PRESENTATION ON LEAD SCORE CASE STUDY

- Submitted By:
- Deeksha
- Dhiraj
- Rinu

BUSINESS PROBLEM

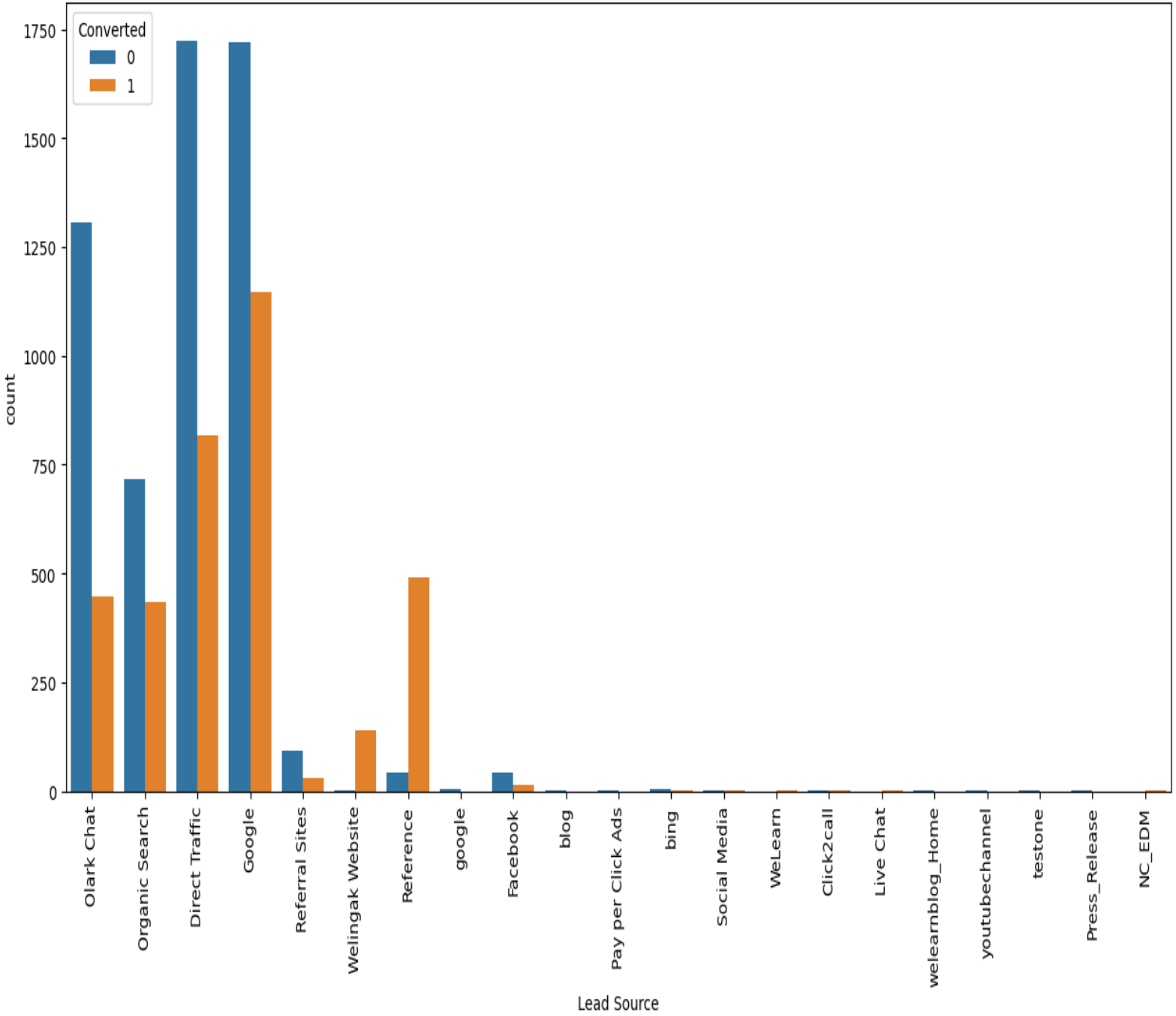
- X education is an organization which provides online courses for industry professional. The company markets its courses on several websites and search engines like Google.
- X education wants to select most promising leads that can be converted into paying customers.
- Although the company generates a lot of leads only a few are converted into paying customers, wherein the company wants a higher leads conversion.

BUSINESS OBJECTIVE

- The company needs a model to be built for selecting most promising leads.
- The company requires to build 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(hot lead) and the customers with a lower lead score have a lower conversion chance(cold lead).

