

Lead Score Case study

-Ruta Moharil

-Rutuja Fulari

Problem Statement

- An education company named X Education sells online courses to industry professionals. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos.
- When these people fill up a form providing their email address or phone number, they are classified to be a lead. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

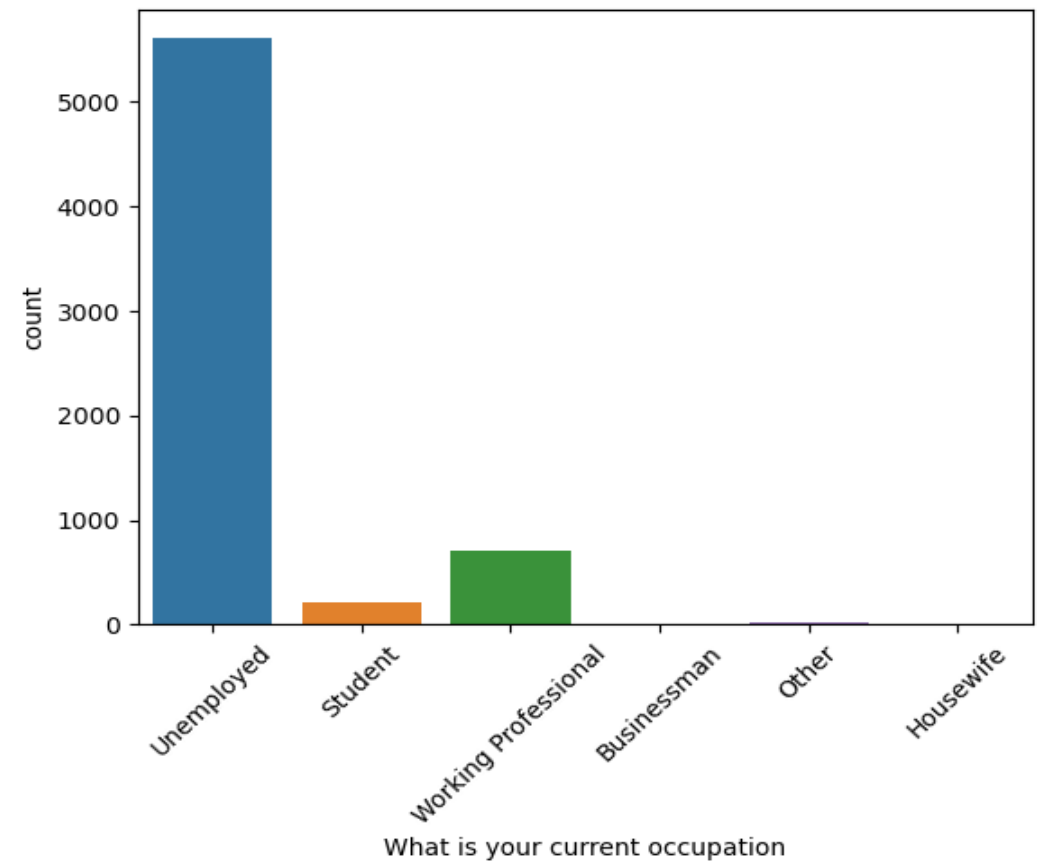
Objective

- The company needs a model wherein we need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a 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%.

