Lead Score case study

Johi Shah

Yerrapragada Sriram

Shobhit Srivastava

Business Problem

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you 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 the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

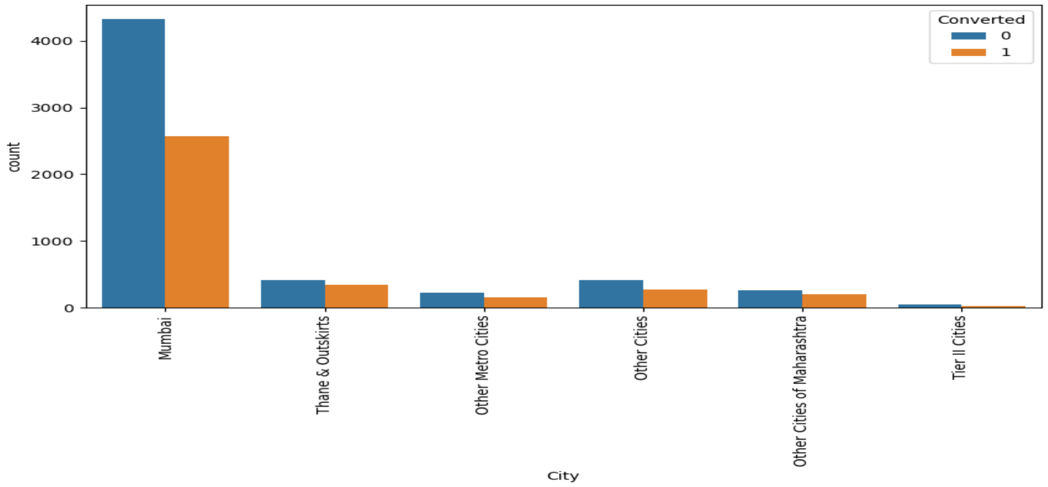
The input File

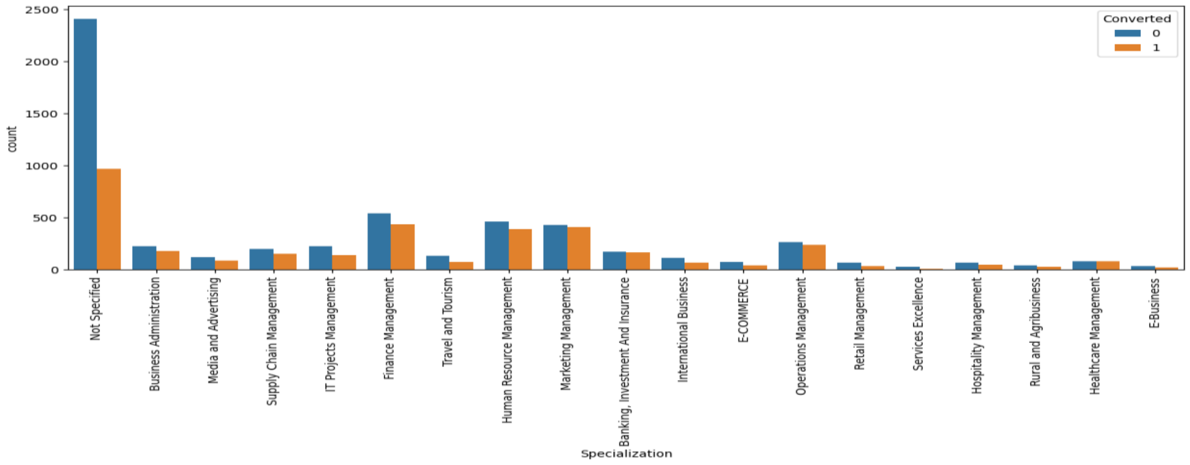
We have been provided with following file as input

