Lead Scoring case study

GROUP MEMBERS:

NEHA YADAV, AKULJOT SINGH MAKKERR , AAKANKSHA MISHRA

Problem Statement

- X Education , An education company named sells online courses to industry professionals.
- Many interested professionals land on their website.
- X Education gets a lot of leads but 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'.
- 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:

- X education wants to know most promising leads.
- For that they want to build a model which identifies the hot leads.
- Deployment of the model for future use.

Proposed Solution

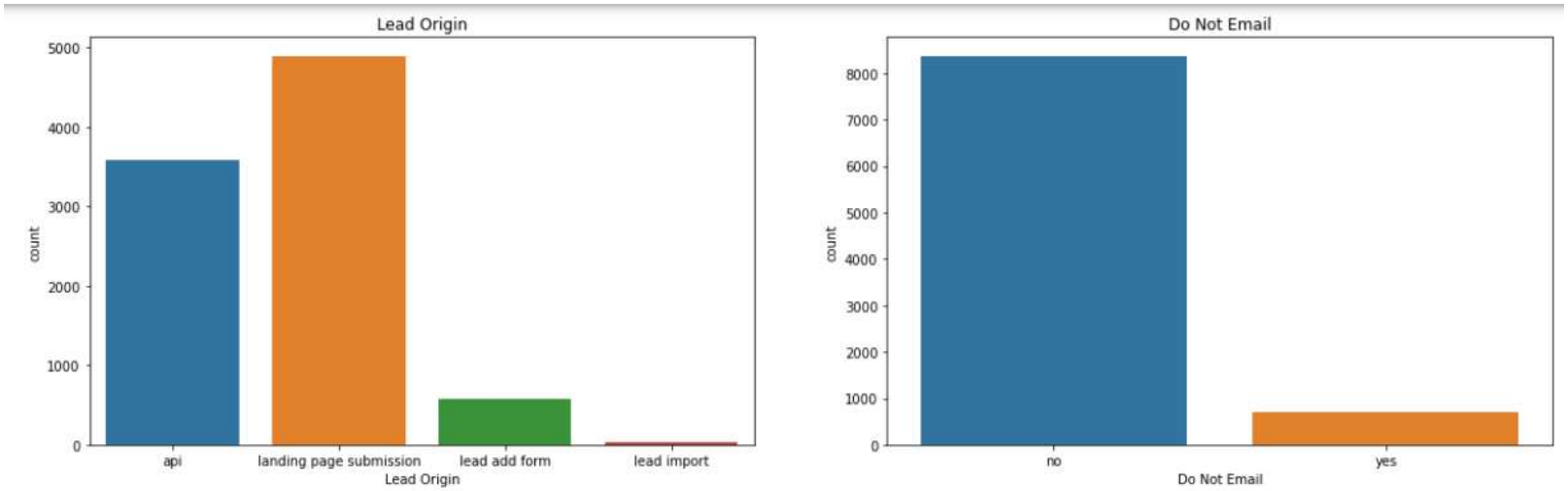
- For our Problem Solution, the crucial part is to accurately identify hot leads.
- The more accurate we obtain the hot lead, the more chance we get of higher conversion ratio.
- Since we have a target of 80% conversion rate, we would want to obtain a high accuracy in obtaining hot leads.