STEPS:

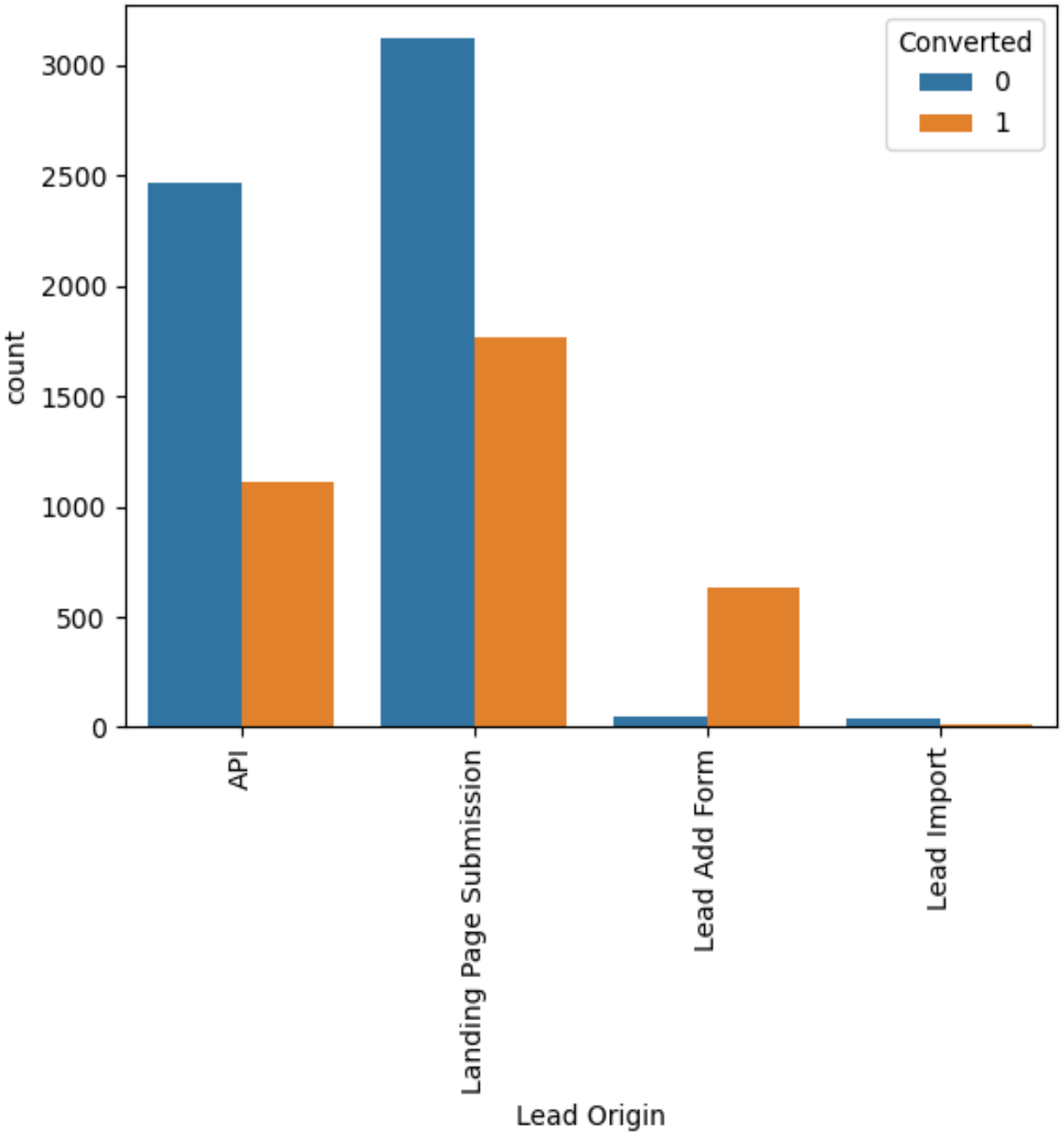
- Import Data
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis
- Scaling features
- Prepare the data for model building
- Build a Logistic Regression Model
- Assign a lead score foe each leads
- Test the model on Train set
- Evaluate the model
- Test the model on Test set
- Measure the accuracy of the model and other metrics



Exploratory data analysis:

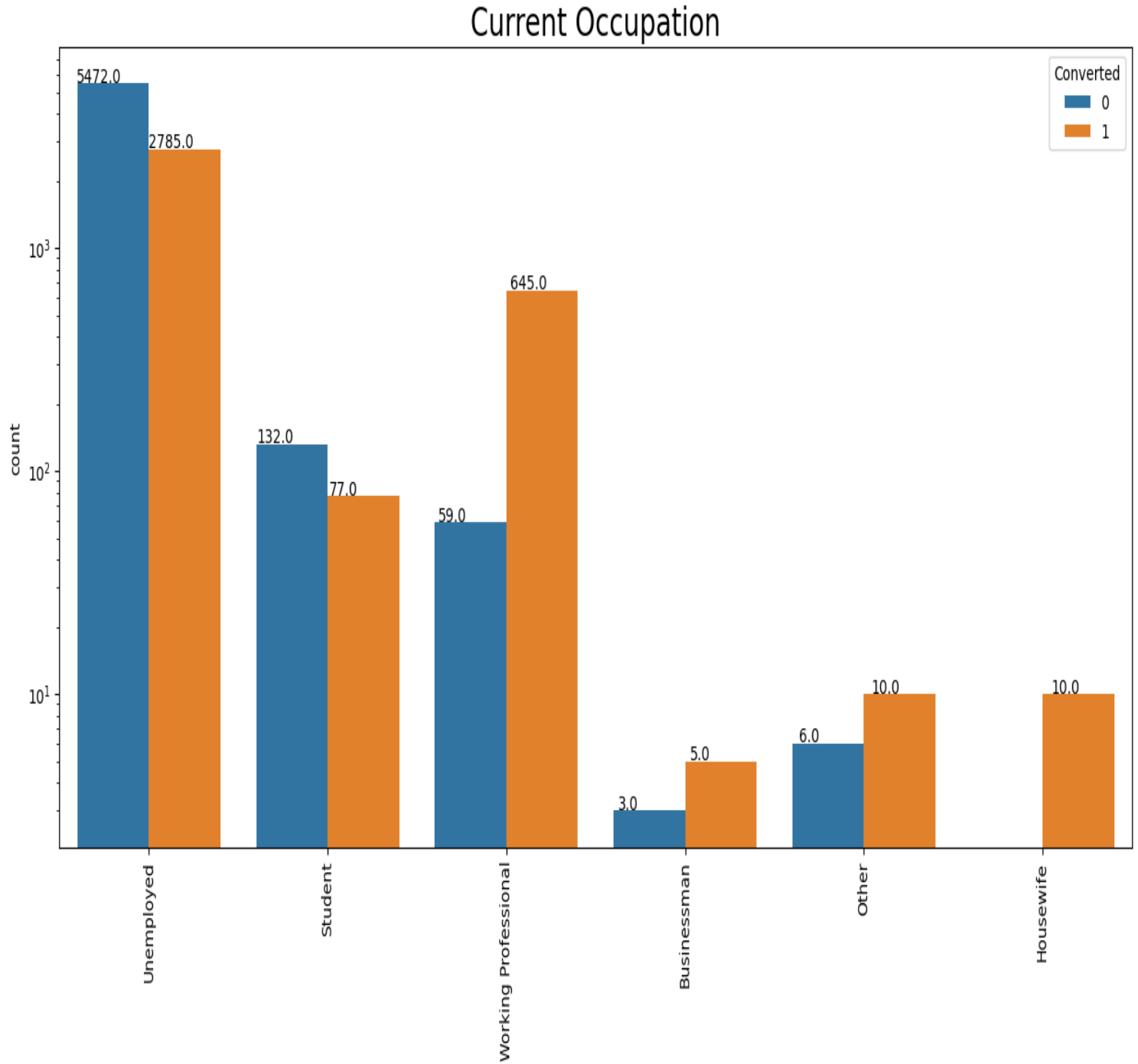
Graph's conclusion

- 1.Google and Direct traffic generate most leads.
- 2.Conversion Rate of reference leads and leads through welingak website is high.



