

# LEAD SCORE CASE STUDY

BY:

- KHWAHISH PATEL
- YASH THAKUR
- ROHIT NAYAK

# PROBLEM STATEMENT

An education company named X Education sells online courses to industry professionals. The company markets its courses on several websites and search engines like Google.

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%.

# GOALS OF THE CASE STUDY

Find leads that have high chances of converting i.e buying the course.

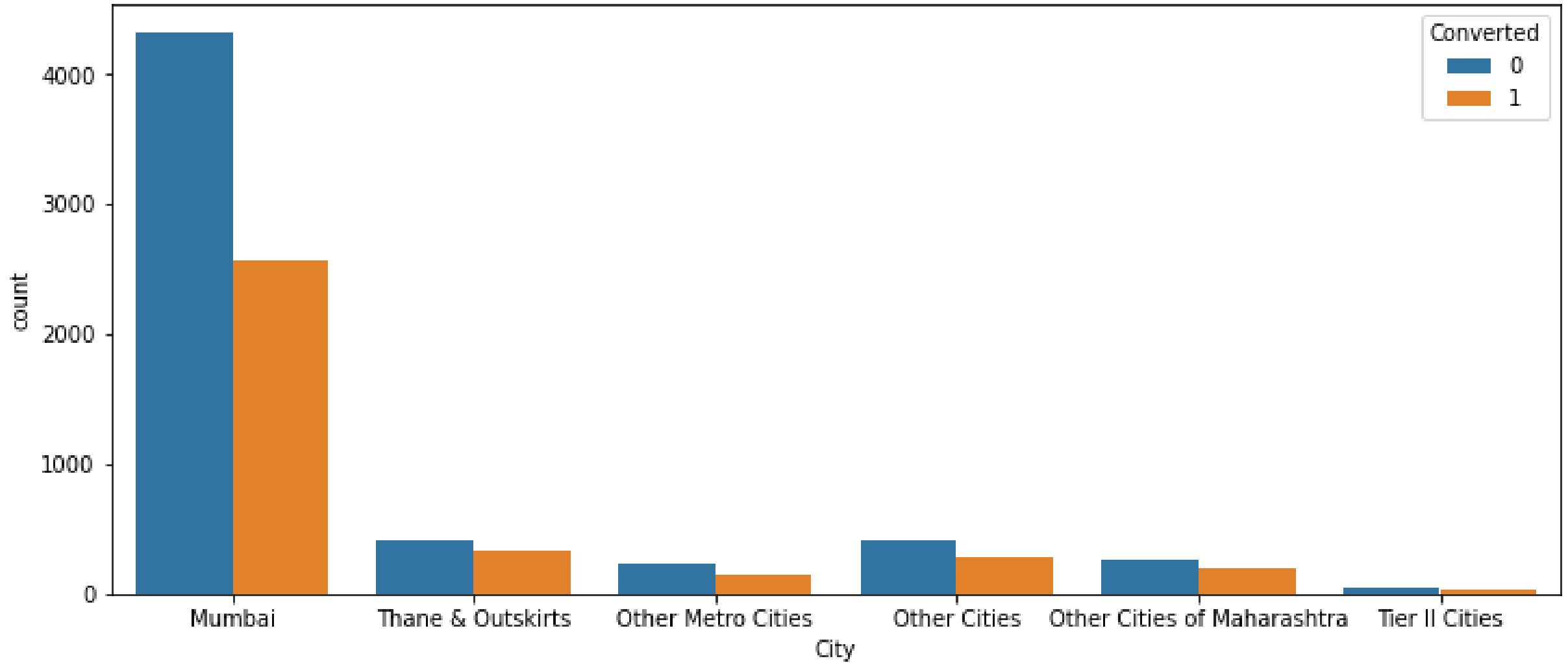
Build a model that assigns a score to leads between 0 to 100 which can be used by the company to target potential leads.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