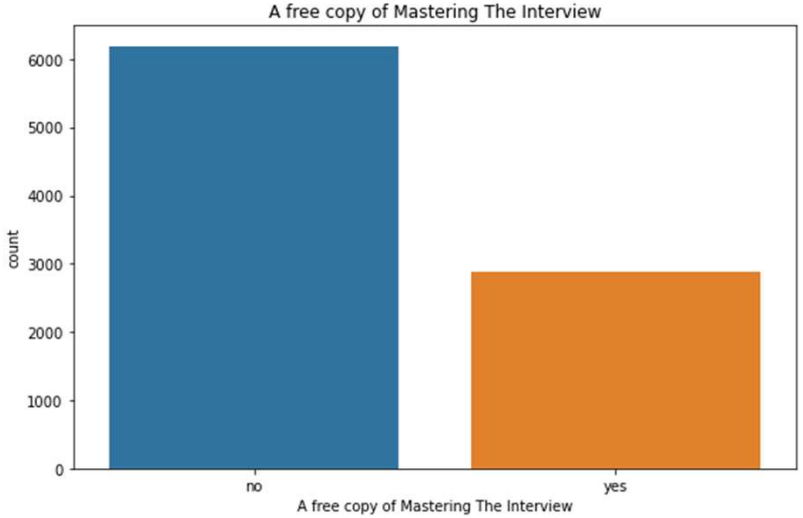
Implementation of Solution

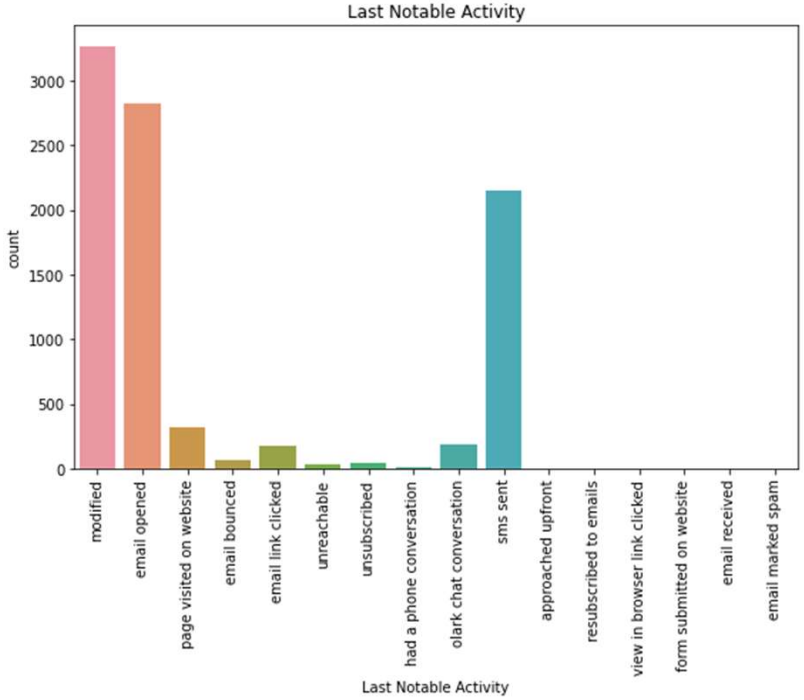
- Step 1: Data Cleaning & Data Manipulation
- Step 2: EDA
- Step 3: Data Preparation
- Step 4: Feature Selection
- Step 5: Model Building
- Step 6: Verifying Model

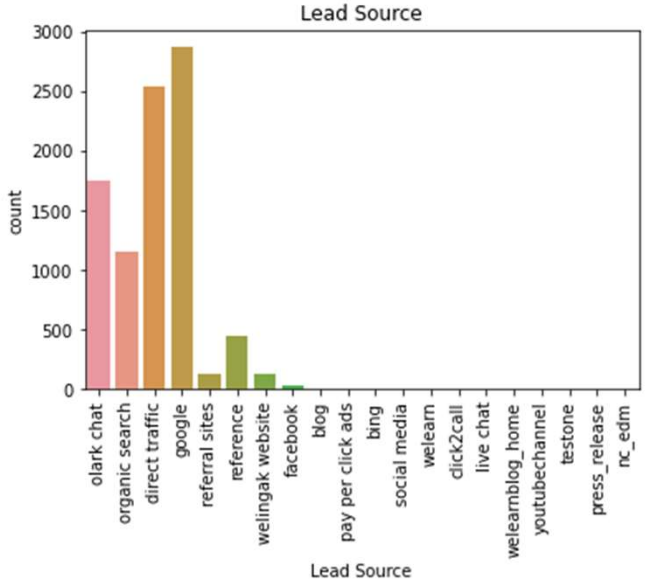
EDA - Plots

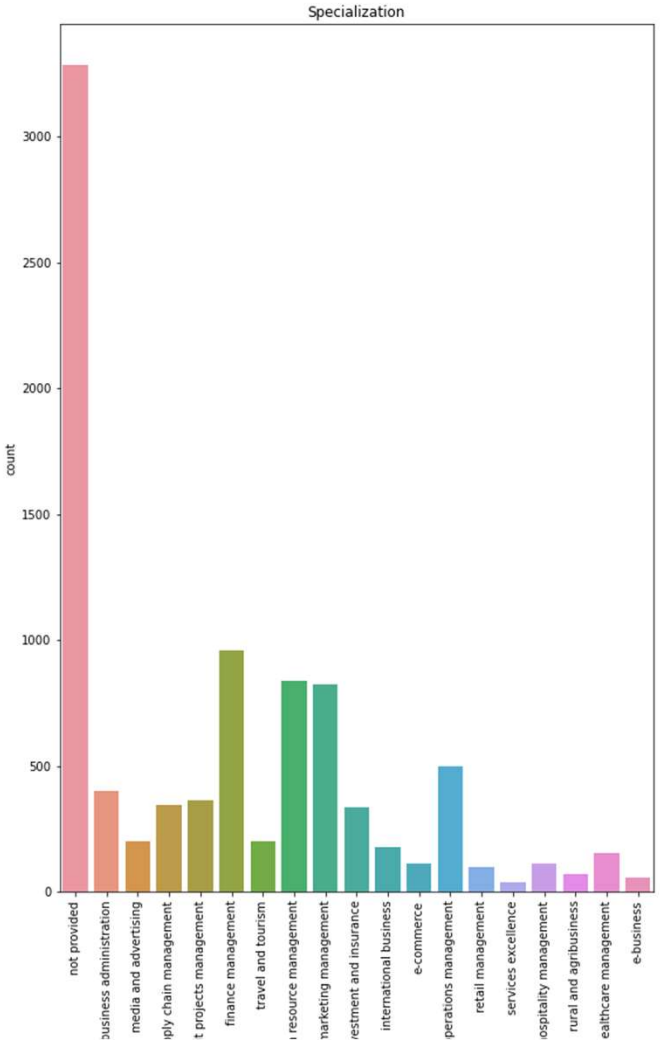
- **Univariate Analysis**
- 1. Count Plots

