PGDDS – Sep 2018

Lead Scoring Case Study

Group Members:

1. Ankit Saini
2. Diwakar Sharma
3. Arihant Sethia
4. Aditi Nag

March 3, 2019

PROBLEM STATEMENT, BUSINESS GOAL, ASSUMPTIONS

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.

BUSINESS GOAL

- The company requires to build a model wherein we 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%.

ASSUMPTIONS

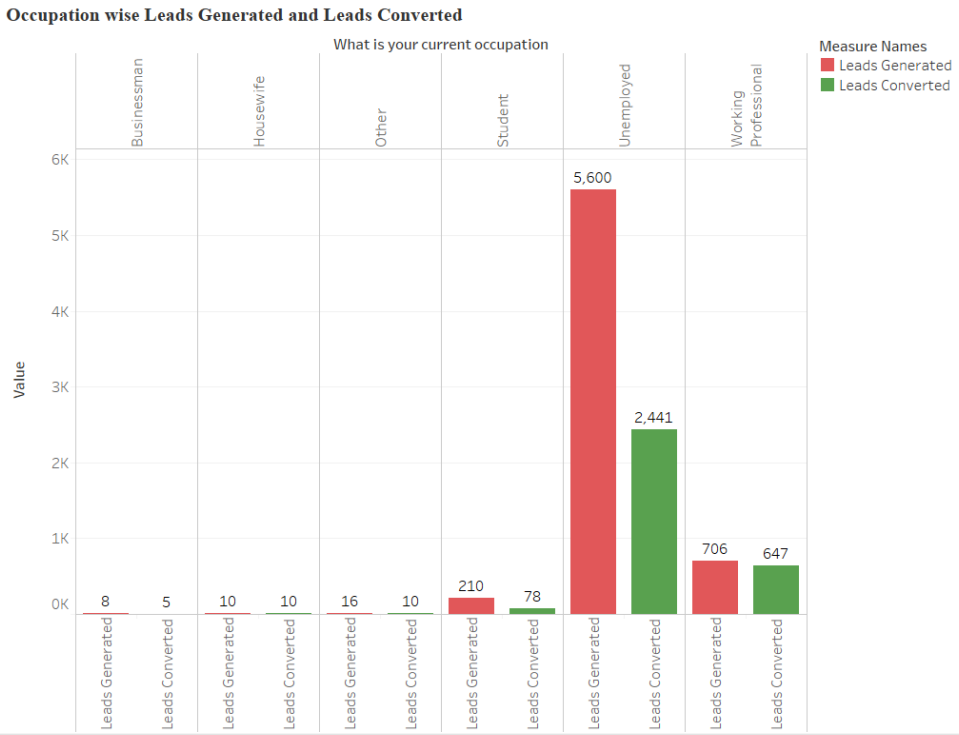
- There are many columns in the dataset which have all values as null or just have a single value populated across it. We have not considered these in our analysis and dropped such columns from the data frame.

Problem Solving Methodology

1. Data Exploration.
2. Dropping the columns with greater than 30 % null values and single values.
3. Converting ‘Select’ to null and dropping columns with greater than 30% null values.
4. Imputing some columns having null values and dropping columns which are not relevant for the analysis.
5. Removing outliers and creating Dummy variables.
6. Removal of correlation and Model building for prediction of potential leads.
7. Based on the above analysis identifying the potential leads and assigning the lead score.

