Lead Score Case Study

Shreyansha Tai Sonwane

Bhomik Sharma

Problem Statement

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to assign a lead score to each of the leads 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. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Goals and Objectives

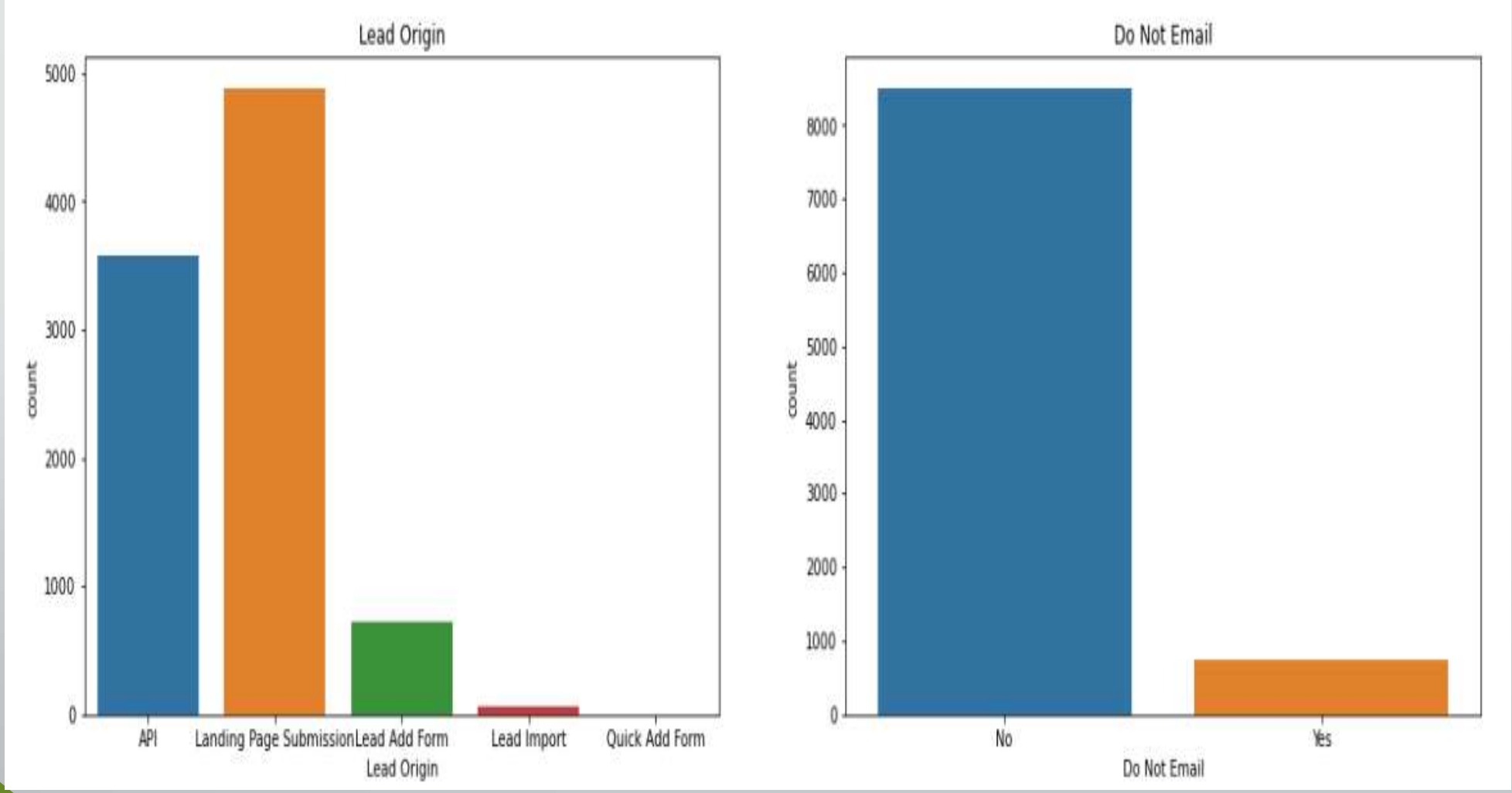
There are quite a few goals for this case study.