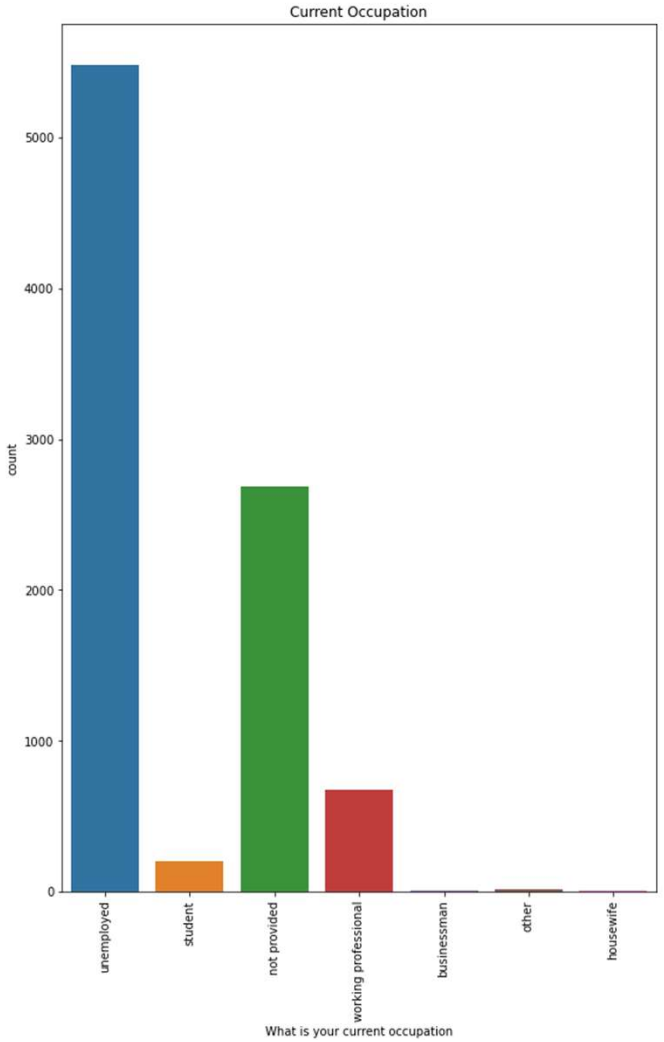


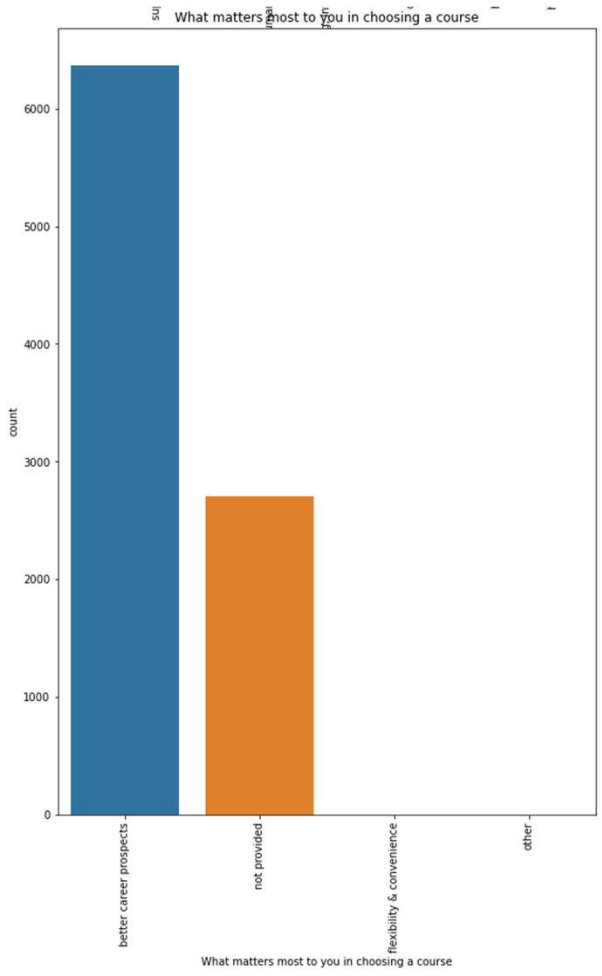

A free copy of Mastering The Interview
count
6000
5000
4000
3000
2000
1000
0
no
yes
A free copy of Mastering The Interview

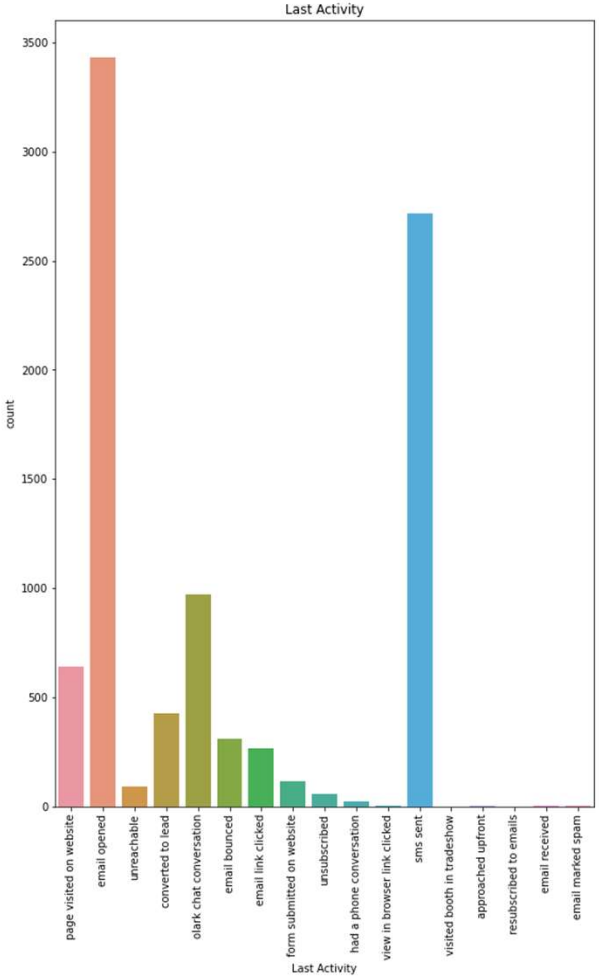

Last Notable Activity
count
3000
2500
2000
1500
1000
500
0
modified
email opened
page visited on website
email bounced
email link clicked
unreachable
unsubscribed
had a phone conversation
olark chat conversation
sms sent
approached upfront
resubscribed to emails
view in browser link clicked
form submitted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