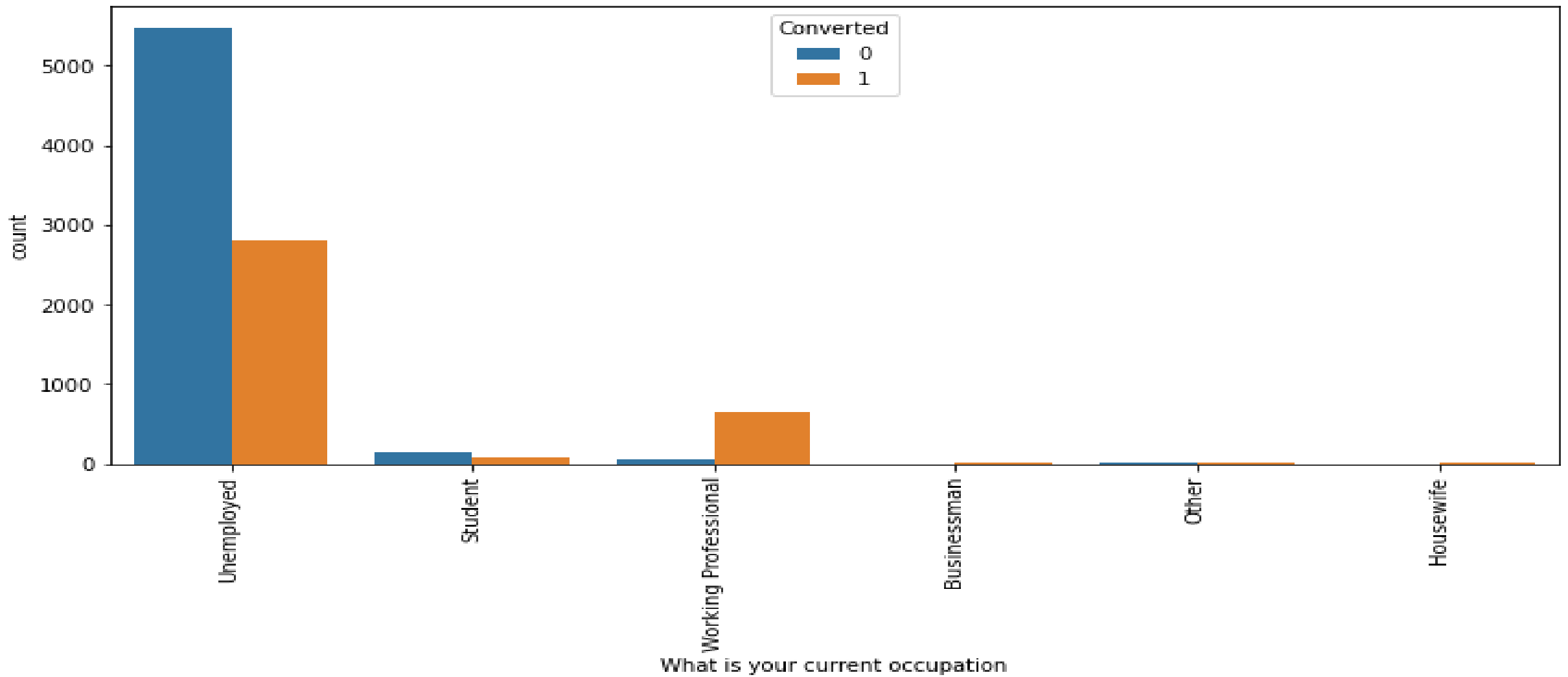
# METHODOLOGY

- Data Sourcing
- Data Manipulation and Cleaning
- EDA
- Feature Scaling, Dummy Variable Creation and Splitting the data into Train and Test Sets
- Model Building and Validation
- Model Evaluation
- Conclusion

