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. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.
- 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%.

GOALS OF THE CASE STUDY

- 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

S T E P S

1.Data Preparation

2.Data Cleaning

3.Univariate Analysis

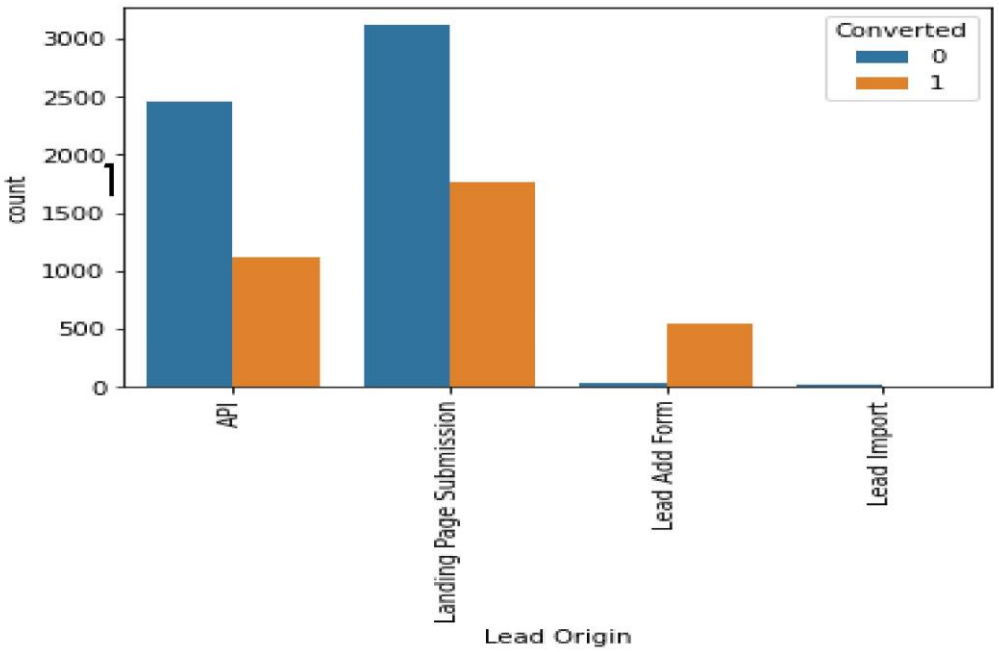
4.Data Splitted to Train and Test data

5.Feature Scaling

6.Model Building

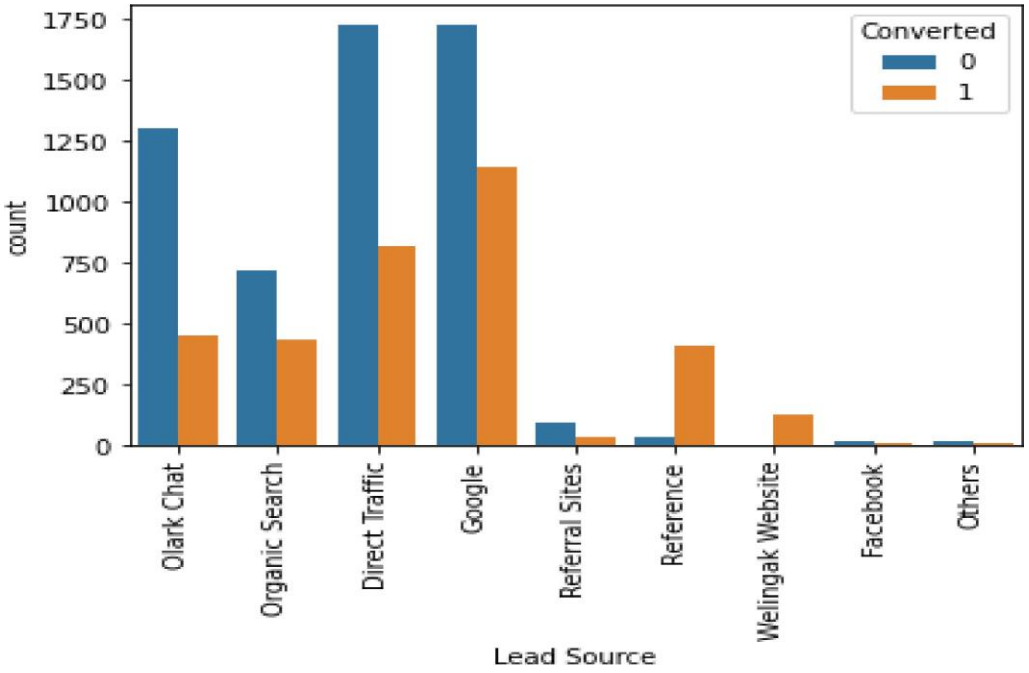
UNIVARIATE ANALYSIS

1.Lead Origin



- API and Landing Page Submission has around 30 to 35 % conversion rate and Lead originated from here are considerable.
- Land Add Form has 90% conversion rate but Lead Add form counts are very less
- Lead Import are very less in count

