# LEAD SCORING CASE STUDY

LOGISTIC REGRESSION

# PROBLEM STATEMENT

X education is an organization which provides online courses for industry professional .The company marks its courses on several popular websites like google.

X education wants to select most promising leads that can be converted  to paying customers.

Although the company generates a lot of leads only a few are converted into paying customers , wherein the company wants a higher lead  conversion. Lead comes through numerous mode like email , advertisements on websites , google searches etc.

The company has had 30% conversion rate through the whole process of turning leads into customers by approaching those leads which are to be found having interests in taking the course. The implementation process of lead generating attributes are not efficient in helping conversions.

# BUSINESS GOAL

The company requires a model to be built for selecting most promising lists.

Lead score to be given to each leads such that it indicates how promising the lead could be. The higher the lead score the most promising the lead to get converted , the lower it is the lesser the chances of conversion.

The model to be built in lead conversion rate around 80% or more.

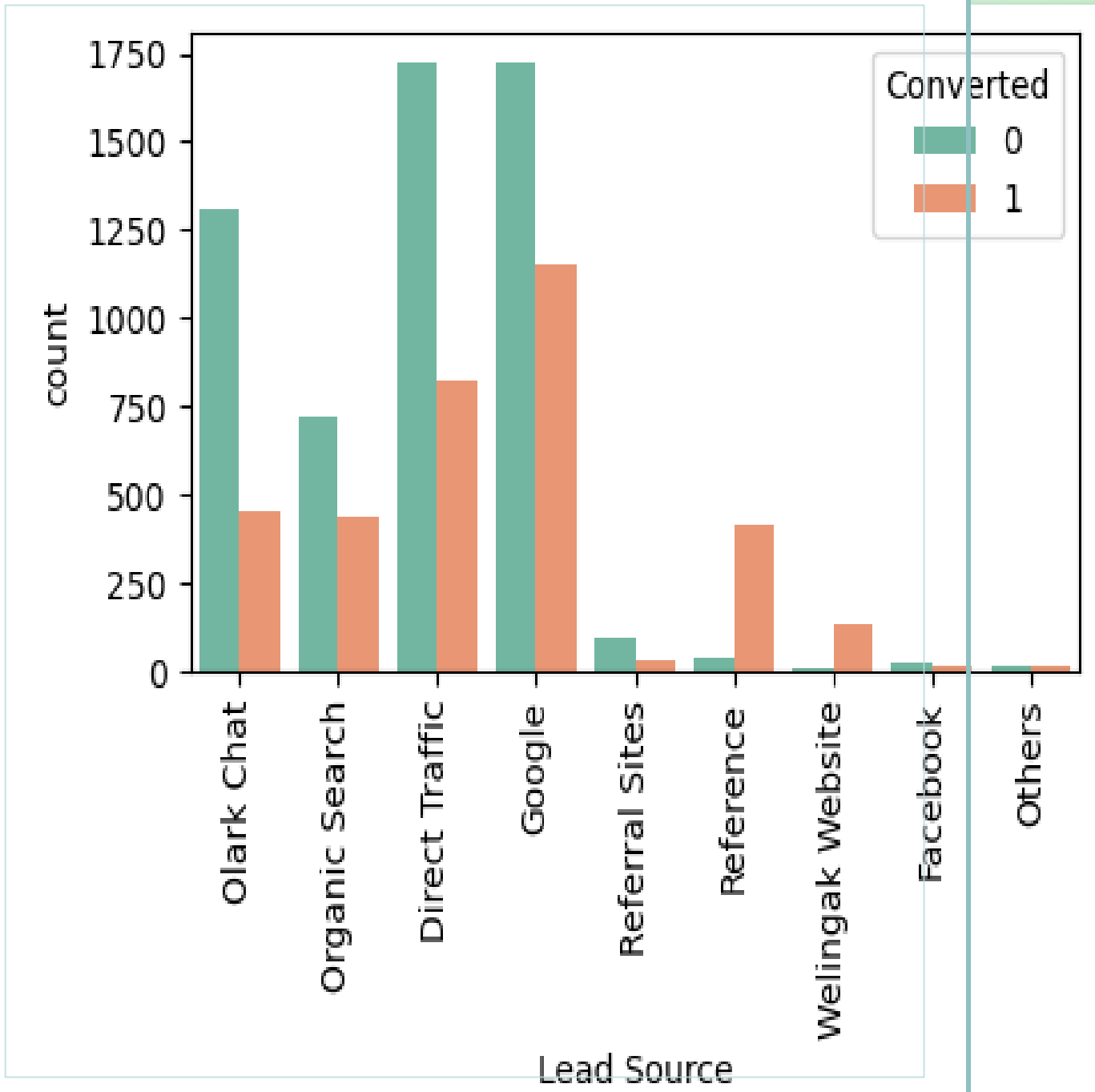
# STRATEGY

- Import data.
- Clean and prepare the acquired data for further analysis.
- EDA data analysis for figuring out most helpful attributes for conversion.
- Scaling features.
- Prepare the data for model building.
- Build a logistic regression model.
- Assign a lead score for each leads.
- Test the model on train set
- Evaluate model by different measures and metrices.
- Test the model on test set.
- Measure the accuracy of the model and other metrices for evaluation.

# EXPLORATORY DATA ANALYSIS

- **LEAD SOURCE VS CONVERTED**

- Google searches has had high conversion compared to other modes, whitelist has had high conversion rate.