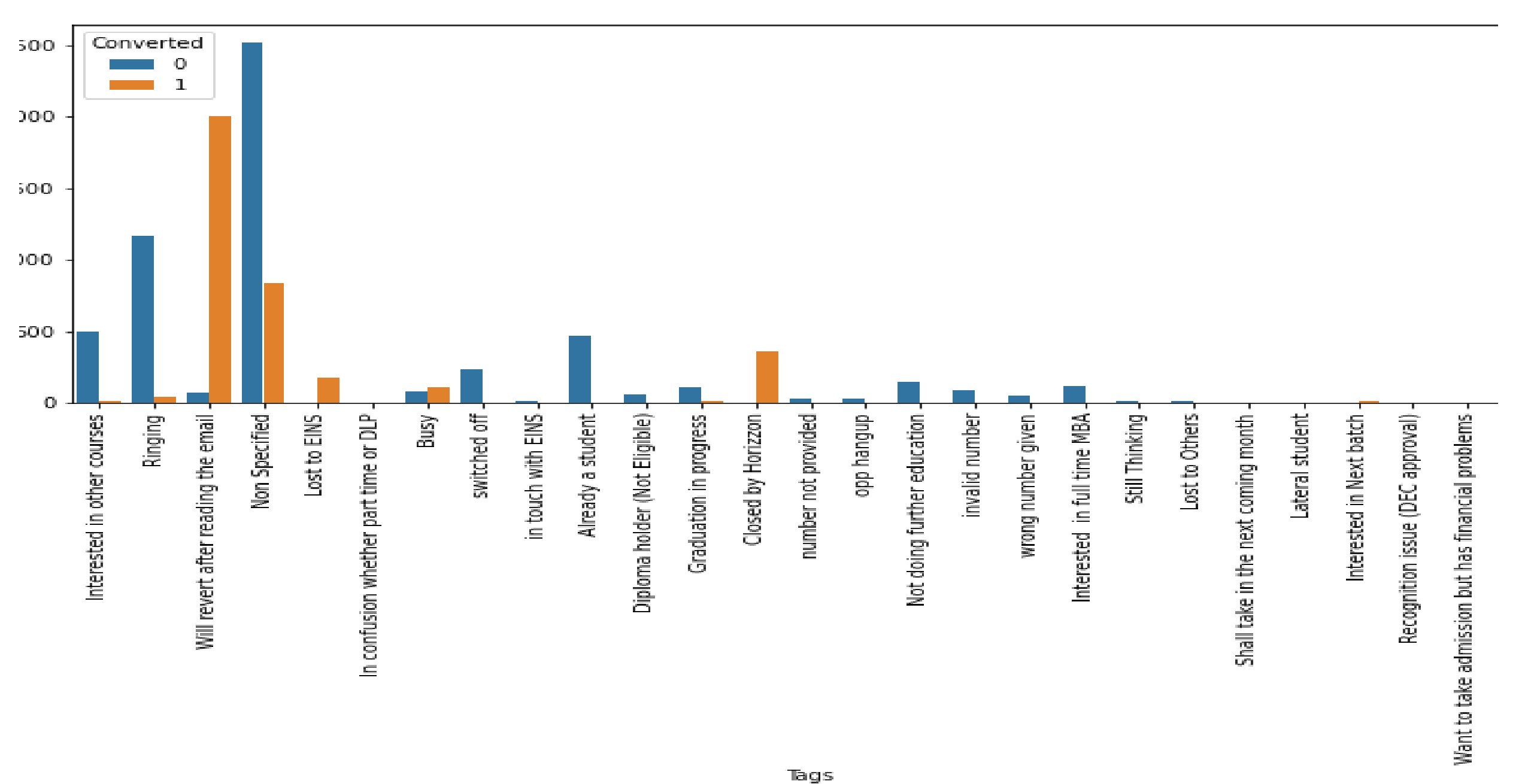
# EDA



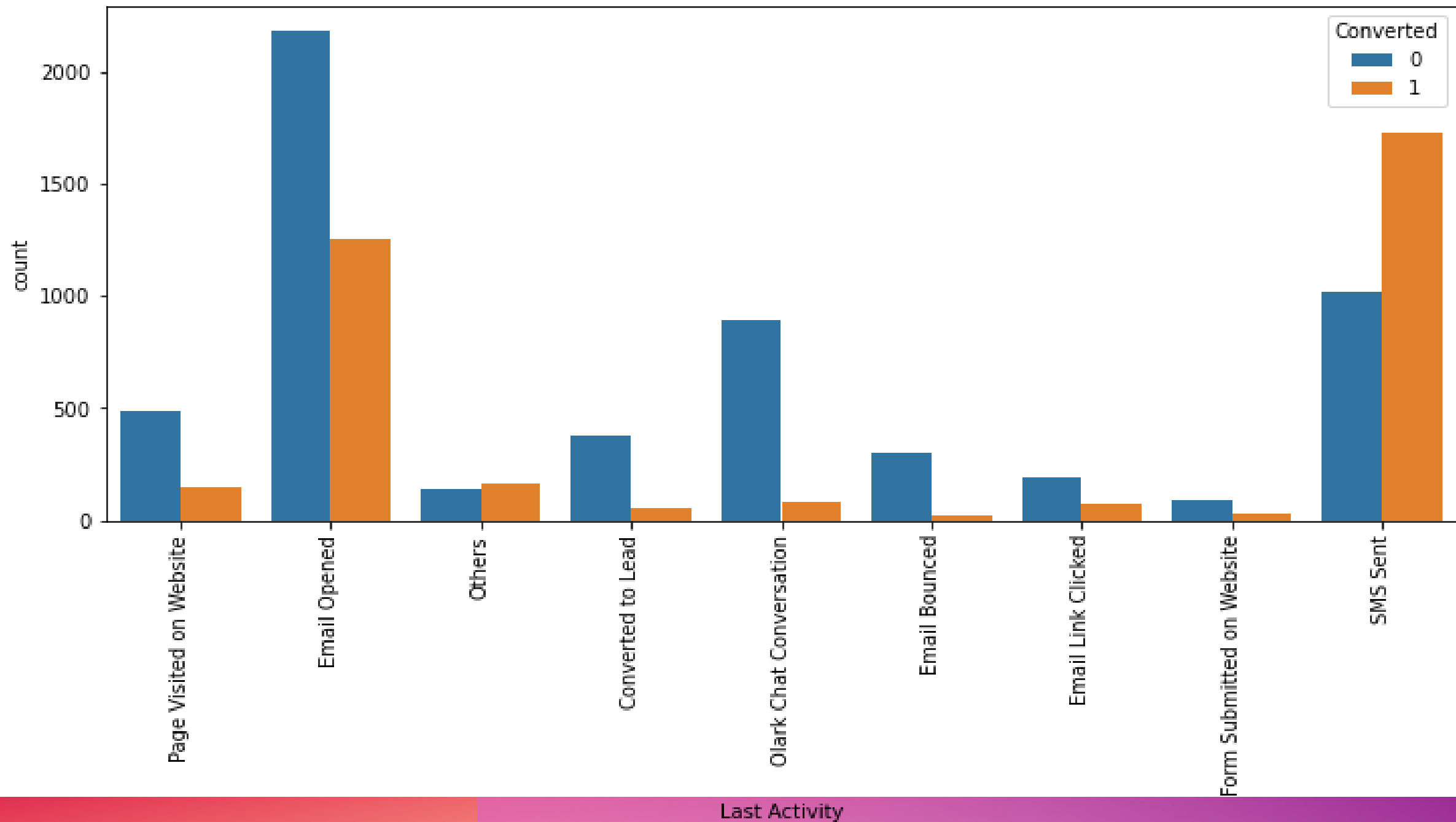
- Most of the leads are from Mumbai.



