

LEAD SCORING CASE STUDY

PRATIKHYA MOHAPATRA

PROBLEM STATEMENT

X Education is an organization which provides online courses for industry professionals. The company shows the course on several popular websites. X Education wants to select the most promising leads to be converted to paying customers. Although the company generates a lot of leads only a few are converted to paying customers by numerous modes like E-mail, Advertisements, searches over the internet, etc.

The company has had 30% conversion rate through the whole process. The implementation process of lead generating attributes are not efficient in helping conversions.

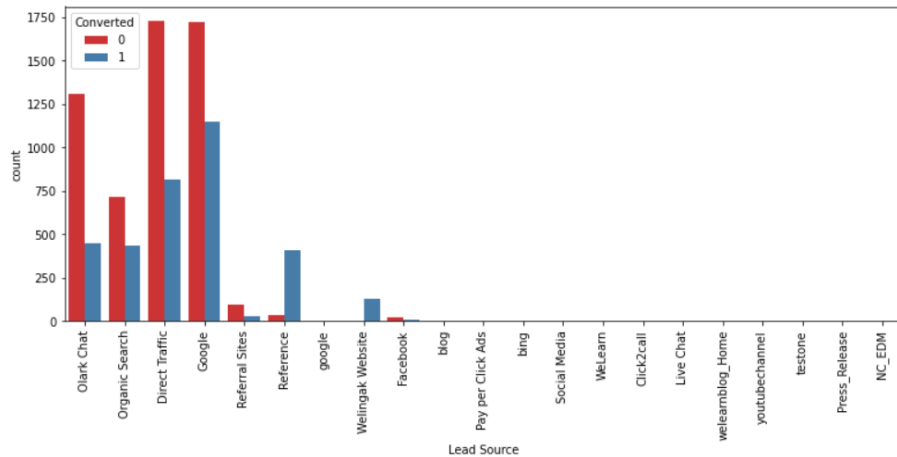
STEPS TO BE TAKEN

- IMPORT DATA
 - DATA CLEANING
 - EDA
 - SCALING FEATURES
 - DATA PREPARATION
 - BUILD A LOGISTIC REGRESSION MODEL
 - ASSIGN LEAD SCORE FOR EACH LEADS
 - TEST THE MODEL ON TRAIN SET
 - MODEL EVALUATION
 - MODEL TESTING
 - MEASURE ACCURACY OF THE MODEL AND OTHER METRICS
 - OBSERVATIONS
-

EDA

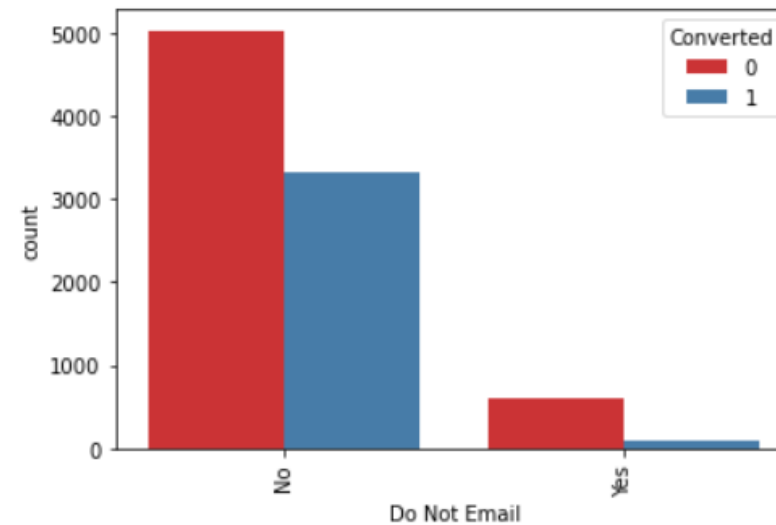
LEAD SOURCE VS CONVERTED

GOOGLE SEARCHES HAS HAD HIGH CONVERSIONS COMPARED TO MODES, WHILST REFERENCES HAS HAD HIGH CONVERSION RATE.