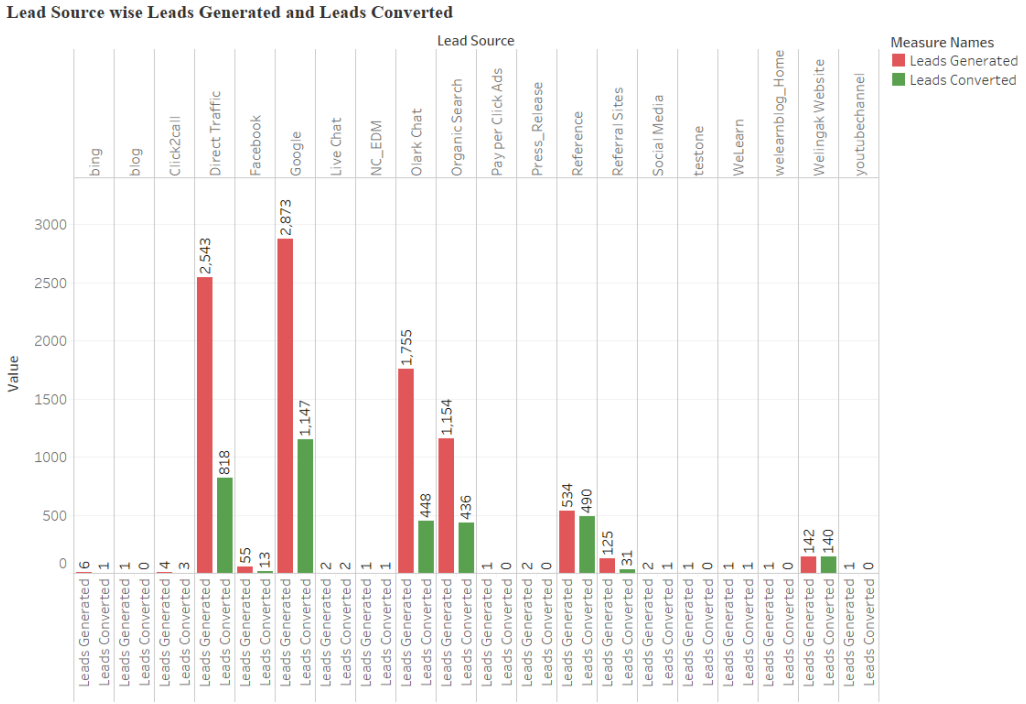
Occupation wise Leads Generated and Leads Converted

There is a good amount of Leads generated across working professionals and has a very high conversion percentage.



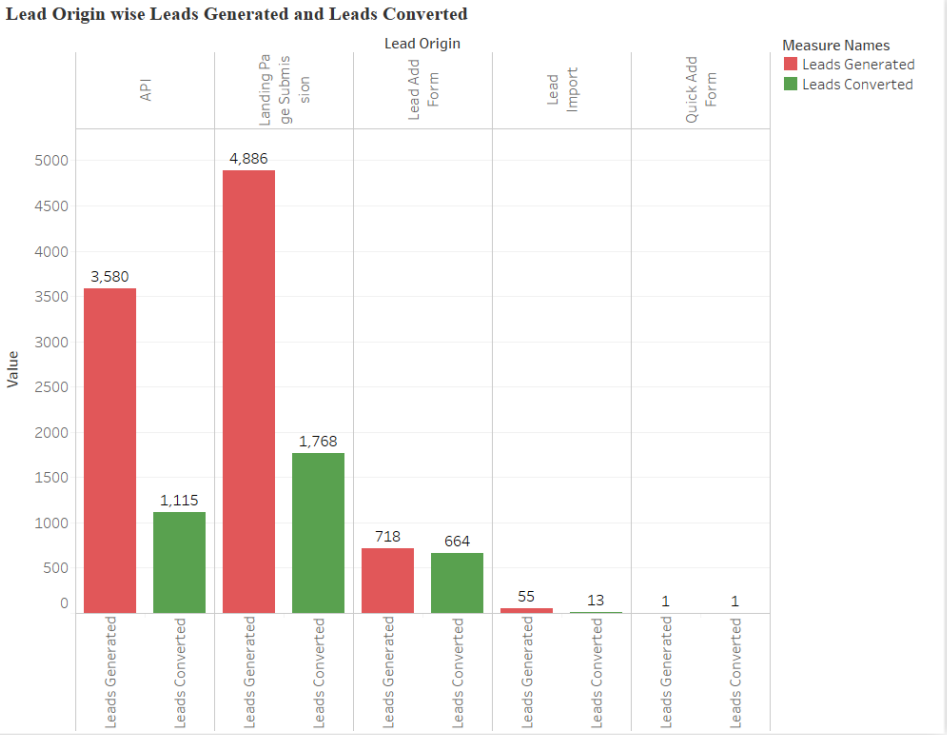
Lead Source wise Leads Generated and Leads Converted

There is a good amount of Leads generated from Reference & Welingak website and has a very high conversion percentage.



