LEAD SCORE CASE STUDY

LOGISTIC REGRESSION

PROBLEM STATEMENT

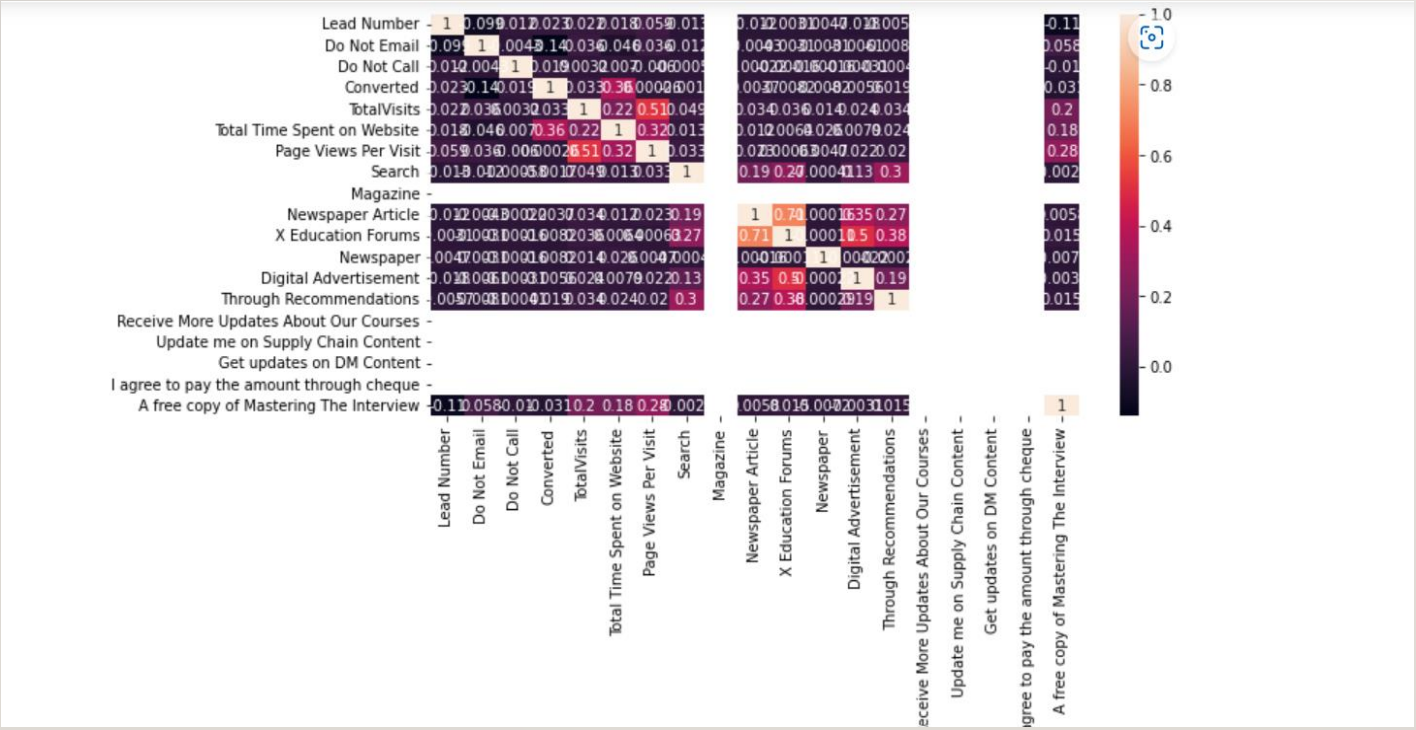
- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.
- The company markets its courses on several websites and search engines like Google. Once these people land on the website, these people fill up a form providing their email address or phone number, they are classified to be a lead and the company also gets leads through past referrals.
- 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%. The implementation process of lead generation are not efficient.

BUSINESS GOAL

- The company requires model for the most promising leads.
- Lead score should be given such that the higher the score, more promising will be the leads, the lower is the lead score less chances of conversion.
- The model to be built in lead conversion is 80% or more.