Microsoft Excel
ma Separated Valu

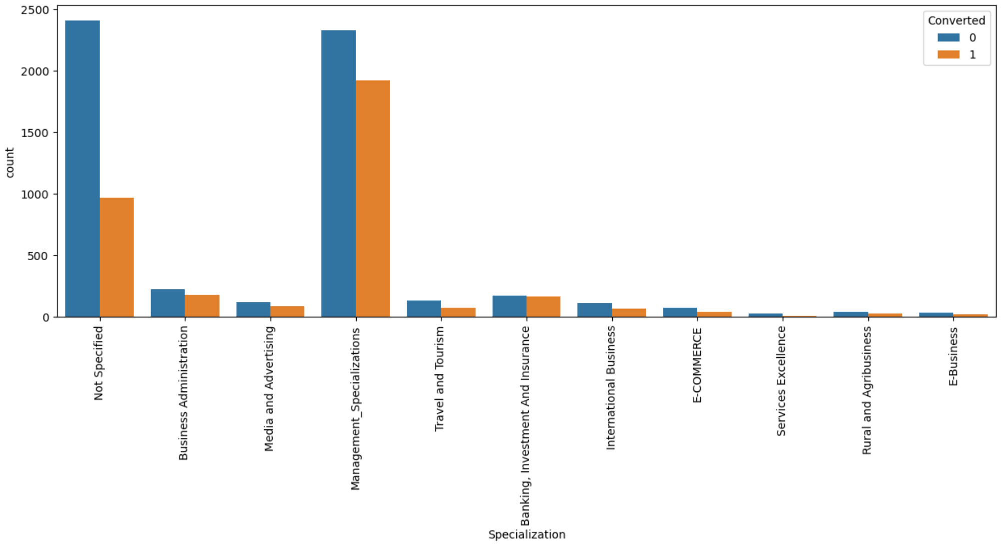
EDA

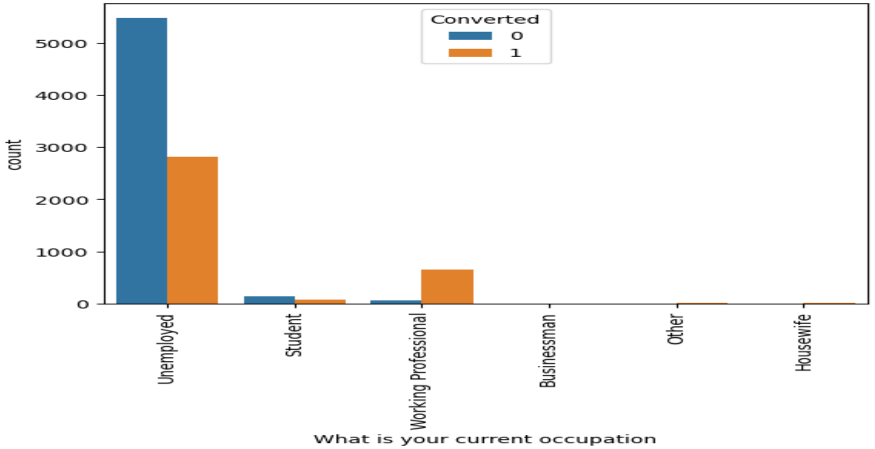
We will do the EDA analysis of data and Following are inferences drawn



