LEAD SCORING CASE STUDY

NEETHU SANTHOSH KUMAR

R KEERTHIKA DEVI

PROBLEM STATEMENT

- AN EDUCATION COMPANY NAMED X EDUCATION SELLS ONLINE COURSES TO INDUSTRY PROFESSIONALS.
- MANY PROFESSIONALS WHO ARE INTERESTED IN THE COURSES LAND ON THEIR WEBSITE AND BROWSE FOR COURSES.
- ALTHOUGH X EDUCATION GETS A LOT OF LEADS, ITS LEAD CONVERSION RATE IS VERY POOR

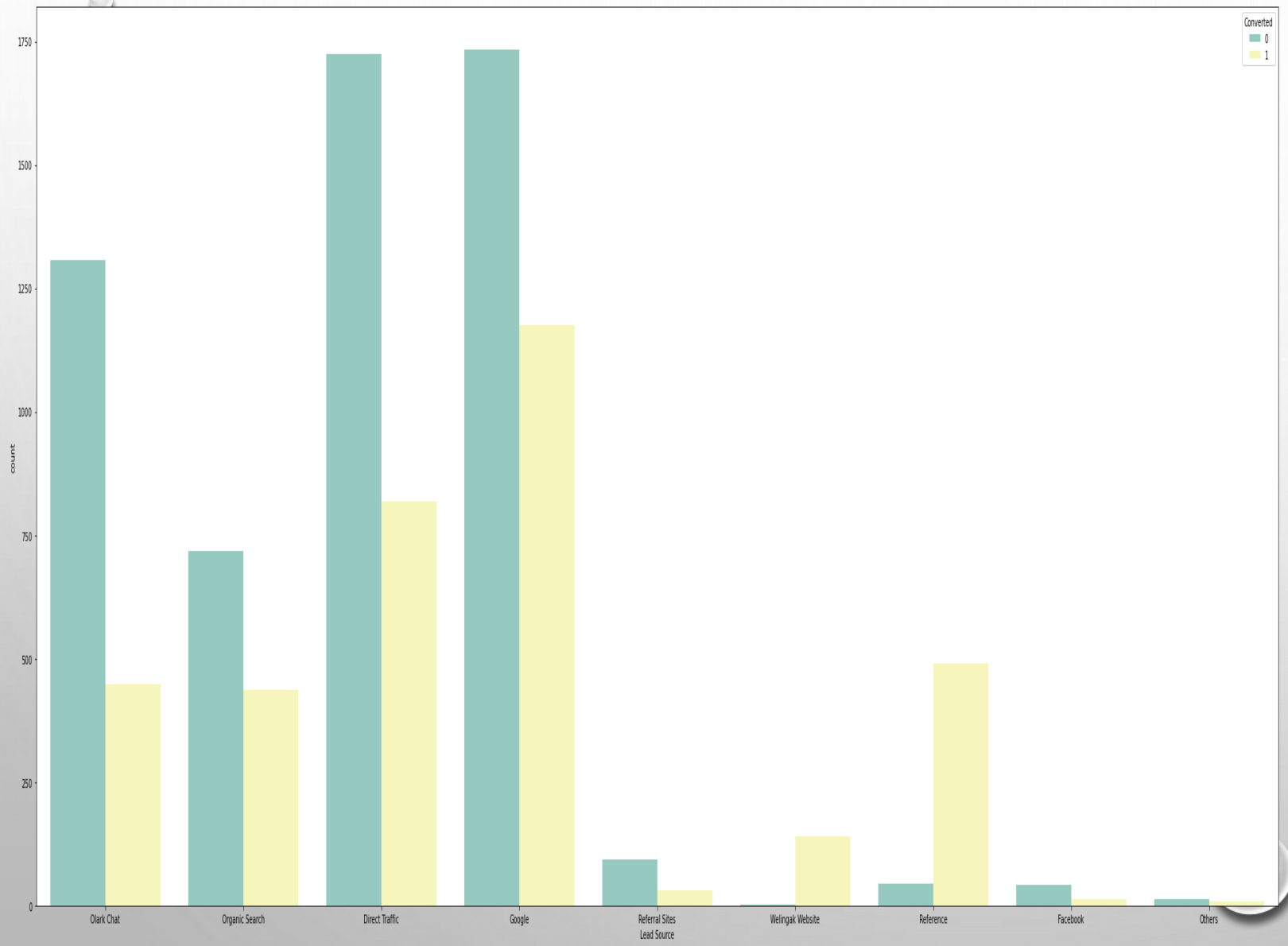
OBJECTIVE

- TO FIND OUT LEADS THAT ARE MOST LIKELY TO CONVERT INTO PAYING CUSTOMERS.