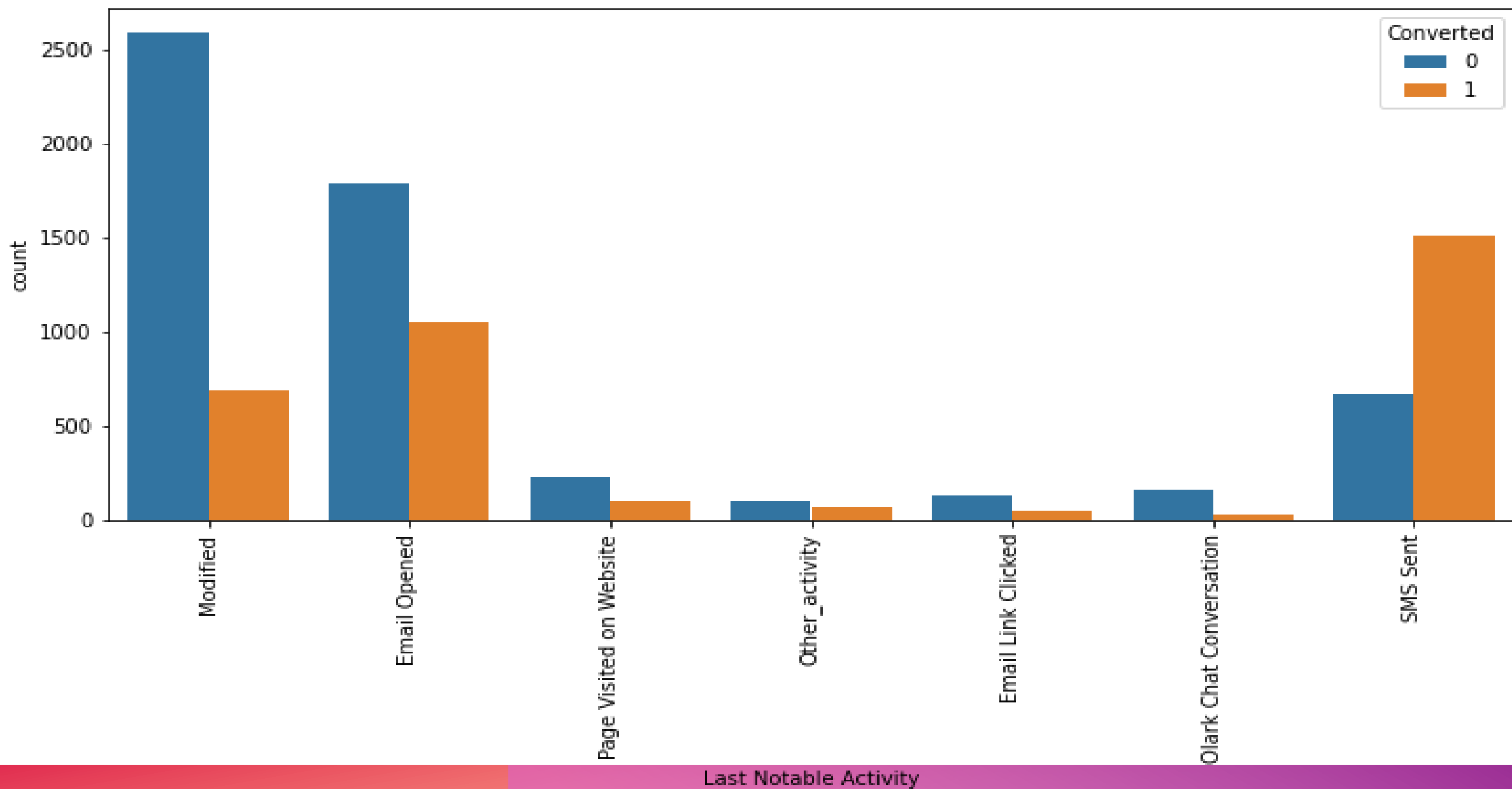
Count for Unemployed is high but Working Professionals have a high conversion rate

