Lead scoring case study

Team:

- Likhitha C
- Wahid
- Venkata Koushik Akella

Problem statement:

- X Education is an organization which provided online courses for industry professional the company marks its courses on several popular websites like google
- X Education company needs help to identify the most promising leads.
- This has to be done by building a Logistic regression model that can be trained with the Leads data of the past and then deployed to predict whether a particular lead will lead to payment or not. Each lead has to be given a lead score between 0 and 100 where less score means not a promising lead and vice-versa. the CEO has also given a ballpark target lead conversion rate to be 80%.

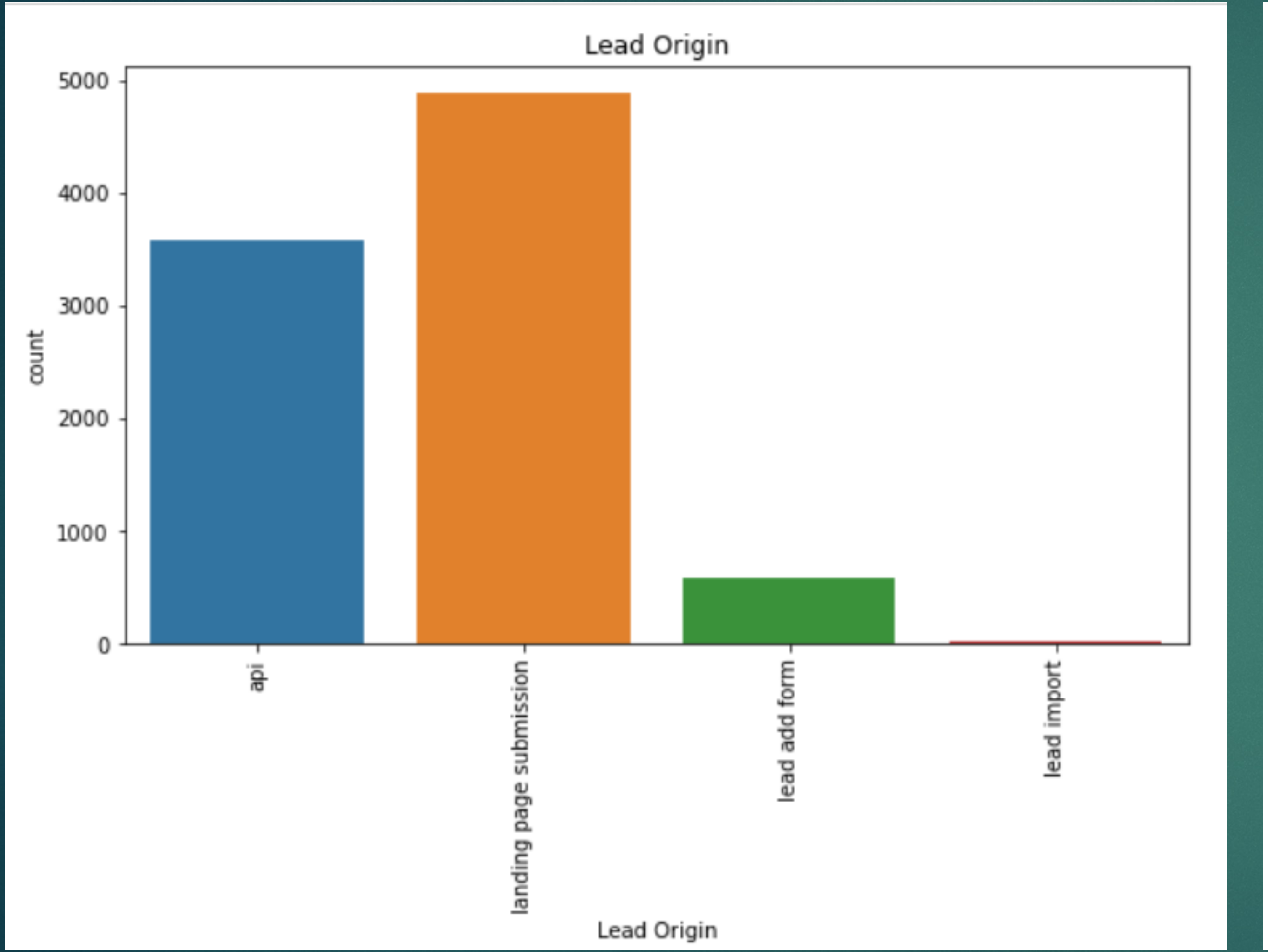
Business goal:

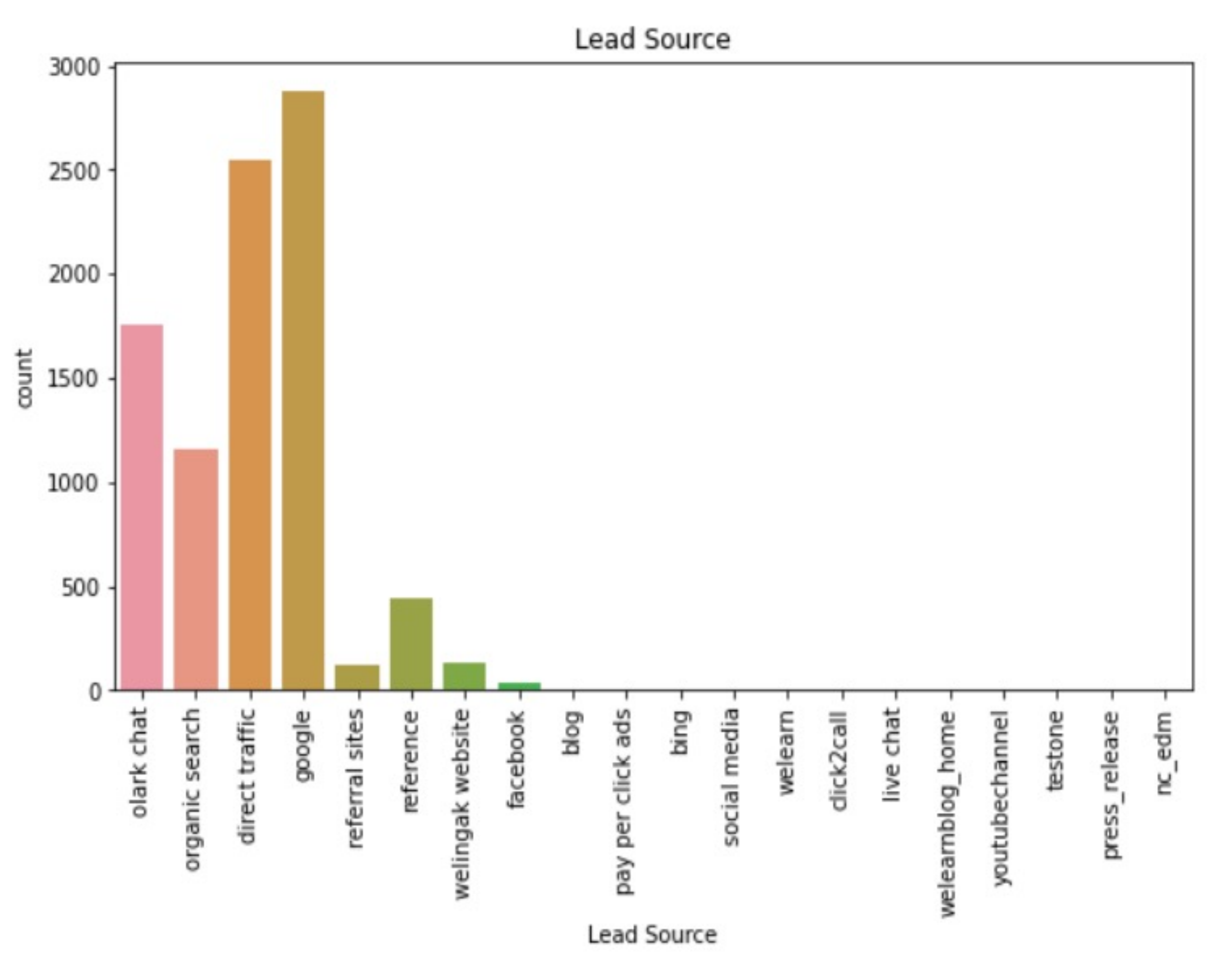
- The company requires a model to be built for selecting most promising leads
- build a logistic regression model to assign a lead score between 0 and 100 to each lead and based on the lead, classify them as hot leads or not so hot leads.
- The model is expected to fit certain other problems put forth by the X Education company in the form of a separate word document. fill the solutions to those problems based on the model we get in the end of this analysis.