- THE COMPANY REQUIRES TO BUILD A MODEL WHEREIN EACH CUSTOMER IS ASSIGNED A LEAD SCORE SUCH THAT
- THE CUSTOMERS WITH HIGHER LEAD SCORE HAVE A HIGHER CONVERSION CHANCE
- AND THE CUSTOMERS WITH LOWER LEAD SCORE HAVE A LOWER CONVERSION CHANCE.

APPROACH

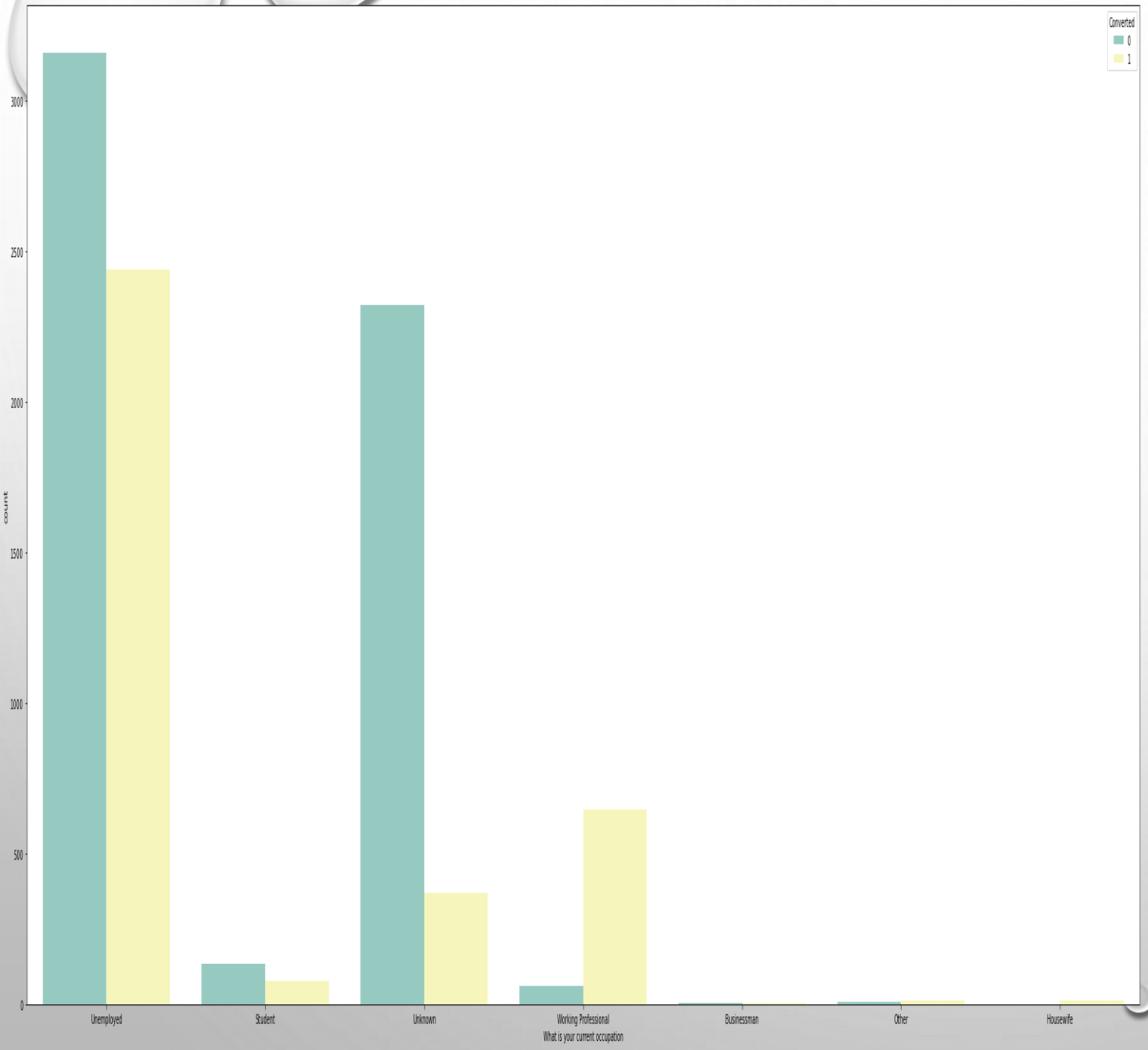
- DATA COLLECTION
- DATA CLEANING
- EXPLORATORY DATA ANALYSIS
- DATA PREPARATION
- TRAIN-TEST SPLIT
- SCALING
- BUILDING A MODELS

