Lead Score Case Study

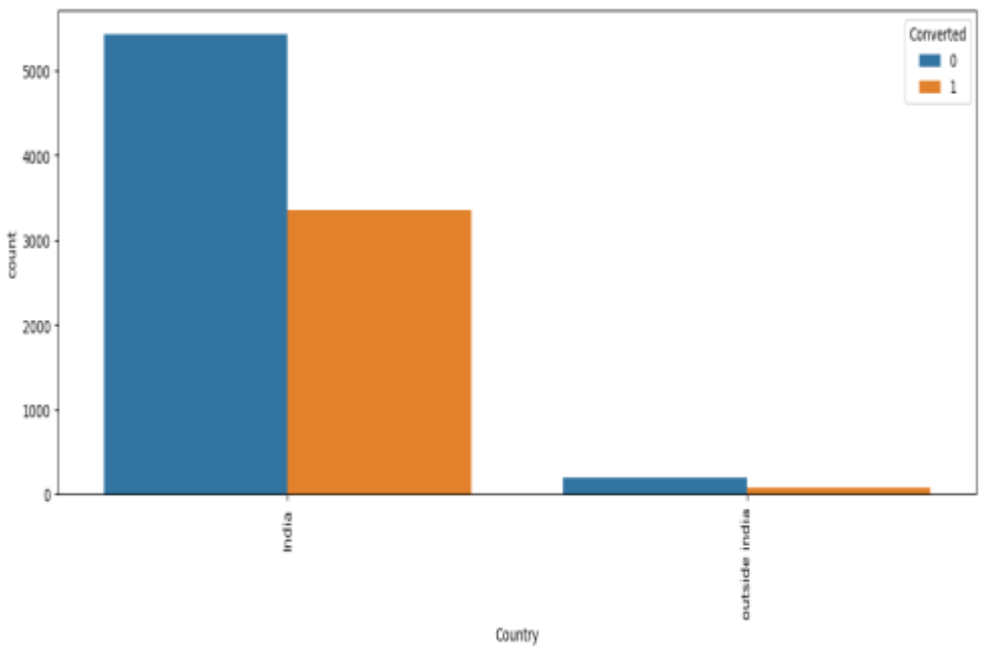
Submitted by:

Koushal Jaral

Akhil Kodali

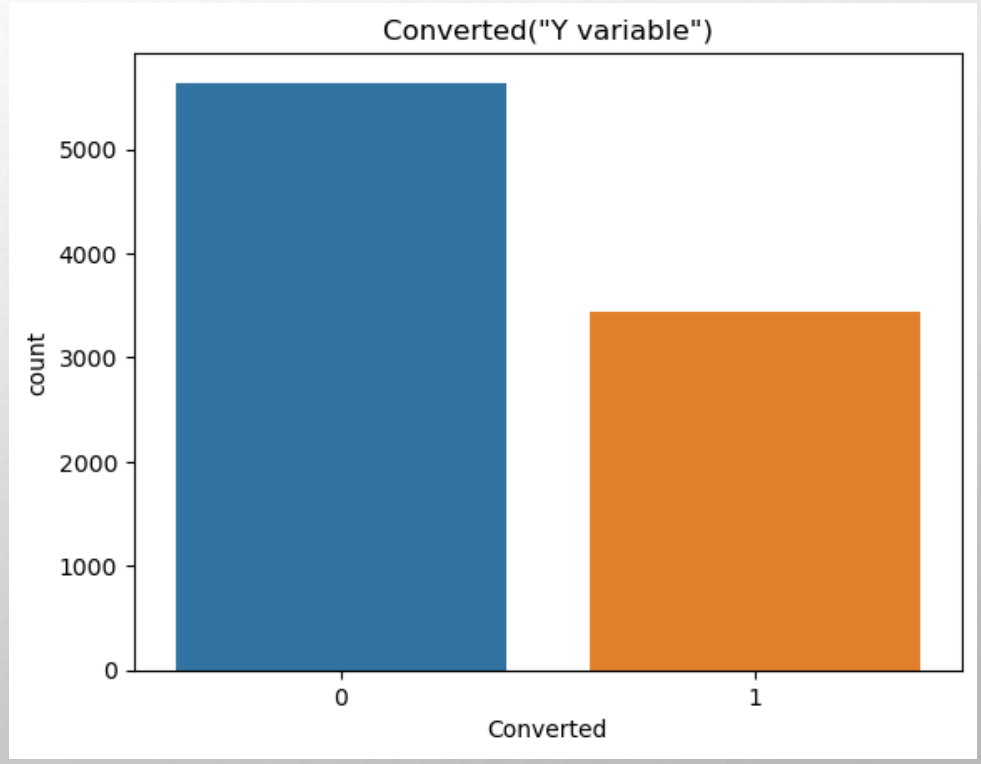
Konark Kashyap

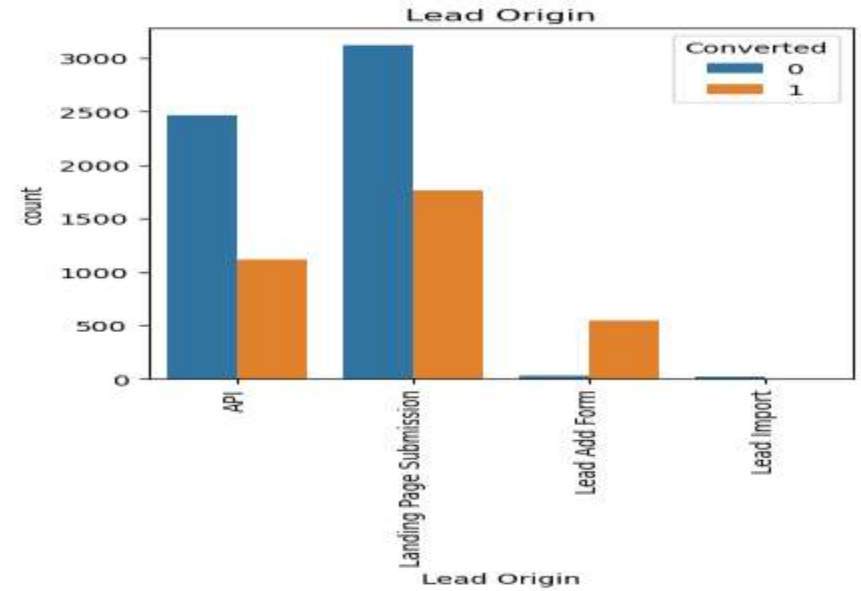
EXPLORATORY DATA ANALYSIS

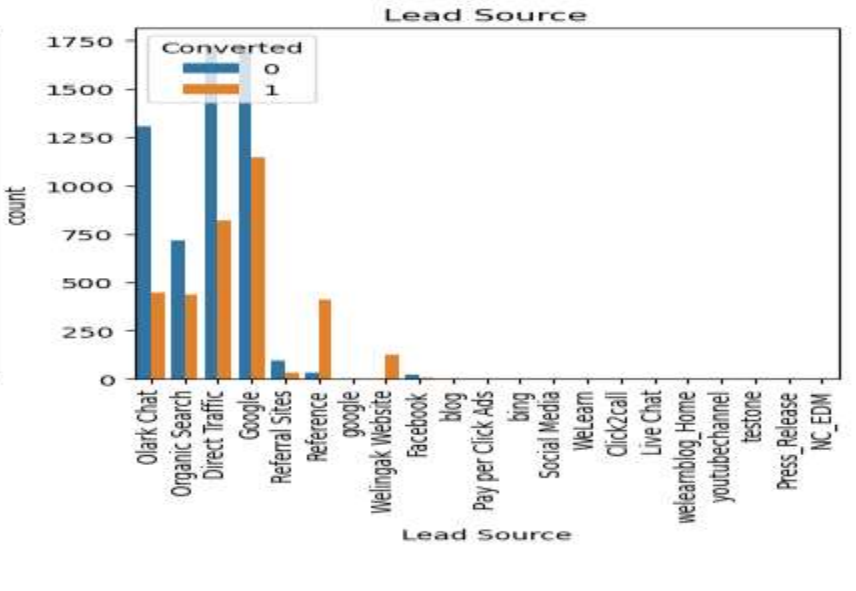


