LEAD SCORE - CASE STUDY

Presented by:

- Sharon Benison
- Shikha Upadhyay
- Sheik Alauddeen

OBJECTIVE:

Problem Statement:

An X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

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. Moreover, the company also gets leads through past referrals. Once the 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%.**

AGENDA:

X Education gets a lot of leads, its lead conversion rate is very poor. To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’.

The company requires us to build a model wherein we 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:

There are quite a few goals for this case study.

1. 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.

2. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

“BUSINESS OPPORTUNITIES ARE LIKE BUSES. THERE'S ALWAYS ANOTHER ONE COMING.”

Richard Branson

STRATEGY:

Ø SOURCE THE DATA FOR ANALYSIS

Ø CLEAN AND PREPARE THE DATA

Ø EXPLORATORY DATA ANALYSIS.

Ø FEATURE SCALING

Ø SPLITTING THE DATA INTO TEST AND TRAIN DATASET.

Ø BUILDING A LOGISTIC REGRESSION MODEL AND CALCULATE LEAD SCORE.

Ø EVALUATING THE MODEL BY USING DIFFERENT METRICS - SPECIFICITY AND SENSITIVITY OR
PRECISION AND RECALL.

Ø APPLYING THE BEST MODEL IN TEST DATA BASED ON THE SENSITIVITY AND SPECIFICITY METRICS.

STEPS:

1. **Data Sourcing , Cleaning and Preparation:**

- Read the Data from the Source
- Convert data into clean format suitable for analysis.
- Remove duplicate data
- Outlier Treatment
- Exploratory Data Analysis (EDA)
- Feature Standardization

2. **Feature Scaling and Splitting Train and Test Sets:**

- Feature Scaling of Numeric data
- Splitting data into train and test set

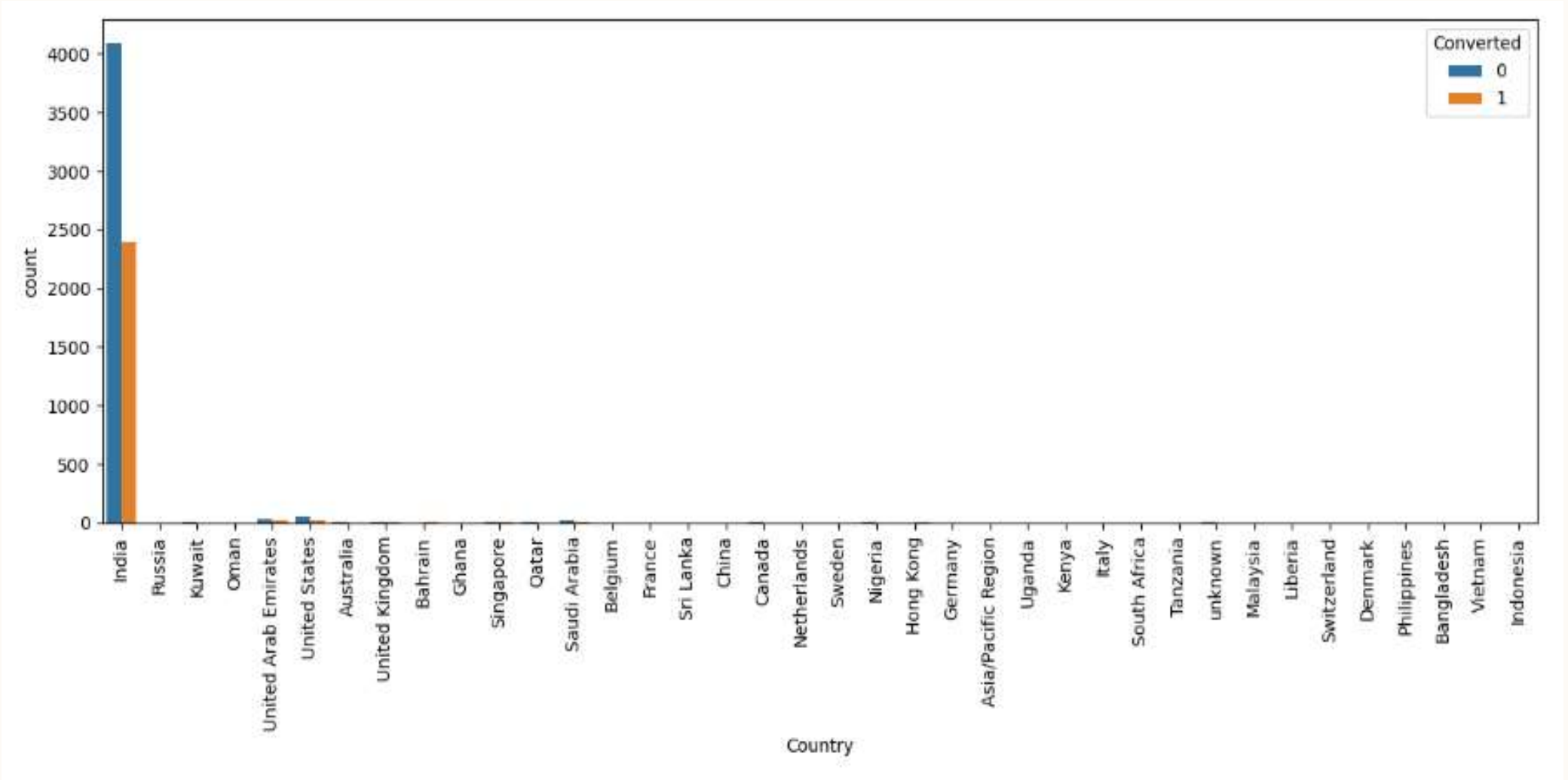
3. Model Building:

- Feature Selection using RFE
- Determine the optimal model using Logistic Regression
- Calculate various metrics like accuracy, sensitivity, specificity, precision and recall and evaluate the model