Lead Origin wise Leads Generated and Leads Converted

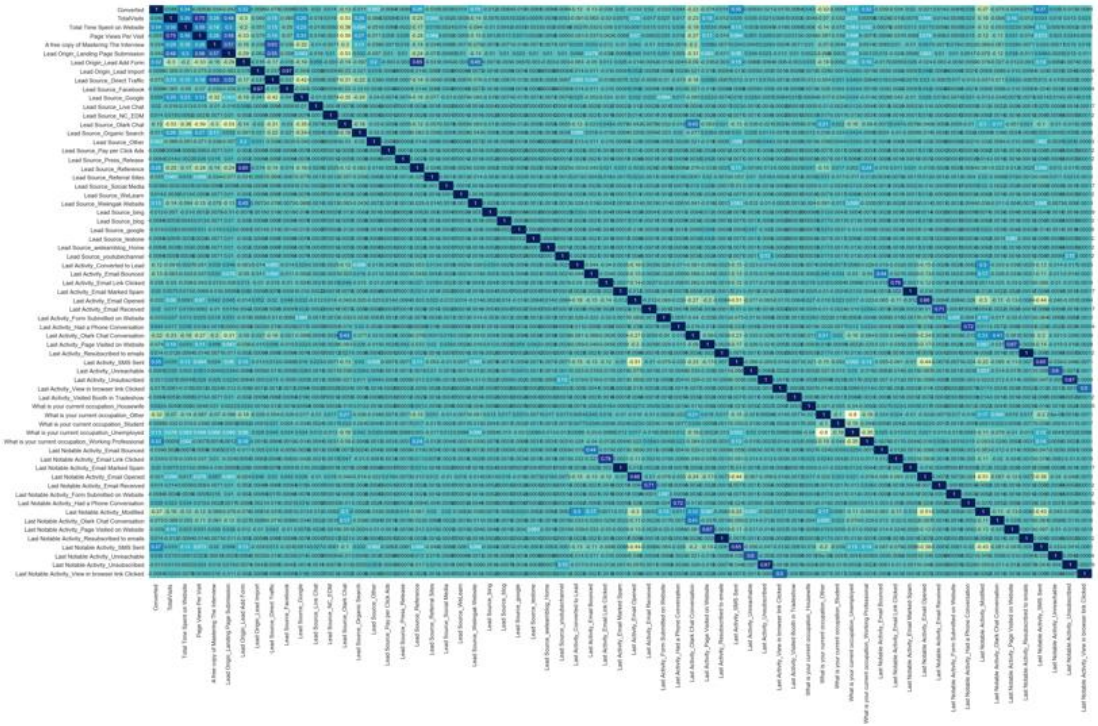
There is a good amount of Leads generated from Lead Add Form and has a very high conversion percentage.