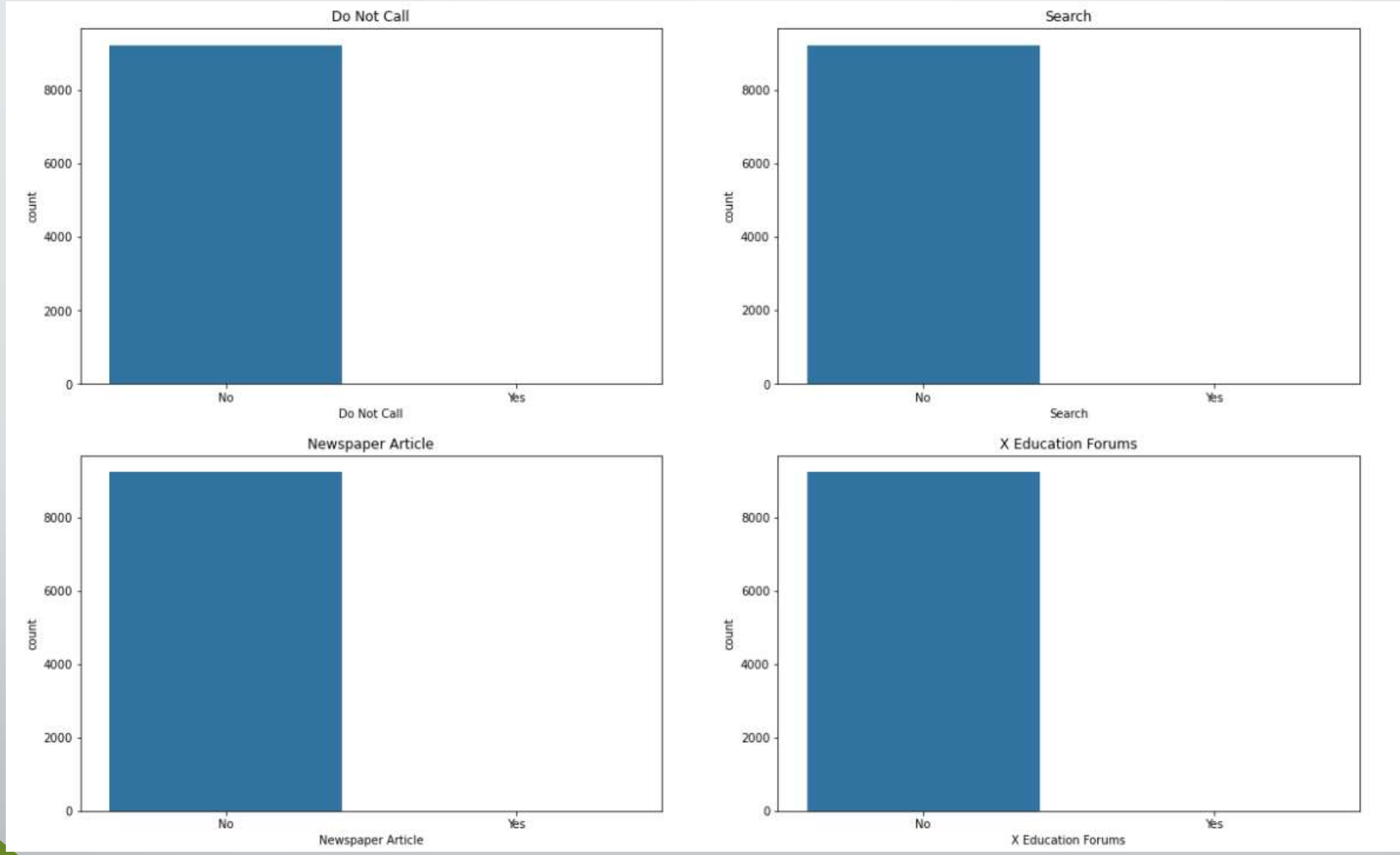
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
- There are some more problems presented by the company which the model should be able to adjust to if the company's requirement changes in the future so we will need to handle these as well.