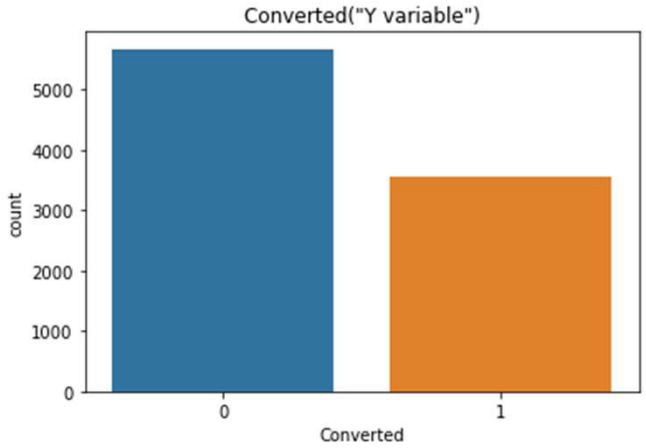

Lead Source
3000
2500
2000
1500
1000
500
0
count
olark chat
organic search
direct traffic
google
referral sites
reference
welingak website
facebook
blog
pay per click ads
bing
social media
welearn
click2call
live chat
welearnblog_home
youtubechannel
testone
press_release
nc_edm
Lead Source

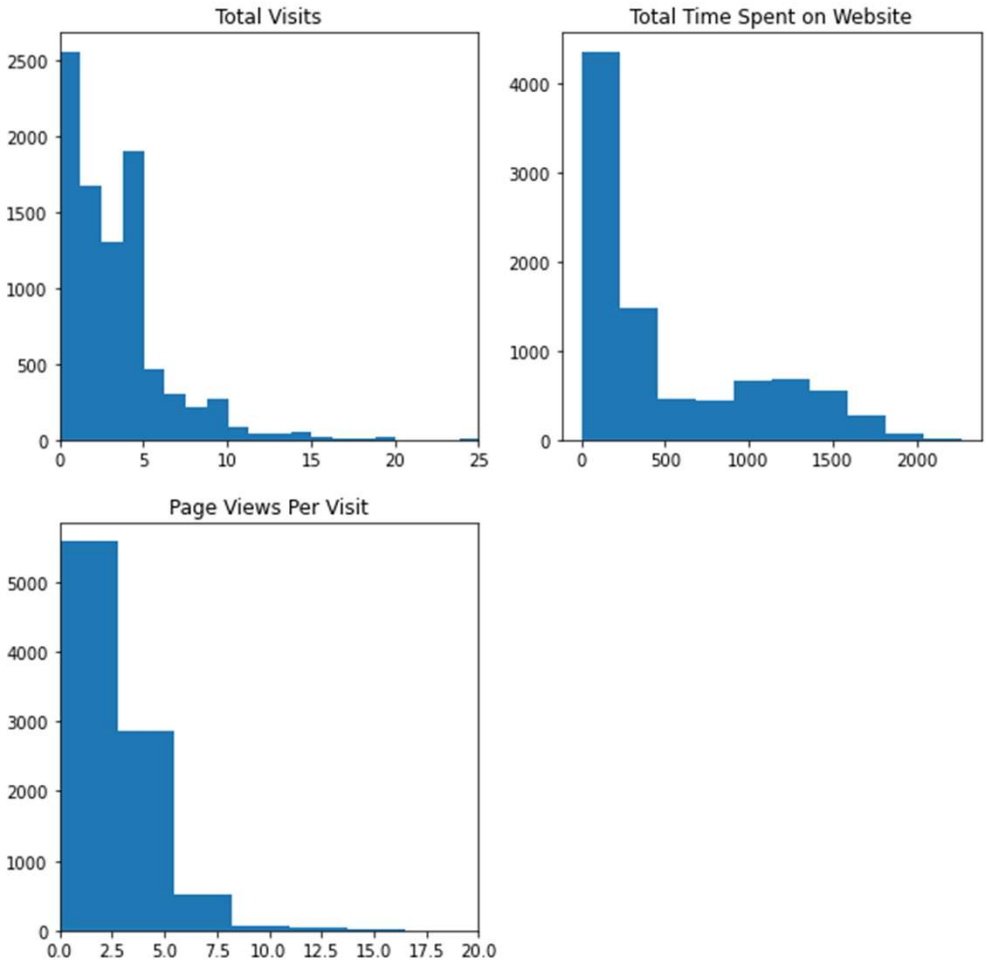

Specialization
count
3000
2500
2000
1500
1000
500
0
not provided
usiness administration
media and advertising
ply chain management
t projects management
finance management
travel and tourism
resource management
marketing management
estment and insurance
international business
e-commerce
perations management
retail management
services excellence
ospitality management
rural and agribusiness
ealthcare management
e-business


Current Occupation
count
5000
4000
3000
2000
1000
0
unemployed
student
not provided
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
not provided
flexibility & convenience
other
What matters most to you in choosing a course