Most of the customers are from India

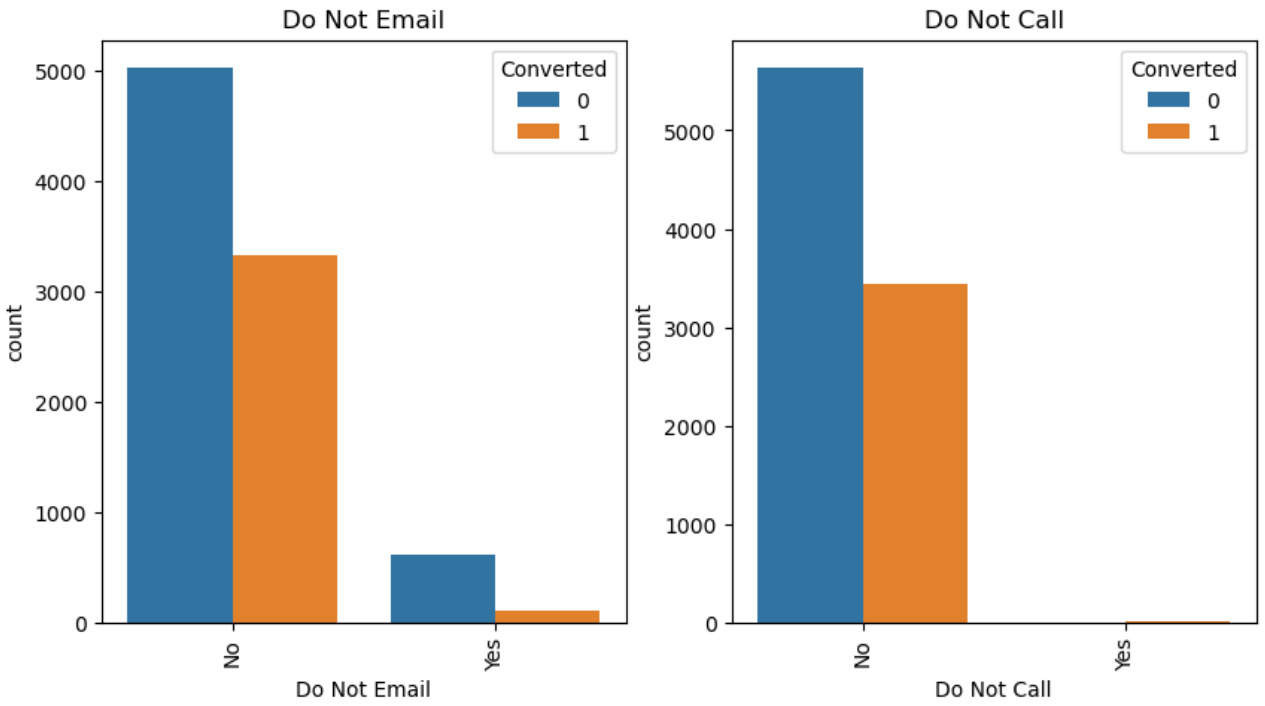
Here we can observe that we have a 38% conversion rate which is pretty less than the target given by the CEO i.e of 80%