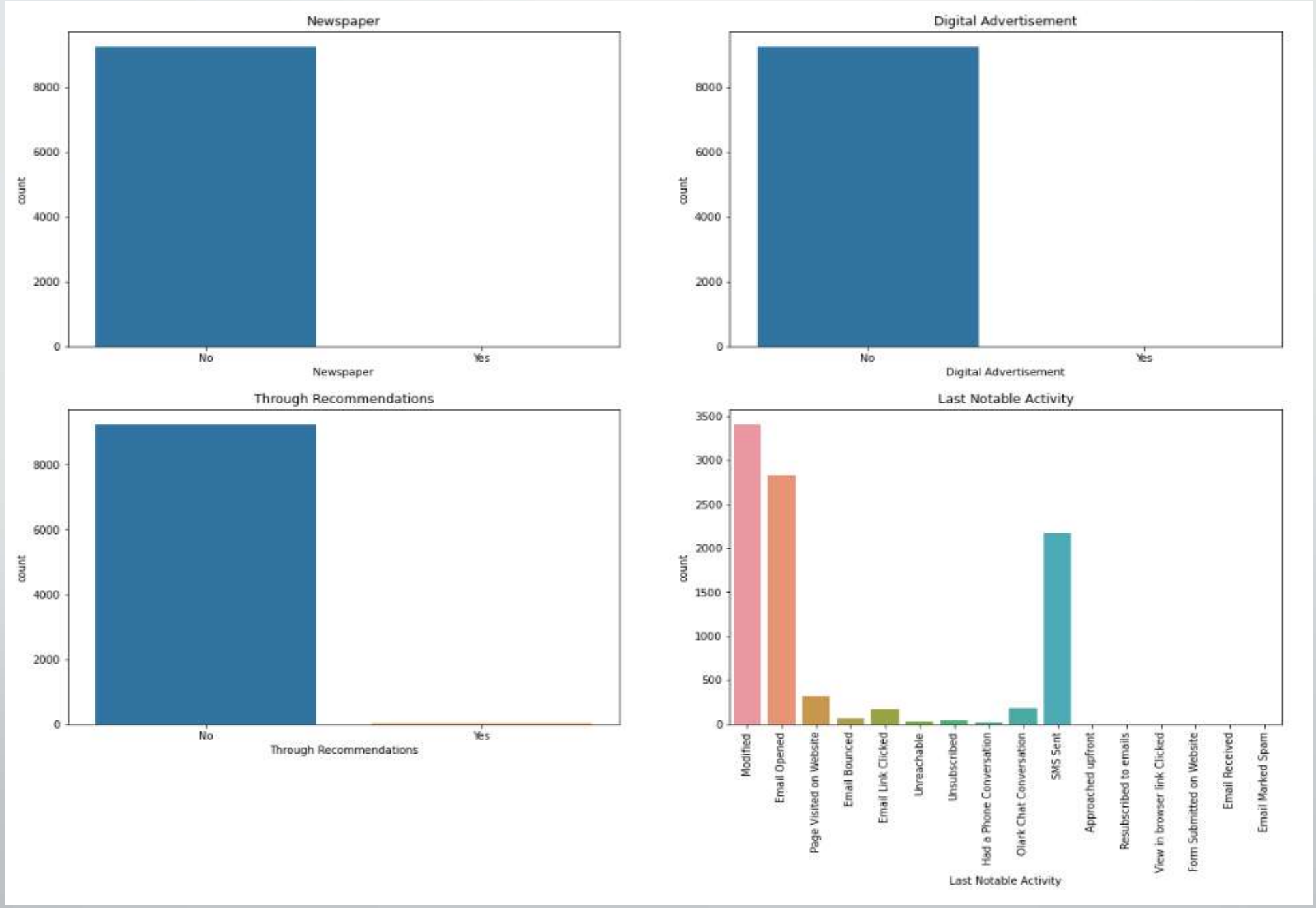
Solution Methodology

1) Data cleaning and data manipulation
2) EDA
3) Feature Scaling & Dummy Variables and encoding of the data
4) Classification technique: logistic regression is used for the model-making and prediction
5) Validation of the model
6) Model presentation
7) Conclusions and recommendations