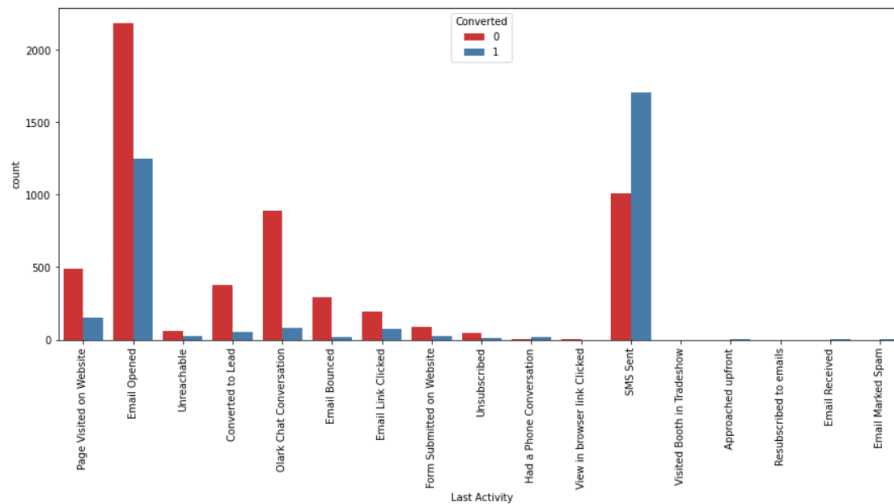
DO NOT EMAIL VS CONVERTED

GOOGLE SEARCHES HAS HAD HIGH CONVERSIONS AS COMPARED TO OTHER MODES, MEANWHILE REFERENCES HAS HAD HIGH CONVERSION RATE.



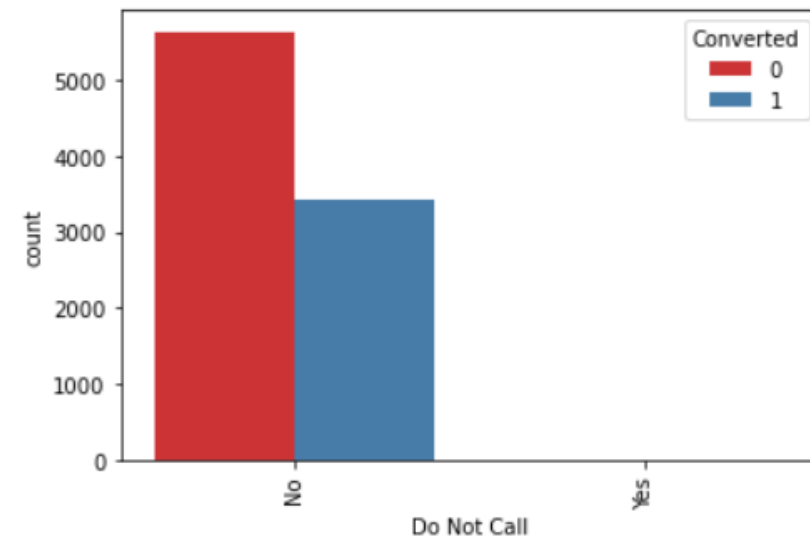
LAST ACTIVITY VS CONVERTED

SMS HAS SHOWN TO BE A PROMISING METHOD FOR GETTING HIGHER CONFIRMED LEADS, EMAILS ALSO HAS HIGH CONVERSIONS.