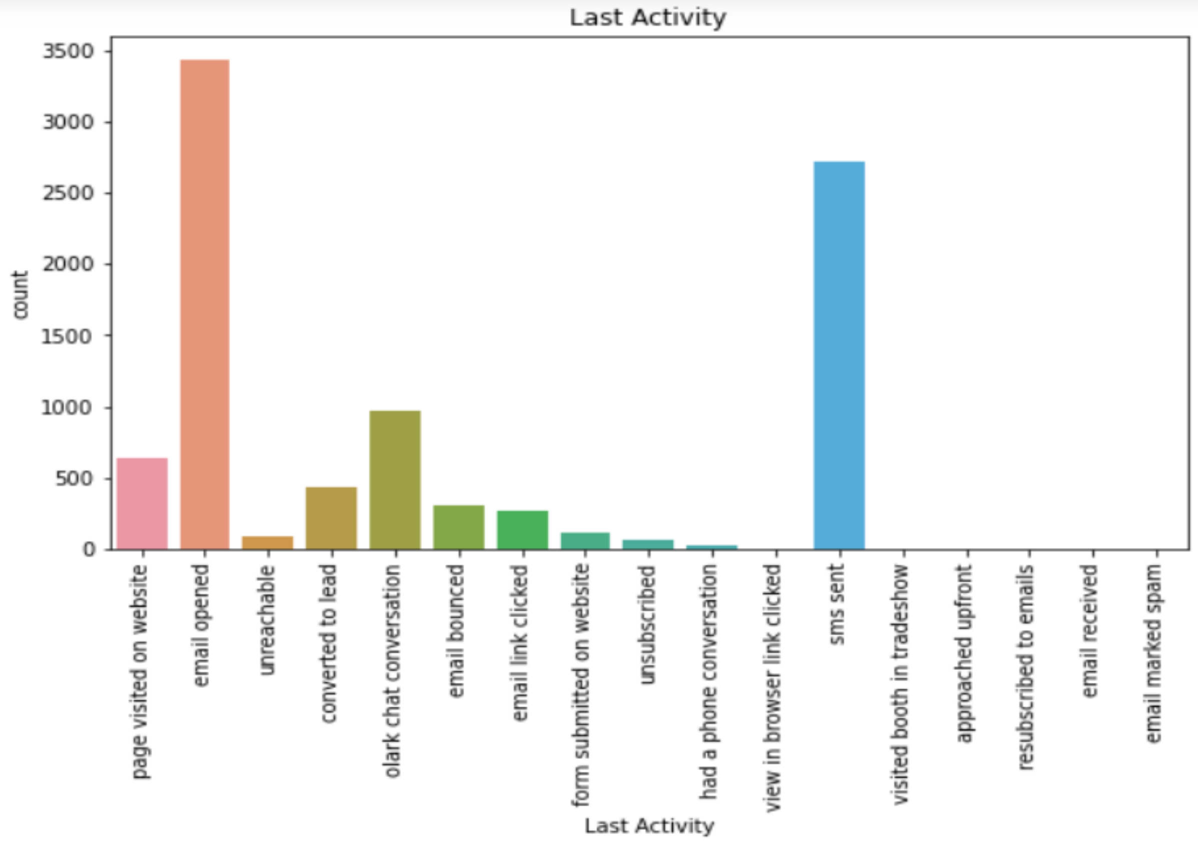
Strategy:

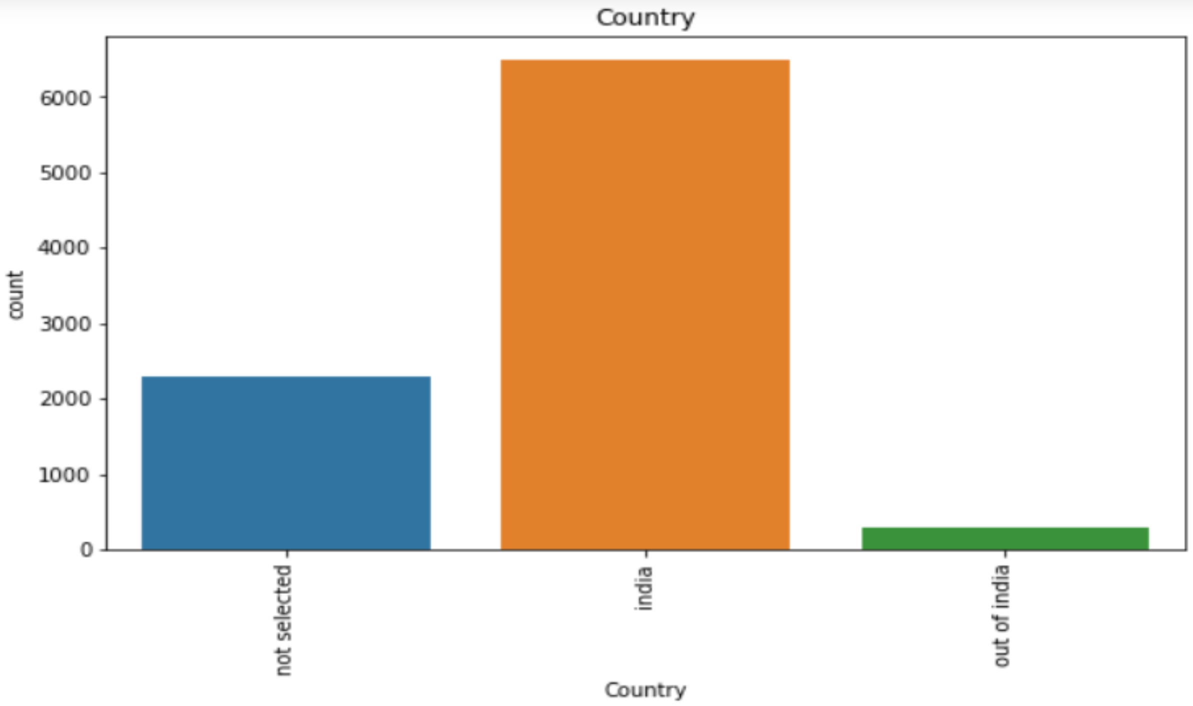
- Importing data
- Prepare the data for model budling
- Build a logistical regression model
- Assign a lead score for each lead
- Test the model on train set
- Evaluate model by different measure and metrics
- Test the model on test set

