# LEAD SCORE CASE STUDY

BY-
DIVYANSH RANA &
SWETANK KUNWAR

# PROBLEM STATEMENT

- The typical lead conversion rate at X education is around 30%.
- There are a lot of leads generated in the initial stage, but only a few of them come out as paying customers. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc. ) in order to get a higher lead conversion.
- X Education has appointed us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

# AGENDA OF THE CASE STUDY

- . We will  **build a model wherein we need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and vice versa.**
- Goal is to 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 our model should be able to adjust to if the company's requirement changes in the future .

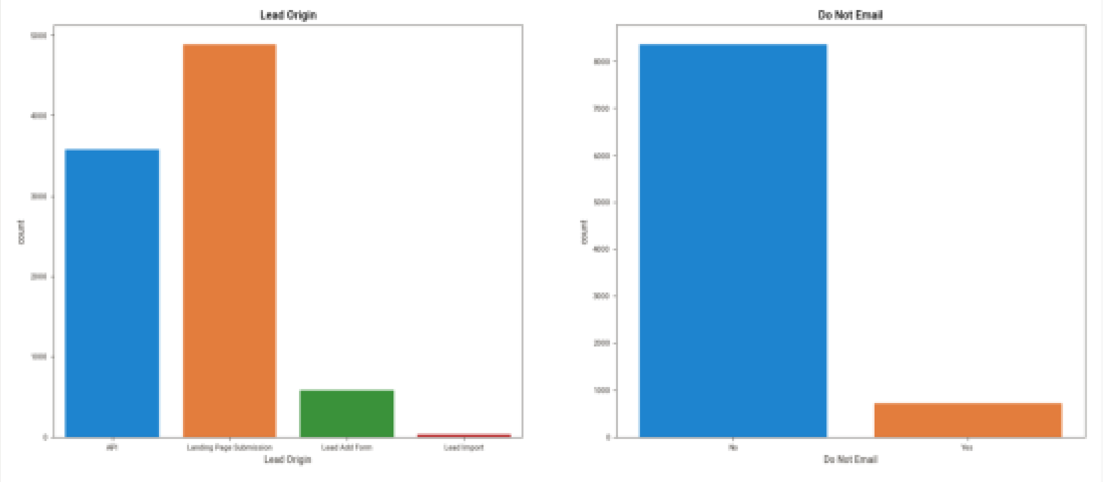
# 1. Data Cleaning

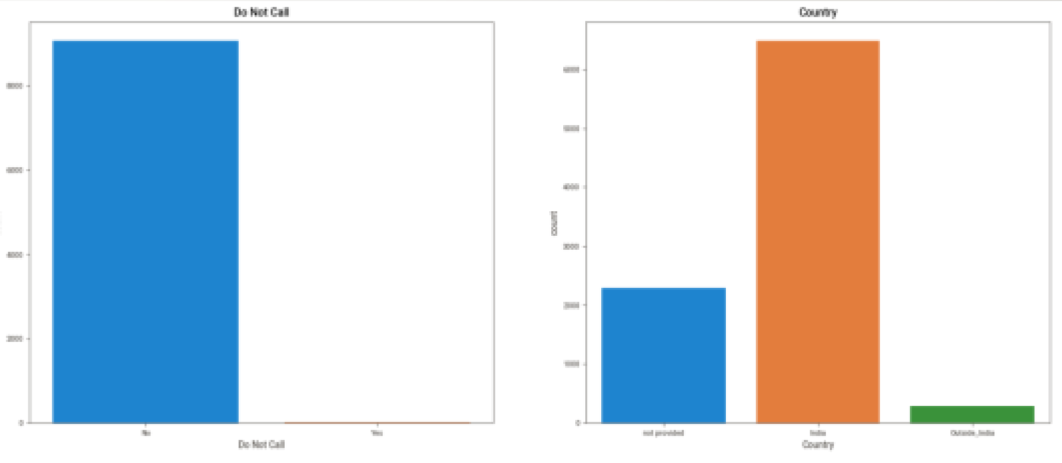
1. Use Sweet viz library we got to know that there is select value which would be filter's select so replace that by nan and lower case select 's s alphabet.
2. Checked for distinct value and remove them as they  can't be used for  analysis .
3. Handle duplicate Value
4. We have dropped columns which has more that 35% row are null.
5. We have replaced few null values by not provided as those columns are required for further analysis.
6. For other countries we renamed them with outside of India as we have very less result for outside of India country.