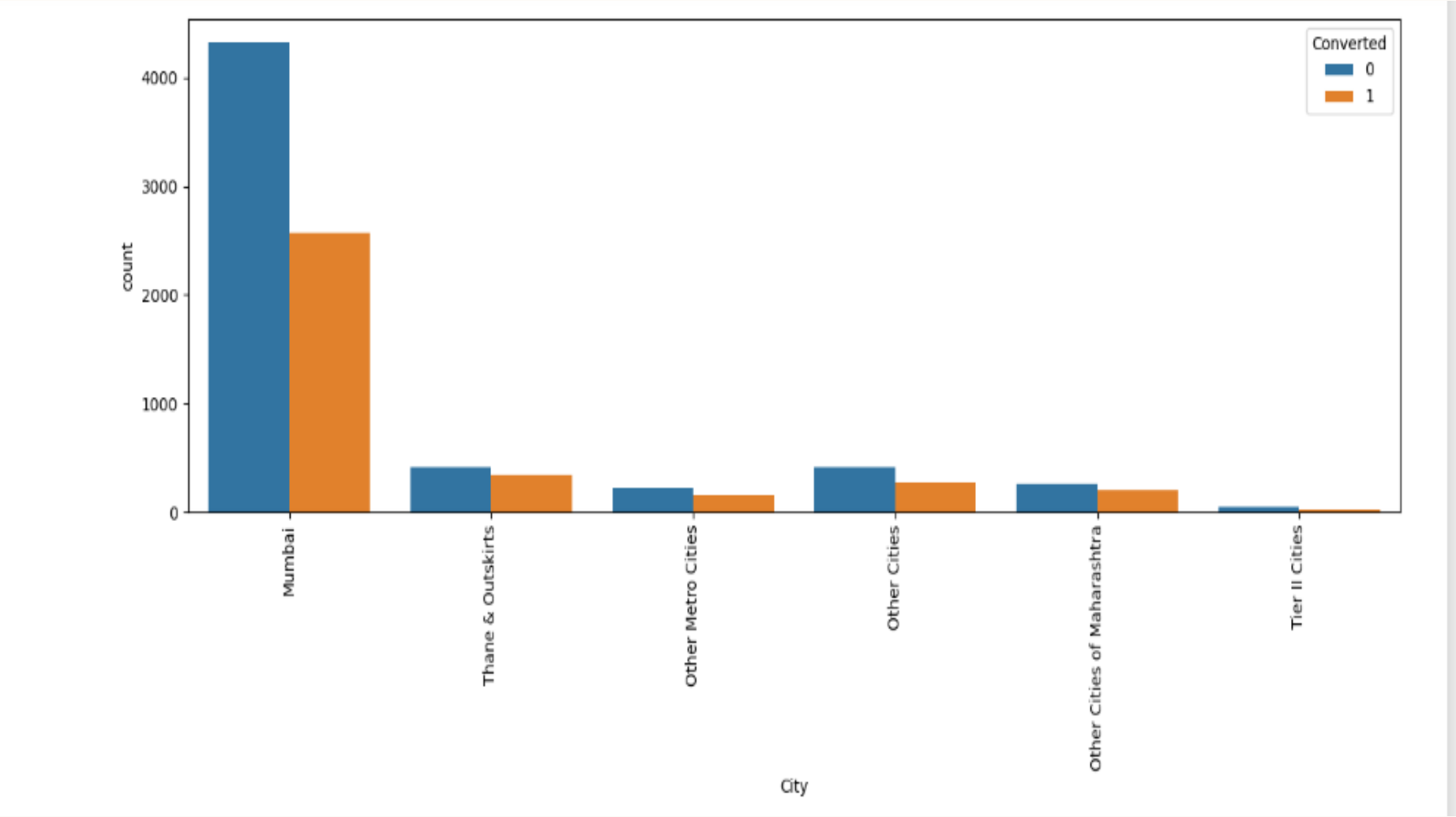
4. Result:

- Determine the lead score and check if target final predictions amounts to 80% conversion rate.
- Evaluate the final prediction on the test set using cut off threshold from sensitivity and specificity metrics.

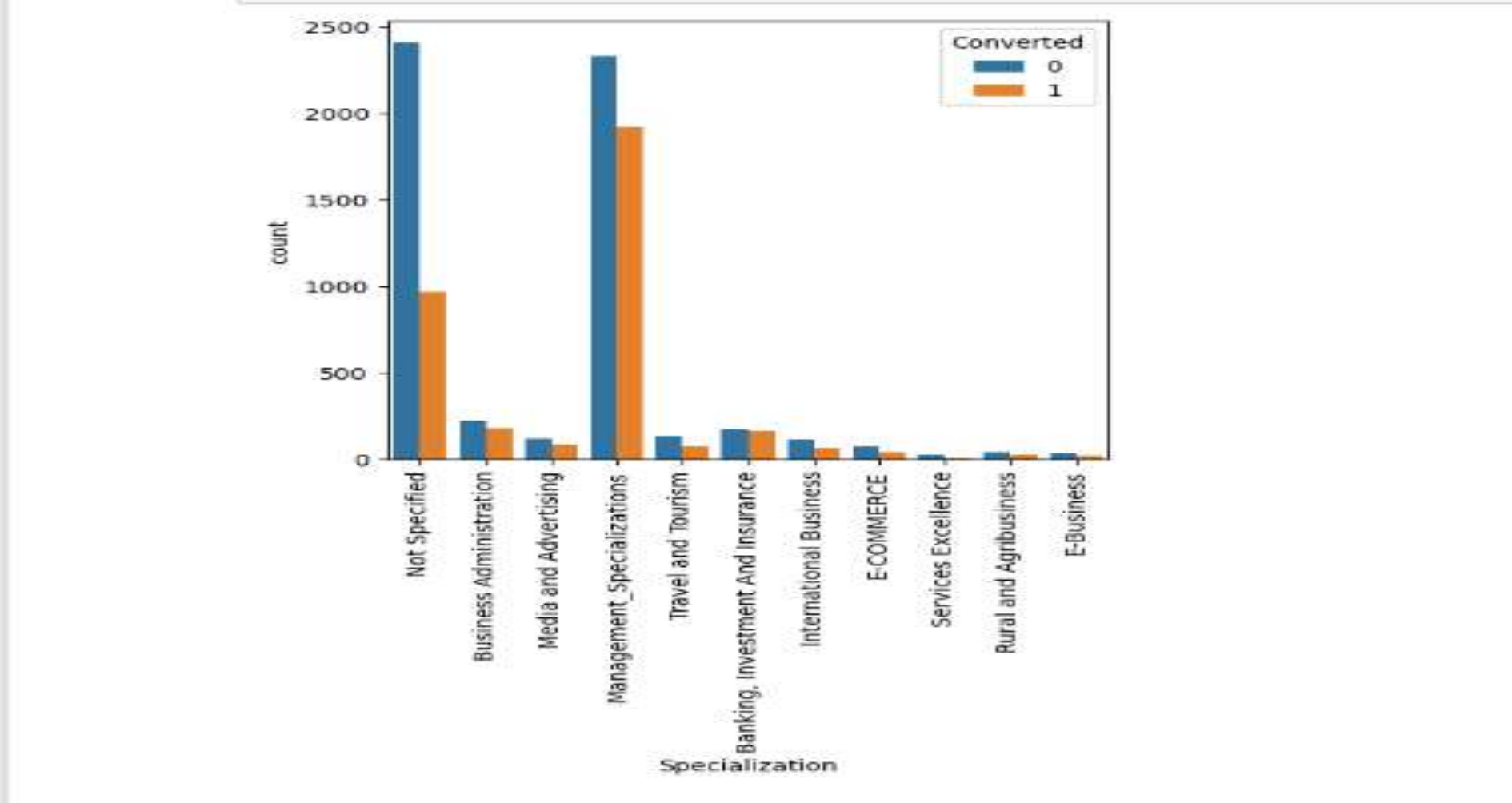
EXPLORATORY DATA ANALYSIS:



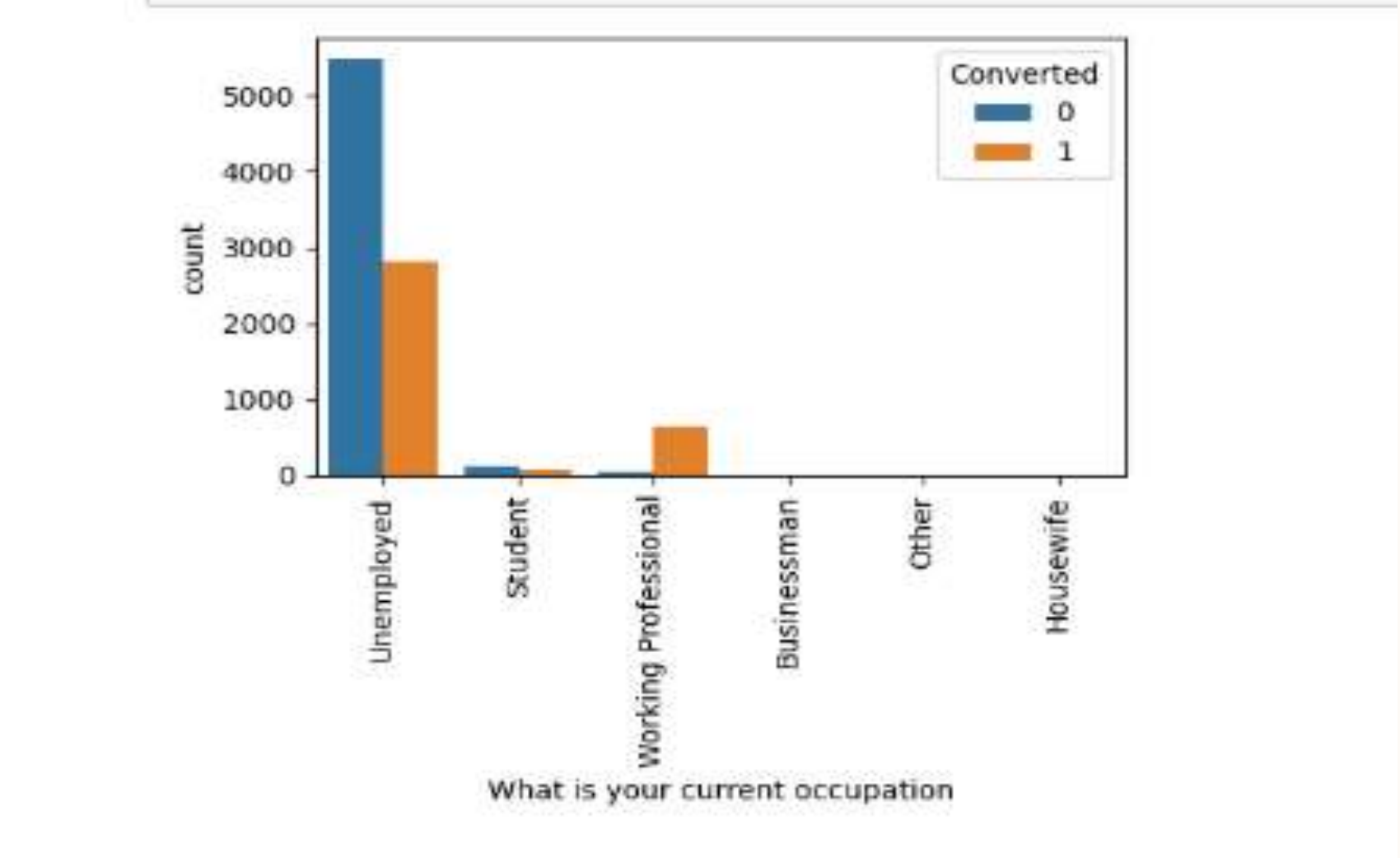
India has the highest number of leads.



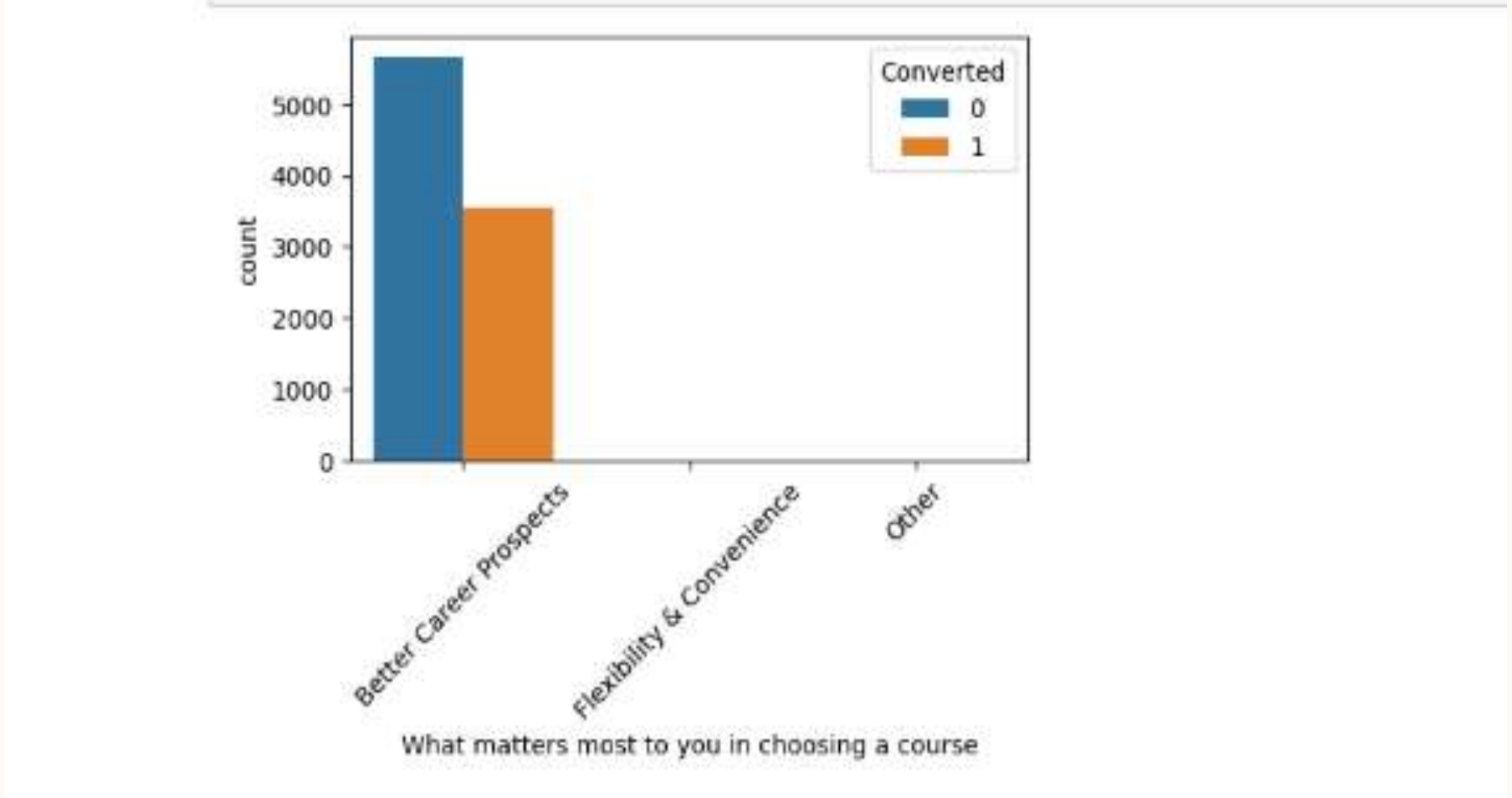
In India, Mumbai has the highest number of leads.