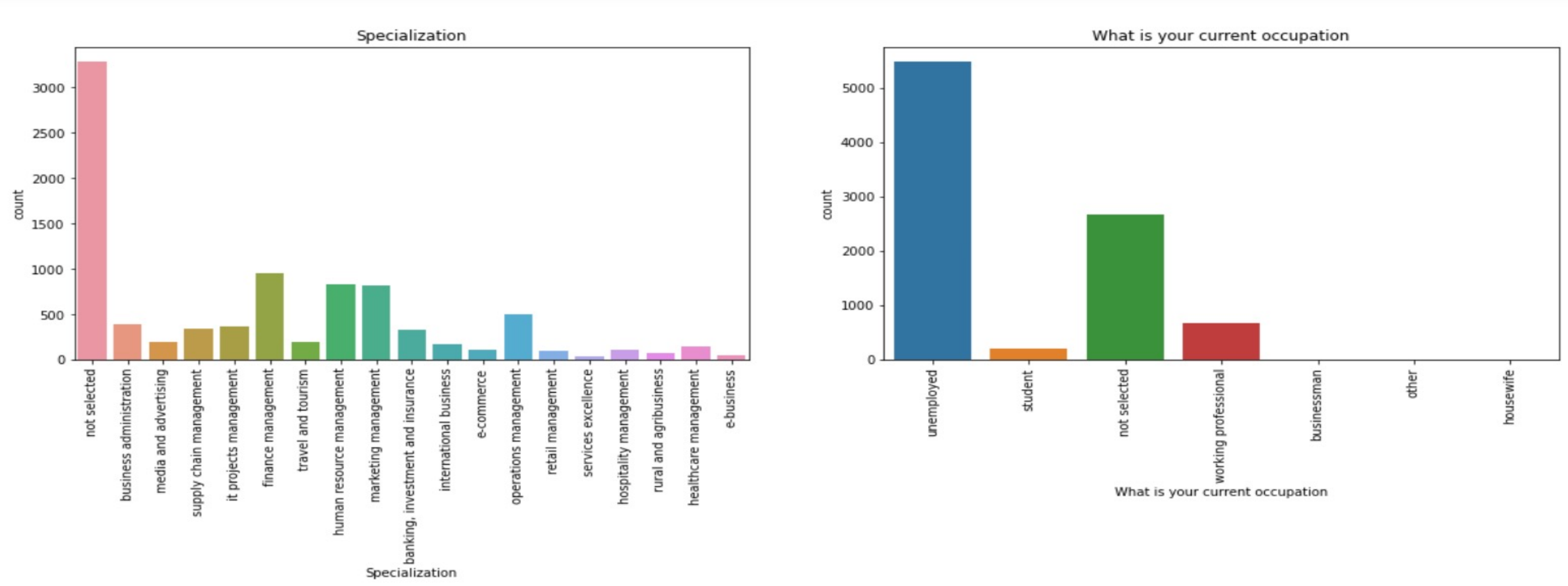
Exploratory data analysis: Categorical Variables

