X-study Lead score case study

Submitted by :-

Ambrish joshi

Shivani

Pompy Mukhopadhyay

Problem - statement

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.

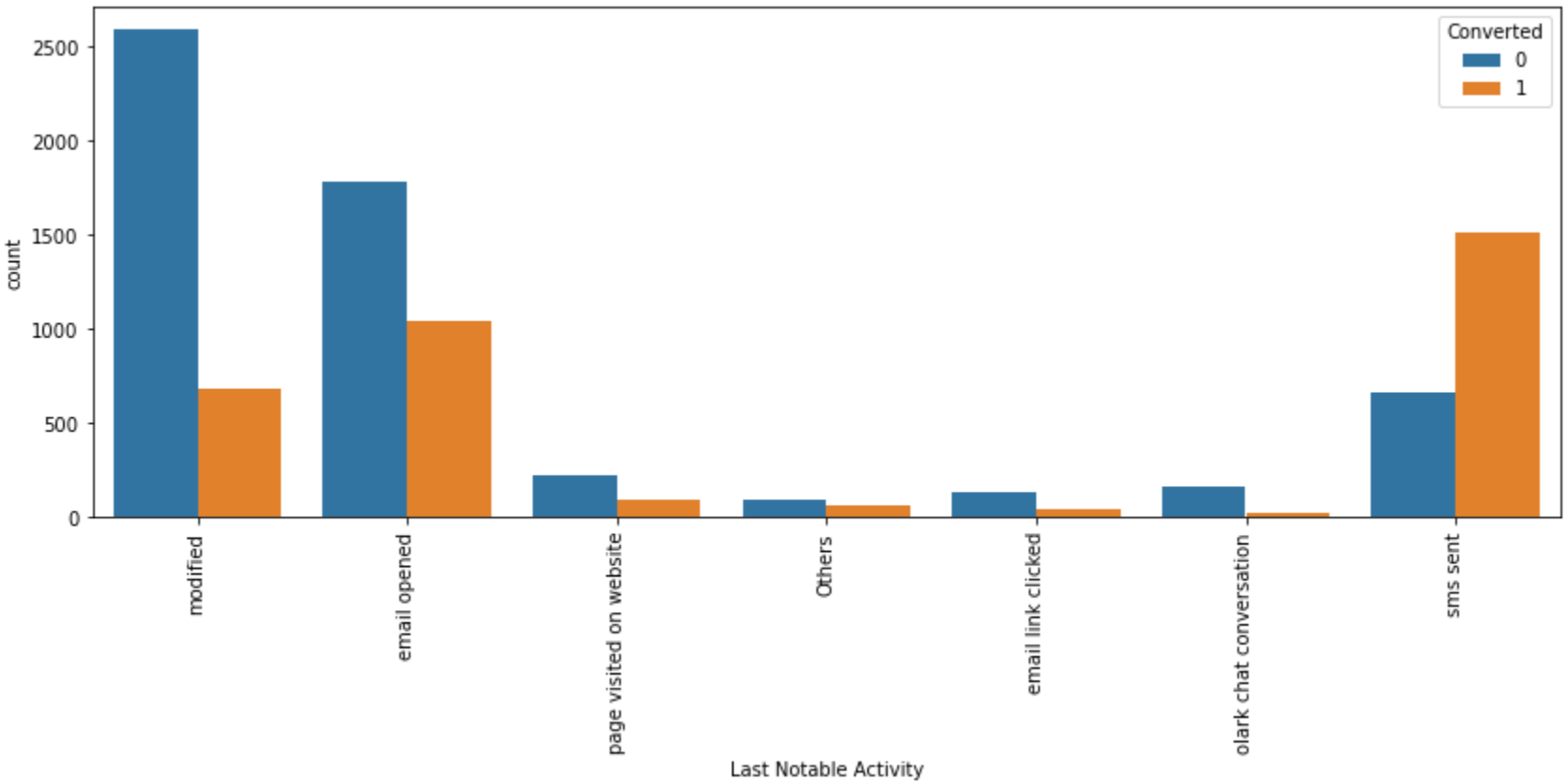
The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified 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%.

Now, although 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’.