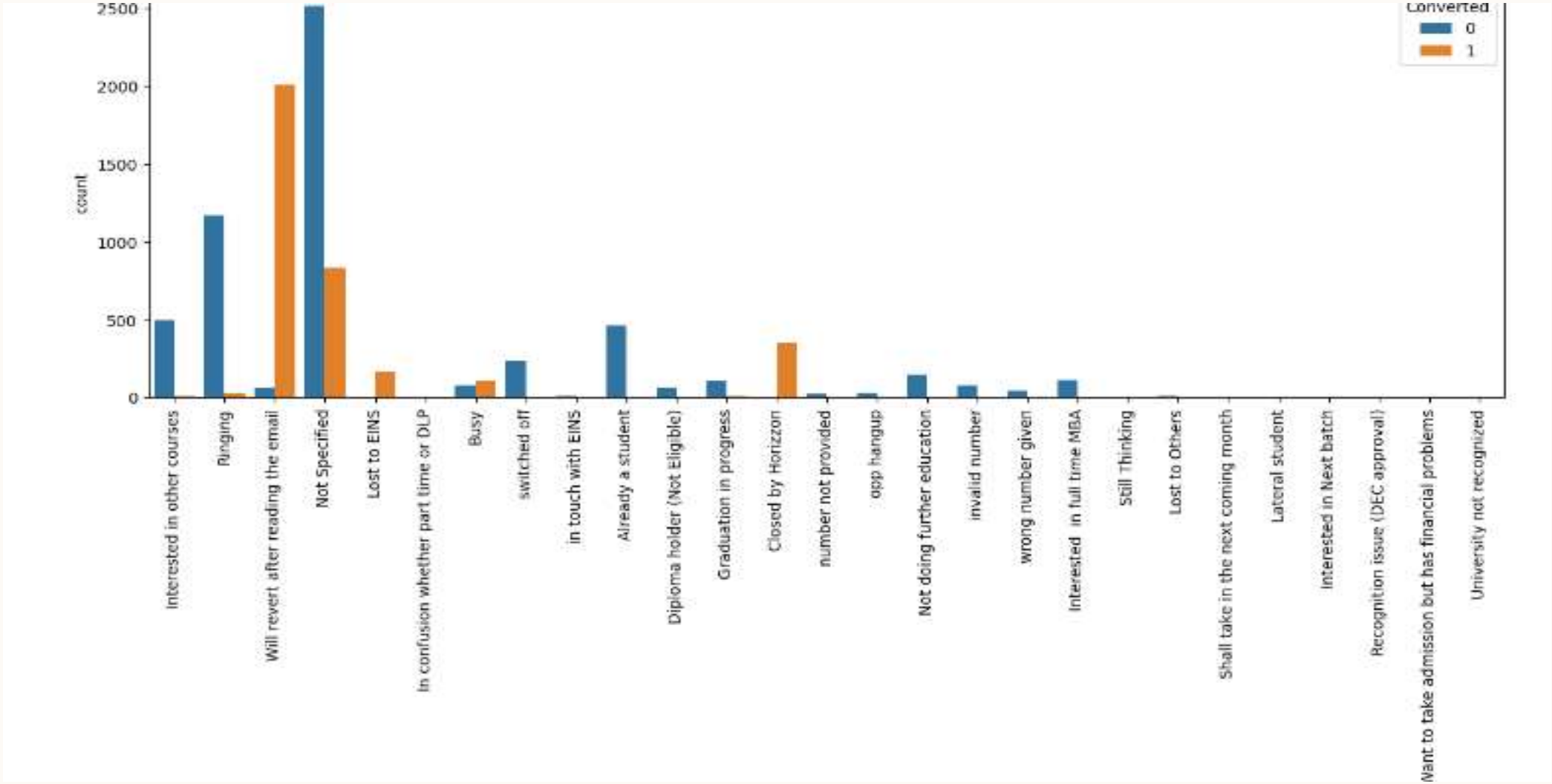
Management specialisations have the highest number of leads and leads-converted.



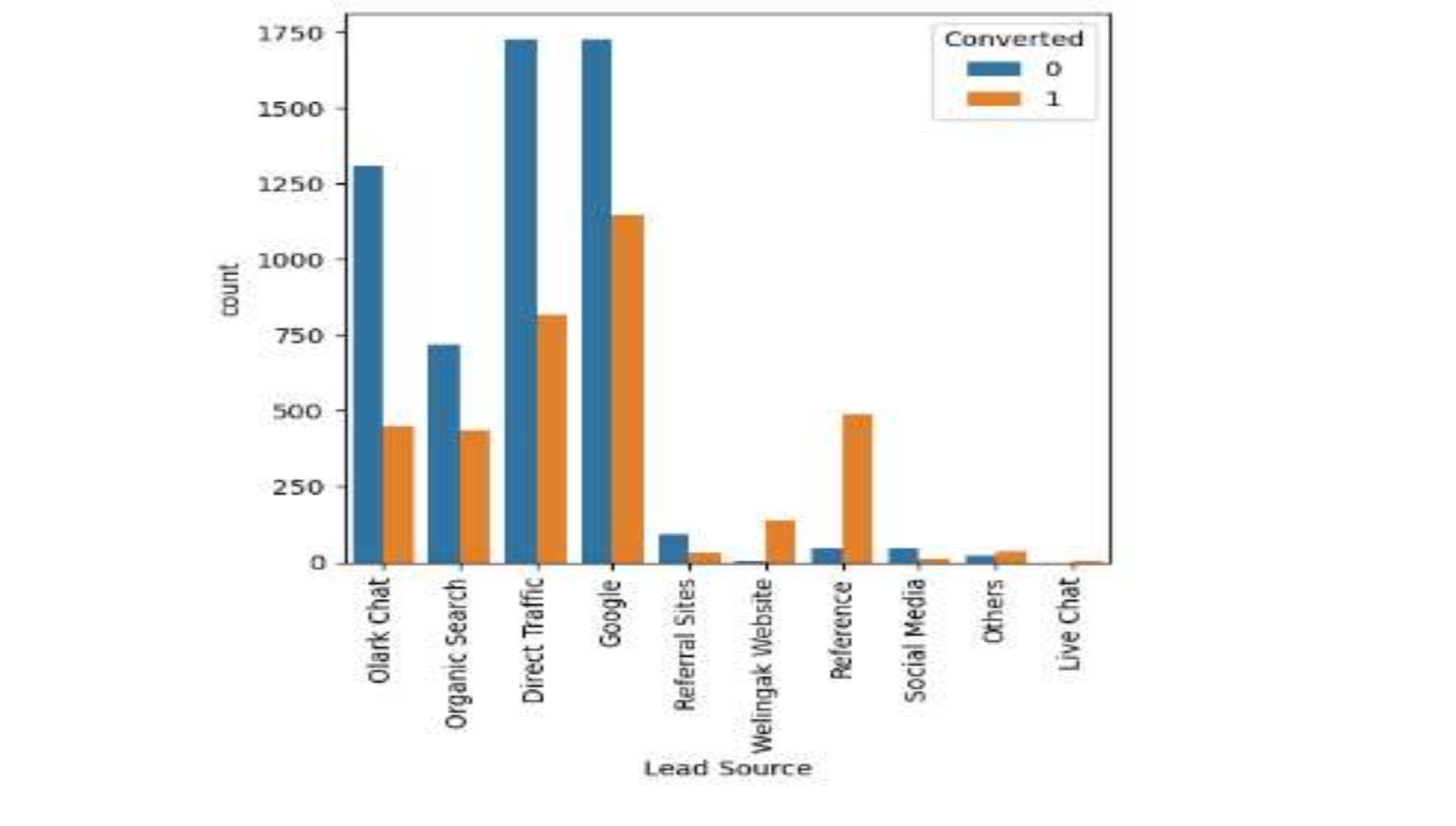
Working professional has the number of leads & leads-converted, followed by Unemployed category .



