LEAD SCORE ASSIGNMENT

Lead Score Case Study for X Education

Problem Statement :

X Education sells online courses to industry professionals. 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%.
Business Goal:
X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.
The company needs a model wherein you a lead score is assigned 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.
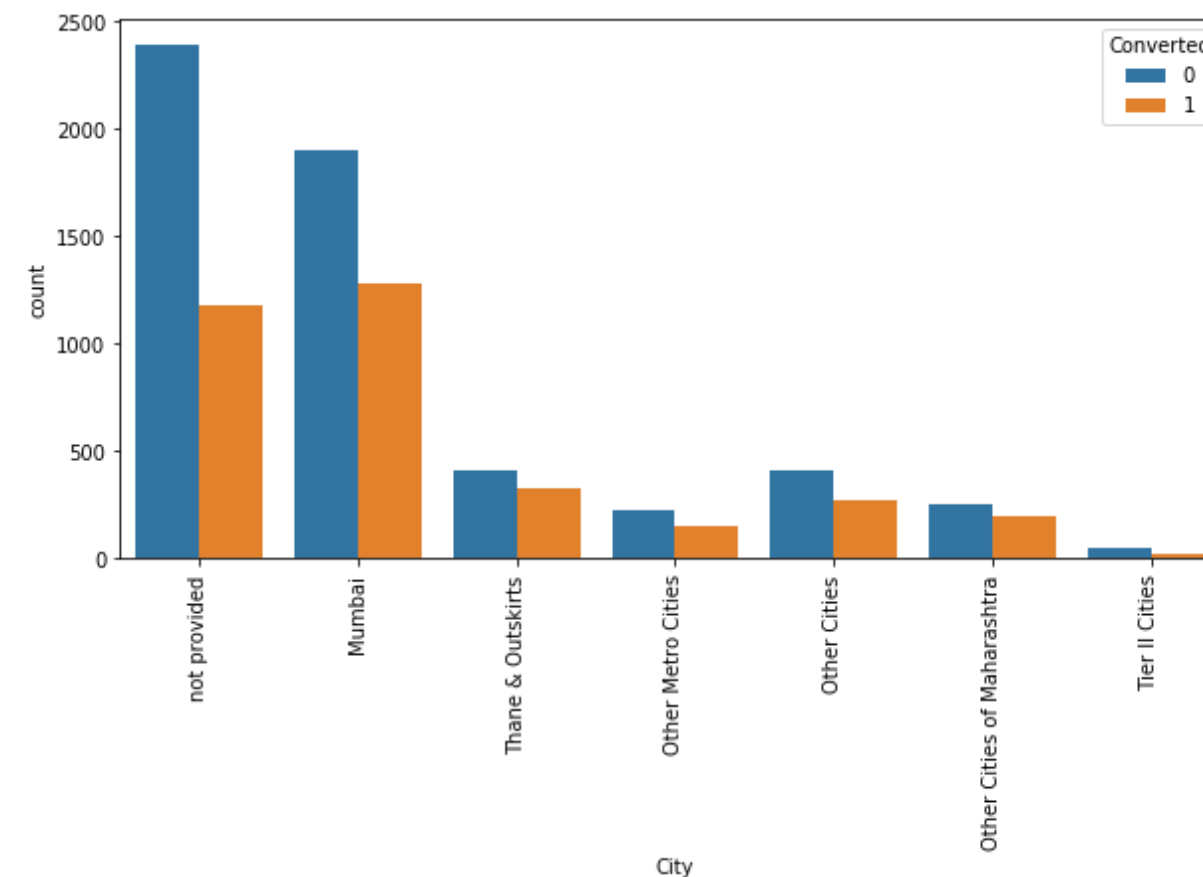