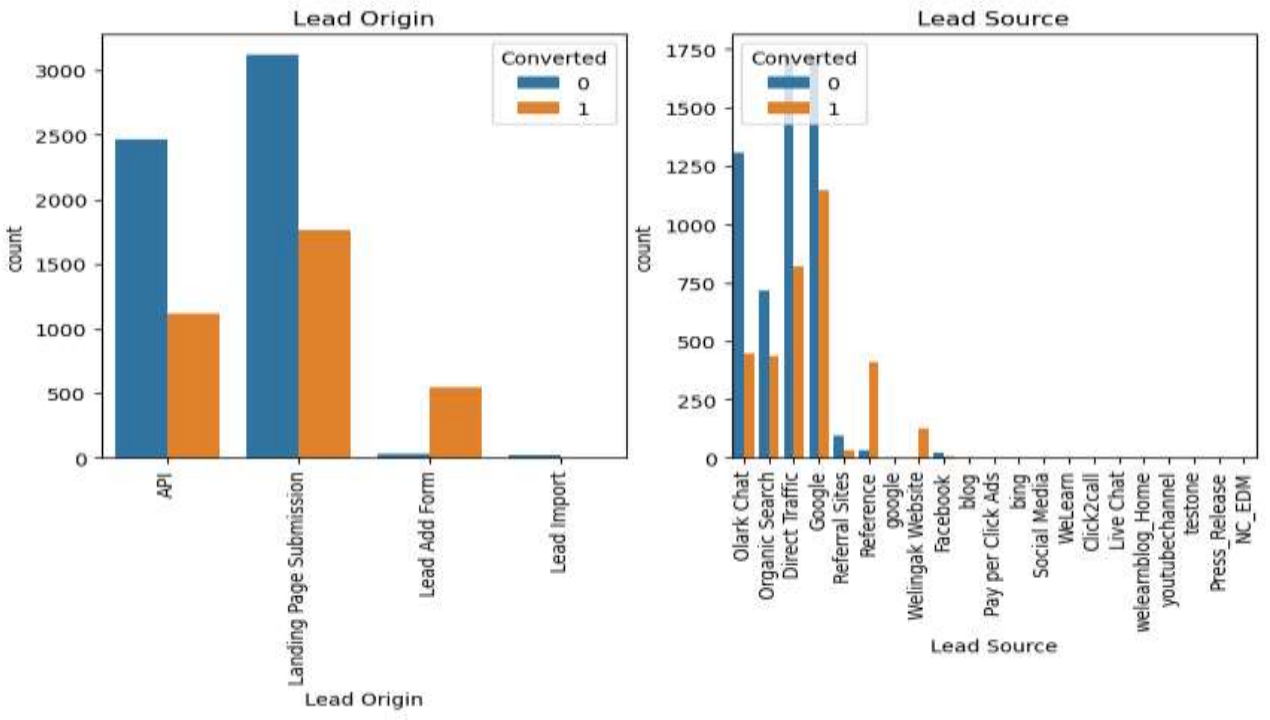


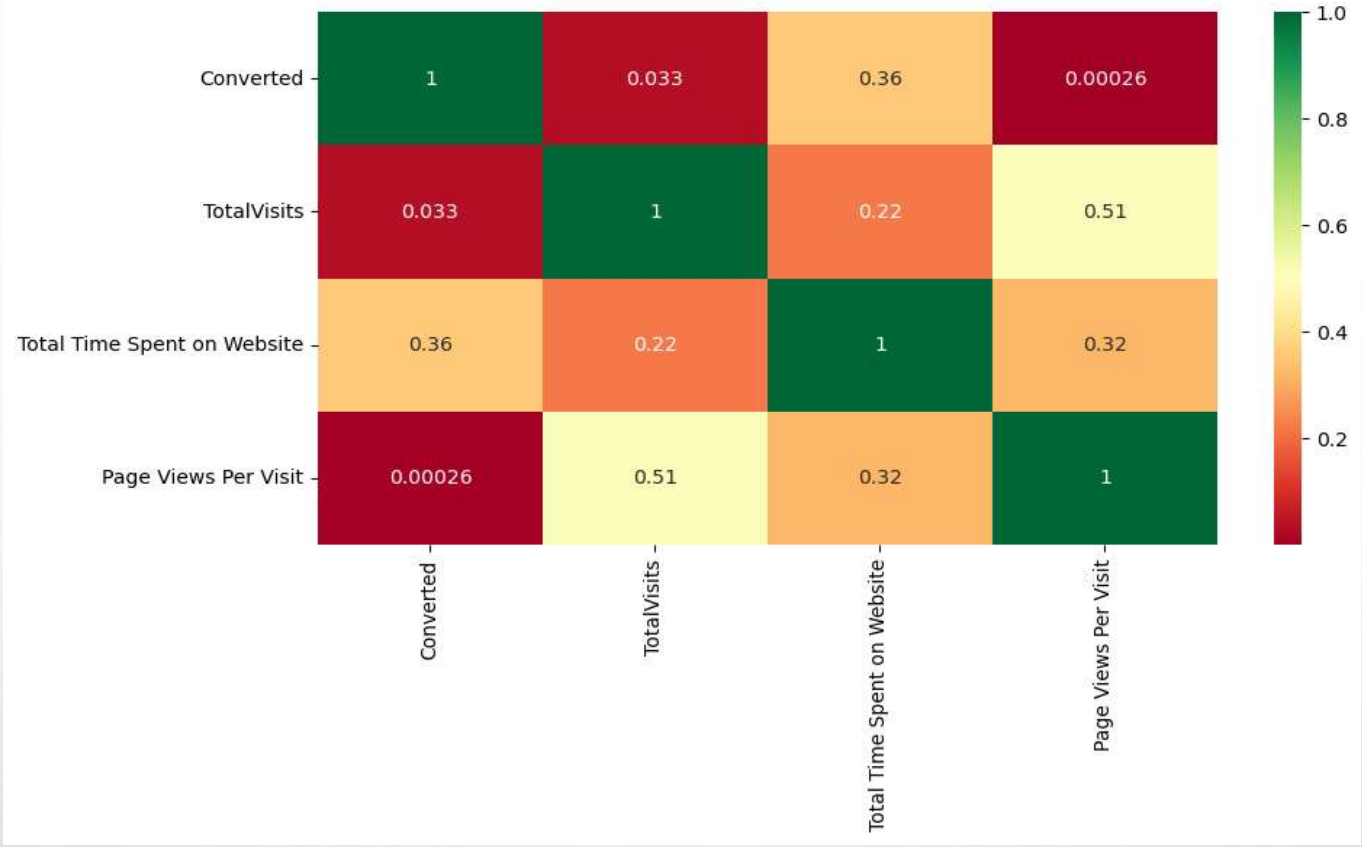
It is evident from the above graphs that the ratio of conversion to visists is highest in refferals along with Welingak Websites and google. Also the lead orignated by the lead add form is performing great as well. Hence, we can focus on such customers for better conversions



In the Lead Origin, maximum number of conversions are from landing page submission

There is a good conversion rate by sending emails and calling the customers.