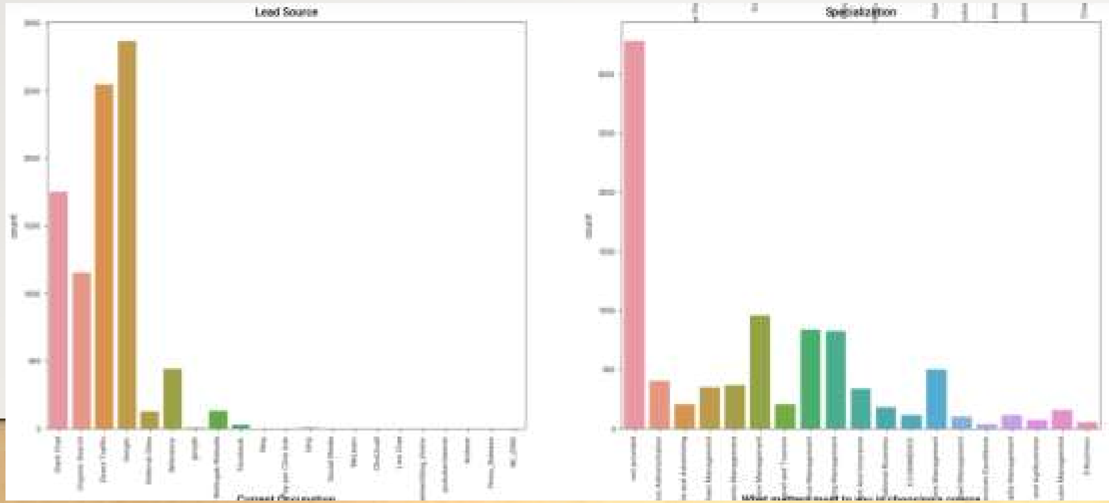
# 2. Exploratory Data Analysis

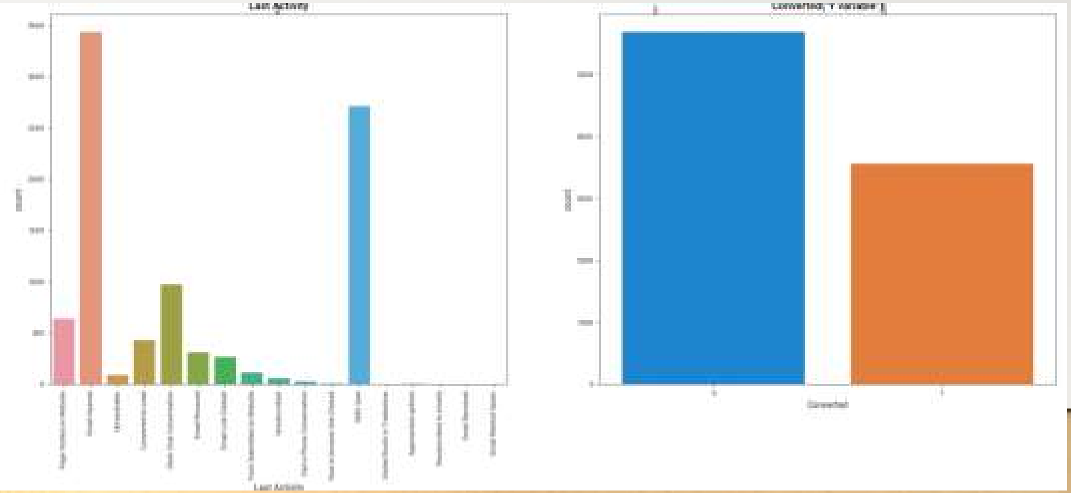
## 2.1 Univariate Analysis

### a) Count the Value.

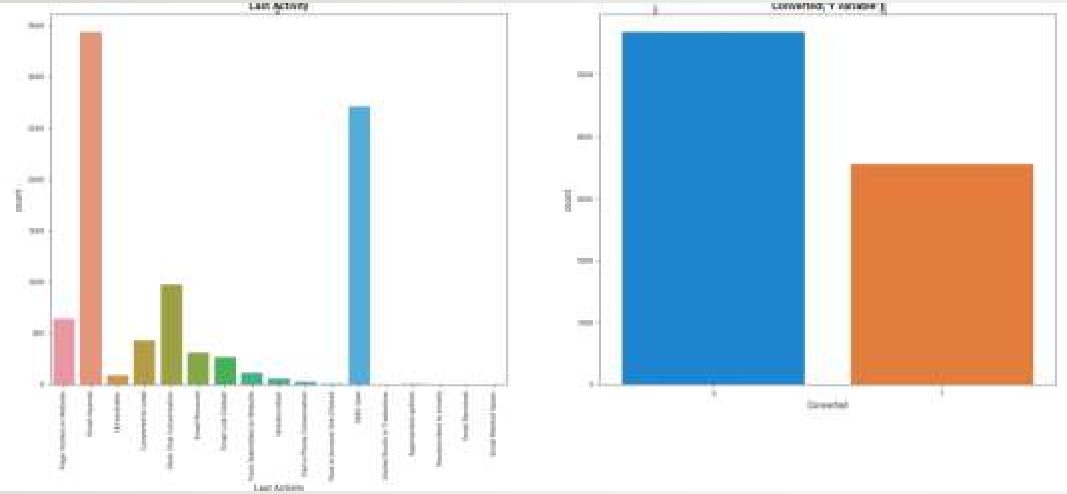




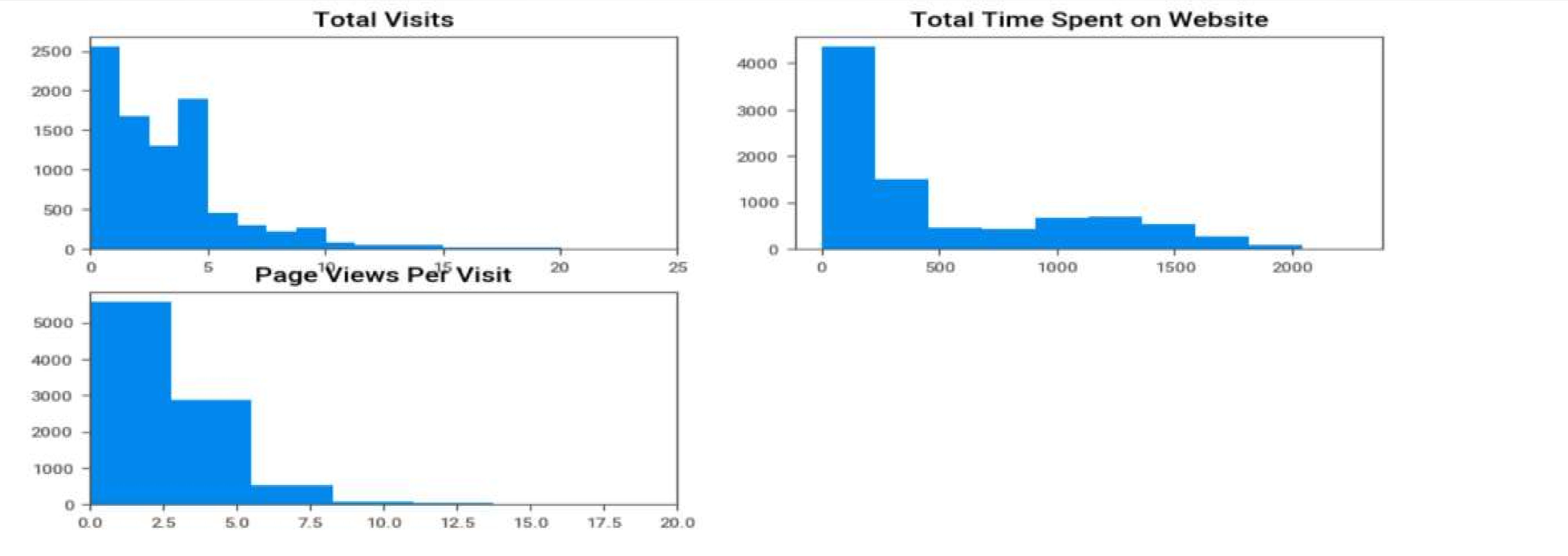