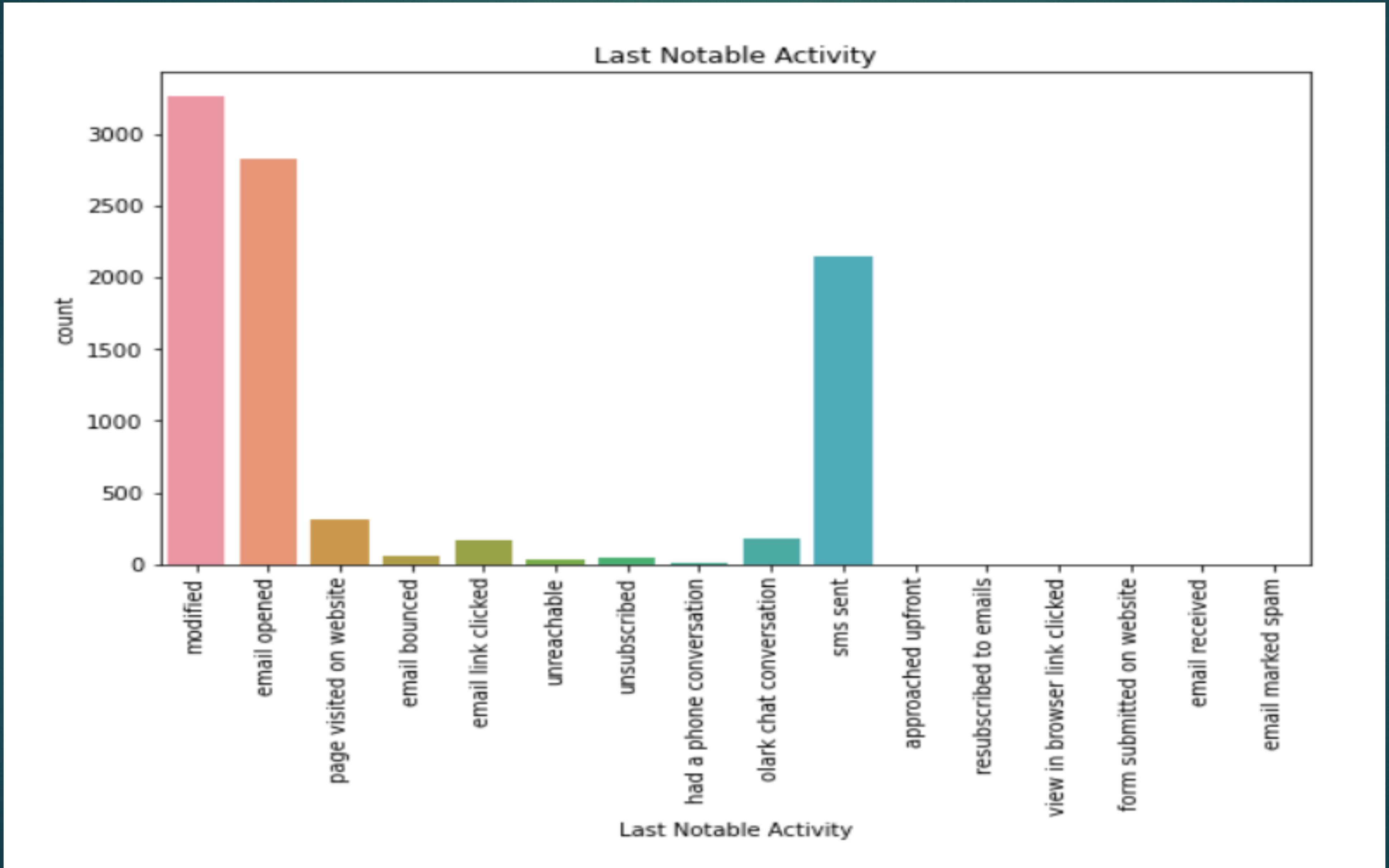




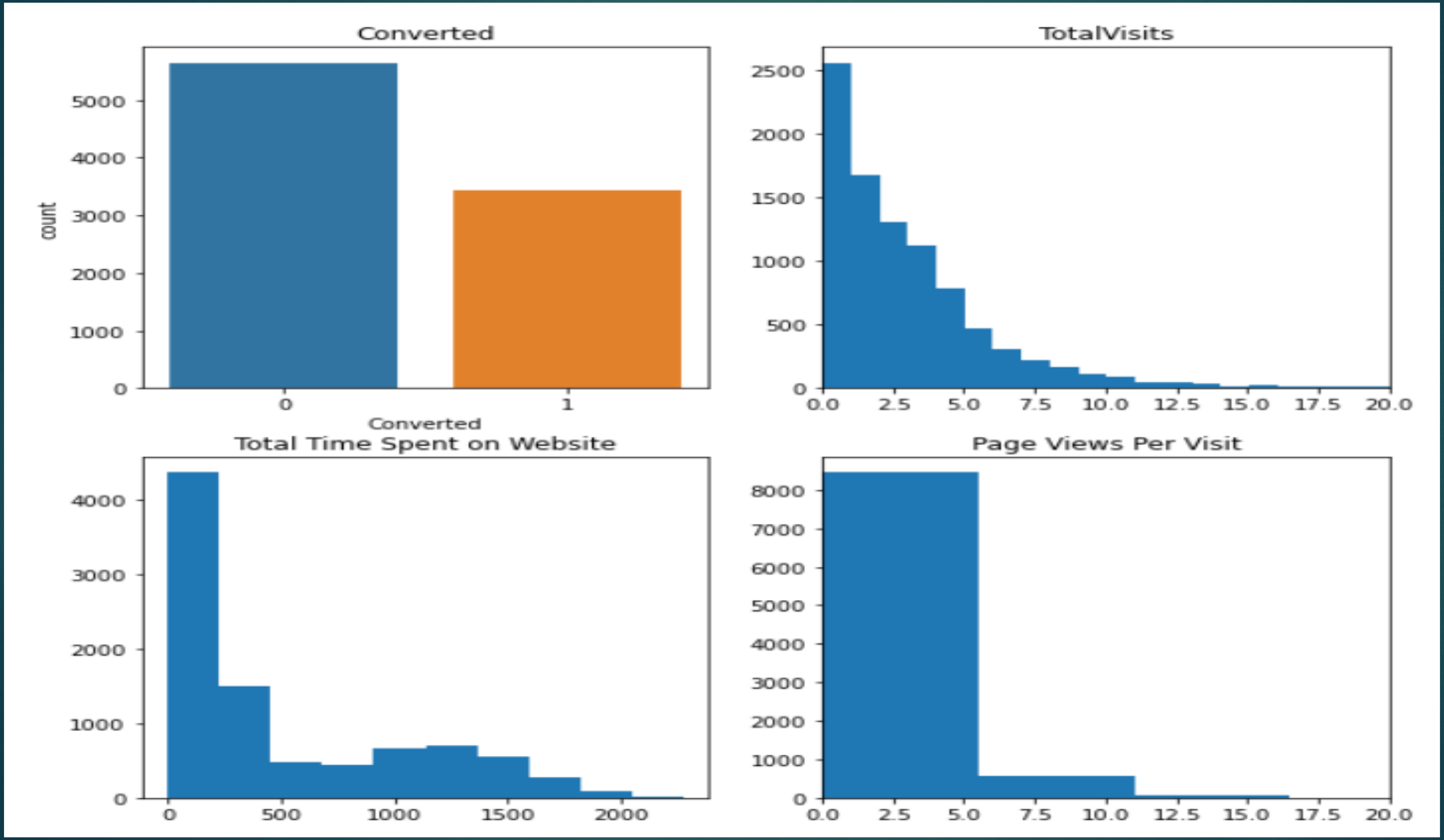

Last Activity
count
0
500
1000
1500
2000
2500
3000
3500
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow
approached upfront
resubscribed to emails
email received
email marked spam
Last Activity


Country
count
0
1000
2000
3000
4000
5000
6000
not selected
india
out of india
Country

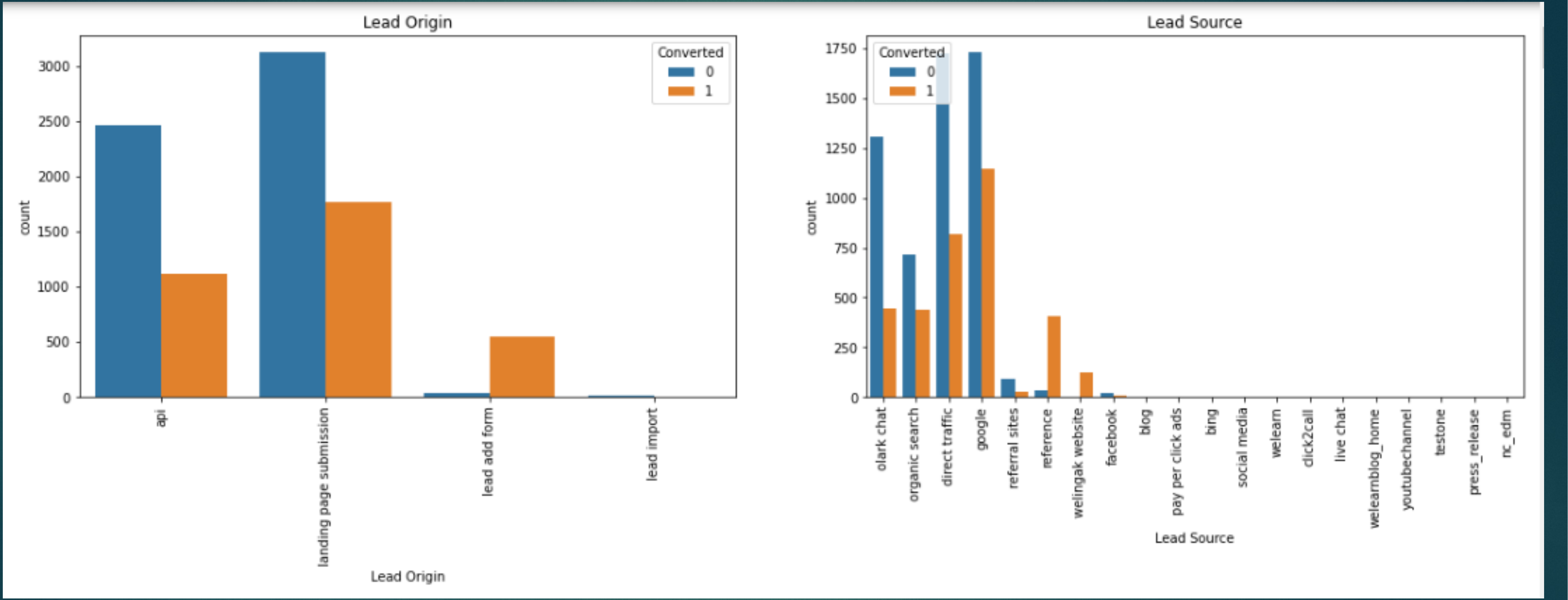

Specialization
count
0
500
1000
1500
2000
2500
3000
not selected
business administration
media and advertising
supply chain management
it projects management
finance management
travel and tourism
human resource management
marketing management
banking, investment and insurance
international business
e-commerce
operations management
retail management
services excellence
hospitality management
rural and agribusiness
healthcare management
e-business
Specialization
What is your current occupation
count
0
1000
2000
3000
4000
5000
unemployed
student
not selected
working professional
businessman
other
housewife
What is your current occupation