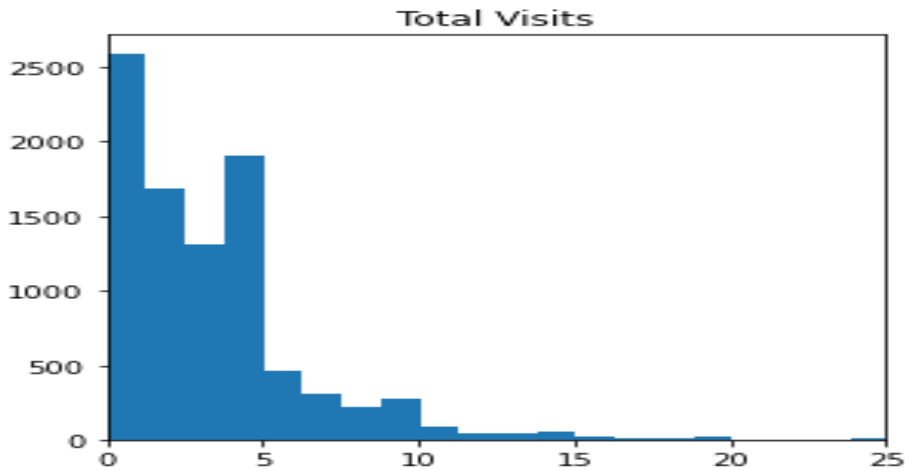
Approach

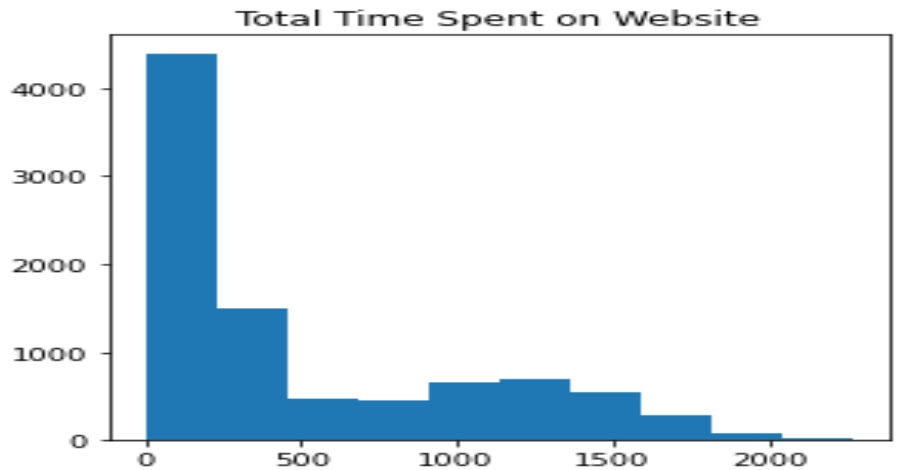
- Loading and cleaning data
- Removal of duplicates
- Cleaning the data frame
- Replacing values to NA
- Checking null values
- Displaying data into graphical form for more understanding
- Understanding the data and key factors for conversion of the leed
- Model building
- Checking and understanding the model
- Checking the accuracy of the model

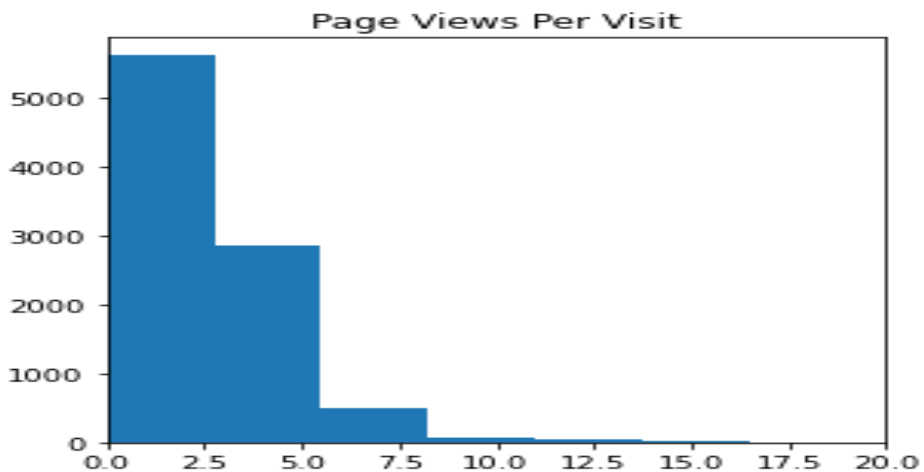
Variables effecting the last Notable activity