1. From the previous slide we can observe that X education system has very less successful rate it is near to 35

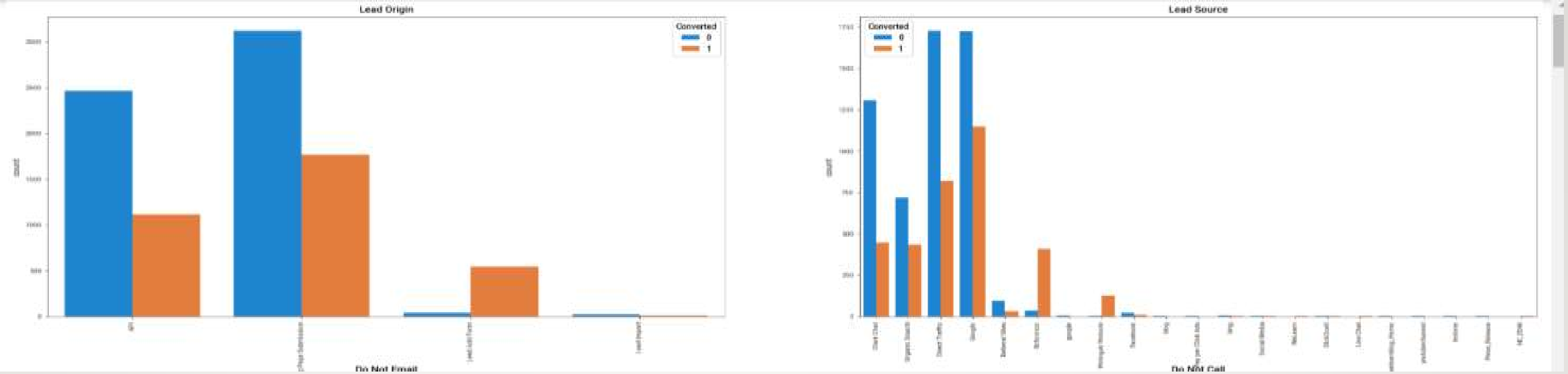


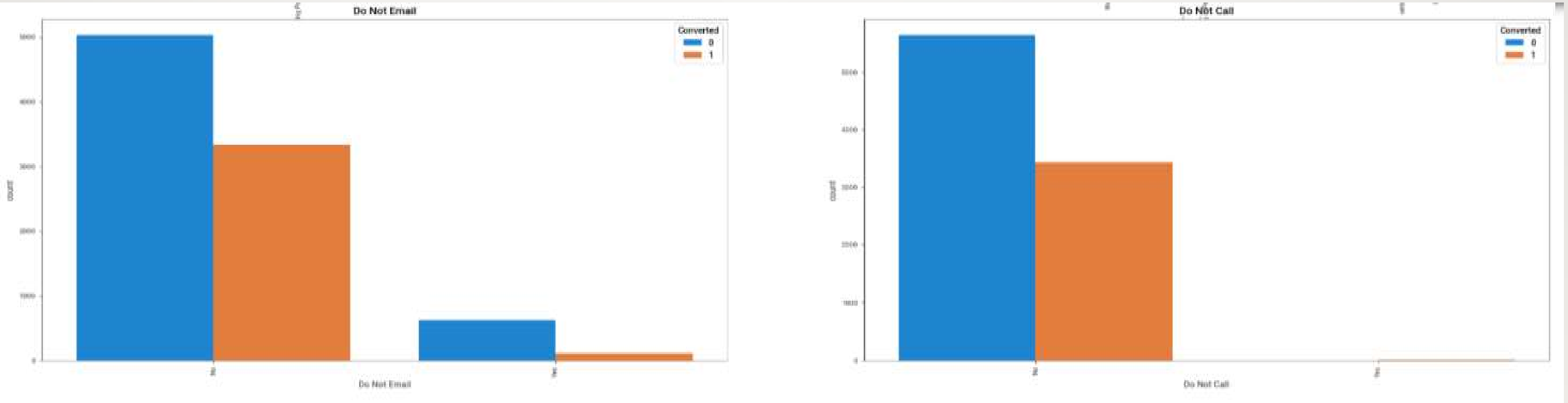
2. Analysis on Numerical  Data