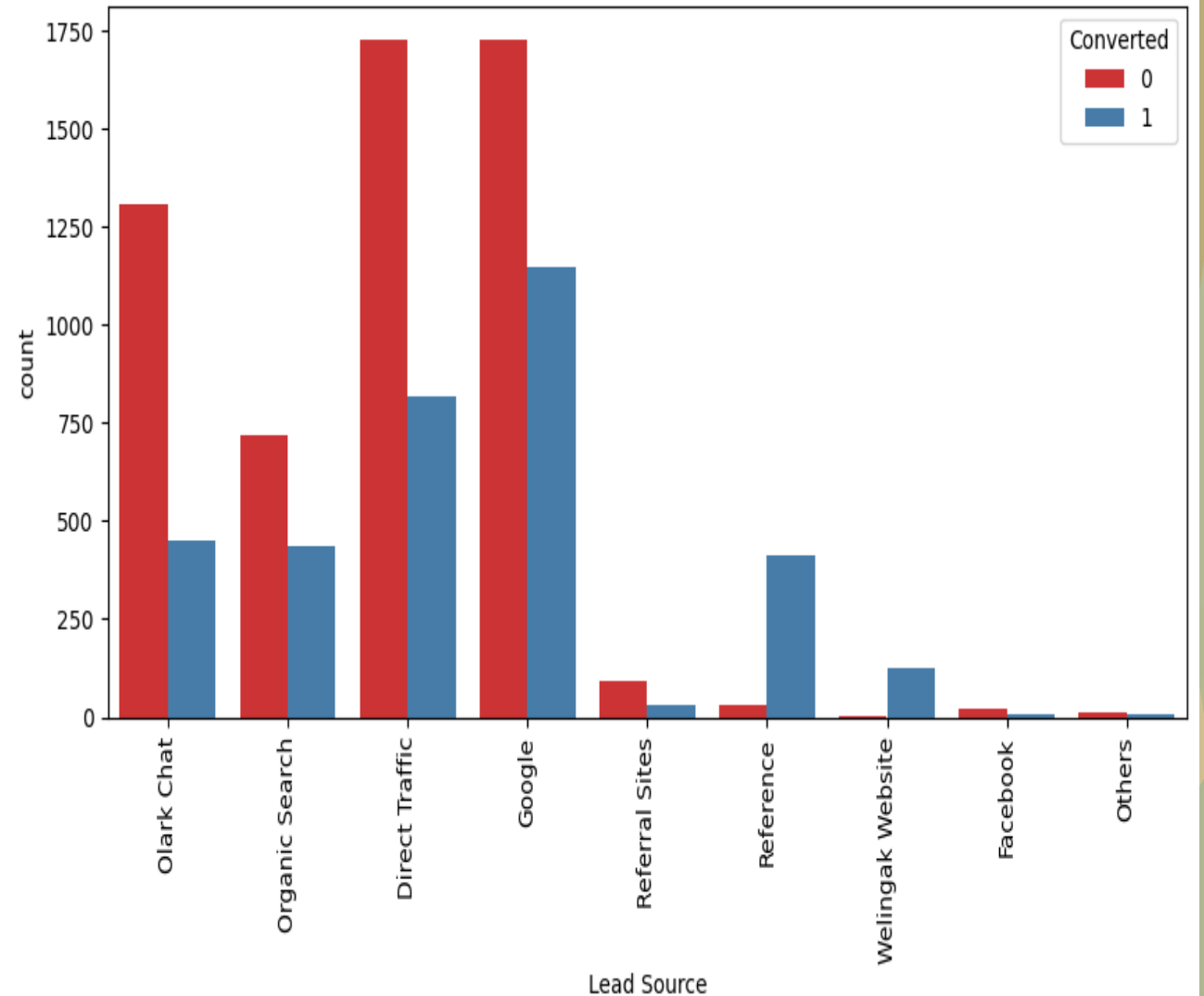
STEPS PERFORMED

- ▶ 1. Data Handling
- ▶ 2. Data cleaning, dropping columns with more 35 percent null values
- ▶ 3. Imputing the missing values
- ▶ 4. Exploratory Data Analysis
- ▶ 5. Model Building
- ▶ 6. Model Evaluation
- ▶ 7. Conclusion