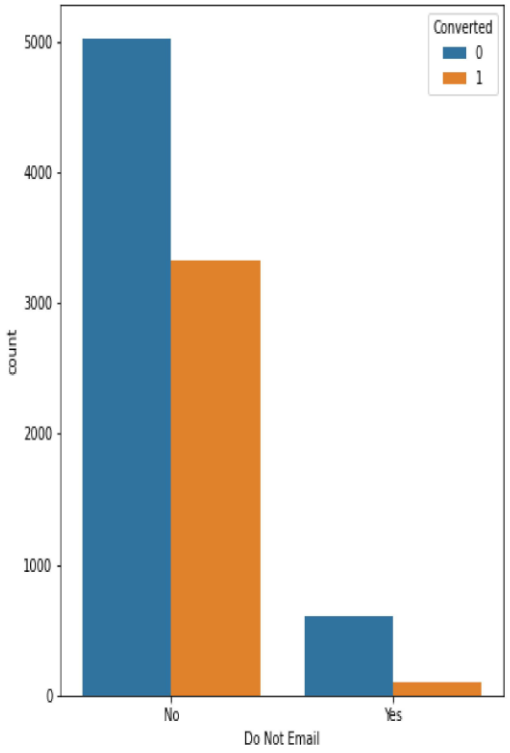
2.Lead Source

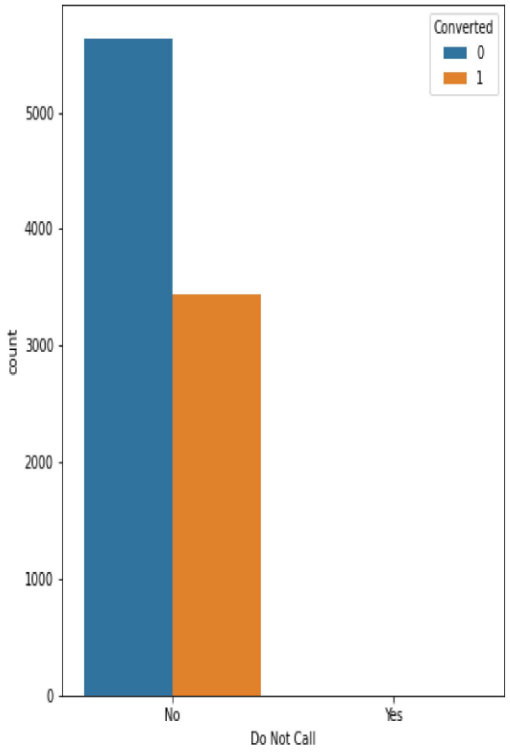


1. Most of the leads have their last activity as "Email opened"

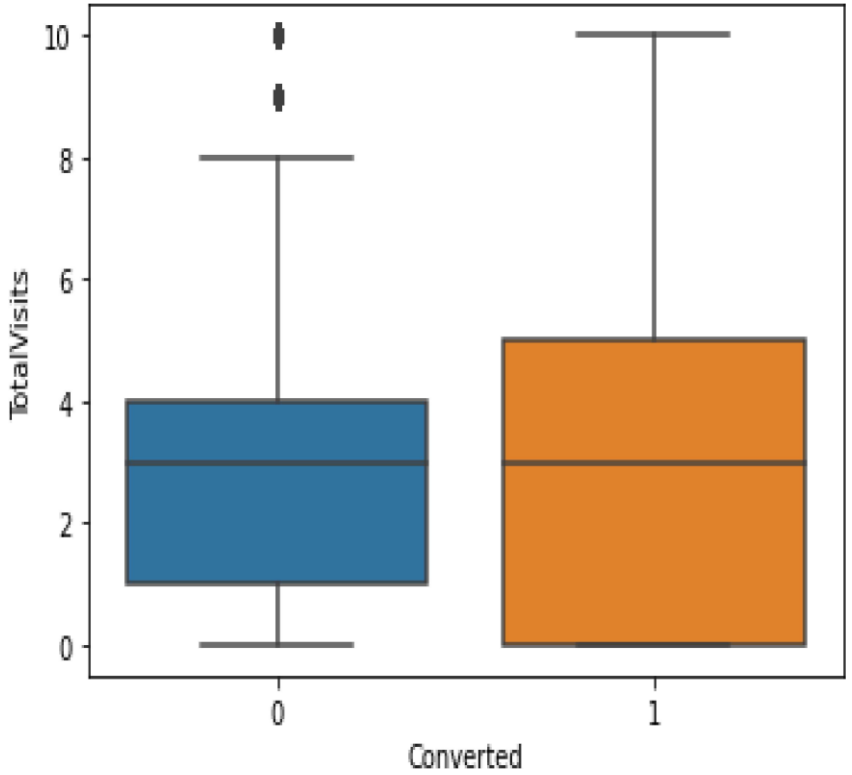
2.Conversion rate of Leads with Last Activity as SMS Sent is high