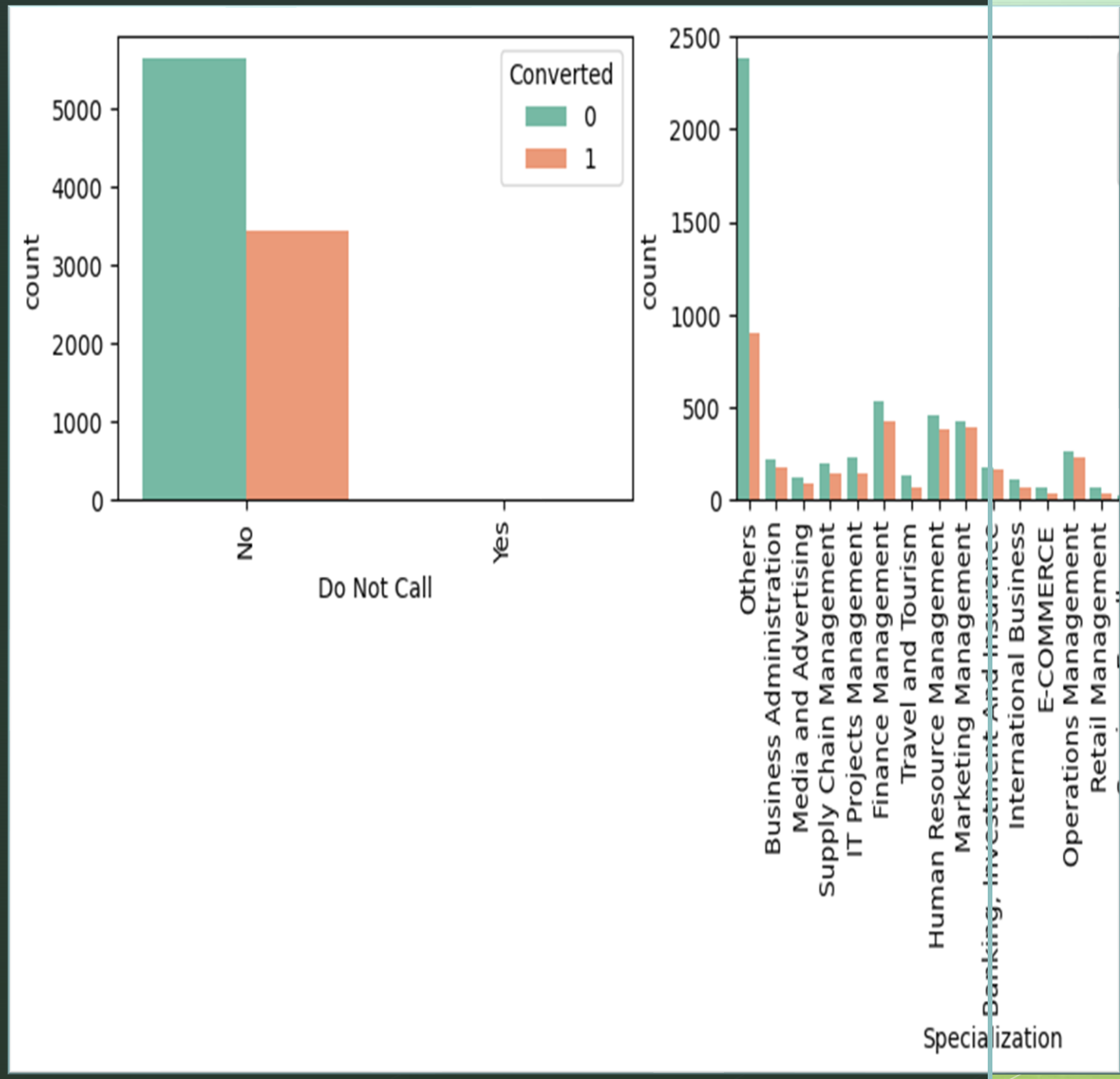


- **DO NOT CALL VS CONVERTED**

- Most entries are no so nothing can do with this parameter.