Model Building :
Correlation Matrix before dropping highly correlated variables

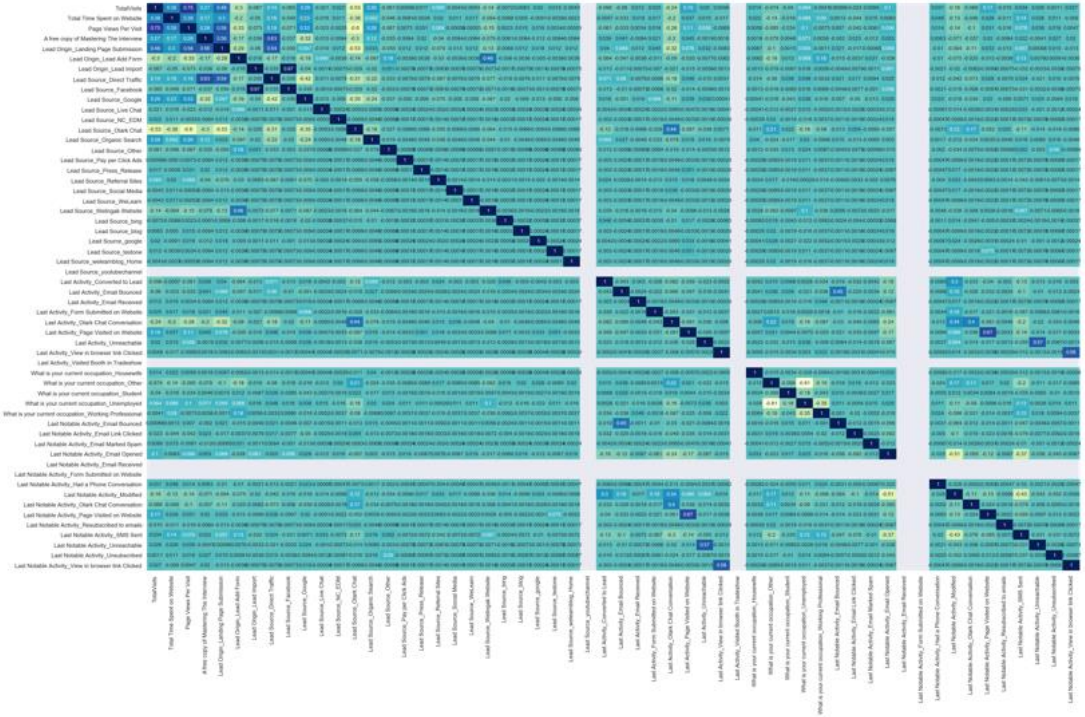
Below are the pairs of highly correlated features:-

- "Last Notable Activity_Email Marked Spam" & "Last Activity_Email Marked Spam"
- "Last Activity_Resubscribed to emails" & "Last Notable Activity_Resubscribed to emails"
- "Last Notable Activity_Unsubscribed" & "Last Activity_Unsubscribed"
- "Last Notable Activity_SMS Sent" & "Last Activity_SMS Sent"
- "Last Activity_Email Opened" & "Last Notable Activity_Email" Opened"
- "Lead Origin_Lead Add Form" & "Lead Source_Reference"
- "Last Notable Activity_Email Link Clicked" & "Last Activity_Email Link Clicked"
- "Last Notable Activity_Had a Phone Conversation" & "Last Activity_Had a Phone Conversation"



Model Building :

Correlation Matrix after dropping highly correlated variables

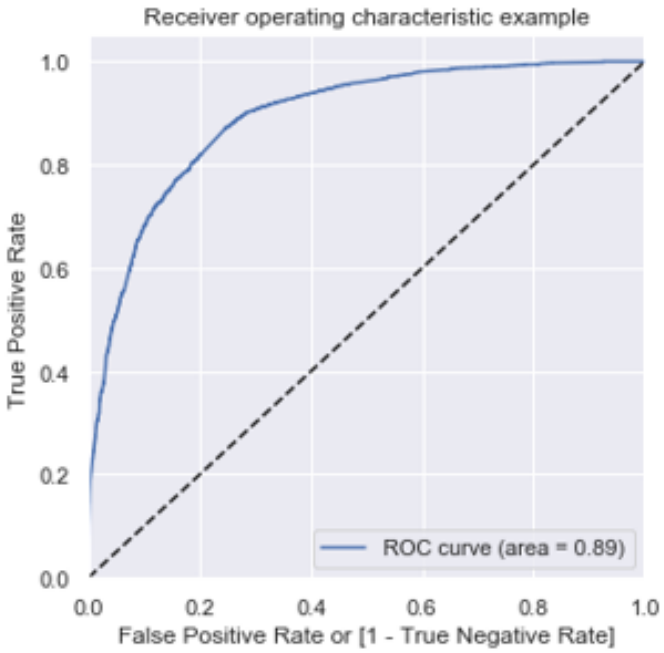


TOP 14 FEATURES OF OUR MODEL

	Features	VIF
1	Lead Origin_Lead Add Form	1.51
10	What is your current occupation_Other	1.46
3	Lead Source_Google	1.39
12	Last Notable Activity_SMS Sent	1.39
2	Lead Source_Direct Traffic	1.34
6	Lead Source_Welingak Website	1.31
0	Total Time Spent on Website	1.24
9	Last Activity_Olark Chat Conversation	1.19
11	What is your current occupation_Working Profes...	1.17
4	Lead Source_Organic Search	1.15
7	Last Activity_Converted to Lead	1.11
8	Last Activity_Email Bounced	1.06
5	Lead Source_Referral Sites	1.01
13	Last Notable Activity_Unreachable	1.01