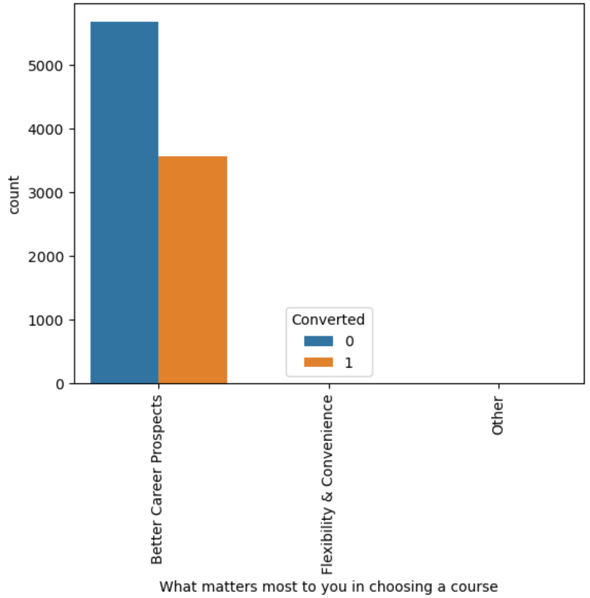


EDA

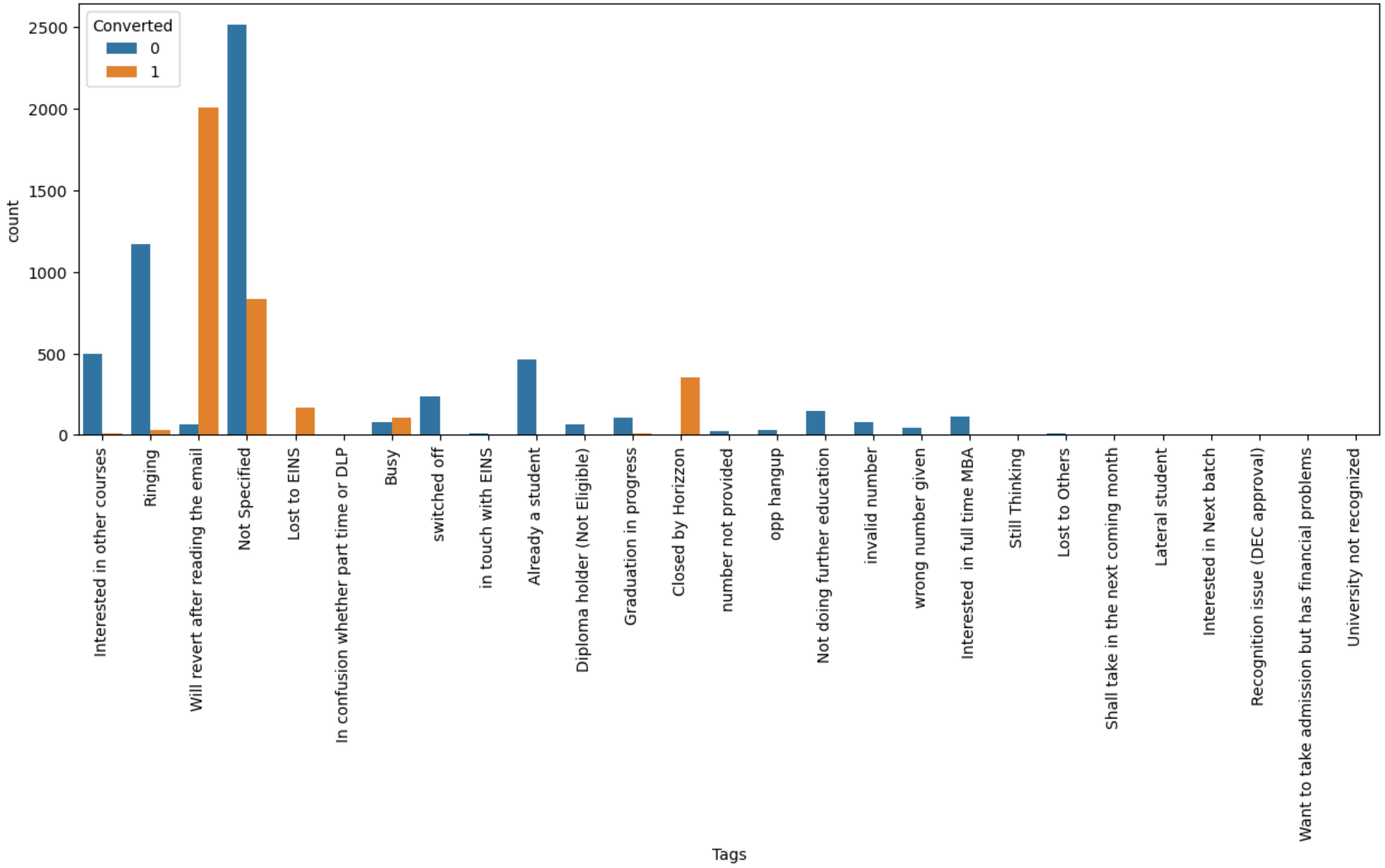

Converted
0
1
count
0
500
1000
1500
2000
2500
Not Specified
Business Administration
Media and Advertising
Management_Specializations
Travel and Tourism
Banking, Investment And Insurance
International Business
E-COMMERCE
Services Excellence
Rural and Agribusiness
E-Business
Specialization


Converted
0
1
count
0
1000
2000
3000
4000
5000
Unemployed
Student
Working Professional
Businessman
Other
Housewife
What is your current occupation