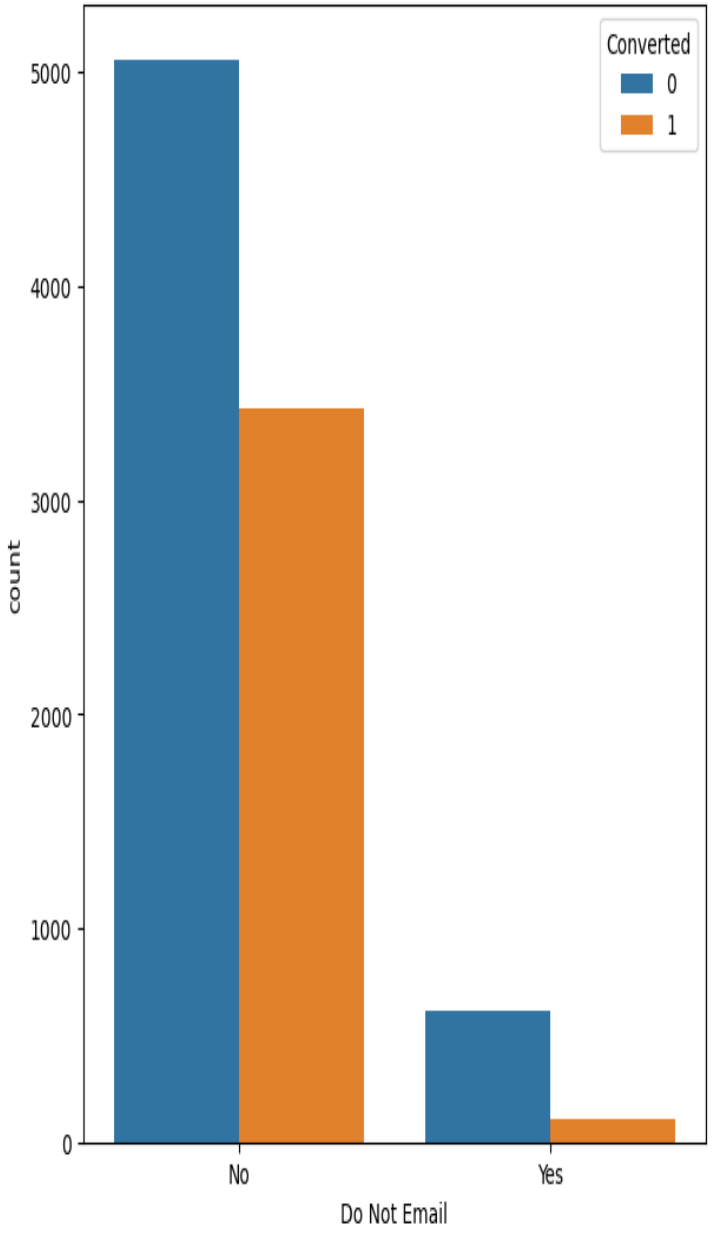
Graph's conclusion

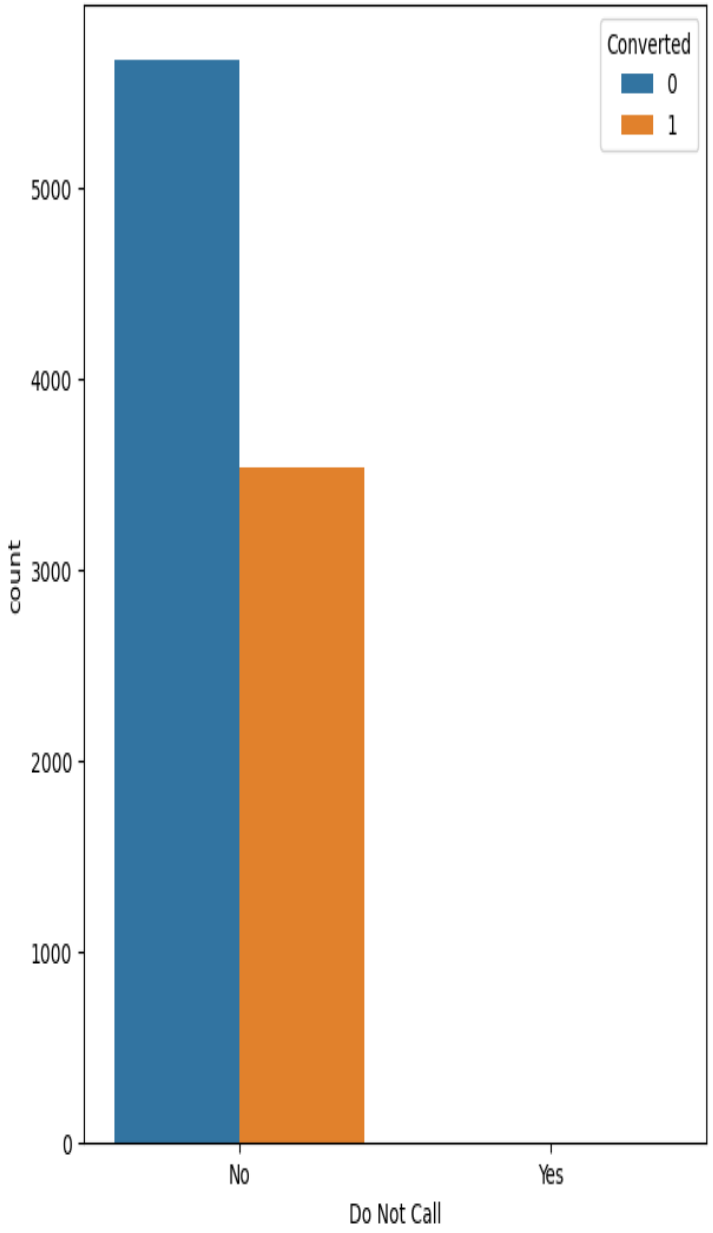
- 1.API and Landing Page Submission have 30-35% conversion rate but count of leads is high.
- 2.Lead Add Form has more than 90% conversion rate but count of leads is on the lower side.
- 3.Lead Import are very less in count.