3.Do Not Email and Do Not Call

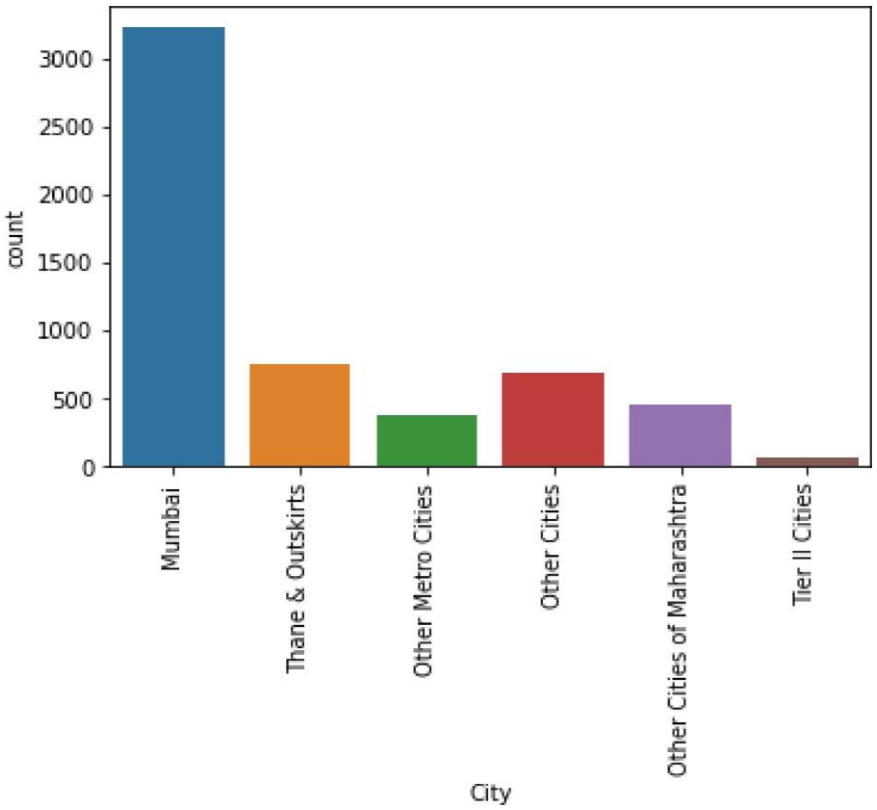




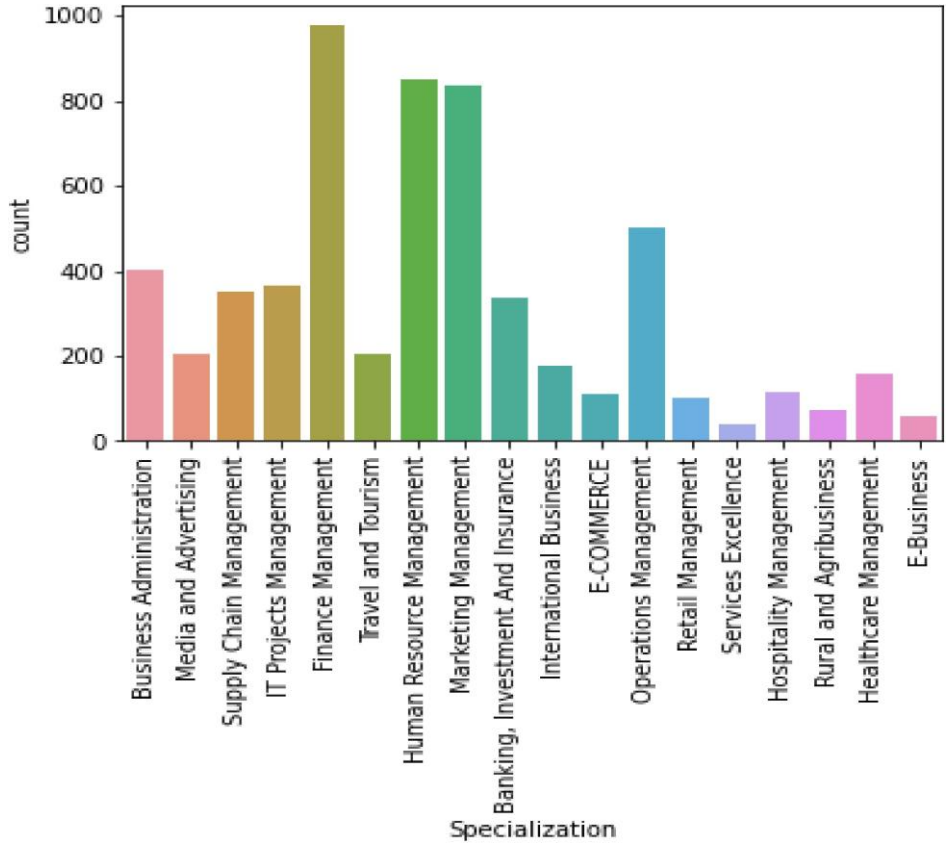
4. Total Visit



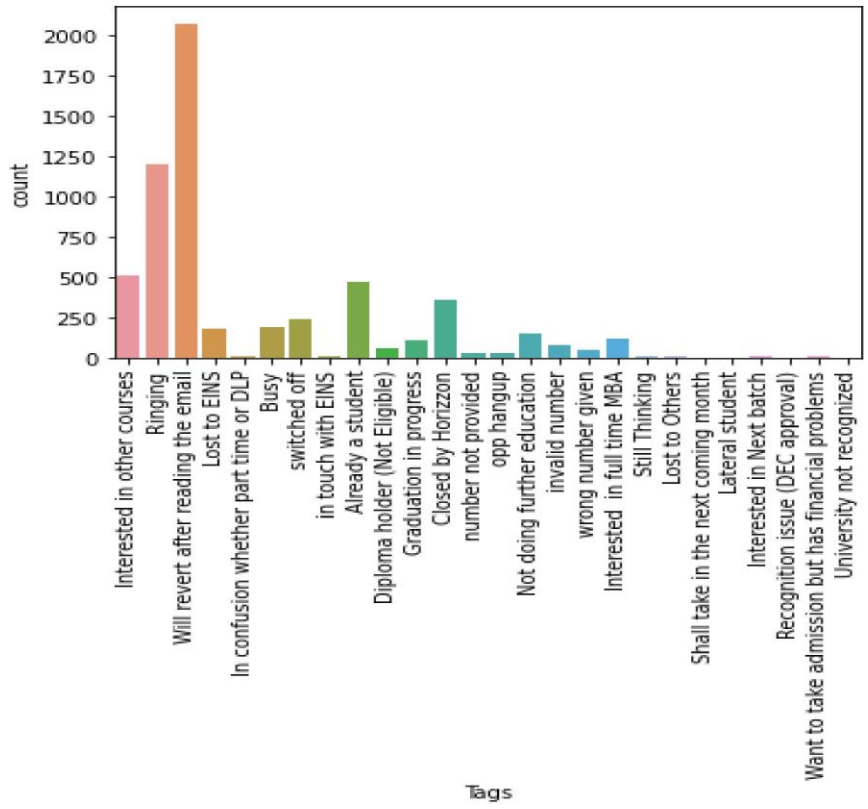
5. City



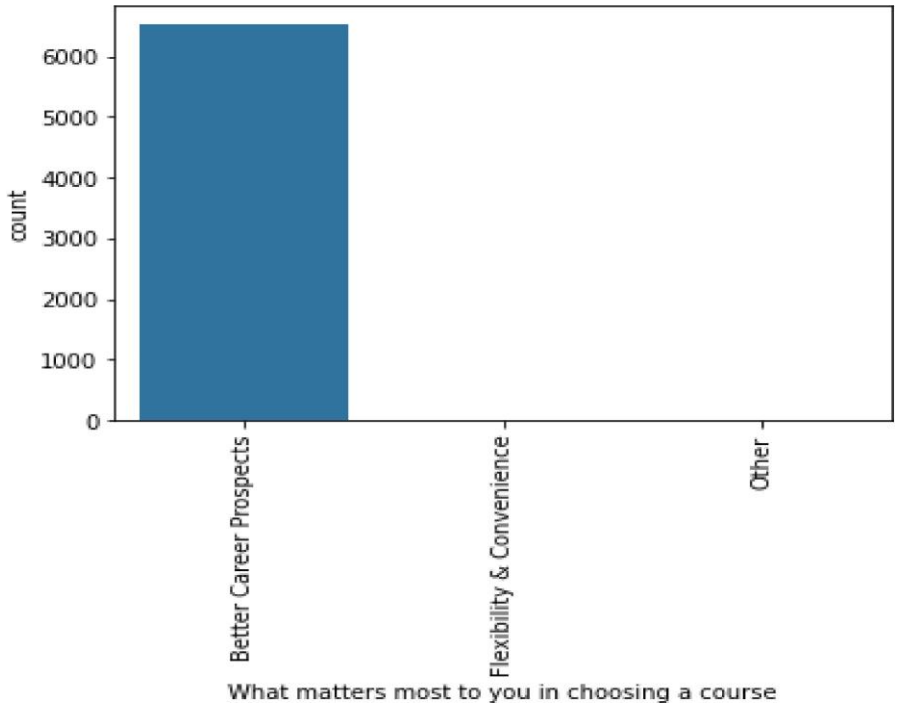
6.Specialization