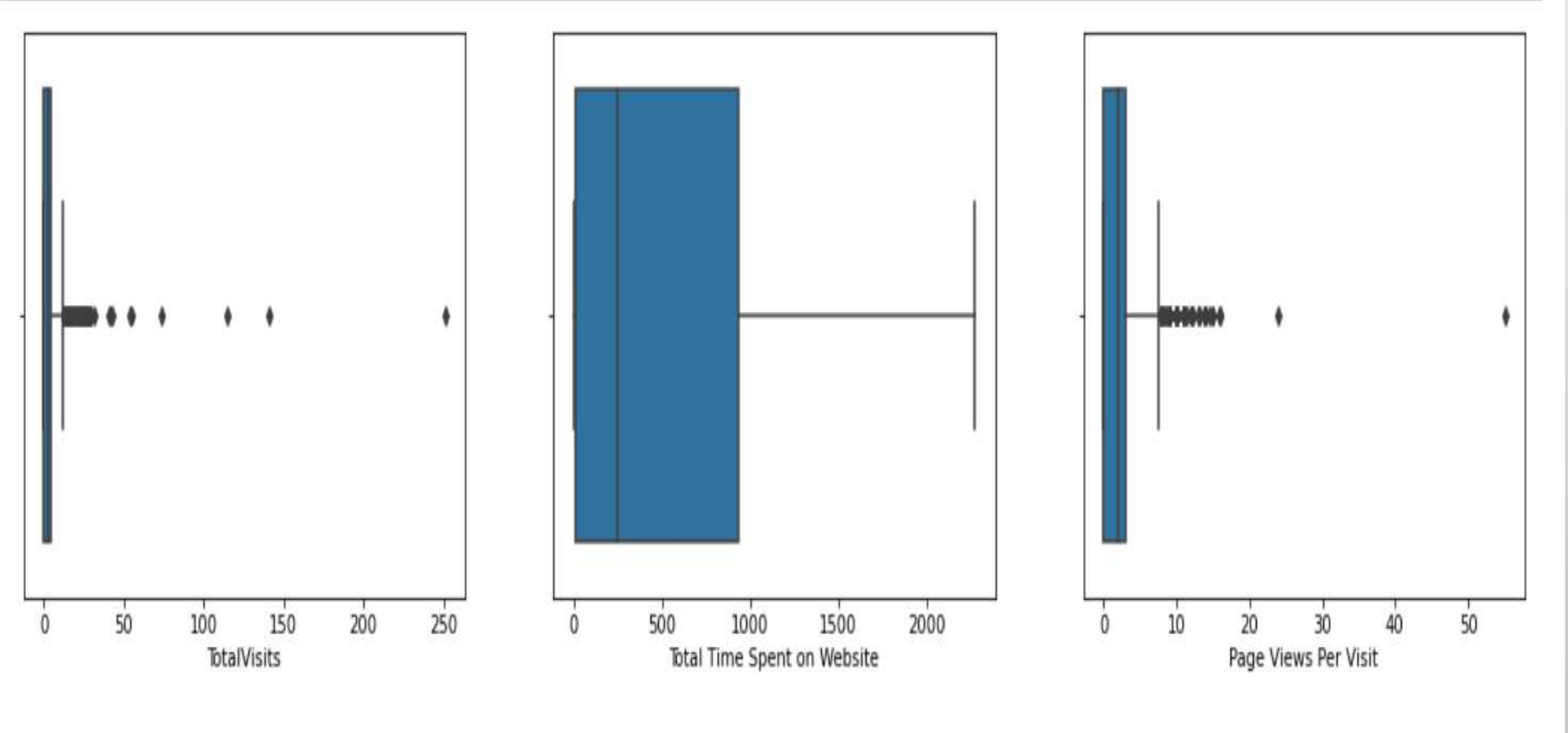
EDA


Lead Origin
count
5000
4000
3000
2000
1000
0
API
Landing Page Submission
Lead Add Form
Lead Import
Quick Add Form
Lead Origin
Do Not Email
count
8000
7000
6000
5000
4000
3000
2000
1000
0
No
Yes
Do Not Email

