LEAD SCORING CASE_STUDY

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’. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Lead Origin: Starting with Lead Origin, Landing Page submissions are highest.

Lead Source: It is found out that Google is the primary and most important place we put it some spending which is going to bring us returns.

Communication: Most of the leads are positive towards the emails that are being sent by us. So I feel Emails are one of the form of communication. Though 99.9 percent people said yes to Phone Call, Phone calls has to primary communication.

Reason to Join Us: Whoever gave the clear response to us, 99.9% leads are joining us for better career prospects. So, Aim is to provide Quality and if possible affordable education to our students.

So, we spend more money on catchy advertisement like the one made by Up Grad team.

Give new hopes to the people who are unwillingly continuing in their careers and help them in career transition.

How they heard of Our X Education: I see most of the students do not read News Paper or any article in News Paper might go our money in vain. It shows most of our learners are social media and mobile frenzy, We shall most concentrate our spending in online advertising rather than any newspapers.


Converted 1 0.023 0.36 -0.015 -0.04 -0.051 0.32 -0.024 0.013 0.026 -0.13 -0.0059 -0.21 -0.19 -0.077 -0.28 0.34 0.38 -0.074 -0.062 0.048 -0.24 -0.074 0.35
TotalVisits 0.023 1 0.21 0.52 0.21 0.3 -0.19 -0.052 -0.0073 0.11 -0.3 0.2 -0.14 0.085 0.036 0.052 -0.048 -0.0051 -0.029 -0.23 0.078 -0.12 0.13 0.0019
Total Time Spent on Website 0.36 0.21 1 0.3 0.17 0.28 -0.14 -0.035 0.033 0.2 -0.38 0.096 -0.19 -0.011 0.0042 -0.056 0.067 0.12 -0.03 -0.23 0.1 -0.11 -0.021 0.13
Page Views Per Visit -0.015 0.52 0.3 1 0.28 0.49 -0.3 -0.078 -0.011 0.21 -0.49 0.31 -0.22 0.028 0.022 0.12 -0.073 -0.024 -0.036 -0.33 0.13 -0.11 -0.025 0.064
A free copy of Mastering The Interview -0.04 0.21 0.17 0.28 1 0.57 -0.18 -0.052 -0.007 -0.32 -0.29 0.15 -0.19 0.07 0.045 0.075 -0.13 -0.005 0.04 -0.36 0.11 -0.085 0.006 0.029
Lead Origin_Landing Page Submission -0.051 0.3 0.28 0.49 0.57 1 -0.31 -0.082 -0.011 0.083 -0.51 0.058 -0.3 0.12 0.047 0.13 -0.15 -0.018 0.00037 -0.62 0.22 -0.1 -0.026 0.058
Lead Origin_Lead Add Form 0.32 -0.19 -0.14 -0.3 -0.18 -0.31 1 -0.022 -0.003 -0.17 -0.14 -0.11 -0.093 -0.075 0.051 -0.082 0.042 0.24 -0.021 0.1 -0.03 -0.013 -0.038 0.093
Lead Origin_Lead Import -0.024 -0.052 -0.035 -0.078 -0.052 -0.082 -0.022 1 0.00081 -0.049 -0.037 -0.029 -0.027 -0.034 -0.013 0.054 -0.018 0.0015 -0.014 -0.055 0.062 0.037 -0.024 -0.04
Lead Origin_Quick Add Form 0.013 -0.0073 0.033 -0.011 -0.007 -0.011 -0.003 -0.00081 1 0.015 -0.005 -0.0039 -0.0036 0.02 -0.0057 -0.004 0.0087 -0.0048 -0.0019 0.01 -0.004 0.014 -0.0033 -0.0058
Lead Source_Google 0.026 0.11 0.2 0.21 -0.32 0.083 -0.17 -0.049 0.015 1 -0.33 -0.26 -0.088 0.017 3.7e-05 0.021 0.02 -0.0085 -0.043 -0.15 0.042 -0.019 -0.026 0.0073
Lead Source_Olark Chat -0.13 -0.3 -0.38 -0.49 -0.29 -0.51 -0.14 -0.037 -0.005 -0.33 1 -0.18 0.43 -0.13 -0.097 -0.099 0.13 -0.11 0.025 0.39 -0.15 0.091 0.06 -0.096
Lead Source_Organic Search -0.0059 0.2 0.096 0.31 0.15 0.058 -0.11 -0.029 -0.0039 -0.26 -0.18 1 -0.074 0.045 0.026 0.048 -0.038 -0.0064 -0.028 0.055 -0.023 -0.018 -0.017 0.038
Last Activity_Olark Chat Conversation -0.21 -0.14 -0.19 -0.22 -0.19 -0.3 -0.093 -0.027 -0.0036 -0.088 0.43 -0.074 1 -0.18 -0.0059 -0.1 0.045 -0.1 0.043 0.26 -0.11 0.32 0.12 -0.19
Last Activity_Others -0.19 0.085 -0.011 0.028 0.07 0.12 -0.075 -0.034 0.02 0.017 -0.13 0.045 -0.18 1 0.13 -0.033 -0.12 -0.067 0.052 -0.072 0.0083 0.33 0.43 -0.29
Tags_Other Reasons -0.077 0.036 0.0042 0.022 0.045 0.047 0.051 -0.013 -0.0057 3.7e-05 -0.097 0.026 -0.0059 0.13 1 -0.21 -0.66 0.022 -0.064 -0.051 0.015 0.2 0.011 -0.14