The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.
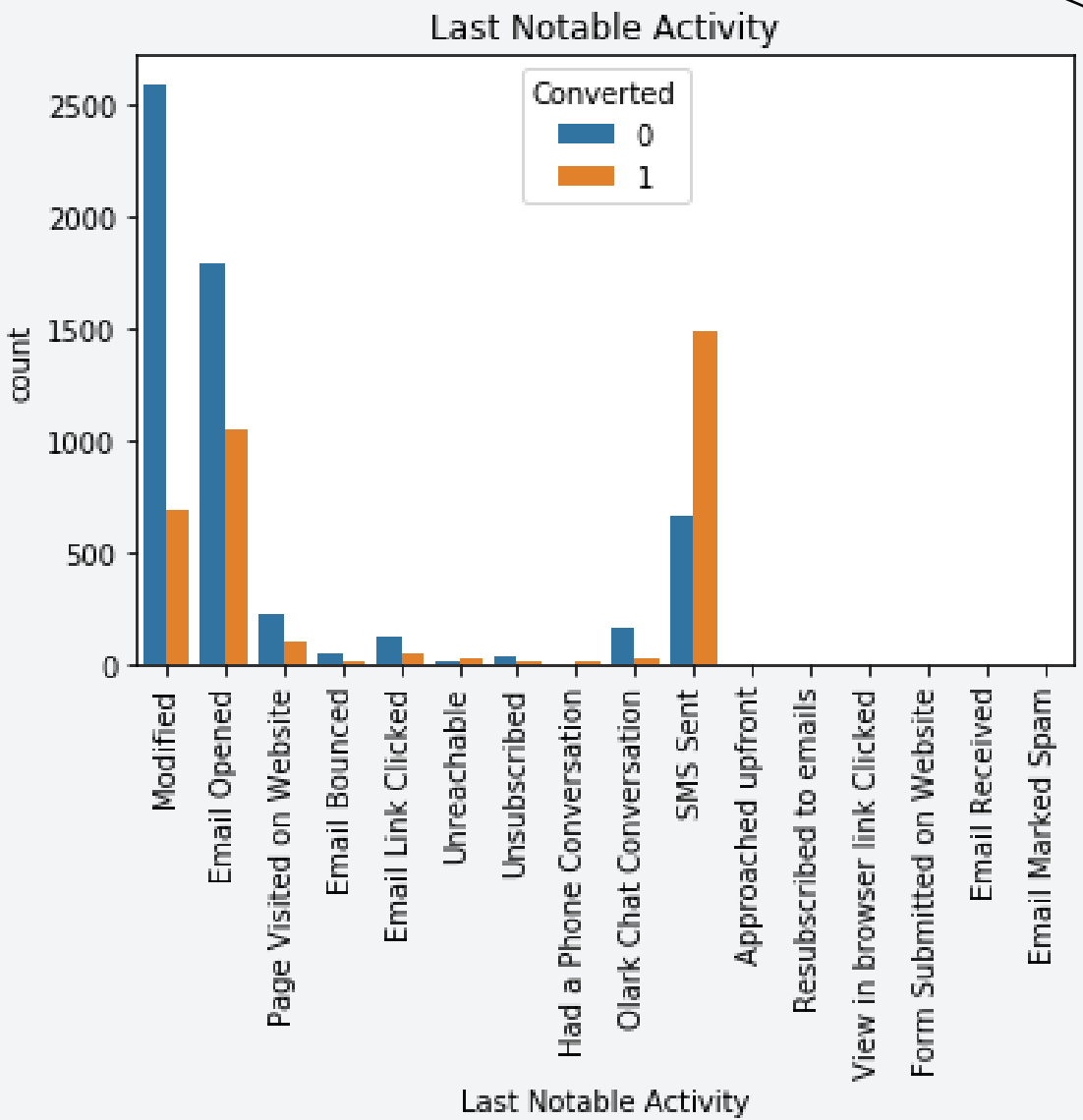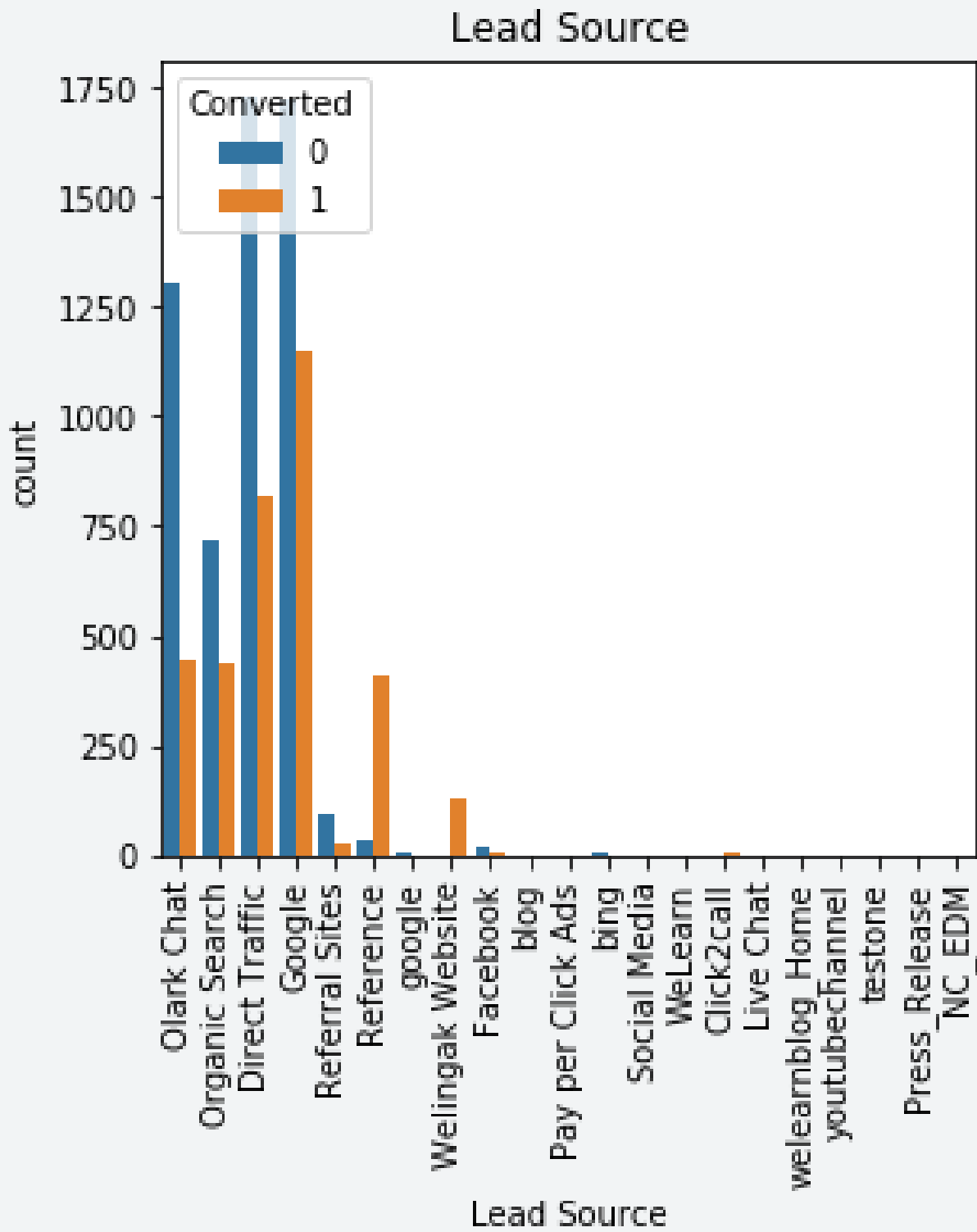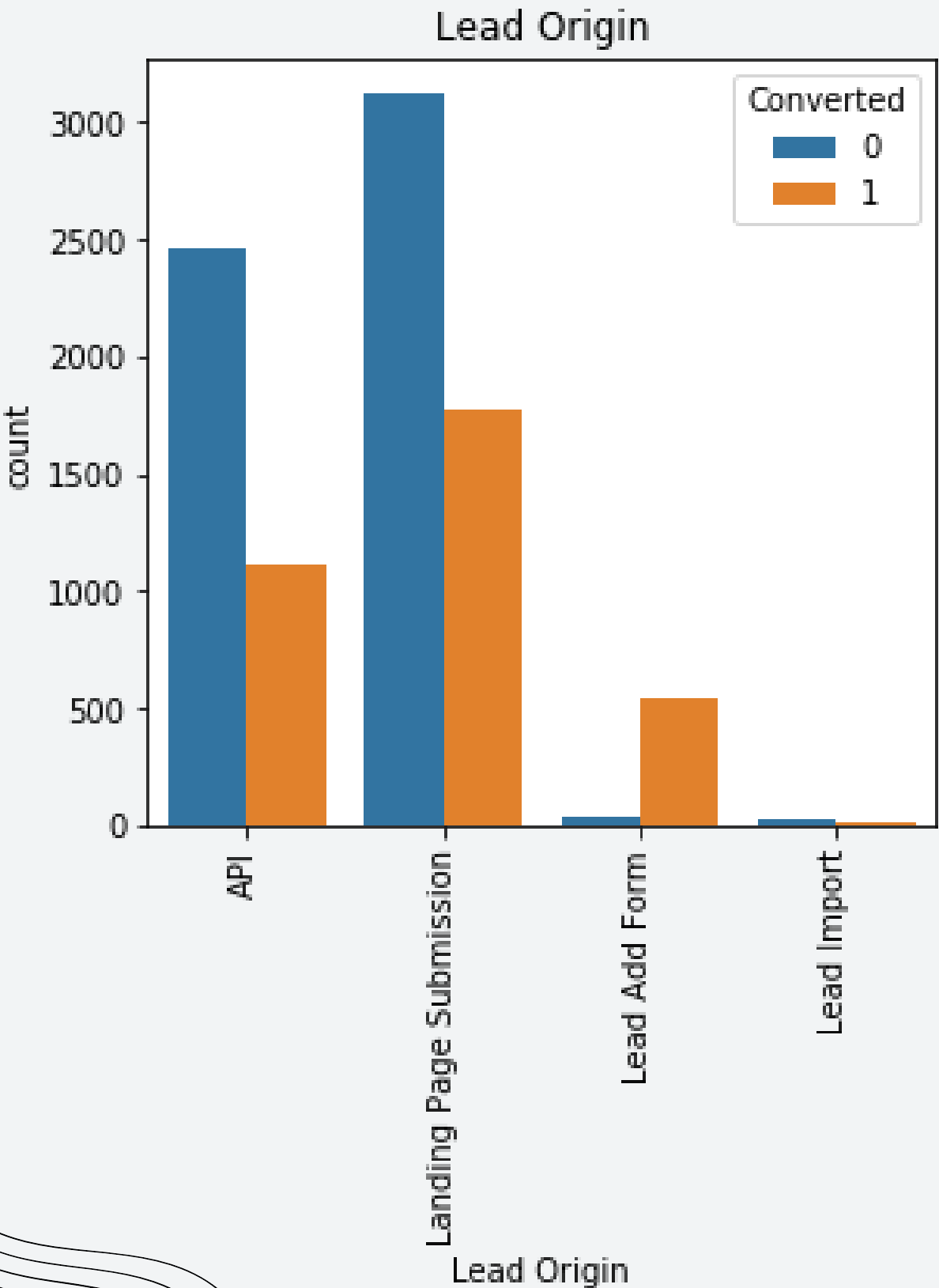
Exploratory Data Analysis:

As we can see the Number of Values for India are quite high (nearly 97% of the Data), this column can be dropped
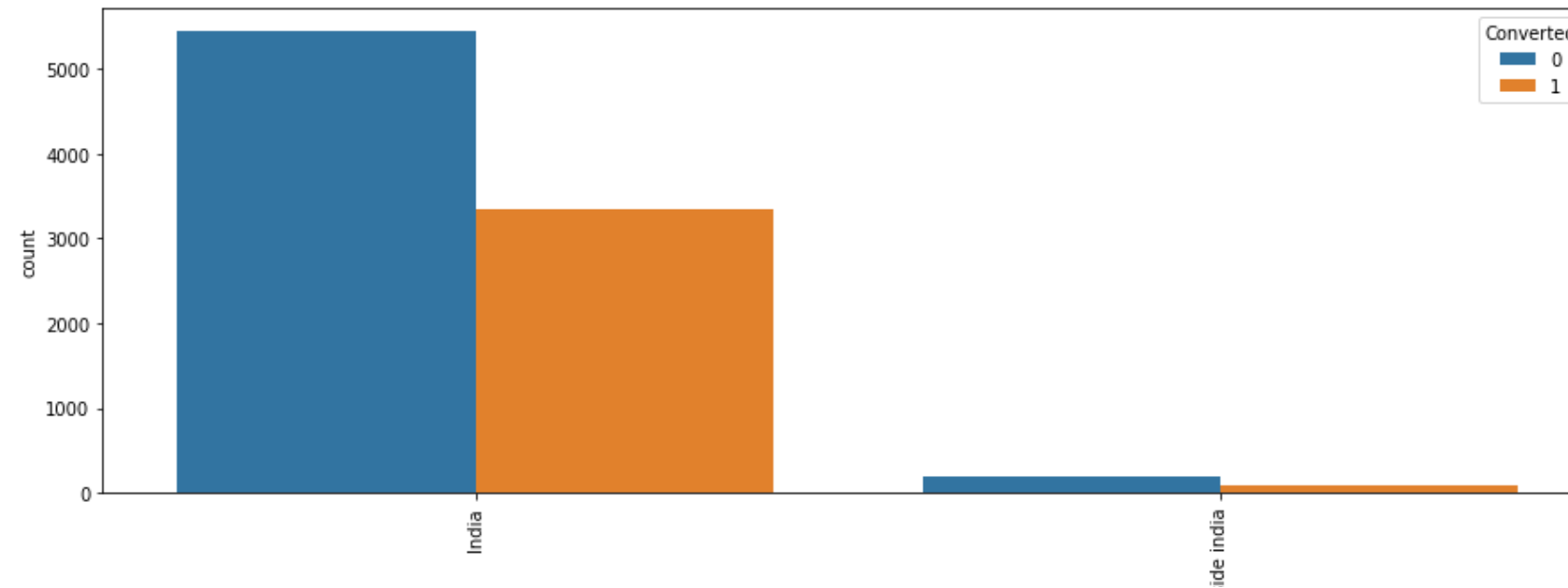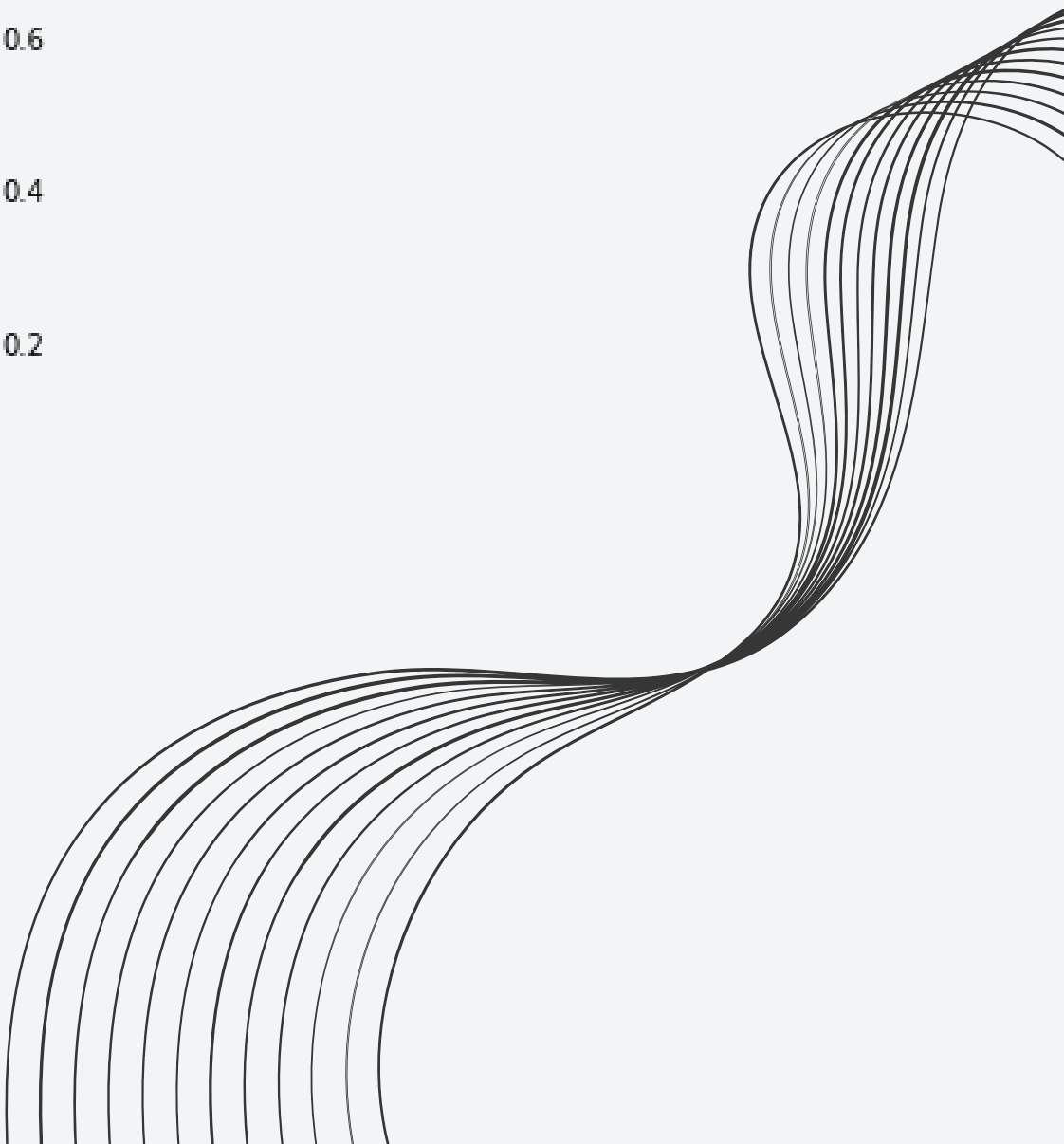