EDA

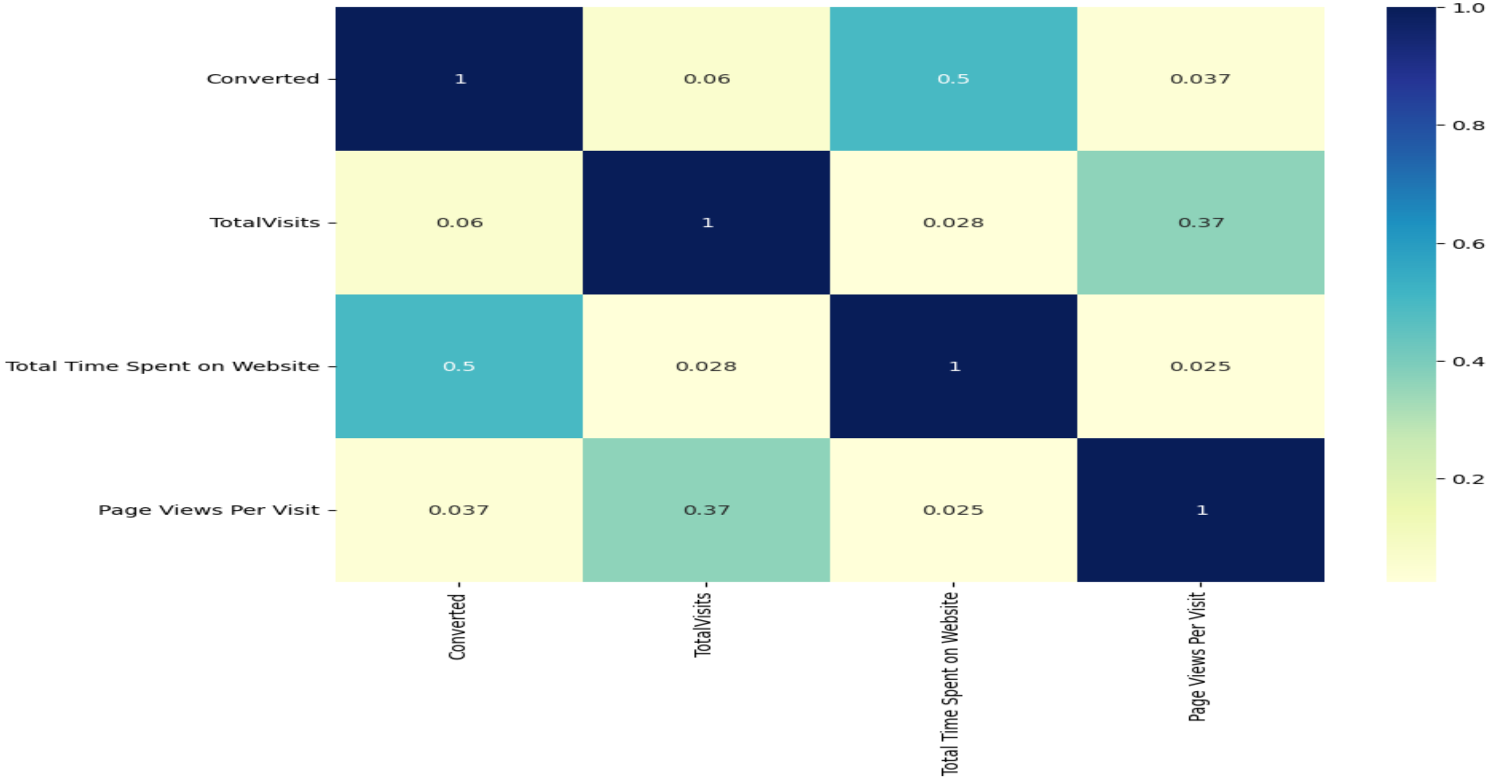


count
0
1000
2000
3000
4000
5000
Converted
0
1
Better Career Prospects
Flexibility & Convenience
Other
What matters most to you in choosing a course



count
0
500
1000
1500
2000
2500
Converted
0
1
Interested in other courses
Ringing
Will revert after reading the email
Not Specified
Lost to EINS
In confusion whether part time or DLP
Busy
switched off
in touch with EINS
Already a student
Diploma holder (Not Eligible)
Graduation in progress
Closed by Horizzon
number not provided
opp hangup
Not doing further education
invalid number
wrong number given
Interested in full time MBA
Still Thinking
Lost to Others
Shall take in the next coming month
Lateral student
Interested in Next batch
Recognition issue (DEC approval)
Want to take admission but has financial problems
University not recognized
Tags