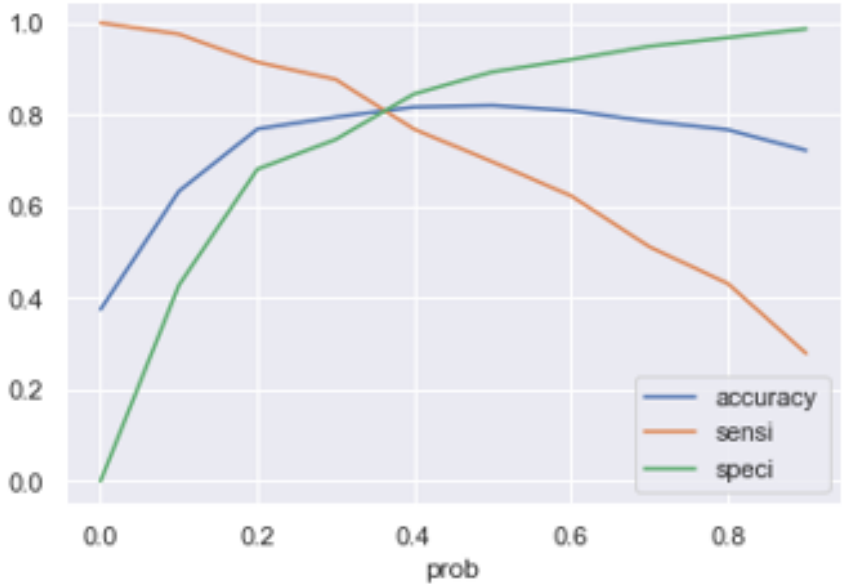
Plotting the ROC curve

The ROC curve is more towards the left-hand border and the top border. Hence , the test that we have performed is nearly accurate.



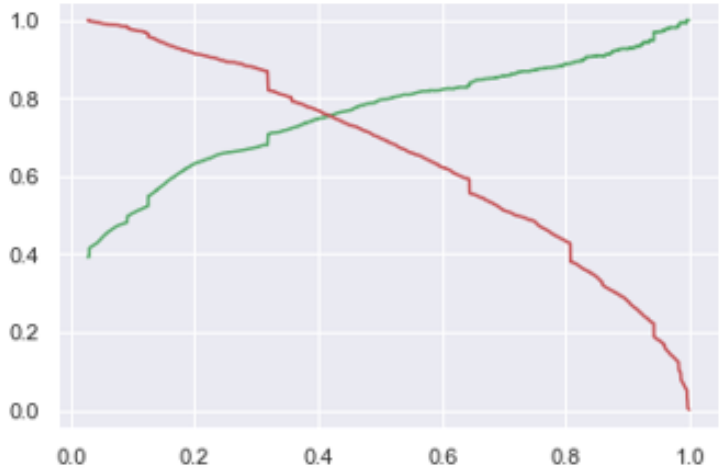
Plotting accuracy, sensitivity and specificity for various probabilities:

The graph shows the cutoff probability which is 0.35. This implies that if the Lead score of the potential customer is 35% or above , they have a high probability of getting Converted.



Precision, Recall and threshold trade off:

Threshold



Green – Precision
Red - Recall

Final Leads DataFrame Head

	Lead Origin	Lead Source	Converted	TotalVisits	Total Time Spent on Website	Page Views Per Visit	Last Activity	What is your current occupation	A free copy of Mastering The Interview	Last Notable Activity	lead_score
0	API	Olark Chat	0	0.0	0	0.0	Page Visited on Website	Unemployed	0	Modified	31.787367
1	API	Organic Search	0	5.0	674	2.5	Email Opened	Unemployed	0	Email Opened	34.159478
2	Landing Page Submission	Direct Traffic	1	2.0	1532	2.0	Email Opened	Student	1	Email Opened	69.058616
3	Landing Page Submission	Direct Traffic	0	1.0	305	1.0	Unreachable	Unemployed	0	Modified	15.507936
4	Landing Page Submission	Google	1	2.0	1428	1.0	Converted to Lead	Unemployed	0	Modified	46.317877

Key Observations

The top three variables which contribute most towards the probability of a lead getting converted are :

- What is your current occupation
- Lead Source
- Lead Origin

The top three categorical/dummy variables which should be focused the most in order to increase the probability of lead conversion :

- Current Occupation – Working professionals : The Company should focus on the Working professionals as they prefer courses which are given via online mode.