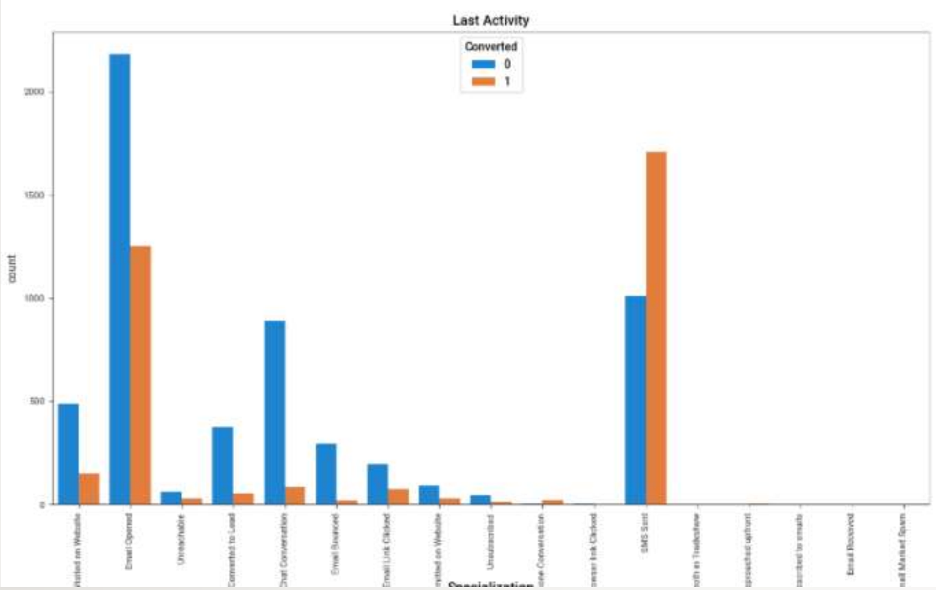


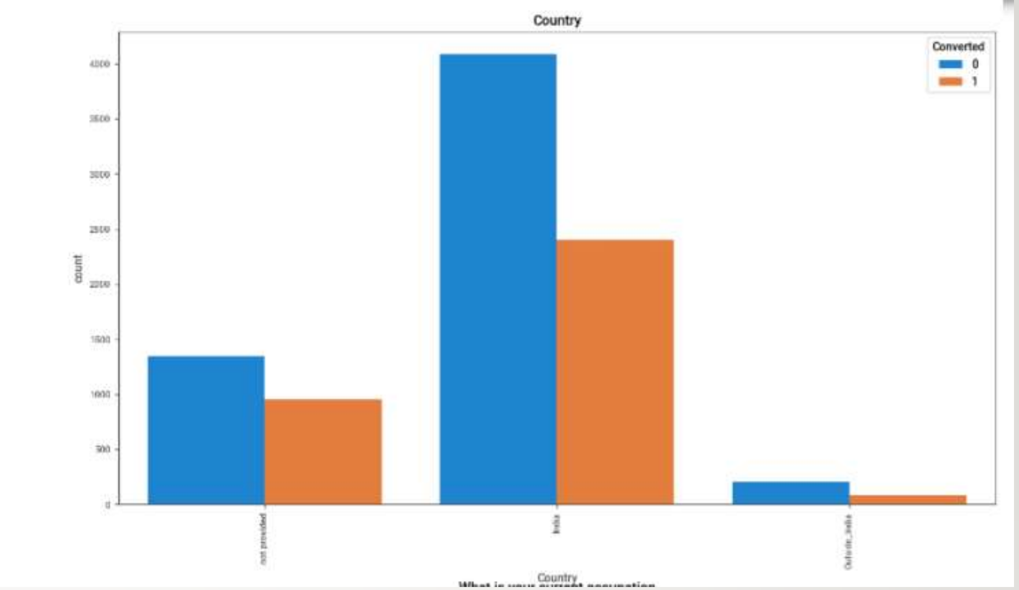
# 2. Categorical Analysis

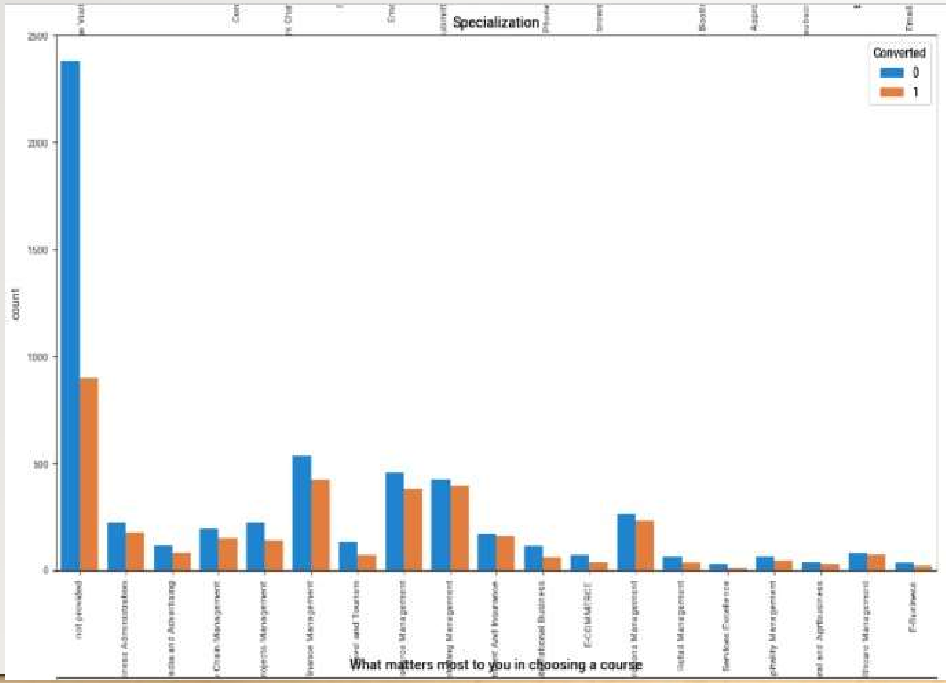
- We can observe from the below charts that conversion rate is comparatively less.