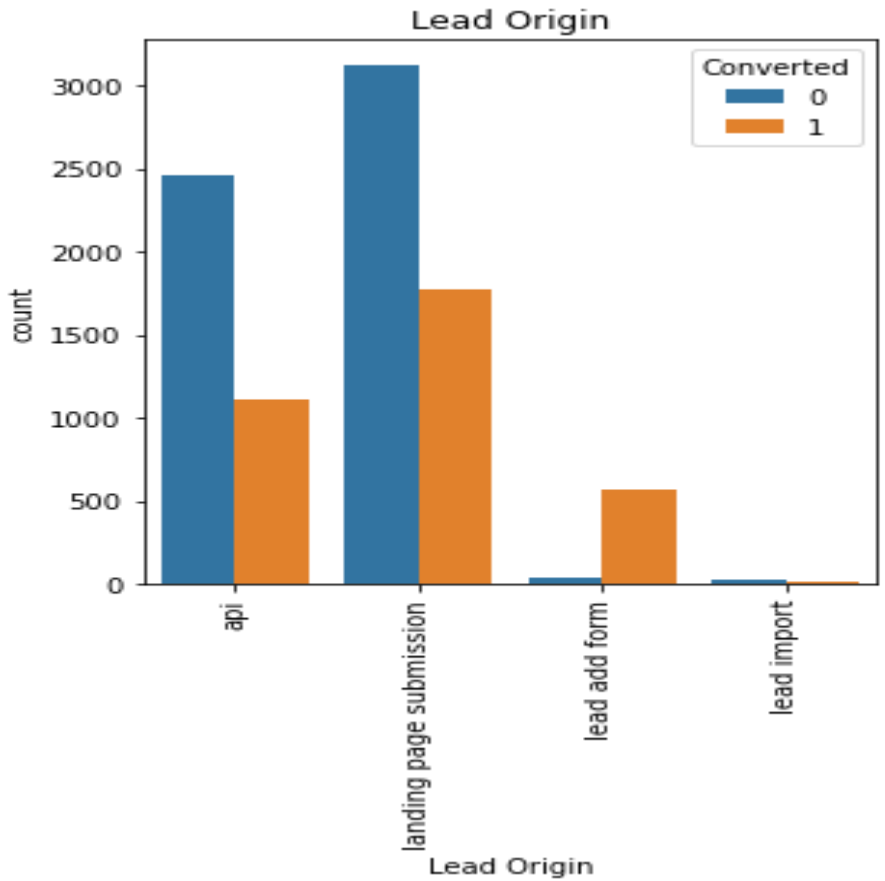
Variable contributing on conversion rate

