LEAD SCORING CASE STUDY

BY: VENKATA L N S V RAGHAVAN VINJAMURI

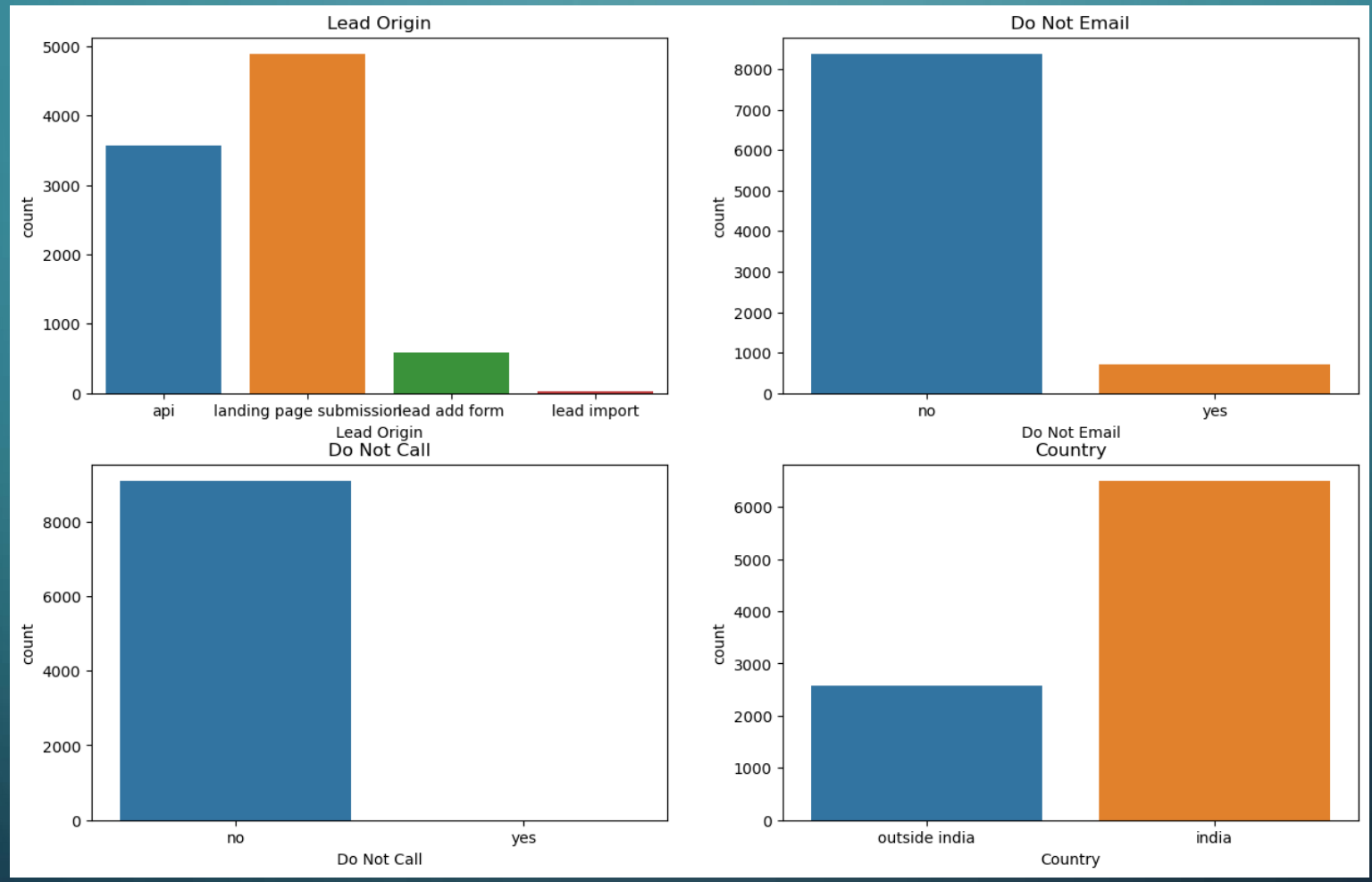
SACHIN KUMAR SINGH

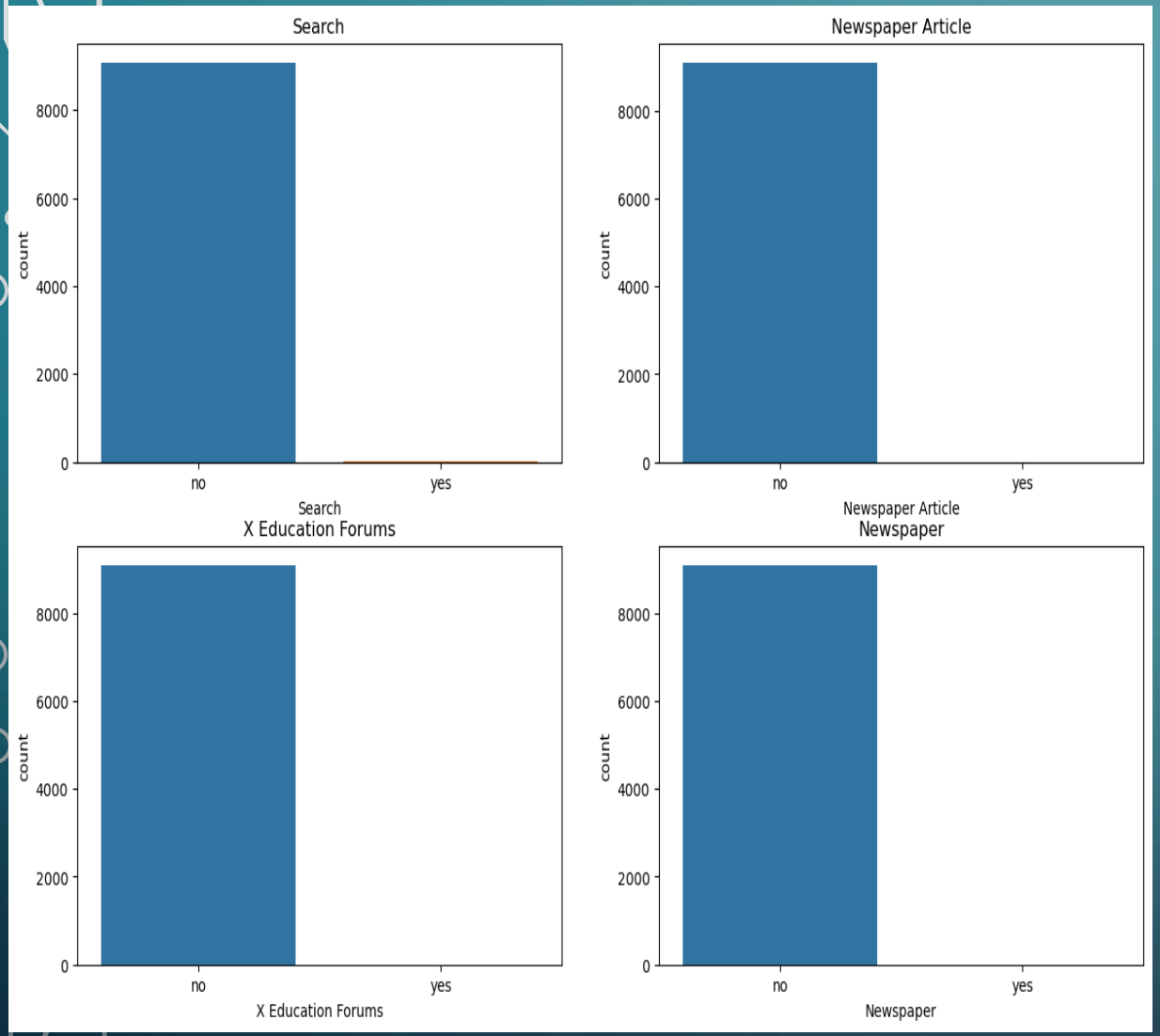
RAGHUNATH V P

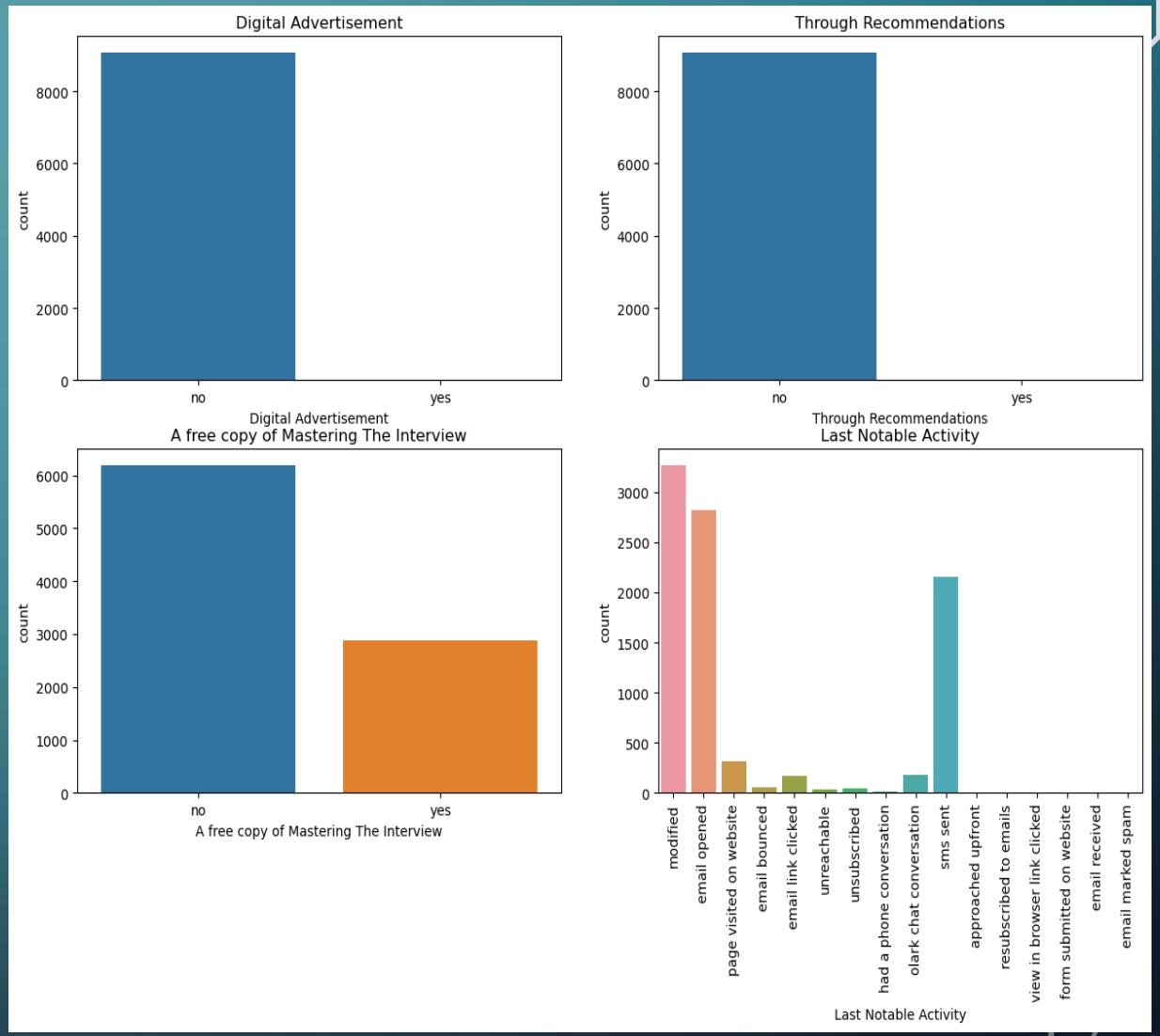
THE PROBLEM

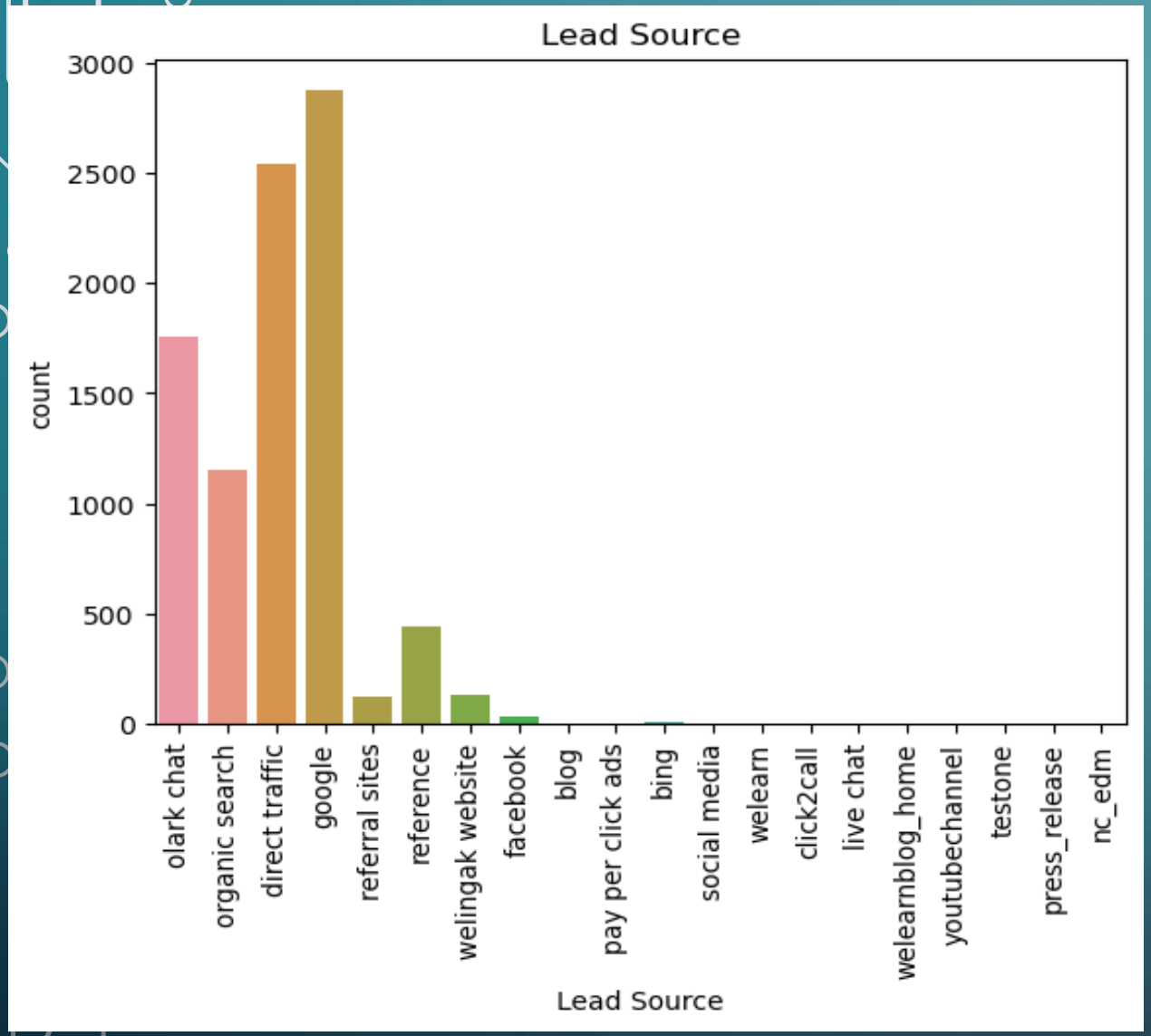
What is the problem?	• To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'.