Graph's conclusion

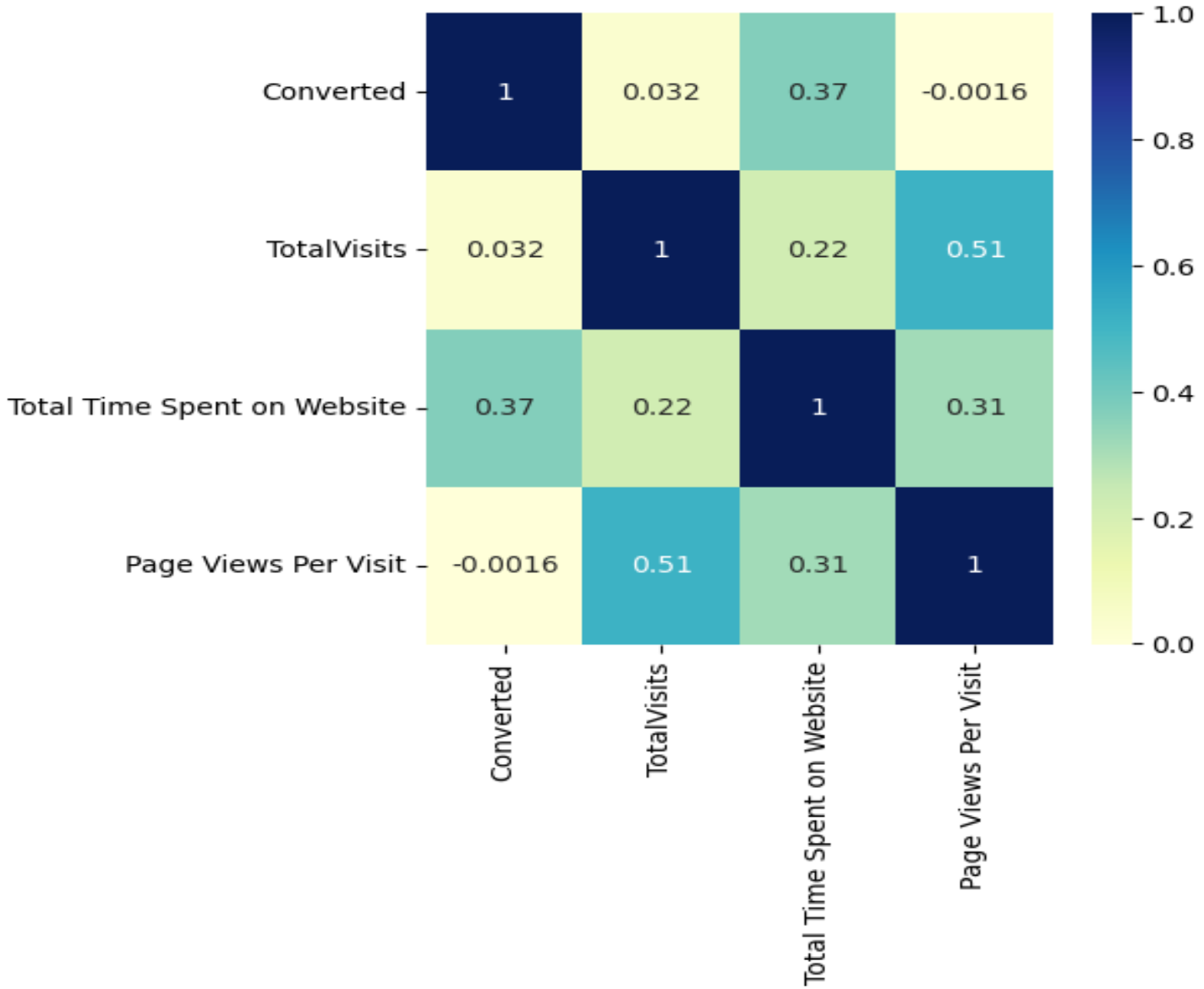
- 1. 'Unemployed' leads are generating more number of leads and having ~34% conversion rate.
- 2. Conversion rate is higher for 'Working Professionals' .