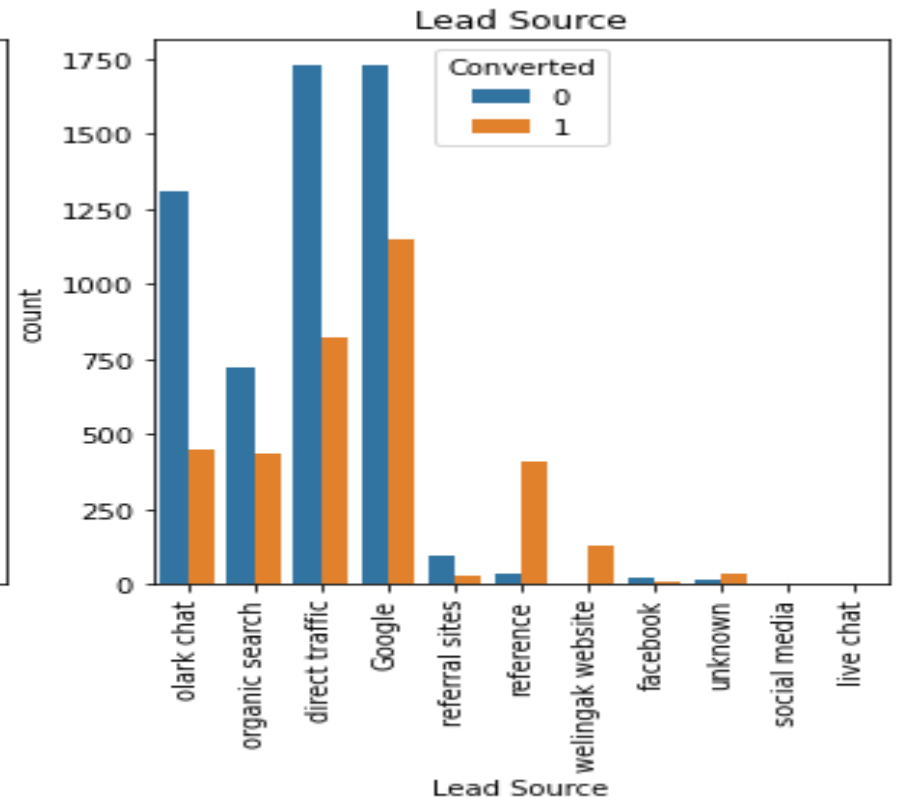


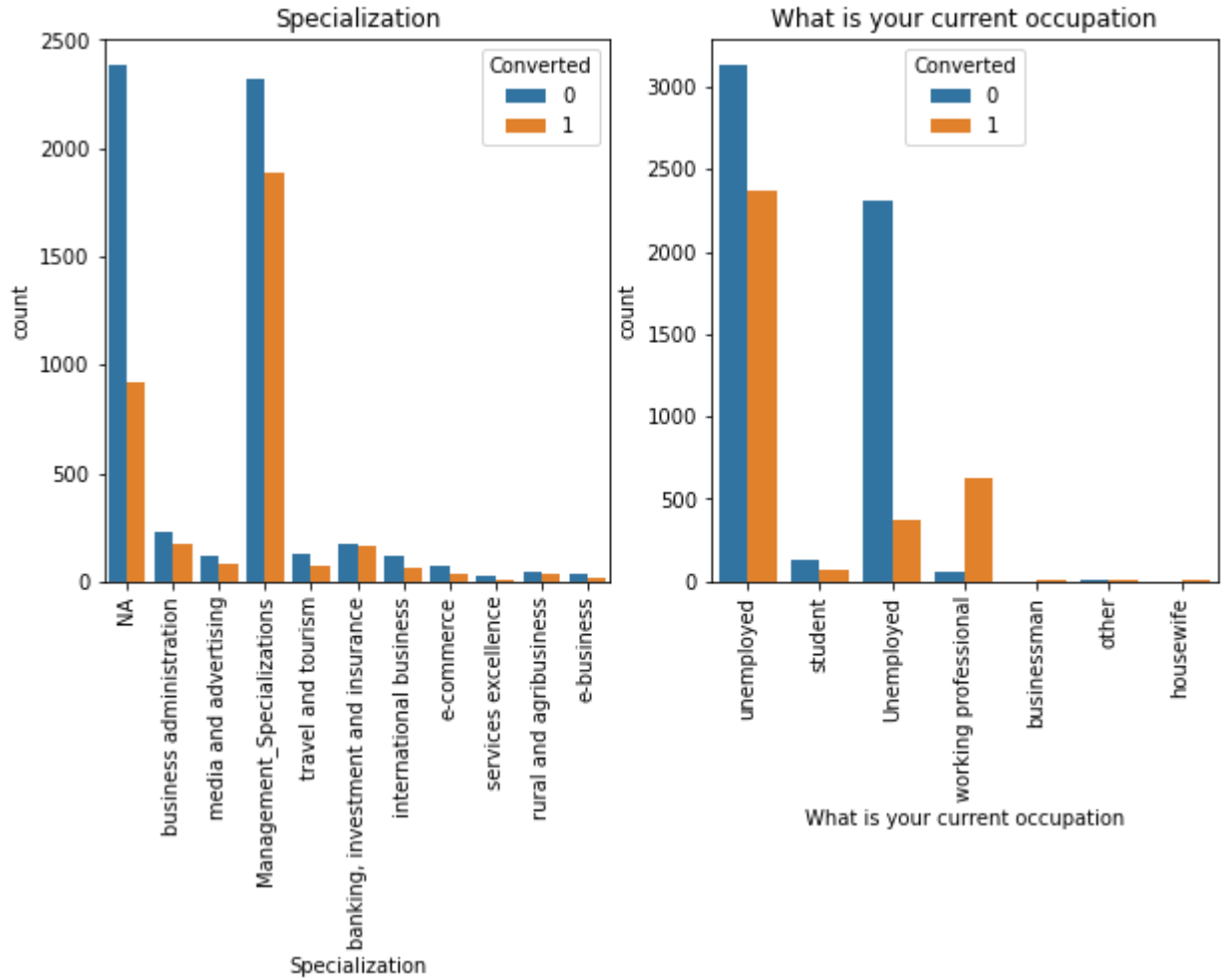