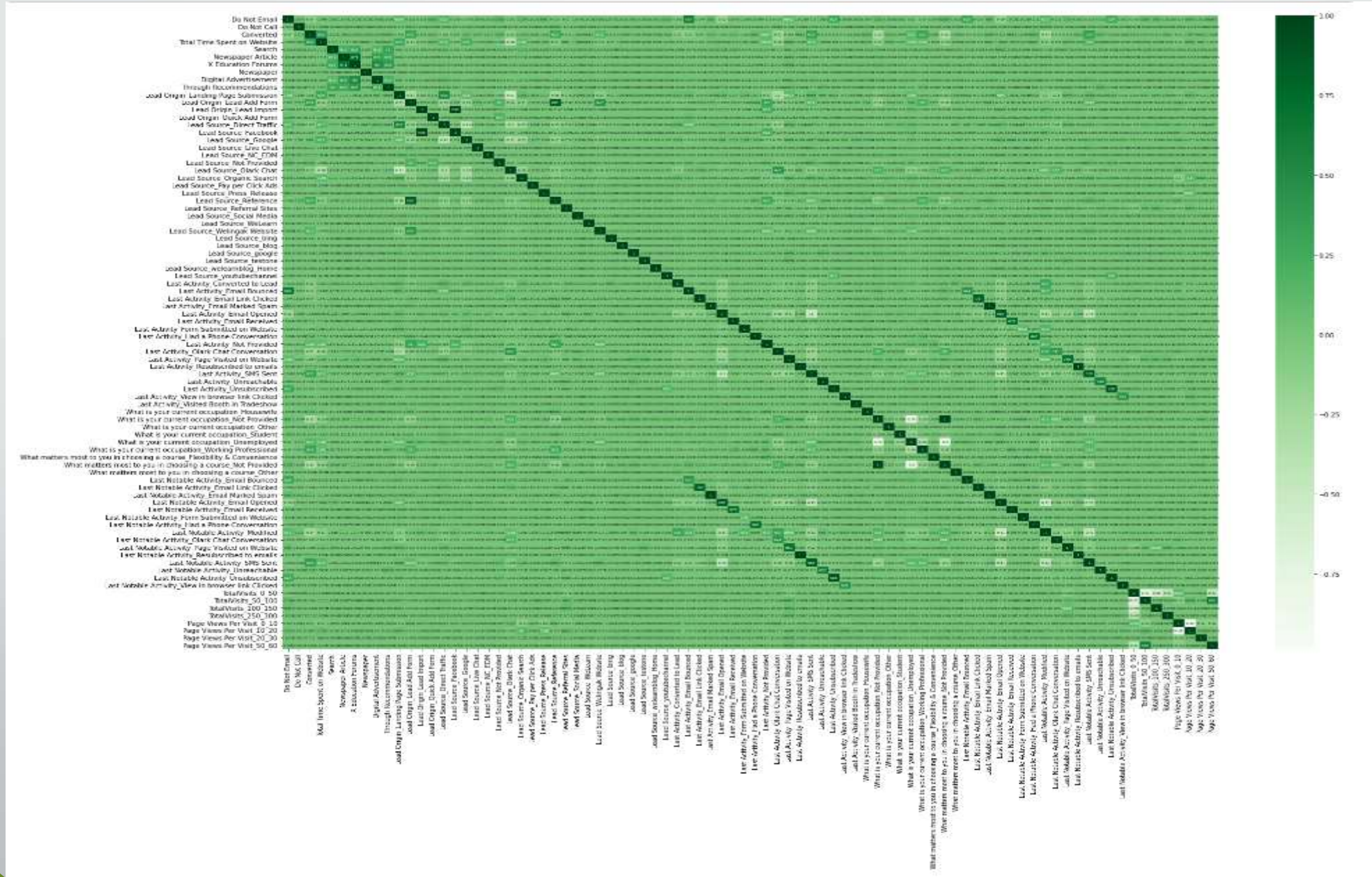

Do Not Call
count
8000
6000
4000
2000
0
No
Yes
Do Not Call
Search
count
8000
6000
4000
2000
0
No
Yes
Search
Newspaper Article
count
8000
6000
4000
2000
0
No
Yes
Newspaper Article
X Education Forums
count
8000
6000
4000
2000
0
No
Yes
X Education Forums


Newspaper
count
8000
6000
4000
2000
0
No
Yes
Newspaper
Digital Advertisement
count
8000
6000
4000
2000
0
No
Yes
Digital Advertisement
Through Recommendations
count
8000
6000
4000
2000
0
No
Yes
Through Recommendations
Last Notable Activity
count
3500
3000
2500
2000
1500
1000
500
0
Modified
Email Opened
Page Visited on Website
Email Bounced
Email Link Clicked
Unreachable
Unsubscribed
Had a Phone Conversation
Olark Chat Conversation
SMS Sent
Approached upfront
Resubscribed to emails
View in browser link Clicked
Form Submitted on Website
Email Received
Email Marked Spam
Last Notable Activity