EDA



Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
1
0.06
0.5
0.037
0.06
1
0.028
0.37
0.5
0.028
1
0.025
0.037
0.37
0.025
1
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
1.0
0.8
0.6
0.4
0.2

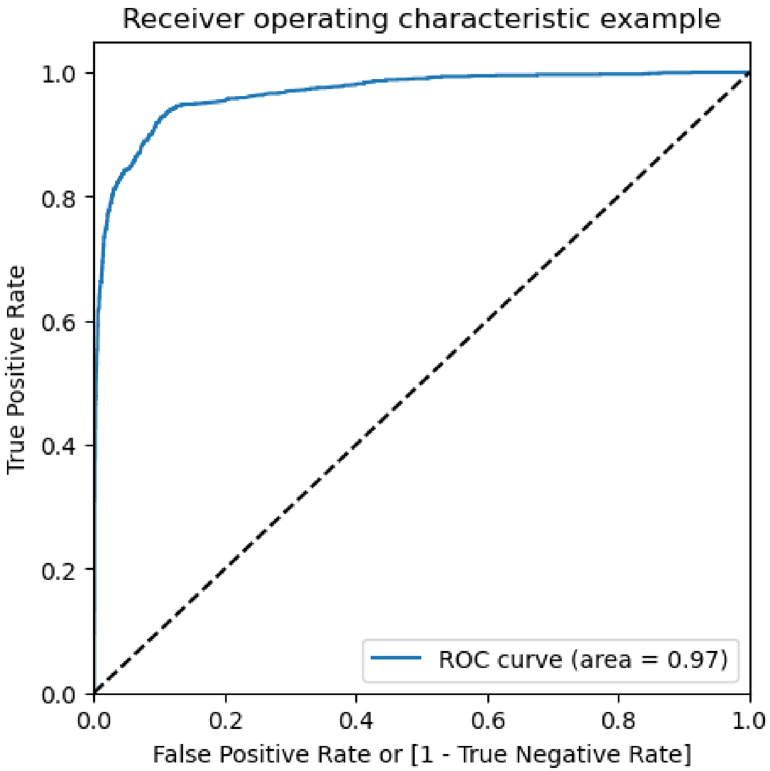
Train-Test Split & Logistic Regression Model Building

We will use logistic regression model

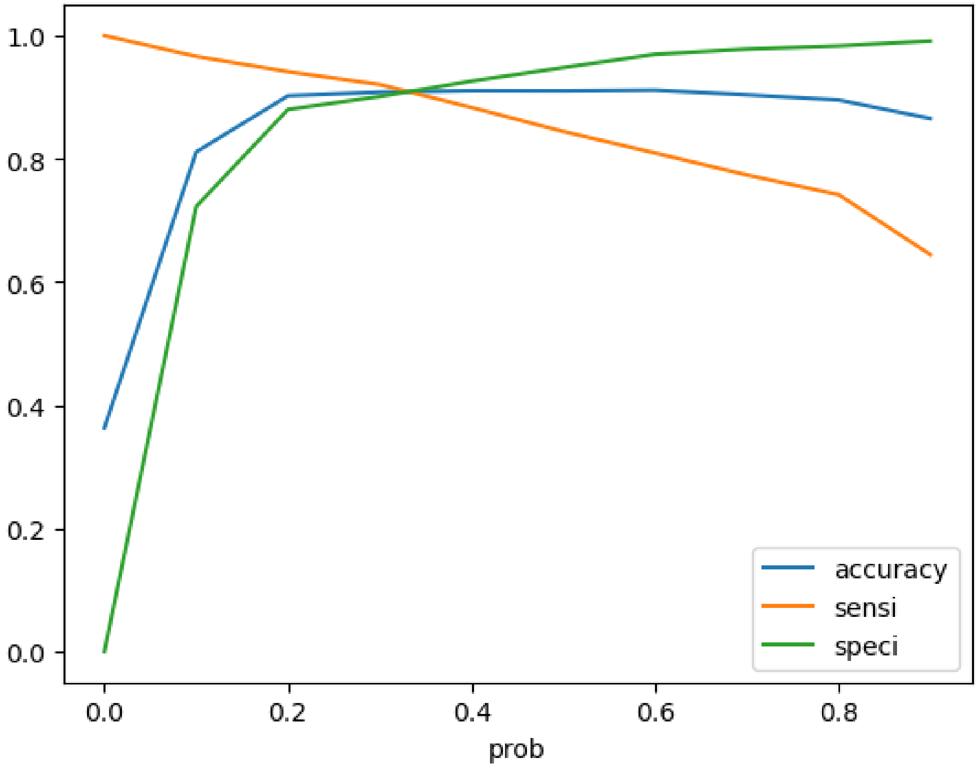
```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression()

from sklearn.feature_selection import RFE
rfe = RFE(logreg, n_features_to_select=15)          # running RFE with 15 variables as
output
rfe = rfe.fit(X_train, y_train)
```

