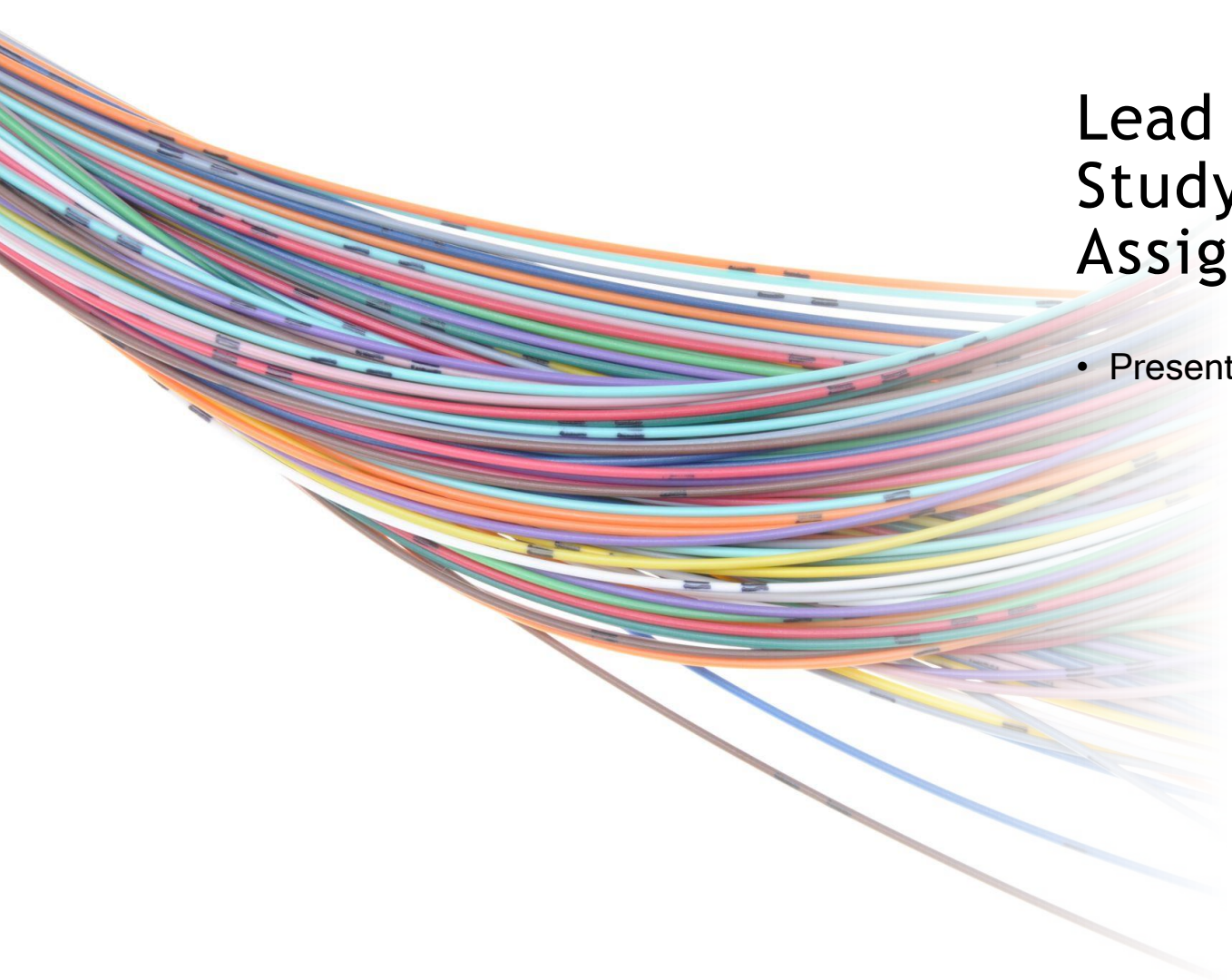


Lead Scoring Case Study - Assignment

- Presenting by
  - Naveen B Gotur
  - Pilla Naveen Kumar
  - Naveen Soni

Problem Statement

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.