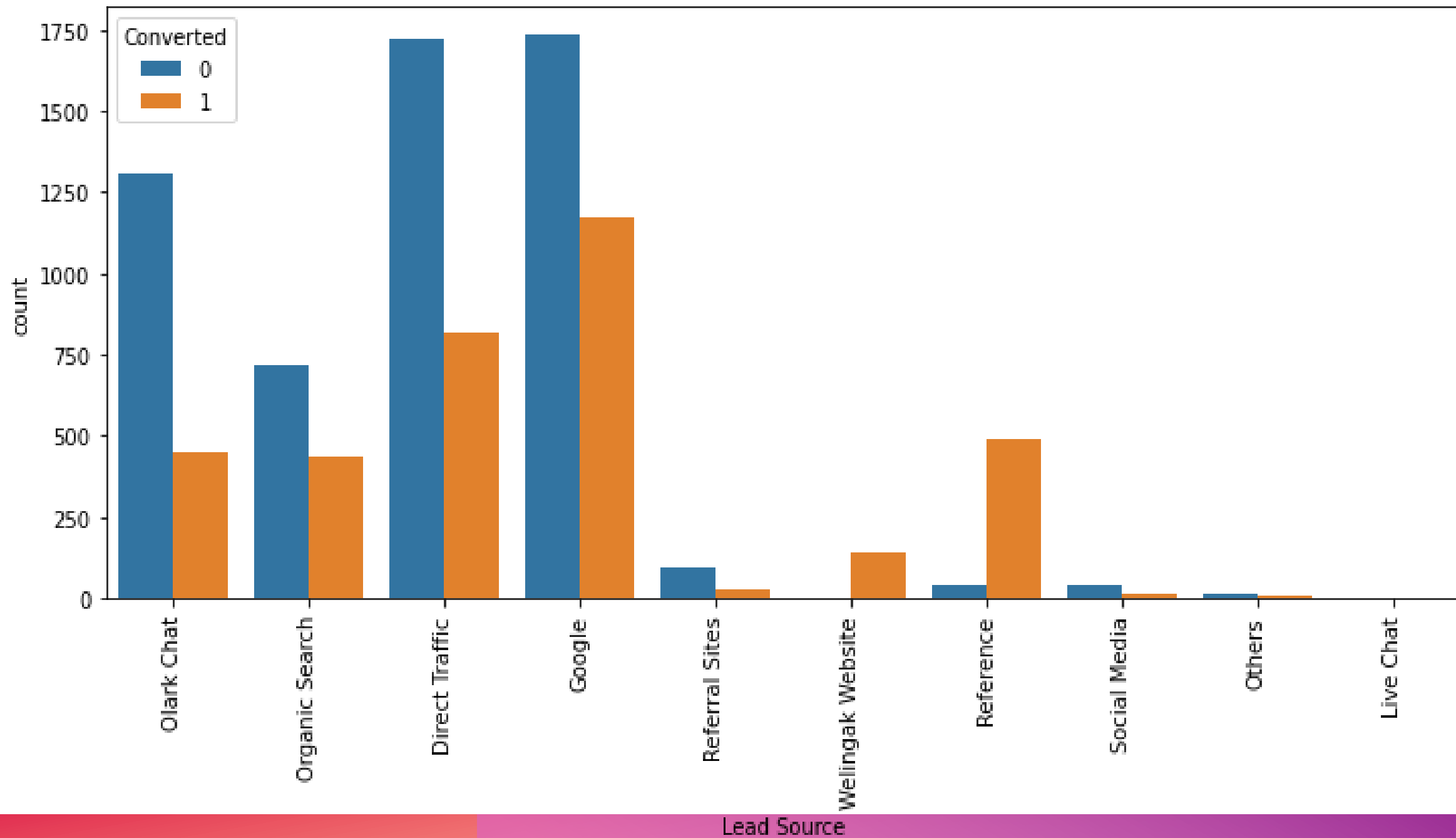


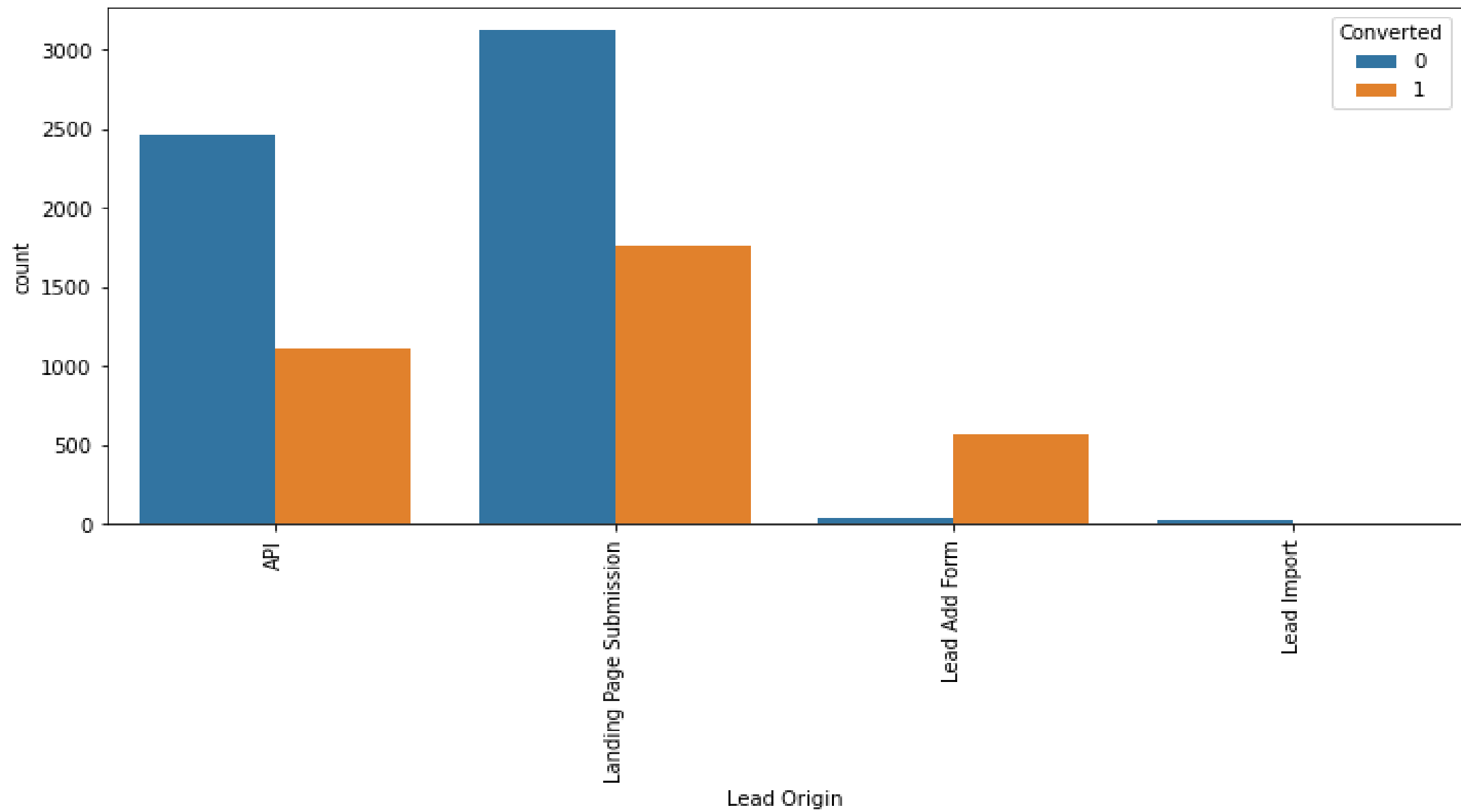
Will revert after reading the email has a very high amount of conversion.