Who has this problem?	• An education company named X Education sells online courses to industry professionals.
Why should this problem be solved?	• Although X Education gets a lot of leads, its lead conversion rate is very poor.
How will I know this has been solved?	• The target lead conversion rate to be around 80%.

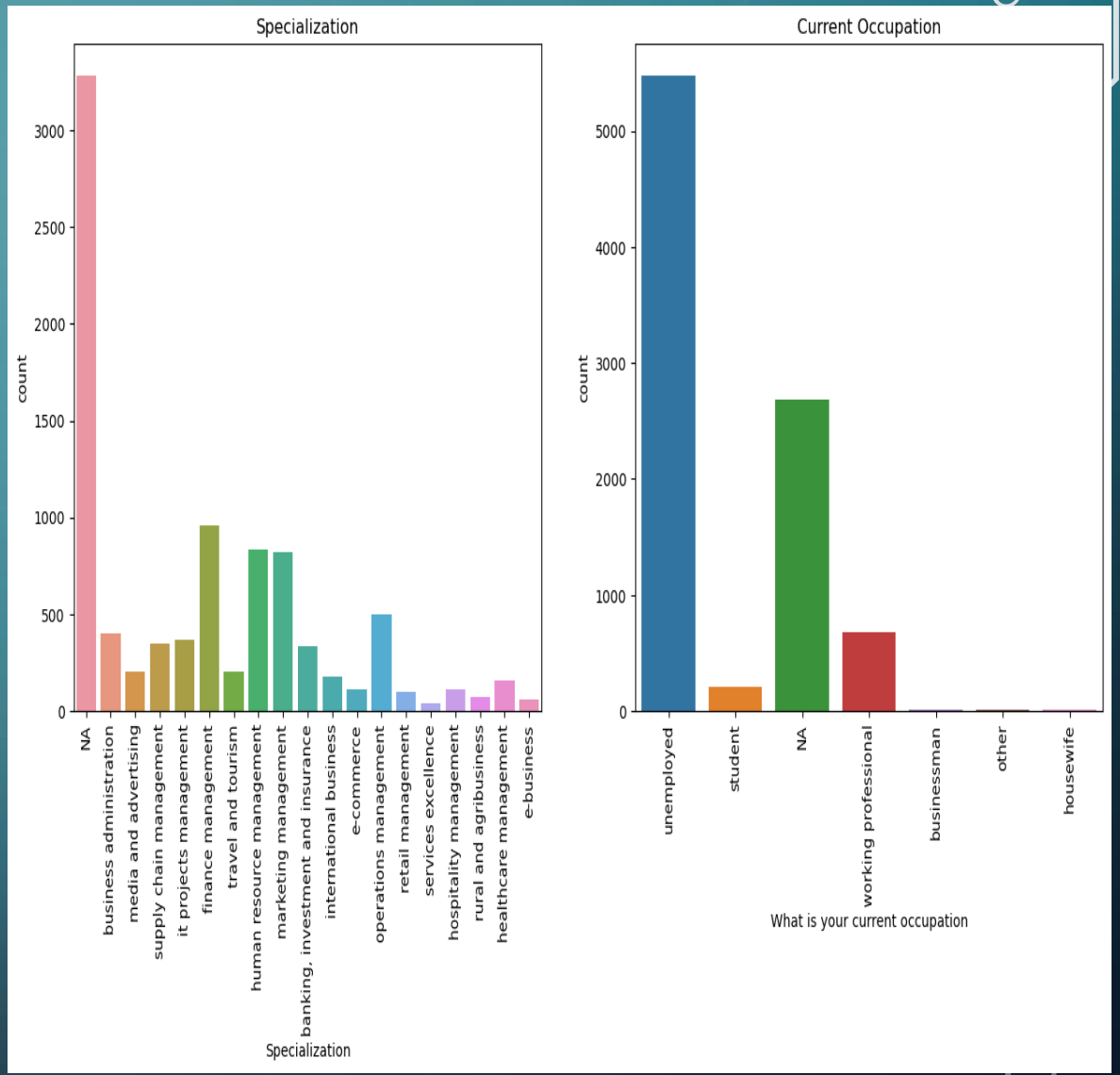
CATEGORICAL VARIABLES - UNIVARIATE ANALYSIS

