# LEAD SCORE CASE STUDY

PRESENTED BY:

DHRUV SINHA

SUDEEP MENON

STRATEGY:

1. SOURCE THE DATA FOR ANALYSIS

2. CLEAN AND PREPARE THE DATA

3. EXPLORATORY DATA ANALYSIS.

4. FEATURE SCALING

5. SPLITTING THE DATA INTO TEST AND TRAIN DATASET.

6. BUILDING A LOGISTIC REGRESSION MODEL

7. EVALUATING THE MODEL BY USING DIFFERENT METRICS - SPECIFICITY AND SENSITIVITY OR PRECISION

AND RECALL.

8. APPLYING THE BEST MODEL IN TEST DATA BASED ON THE SENSITIVITY AND SPECIFICITY METRICS.
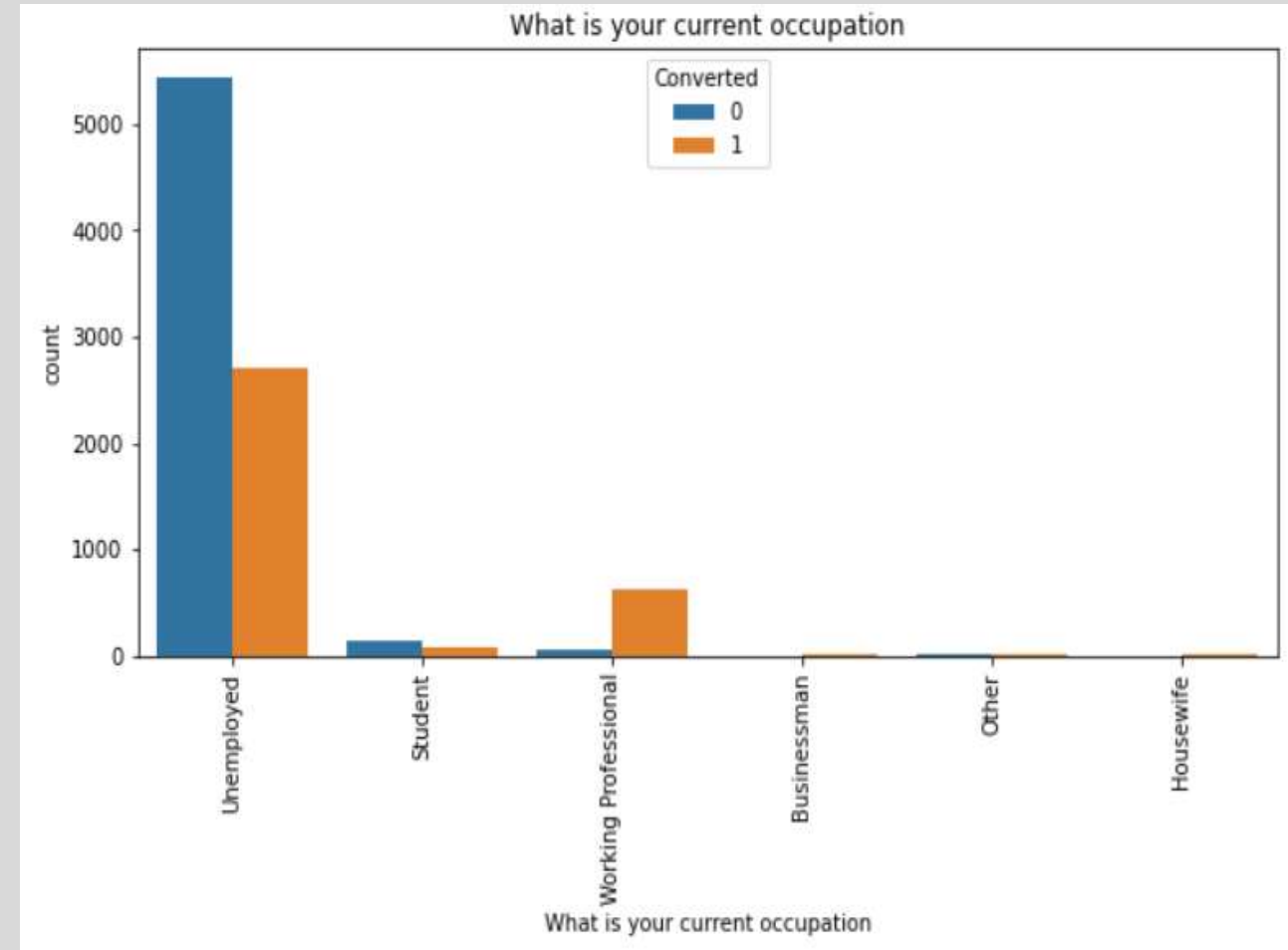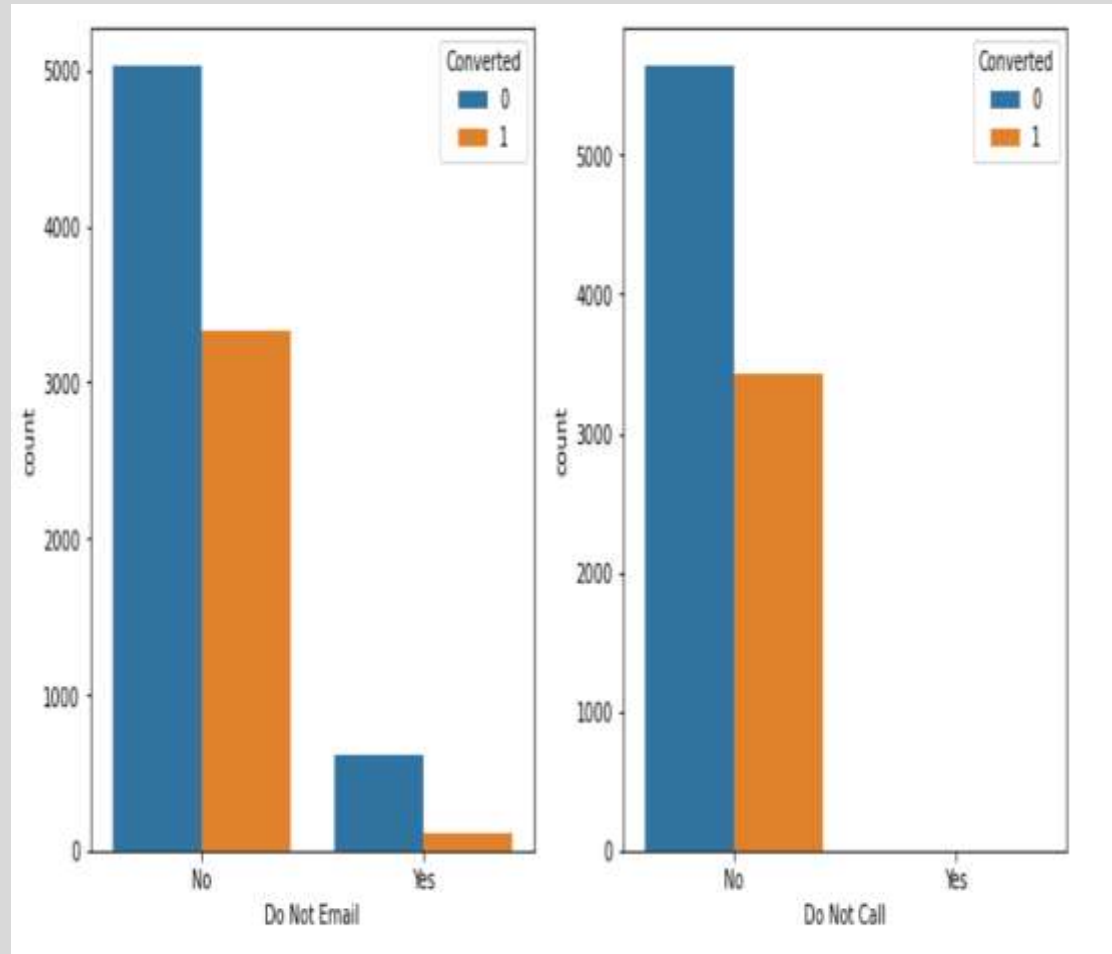
DATA INSIGHTS:

The conversion rates were high for lead origin of add form, lead source of reference and working professionals.
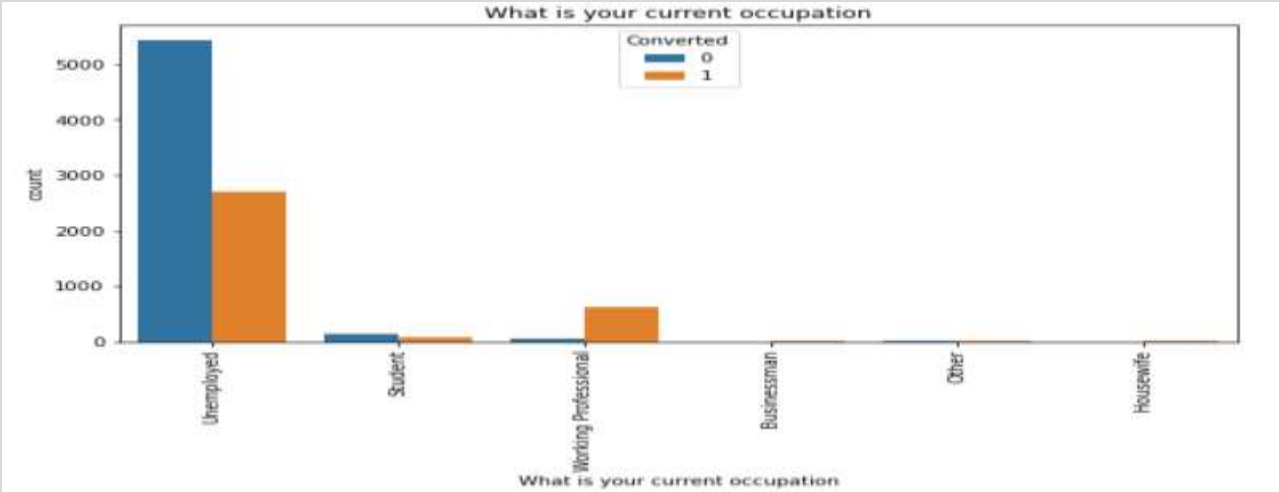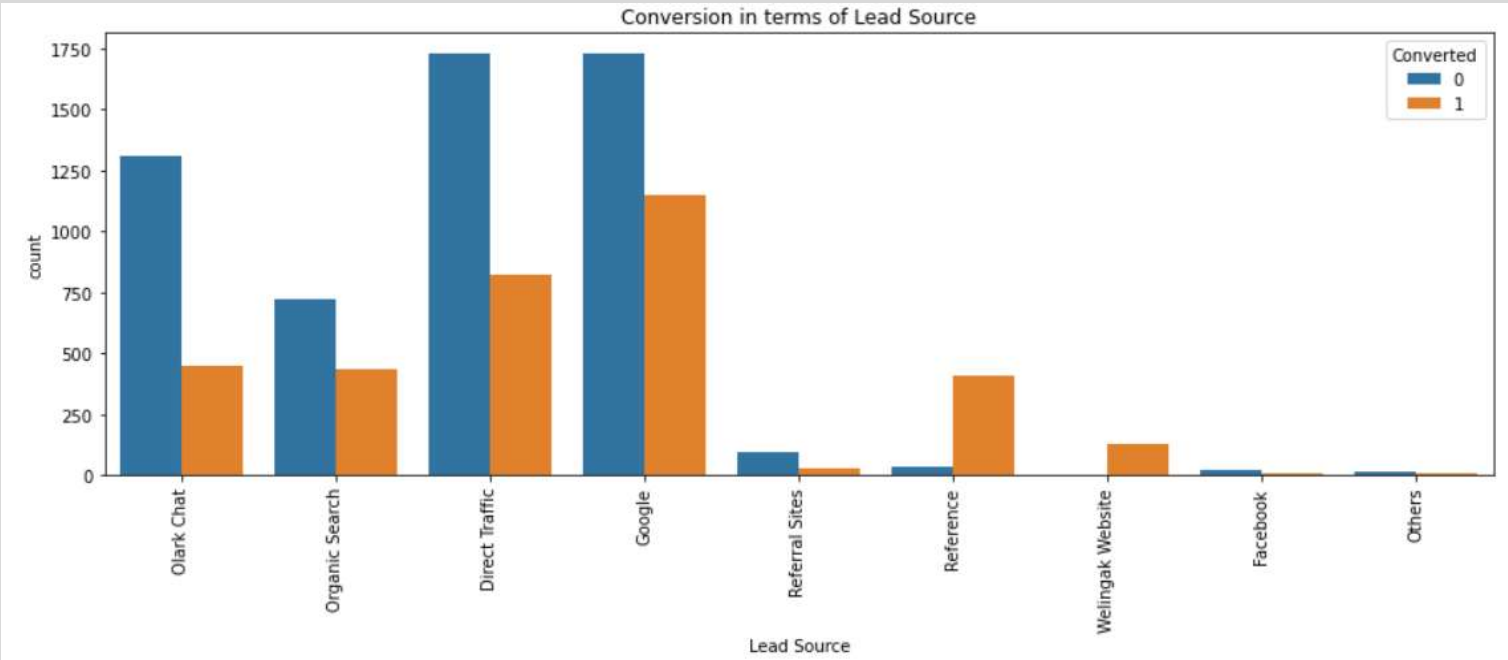