EDA



1.LEAD SOURCE INSIGHT

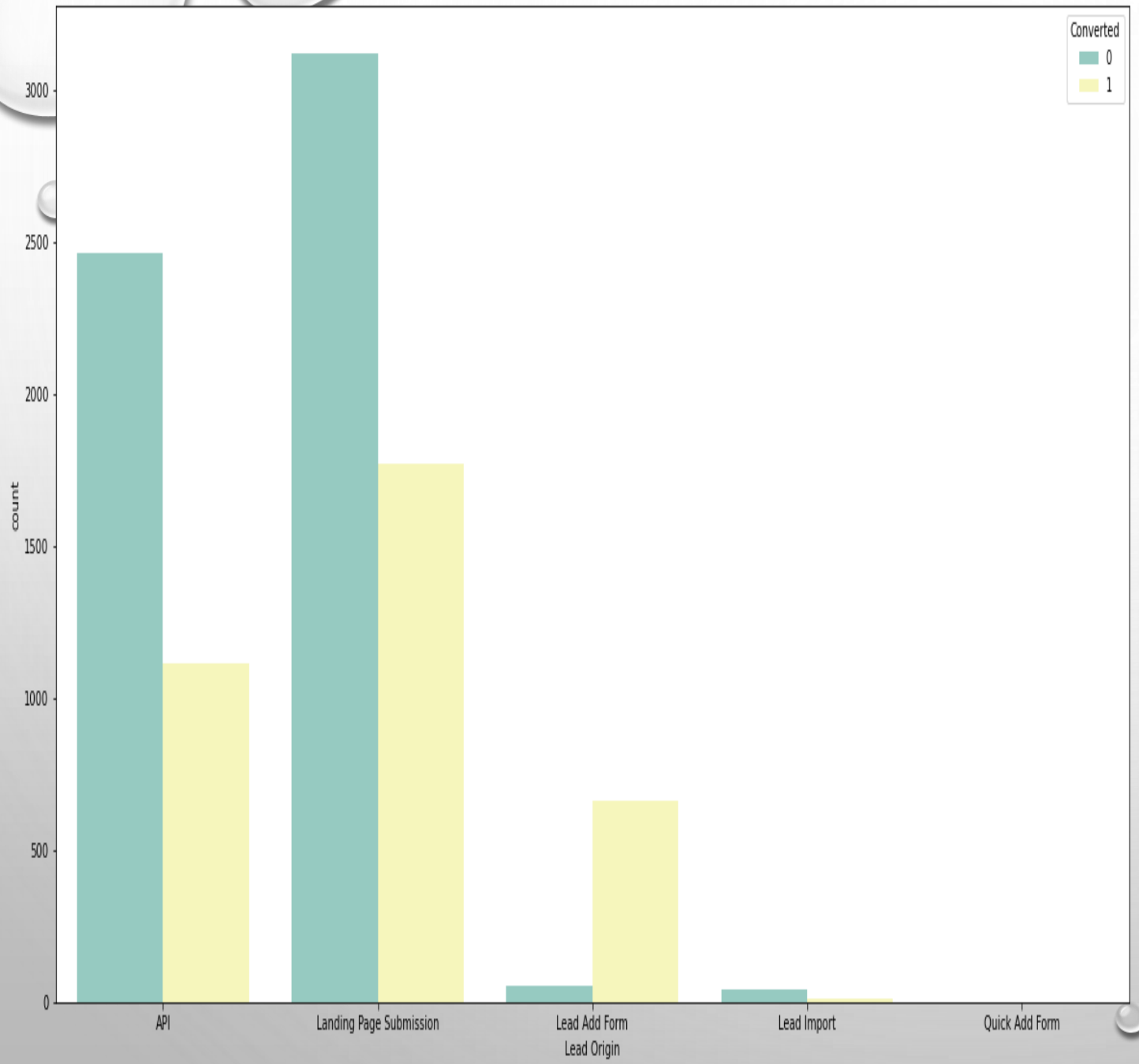
- MAXIMUM CONVERSION OCCURS THROUGH LEADS GENERATED FROM GOOGLE, DIRECT TRAFFIC AND OLARK CHAT.