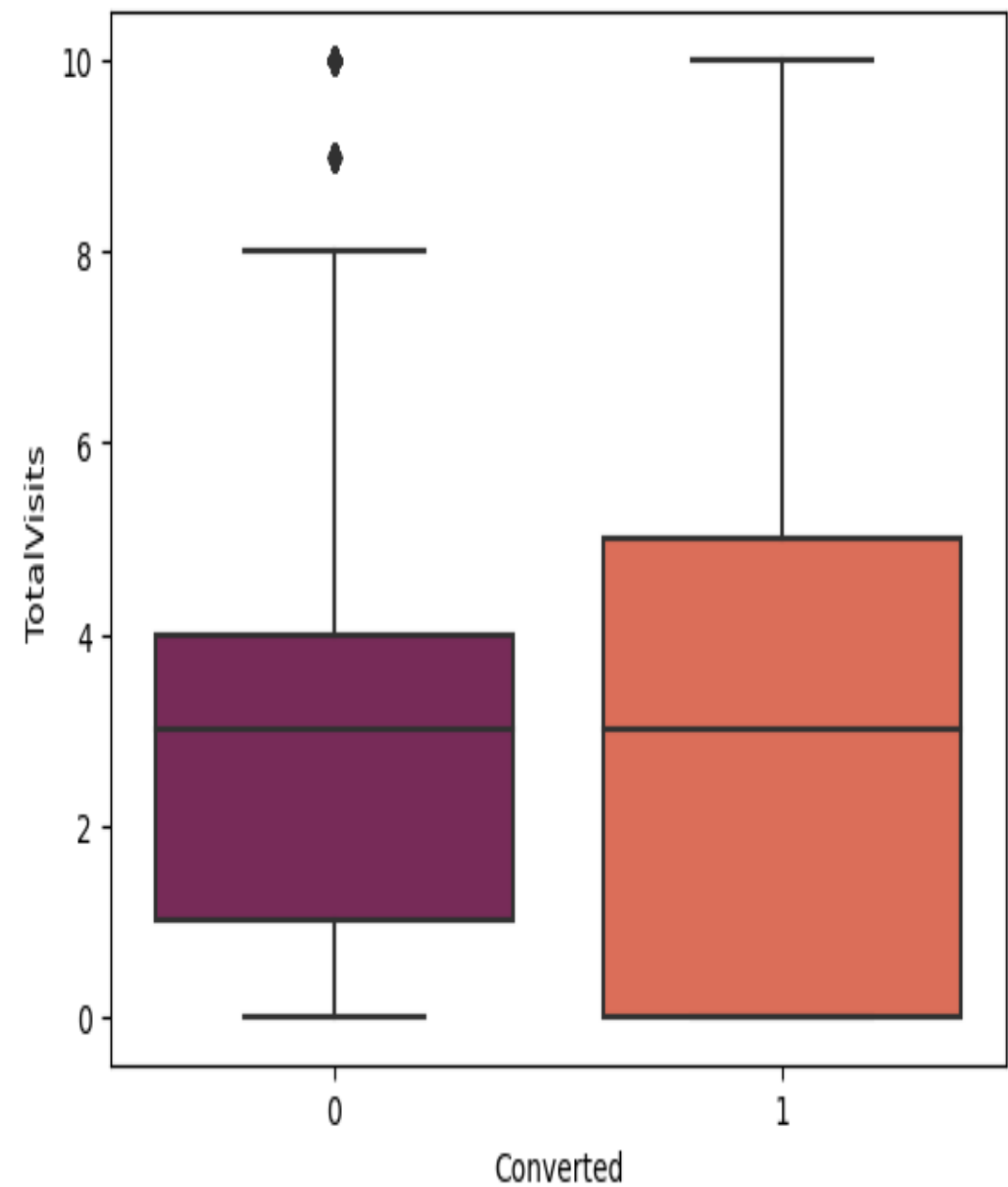
As observed from the graph maximum number of leads are unemployed or working professionals