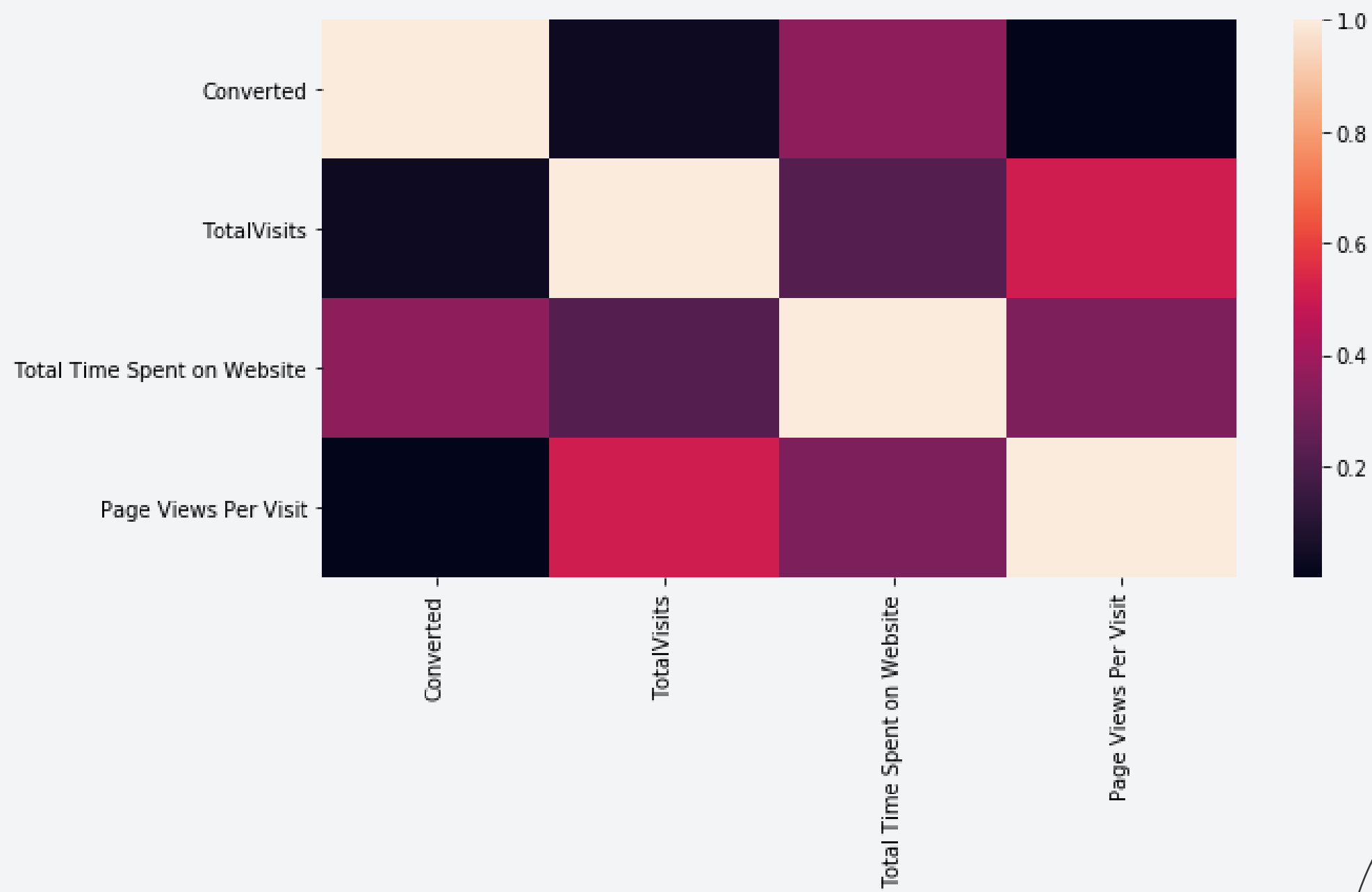
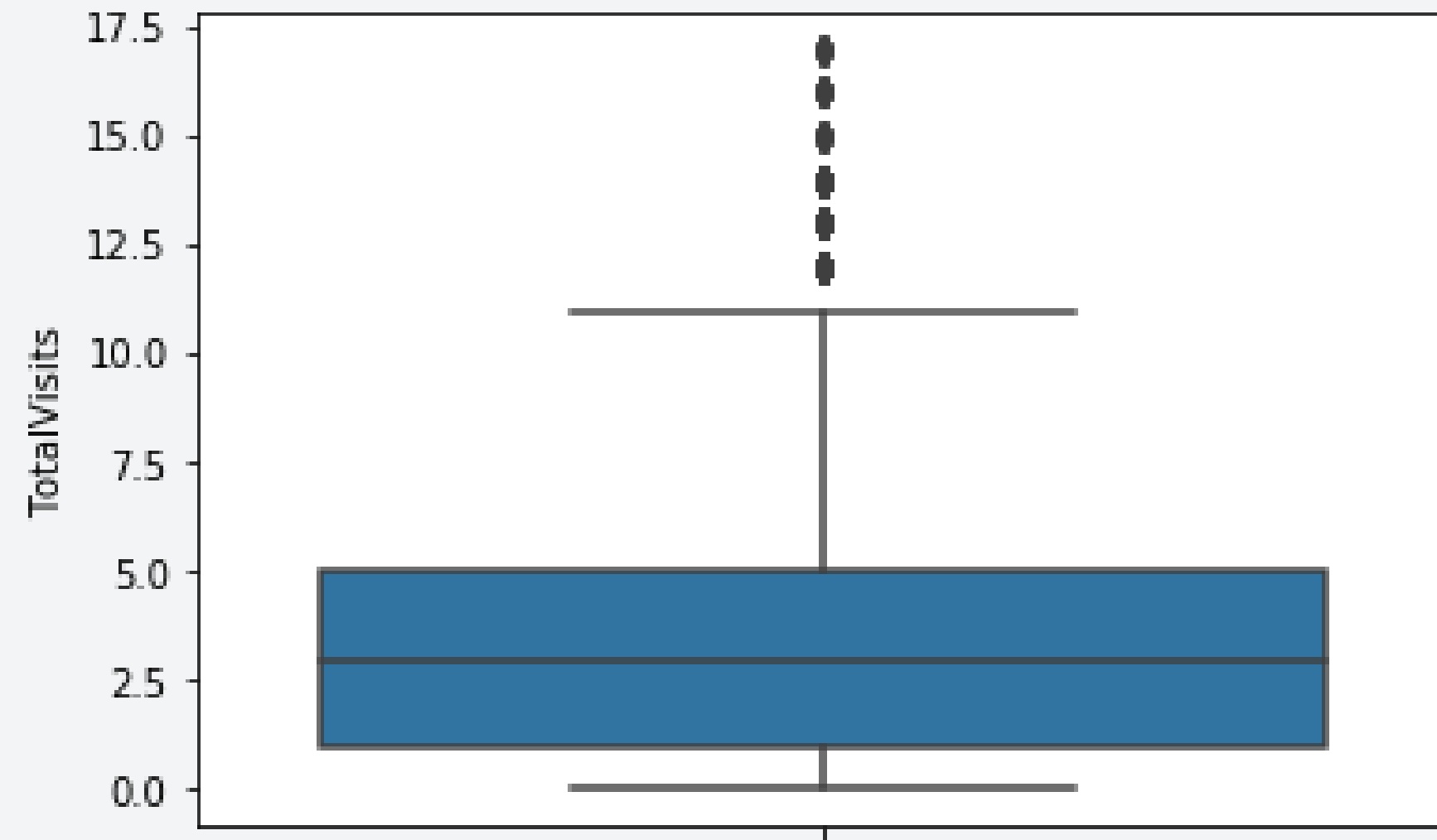