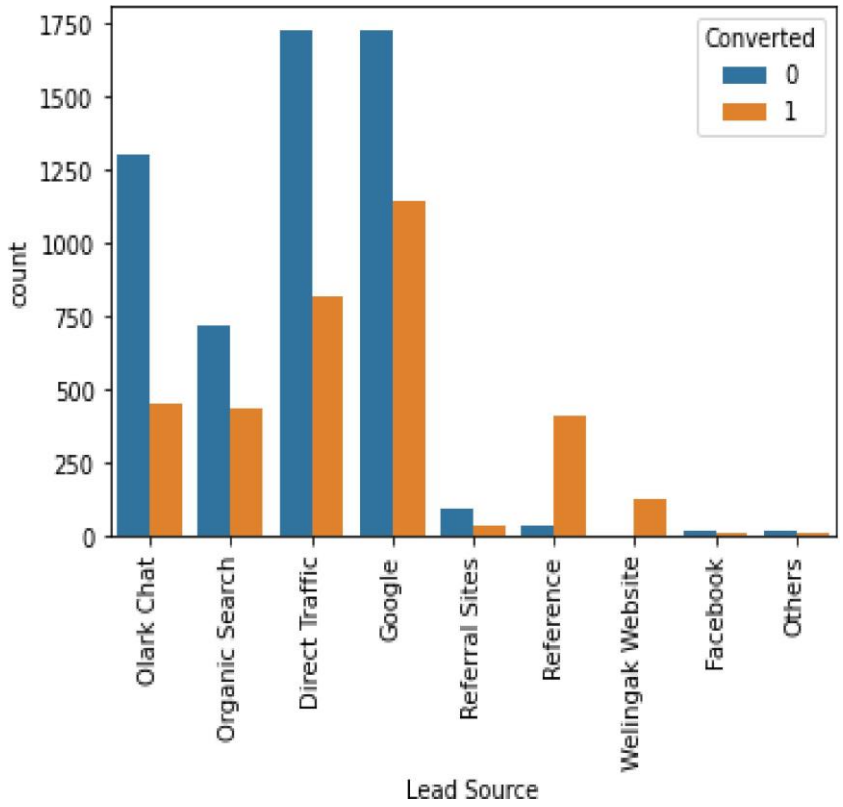
7. Tags



8.What matters most to you in choosing a course



9.Lead Source

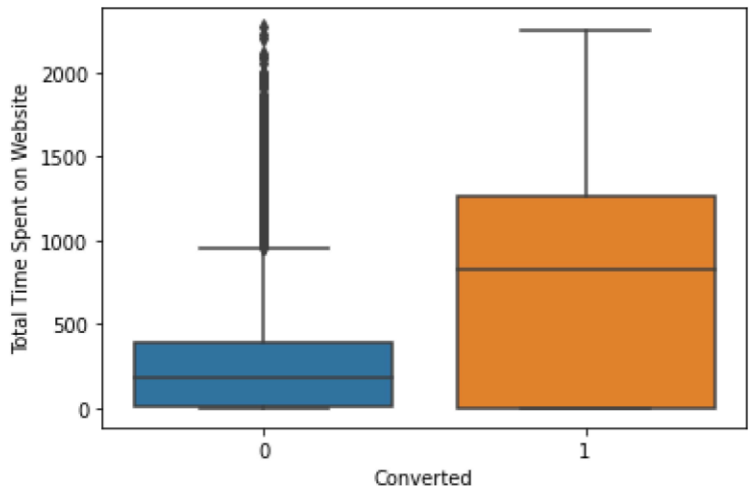


1.Lead Conversion from Google and Direct Traffics are quite good