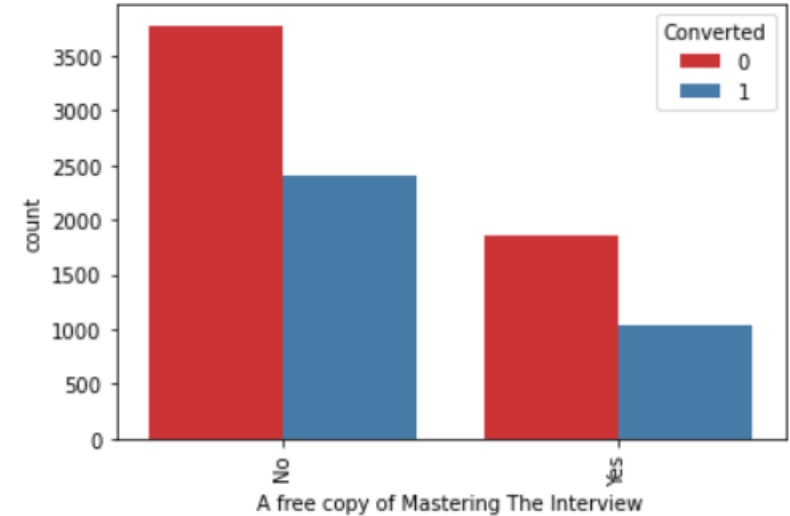


DO NOT CALL VS CONVERTED

MOST LEADS PREFER NOT TO INFORMED THROUGH PHONE.

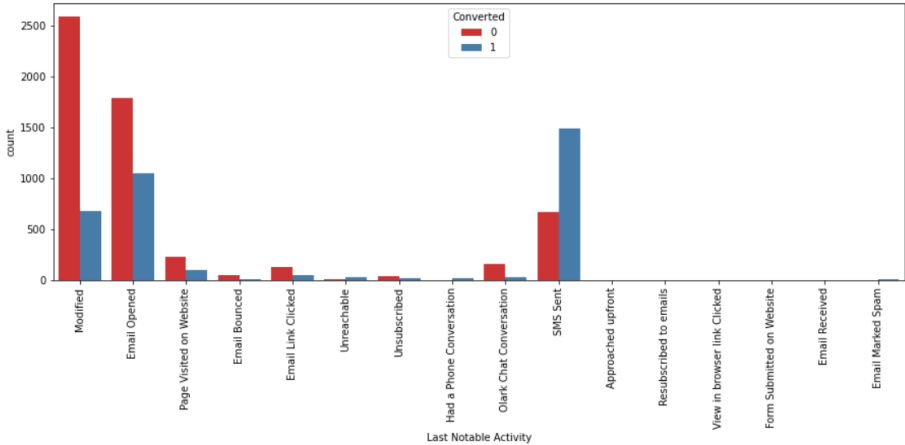


A FREE COPY OF MASTERING THE INTERVIEW VS CONVERTED
LEADS PREFER LESS COPIES OF INTERVIEWS.



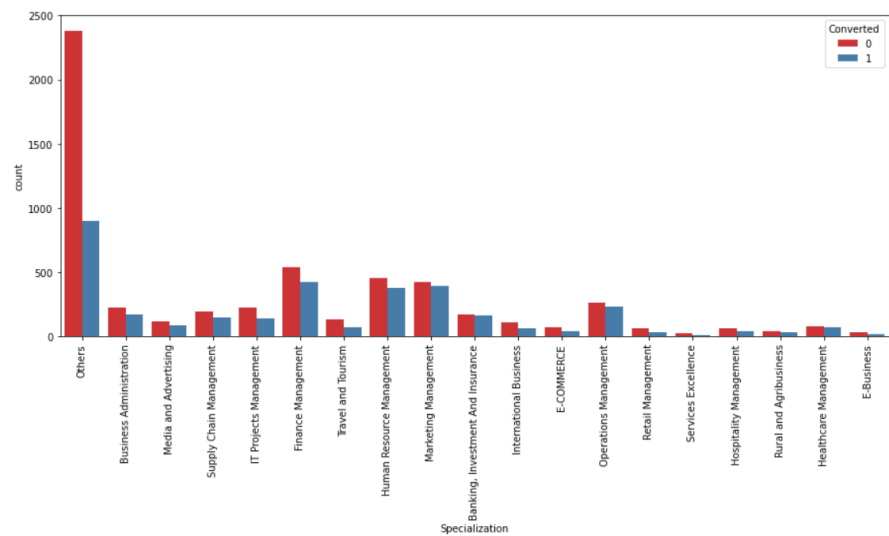
LAST NOTABLE ACTIVITY VS CONVERTED

MOST LEADS ARE CONVERTED WITH MESSAGES. EMAILS ALSO INCLUDE LEADS.