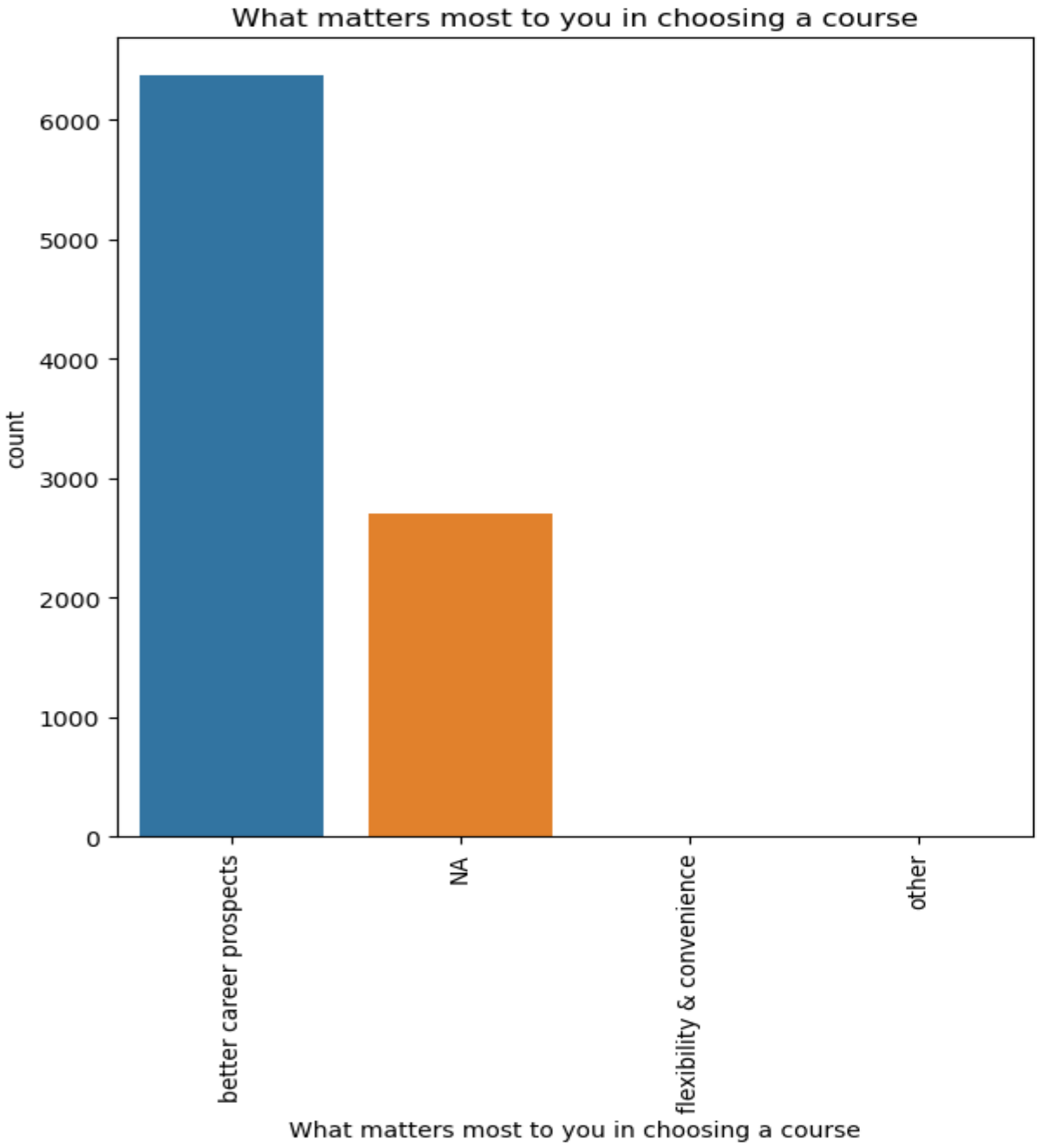


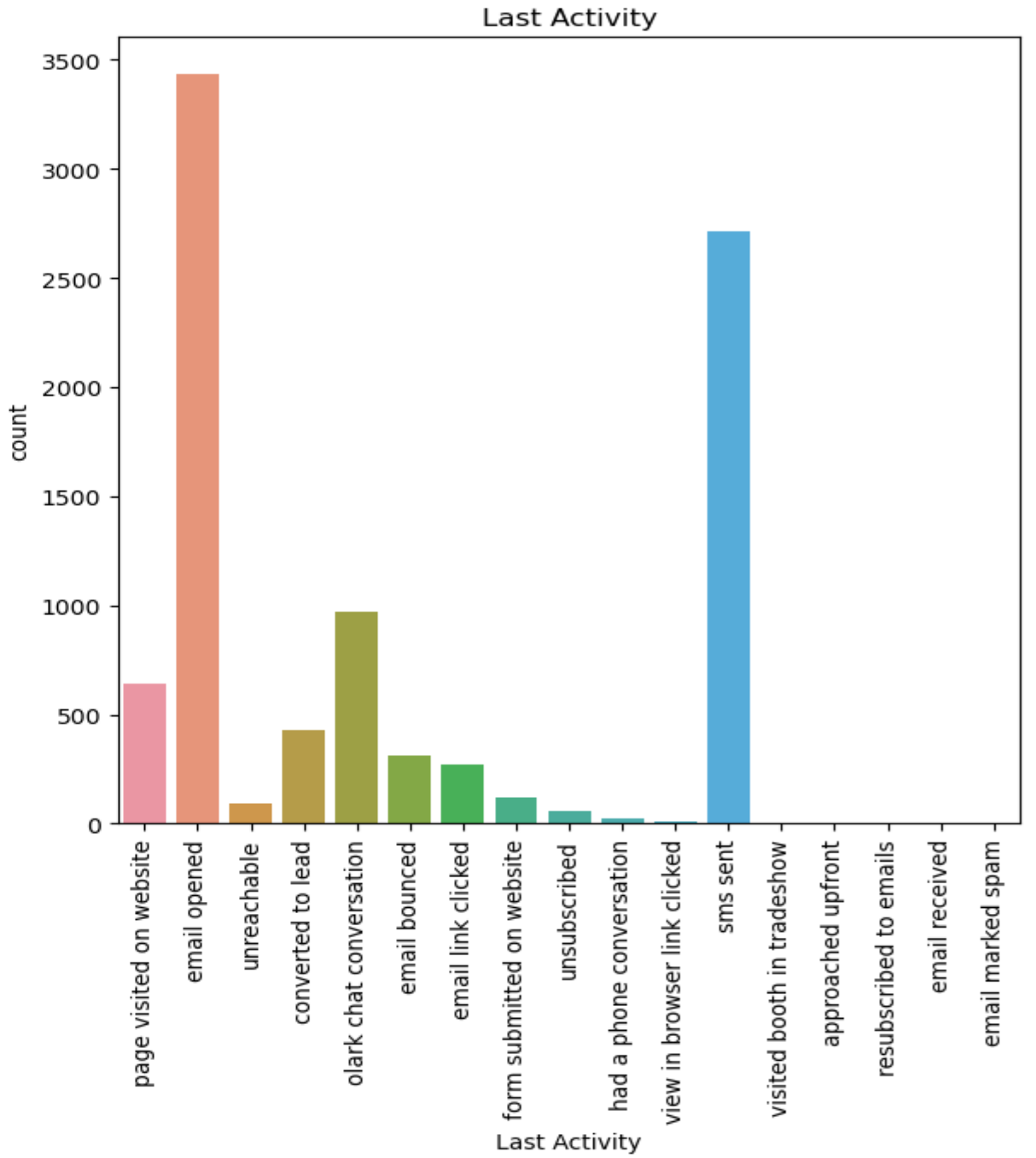

Search
count
no
yes
Search
Newspaper Article
count
no
yes
Newspaper Article
X Education Forums
count
no
yes
X Education Forums
Newspaper
count
no
yes
Newspaper


Digital Advertisement
count
no
yes
Digital Advertisement
Through Recommendations
count
no
yes
Through Recommendations
A free copy of Mastering The Interview
count
no
yes
A free copy of Mastering The Interview
Last Notable Activity
count
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

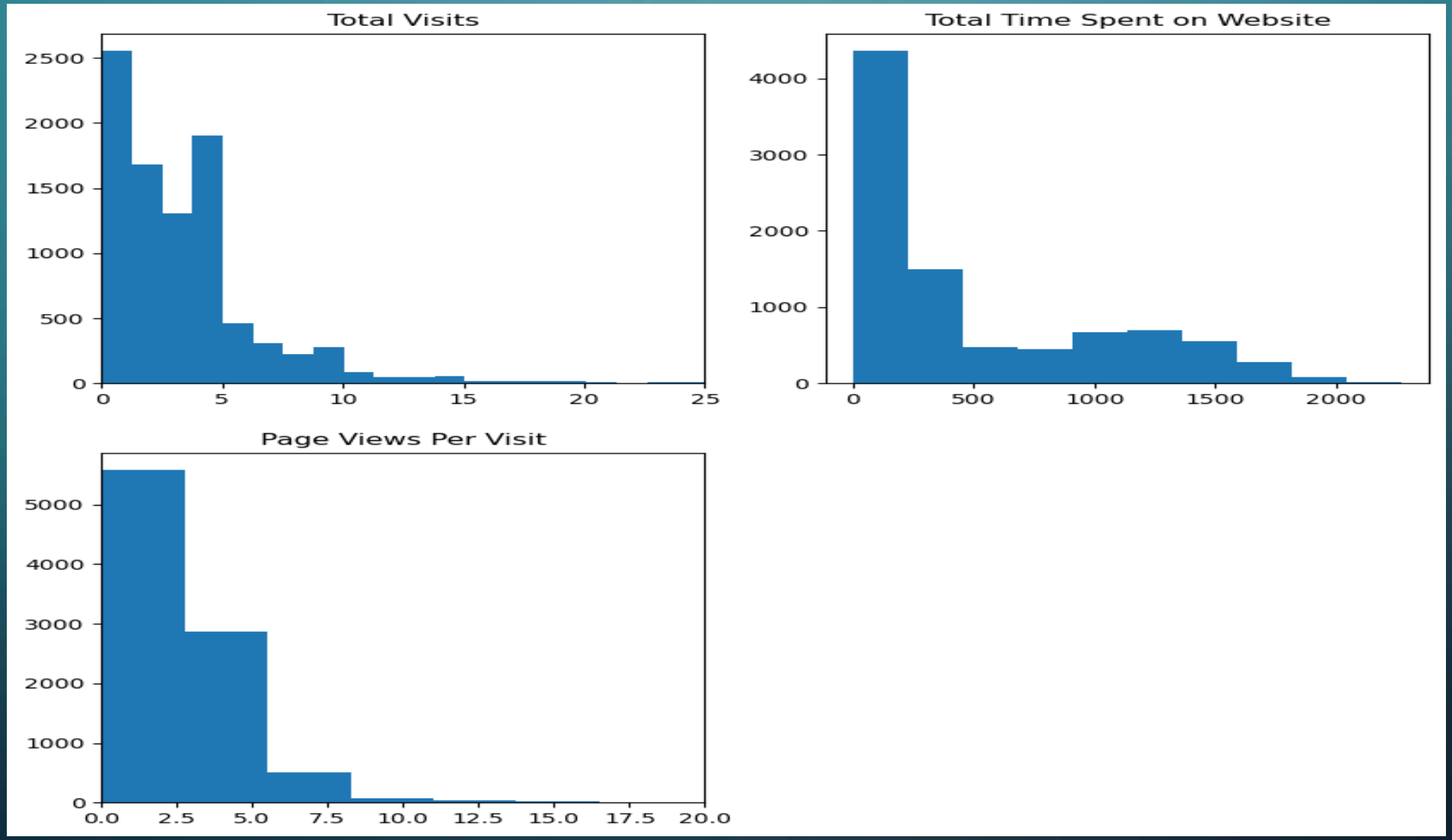

Lead Source
count
3000
2500
2000
1500
1000
500
0
olark chat
organic search
direct traffic
google
referral sites
reference
welingak website
facebook
blog
pay per click ads
bing
social media
welearn
click2call
live chat
welearnblog_home
youtubechannel
testone
press_release
nc_edm
Lead Source


Specialization
count
3000
2500
2000
1500
1000
500
0
NA
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
Current Occupation
count
5000
4000
3000
2000
1000
0
unemployed
student
NA
working professional
businessman
other
housewife
What is your current occupation