plotting spread of City columnn

Looking between Lead Origin and lead Source below :

It is understandable from the above EDA that there are many elements that have very little data and so will be of less relevance to our analysis.
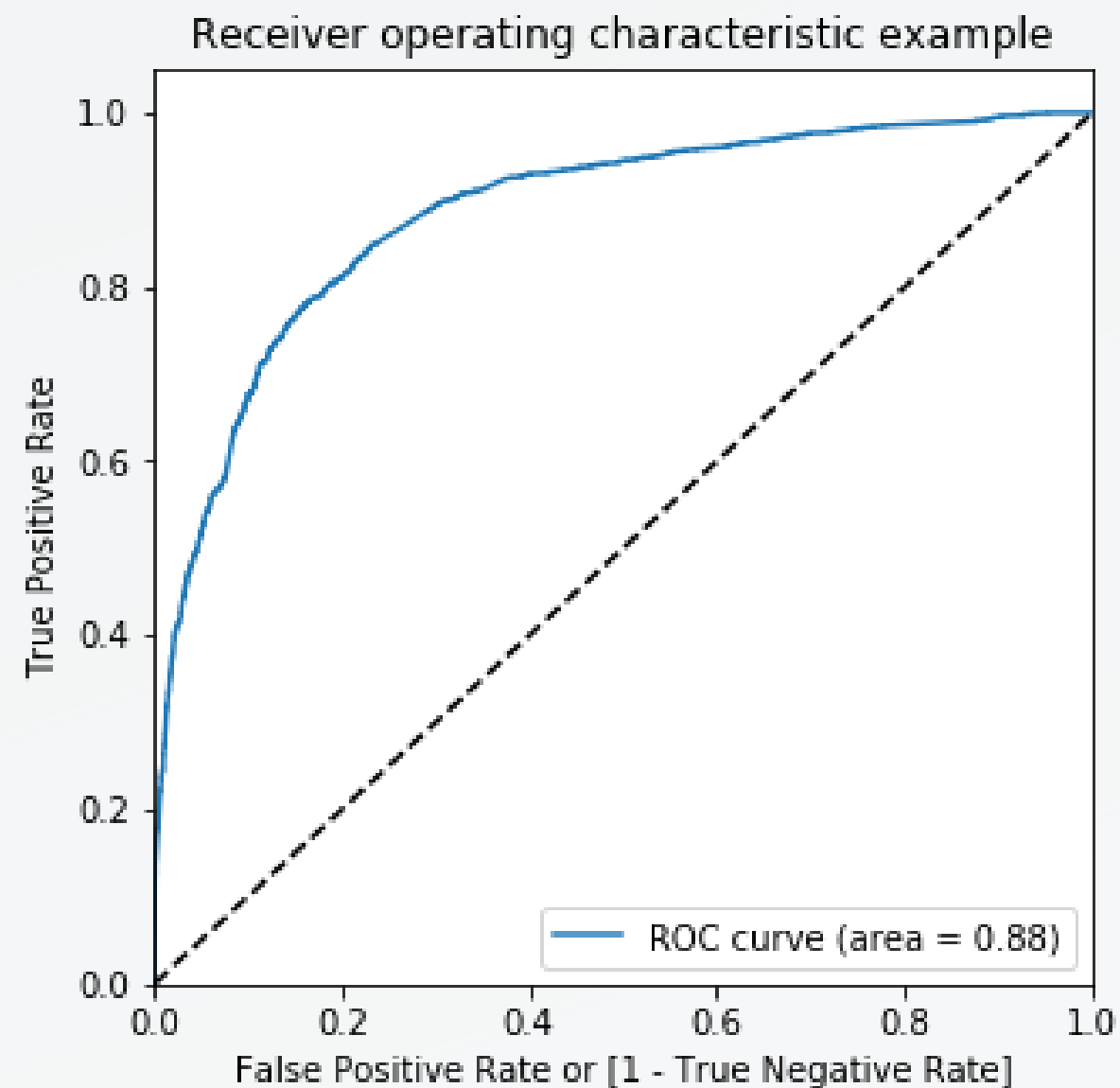
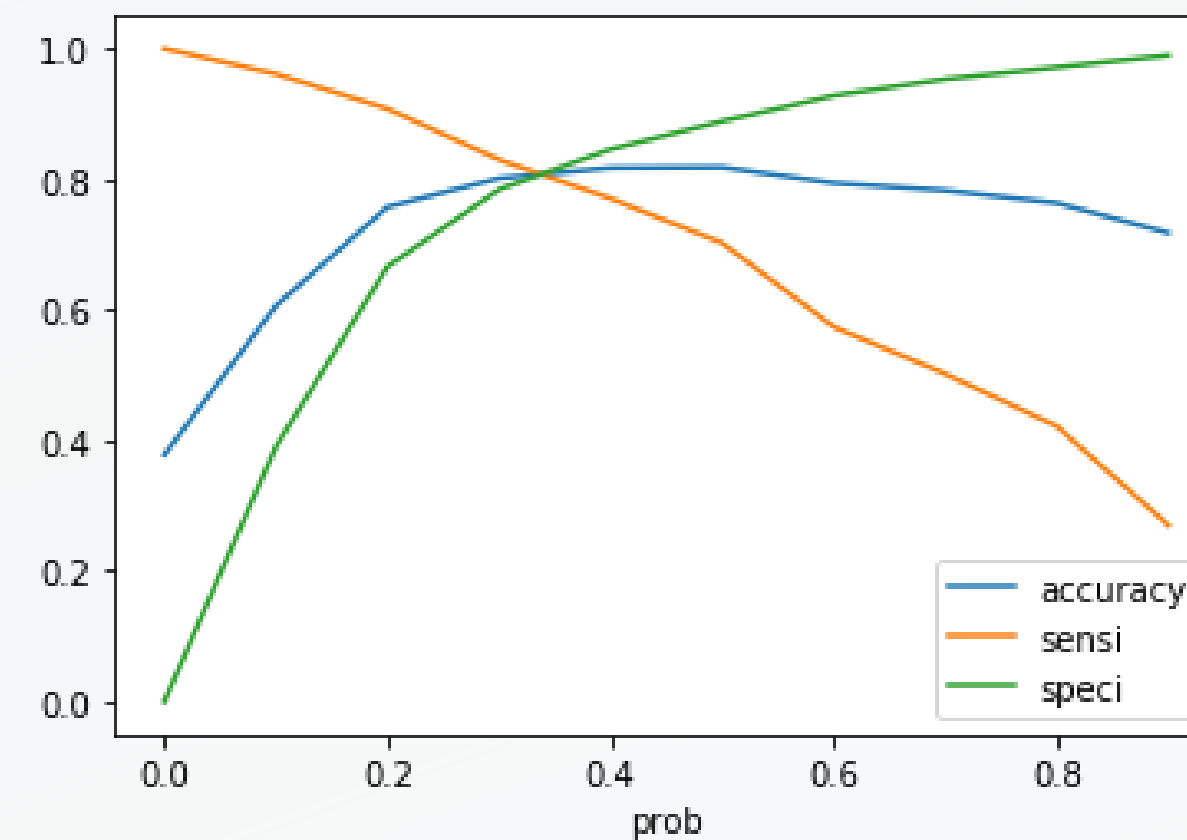Outlier Treatment: Remove top & bottom 1% of the Column Outlier values