EXPLORATORY DATA ANALYSIS

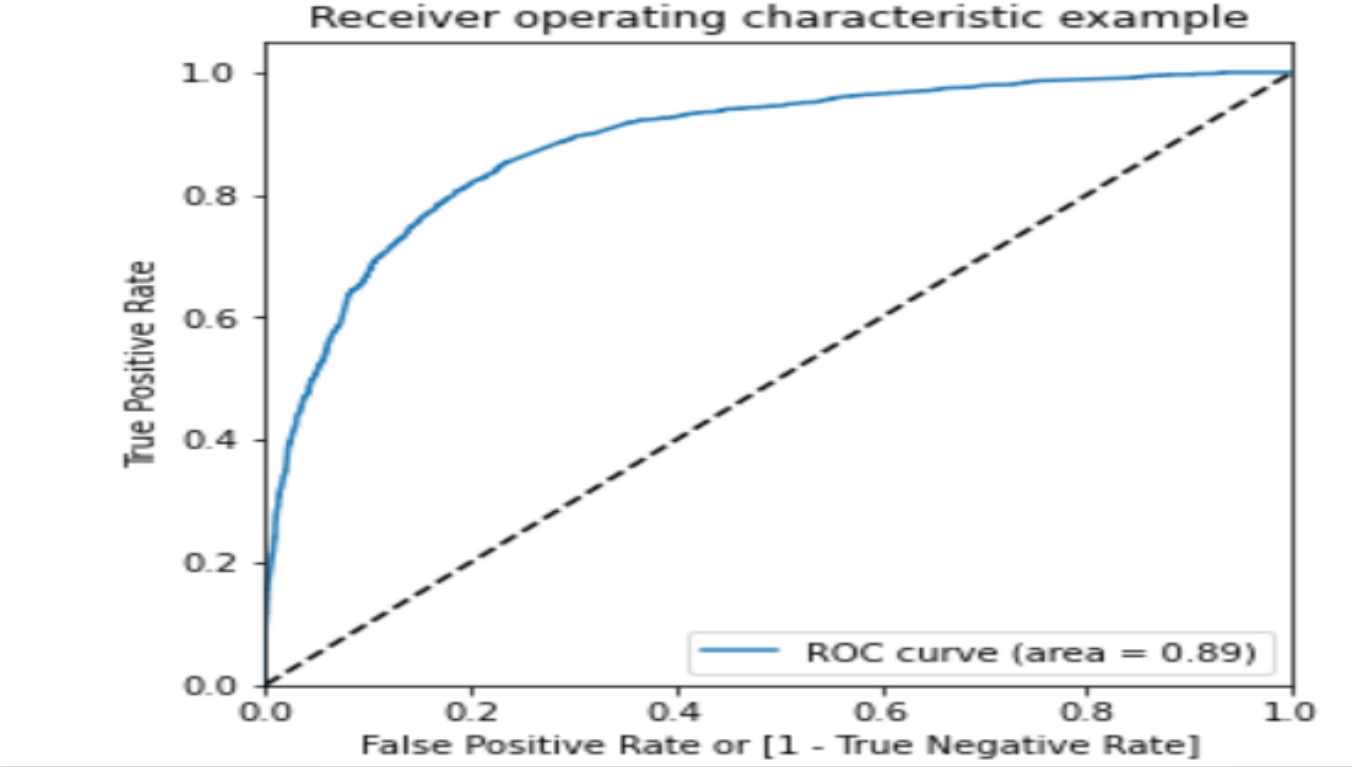
- Correlation Matrix



MODEL BUILDING

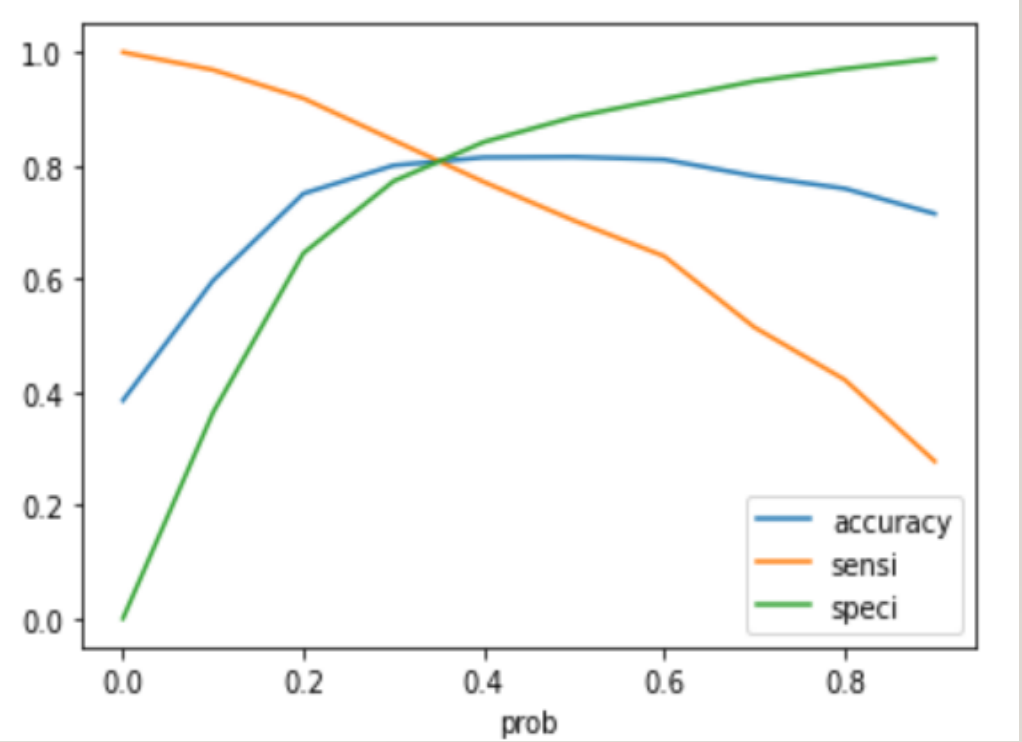
- Splitting into test and train dataset
- Scale variable in train set
- Build first model
- Use RFE to remove less relevant variable
- Build next model
- Assesing model with statsmodel
- Evaluate accuracy and other metric
- Predict using test set
- Precison and recall analysis on test predictions

ROC CURVE


Receiver operating characteristic example
True Positive Rate
False Positive Rate or [1 - True Negative Rate]
