Lead Scoring Case Study

Jacob Asir

Problem Statement

- X Education sells online courses to industry professionals.
- 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.

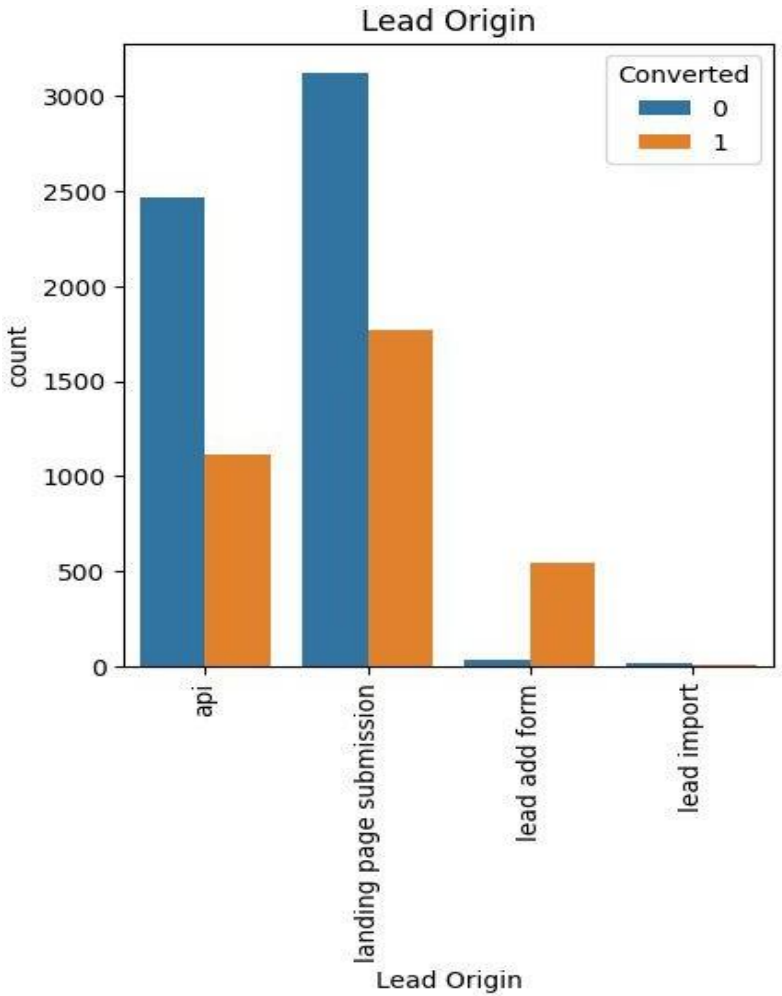
Business Objective

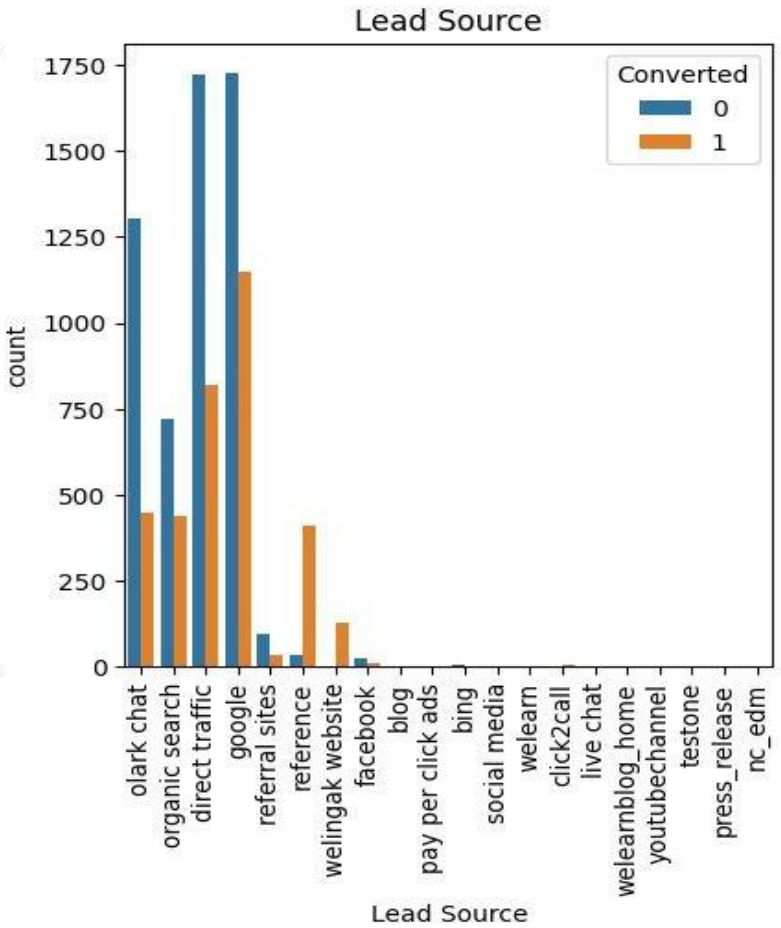
- Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

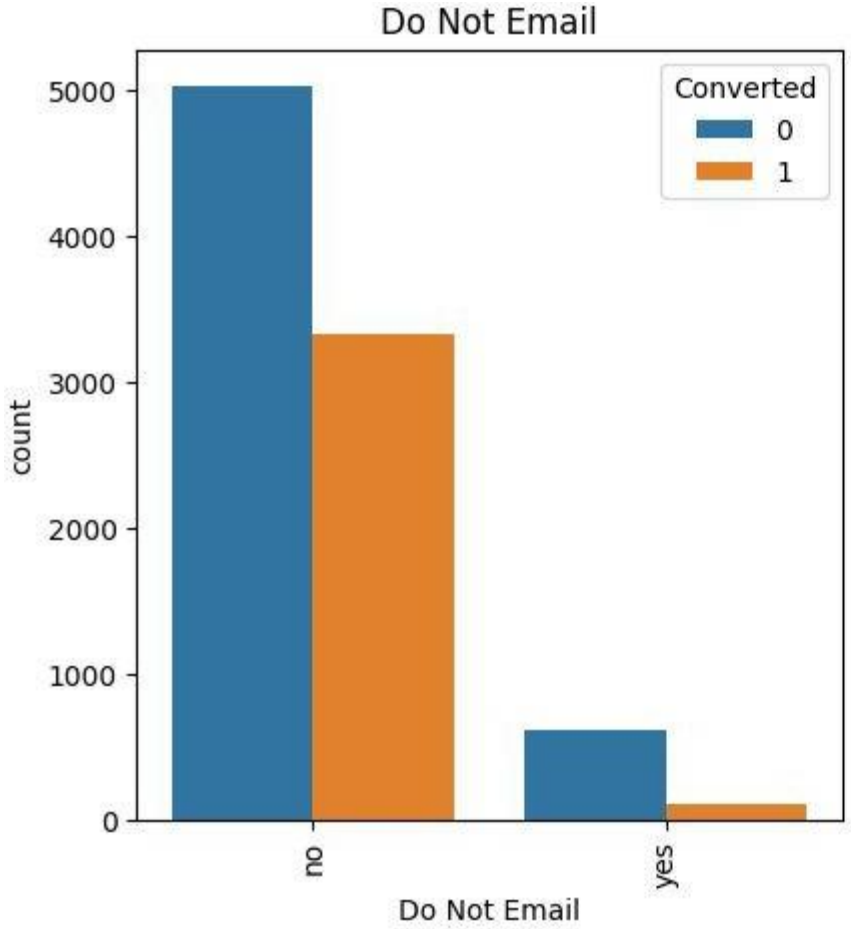
- There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you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