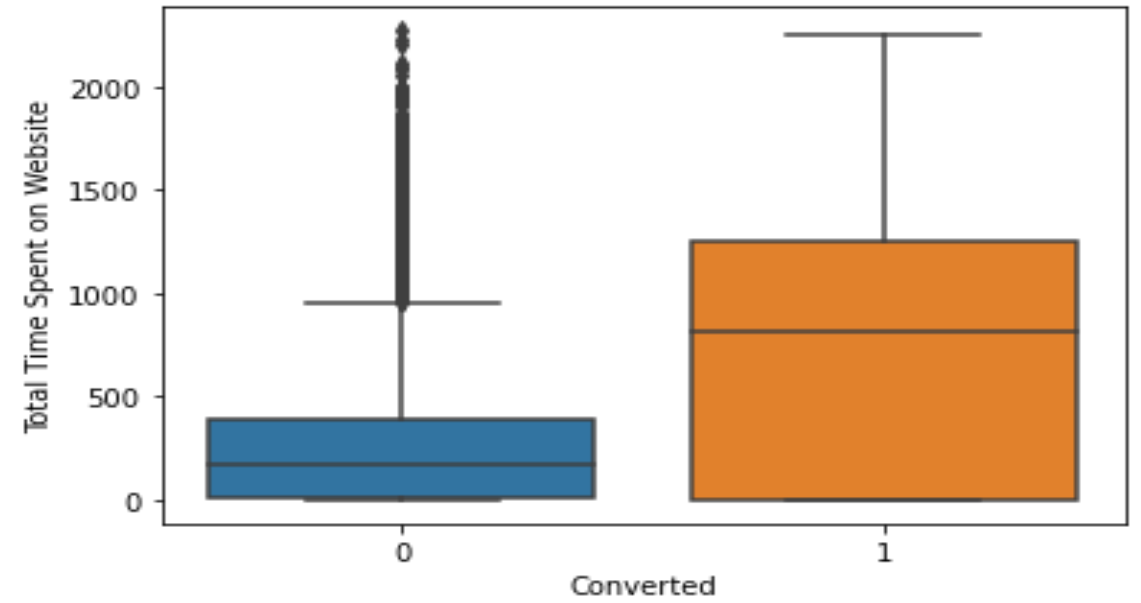
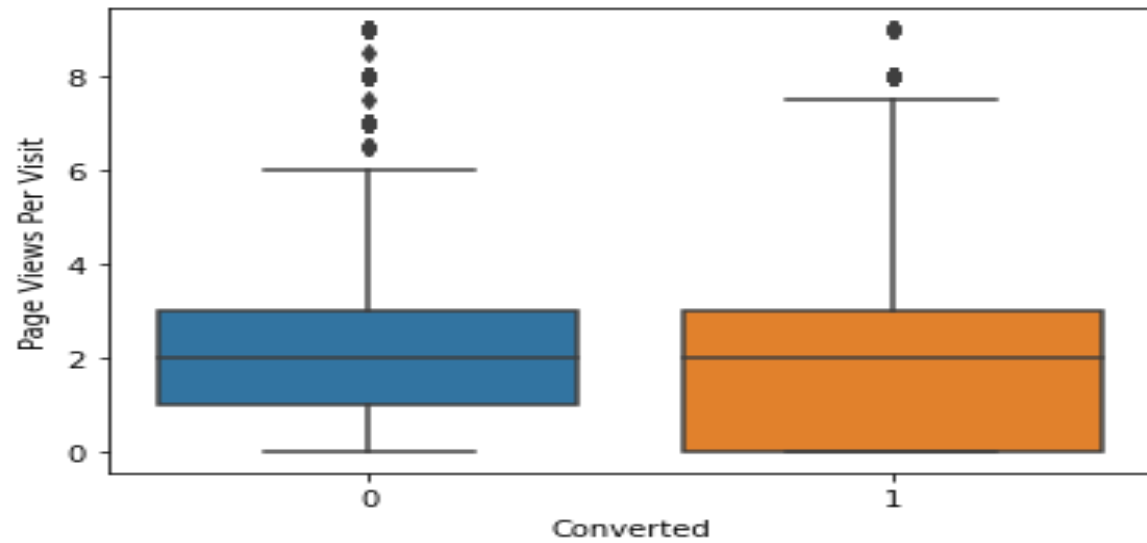
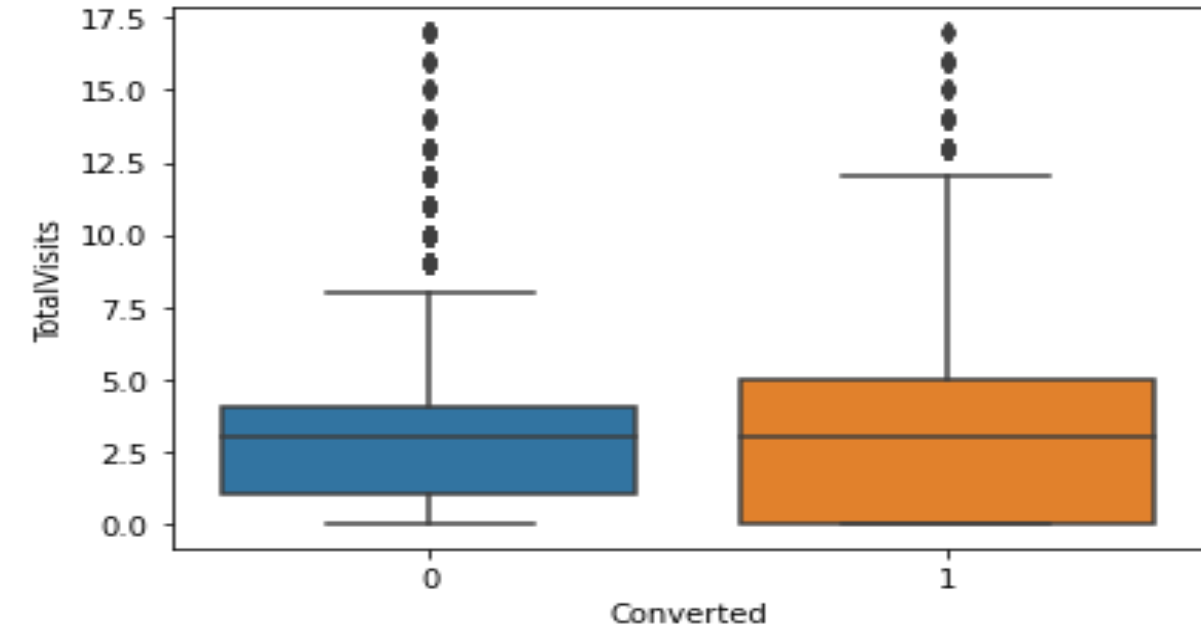