Goals of the Case Study

There are quite a few goals for this case study:

1. 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.
2. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

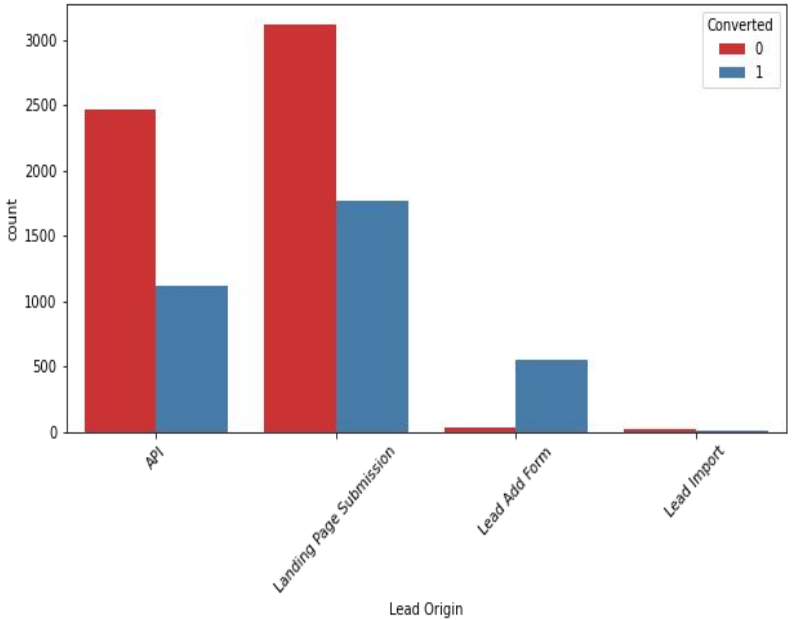
Lead Origin

1)API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.

2)Lead Add Form has more than 90% conversion rate but count of lead are not very high.

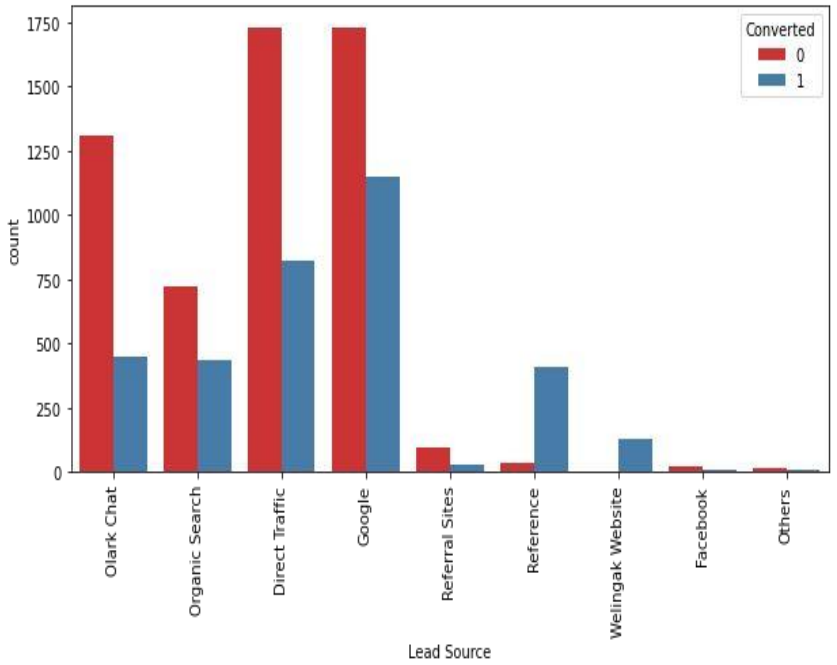
3)Lead Import are very less in count.