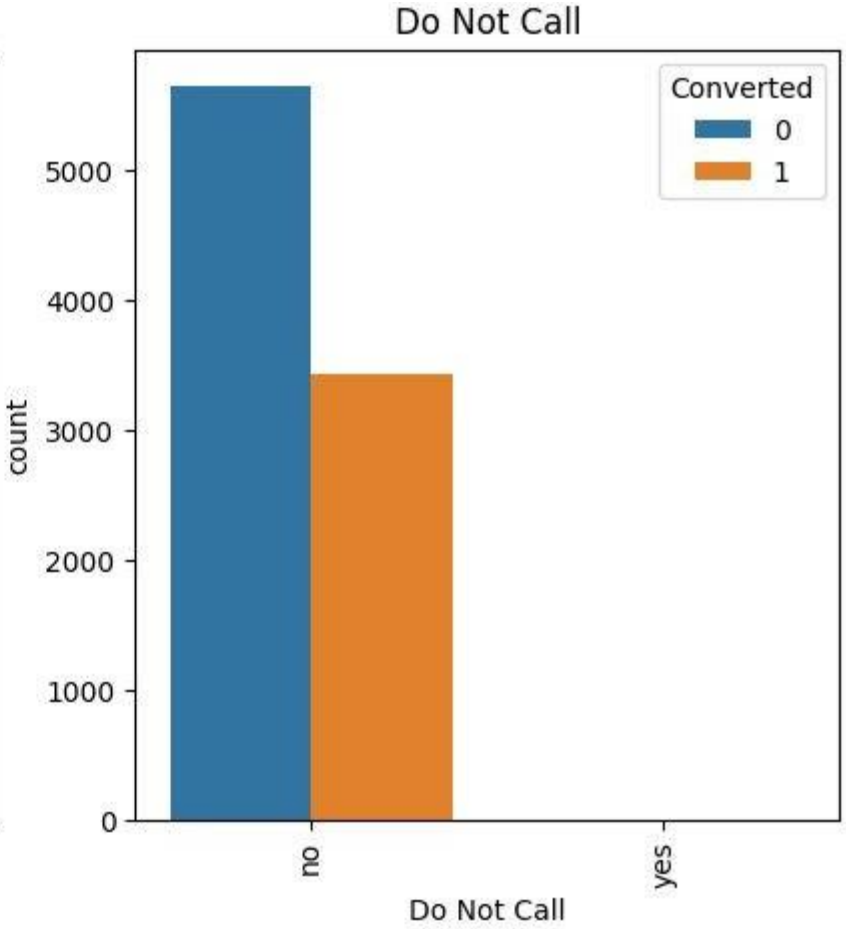
Solution Methodology

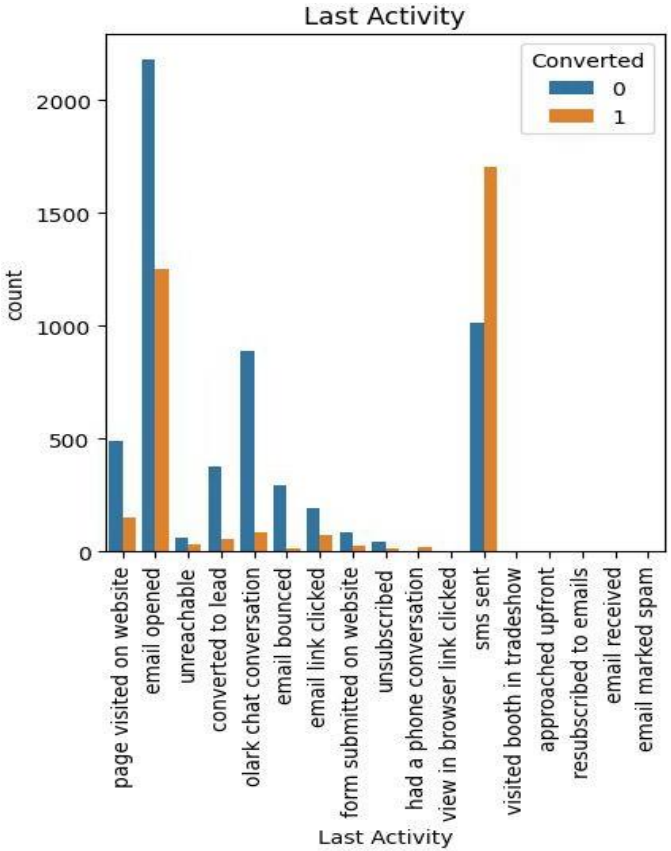
- Importing and understanding of data
- Missing value check
- Correlation Check
- Data Reduction
- Data Cleaning/Wrangling
- Feature Engineering
- Univariate Analysis
- Bivariate Analysis

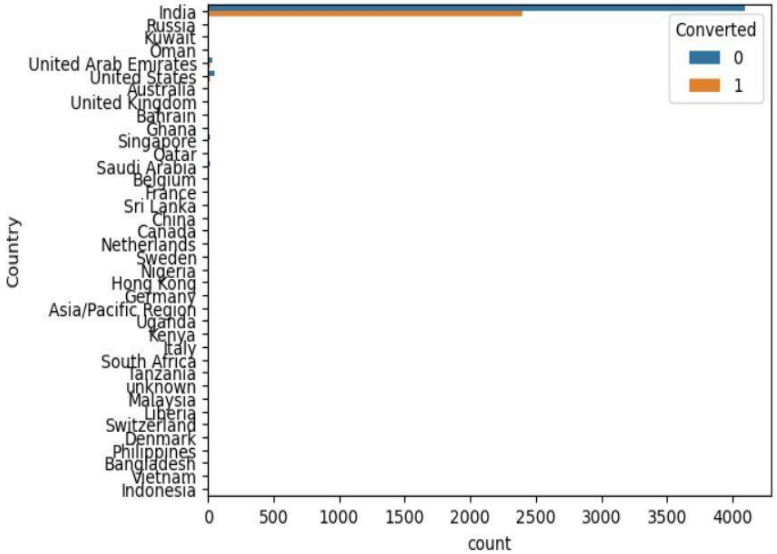

Lead Origin
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
api
landing page submission
lead add form
lead import
Lead Origin