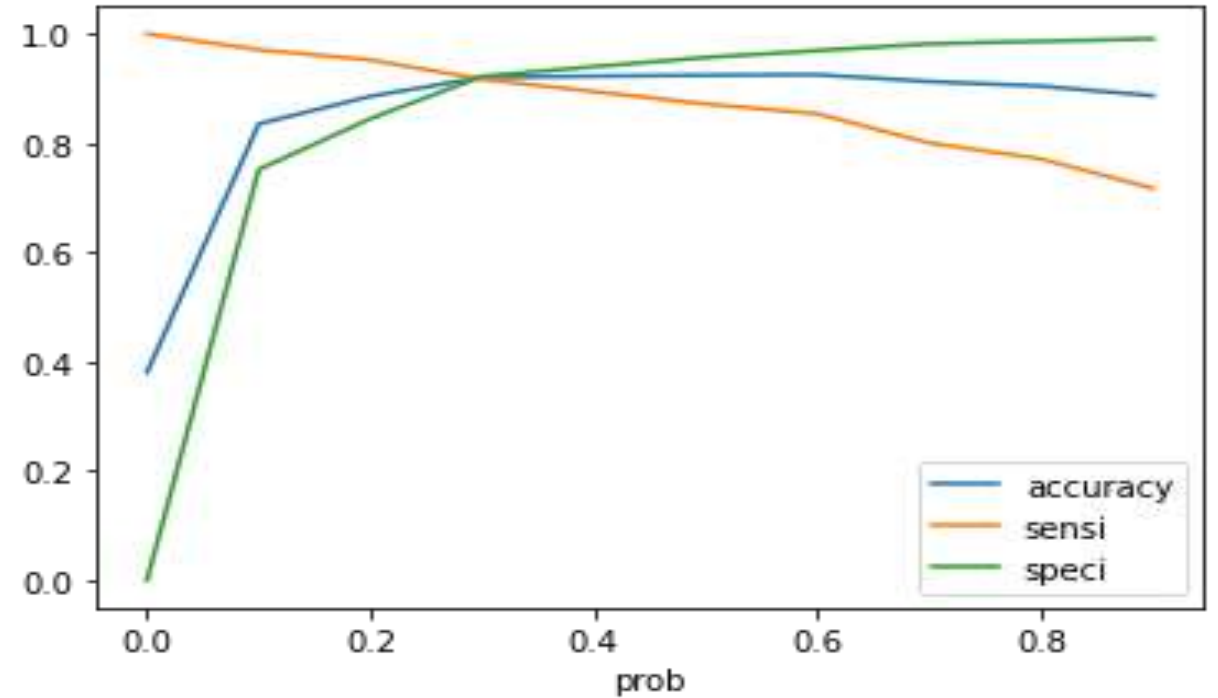
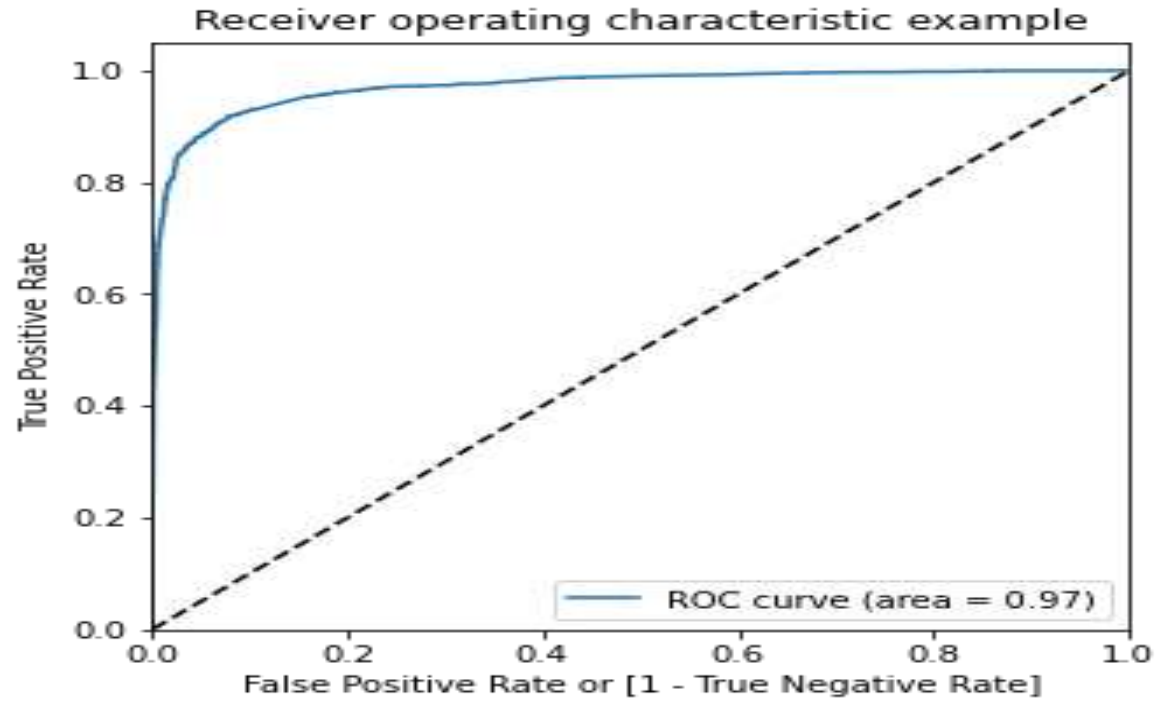