2.WHAT IS YOUR CURRENT OCCUPATION

- HIGHEST LEADS ARE FROM UNEMPLOYED AND THEIR CONVERSION RATE IS ALSO HIGH.

- HIGHEST CONVERSION RATE IS FOR WORKING PROFESSIONALS.

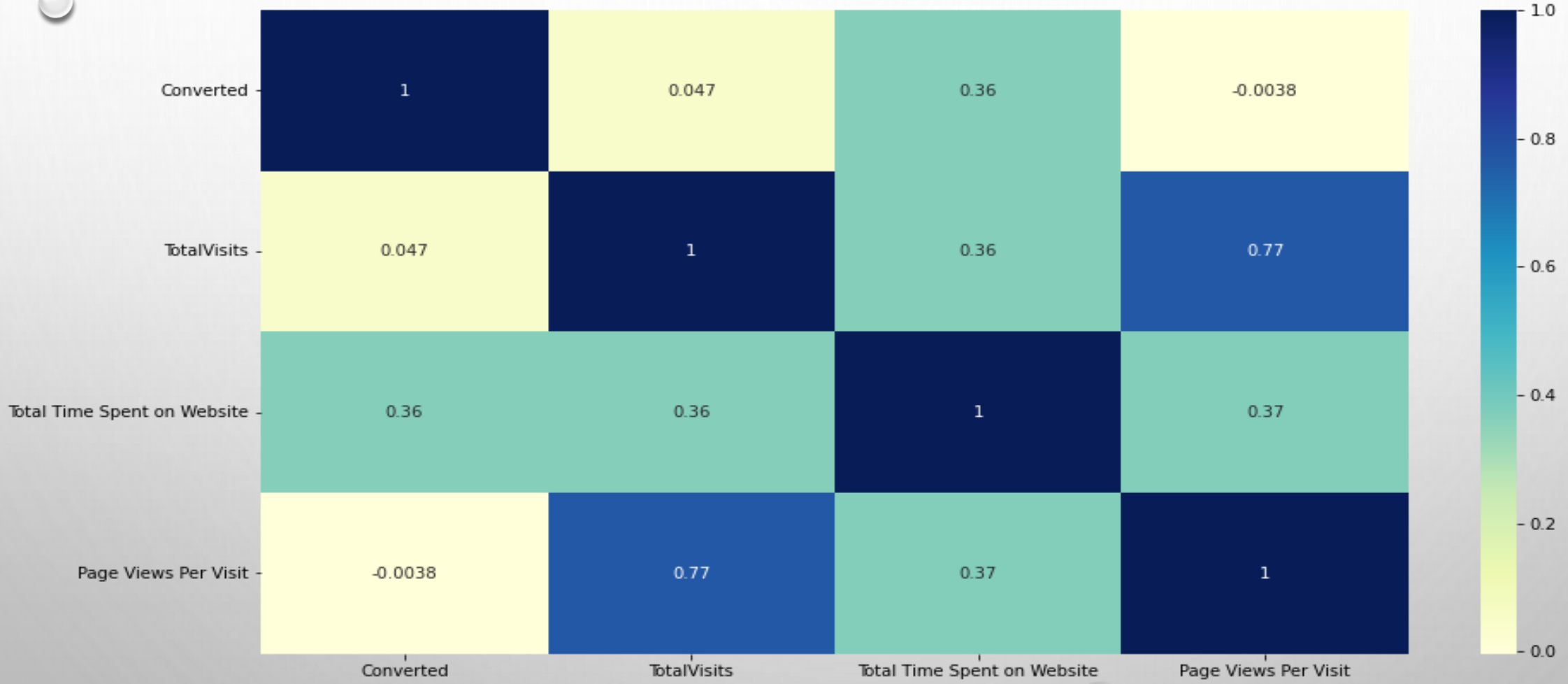


3. LEAD ORIGIN