To improve overall lead conversion rate, we need to focus more on improving lead converion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



Lead Source

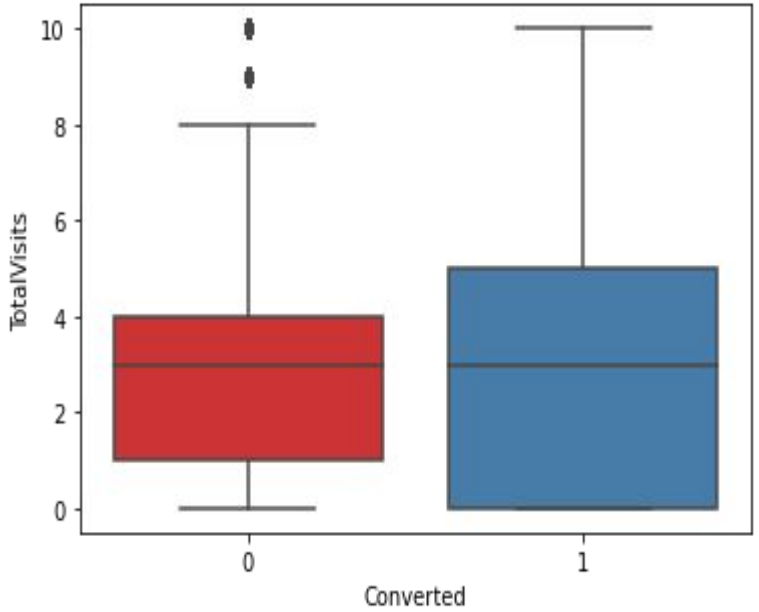
1) Google and Direct traffic generates maximum number of leads. Conversion Rate of reference leads and leads through welingkar website is high
2) To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and generate more leads from reference and welingkar website.



Total Visits

1)Median for converted and unconverted leads are the same.

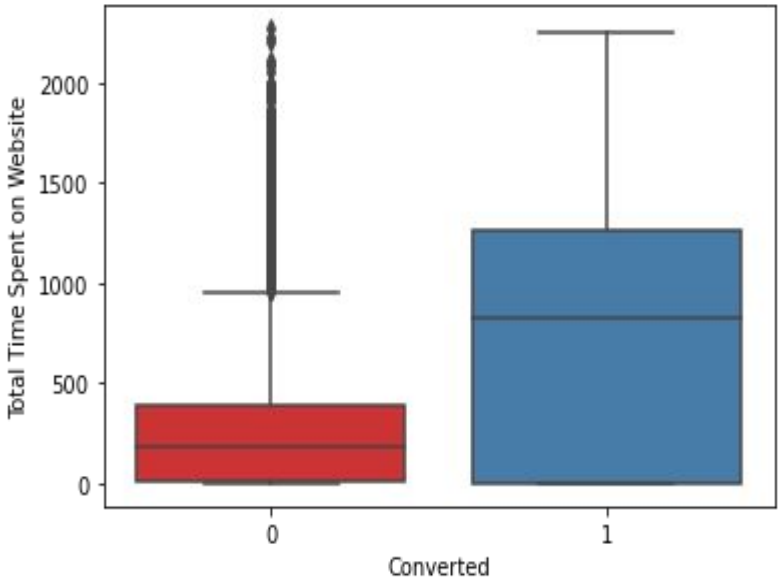
2)Nothing can be concluded on the basis of Total Visits.



Total Time Spent on Website

1)Leads spending more time on the weblise are more likely to be converted.

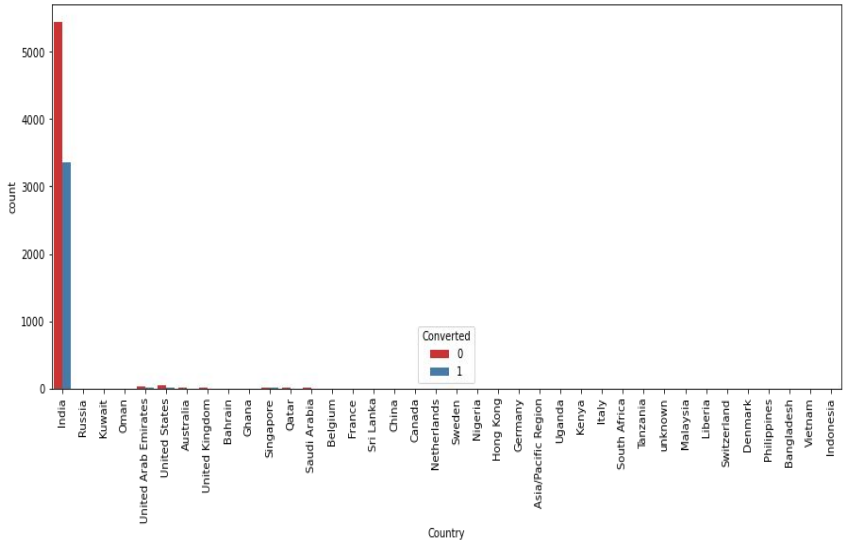
2)Website should be made more engaging to make leads spend more time



Last Activity

1)Most of the lead have their Email opened as their last activity.