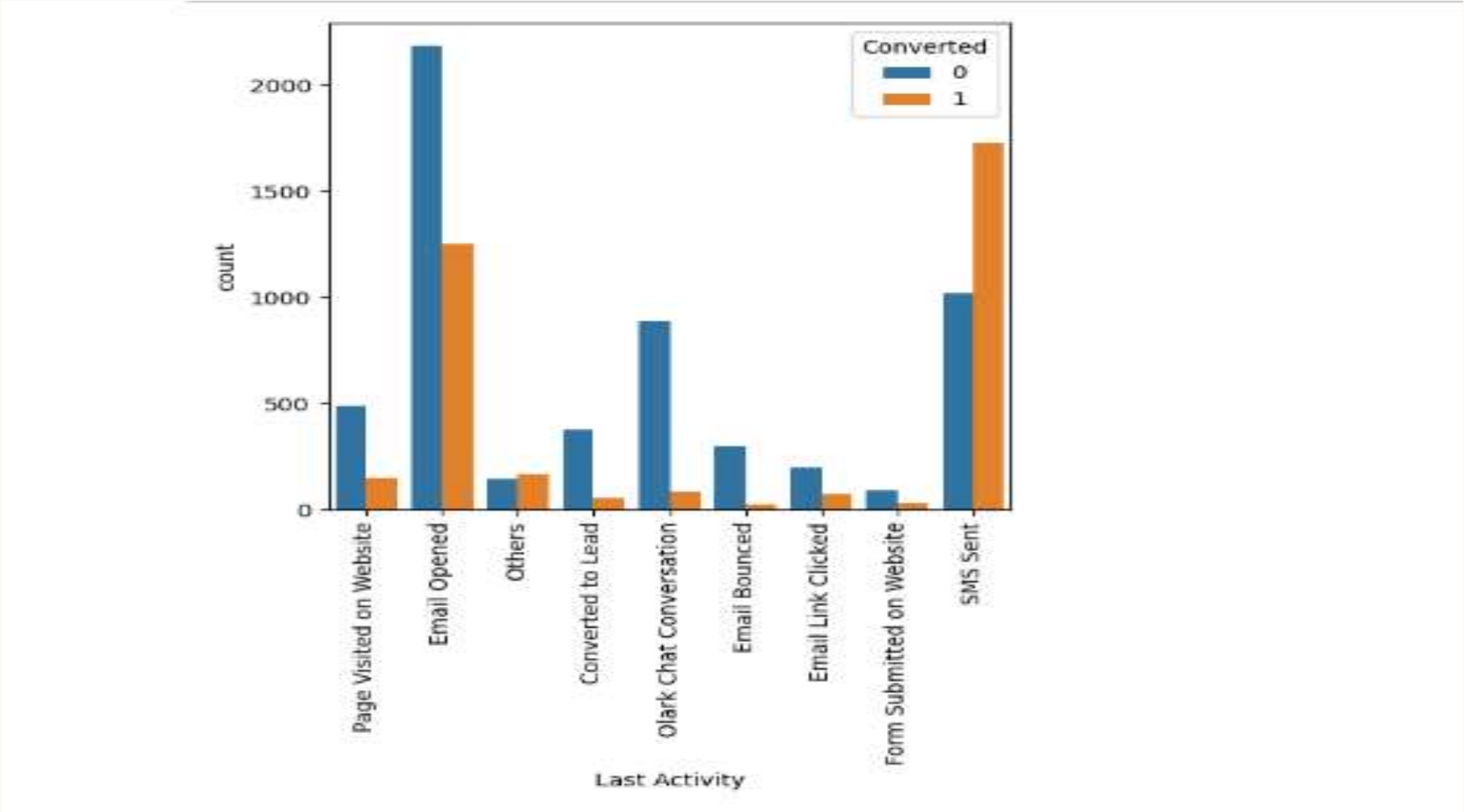
‘Better Career Prospects’ has the highest number of value counts.



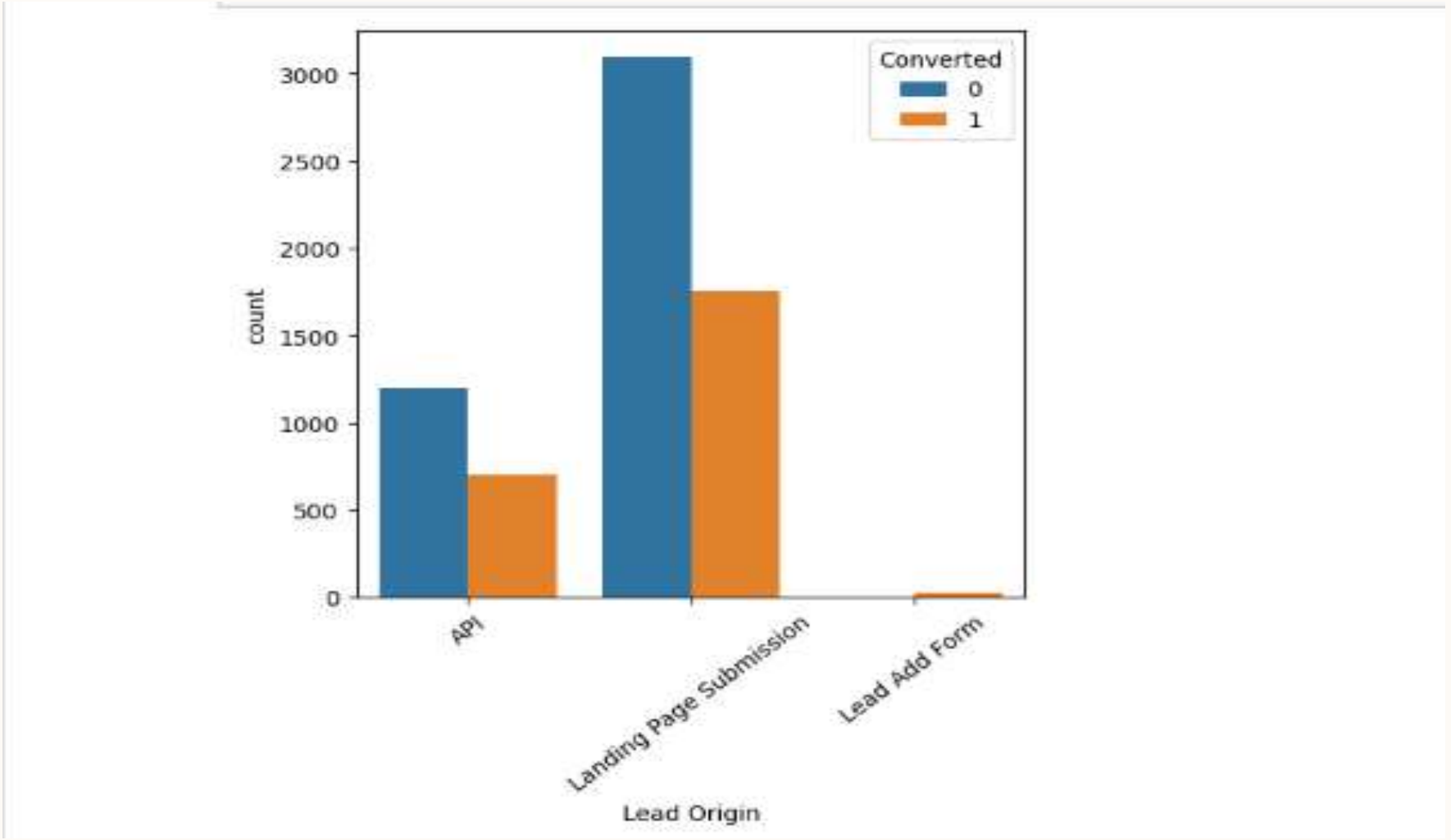
Among Tags, Will revert after reading the email' has the highest number of conversion



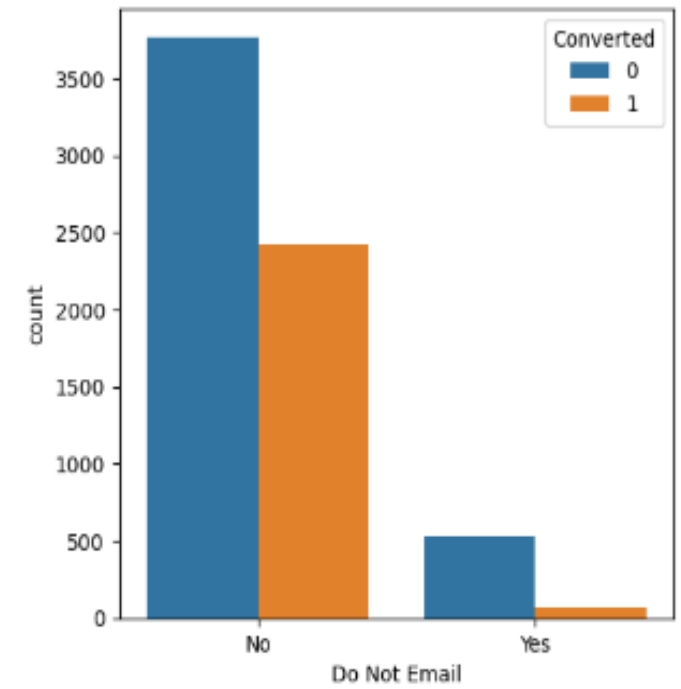
‘Google’ & ‘Direct Traffic’ bring the highest number of Leads, followed by ‘Olark Chat’