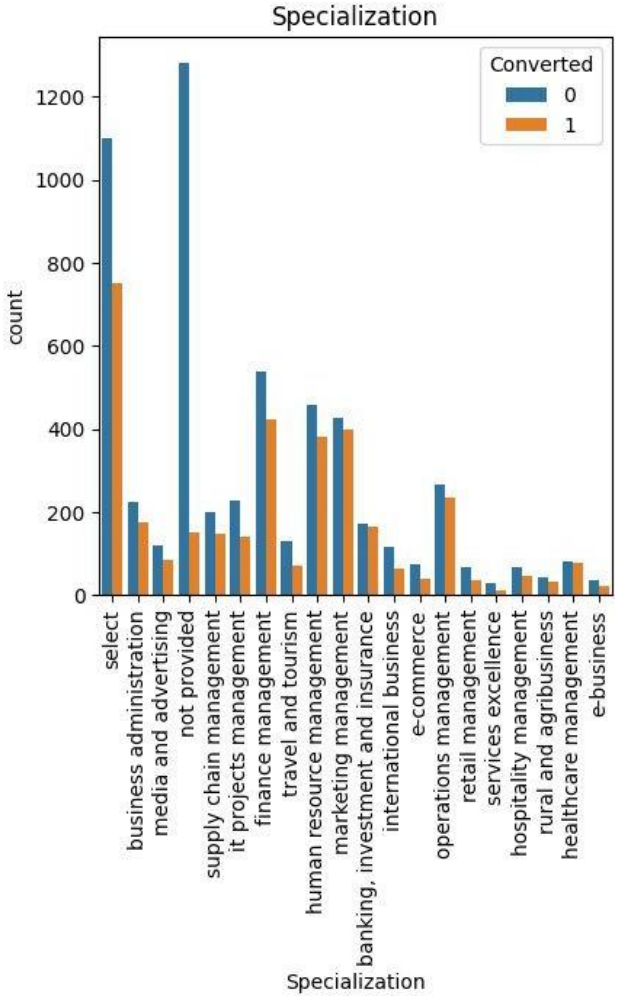

Lead Source
Converted
0
1
count
1750
1500
1250
1000
750
500
250
0
olark chat
organic search
direct traffic
google
referral sites
reference
welingak website
facebook
blog
pay per click ads
bing
social media
welearn
click2call
live chat
welearnblog_home
youtubechannel
testone
press_release
nc_edm
Lead Source

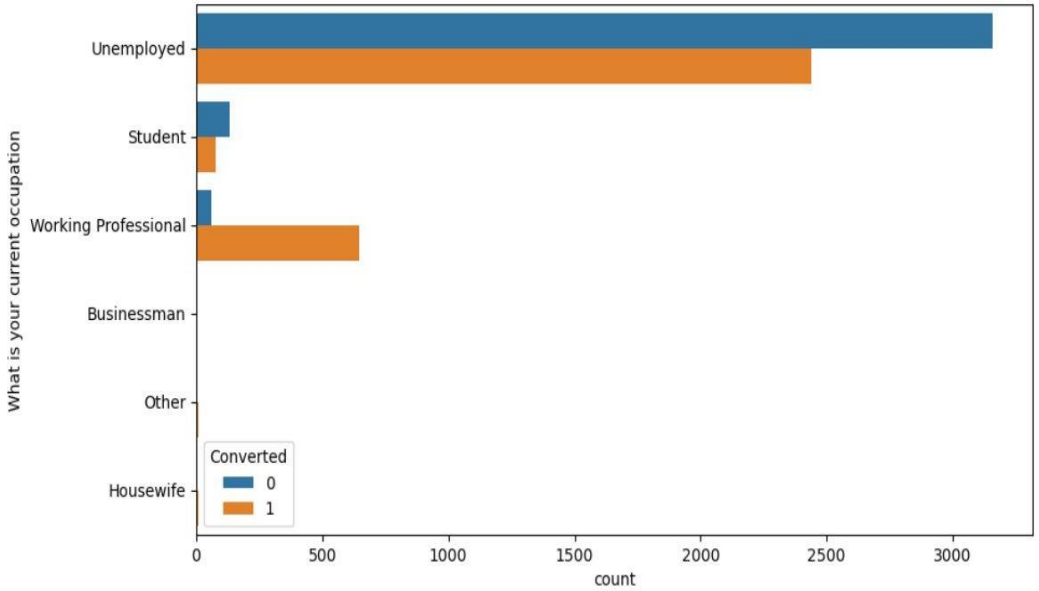

Do Not Email
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Email

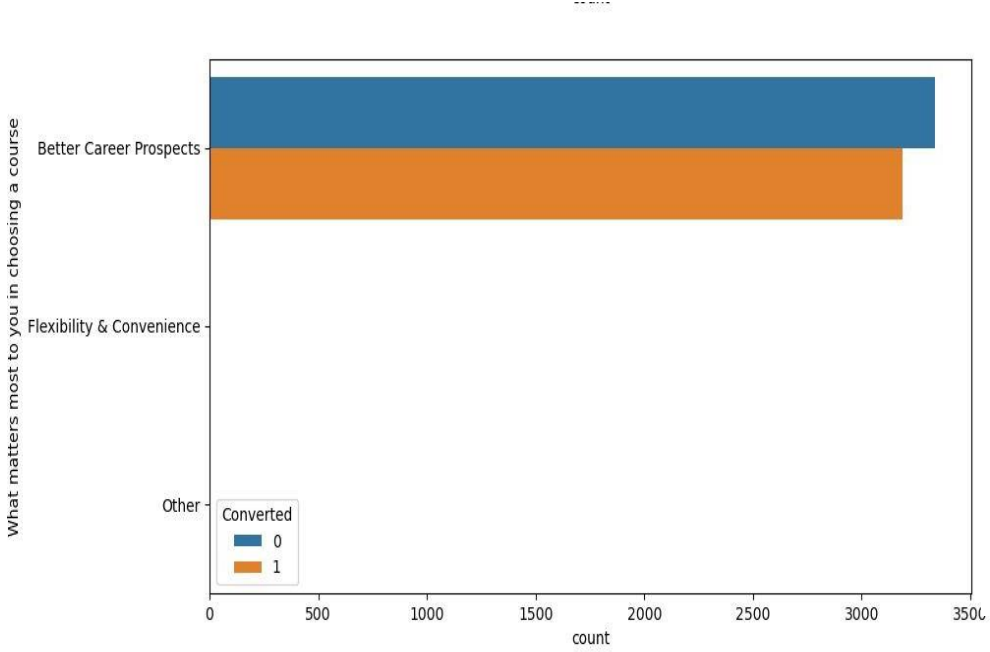

Do Not Call
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Call

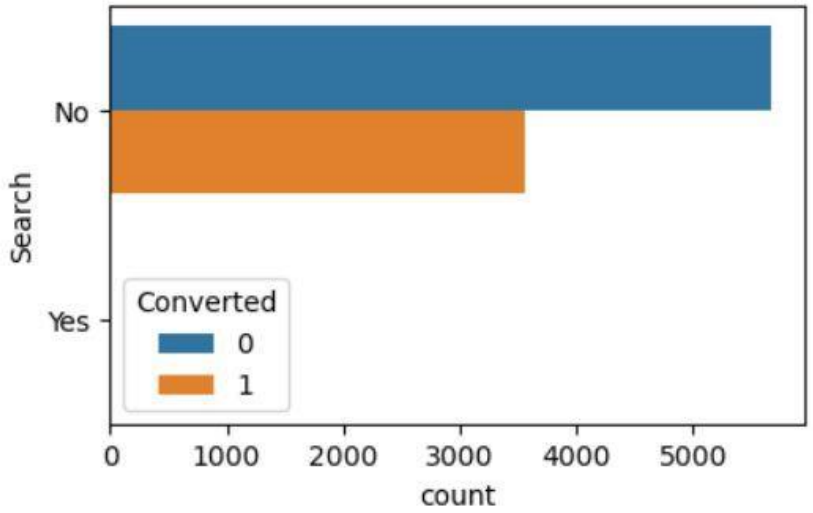

Last Activity
Converted
0
1
count
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow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