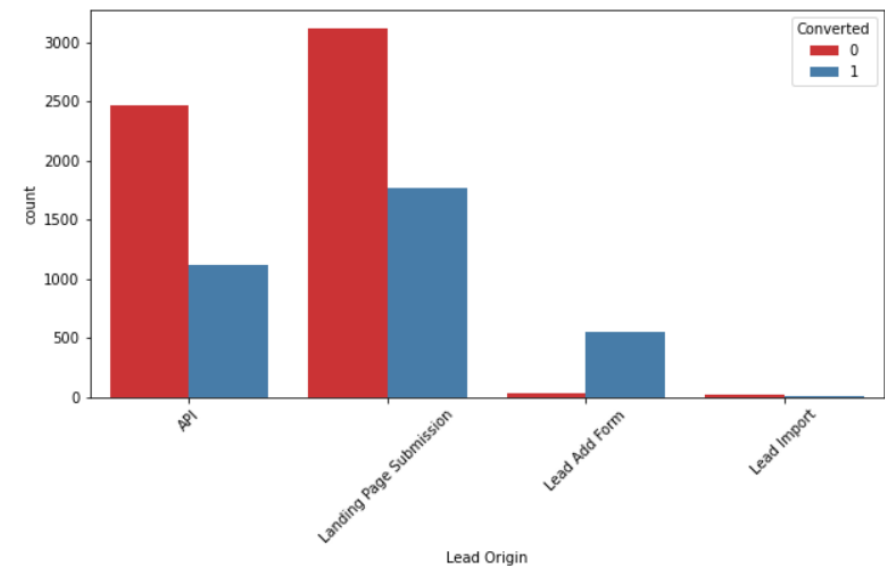
SPECIALIZATION VS CONVERTED

MOST OF THE LEADS HAVE NO INFORMATION. ON THE OTHER HAND, MARKETING MANAGEMENT, HR MANAGEMENT HAS HIGH CONVERSION RATES.



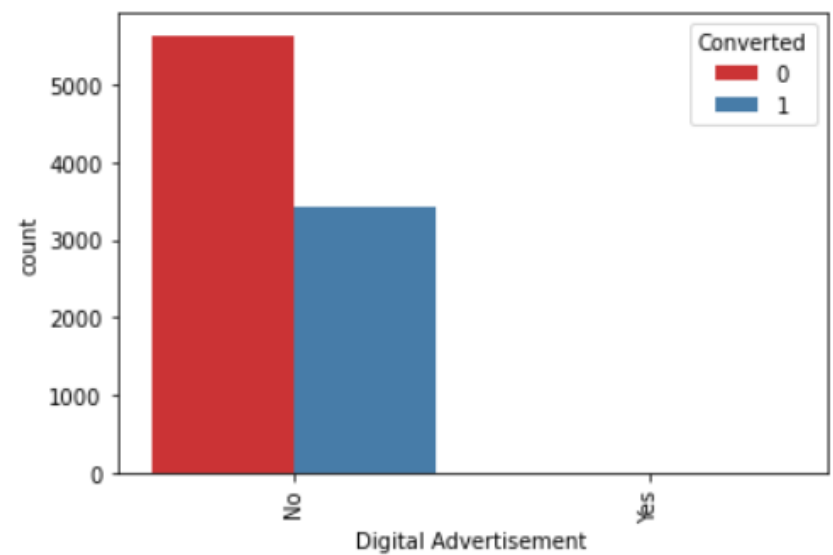
LEAD ORIGIN VS CONVERSION

LANDING PAGE HAS HAD HIGH LEAD CONVERSIONS.