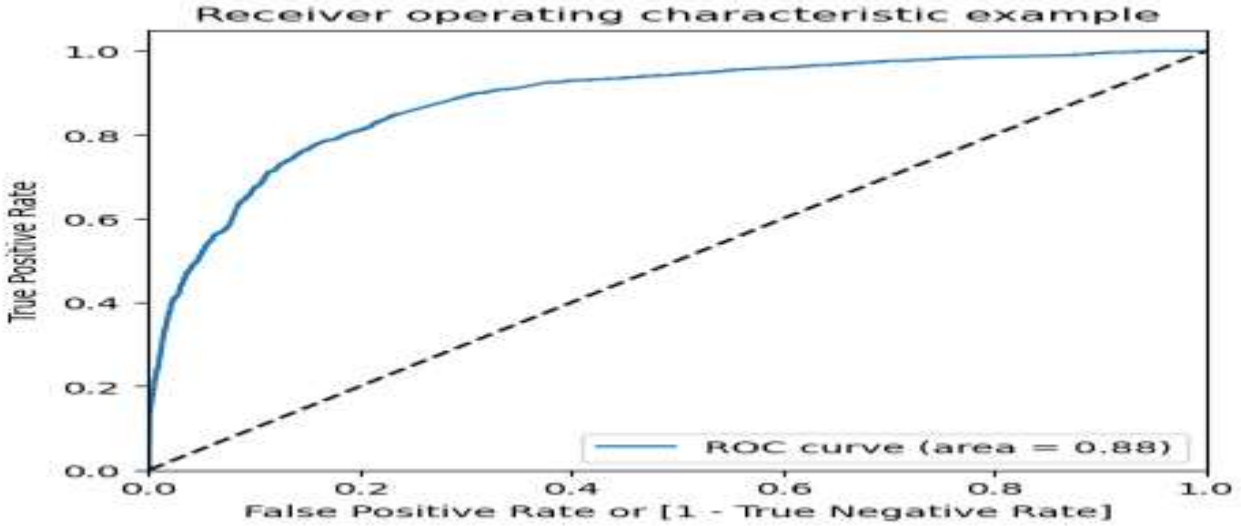


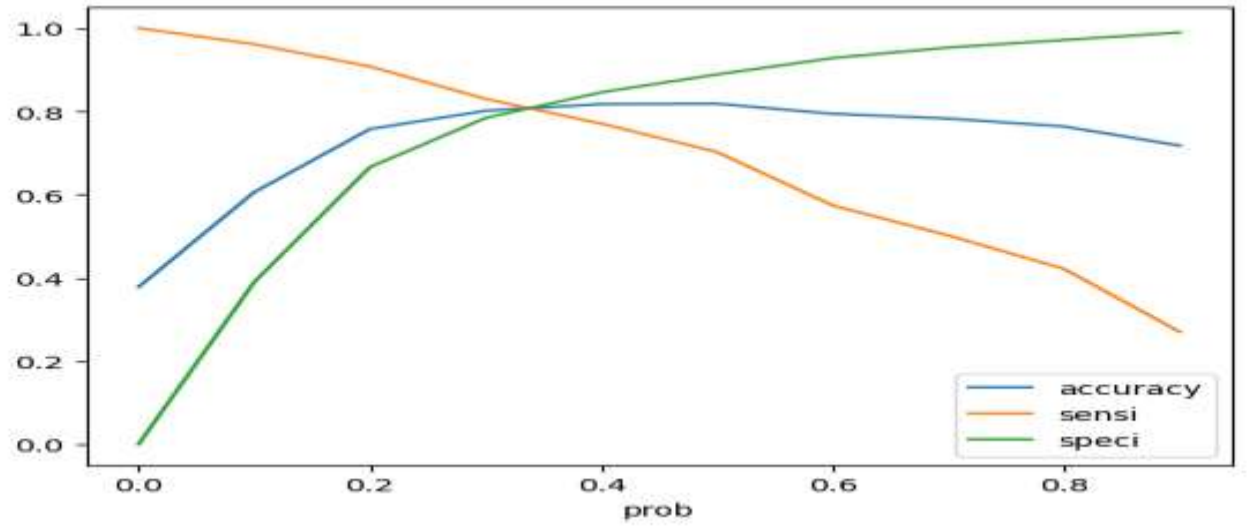


Here we can observe that people who have spend a considerable amount of time in our website are likely to convert which is an evident factor. Hence, approaching them will be highly lucrative.

The provided exploratory data analysis (EDA) suggests that certain elements within the dataset contain limited data, which could potentially result in reduced relevance for our subsequent analysis. Hence, removing them from our dataset will be beneficial.