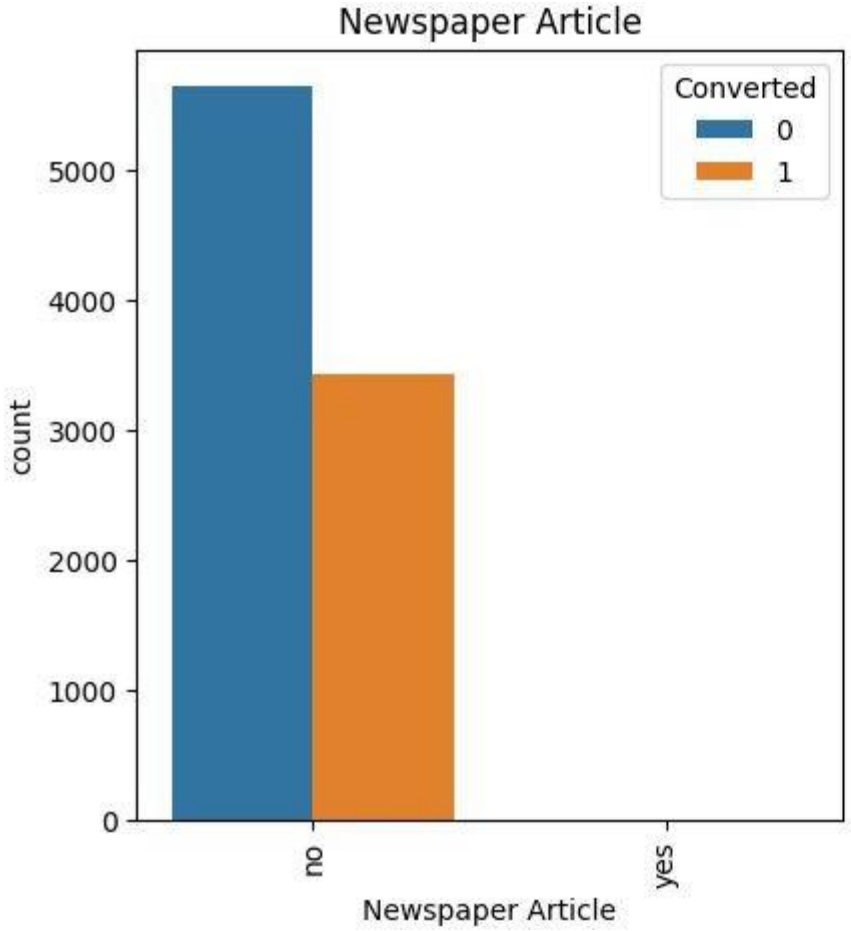

Converted
0
1
Country
India
Russia
Kuwait
Oman
United Arab Emirates
United States
Australia
United Kingdom
Bahrain
Ghana
Singapore
Qatar
Saudi Arabia
Belgium
France
Sri Lanka
China
Canada
Netherlands
Sweden
Nigeria
Hong Kong
Germany
Asia/Pacific Region
Uganda
Kenya
Italy
South Africa
Tanzania
unknown
Malaysia
Liberia
Switzerland
Denmark
Philippines
Bangladesh
Vietnam
Indonesia
0
500
1000
1500
2000
2500
3000
3500
4000
count

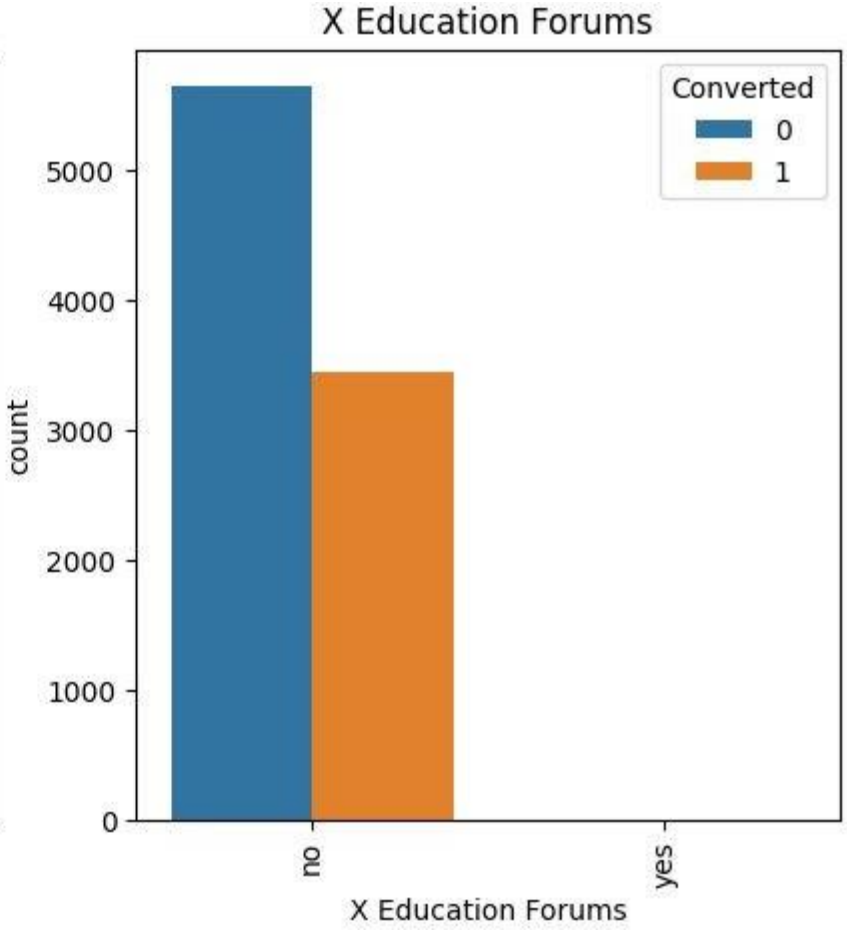

Specialization
Converted
0
1
count
1200
1000
800
600
400
200
0
select
business administration
media and advertising
not provided
supply chain management
it projects management
finance management
travel and tourism
human resource management
marketing management
banking, investment and insurance
international business
e-commerce
operations management
retail management
services excellence
hospitality management
rural and agribusiness
healthcare management
e-business
Specialization