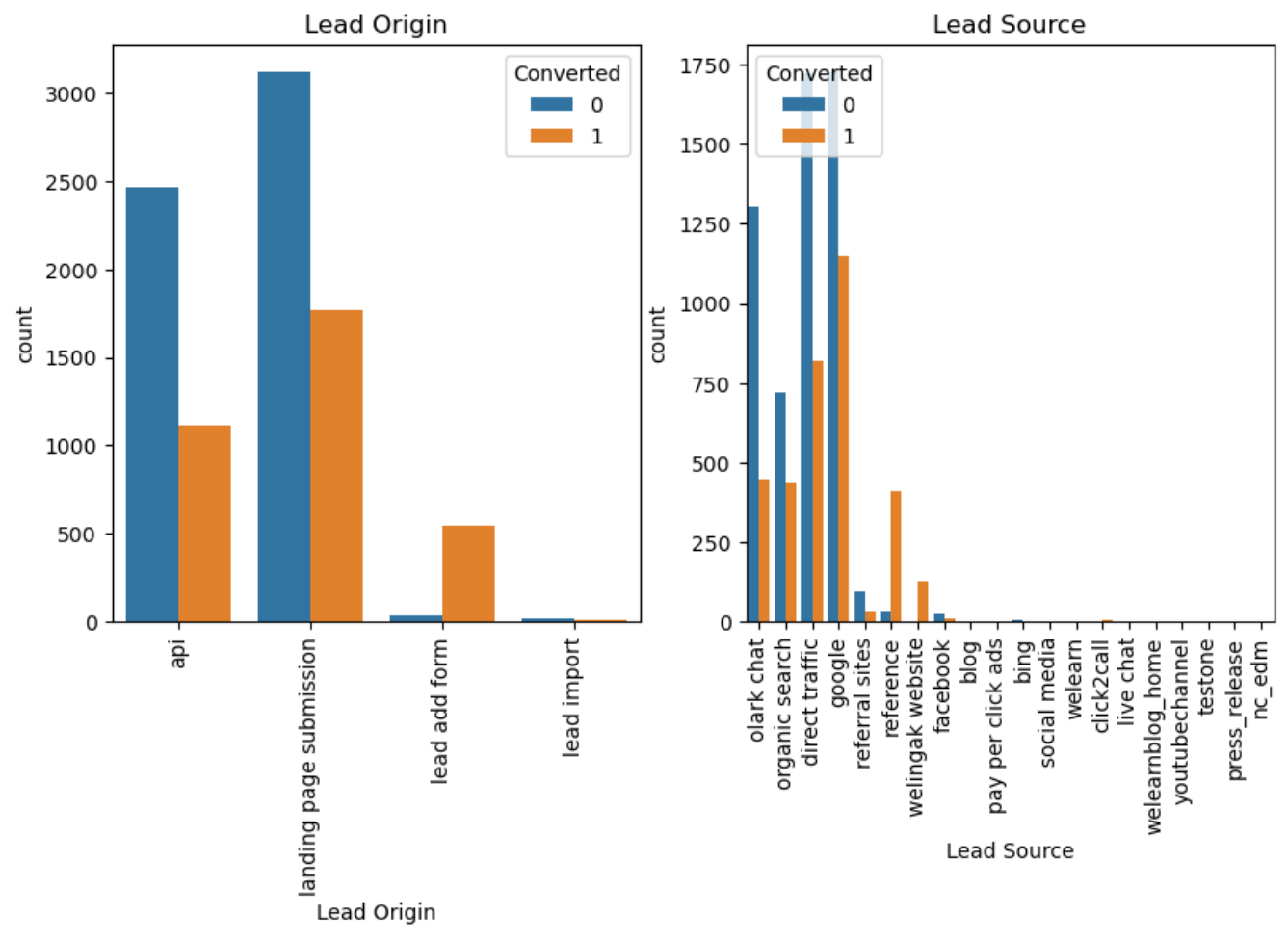

What matters most to you in choosing a course
count
6000
5000
4000
3000
2000
1000
0
better career prospects
NA
flexibility & convenience
other
What matters most to you in choosing a course

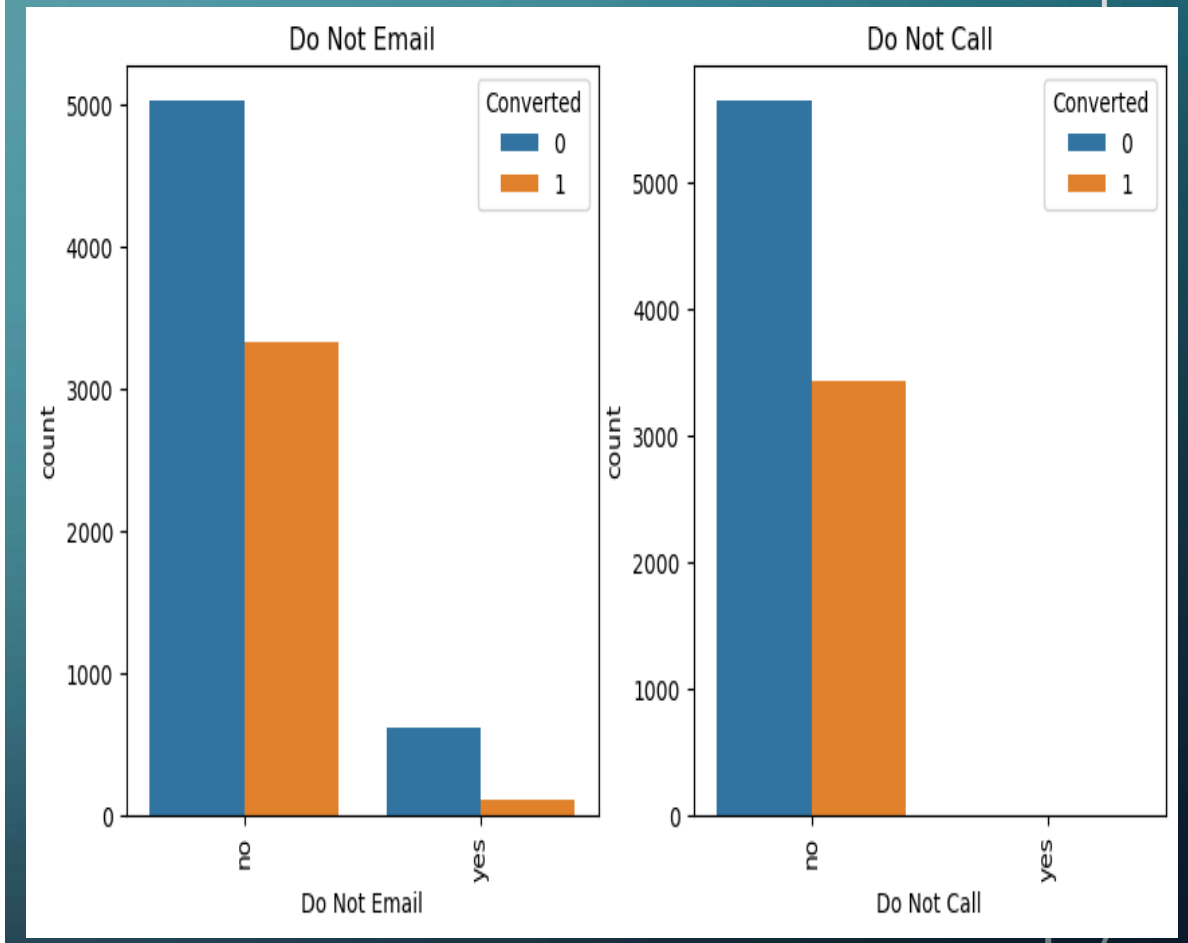

Last Activity
count
3500
3000
2500
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow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

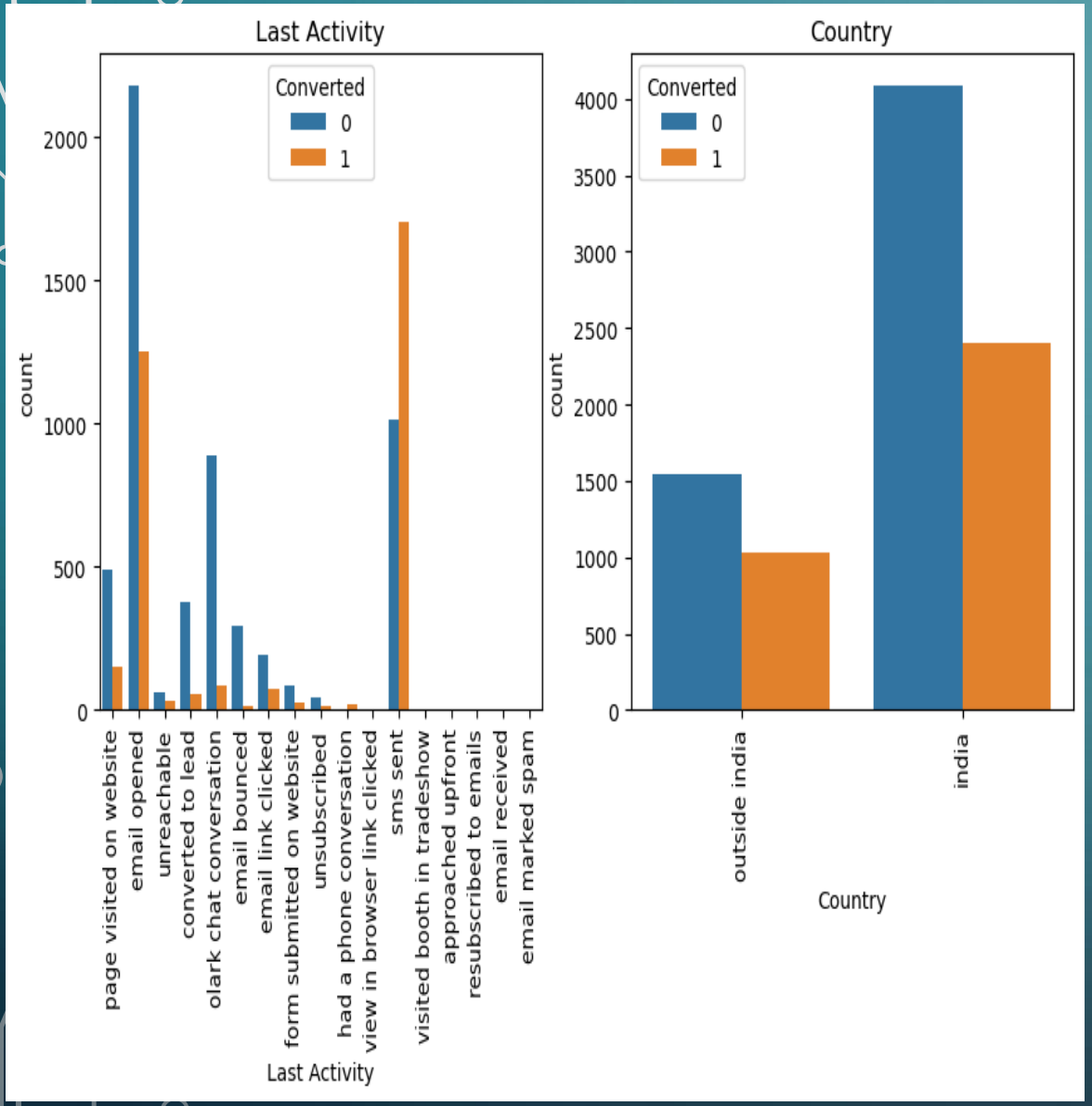
NUMERICAL VARIABLES - UNIVARIATE ANALYSIS