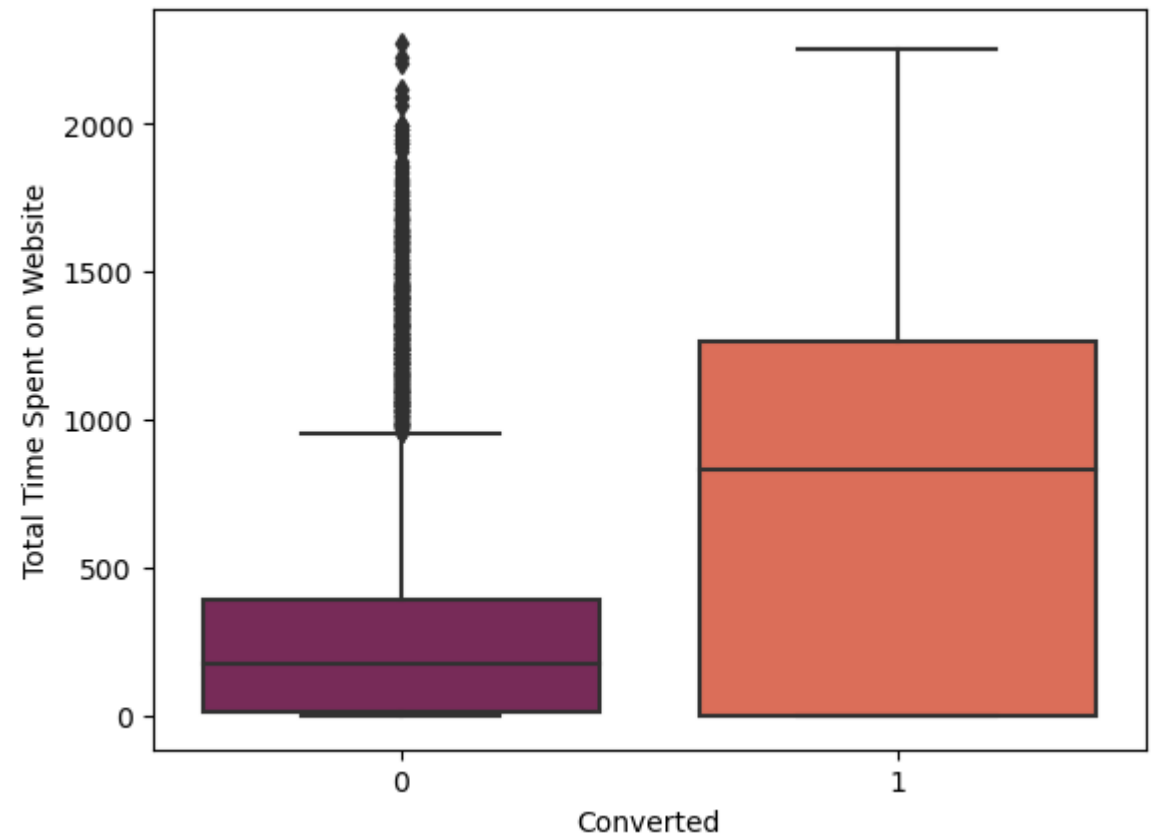
- **Google and Direct traffic** generates maximum number of leads.
- **Conversion Rate** of reference leads and leads through **welingak website** is high.
- To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingak website.



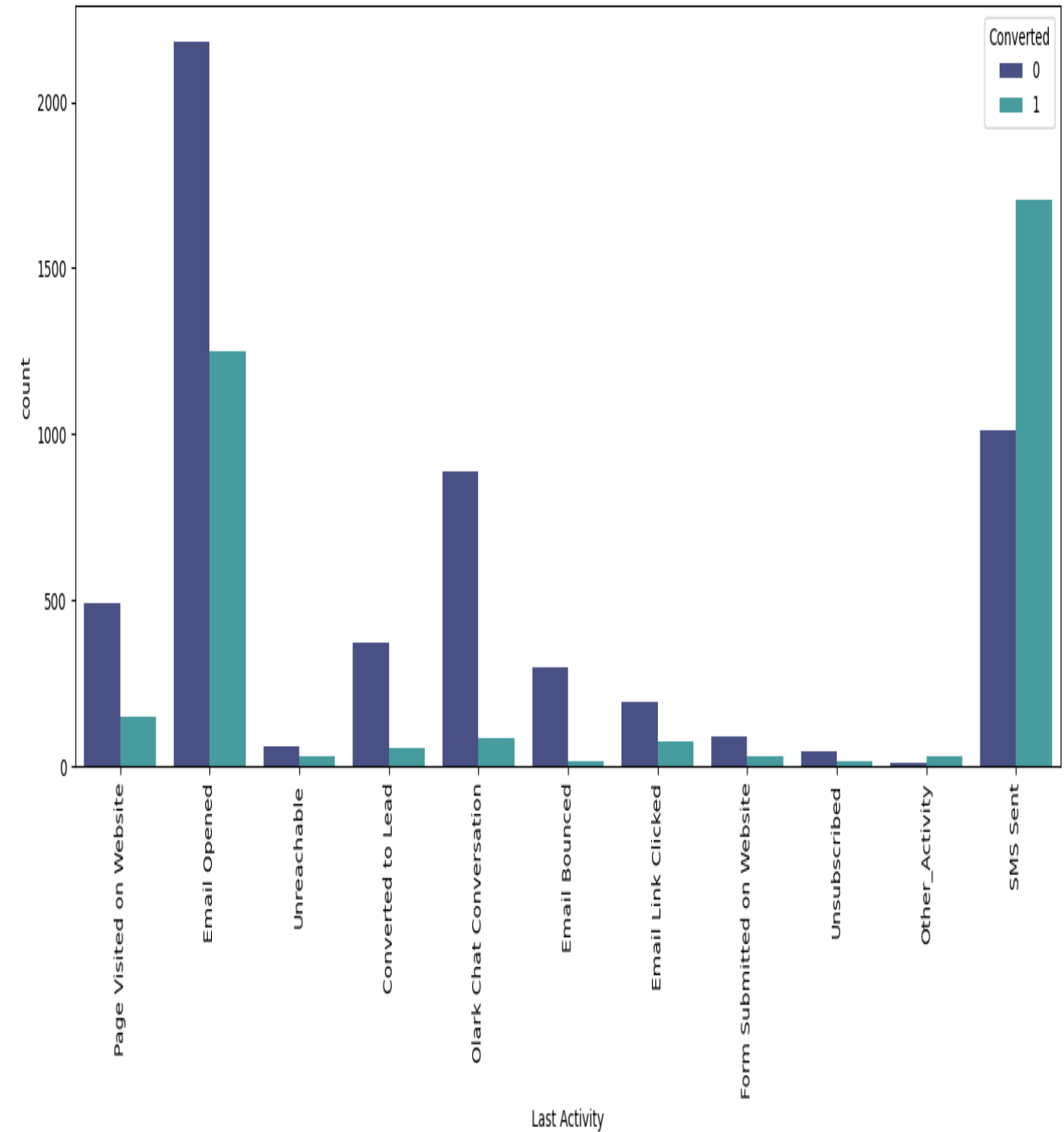
Median for converted and not converted leads are the same.
Nothing can be concluded on the basis of **Total Visits**.