Outliar Analysis



Correlation Matrix

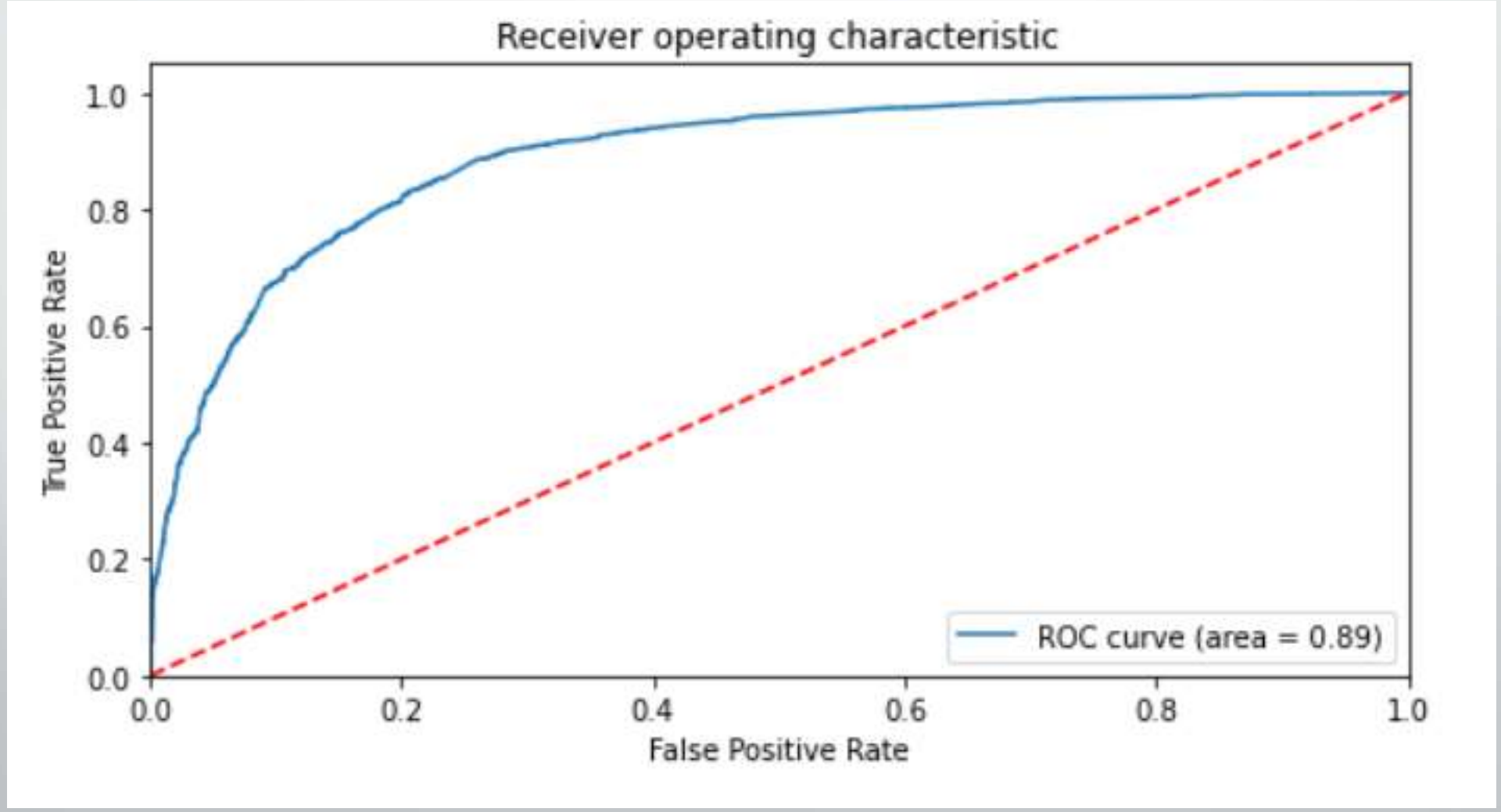


Model Building

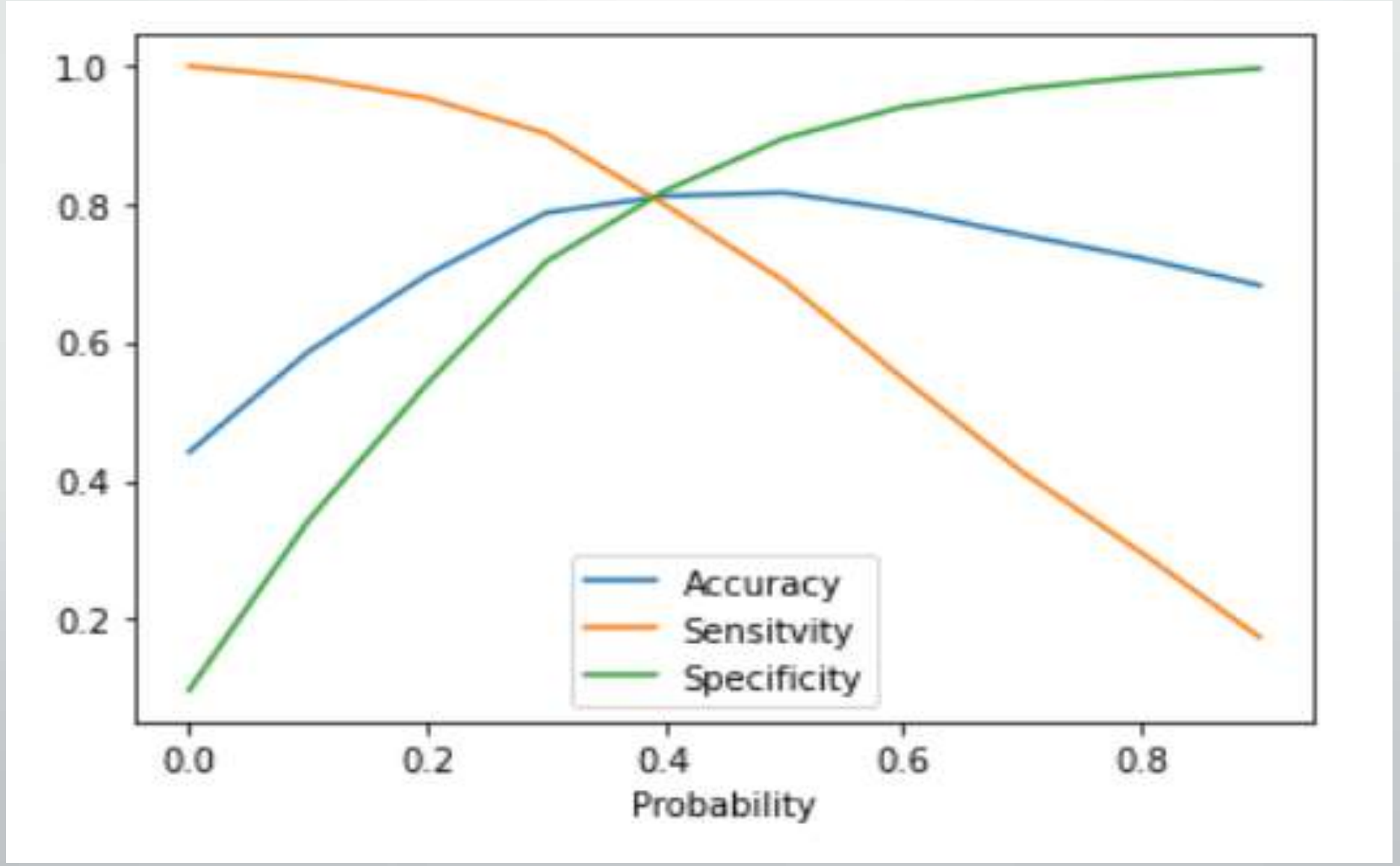
- We built a model with all the features included and found there were many insignificant variables present in our model.
- We need to drop them, but we can’t do it one by one as it is time consuming and not an efficient way to do so.
- Therefore, we started with RFE method to deduct those insignificant variables. We choose with RFE count 19 and 15.

- We did two rfe count because we want to find out our final model stability.
- We started creating our model with rfe count 19 and went dropping variables one by one until we reach the point where the model is having all significant variables and low VIF values.
- Now we evaluated our model by first predicting it. We created new dataset with original converted values and the prediction values.

Model Evaluation



Optimal Cutoff



Conclusion

- The Accuracy, Precision and Recall/Sensitivity are showing promising scores in test set which is as expected after looking the same in train set evaluation steps. Means the recall is having high score value than precision which is acceptable for business needs.
- In business terms, this model has an ability to adjust with the company’s requirements in coming future. This concludes that the model is in stable state.
- Important variables from the model:

    1) Last Notable Activity_Had a Phone Conversation

    2) Lead Origin_Lead Add Form

    3) What is your current occupation_WorkingProfessional

THANK YOU