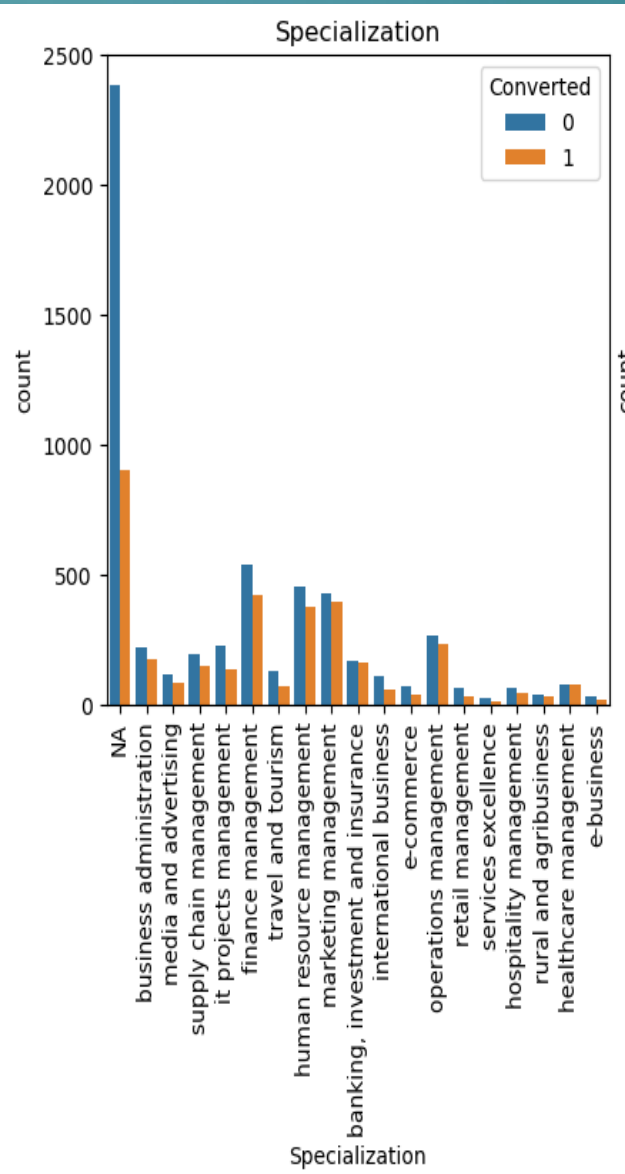
BIVARIATE ANALYSIS WITH "CONVERTED" VARIABLE