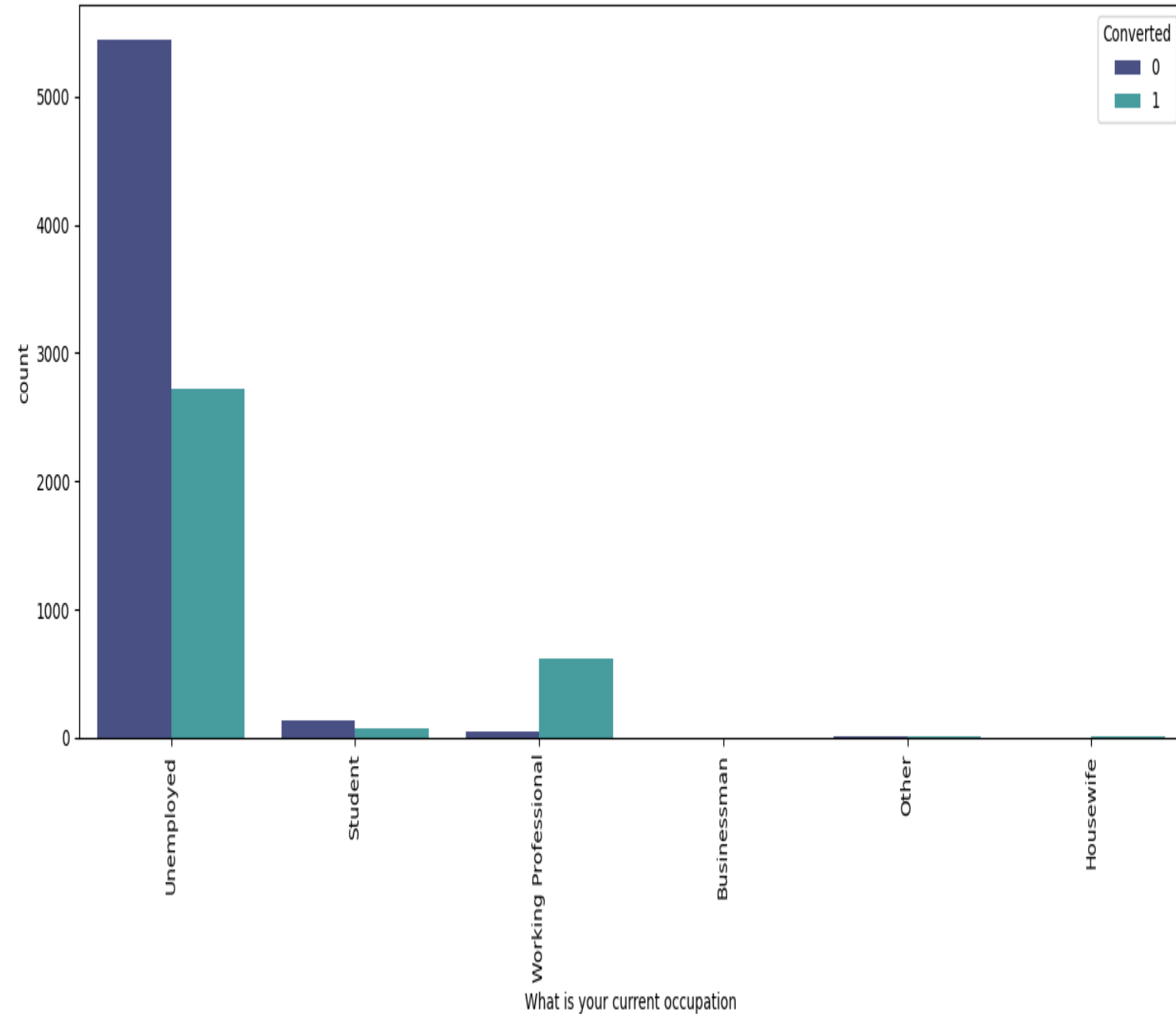
- Leads spending more time on the website are more likely to be converted.