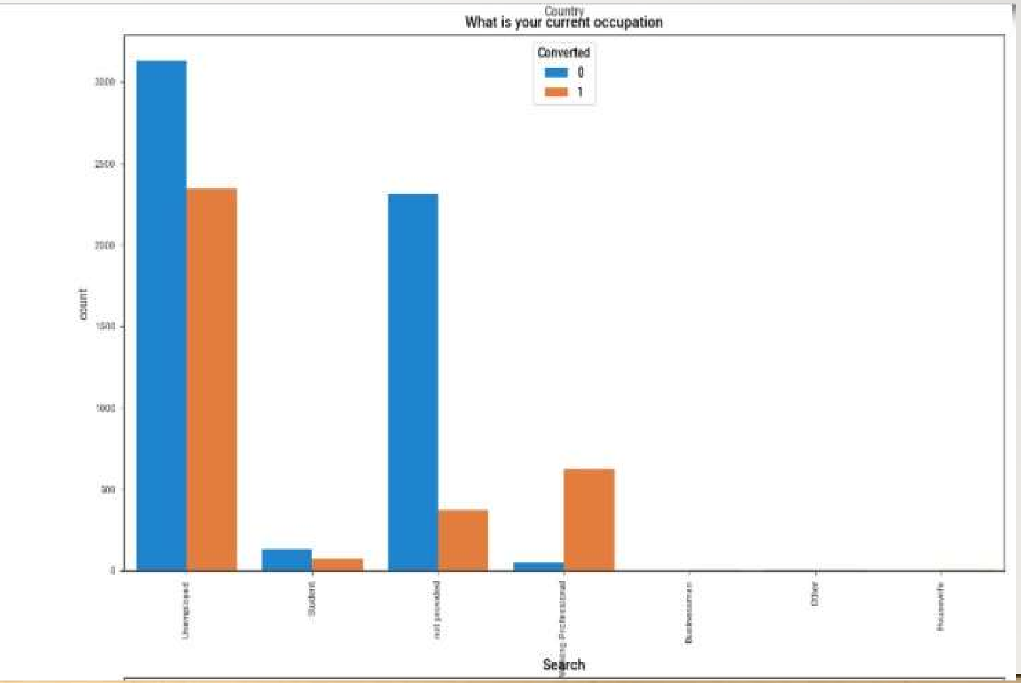


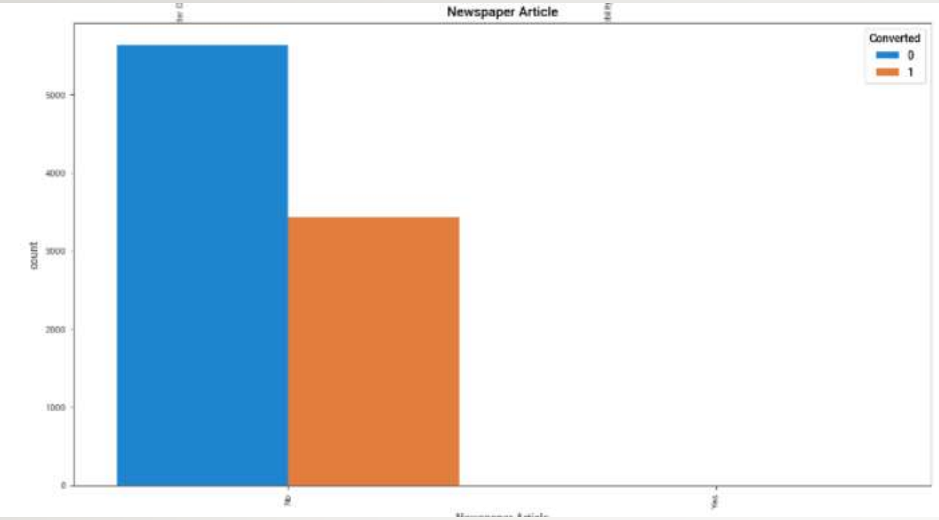



Last Activity
Converted
0
1
count
2000
1500
1000
500
0
isited on Website
Email Opened
Unreachable
Converted to Lead
Chat Conversation
Email Bounced
Email Link Clicked
rmitted on Website
Unsubscribed
one Conversation
owser link Clicked
SMS Sent
oth at Tradeshow
proached upfront
scribed to emails
Email Received
ail Marked Spam
Specialization

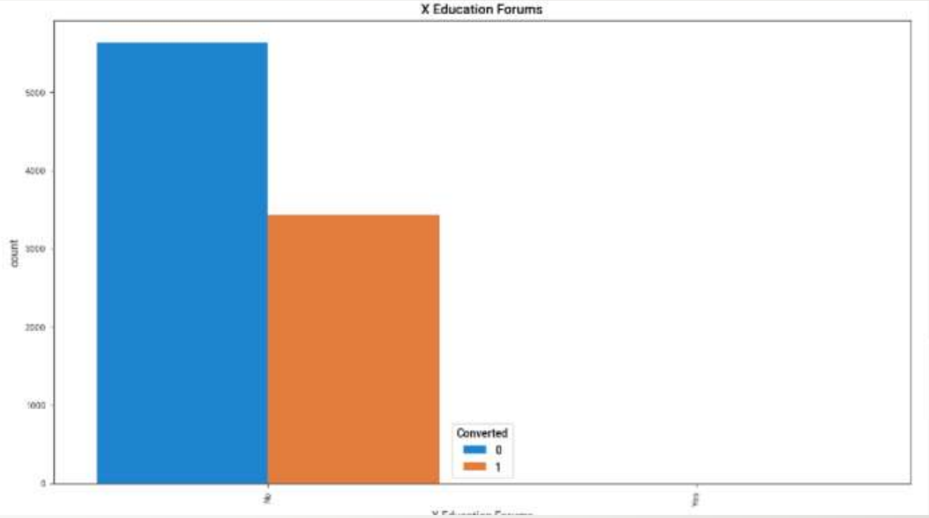

Country
Converted
0
1
count
4000
3500
3000
2500
2000
1500
1000
500
0
not provided
India
Outside_India
Country
What is your current occupation


Specialization
Converted
0
1
count
2500
2000
1500
1000
500
0
not provided
iness Administration
edia and Advertising
y Chain Management
ources Management
inance Management
vel and Tourism
tions Management
ting Management
ing And Insurance
ational Business
E-COMMERCE
tions Management
Retail Management
Services Excellence
itality Management
ral and Agribusiness
thcare Management
E-Business
What matters most to you in choosing a course

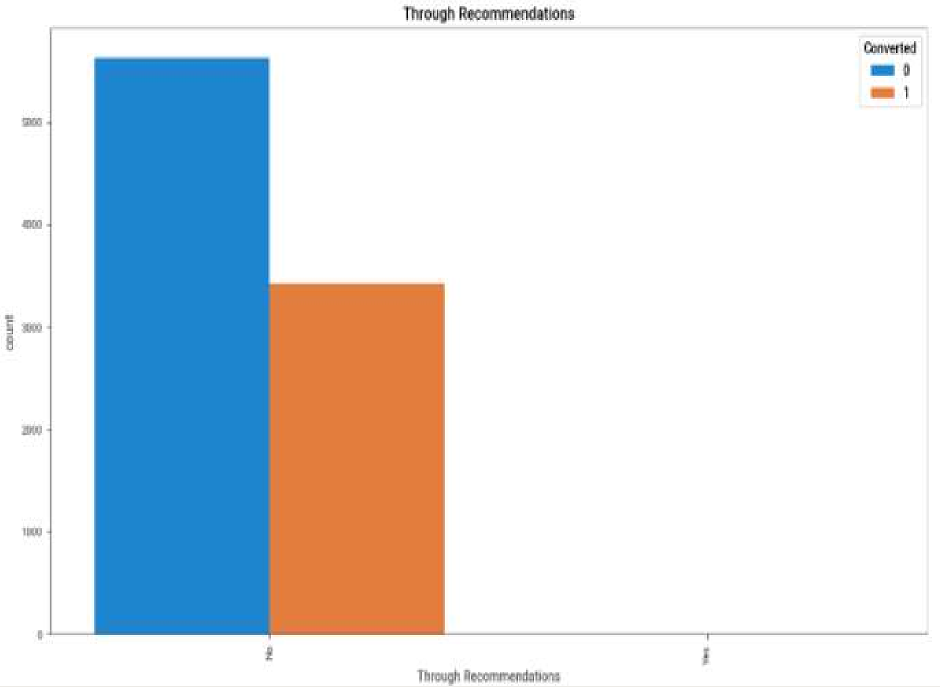

Country
What is your current occupation
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
Unemployed
Student
not provided
Working Professional
Businessman
Other
Housewife
Search