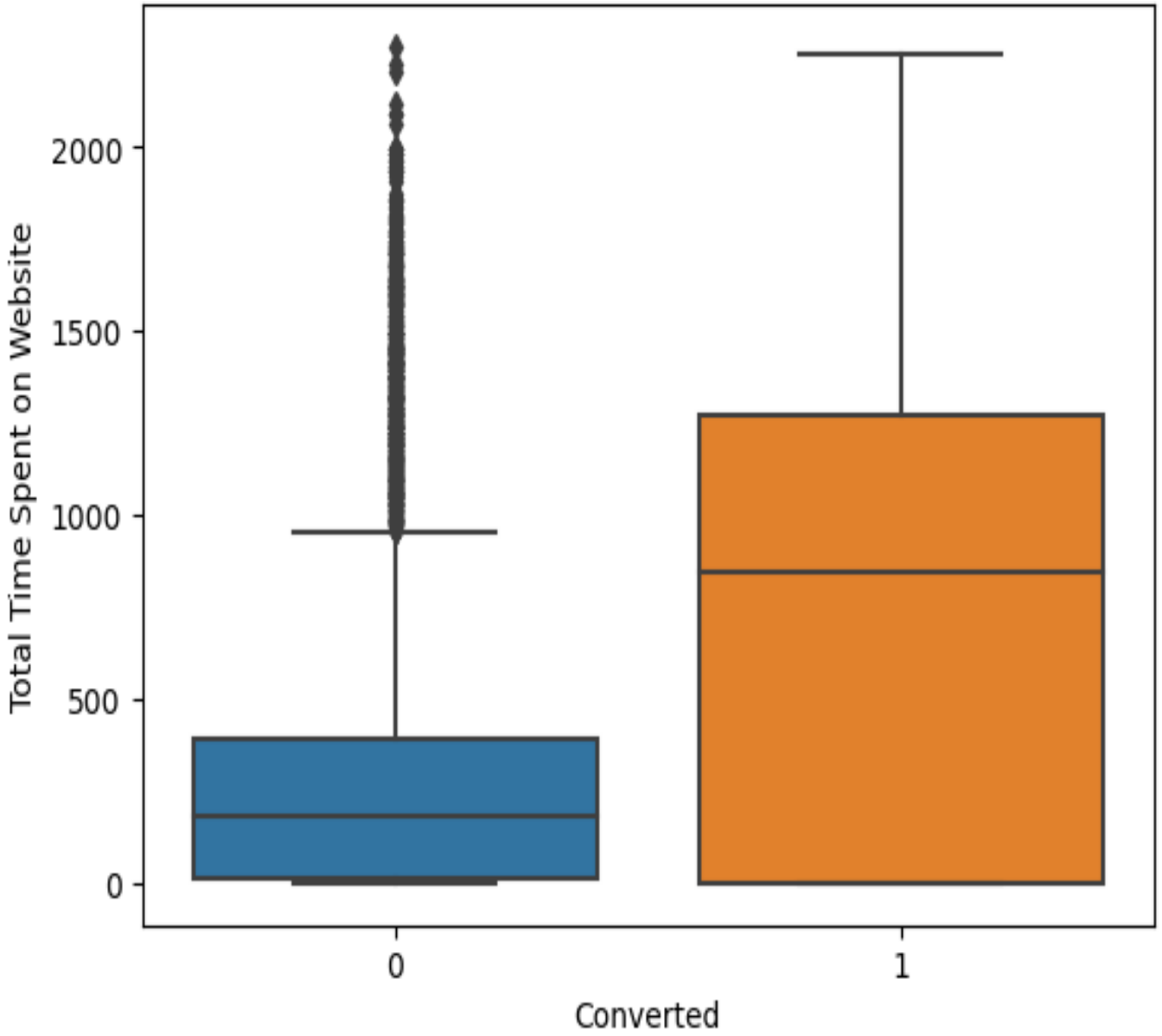
Graph's conclusion

- As we can see so many customers choose the option for Do Not Emails and Do Not Call, which means most leads prefer not to informed neither through phone nor to be mailed.



