Lead Score – Case Study

Prepared by :-

Mukesh Chaudhary,

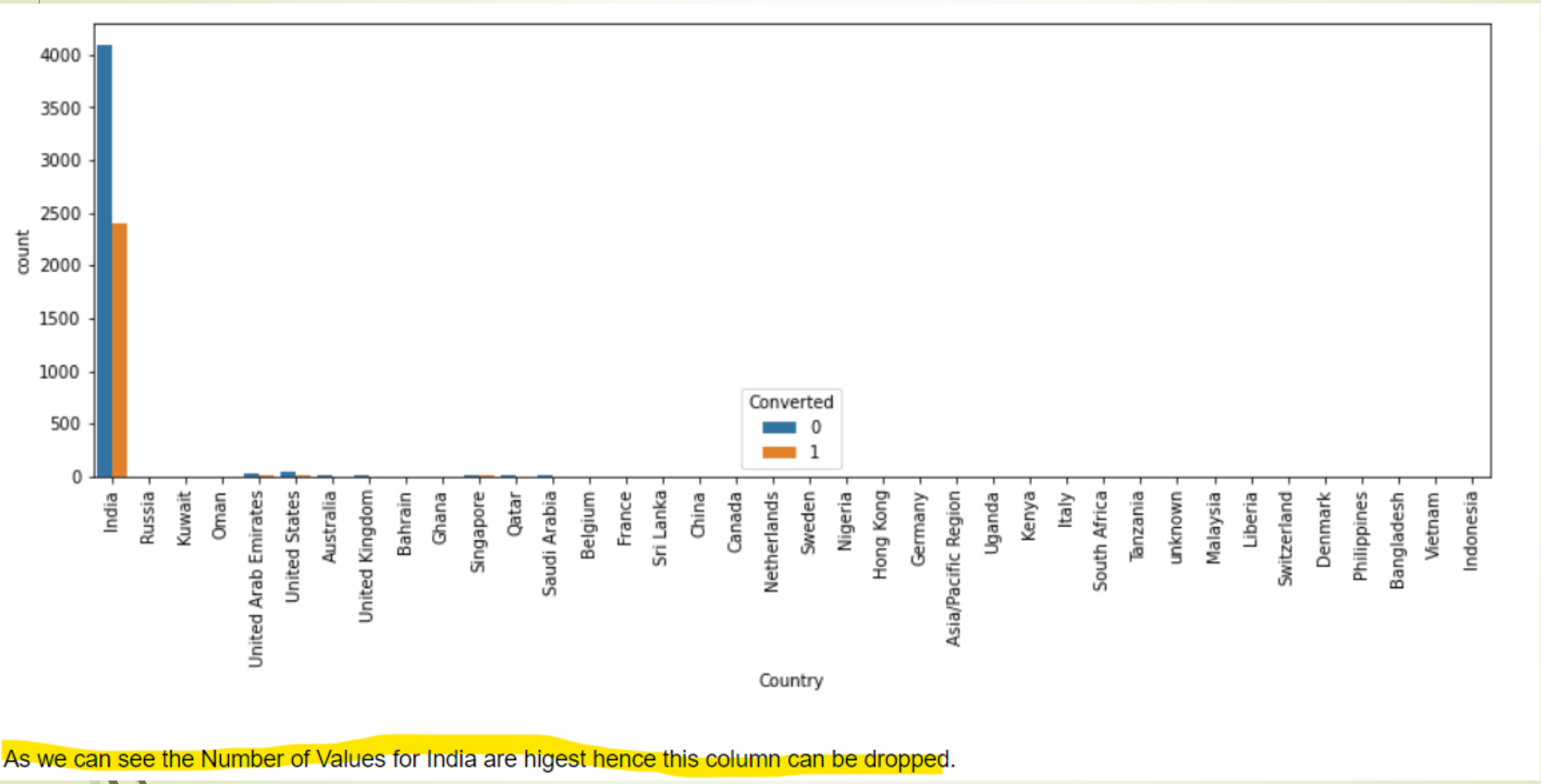
Mohammed Muddasir,

Madhu Sudhan

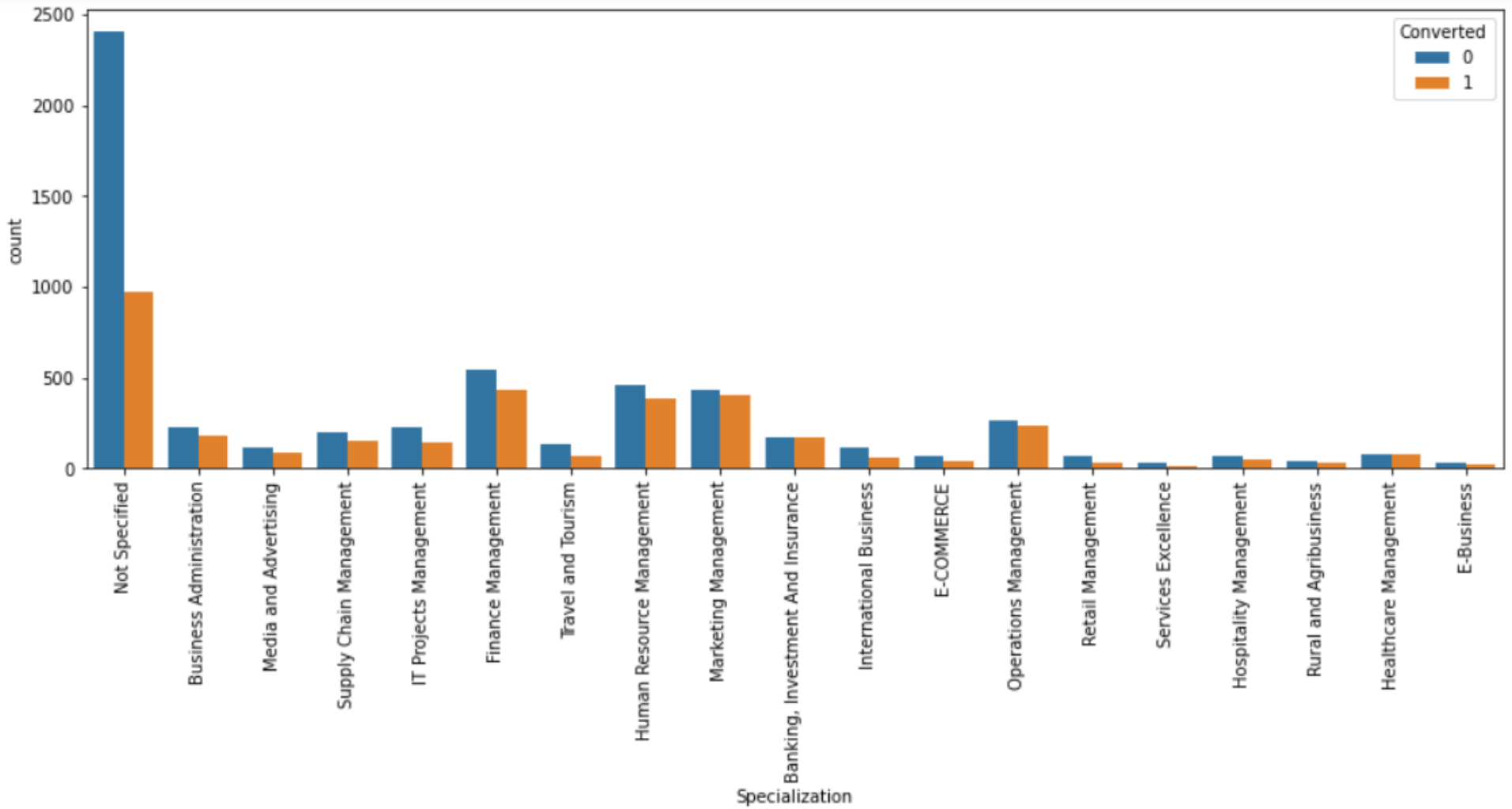
Problem Statement

- An education company named 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 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%
- Now, although 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'
- X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you 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%.**