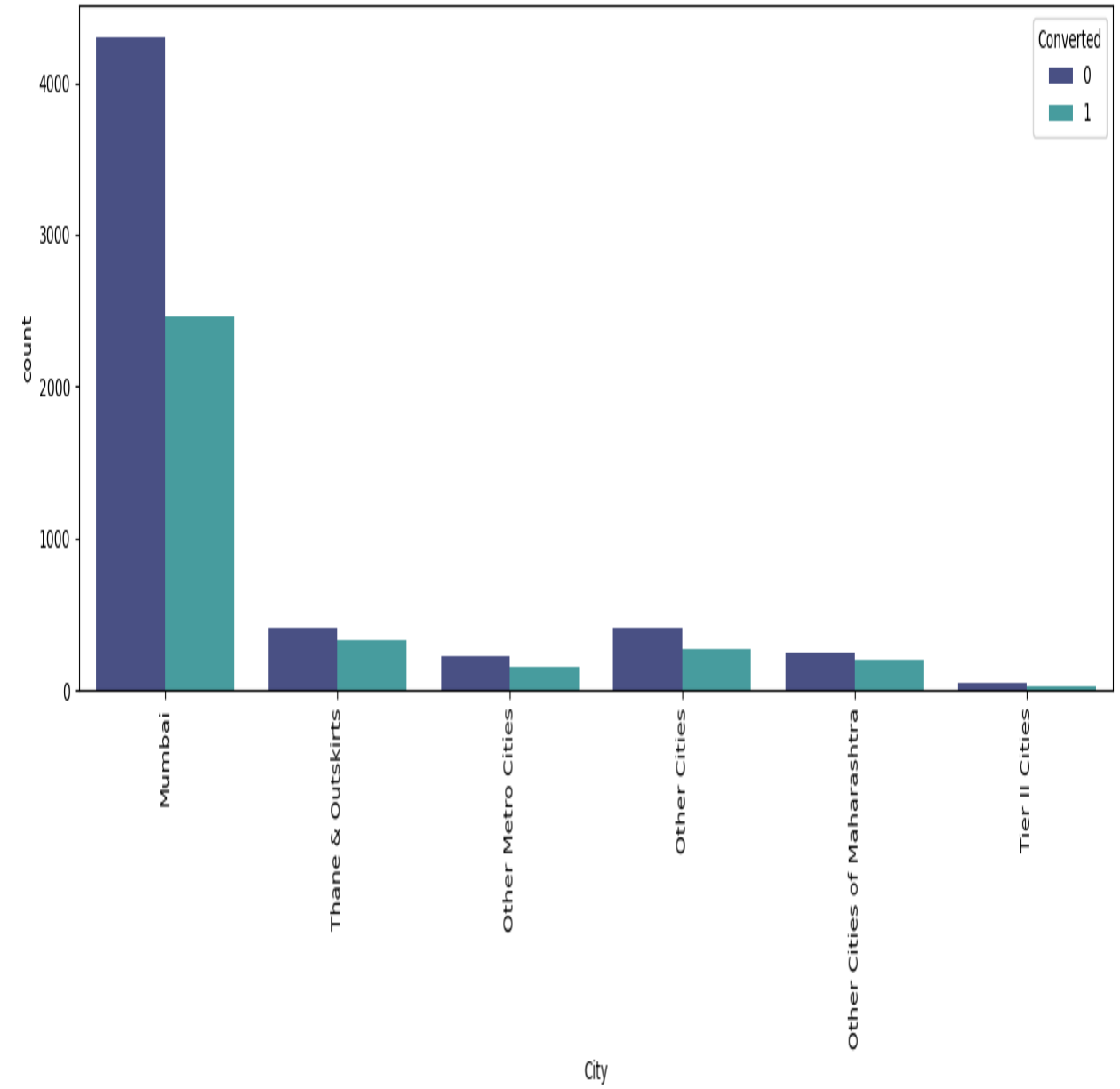
- ▶ Most of the lead have their Email opened as their last activity.
- ▶ Conversion rate for leads with last activity as SMS Sent is higher.