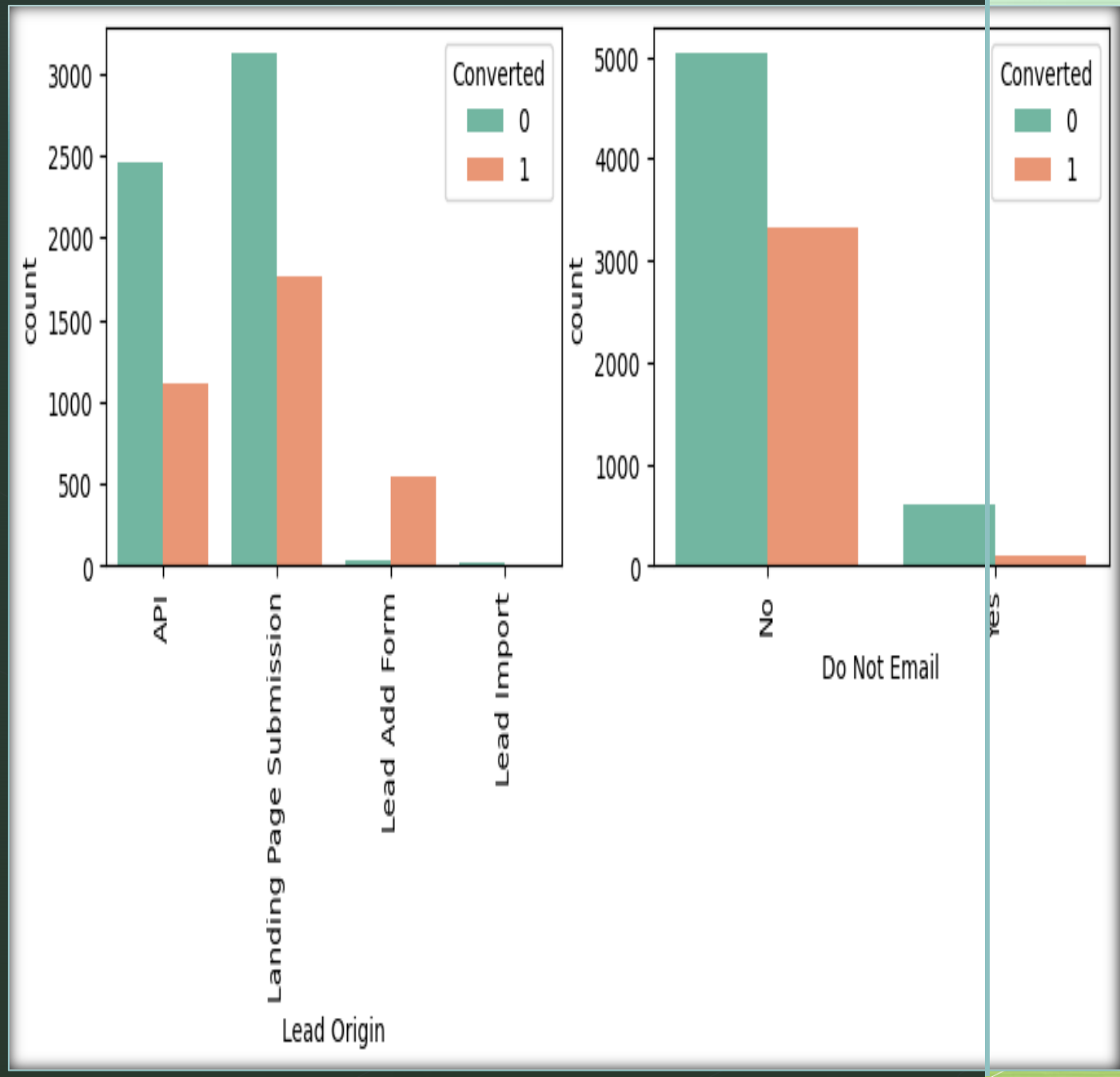
- **SPECIALISATION**

- Other categories has high conversion rate so focus would be more in this specialization.