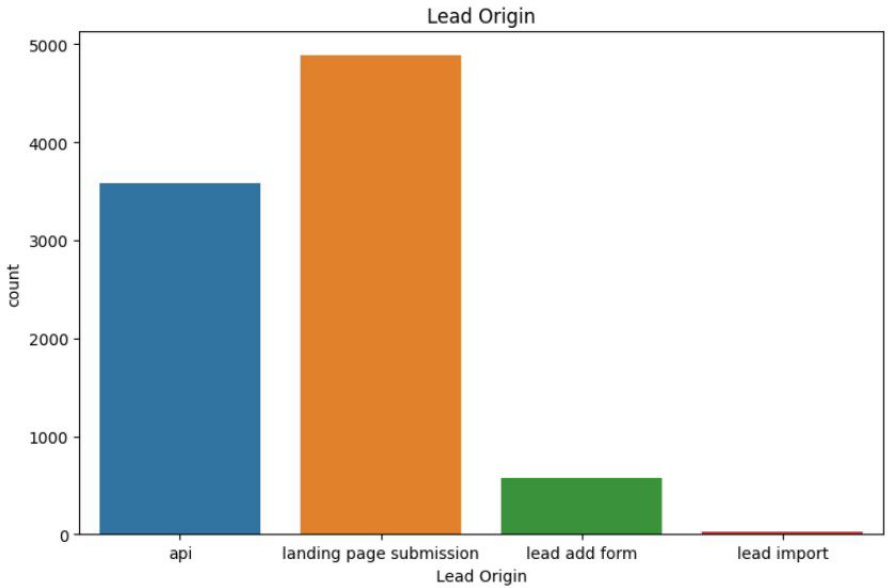

What is your current occupation
Unemployed
Student
Working Professional
Businessman
Other
Housewife
Converted
0
1
0
500
1000
1500
2000
2500
3000
count

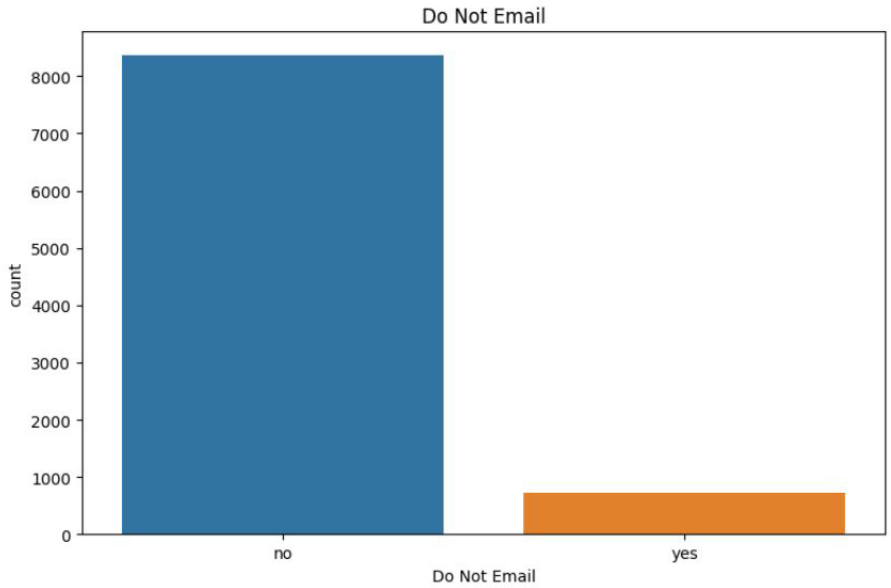

What matters most to you in choosing a course
Better Career Prospects
Flexibility & Convenience
Other
Converted
0
1
0
500
1000
1500
2000
2500
3000
count