- Lead spending more time on the website have a high chance for conversion.
- Total Visits and Page Views Per Visit nothing conclusive can be observed.

# ROC AND FINDING OPTIMAL CUT-OFF



- ROC is 0.97 which is good.
- Optimal cut-off is 0.3.

# MODEL EVALUATION ON TRAIN SET

- **Train Data:**
- Accuracy : 92%
- Sensitivity : 91.69%
- Specificity : 92.22%
- False Positive Rate : 4.32%
- Positive Predicted Value : 92.52%
- Negative Predicted Value : 92.38%
- Precision : 87.86%
- Recall : 91.69%

# MODEL EVALUATION ON TEST SET

- **Test Data:**
- Accuracy : 92.10%
- Sensitivity : 90.89%
- Specificity : 92.84%

# CONCLUSIONS

	coef
const	-0.4709
Total Time Spent on Website	1.0425
Lead Source_Direct Traffic	-1.6601
Lead Source_Google	-1.0360
Lead Source_Organic Search	-1.1346
Lead Source_Welingak Website	4.9715
Last Activity_Olark Chat Conversation	-1.0254
Last Activity_SMS Sent	2.0650
Tags_Closed by Horizzon	7.5370
Tags_Interested in other courses	-2.1876
Tags_Lost to EINS	5.9659
Tags_Other	-2.2780
Tags_Ringing	-3.3331
Tags_Will revert after reading the email	4.7111
Last Notable Activity_Modified	-1.4729

- These are the features which affect the chances of a lead converting or not the most, so the company can make appropriate business decisions based on those.