MODAL EVALUATION ON TRAIN DATA SET

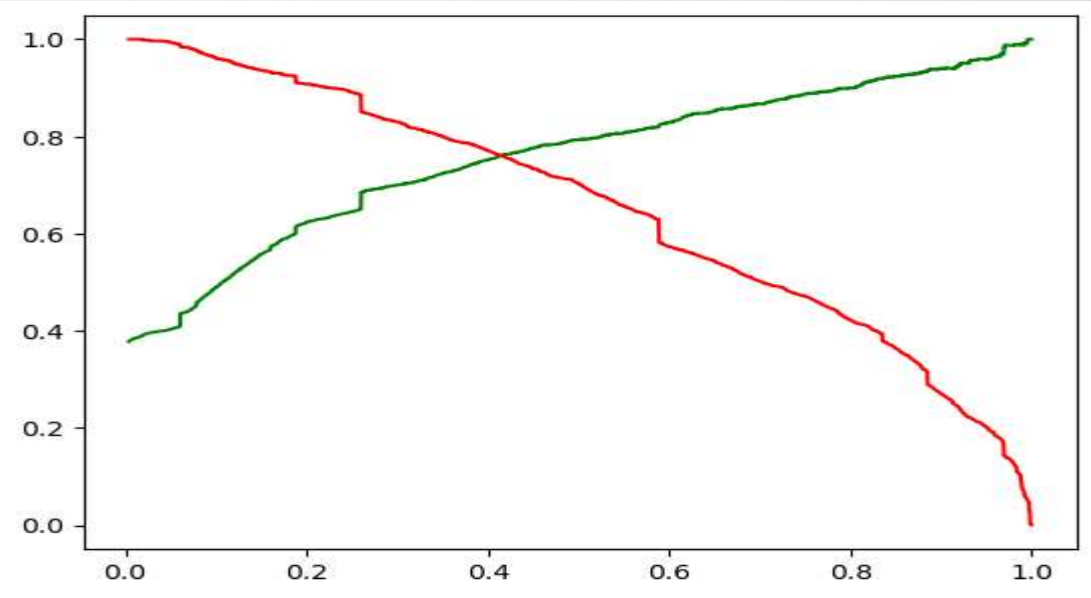




→ The area under ROC curve is 0.88 which depicts that the model is working excellent.
→ From the graph it is visible that the optimal cut off is approximately 0.35.

PRECISION AND RECALL ON TRAIN DATASET

array([[3333, 582], [563, 1815]], dtype=int64)
→ Precision = TP / TP + FP = 0.757196495619524 4
→ Recall = TP / TP + FN=0.763246425567704



MODAL EVALUATION ON TEST DATA SET

- CONFUSION MATRIX => ARRAY([[1426, 251], [273, 748]], DTYPE=INT64)
- → PRECISION = TP / TP + FP = 0.748748748748748
- → RECALL = TP / TP + FN=0.732615083251714

WITH THE CURRENT CUT OFF AS 0.41 WE HAVE PRECISION AROUND 75% , RECALL AROUND 73% AND ACCURACY 80.5%.

THE MODEL SEEMS TO PREDICT THE CONVERSION RATE VERY WELL AND WE SHOULD BE ABLE TO GIVE THE CEO CONFIDENCE IN MAKING GOOD CALLS BASED ON THIS MODEL

CONCLUSION

IT WAS FOUND THAT THE VARIABLES THAT MATTERED THE MOST IN THE POTENTIAL BUYERS ARE (IN DESCENDING ORDER)

- TOTALVISITS
- THE TOTAL TIME SPEND ON THE WEBSITE.
- LEAD ORIGIN_LEAD ADD FORM
- LEAD SOURCE_DIRECT TRAFFIC
- LEAD SOURCE_GOOGLE
- LEAD SOURCE_WELINGAK WEBSITE
- LEAD SOURCE_ORGANIC SEARCH
- LEAD SOURCE_REFERRAL SITES
- LEAD SOURCE_WELINGAK WEBSITE
- DO NOT EMAIL_YES
- LAST ACTIVITY_EMAIL BOUNCED
- LAST ACTIVITY_OLARK CHAT CONVERSATION

WITH THE ABOVE ANALYSIS THE ORGANISATION CAN HAVE A CONVERSION ACCURACY OF MORE THAN 80% WHICH IS PRETTY ROBUST THE ORGANISATION TO FLOURISH.