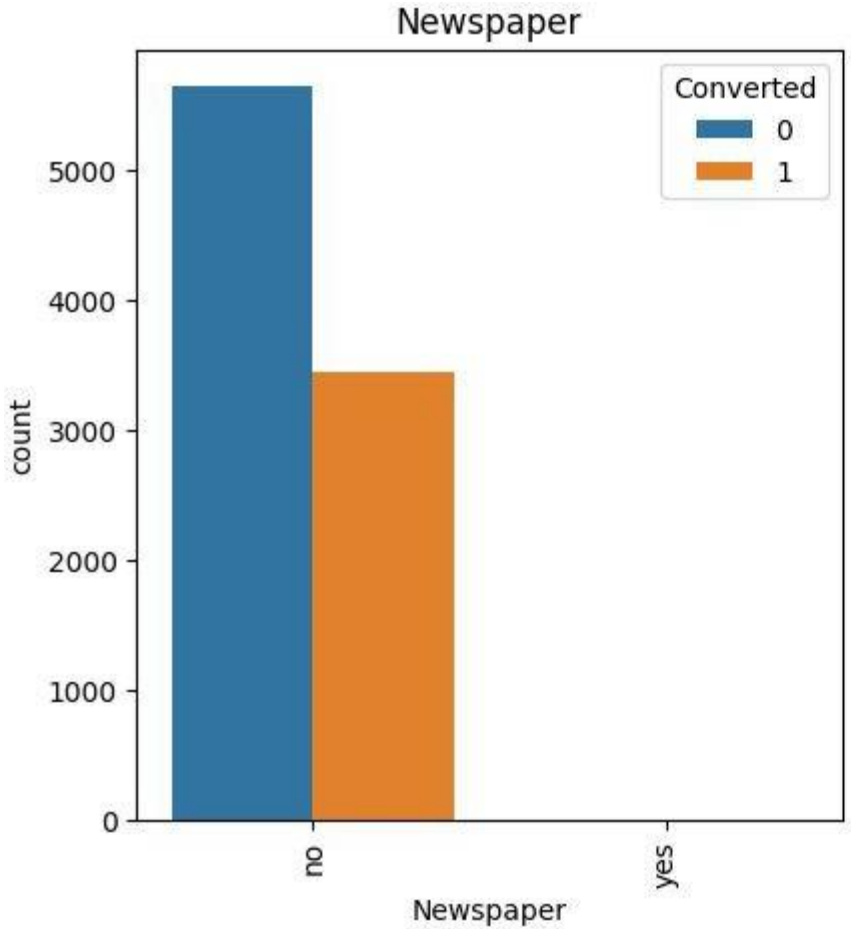

Search
No
Yes
Converted
0
1
0
1000
2000
3000
4000
5000
count

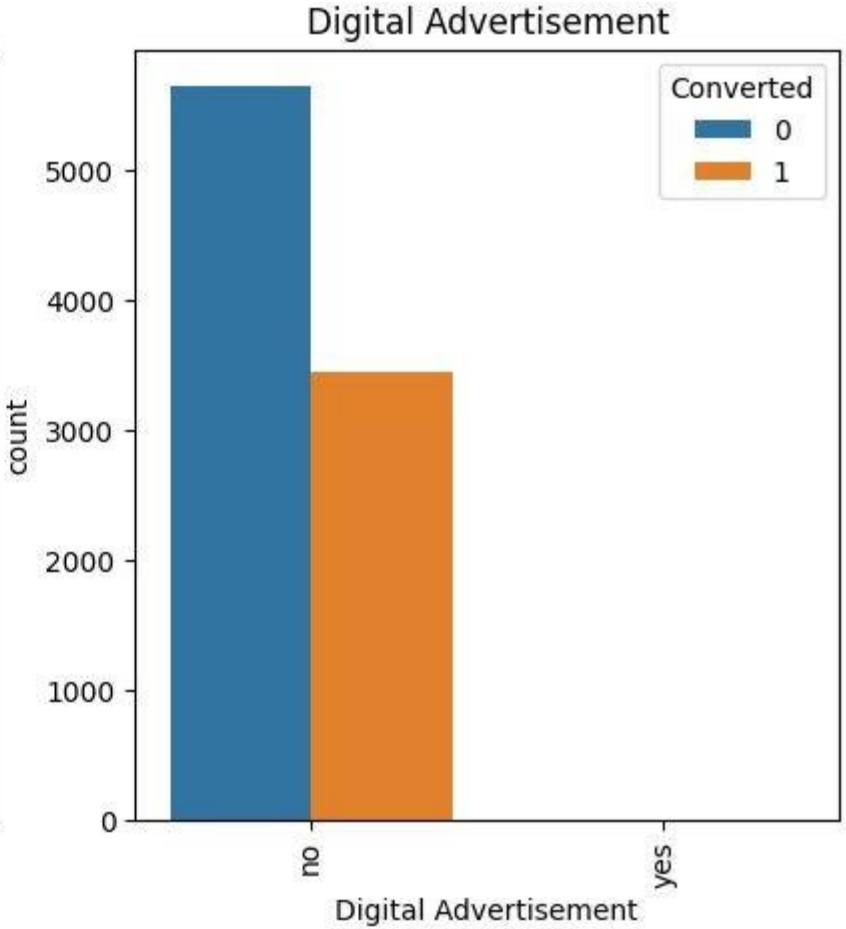

Newspaper Article
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
Newspaper Article

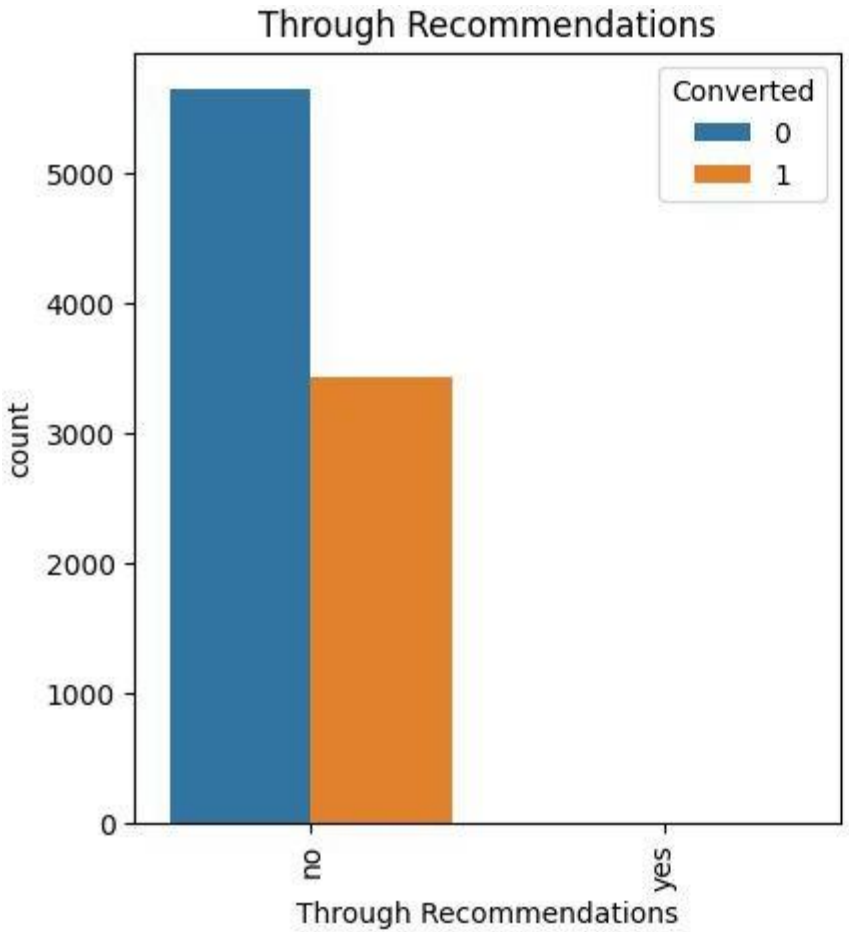

X Education Forums
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
X Education Forums

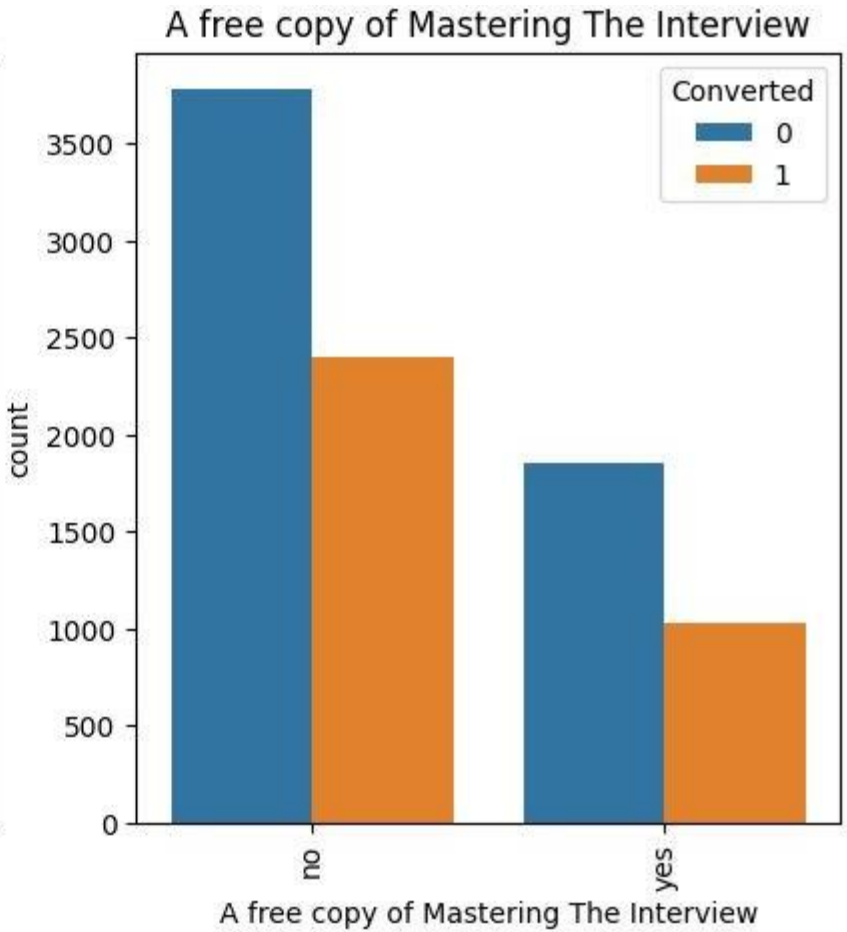

Lead Origin
5000
4000
3000
2000
1000
0
count
api
landing page submission
lead add form
lead import
Lead Origin