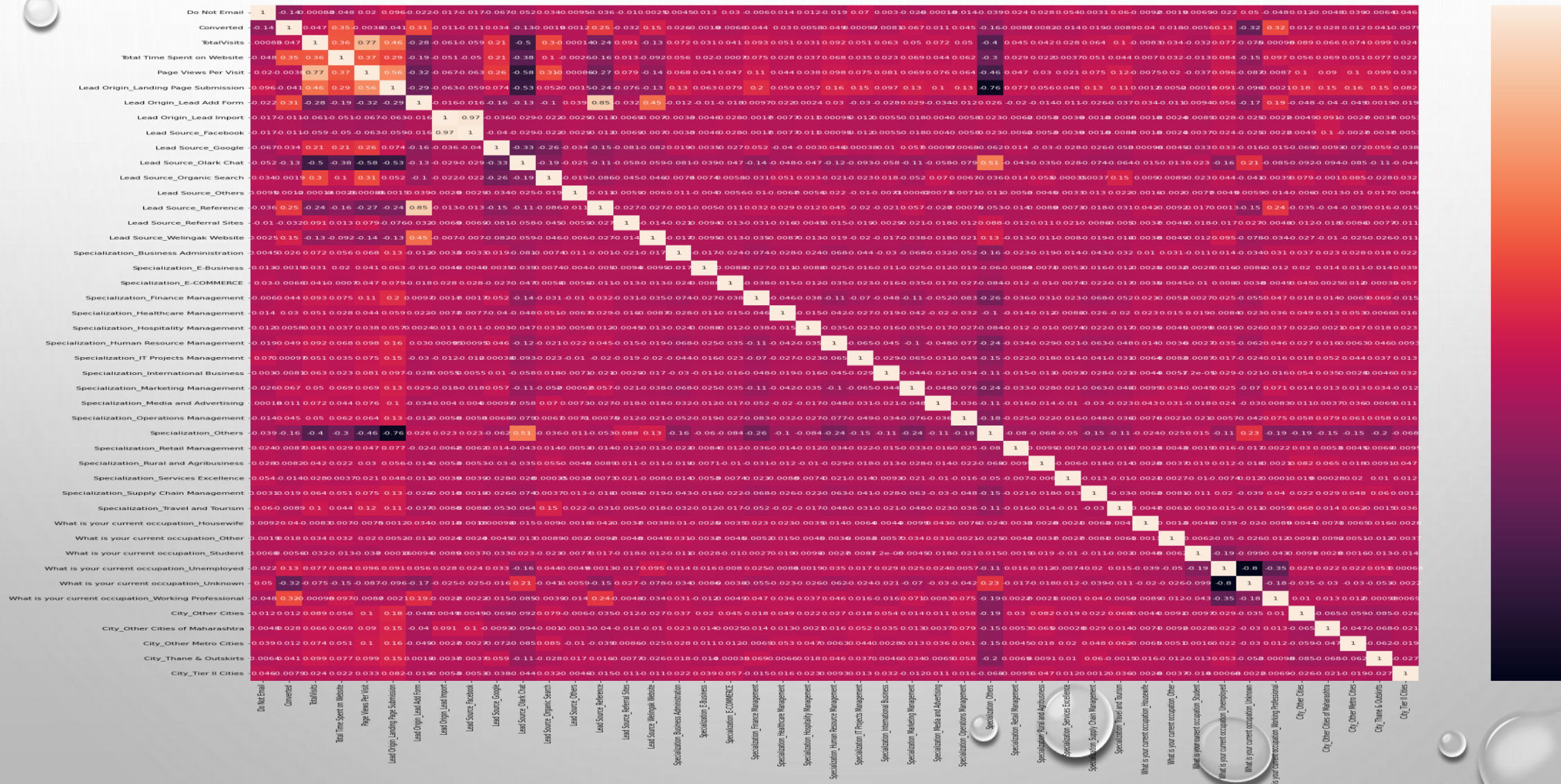
INSIGHTS

- MORE LEADS ARE GENERATED IN LANDING PAGE SUBMISSION AND API AND THEIR CONVERSION RATE IS HIGH WHEN COMPARED TO OTHERS
- IN LEAD ADD FORM CONVERSION RATE IS HIGH BUT LESS NUMBER OF LEADS ARE GENERATED FROM THERE.