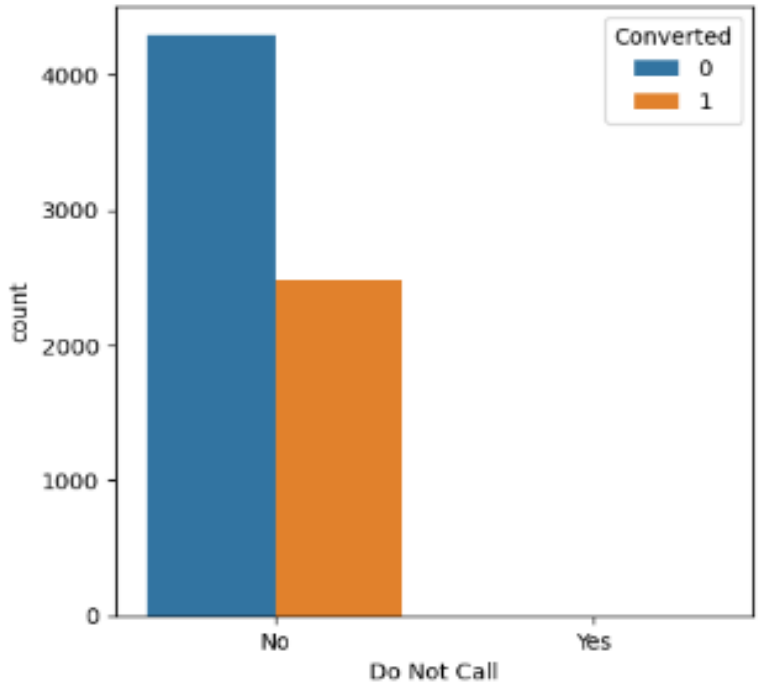


Considering Last Activity, ‘Email Opened’ brings the highest no. of leads but ‘SMS Sent’ has the highest conversion rate



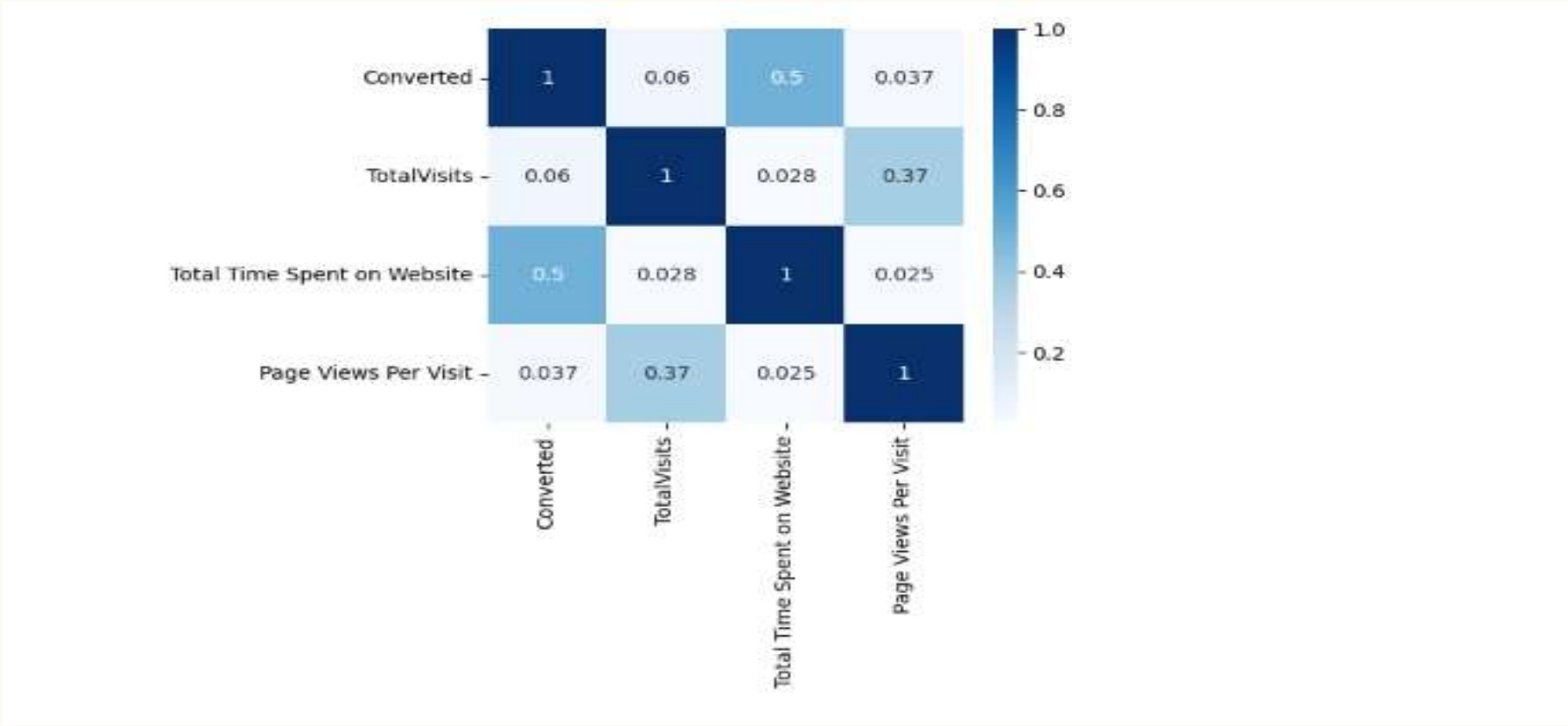
Conversion rate of leads from API and Landing Page submission seems to be lesser than the number of leads they land. This needs to improve.