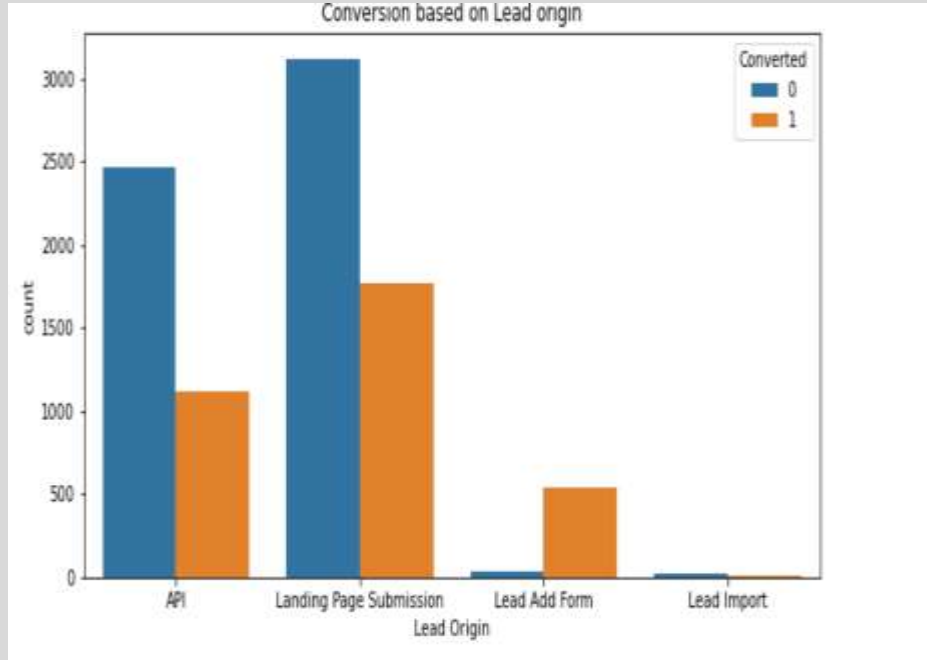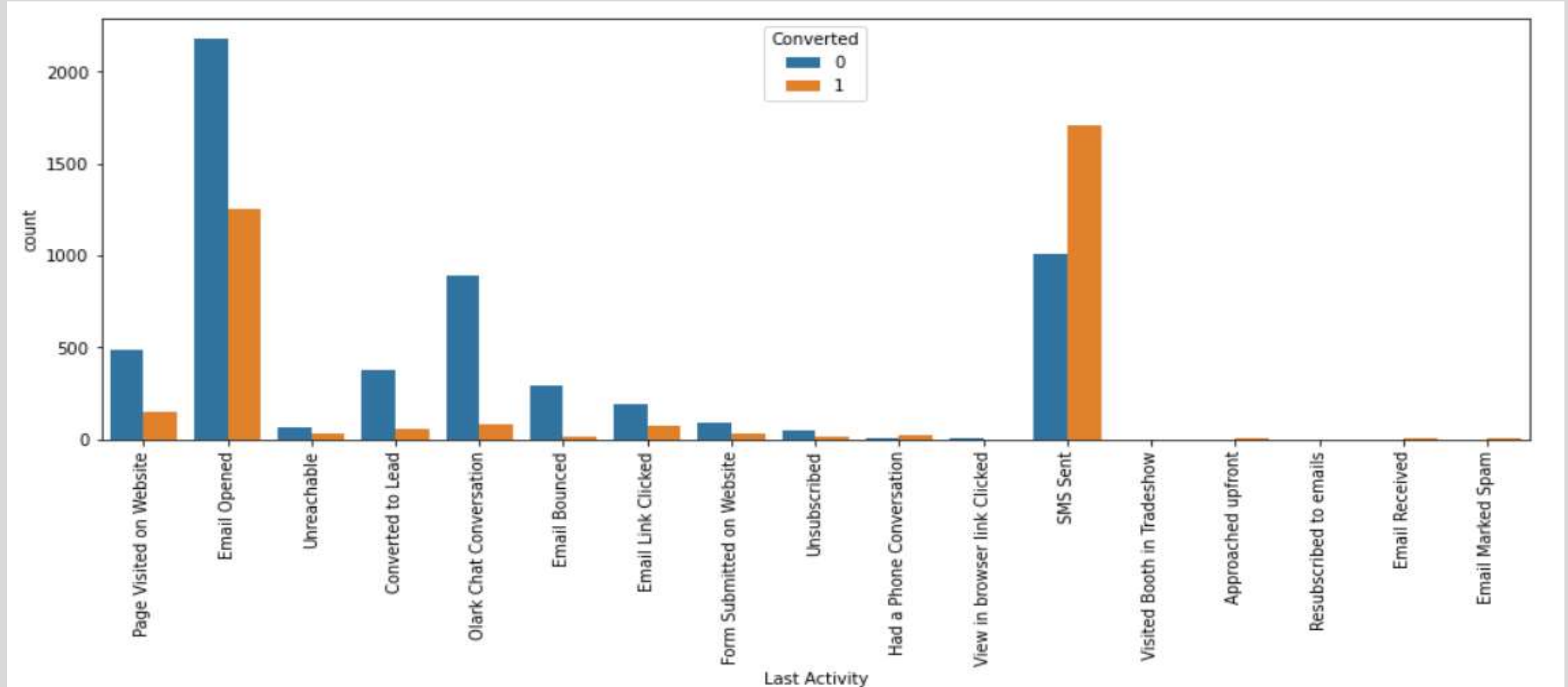
DATA INSIGHTS:

Major leads were generated from emails sent and calls made and more conversion happened with people who are working professionals but unemployed people generated the most leads.

DATA INSIGHTS:

Last activity value of SMS sent had more conversion.

VARIABLES IMPACTING THE CONVERSION RATE AND THE MODEL:

1. DO NOT EMAIL

2. TOTAL TIME SPENT ON WEBSITE.

3. LEAD SOURCE - OLARK CHAT, REFERENCE, WELINGAK WEBSITE.

4. LAST ACTIVITY - CONVERTED TO LEAD, EMAIL BOUNCED, OLARK CHAT CONVERSATION, PAGE VISITED ON WEBSITE.

5. WHAT IS YOUR CURRENT OCCUPATION - HOUSEWIFE, STUDENT, UNEMPLOYED.

6. LAST NOTABLE ACTIVITY - EMAIL LINK CLICKED, EMAIL OPENED, MODIFIED, OLARK CHAT CONVERSATION.

MODEL EVALUATION:

The following graph depicts the optimal cut-off of 0.36 based on accuracy, specificity and sensitivity.
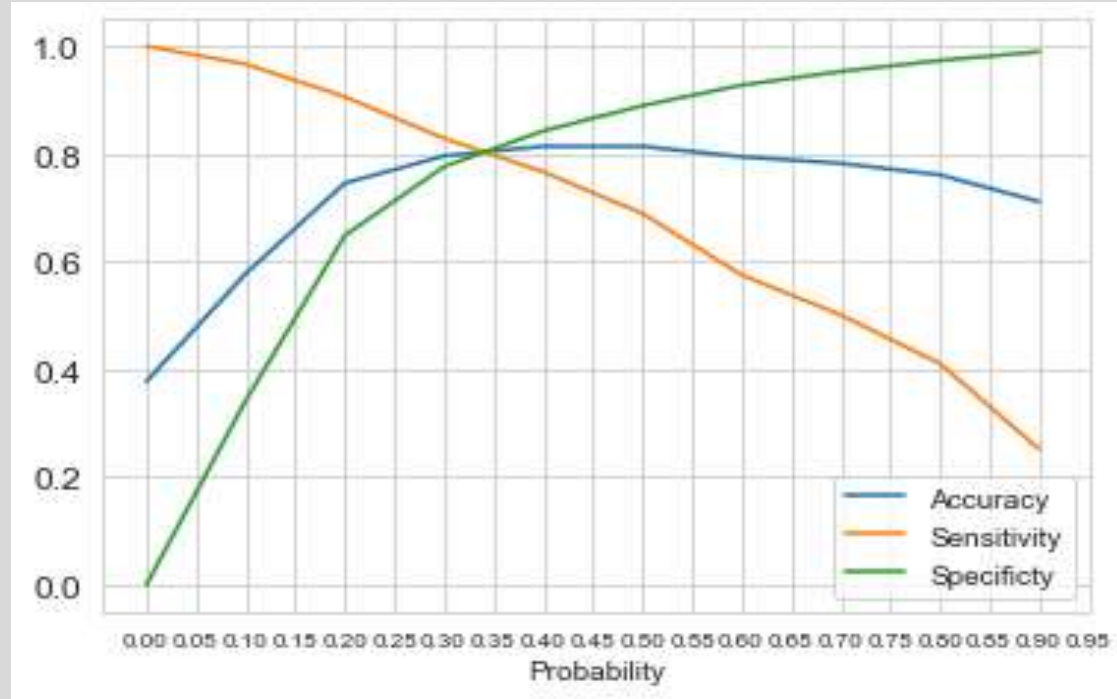
Confusion Matrix:

3235	719
497	1900

MODEL EVALUATION:

The graph depicts an optimal cut-off of 0.41based on precision and recall.

CONCLUSION:

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. –

1. Accuracy, sensitivity and specificity values of test set are around 81.16%, 80.3% and 81.7% which are approximately closer to the respective values calculated using trained set.
2. The top 3 variables that contribute for lead getting converted in the model are
a. Current occupation of housewife
b. Lead source from welingak website.
c. Lead source as reference
d. Total time spent on website