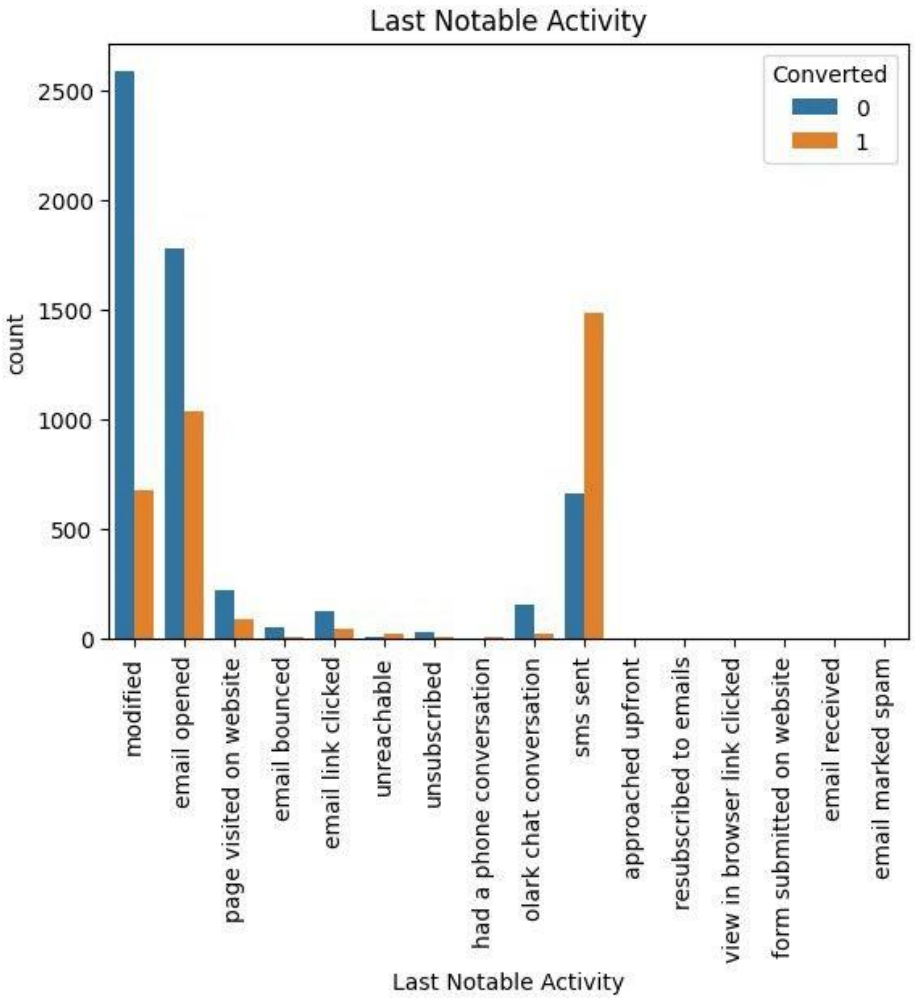

Do Not Email
8000
7000
6000
5000
4000
3000
2000
1000
0
count
no
yes
Do Not Email

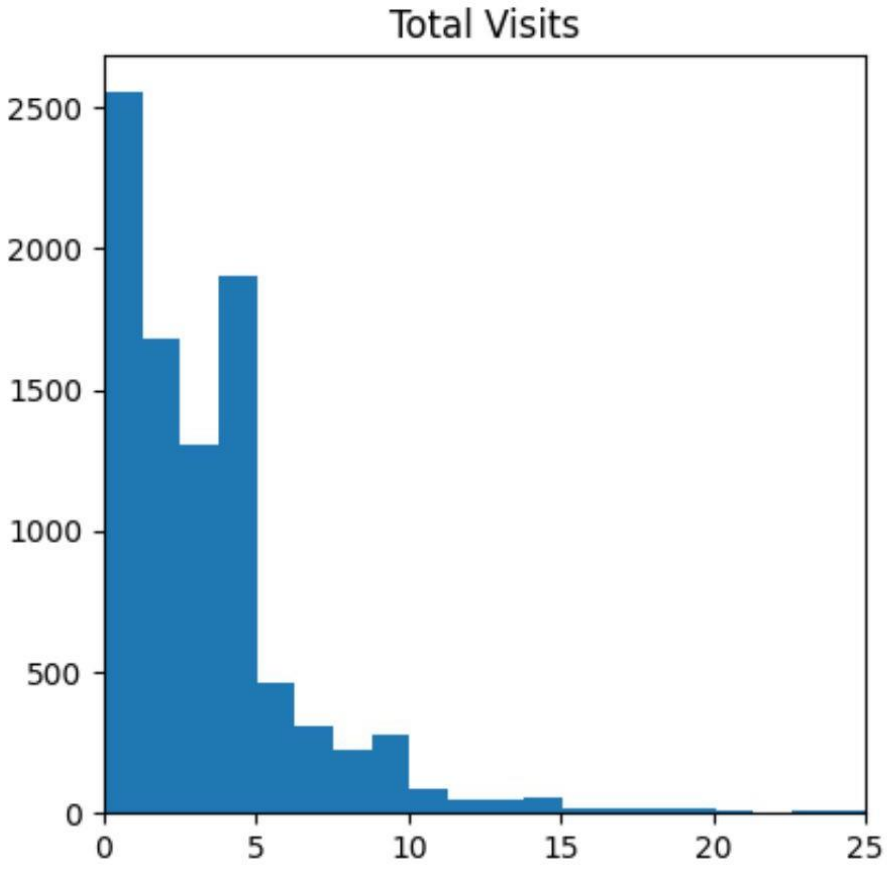

Newspaper
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
Newspaper

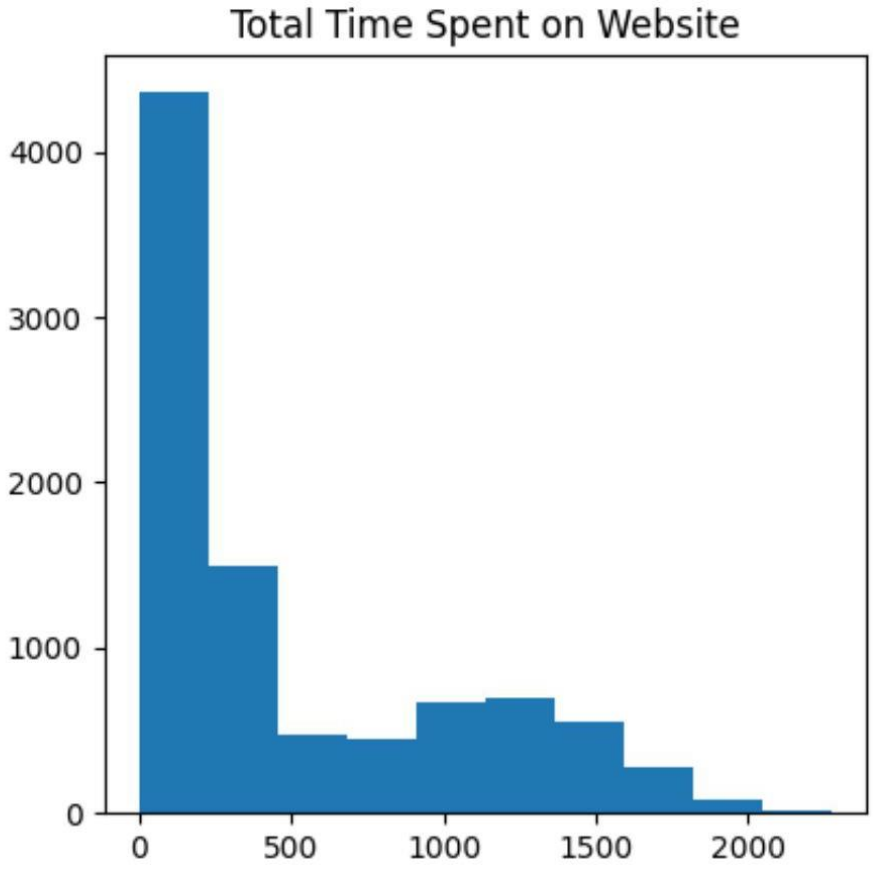

Digital Advertisement
Converted
0
1
5000
4000
3000
2000
1000
0
count
no
yes
Digital Advertisement


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


Total Visits
2500
2000
1500
1000
500
0
0
5
10
15
20
25

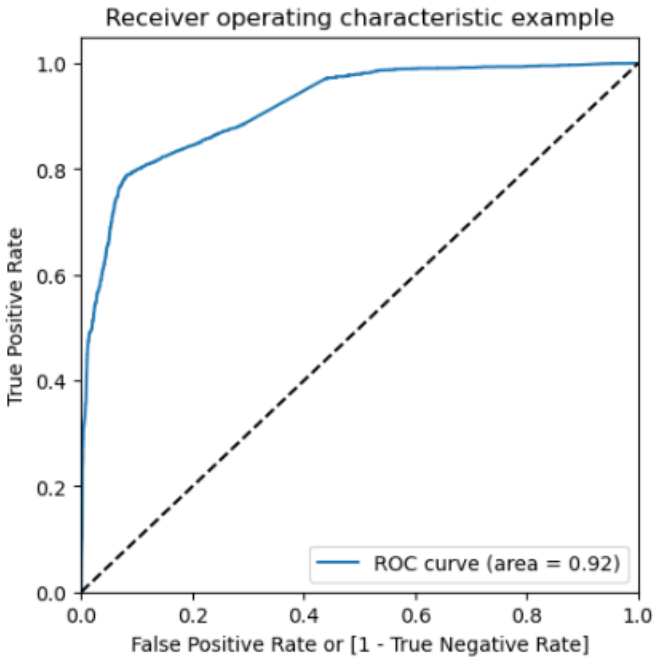

Total Time Spent on Website
4000
3000
2000
1000
0
0
500
1000
1500
2000

Model Building

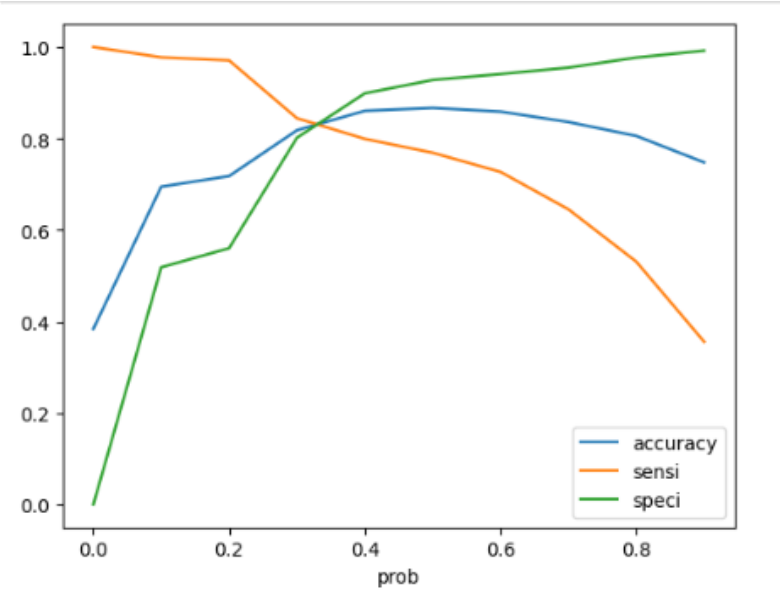
- Splitting data into training and testing data
- Feature Scaling
- Feature Selection
- Building and training the model
- Variance Inflation Factor
- Making predictions from the model
- Testing the performance of the model
- ROC curve
- Accuracy Score