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.89)

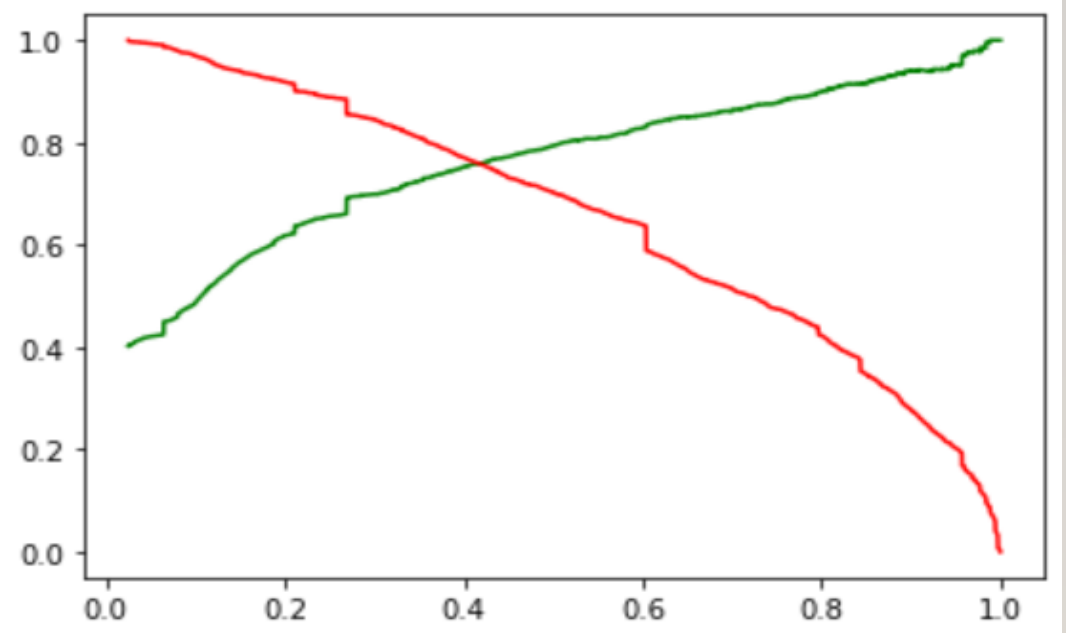
MODEL EVALUATION-TRAIN SET

- Accuracy = 80.02
- Sensitivity = 84.46
- Specificity = 77.23



MODEL EVALUATION-TEST SET

- Accuracy = 81.49
- Sensitivity = 75.43
- Specificity = 84.95



CONCLUSION

- The model shows high close to 81% of accuracy.
- Threshhold can be selected from Accuracy, Specificity, Sensitivity and precision, Recall curves
- Model shows about 75% sensitivity and 85% specificity.
- Overall this model proves to be accurate.