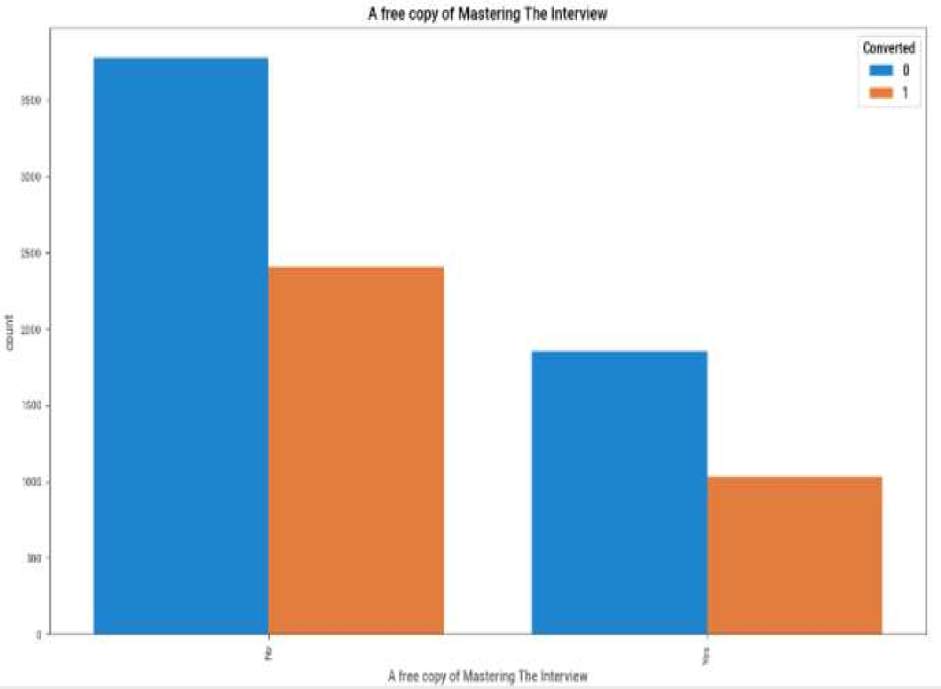
Newspaper Article



X Education Forums



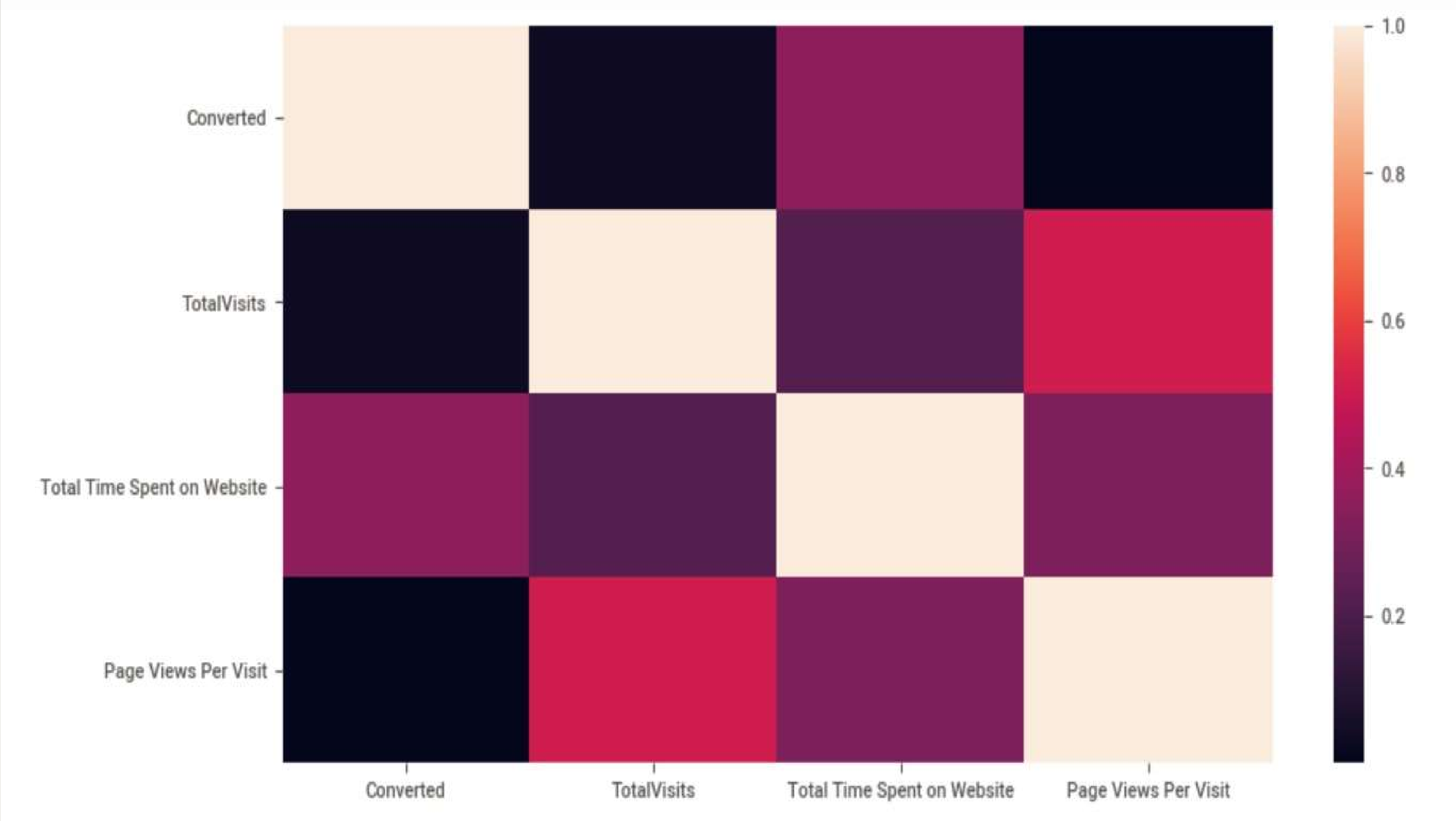
Through Recommendations



A free copy of Mastering The Interview

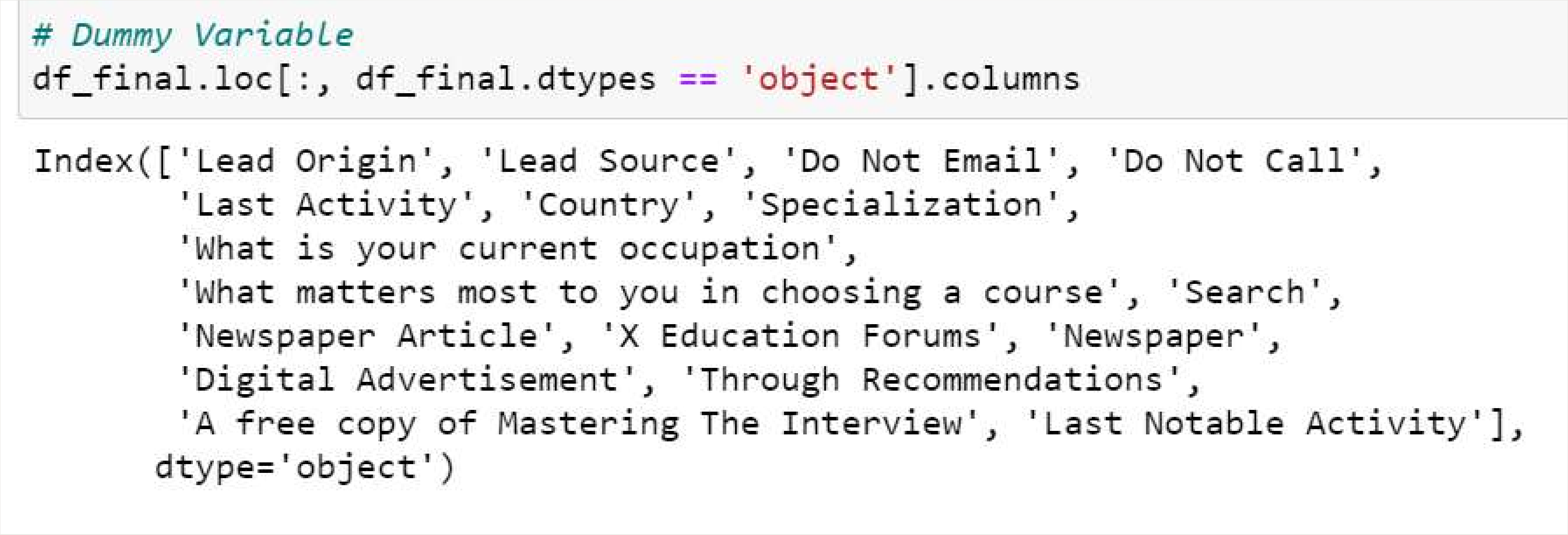
1. We can observe that converted and Total time spent on website has better correlation.

2. Also Page Views per visit and TotalVisits does not have much impact on conversion rate.



# 4. Creating Dummy Variable

- Creating Dummy Variable for categorical data
- Below are the attributes for creating the dummy variable
- Also make sure that first column of dummy variable has been dropped as it 7is very obvious scenario.