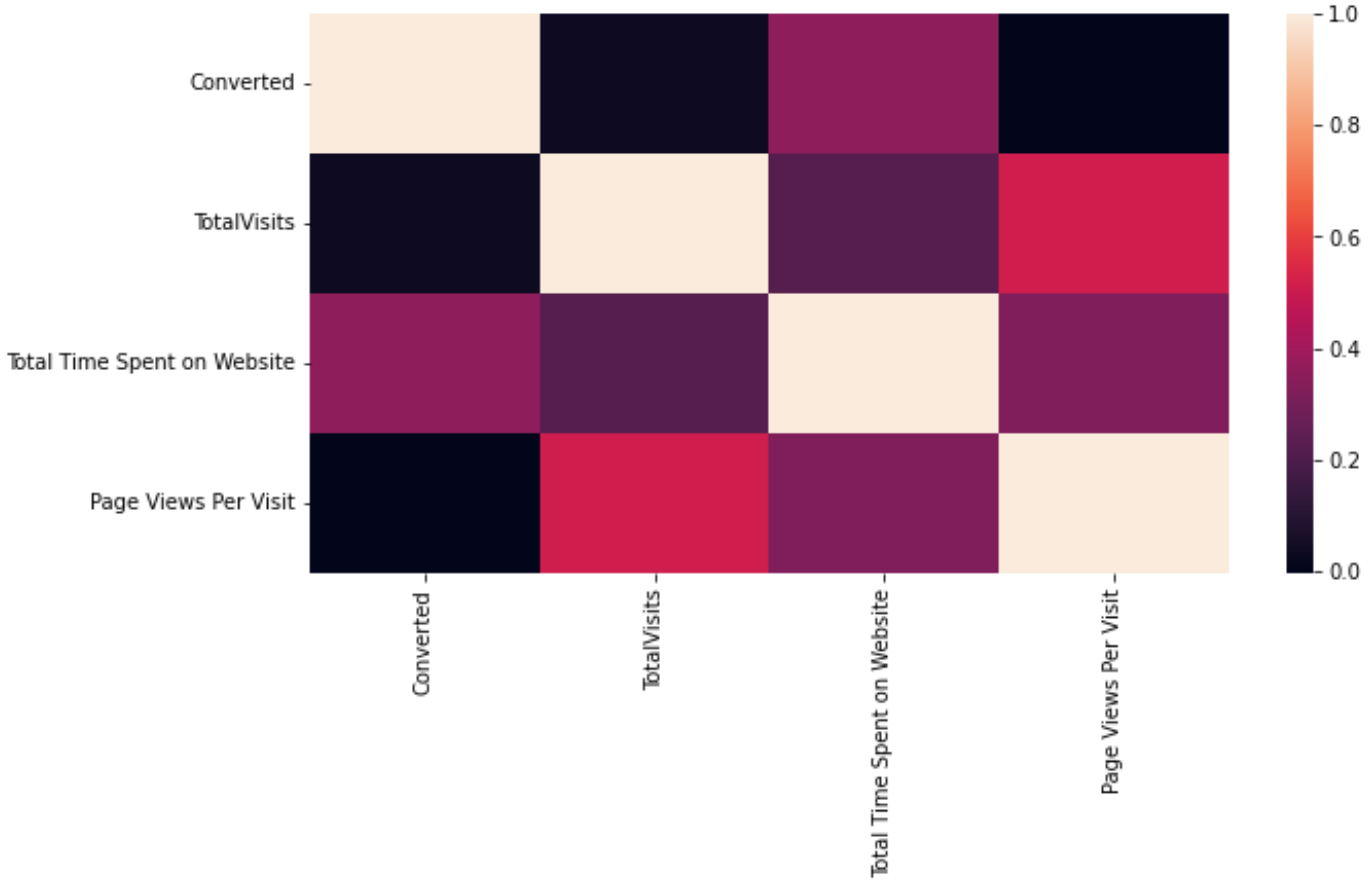
Variables effecting Lead origin and Lead source