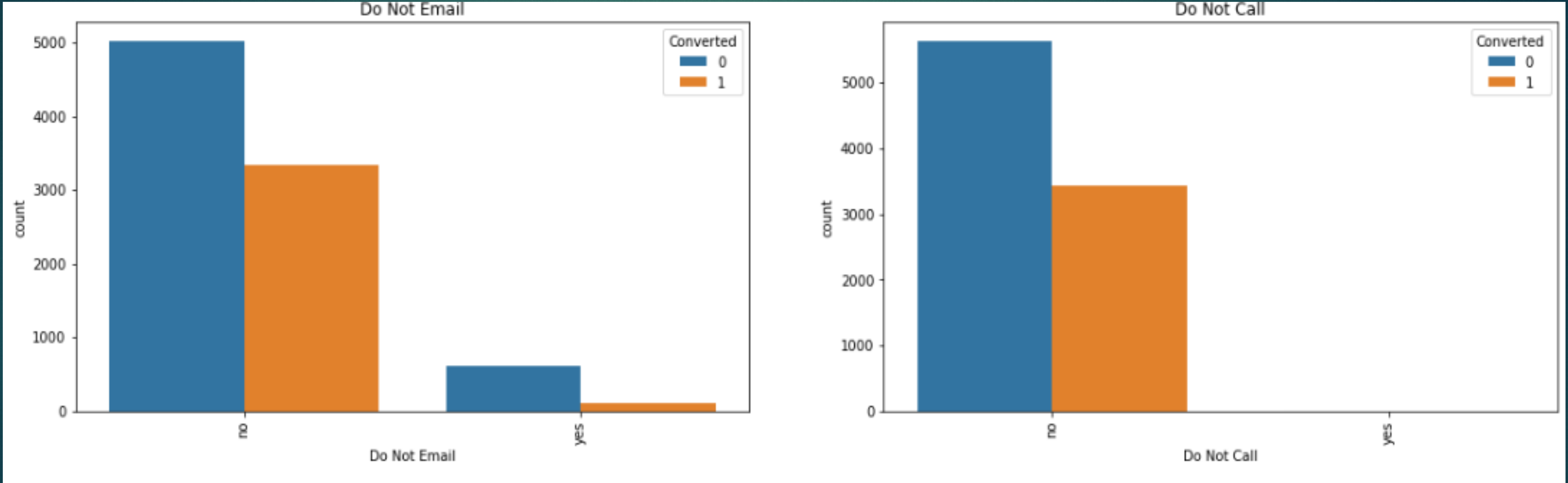

Last Notable Activity
count
0
500
1000
1500
2000
2500
3000
modified
email opened
page visited on website
email bounced
email link clicked
unreachable
unsubscribed
had a phone conversation
olark chat conversation
sms sent
approached upfront
resubscribed to emails
view in browser link clicked
form submitted on website
email received
email marked spam
Last Notable Activity

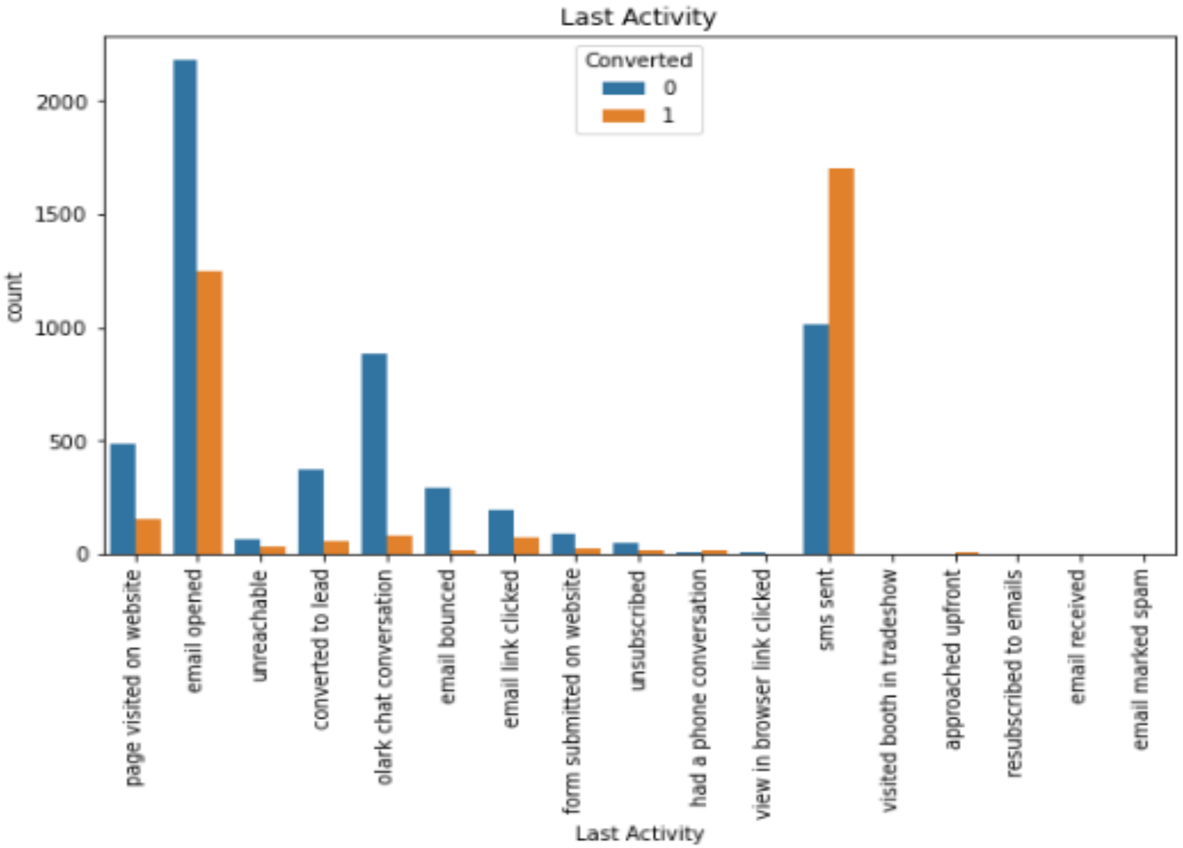
EDA: Continuous Variables:

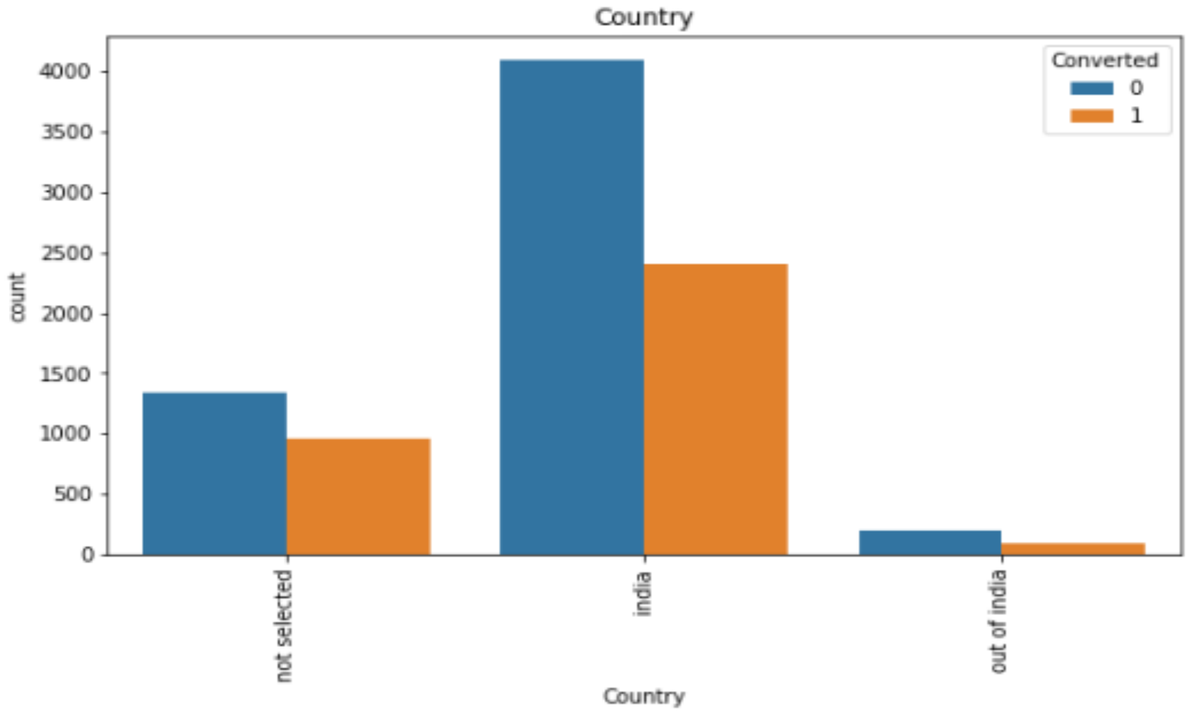


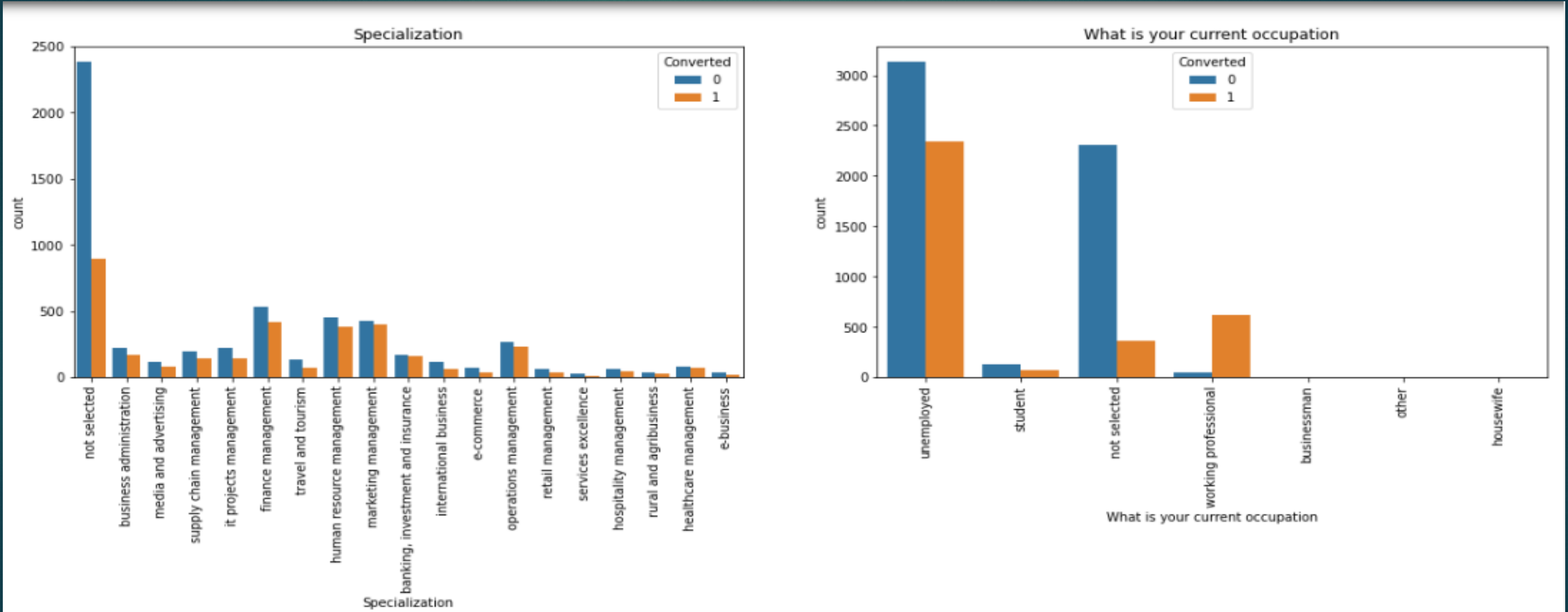
Bivariate Analysis: Categorical Variables against Conversion rate:

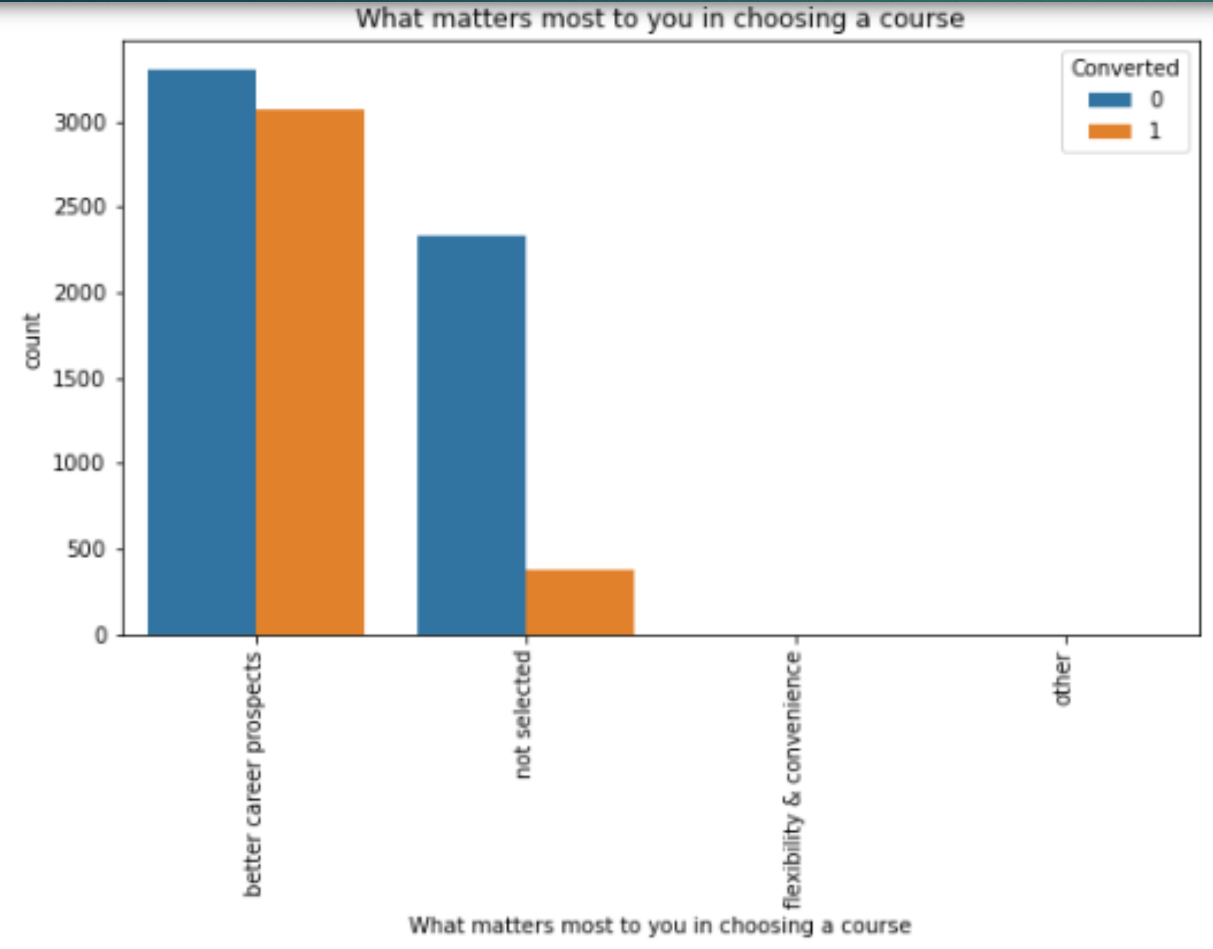


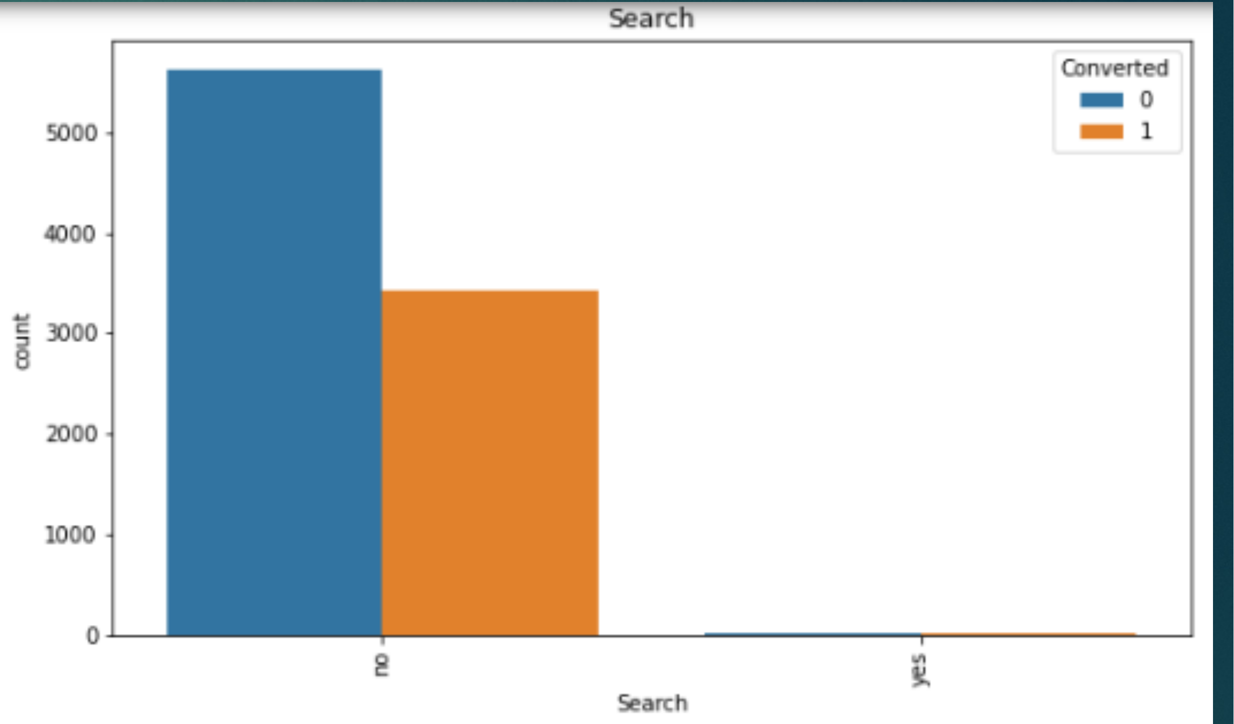

Do Not Email
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Email
Do Not Call
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Call