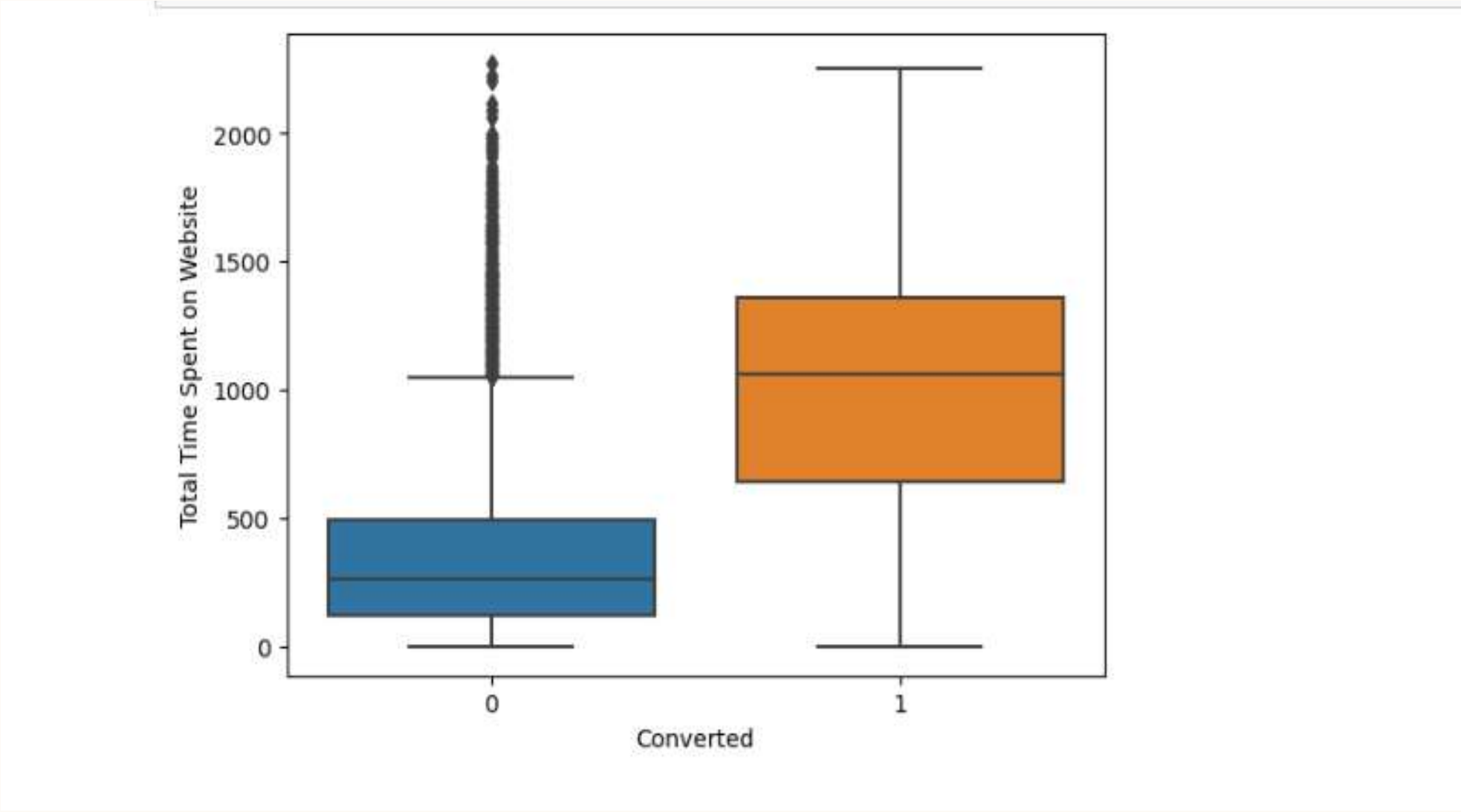


We dropped 'Do not call' column as more than 90% has only one value

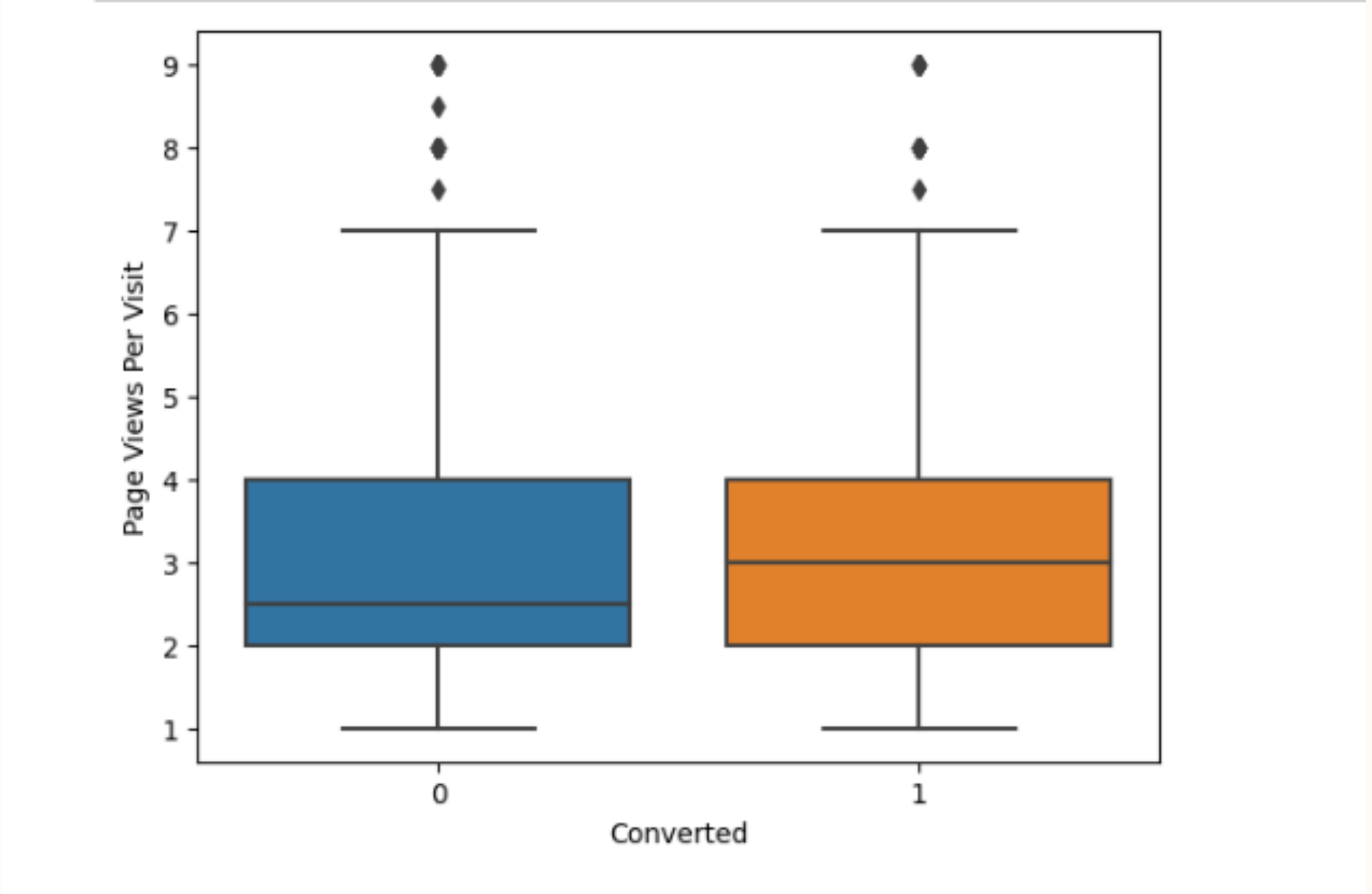
ANALYSIS OF NUMERICAL VARIABLES :



Heatmap shows that the highest correlation of Converted is with 'Total Time Spent on Website'