Tags_Ringing -0.28 0.052 -0.056 0.12 0.075 0.13 -0.082 0.054 -0.004 0.021 -0.099 0.048 -0.1 -0.033 -0.21 1 -0.46 -0.084 -0.069 -0.072 0.018 -0.12 -0.033 0.072
Tags_Will revert after reading the email 0.34 -0.048 0.067 -0.073 -0.13 -0.15 0.042 -0.018 0.0087 0.02 0.13 -0.038 0.045 -0.12 -0.66 -0.46 1 0.056 -0.17 0.1 -0.021 -0.12 -0.0051 0.13
Lead Profile_Potential Lead 0.38 -0.0051 0.12 -0.024 -0.005 -0.018 0.24 0.0015 -0.0048 -0.0085 -0.11 -0.0064 -0.1 -0.067 0.022 -0.084 0.056 1 -0.082 -0.14 0.1 -0.066 -0.033 0.083
Lead Profile_diploma_dual_n_SomeSchool_lead -0.074 -0.029 -0.03 -0.036 0.04 0.00037 -0.021 -0.014 -0.0019 -0.043 0.025 -0.028 0.043 0.052 -0.064 -0.069 -0.17 -0.082 1 -0.0079 -0.0024 0.04 0.031 -0.089
City_Other Cities -0.062 -0.23 -0.23 -0.33 -0.36 -0.62 0.1 -0.055 0.01 -0.15 0.39 0.055 0.26 -0.072 -0.051 -0.072 0.1 -0.14 -0.0079 1 -0.4 0.07 0.037 -0.022
City_Other Cities of Maharashtra 0.048 0.078 0.1 0.13 0.11 0.22 -0.03 0.062 -0.004 0.042 -0.15 -0.023 -0.11 0.0083 0.015 0.018 -0.021 0.1 -0.0024 -0.4 1 -0.041 -0.03 0.0074
Last Notable Activity_Modified -0.24 -0.12 -0.11 -0.11 -0.085 -0.1 -0.013 0.037 0.014 -0.019 0.091 -0.018 0.32 0.33 0.2 -0.12 -0.12 -0.066 0.04 0.07 -0.041 1 -0.24 -0.42
Last Notable Activity_Other_Activity -0.074 0.13 -0.021 -0.025 0.006 -0.026 -0.038 -0.024 -0.0033 -0.026 0.06 -0.017 0.12 0.43 0.011 -0.033 -0.0051 -0.033 0.031 0.037 -0.03 -0.24 1 -0.17
Last Notable Activity_SMS Sent 0.35 0.0019 0.13 0.064 0.029 0.058 0.093 -0.04 -0.0058 0.0073 -0.096 0.038 -0.19 -0.29 -0.14 0.072 0.13 0.083 -0.089 -0.022 0.0074 -0.42 -0.17 1
Converted
TotalVisits
Total Time Spent on Website
Page Views Per Visit
A free copy of Mastering The Interview
Lead Origin_Landing Page Submission
Lead Origin_Lead Add Form
Lead Origin_Lead Import
Lead Origin_Quick Add Form
Lead Source_Google
Lead Source_Olark Chat
Lead Source_Organic Search
Last Activity_Olark Chat Conversation
Last Activity_Others
Tags_Other Reasons
Tags_Ringing
Tags_Will revert after reading the email
Lead Profile_Potential Lead
Lead Profile_diploma_dual_n_SomeSchool_lead
City_Other Cities
City_Other Cities of Maharashtra
Last Notable Activity_Modified
Last Notable Activity_Other_Activity
Last Notable Activity_SMS Sent
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Heat Map Inferences:

The First Column ‘Converted’, gives you correlations between all the potential variables, but in that in the business perspective, we shall first concentrate on what are the variables which contribute the most in model building. The Thick Green color indicates most important variable.

The potential leads in Lead Profile column are the first dummy column on which we should concentrate the highest on.

And then concentrate most on students who spent more time on website.

And on the dummy column Tazewell revert you back after reading your email.

Results: CONFUSION MATRIX:

Predicted Actual	Not Converted	Converted
Not Converted	3415	587
Converted	321	2145

CONFUSION MATRIX:

TRAINING SET

Predicted	Not Converted	Converted
Actual		
Not Converted	3415	587
Converted	321	2145

- The Probability Threshold value is chosen by taking the tradeoff between Accuracy, Sensitivity and specificity

Test SET

Predicted	Not Converted	Converted
Actual		
Not Converted	1494	183
Converted	199	896

- The Probability threshold value is chosen by taking the tradeoff between Precision and Recall

Results and Conclusions:

- The model that we have developed, has an accuracy around 85.96% with training set and with test set also it is almost same (86.21%)
- Most importantly the precision Score on test set is around 83%, which is more than the CEO expected from a model. The precision Score defines that out of 100 assumed positive leads, 83 of them are converted into which is wonderful.