- Working Professionals going for the course have high chances of joining it.
- Unemployed leads are the most in numbers but has around lesser conversion rate.



- Most leads are from Mumbai with higher conversion rate.

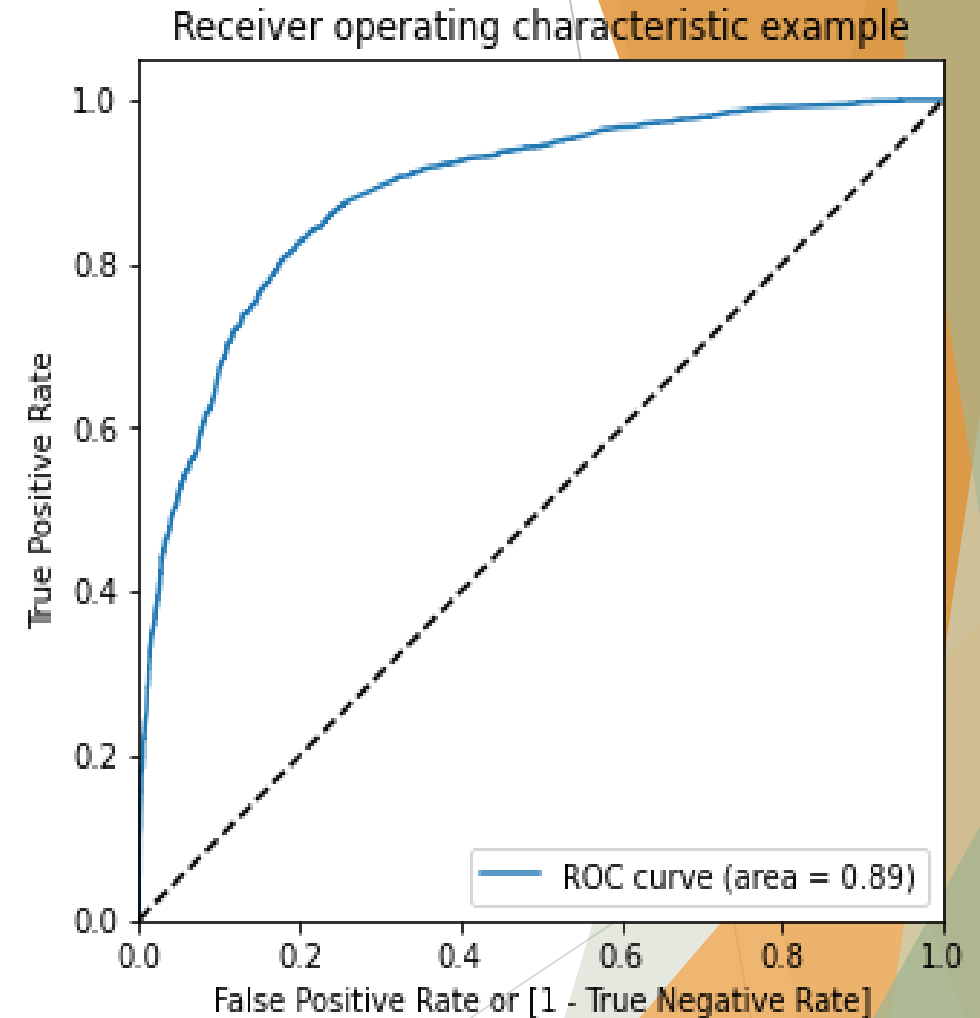


Model Building

- ▶ 1. We have splitted the data into train and test sets
- ▶ 2. We have chosen the train test split ratio as 70:30.
- ▶ 3. We have build the model by removing the variables by using RFE whose VIF greater than 5 and p-value greater than 0.05.
- ▶ 4. Running RFE with 12 variable as final model.