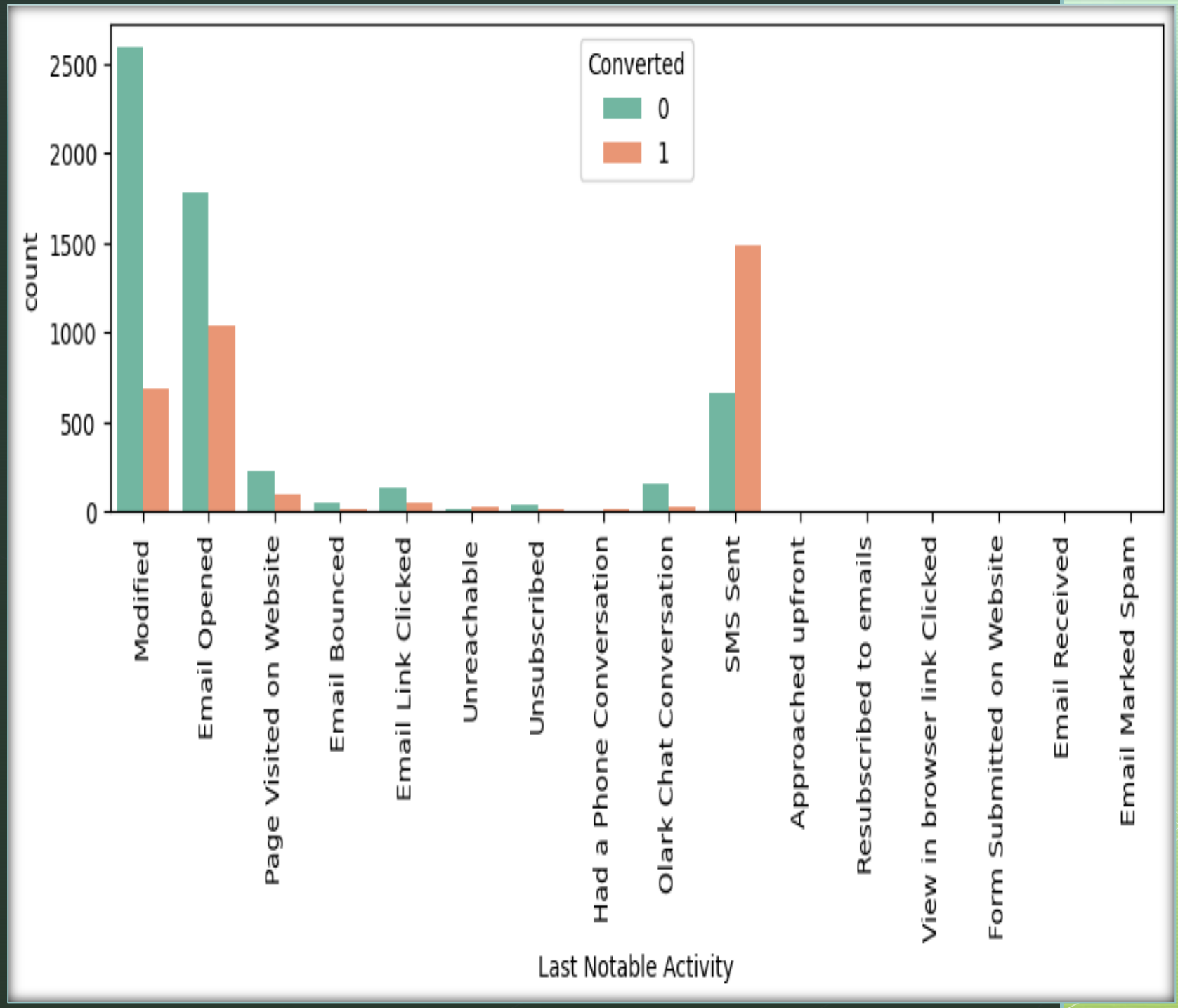
- **LEAD ORIGIN VS CONVERTED**
- Lead page submission have 30-35% conversion rate and the count of leads originated from them are quite high.