- Lead Source – Welingak Website : The Company should put up ads on the Welingak website as it has high probability of conversion and thus generates higher number of potential leads.

- Lead Origin – Lead Add Form : The customers who are identified with Lead Add Form is a potential lead and thus the Company should have a lookout for these customers so as to increase their lead conversion.

X Education has a period of 2 months every year during which they hire some interns. The sales team, in particular, has around team of 10 interns allotted to them. So during this phase, they wish to make the lead conversion more aggressive. So they want almost all of the potential leads (customers who have been predicted as 1 by the model) to be converted and hence, want to make phone calls to as much of such people as possible. Suggest a good strategy they should employ at this stage.

During this period, to make our lead conversion more aggressive, X education should ask their data science team to little lower down the conversion probability cut-off to include even leads having lower conversion lead score, so that more leads can be predicted as converted by model and can be targeted by team of interns. When there is more target leads, there will be more conversion. At this stage, those which are low potential leads should be sent email with links of Welingak website advertisement of program, so that if they visit the same, there will be better chance of conversion.

Also, team should focus on leads in order of the variables which has better probabilities of lead conversion. From the answer above, we got 3 categorical variables which would increase the probability of lead conversion better than the rest. X Education interns should target potential leads who are working professionals, whose lead source is from Welingak website and the lead origin for whom is Lead Add Form.

To increase conversion rate, interns make sure that it is as easy as possible for people to complete their purchases. Same can be done by:

- Providing one-click signup and sign-in options
- Minimizing the form fields
- Simplifying your checkout and payment processes

Also, the leads which has not converted to customers, interns should focus to understand what are their pain points, so that same can be tried to improve.

Similarly, at times, the company reaches its target for a quarter before the deadline. During this time, the company wants the sales team to focus on some new work as well. So during this time, the company's aim is to not make phone calls unless it's extremely necessary, i.e. they want to minimize the rate of useless phone calls. Suggest a strategy they should employ at this stage.

As the Company has already reached its target and found those customers who have already converted, they should try to improvise their experience by doing more interactive sessions such that they do not churn (i.e. do not ask to de-enroll before commencement of course).

Additionally, the sales team can think of strategies so as to implement the following things :

- Providing discounts
- Providing referral bonus by asking the existing leads to bring in more customers who are interested
- Run competition for leads asking them to bring in people who they think might be interested in taking this course
- Running email campaigns to get more extensive feedback.
- Publish testimonials of previous students on their website and other portals like LinkedIn
- Sent emails to candidates of the career transitions or placements through this course,
- ranking and recognition of the course
- Stay in Touch with Converted Leads through Remarkable Presence on Different Social Networking Sites
- Use of Videos to communicate with Students
- Become a 'page one' search result on Google (This means a lot to the one who enrolls as to the people they tells about the course also this makes a big impact)

CONCLUSION

Based on the analysis performed , we can conclude that the following variables are influencing the conversion rate for the X Education Company to a large extent. The Company should analyze and have a lookout for the following traits which could serve as a potential for lead conversion.

Variables

- Current Occupation – The applicants who are currently Working Professionals prefer these online courses as compared to students and those who are unemployed and looking for jobs.
- Lead Source – Welingak website stood out as best source of lead and the company should definitely give ads on this site so as to generate higher number of potential leads.
- Lead Origin – Mostly the candidates who were identified with Lead Add form generated potential leads who ultimately got converted.

Thus , X Education should definitely pay attention and improvise on the above variables as well as create new ways as already mentioned previously such that it can increase its lead conversion rate from around 30% to 80%.