- Confusion Metrics
- Precision and Recall

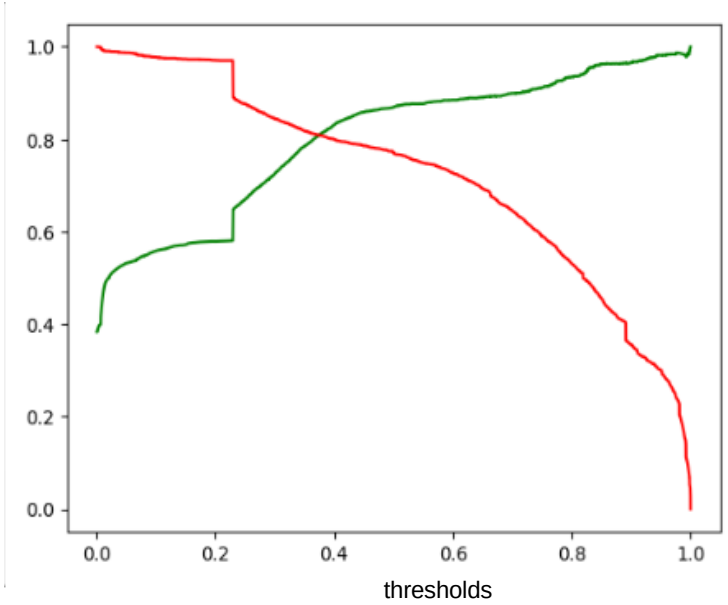
ROC Curve

- Finding Optimal Cut off Point
- Optimal cut off probability is that
- probability where we get balanced sensitivity and specificity.
- From the second graph it is visible that the optimal cut off is at 0.35.



Optimal Cut Off





Conclusion

X Education should focus on clientele meeting the following criteria::

- If someone has interacted with Horizon, lost tags to Eins, and reads our email, they're more likely to convert.
- People who spend a lot of time on our platform could become clients.
- Leads from the Welingak website, especially high school students using our forms, are more likely to convert.
- Working professionals and the unemployed looking to upgrade themselves are great leads for X Education.
- Those who ask X Education to email them are more likely to become customers.

Thankyou