

#### **Problem Statement**

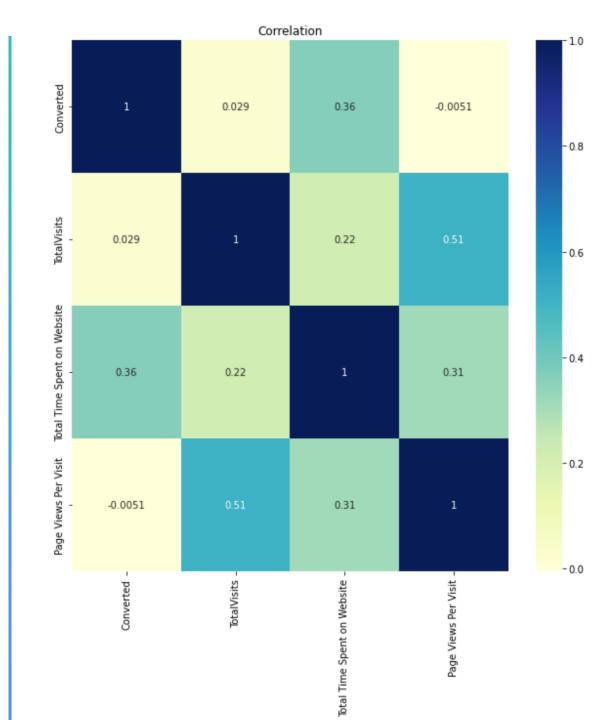
- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.
- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

#### **Business Goal**

- The company requires you 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, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

## Strategy

- Import dataset
- Data Cleaning
- Exploratory Data Analysis
- Scaling Features
- Prepare the data for model building
- Logistic Regression Model
- Assign a lead score for each leads
- Train and test Model
- Evaluate the model
- Test the Model in test set
- Measure the accuracy of the model



#### Correlation

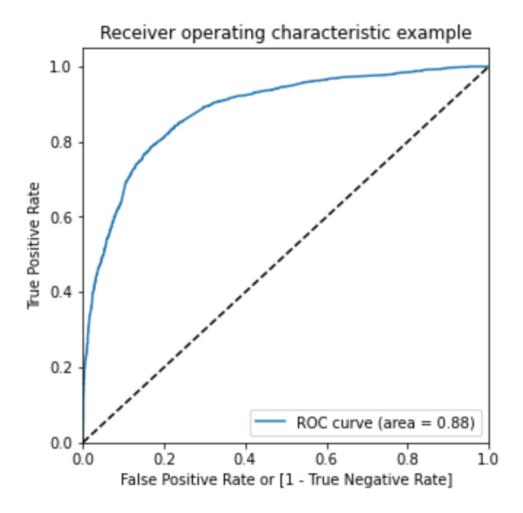
- Total Visits
- Converted
- Total time spent on website

## Heat Map to check for correlation

Total Time Spent on Website X Education Forums Digital Advertisement A free copy of Mastering The Interview Lead Origin Lead Add Form Lead Origin Quick Add Form Lead Source\_Google Lead Source Organic Search Lead Source Reference Lead Source Welingak Website Last Activity Email Link Clicked Last Activity Form Submitted on Website Last Activity\_Others Last Activity Last Activity\_SMS Sent Last Activity Unsubscribed What is your current occupation Unemployed Last Notable Activity Email Link Clicked Last Notable Activity Modified Last Notable Activity Other Activity Last Notable Activity SMS Sent Specialization\_Business Administration Specialization E-COMMERCE Specialization Healthcare Management Specialization Human Resource Management Specialization International Business Specialization Media and Advertising Specialization Retail Management Specialization Services Excellence Specialization\_Travel and Tourism City Other Cities of Maharashtra City Thane & Outskirts

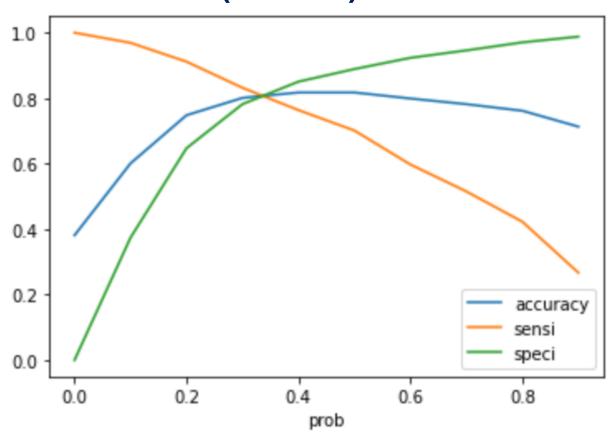
-1.00 - 0.75 - 0.50 0.25 0.00 -0.25

## **ROC Curve**



Optimal cutoff probability is that probwhere we get balanced sensitivity and specificity.

# Model evaluation(TRAIN)c



Accuracy sensitivity and specificity

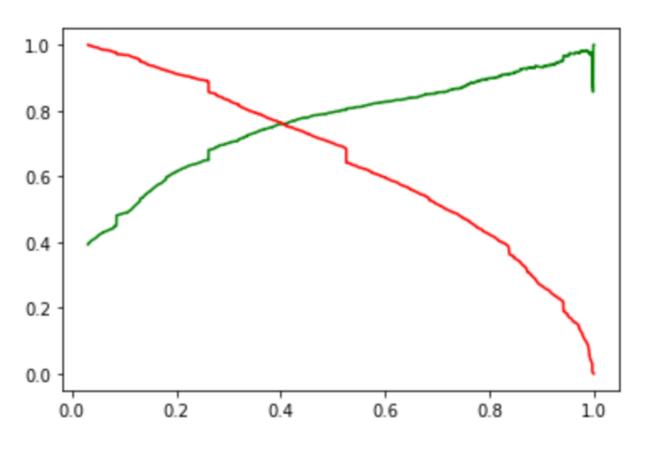
Train Data:

**Accuracy : 81.07%** 

Sensitivity: 79.3%

Specificity: 82.1%

# Model evaluation (TEST)



Accuracy sensitivity and specificity

**Test Data:** 

Accuracy: 79.68% Sensitivity: 81%

Specificity: 78.47%

### Conclusion

The model seems to be performing well.

**Train Data:** 

Accuracy: 81.07% Sensitivity: 79.3% Specificity: 82.1%

**Test Data:** 

Accuracy: 79.68% Sensitivity: 81%

Specificity: 78.47%

The model seems to be performing well. Can be recommend this model in making good calls based on this model.

Thank You.