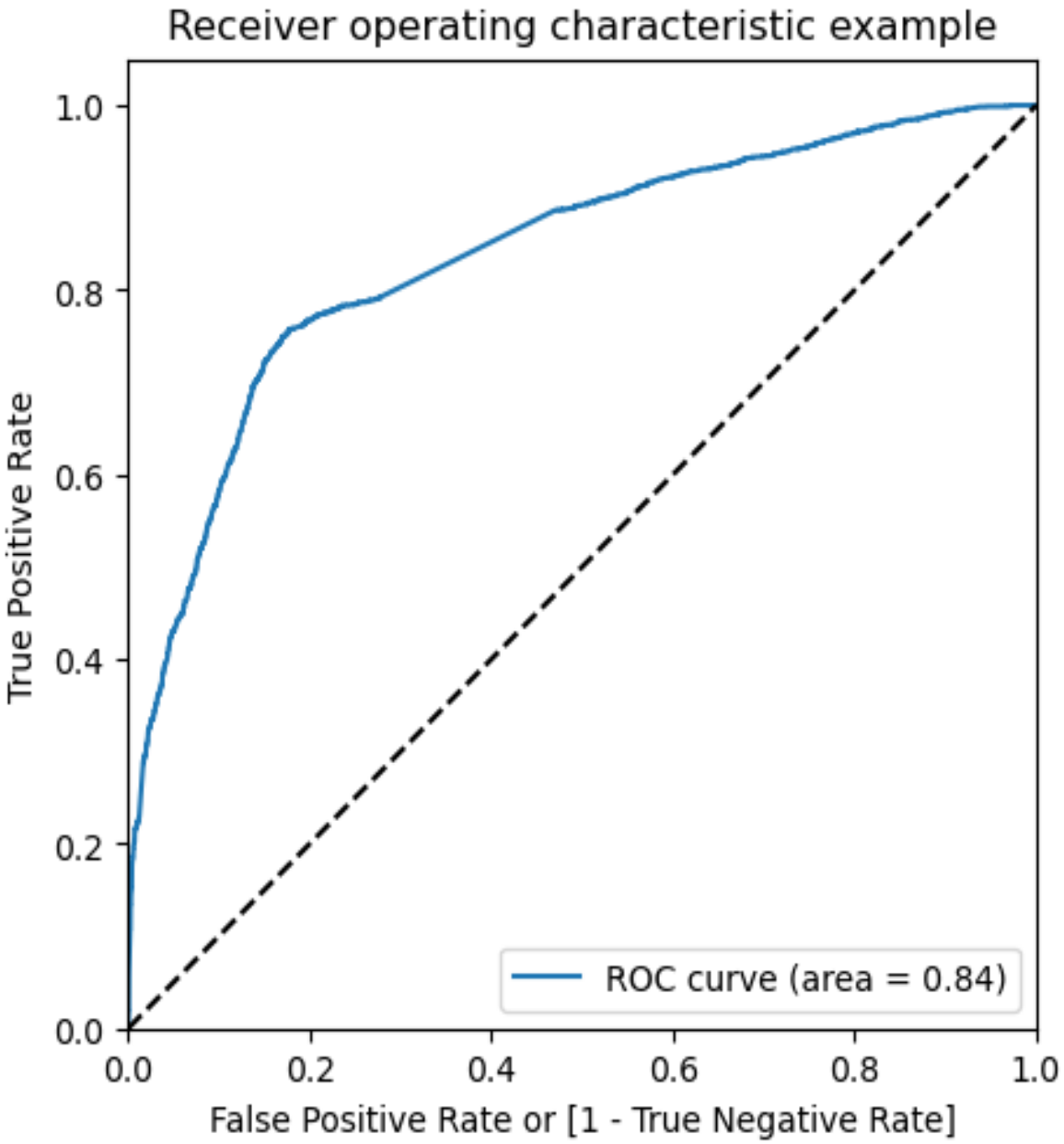
Graph's conclusion

- 1.There is positive correlation between Total Time Spent on Website and Conversion
- 2.There is almost no correlation in Page Views Per Visit and Total Visits with Conversion.