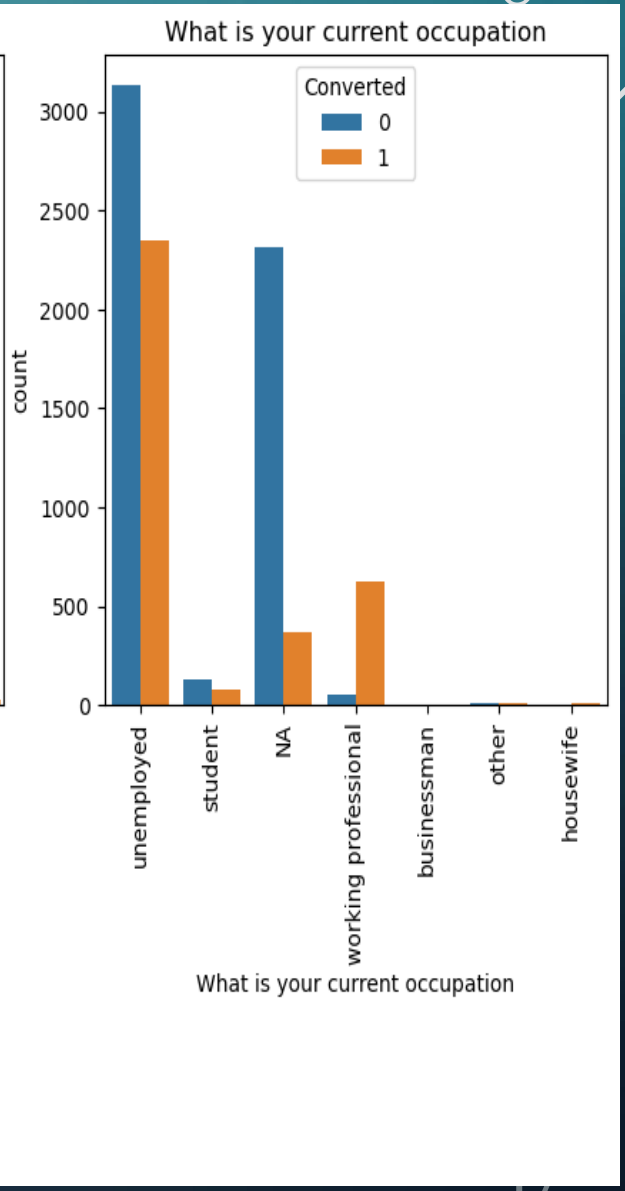


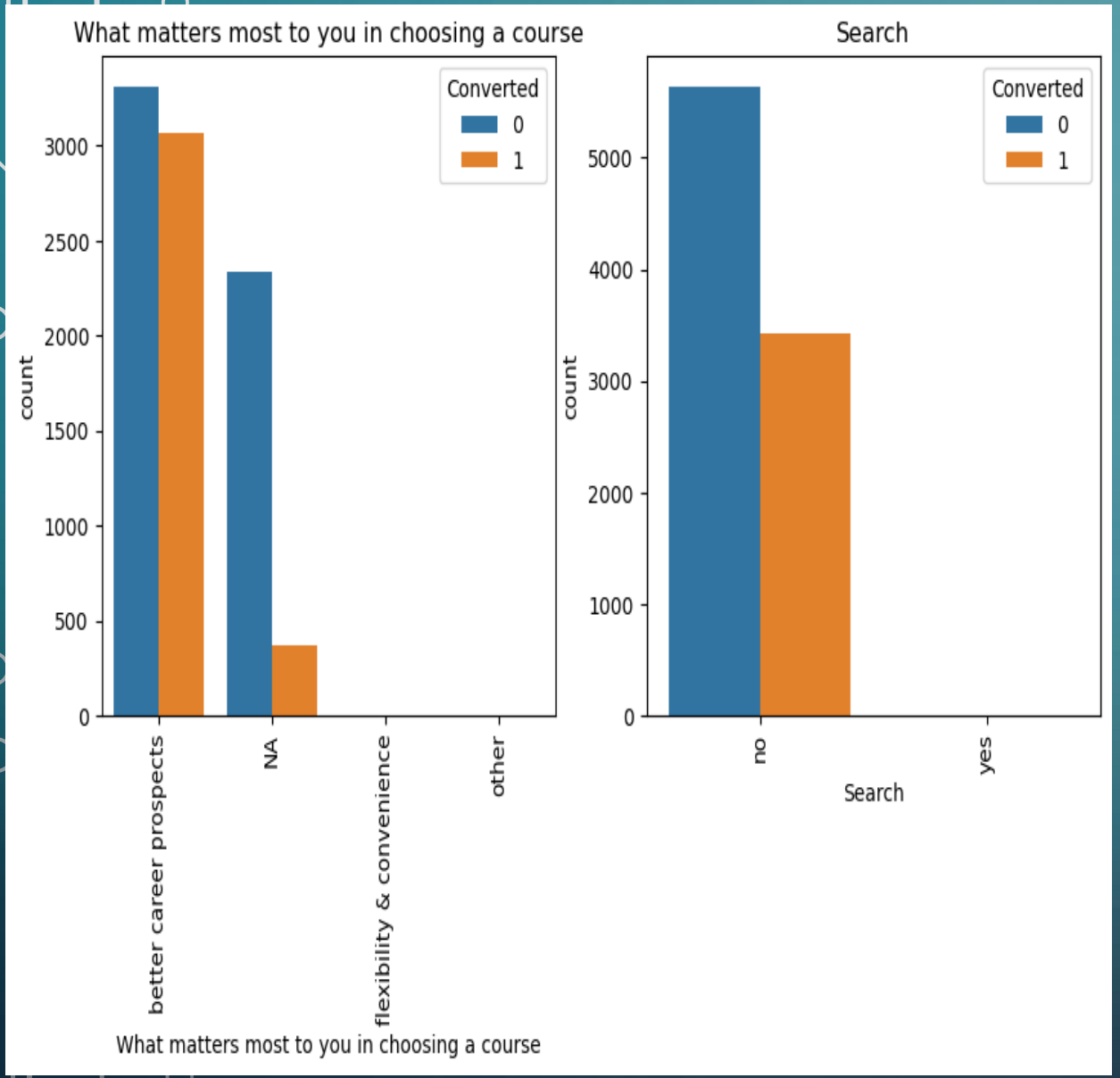


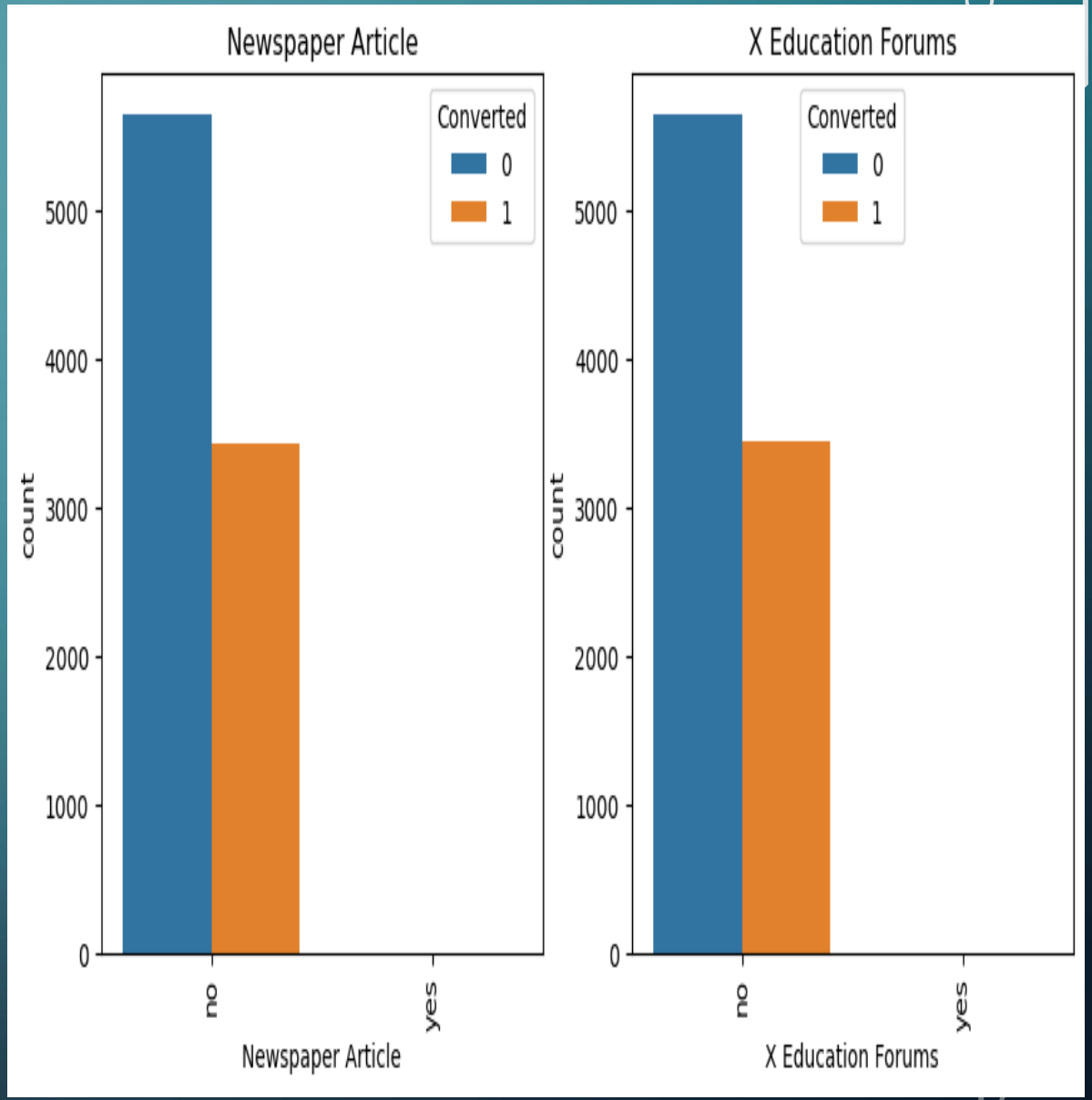

Last Activity
Converted
0
1
count
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow
approached upfront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

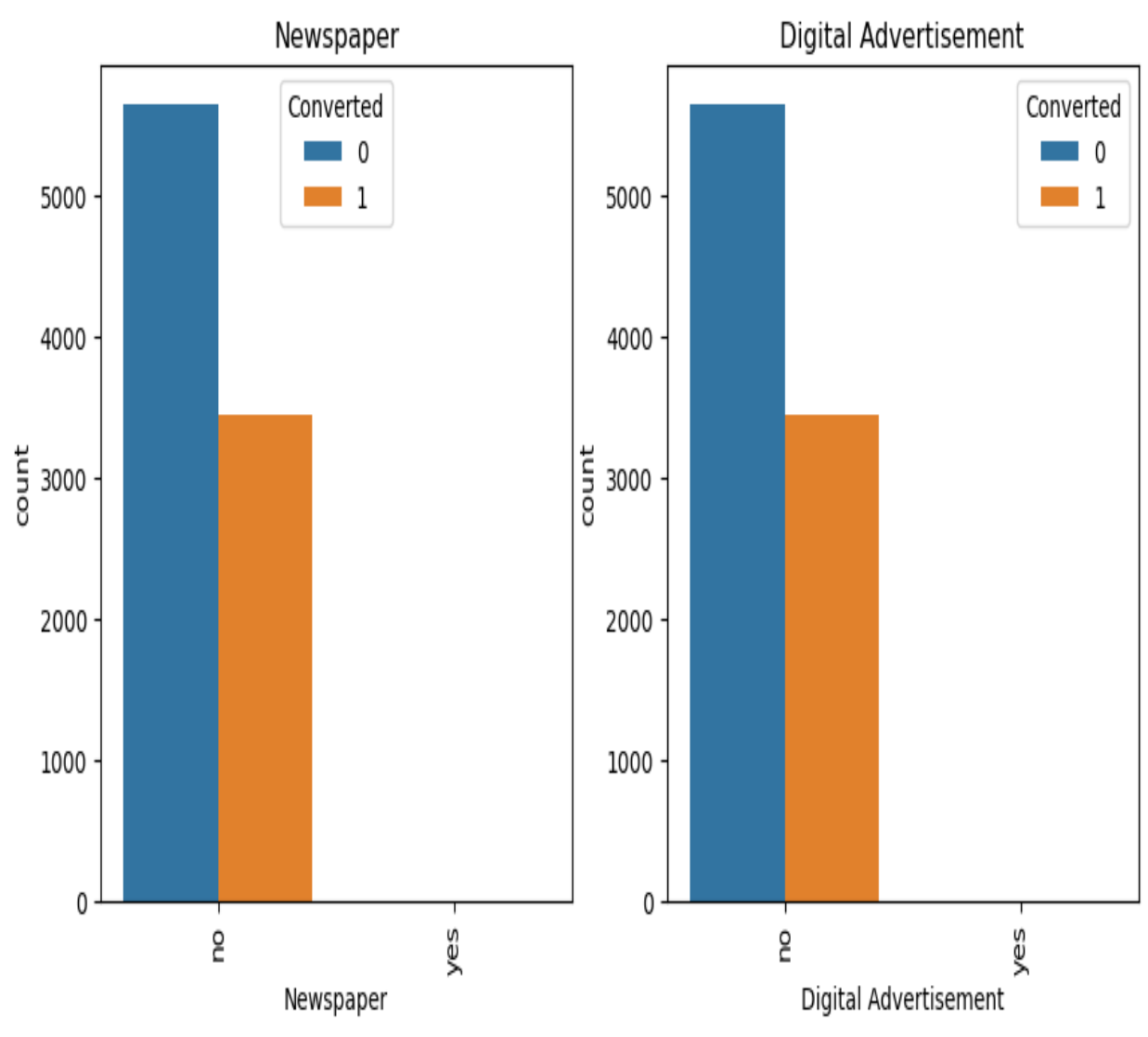
Country
Converted
0
1
count
outside india
india
Country


Specialization
Converted
0
1
count
NA
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

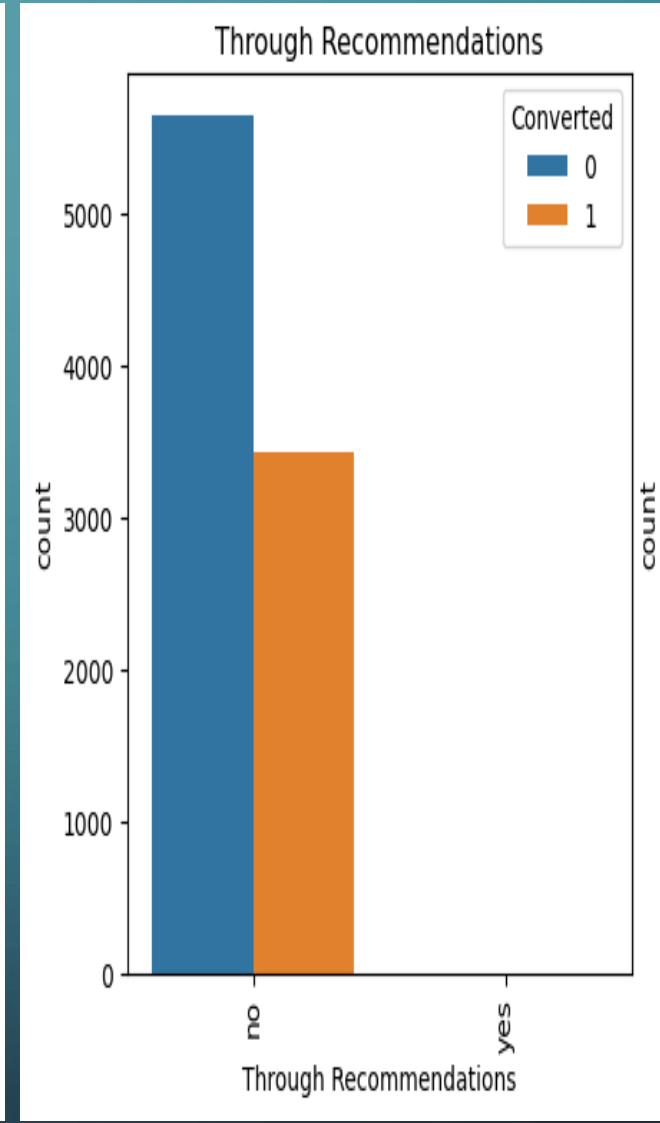

What is your current occupation
Converted
0
1
count
unemployed
student
NA
working professional
businessman
other
housewife
What is your current occupation

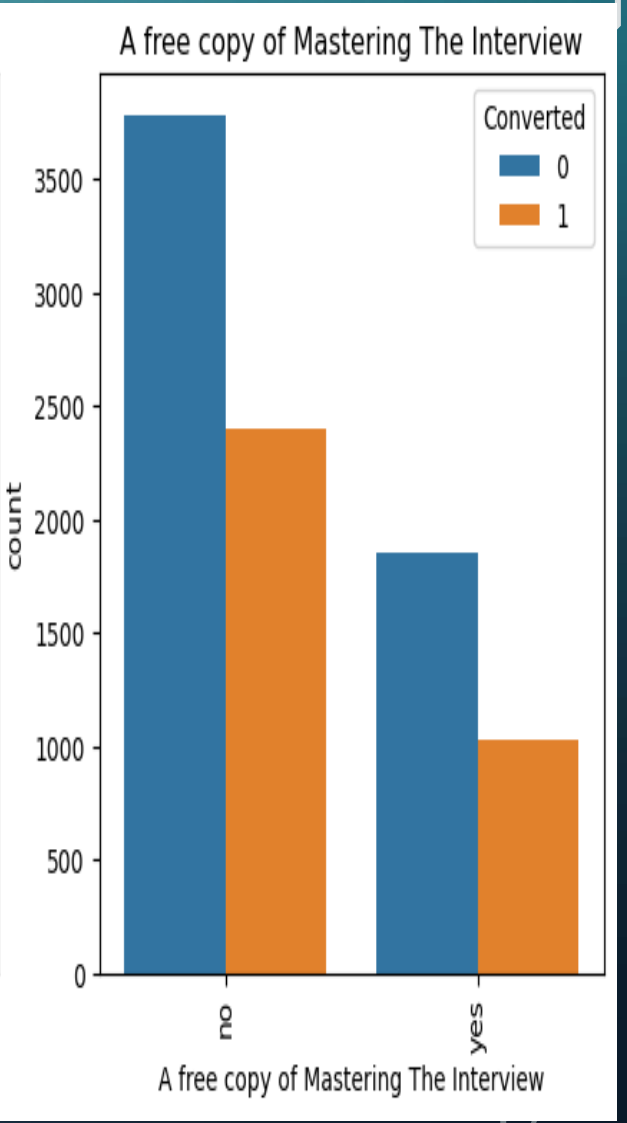

What matters most to you in choosing a course
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
better career prospects
NA
flexibility & convenience
other
What matters most to you in choosing a course
Search
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Search