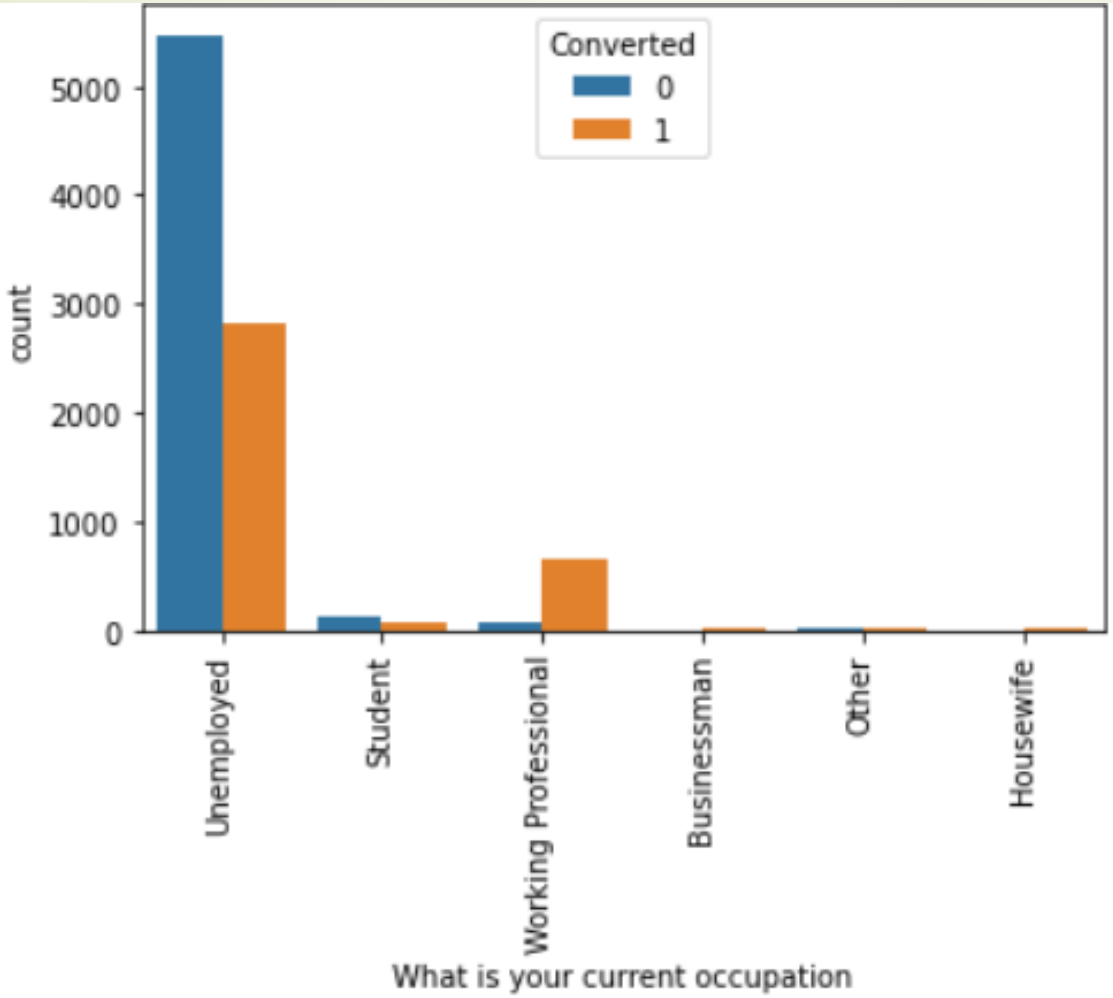
Exploratory Data Analysis



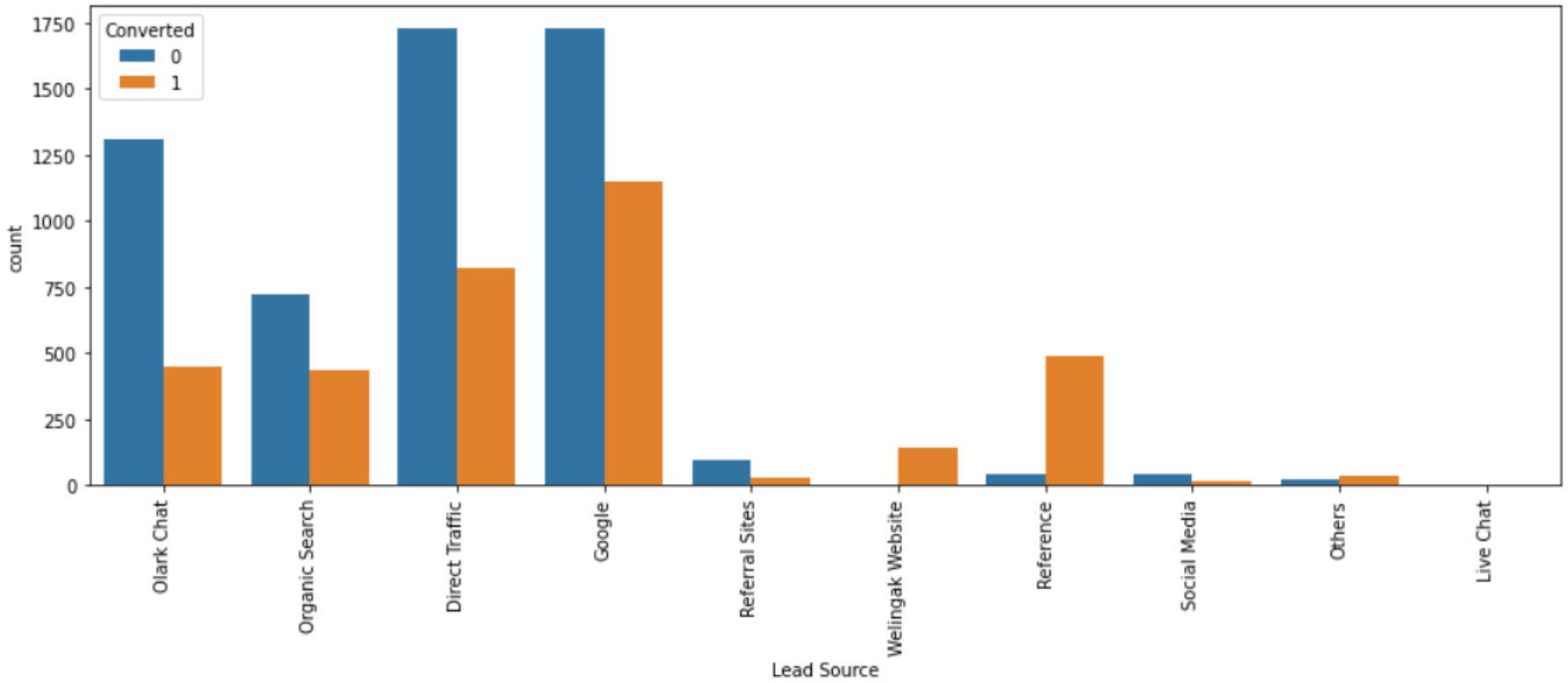
As we can see the Number of Values for India are higest hence this column can be dropped.



From above plot We have observed that specialization with Management have higher number of leads as well as leads converted. So this is a significant variable and should not be dropped.