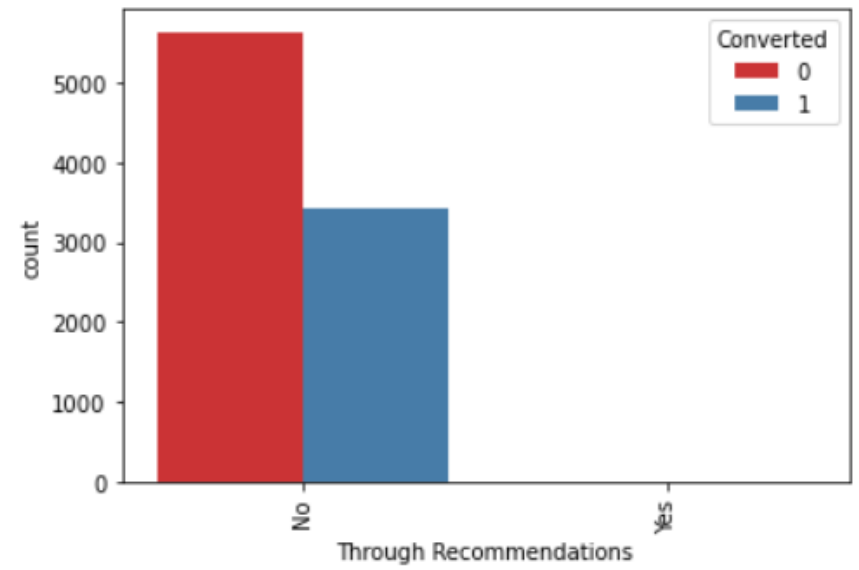
DIGITAL ADVERTISEMENTS VS CONVERTED

BASED ON THE BELOW GRAPH DIGITAL ADVERTISEMENTS DON'T FULFILL THE REQUIREMENTS.



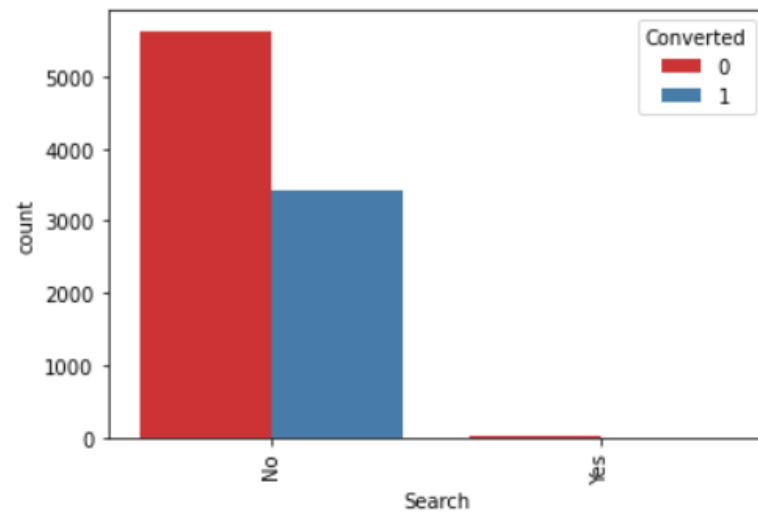
THROUGH RECOMMENDATIONS VS CONVERTED

FROM THE BELOW GRAPH THROUGH RECOMMENDATIONS DON'T FULFILL THE REQUIREMENT.