```
# Dummy Variable
df_final.loc[:, df_final.dtypes == 'object'].columns
```

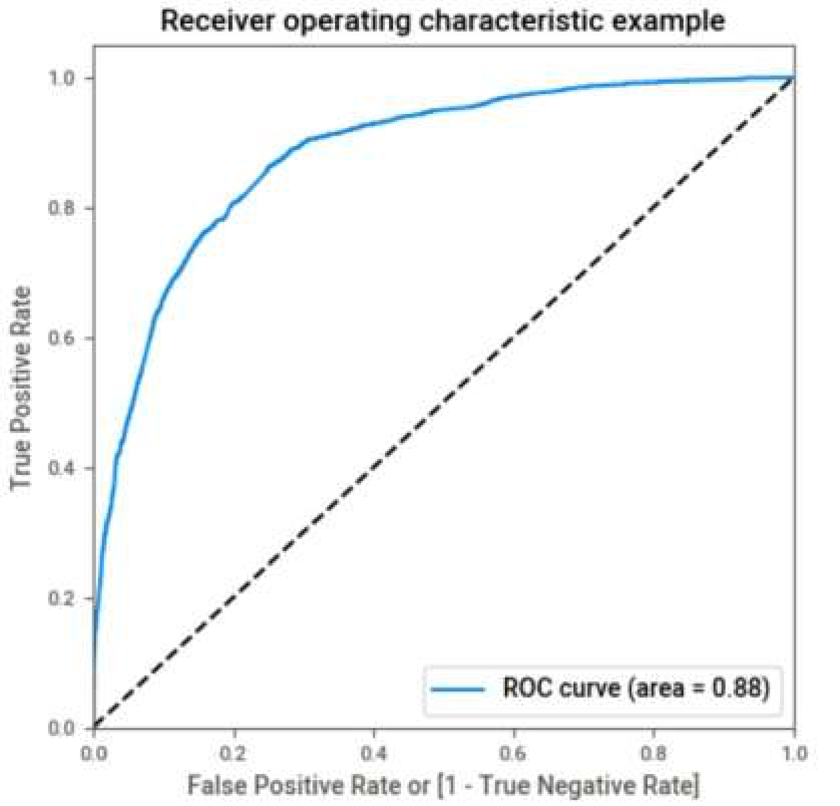
```
Index(['Lead Origin', 'Lead Source', 'Do Not Email', 'Do Not Call',
       'Last Activity', 'Country', 'Specialization',
       'What is your current occupation',
       'What matters most to you in choosing a course', 'Search',
       'Newspaper Article', 'X Education Forums', 'Newspaper',
       'Digital Advertisement', 'Through Recommendations',
       'A free copy of Mastering The Interview', 'Last Notable Activity'],
      dtype='object')
```