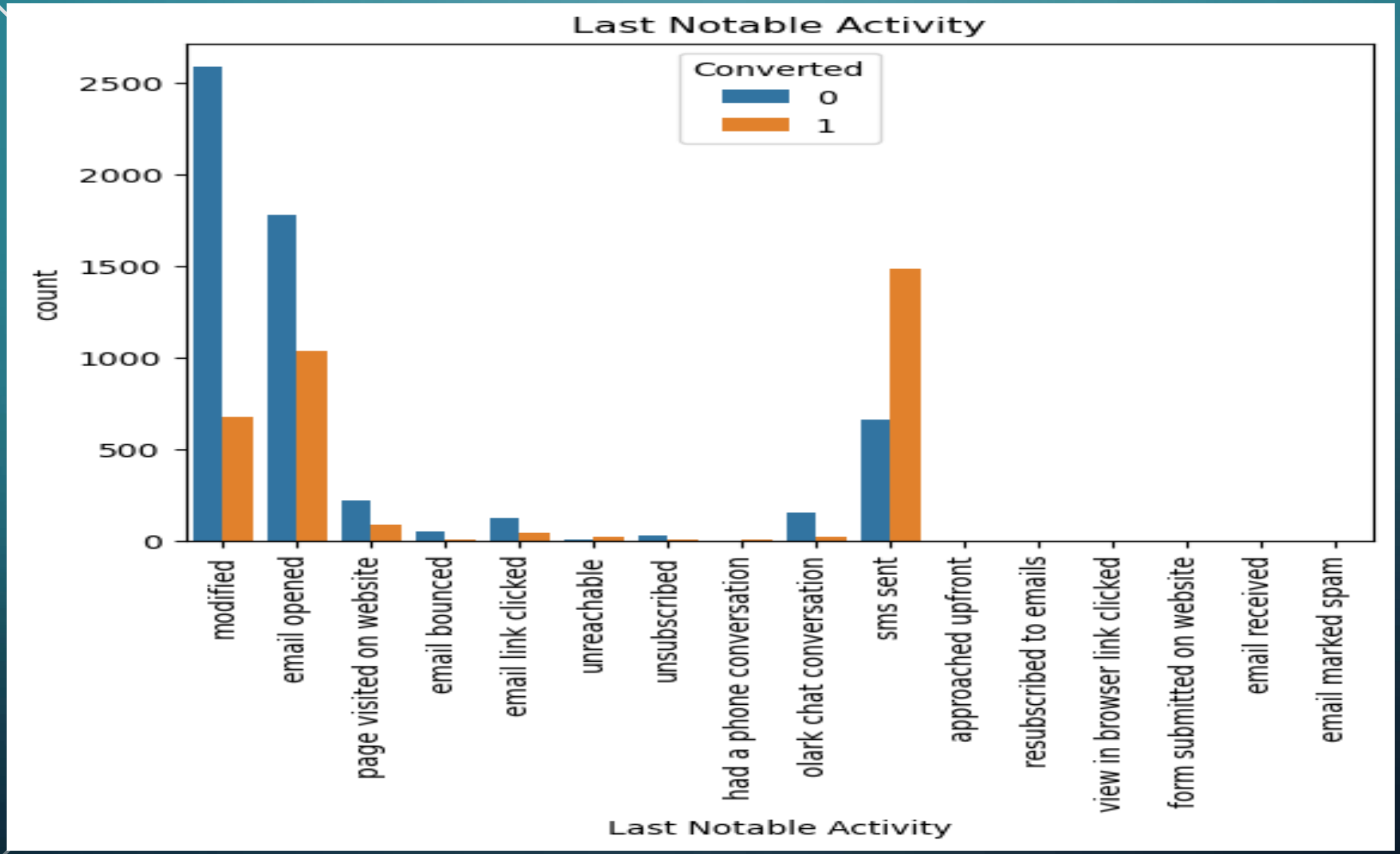

Newspaper Article
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Newspaper Article
X Education Forums
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
X Education Forums


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

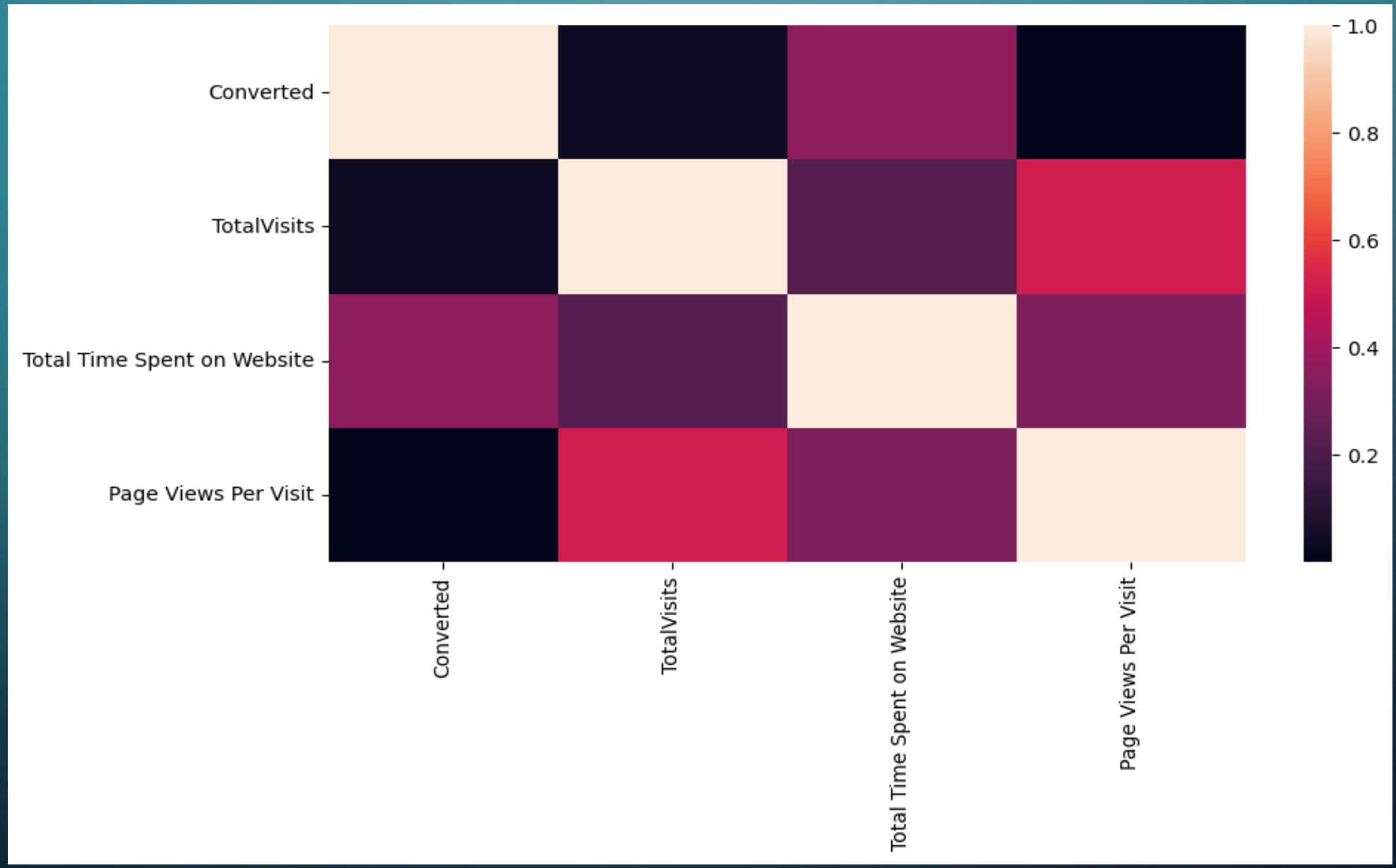
Digital Advertisement
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Digital Advertisement


Through Recommendations
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Through Recommendations


A free copy of Mastering The Interview
Converted
0
1
count
3500
3000
2500
2000
1500
1000
500
0
no
yes
A free copy of Mastering The Interview