CORRELATION BETWEEN NUMERICAL VARIABLES



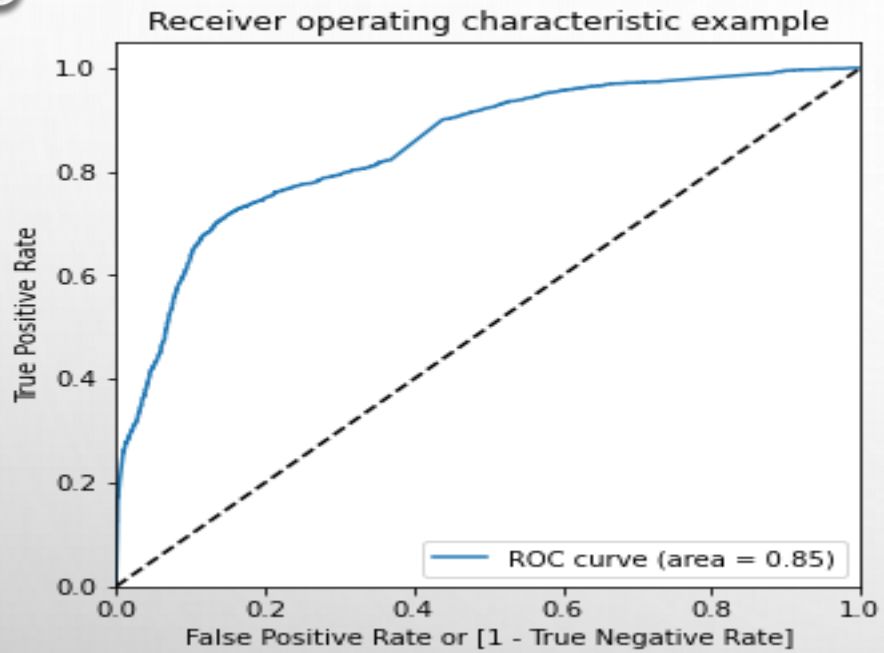
CORRELATION MATRIX



IMPORTANT VARIABLES SELECTED BY ALGORITHM

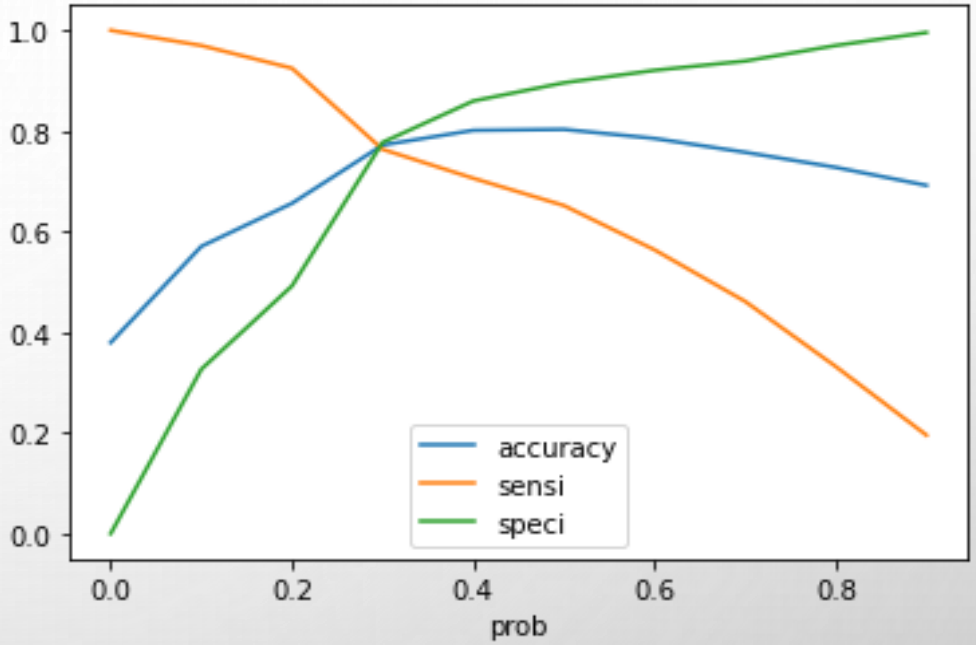
Features	VIF
Lead Origin_Landing Page Submission	1.98
What is your current occupation_Unemployed	1.91
Lead Origin_Lead Add Form	1.61
Lead Source_Welingak Website	1.32
What is your current occupation_Working Profes...	1.23
Specialization_Marketing Management	1.14
Total Time Spent on Website	1.10
What is your current occupation_Student	1.03
Do Not Email	1.02
Lead Source_Referral Sites	1.01
What is your current occupation_Other	1.01

ROC CURVE



ROC Curve value= 0.85 and it means the model is good.
As ROC value should be close to 1

OPTIMAL CUT-OFF POINT



