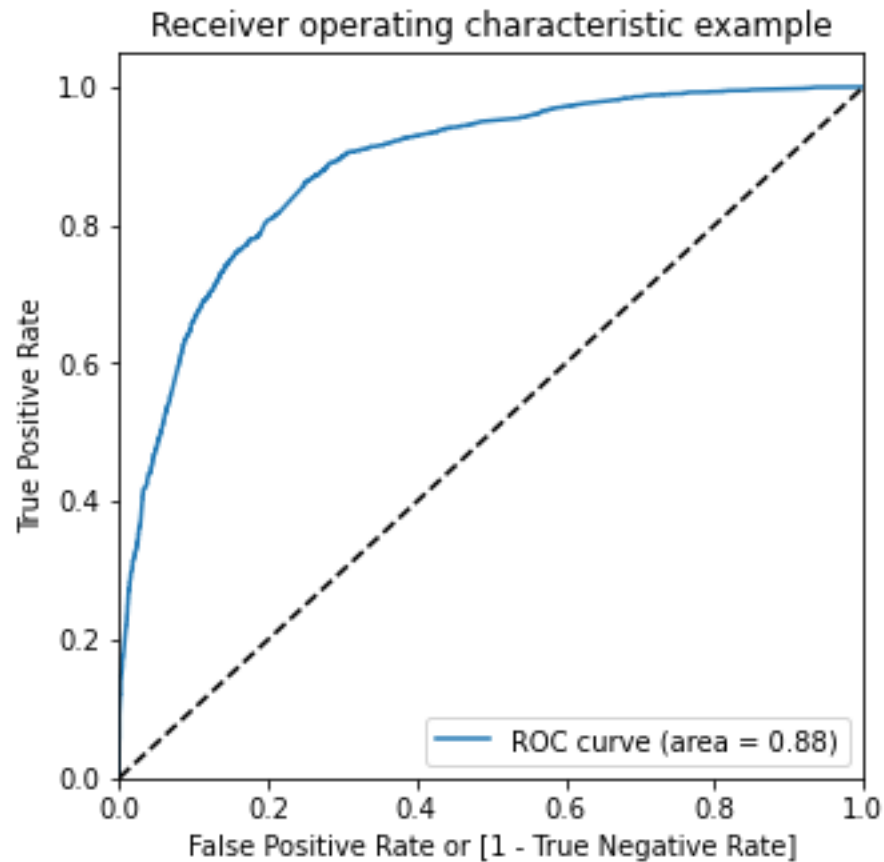
The area under ROC curve is 0.87 which is a very good value

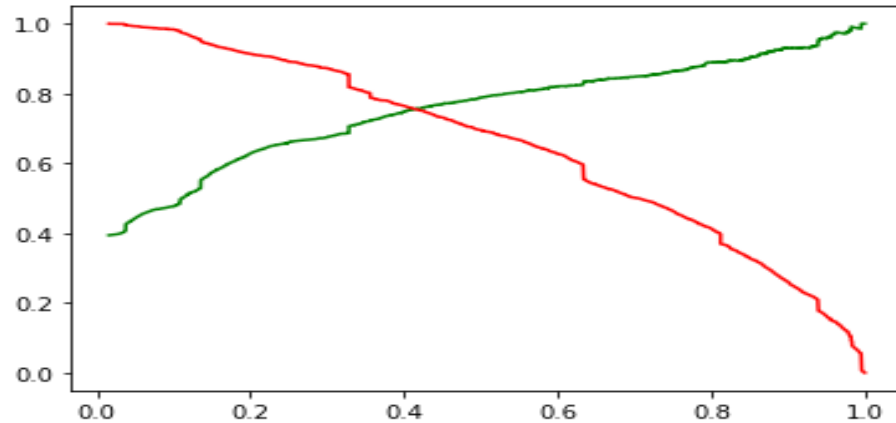# Model Evaluation on Train Data – Sensitivity and Specificity

Confusion Metrix

3438	**457**
748	1708

Accuracy - 81%
Sensitivity – 70%
Specificity – 88%

Model Evaluation on Train Data – Precision and Recall

Confusion Metrix

3289	**606**
593	1863

Precision – 75%
Recall – 76%

Model Evaluation on Test Data

Confusion Metrix

1406	**338**
181	798

Accuracy - 81%
Sensitivity – 82%
Specificity – 81%

Confusion Metrix

1472	**272**
232	747

Precision – 73%
Recall – 75%

Conclusion

It was found that the variables that mattered the most in the potential buyers are (In descending order) :

- The total time spend on the Website.

- Total number of visits.

- When the lead source was:
  a. Google
  b. Direct traffic
  c. Organic search
  d. Welingak website

- When the last activity was:
  a. SMS
  b. Olark chat conversation

- When the lead origin is Lead add format.

- When their current occupation is as a working professional.