- **DO NOT EMAIL VS CONVERTED**
- Most entries are no ...



- **LAST NOTABLE ACTIVITY VS CONVERTED**

- Most leads are converted with messages. Email also induce leads.

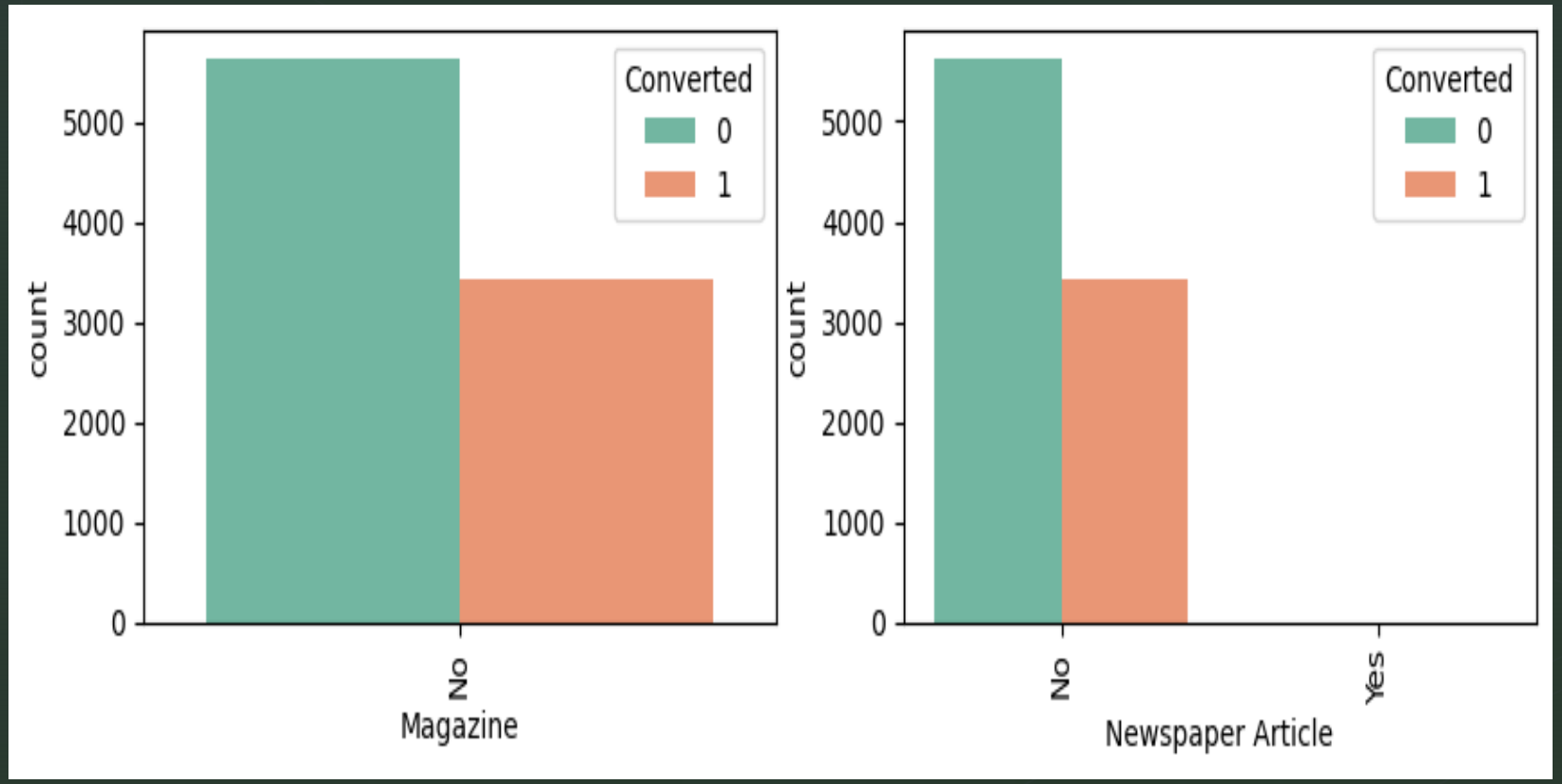


## MAGAZINE VS CONVERTED

Magazine do not have high conversion rate.

## NEWSPAPER ARTICLE VS CONVERTED

Newspaper do not have high conversion rate



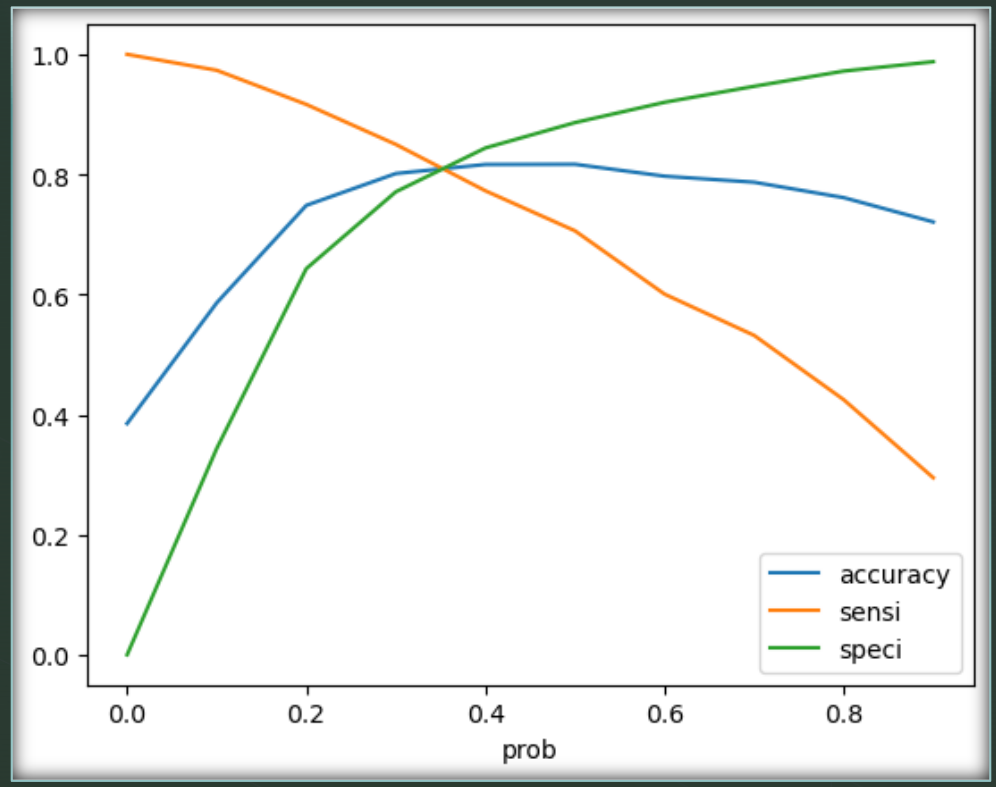
# MODEL BUILDING

- Splitting into train and test data .
- Scale variables in train set.
- Build the first model.
- Use RFE to eliminate less relevant variables.
- Build the next model.
- Eliminates variable based on high p-values.
- Check VIF values for all the existing columns.
- Predict using train set.
- Evaluate accuracy and other metric.
- Predict using test set.
- Precision and recall analysis on train predictions.