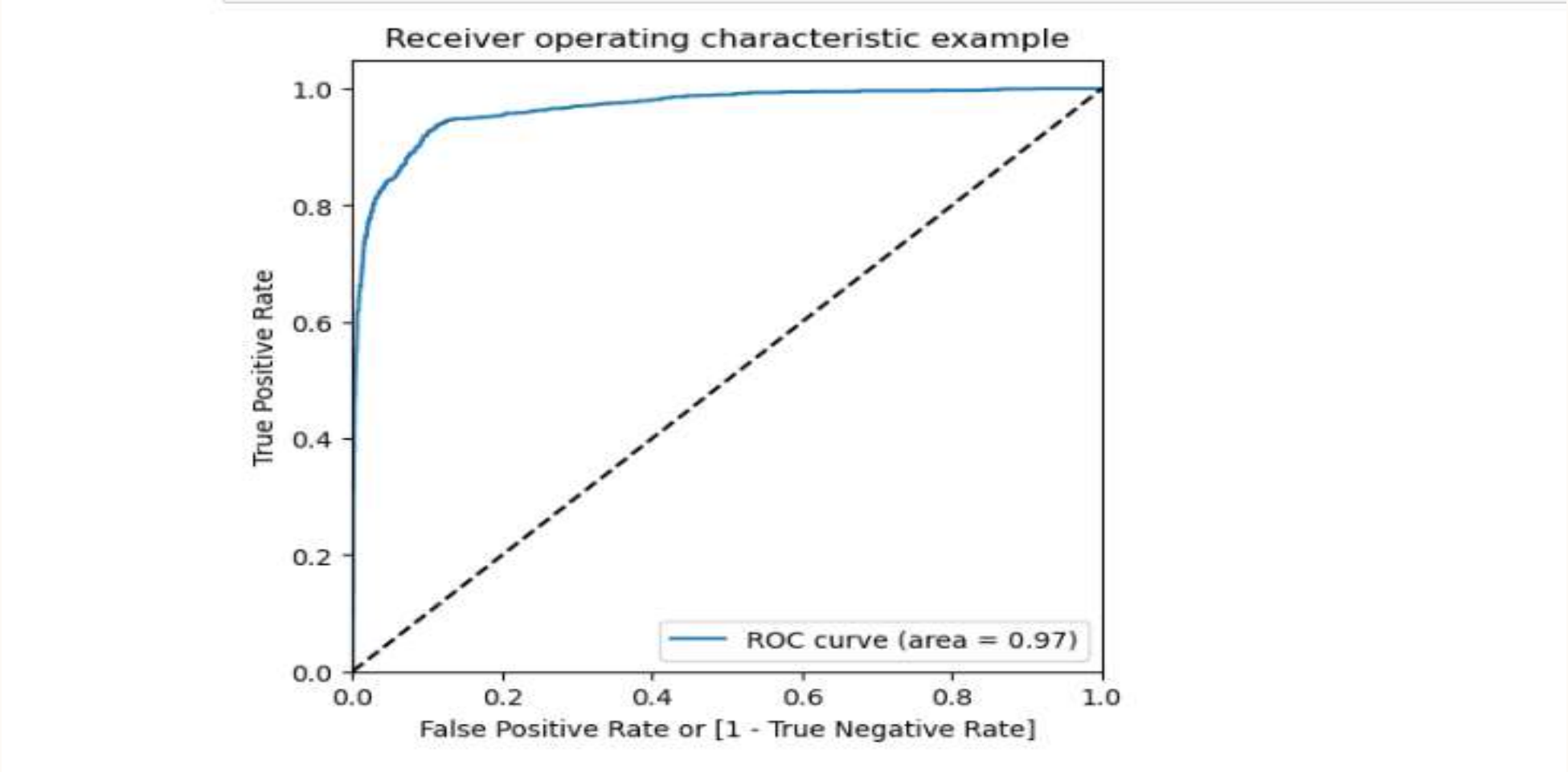


- The more time leads spend on the website, higher the chances of them being converted.
- Hence, the website needs to be more engaging for higher chances of conversion.



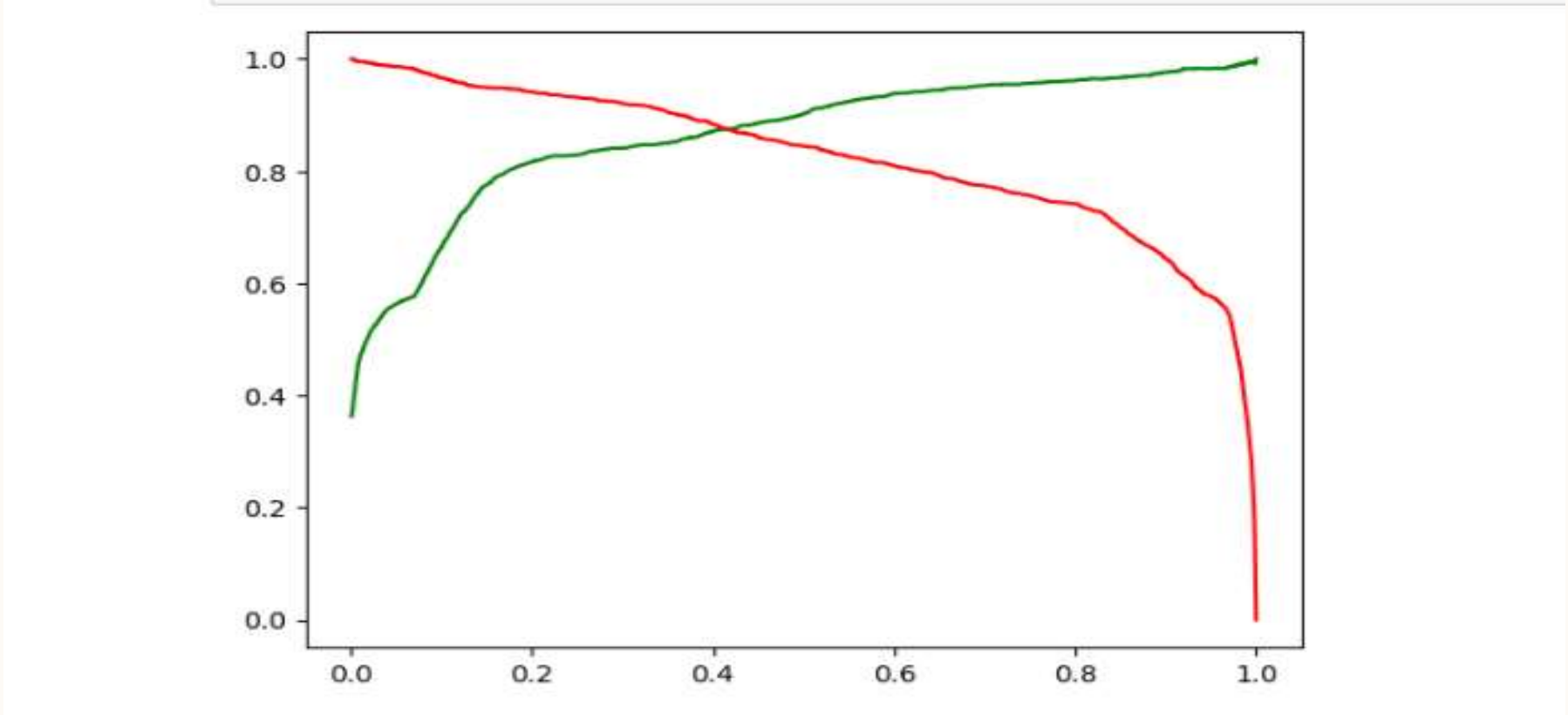
Nothing conclusive can be said for lead conversions from ‘Page Views Per Visit’

ROC CURVE:



This indicates a good predictive model with a value of 0.97. The ROC Curve should be a value close to 1.

ROC CURVE:



From the curve we infer that 0.3 is the optimum point to take as a cutoff probability

FINAL MODEL OBSERVATIONS:

Final Observations:

Comparison of the values obtained for Train & Test datasets:

Train Data:

Accuracy : 90.81%

Sensitivity : 92.05%

Specificity : 90.10%

Test Data:

Accuracy : 90.92%

Sensitivity : 91.41%

Specificity : 90.62%

RESULT:

The Model seems to predict the Conversion rate excellently.

RECOMMENDATIONS

RECOMMENDATION

X Education Company needs to focus on following key aspects to improve the overall conversion rate:

- Increase user engagement on their website
- Increase on sending SMS notifications and email with the details of the course
- Get Total visits increased by advertising etc
- Improve the Olark Chat service as this is affecting the conversion negatively
- Focus more on the Management specialisations which has high potential to convert the leads