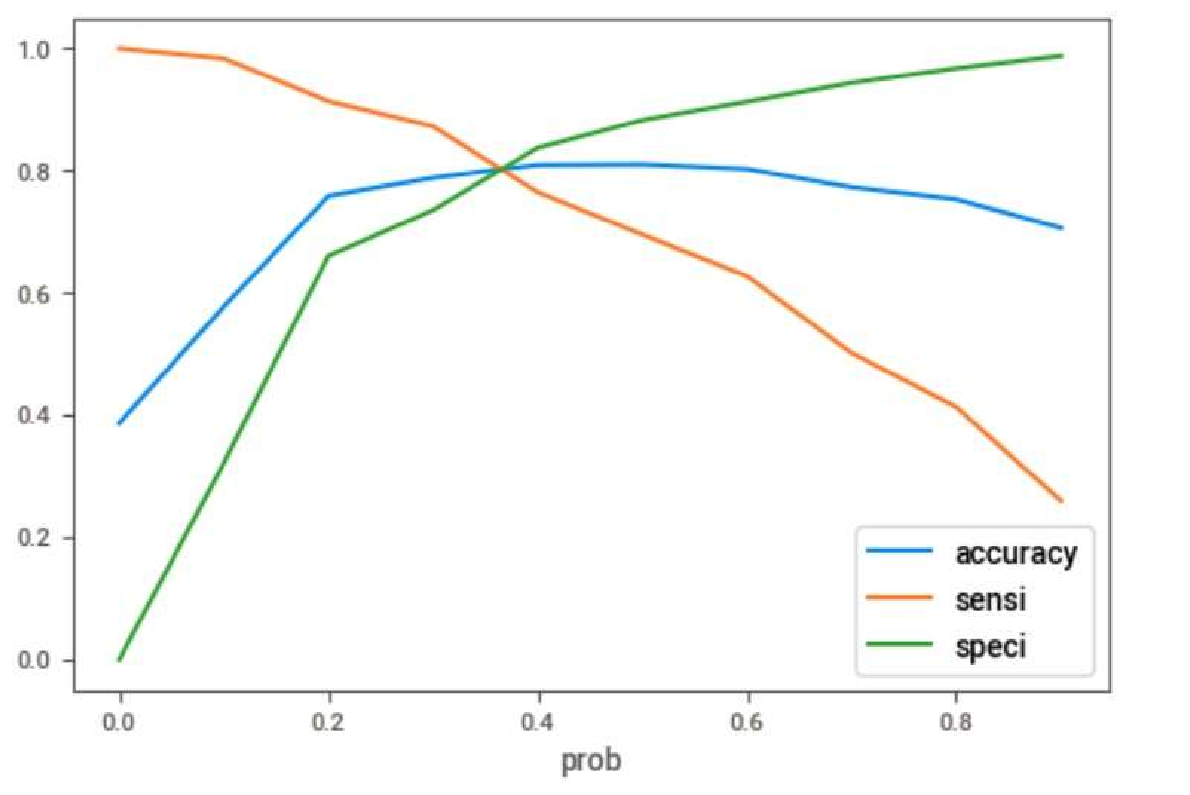
# 5. Building The Model using Train and Test split

- We have split the data between Train and test in proportion of 70:30
- Target Variable is converted and rest of them are factors variable.
- With the help of MinMaxScaler scalling “TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website'”.
- Use Logistic regression
- Create the model with 15 Variable for RFE
- Removed all the variable having P values greater than 0.05.
- We have dropped “Last Notable Activity_Had a Phone Conversation”, “What is your current occupation_Housewife” and “What is your current occupation_Other” one by one and run the model.
- Made the predication on Test data
- We have achieved accuracy around 81.02%

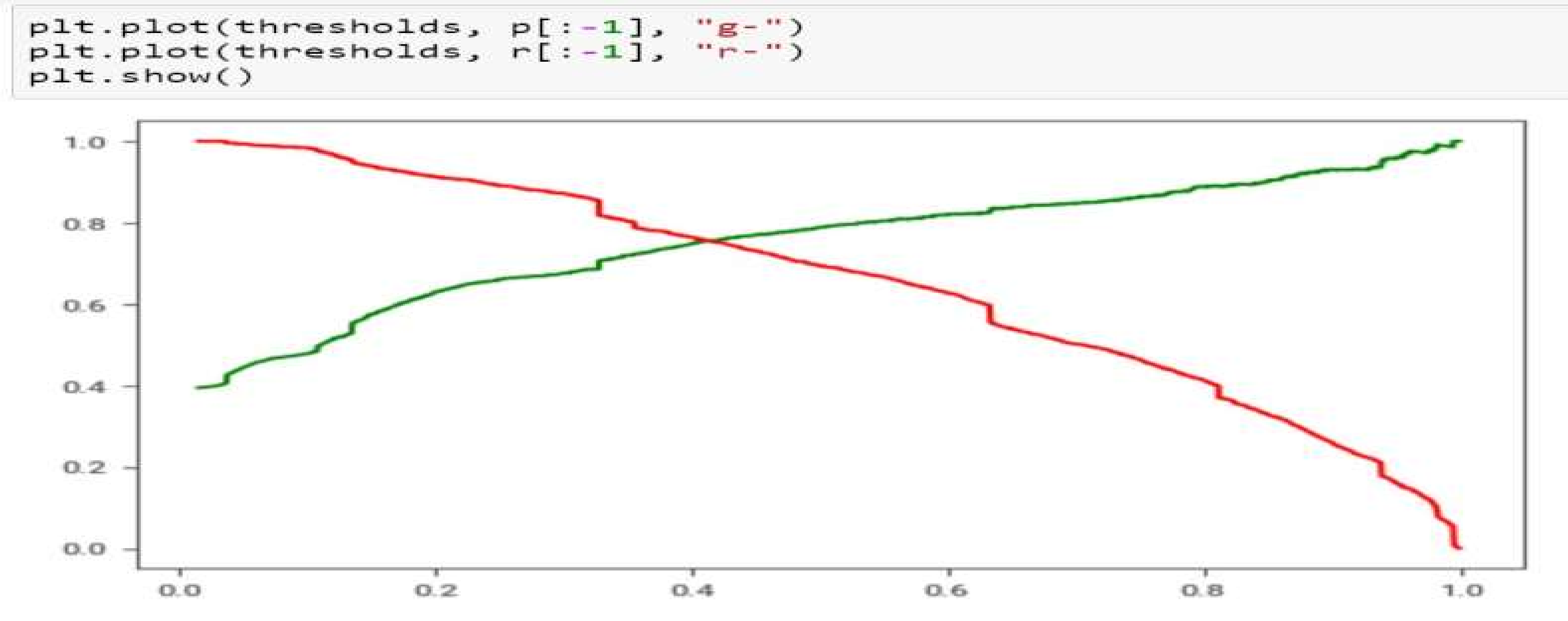
# ROC Curve

- We can see that ROC curve is very far from the centre
- We can see that optimal point near to .37





# Threshold



```
plt.plot(thresholds, p[:-1], "g-")
plt.plot(thresholds, r[:-1], "r-")
plt.show()
```

# Analysis

Below are the factors which has high impact on conversion rate.

1. The total time spent on Website.
2. If the source are following below

    2.1 Google

    2.2 Direct Traffic

    2.3 Organic Search

3. Last Activity

SMS

Olark chat conversation

4. When the current occupation is a working professional.

# RECOMMENDATIONS

- Make website more attractive as most leads are converting through Website
- Last activity is coming through Olark chat and SMS
- Mostly working professional are attractive so tailor make the courses as per their needs for tagertting the potential clientele.
- We can also some education benefits for unemployed or students to churn higher rate and provide them more information regarding the course and how it can help them in pocketing higher package.
- They Can create the chat robots for genuine calls and chat and they can do further analysis based on data in chat robots.