# MODEL EVALUATION

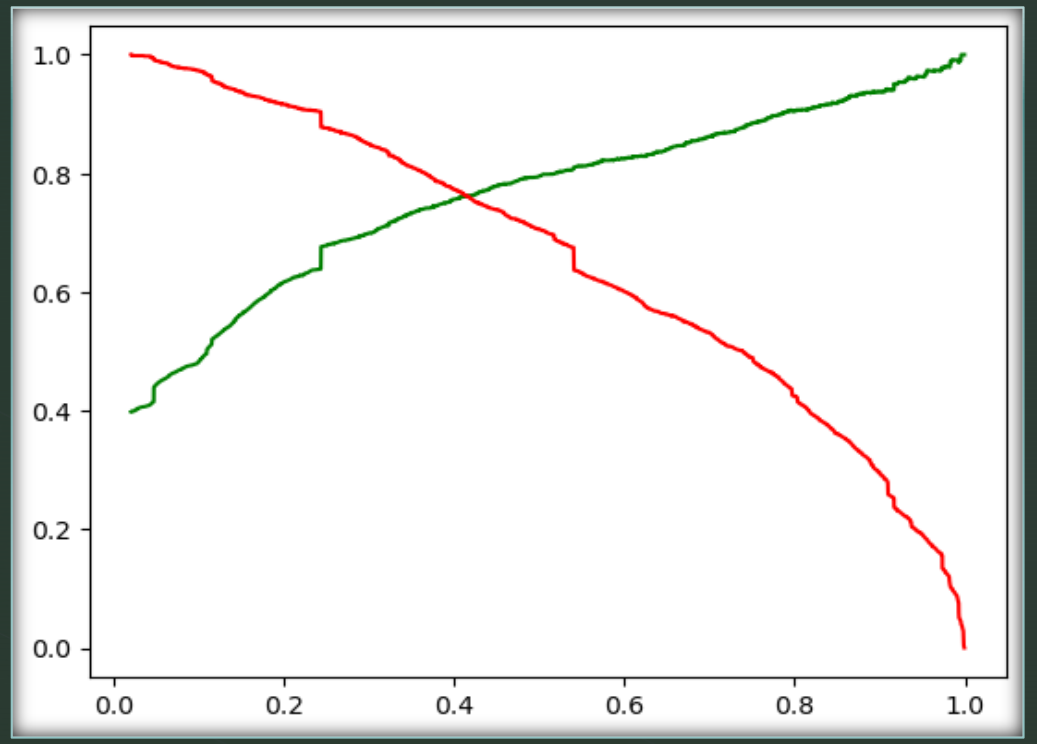
- **ACCURACY, SENSITIVITY AND SPECIFICITY**

- Accuracy- 81.0%
- Sensitivity-81.7%
- Specificity-80.6%



- **PRECISION AND RECALL**

- Precision- 79.5%
- Recall- 70.6%



# CONCLUSION

1. Our model is fit as area under curve is higher 0.89

2. Train Data: Accuracy : 81.0 %  Sensitivity : 81.7 %  Specificity : 80.6 %

   Test Data: Accuracy : 80.4 %  Sensitivity : 80.49 %  Specificity : 80.5

   We have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . The Model seems to predict the Conversion Rate according to the requirement. Business can be confident on using this model.

3.The top 3 models that contribute for lead getting converted in the model are:

a) Lead source in the data(Welingak websites and references)

b) Last activity

c) More time on the websites

4. Hence this overall model seems to be good.

# RECOMMENDATIONS

1. The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.

2. The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.

3. The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.

4. The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.

5. The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.

6. The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.

7. The company should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.

8. The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.

9. The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.

10. The company should not make calls to the leads whos Last Notable Activity is 'Modified' as they are not likely to get converted.