2.Reference Sites and Welingak Website's conversion rate are very high

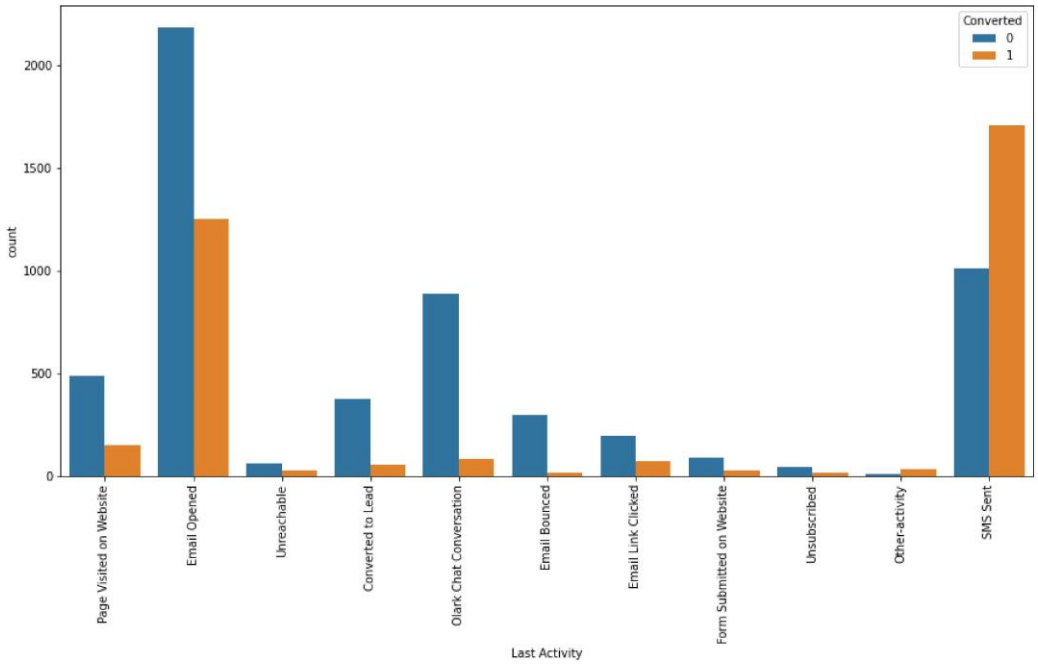
To improve overall conversion rate should focus on Clark Chat , Organic Search ,Direct Traffic , Google leads and generate more leads from Reference Sites and Welingak Website