2)Conversion rate for leads with last activity as SMS Sent is almost 60%.

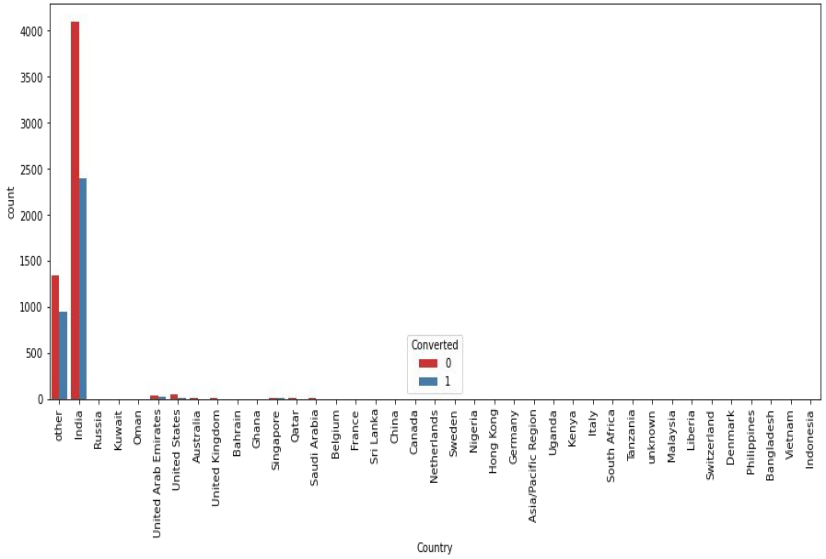


Country

1) Most values are 'India' no such inference can be drawn

2) Other country no interest in online study

3) we have to more focus others country to reach our education system to people

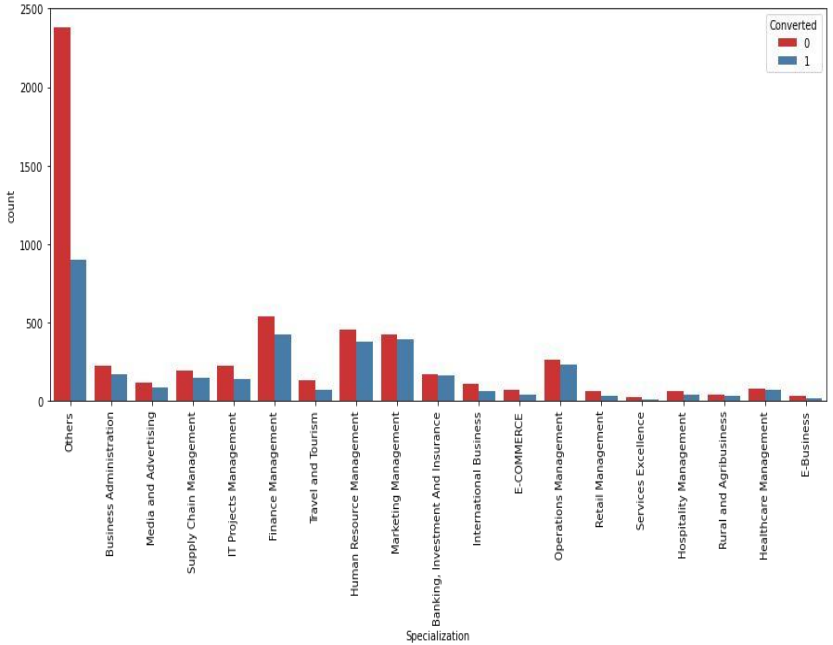


Specialization

Focus should be more on the Specialization with high conversion rate

1) These three people choose specialization
   a) Finance Management
   b) Human Resource Management
   c) Marketing Management