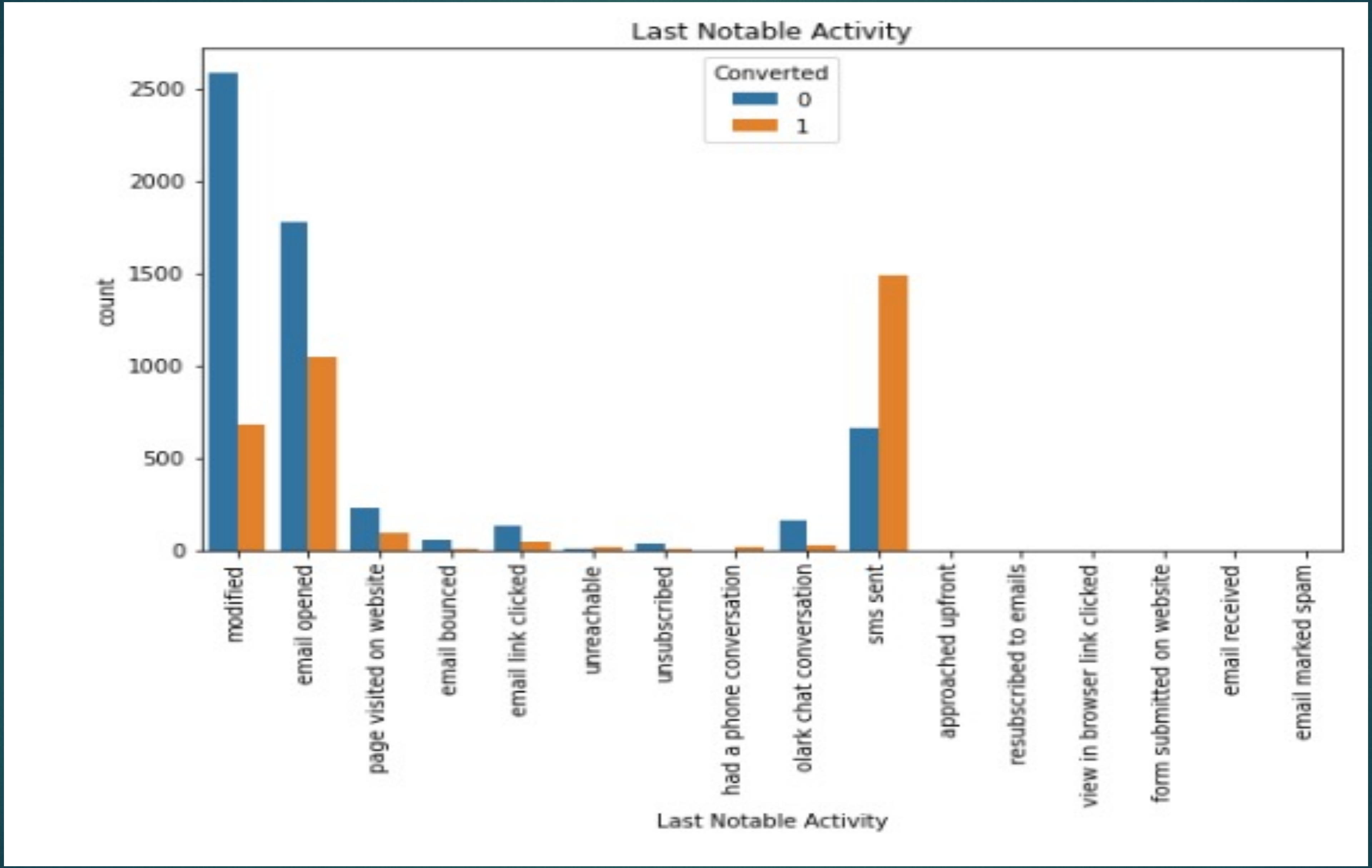

Last Activity
Converted
0
1
count
0
500
1000
1500
2000
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow
approached upfront
resubscribed to emails
email received
email marked spam
Last Activity


Country
Converted
0
1
count
0
500
1000
1500
2000
2500
3000
3500
4000
not selected
india
out of india
Country

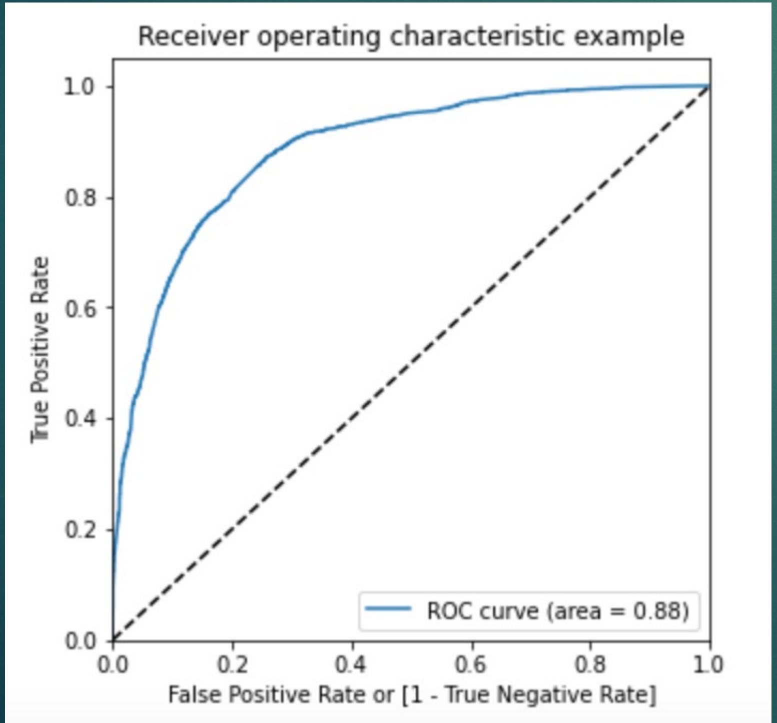

Specialization
Converted
0
1
count
2500
2000
1500
1000
500
0
not selected
business administration
media and advertising
supply chain management
it projects management
finance management
travel and tourism
human resource management
marketing management
banking, investment and insurance
international business
e-commerce
operations management
retail management
services excellence
hospitality management
rural and agribusiness
healthcare management
e-business
Specialization
What is your current occupation
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
unemployed
student
not selected
working professional
businessman
other
housewife
What is your current occupation

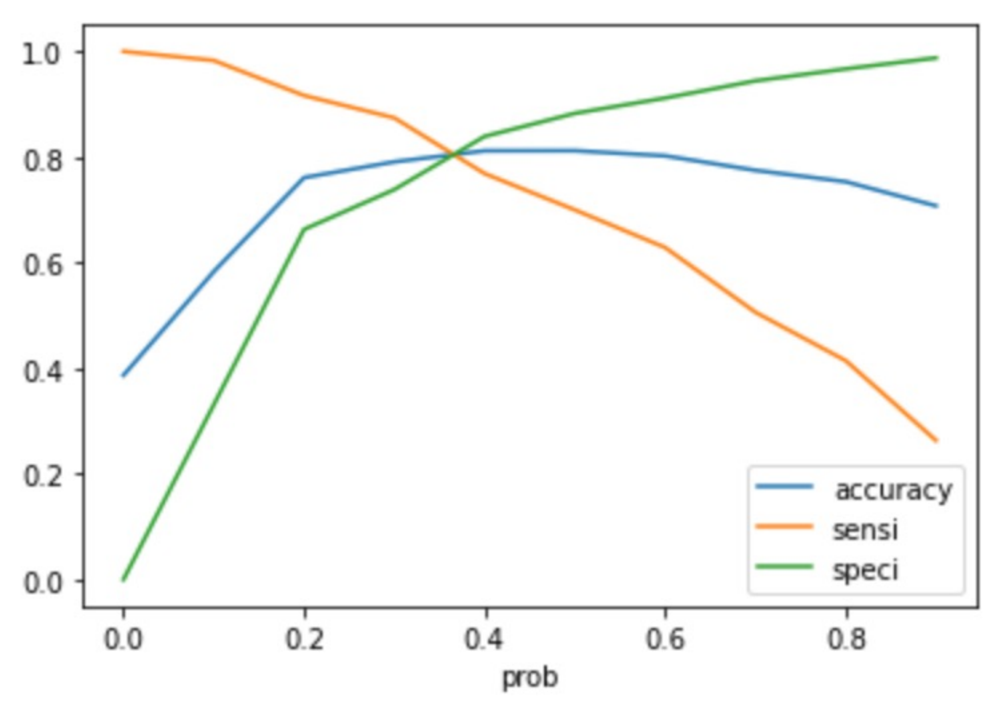

What matters most to you in choosing a course
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
better career prospects
not selected
flexibility & convenience
other
What matters most to you in choosing a course


Search
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Search


Last Notable Activity
Converted
0
1
count
2500
2000
1500
1000
500
0
modified
email opened
page visited on website
email bounced
email link clicked
unreachable
unsubscribed
had a phone conversation
olark chat conversation
sms sent
approached upfront
resubscribed to emails
view in browser link clicked
form submitted on website
email received
email marked spam
Last Notable Activity

- After the EDA, we go ahead with creating dummy variables for categorical variables.
- Then we do splitting of the data into train and test sets in 70:30 ratio.
- Now we go with building the model. We use RFE for feature selection and then train the model using those features. In this respect, our third model has been finalized as our final model with 80% accuracy.
- The ROC curve is as shown in figure which is acceptable. Also the accuracy-sensitivity and specificity tradeoff has given a cutoff of 0.35.





Conclusion:

It was found that the following parameters are effecting the conversion rate of potential hot leads:

- The total time spend on the Website.
- Total number of visits.
- When the lead source was:Google, Direct traffic, Organic search
- When the last activity was: SMS, Olark chat conversation
- When the lead origin is Lead add format.
- When their current occupation is as a working professional.

by focussing on these potential parameters, X Education can focus on these variables to achieve the best conversion rate

Thank you