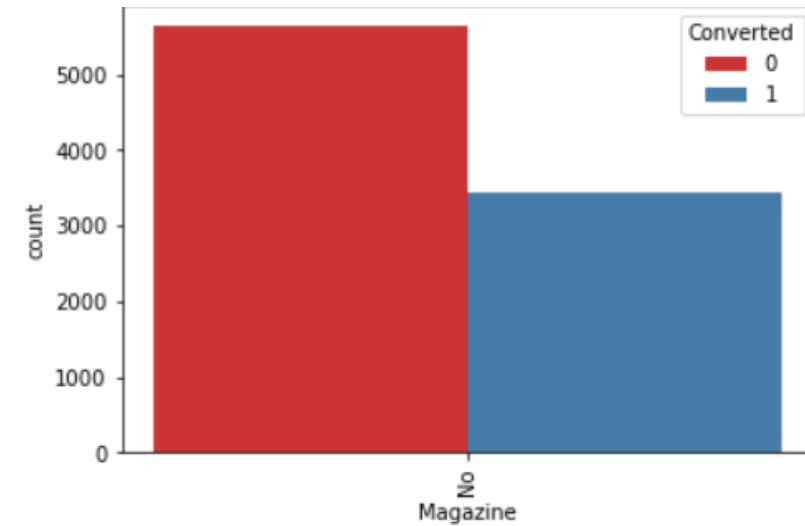
SEARCH VS CONVERTED

THE BELOW GRAPH SHOWS SEARCHES AREN'T GOOD SOURCE OF PROMISING LEADS.



MAGAZINE VS CONVERTED

MAGAZINES DON'T HAVE HIGHER CONVERSION RATE.



MODEL BUILDING:

Necessary Steps To Be Taken:

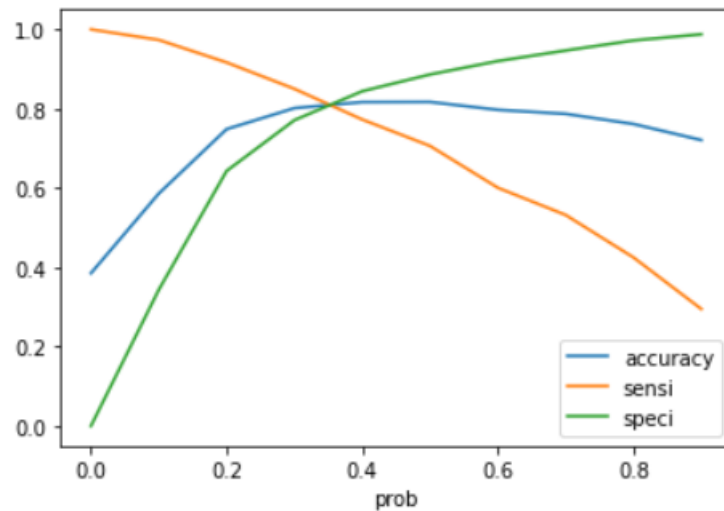
- SPLITTING INTO TRAIN-TEST SET
- SCALE THE VARIABLES IN TRAIN SET
- BUILD THE FIRST MODEL
- USE RFE TO ELIMINATE LESS RELEVANT VARIABLES
- BUILD THE NEXT MODELS
- ELIMINAE VARIABLES BASED ON HIGH P-VALUES
- CHECK VIF VALUES FOR ALL EXISTING COLOMNS
- PREDICT USING TRAIN SET
- EVALUATE ACCURACY AND OTHER METRIC
- PREDICT USING TEST SET
- PRECISION AND RECALL ANALYSIS ON TEST PREDICTIONS

MODEL EVALUATION

TRAIN:

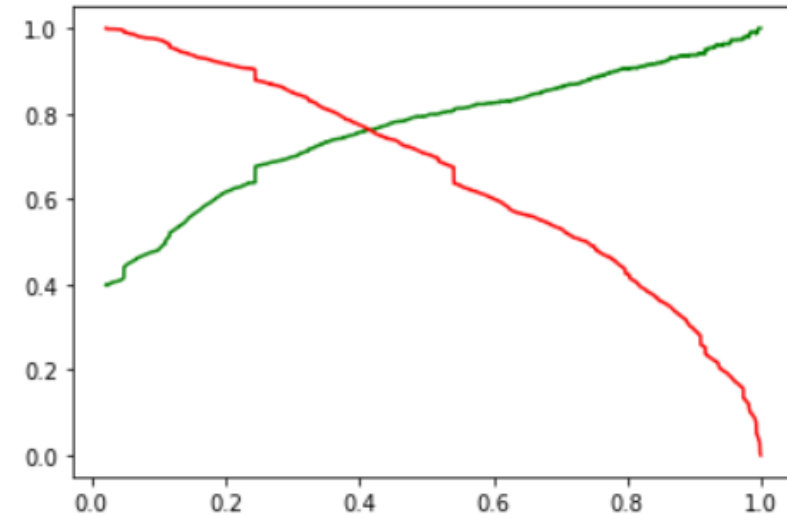
ACCORDING TO THE TRAIN SET

- ACCURACY : 81.0%
- SENSITIVITY : 81.7%
- SPECIFICITY : 80.6%



ACCORDING TO THE TRAIN SET

- PRECISION : 79.5%
- RECALL : 70.6%

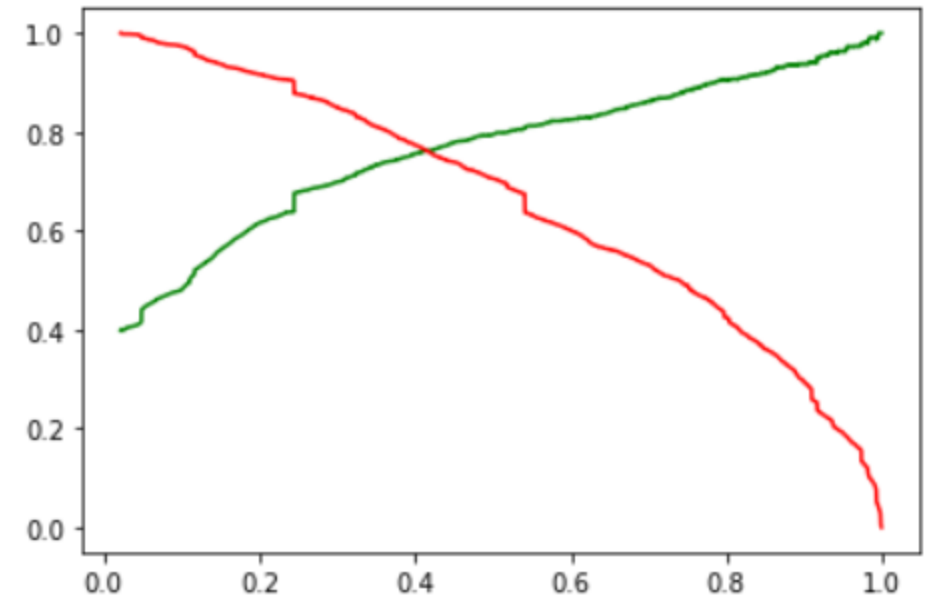


MODEL EVALUATION: TEST:

- Precision : 74.4%
- Recall : 75.5%

Test set threshold has been set as 0.41

- Accuracy : 80.4%
- Sensitivity : 80.4%
- Specificity : 80.5%



CONCLUSION

EDA:

- People spending higher than average time are promising leads. SMS's can have high impact on lead conversions.
- Marketing management, HR management has high conversion rate. People from specializations can also be promising leads.
- References and offers for referring a lead can be good source for high conversions.
- Aler texts or info seem to have high lead conversion rates.

LOGISTIC REGRESSION:

- The threshold has been selected from Accuracy, Sensitivity, Specificity measures and precision, recall curves.
- The model shows 81% Accuracy, 76% Sensitivity and 83% Specificity.
- Therefore, the overall model proves to be accurate.

THANK YOU