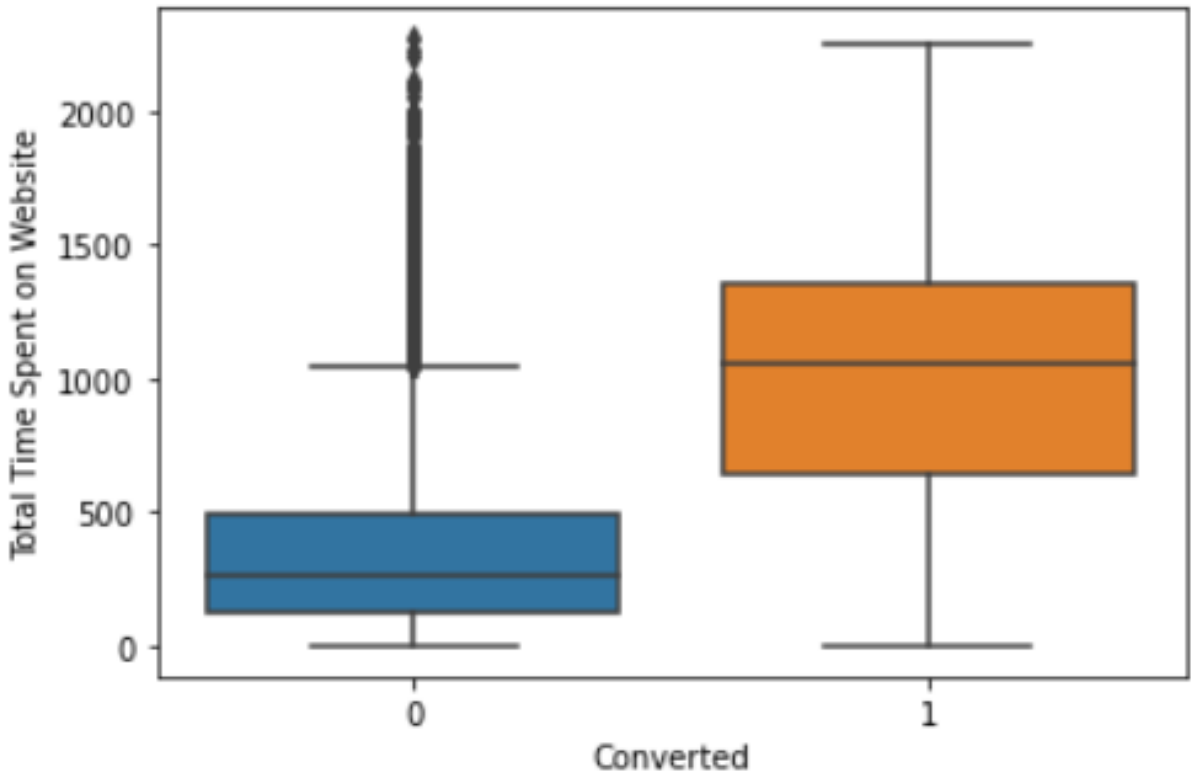


1. Working Professionals having high chances of joining the courses Whereas
2. Unemployed leads are the most in terms of Absolute numbers.



###Observartions :-

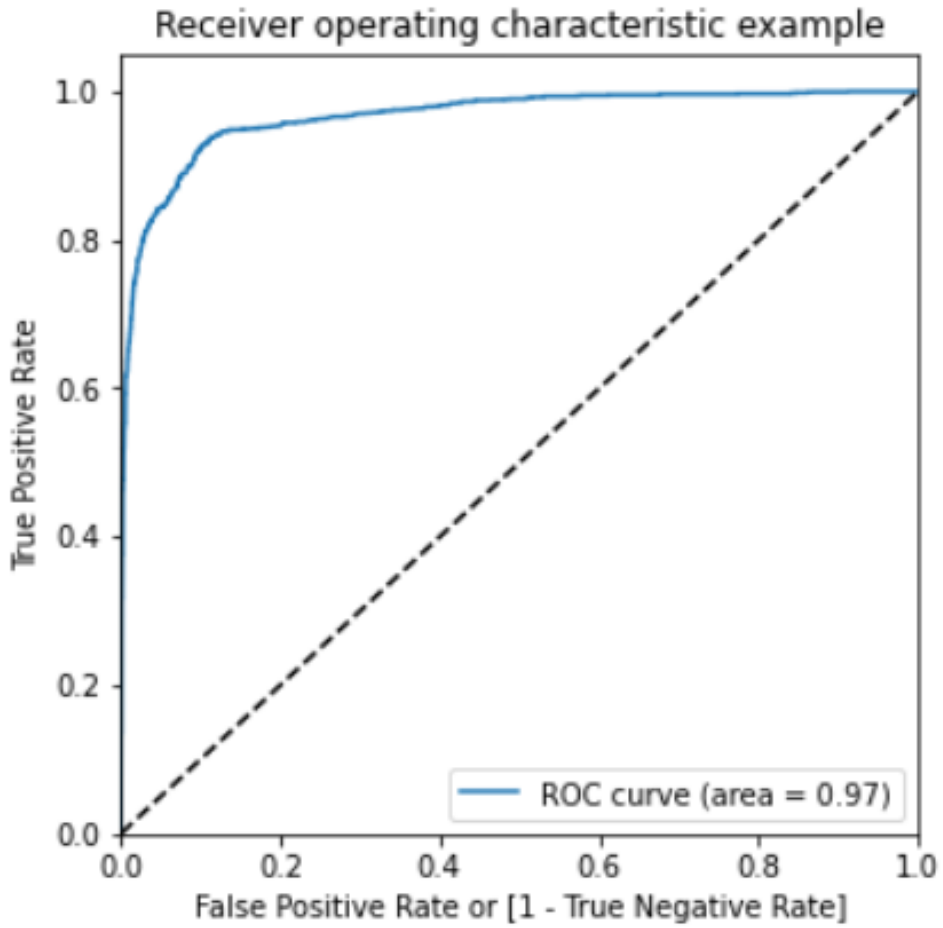
1. Top focus areas olark chat, organic search, direct traffic, and google leads
2. Maximum number of leads are generated by Direct traffi and Google.
3. Higher conversion Rate of reference leads and leads through welingak.



Inference

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.

Model Evaluation



The ROC Curve should be close to 1. We have got 0.97 indicating a good predictive model.

We have following Observations:

- We have 0.97 ROC value
- Accuracy : 90.81%
- Sensitivity : 92.05%
- Specificity : 90.10%

>>>>>We have follwing final observations which can be presented to CEO

Parameters	Test Data	Train Data	Delta (Test - Train)
Accuracy	90.92%	90.81%	0.11%
Sensitivity	91.41%	92.05%	-0.64%
Specificity	90.62%	90.10%	0.52%