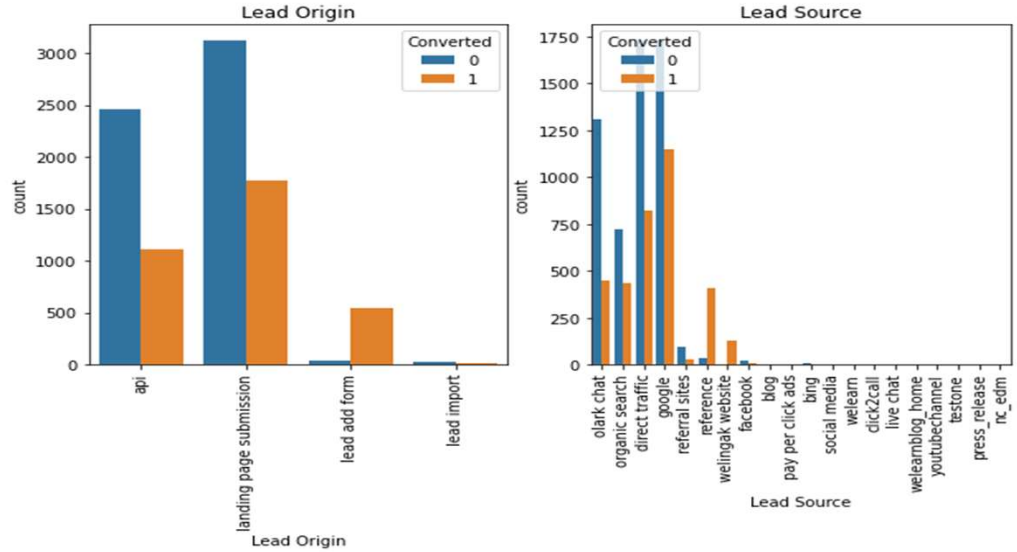

Last Activity
count
3500
3000
2500
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow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

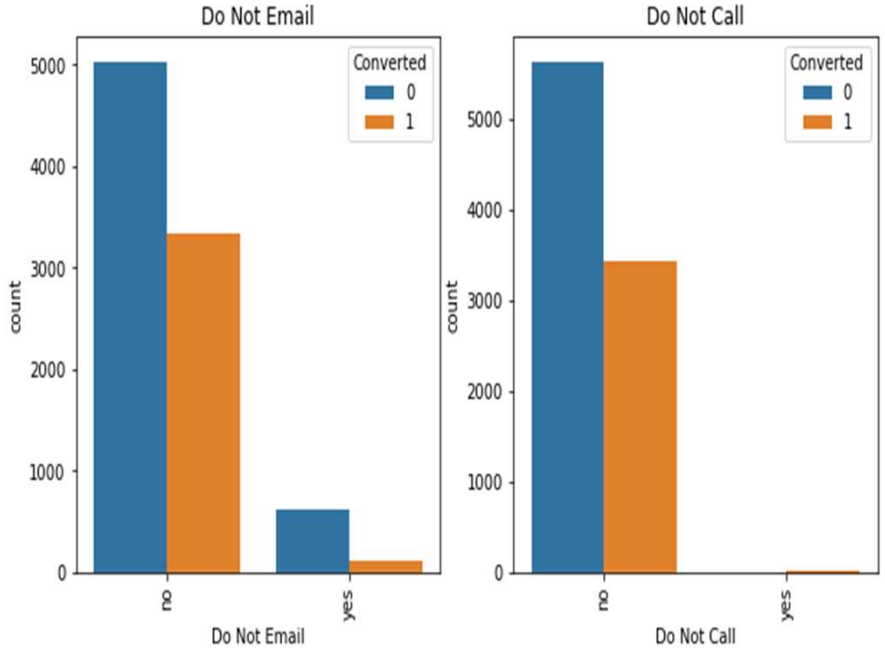

Converted("Y variable")
count
5000
4000
3000
2000
1000
0
0
1
Converted


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
Page Views Per Visit
5000
4000
3000
2000
1000
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

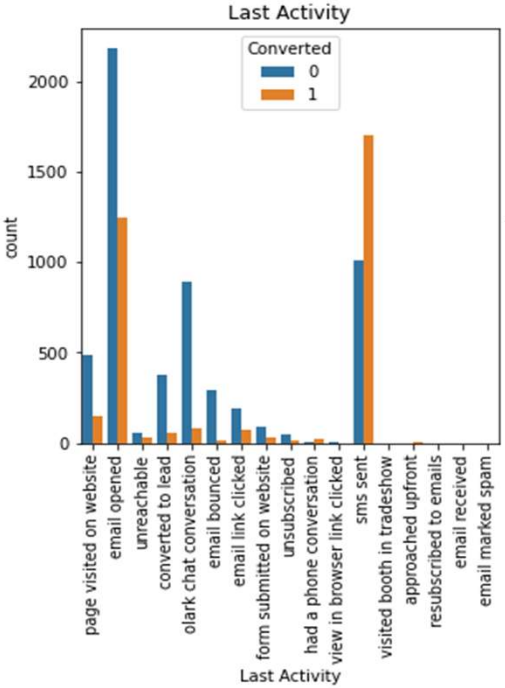
EDA - Plots

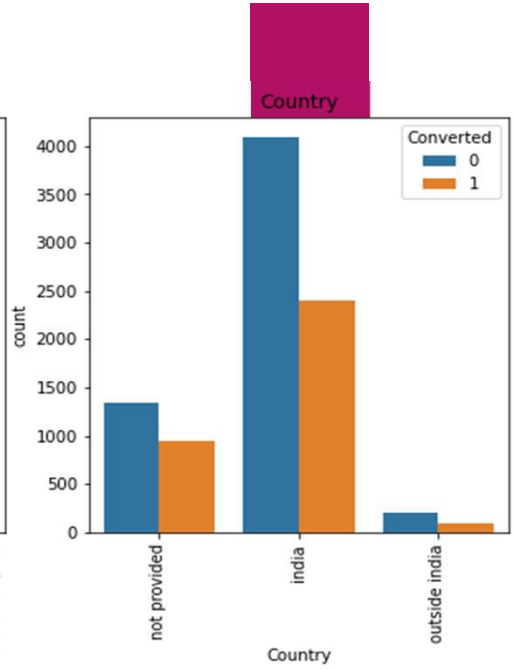
- **Categorical variables to Converted**