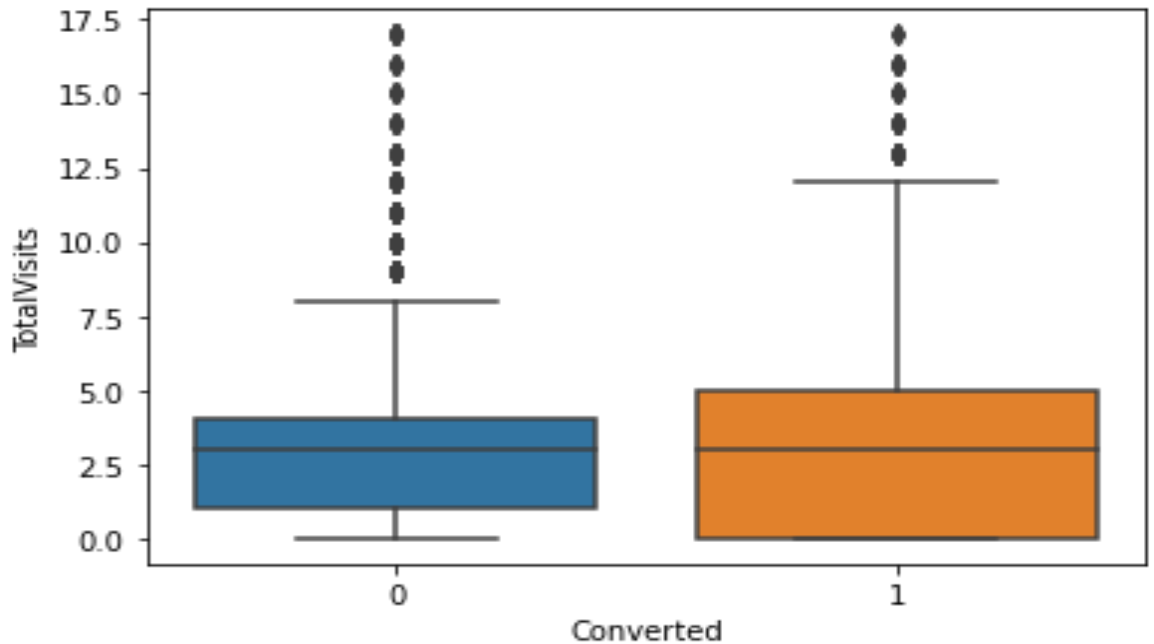



Specialization
Converted
0
1
count
2500
2000
1500
1000
500
0
NA
business administration
media and advertising
Management_Specializations
travel and tourism
banking, investment and insurance
international business
e-commerce
services excellence
rural and agribusiness
e-business
Specialization
What is your current occupation
Converted
0
1
count
3000
2500
2000
1500
1000
500
0
unemployed
student
Unemployed
working professional
businessman
other
housewife
What is your current occupation

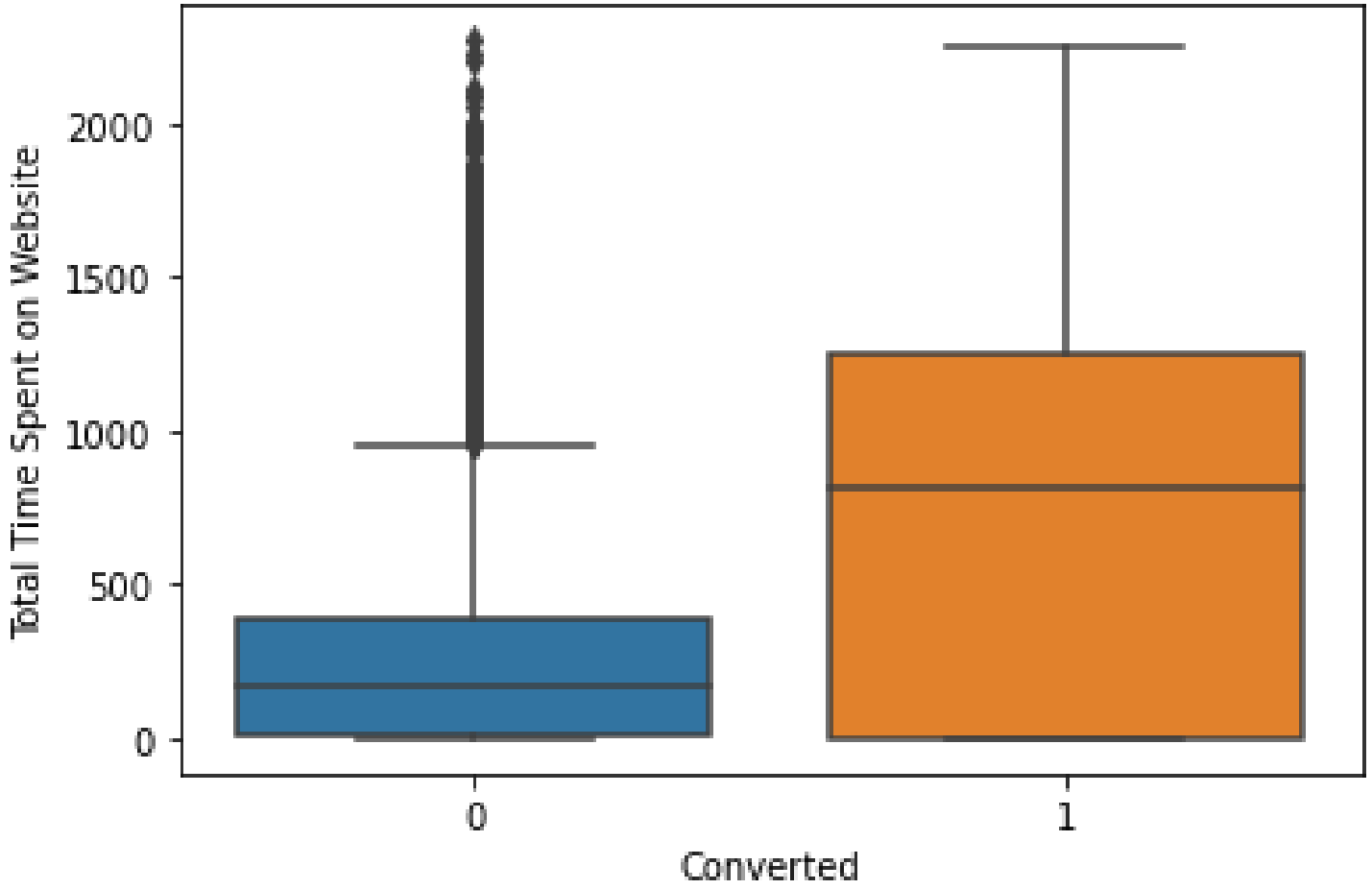
Heat map diagram representing the various variable interdependency over each other.