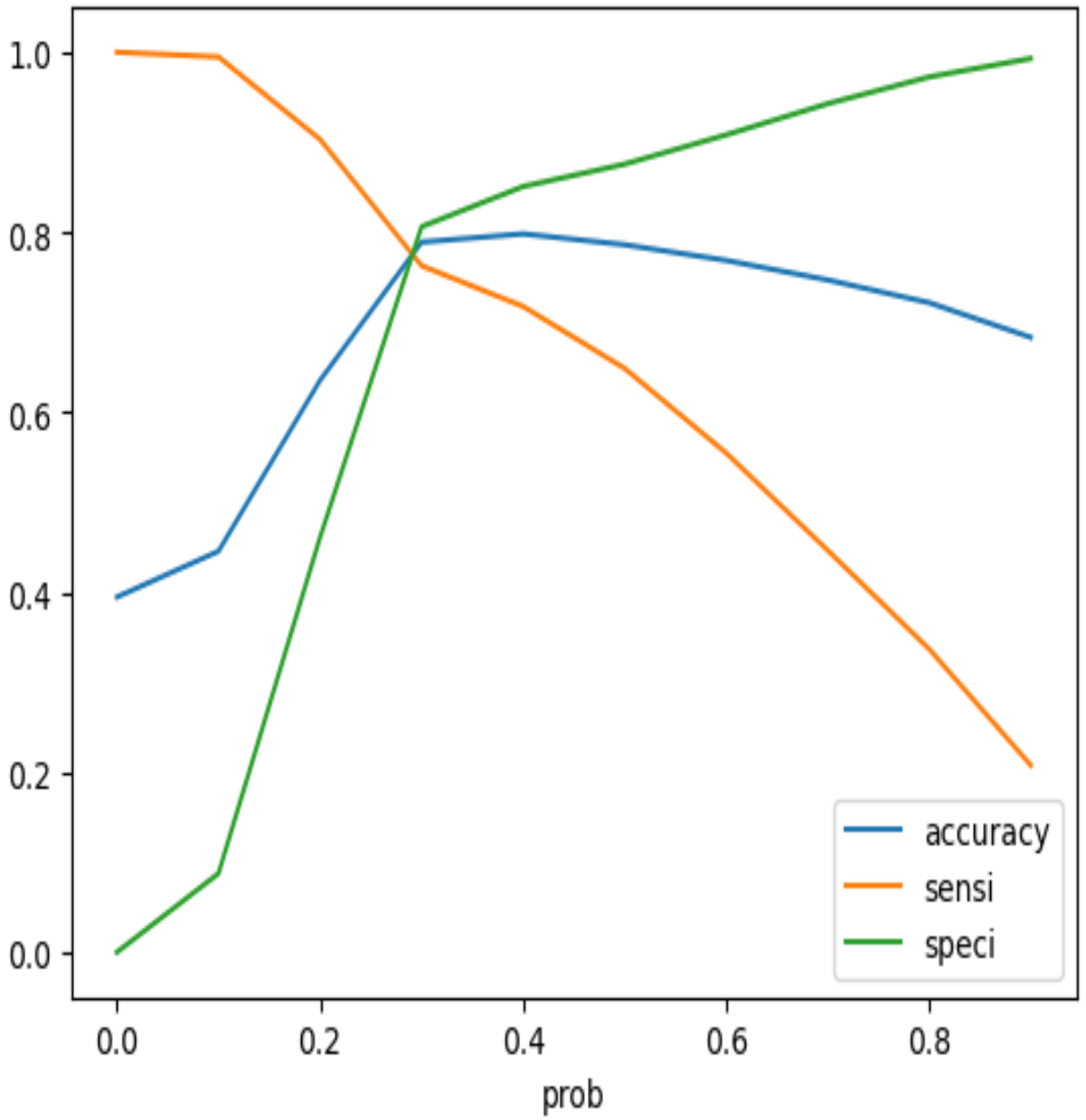
Graph’s conclusion

- 1.Leads spending more time on the website are more likely to be converted.
- 2.Website should be made more engaging to make leads spend more time.



Graph's conclusion

- The area under the curve of the ROC is 0.84 which is quite good. So we seem to have a good model.



Graph's conclusion

- As we get the optimal values of the three metrics at 0.3, so the cutoff value would be 0.3 .

CONCLUSION

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.
- Accuracy, Sensitivity and Specificity values of test set are around 79%, 76% and 81% which are approximately closer to the respective values calculated using trained set.
- Also the lead score calculated shows the conversion rate on the final predicted model is around 72% precision and 79% recall.
- The top 3 variables that contribute for lead getting converted in the model are :

I. Total time spent on website.

II. Lead source reference

III. If the lead is already a student, chances are they will not take up another course which is designed for working professionals.

- Hence overall this model seems to be good.

THANK YOU