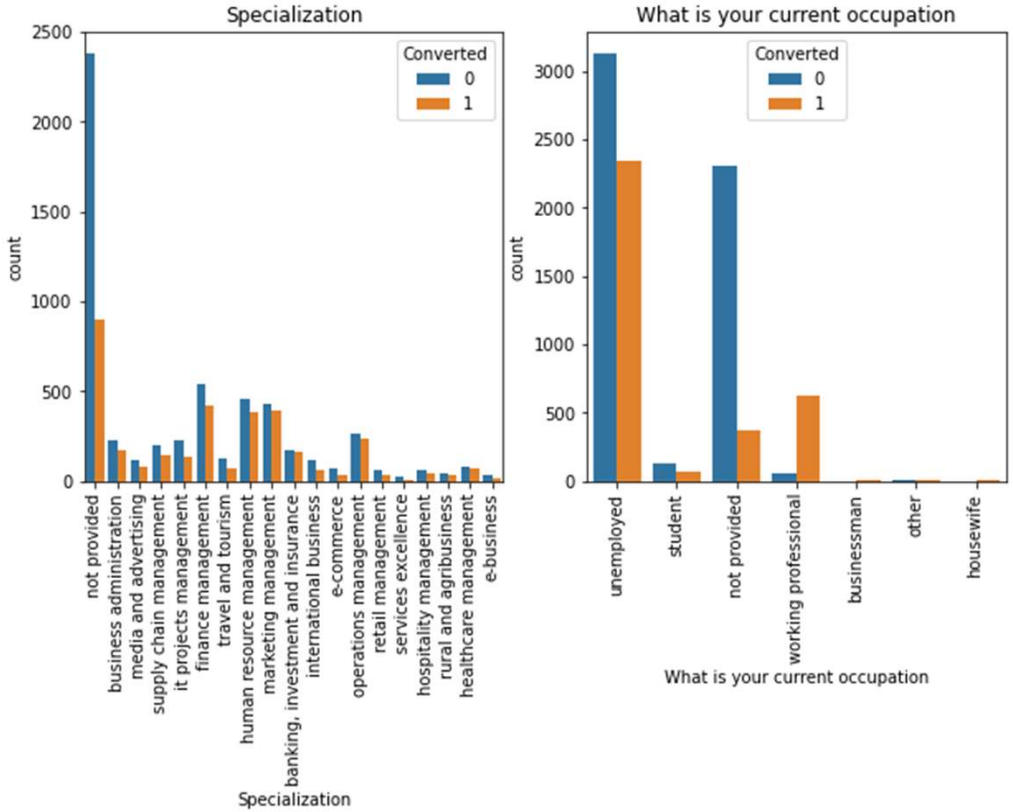



Do Not Email
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Email

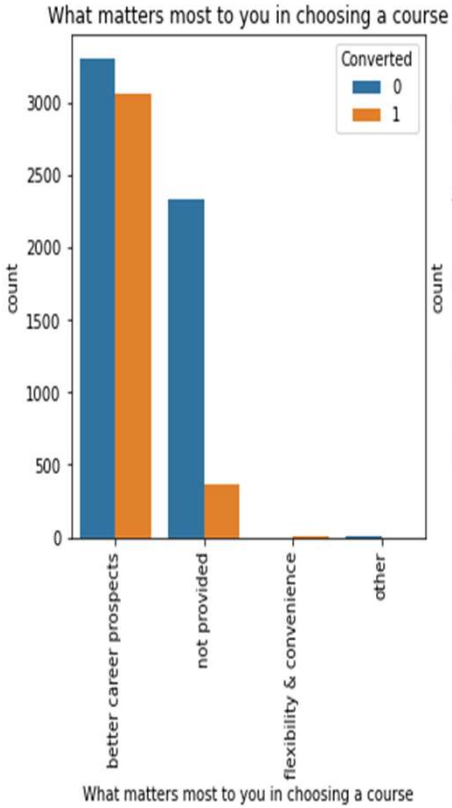
Do Not Call
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Do Not Call

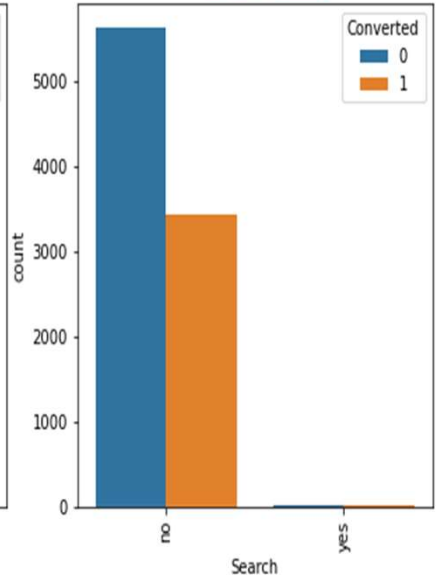

Last Activity
Converted
0
1
count
2000
1500
1000
500
0
page visited on website
email opened
unreachable
converted to lead
olark chat conversation
email bounced
email link clicked
form submitted on website
unsubscribed
had a phone conversation
view in browser link clicked
sms sent
visited booth in tradeshow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