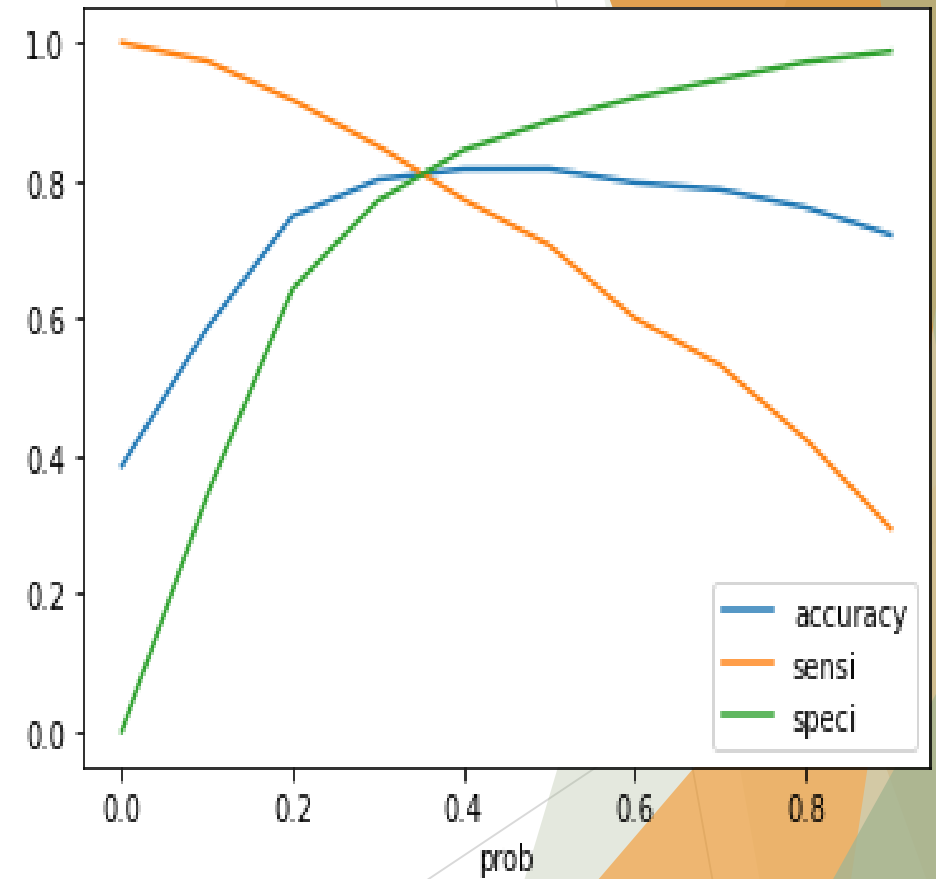
ROC Curve

- It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
- We have **higher (0.89) area under the ROC curve** , therefore our **model is a good one**.



Optimal Cutoff Point

- Optimal cutoff probability is that prob where we get balanced sensitivity and specificity.
- From the curve above, 0.34 is the optimum point to take it as a cutoff probability



Model Evaluation

- ▶ **Train Data:**

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

- ▶ **Test Data:**

- ▶ Accuracy : 80.4 %
- ▶ Sensitivity : 80.4 %
- ▶ Specificity : 80.5 %
- ▶ The precision-recall method was used to re check and cut off of 0.41 was found with precision around 79% and recall around 70% on test data frame.

Conclusion

- 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 very well and we should be able to give the CEO confidence in making good calls based on this model to get a higher lead conversion rate of 80%.
- The company should make calls to
 - leads coming from the lead sources "Welingak Websites" and "Reference"
 - the leads who are the "working professionals"
 - the leads who spent "more time on the websites"
 - the leads coming from the lead sources "Olark Chat"
 - the leads whose last activity was SMS Sent