10.Total Time Spent on Website



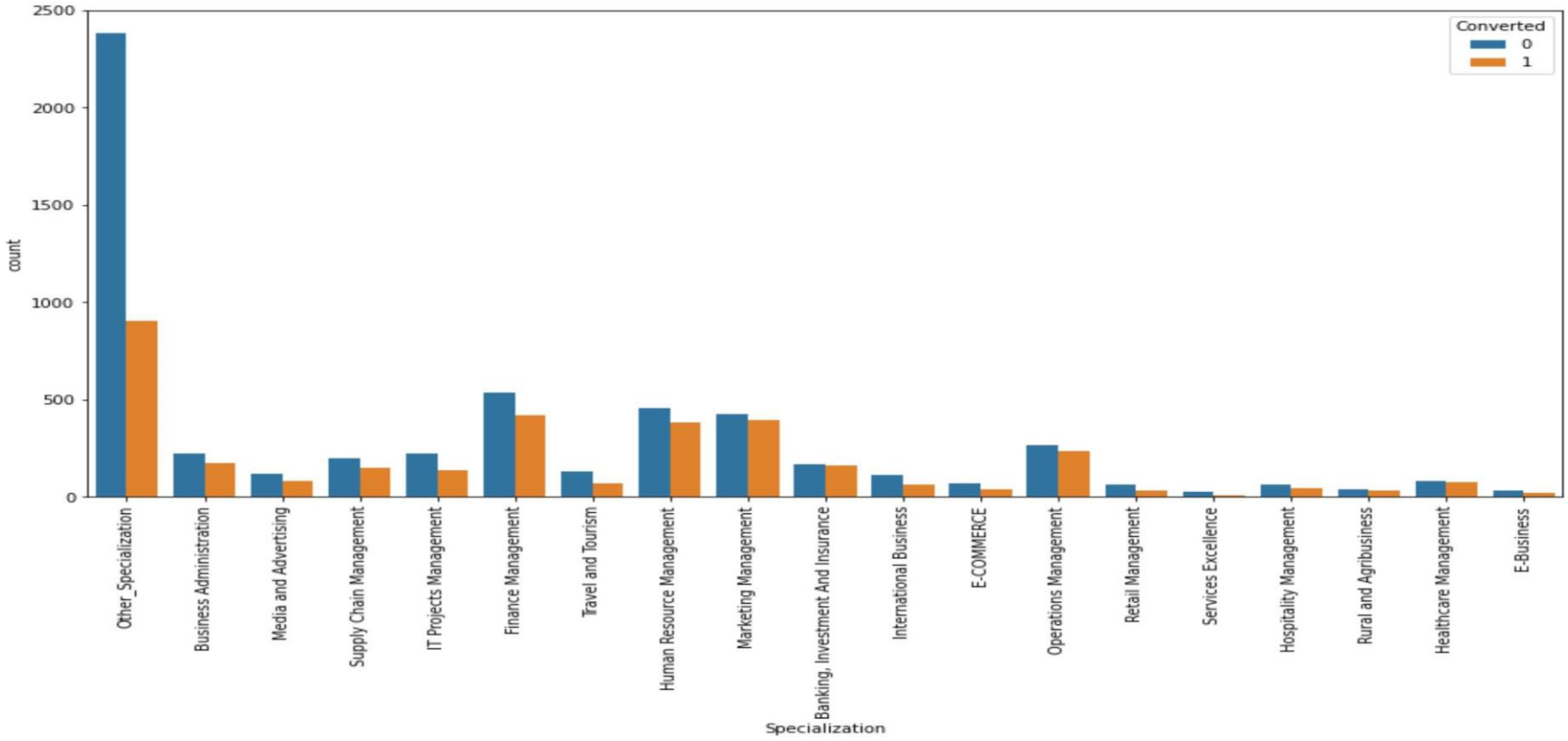
Lead Spending more time on Website are more likely to get converted