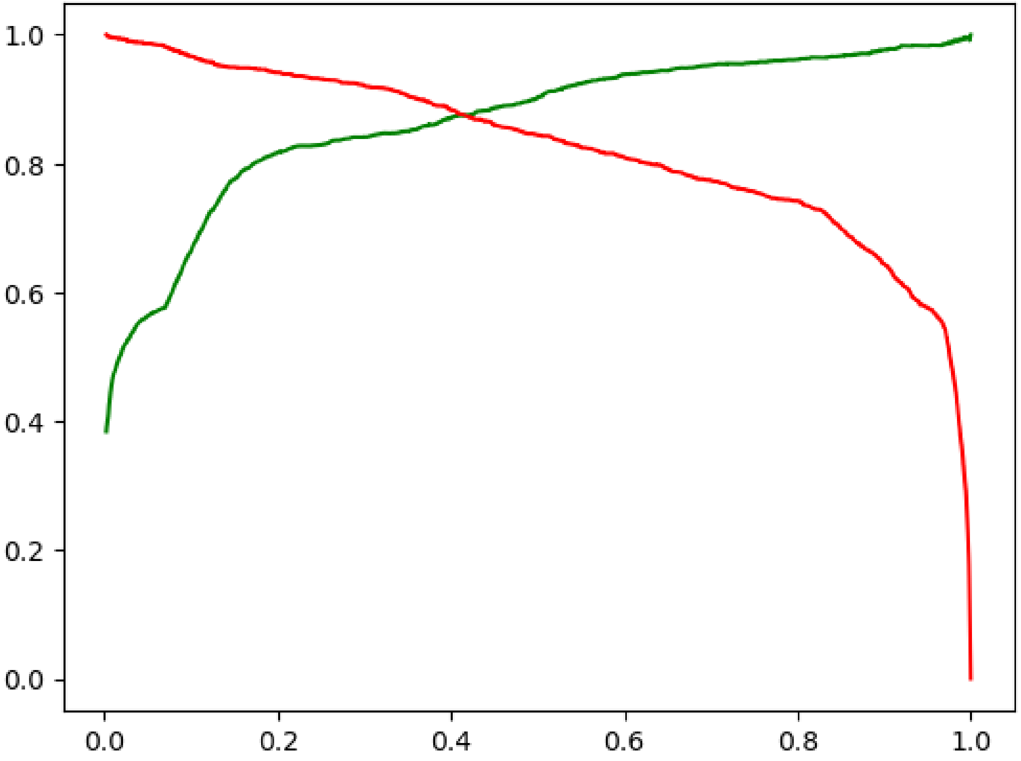
PLOTTING ROC CURVE



Finding Optimal Cutoff Point



Finding Optimal Cutoff Point



PREDICTIONS ON TEST SET

```
y_test_pred = res.predict(X_test_sm)
```

Observation

After running the model on the Test Data these are the figures we obtain:

Accuracy : 90.92%

Sensitivity : 91.41%

Specificity : 90.62%

Final Observation:

Let us compare the values obtained for Train & Test:

Train Data:

Accuracy : 90.81%

Sensitivity : 92.05%

Specificity : 90.10%

Test Data:

Accuracy : 90.92%

Sensitivity : 91.41%

Specificity : 90.62%

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model