ROC Curve

The area under ROC curve is 0.88 which is a very good value
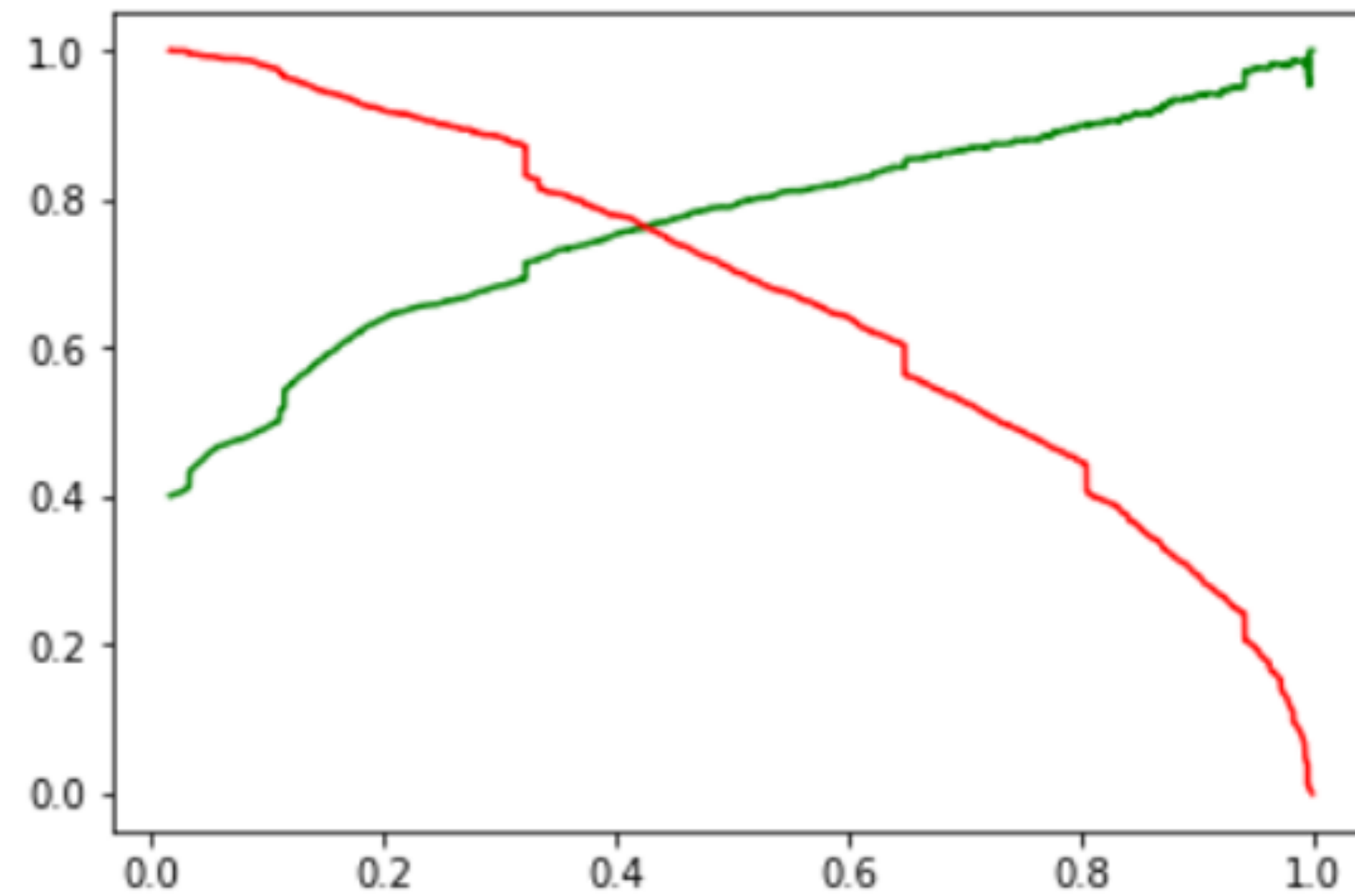
From the graph it is visible that the optimal cut off is at 0.35.

Receiver operating characteristic example

Model Evaluation- Precision and Recall on Train Datase

The graph depicts an optimal cut off of 0.42 based on Precision and
Recall

Creating confusion matrix

[[3333, 582],
 [563, 1815]],

Precision = TP / TP + FP

0.7571964956195244

Recall = TP / TP + FN

0.763246425567704

With the current cut off as 0.44 we have Precision around 76% and Recall around 76.3% and accuracy 82 %.

Creating confusion matrix

[[1426, 251],
 [273, 748]]

Precision = TP / TP + FP

0.7487487487487487

Recall = TP / TP + FN

0.732615083251714

With the current cut off as 0.41 we have Precision around 75% , Recall around 73% and accuracy 80.5%.

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

Conclusion

It was found that the variables that mattered the most in the potential buyers are, TotalVisits,The total time spend on the Website,Lead Origin_Lead Add Form,Lead Source_Welingak Website,Lead Source_Welingak Website, Lead Source_Direct Traffic,Lead Source_Google, Lead Source_Organic Search, Lead Source_Referral Sites, Do Not Email_Yes and so on.

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

TOP VARIABLE CONTRIBUTING TO CONVERSION:
• LEAD SOURCE:
Lead Origin:
• Lead source:
Referral Sites

Last Activity:
• Do Not Email_Yes
• Last Activity_Email Bounced
• Olark chat conversation

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

******Thankyou******