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

HEAT MAP TO CHECK THE CORRELATION AMONG VARIABLES

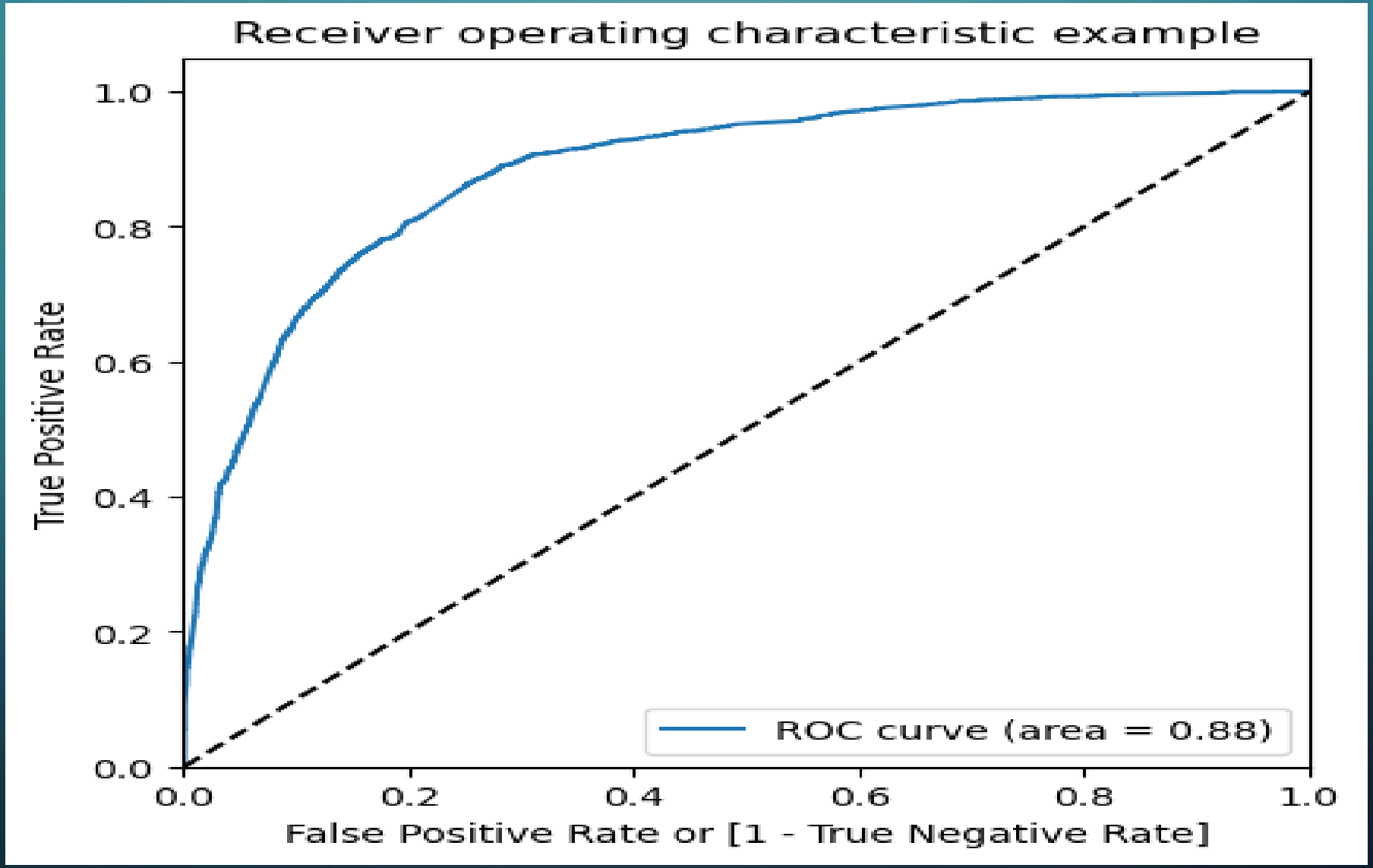


Generalized Linear Model Regression Results

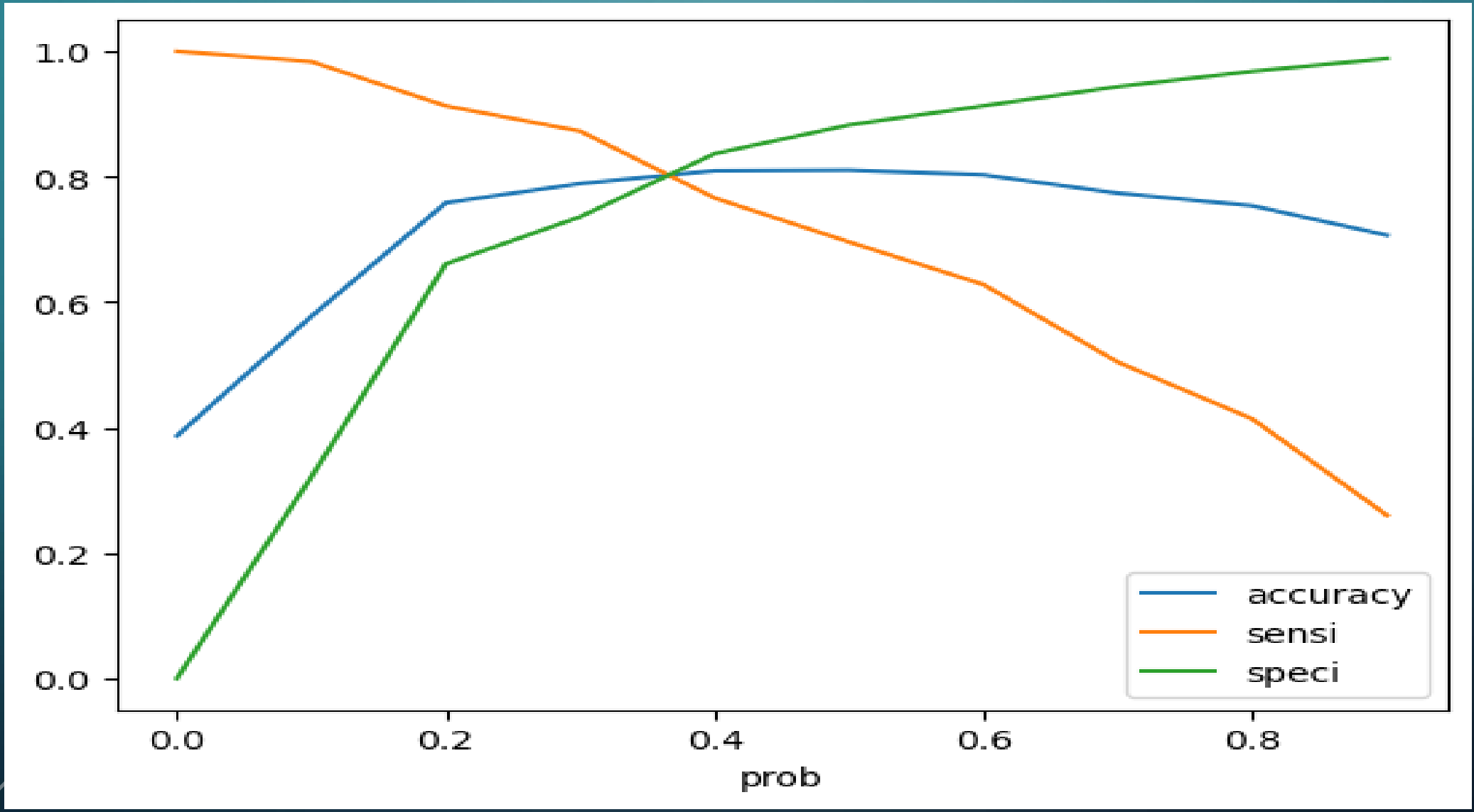
Dep. Variable:	Converted	**No. Observations:**	6351
Model:	GLM	**Df Residuals:**	6337
Model Family:	Binomial	**Df Model:**	13
Link Function:	Logit	**Scale:**	1.0000
Method:	IRLS	**Log-Likelihood:**	-2651.3
Date:	Mon, 14 Aug 2023	**Deviance:**	5302.6
Time:	03:35:00	**Pearson chi2:**	6.50e+03
No. Iterations:	7	**Pseudo R-squ. (CS):**	0.3932
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-3.4533	0.113	-30.579	0.000	-3.675	-3.232
TotalVisits	5.5427	1.444	3.838	0.000	2.712	8.373
Total Time Spent on Website	4.6048	0.166	27.690	0.000	4.279	4.931
Lead Origin_lead add form	3.7501	0.225	16.651	0.000	3.309	4.192
Lead Source_olark chat	1.5802	0.111	14.187	0.000	1.362	1.798
Lead Source_welingak website	2.5821	1.033	2.500	0.012	0.558	4.607
Do Not Email_yes	-1.4360	0.170	-8.437	0.000	-1.770	-1.102
Last Activity_olark chat conversation	-1.3974	0.167	-8.348	0.000	-1.725	-1.069
Last Activity_sms sent	1.2672	0.074	17.164	0.000	1.123	1.412
What is your current occupation_other	2.1567	0.755	2.857	0.004	0.677	3.636
What is your current occupation_student	1.2456	0.226	5.502	0.000	0.802	1.689
What is your current occupation_unemployed	1.1632	0.086	13.582	0.000	0.995	1.331
What is your current occupation_working professional	3.6797	0.204	18.008	0.000	3.279	4.080
Last Notable Activity_unreachable	1.8153	0.601	3.022	0.003	0.638	2.993

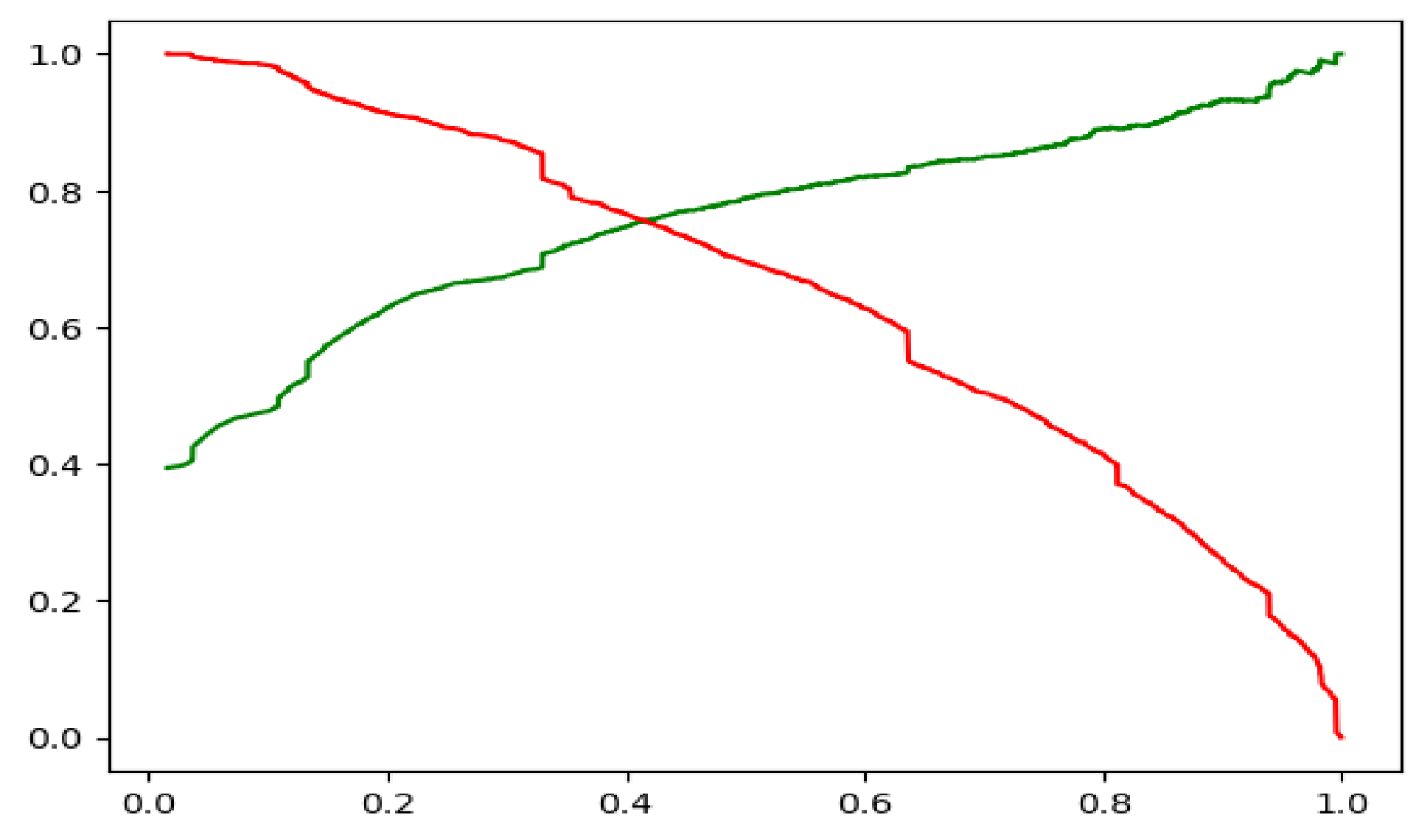
THE ROC CURVE



ACCURACY SENSITIVITY AND SPECIFICITY FOR VARIOUS PROBABILITIES



PRECISION AND RECALL TRADEOFF THRESHOLDS



FINAL PRECISION

With cutoff as 0.41, we have Precision around 73% and Recall around 76%

- We can say the that Precision around 73% and Recall around 76% are most likely to convert into paying customers.

OBSERVATIONS

It was found that the variables that mattered the most in the potential buyers are :

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
    a. Google
    b. Direct traffic
    c. Organic search
    d. Welingak website
4. When the last activity was:
    a. SMS
    b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.