From the curve above, 0.3 is the optimum point to take it as a cutoff probability

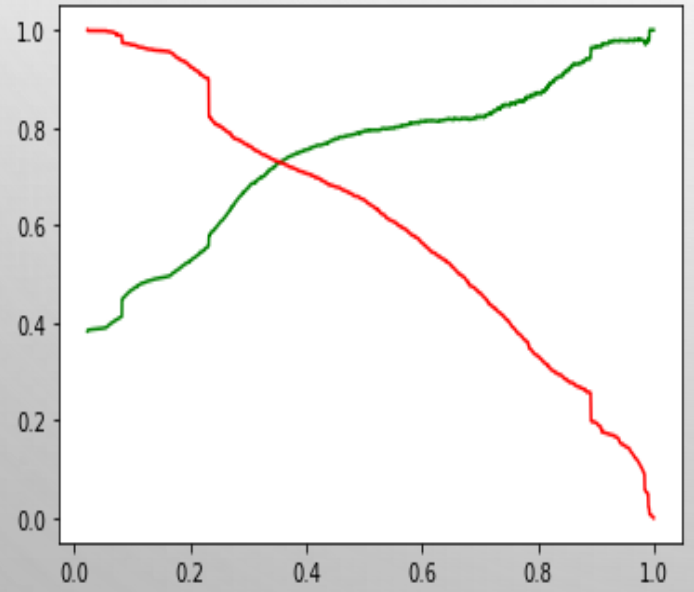
We have the following values for train data

- CONVERSION RATE ON TRAIN DATA IS 76%
- ACCURACY : 77%
- SENSITIVITY :77%
- SPECIFICITY :78%

We have the following values for test data

- CONVERSION TARGET RATE IS 75%
- ACCURACY -77%
- SENSITIVITY-75%
- SPECIFICITY-79%

Precision and recall trade-off



CONCLUSION

- FINAL CONVERSION RATE FOR THE MODEL IS 76% BEFORE BUILDING THE MODEL CONVERSION RATE WAS 38%
- METRICS VALUES FOR BOTH TRAIN AND TEST ARE ALMOST SAME.
- MOST IMPORTANT VARIABLES TO FOCUS ON TO CONVERT MORE LEADS ARE
- 1) WHAT IS YOUR CURRENT OCCUPATION_WORKING PROFESSIONAL.
- 2) LEAD ORIGIN_LEAD ADD FORM.¶
- 3) LEAD SOURCE_WELINGAK WEBSITE.