11.Last Activity

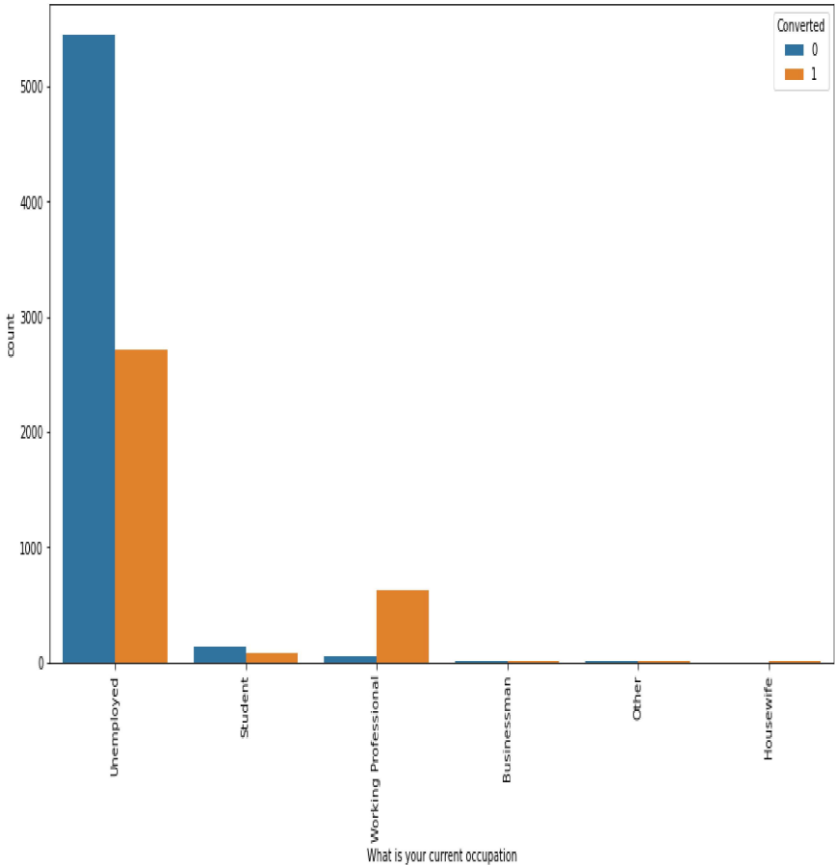


Conversion rate of Leads with Last Activity as SMS Sent is high

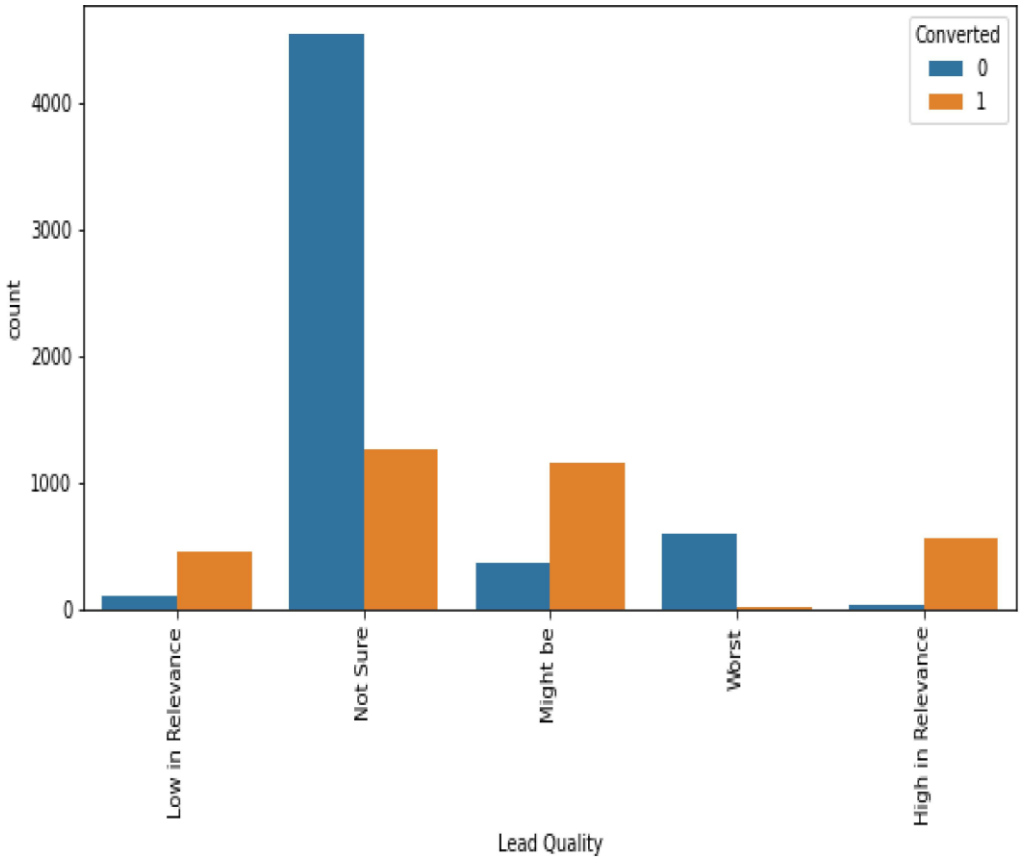
12.Specialization



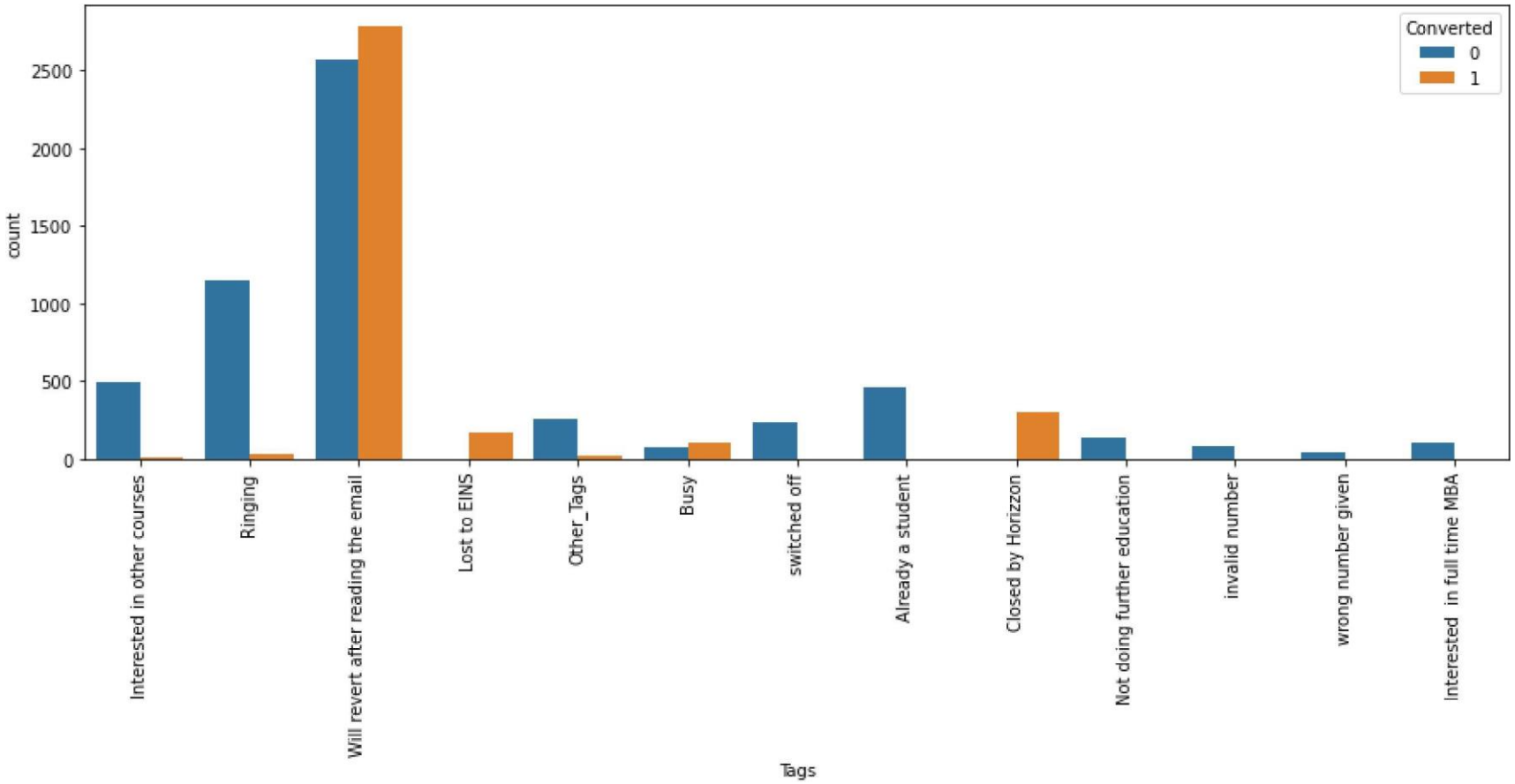
13.What is your current occupation



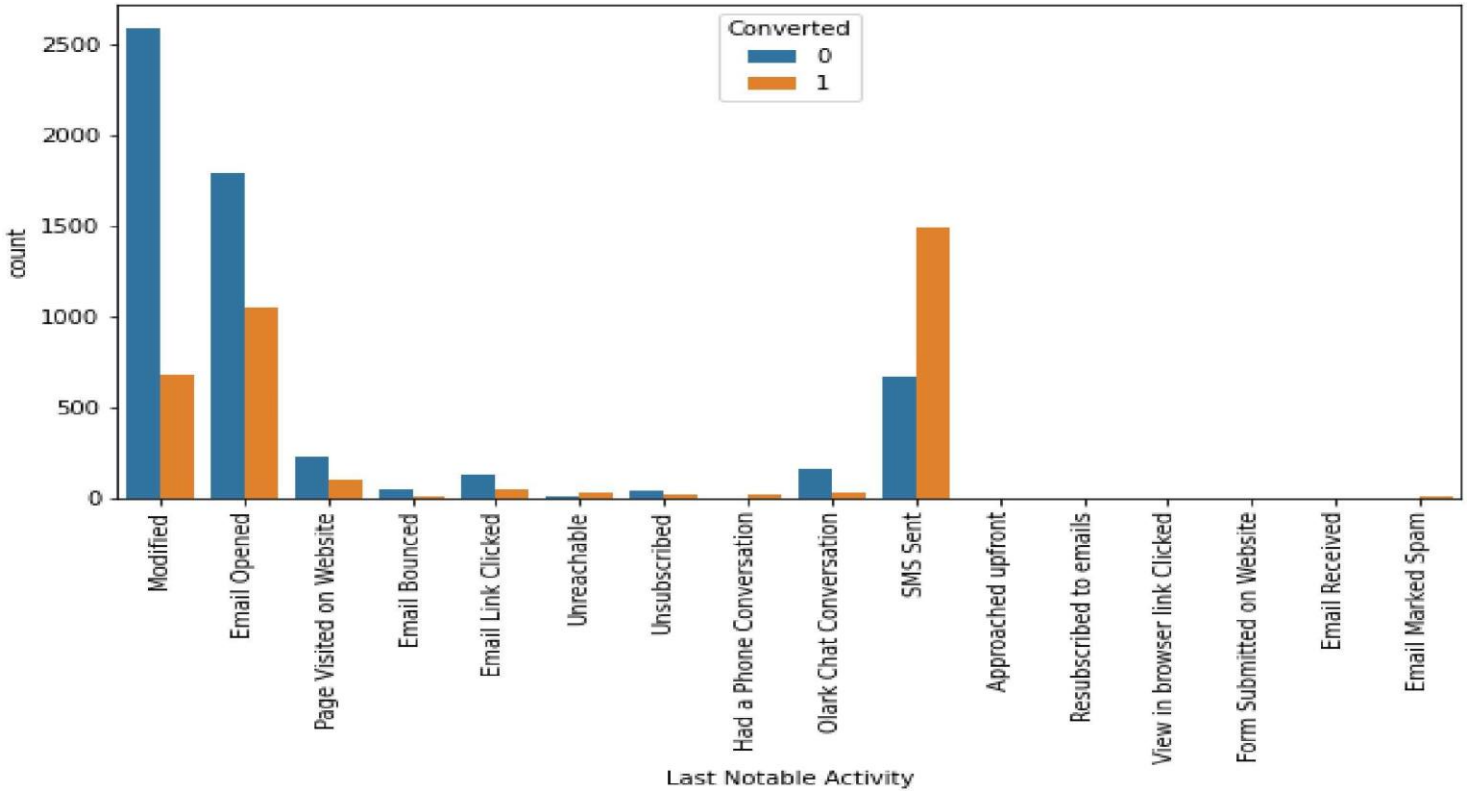
14.Lead Quality



15.Tags


Converted
0
1
count
2500
2000
1500
1000
500
0
Interested in other courses
Ringing
Will revert after reading the email
Lost to EINS
Other_Tags
Busy
switched off
Already a student
Closed by Horizzon
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Tags

16.Last Notable Activity



Train Data:

- Accuracy=91.93%
- Sensitiviy=85.15%
- Specificity=96.18%

Test Data :

- Accuracy=90.45%
- Sensitivity=84.42%
- Specificity=93.88%

Valuable Insights:

- The sales team of X-Education should focus on leads having 1.Lead_Origin:Lead Add Form 2.Occupation : Working Profession 3.Lead Source:Welingak Websit
- 2.Lead Spending more time on Website are more likely to get converted
- 3.The Sensitivity , Specificity and accuracy obtained in test set are almost accurate.
- 4.High Sensitivity ensures that almost all the leads who are likely to convert are correctly predicted.
- High specificity will ensure that leads that are on the brink of probability of getting converted or not are not selected.
- It is better not focus more on customer who do not want to be called about the course.
- 7.If last notable action is modified , he/she may not be a potential lead