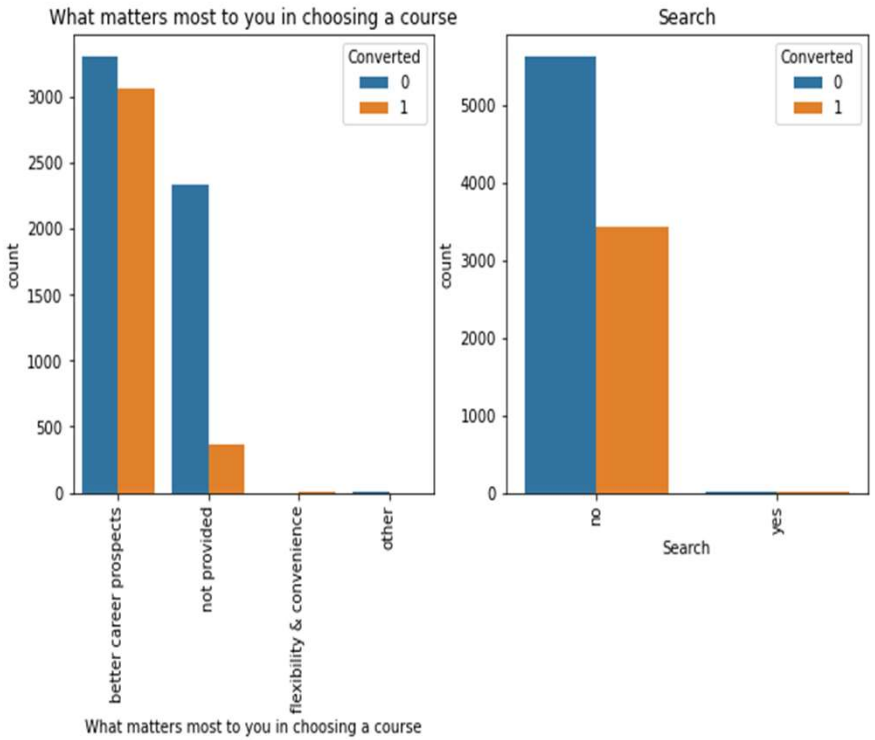

Country
Converted
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
not provided
india
outside india
Country

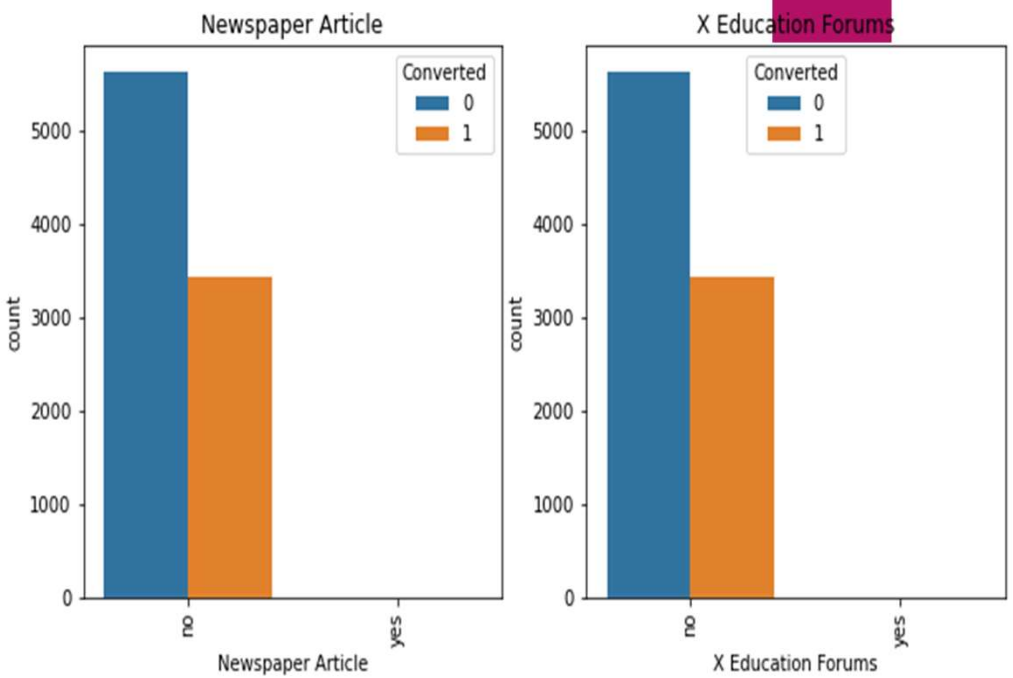

Specialization
Converted
0
1
count
not provided
business administration
media and advertising
supply chain management
it projects management
finance management
travel and tourism
human resource management
marketing management
banking, investment and insurance
international business
e-commerce
operations 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
not provided
working professional
businessman
other
housewife
What is your current occupation