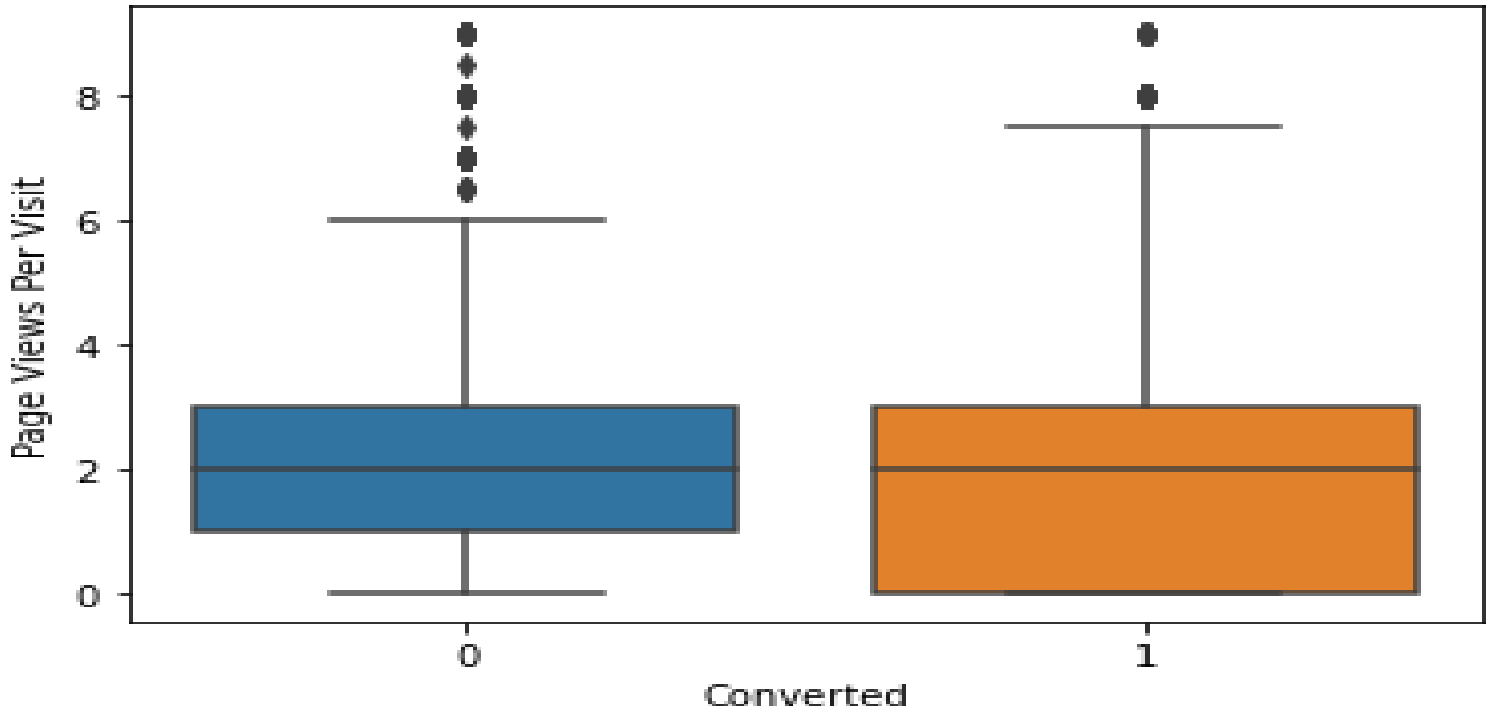
Total visits conversion representations



Lead conversion for Total time conversion

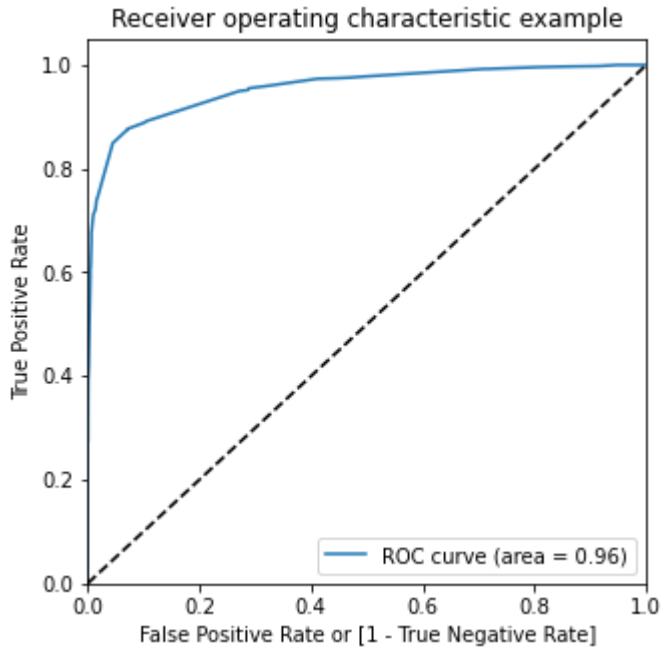


Lead conversion for page per view



ROC - Curve

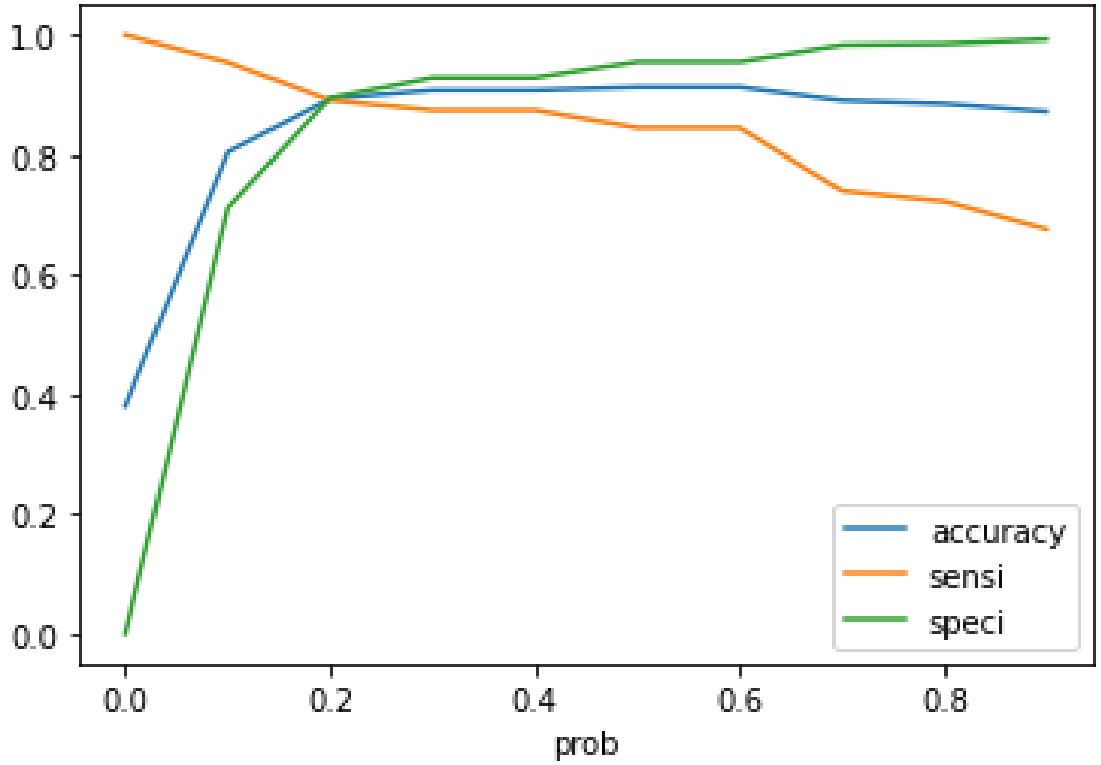
The ROC Curve should be a value close to 1. We are getting a good value of 0.96 indicating a good predictive model.



Observation:

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

Accuracy : 89.94% Sensitivity : 89.20%
Specificity : 90.39%



Final Observation:

Let us compare the values obtained for Train & Test:

Train Data:

Accuracy : 89.38% Sensitivity : 89.14% Specificity : 89.54%

Test Data:

Accuracy : 89.94% Sensitivity : 89.20% Specificity : 90.39%

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