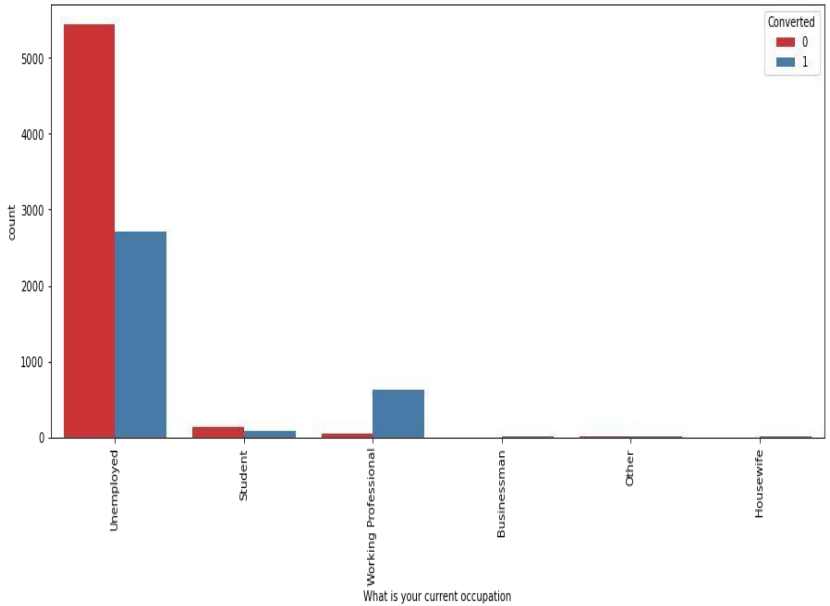
2) Major of the people looking for other course



What is your current occupation

1) Working Professionals going for the course have high chances of joining it.

2) Unemployed leads are the most in numbers but has around 30-35% conversion rate.



Checking for VIF values:

Since the P-values of all variables is 0 and VIF values are low for all the variables, model-9 is our final model. We have 12 variables in our final model.

	Features	VIF
9	Specialization_Others	2.16
3	Lead Source_Olark Chat	2.03
11	Last Notable Activity_Modified	1.78
2	Lead Origin_Landing Page Submission	1.69
6	Last Activity_Olark Chat Conversation	1.59
8	Last Activity_SMS Sent	1.56
1	Total Time Spent on Website	1.29
4	Lead Source_Reference	1.24
10	What is your current occupation_Working Profes...	1.18
0	Do Not Email	1.13
5	Lead Source_Welingak Website	1.09
7	Last Activity_Other_Activity	1.01

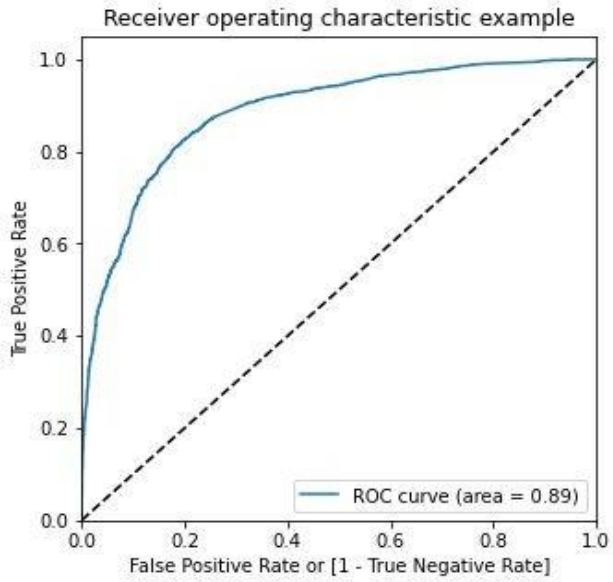
Plotting the ROC Curve

1)It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).

2)The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.

3)The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.

4)Since we have higher (0.89) area under the ROC curve , therefore our model is a good one.



Results :

1) Comparing the values obtained for Train & Test:

Train Data:

1)Accuracy : 81.0 % 2)Sensitivity : 81.7 % 3)Specificity : 80.6 %

Test Data:

1)Accuracy : 80.4 % 2)Sensitivity : 80.4 % 3)Specificity : 80.5 %

Thus we have achieved our goal of getting a ballpark of the target lead conversion rate to be around 80% . 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 to get a higher lead conversion rate of 80%.