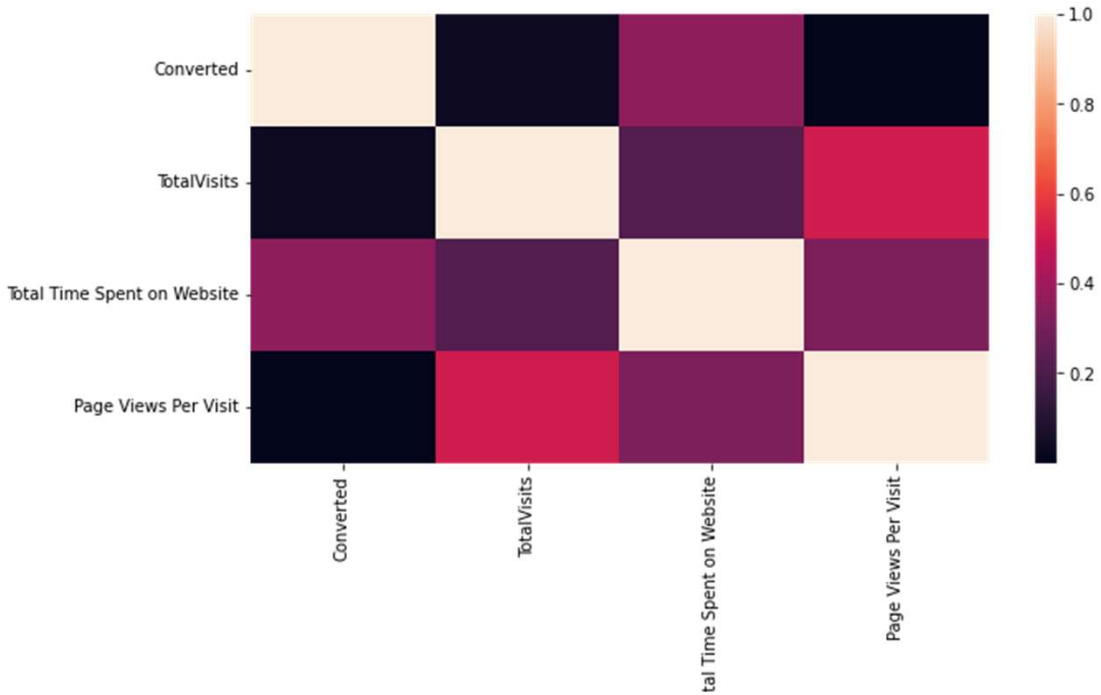

What matters most to you in choosing a course
Converted
0
1
count
better career prospects
not provided
flexibility & convenience
other
What matters most to you in choosing a course


Converted
0
1
count
no
yes
Search

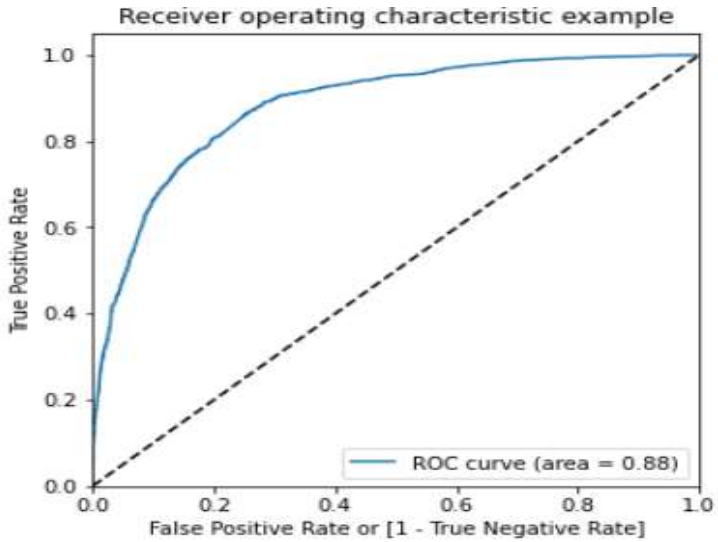

What matters most to you in choosing a course
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
better career prospects
not provided
flexibility & convenience
other
What matters most to you in choosing a course
Search
Converted
0
1
count
5000
4000
3000
2000
1000
0
no
yes
Search


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

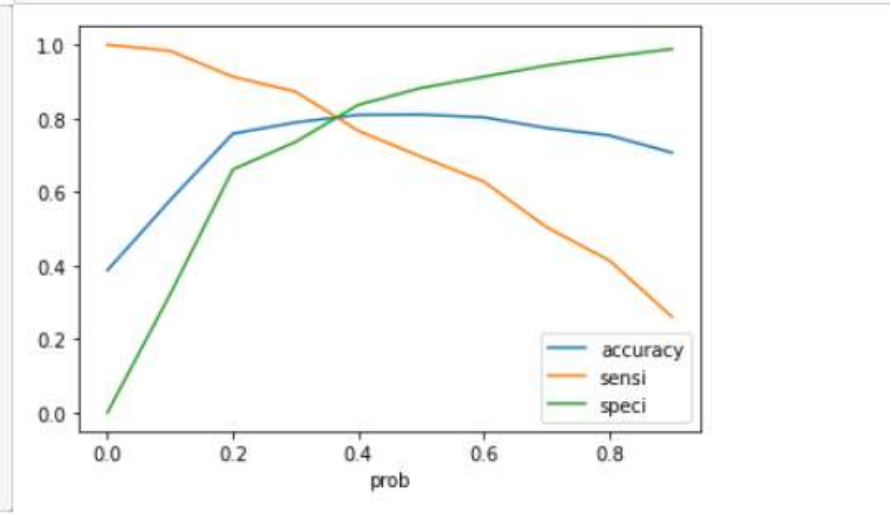
Heat Map:


Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
Converted
TotalVisits
tal Time Spent on Website
Page Views Per Visit
1.0
0.8
0.6
0.4
0.2

Optimise Cut off (ROC Curve)



The area under ROC curve is 0.88 which is a very good value.



From the graph it is visible that the optimal cut off is somewhat at 0.35.

Final Observation:

Final Observation Using Sensitivity – Specificity model:

Let us compare the values obtained for Train & Test:

Train Data:

- Accuracy 80.31%
- Sensitivity 80.37 %
- Specificity 80.28%.

Test Data:

- Accuracy 80.79%
- Sensitivity 81.03%
- Specificity 80.50%

Final Observation Using Precision – Recall:

Let us compare the values obtained for Train & Test:

Train Data:

- Accuracy 81.10%
- Precision 75.40%
- Recall 75.89%

Test Data:

- Accuracy 81.52%
- Precision 73.24%
- Recall 76.60%