LEAD SCORE CASE STUDY PRESENTATION

GROUP PARTNERS:

SHUBHAM

SHWETA

TIRTHA

PROBLEM STATEMENT:

1) X-Education sells online courses to industry professionals.
2) X-Education gets lots of leads but its lead conversion rate is very poor. The typical lead conversion rate of X-Education is around 30 %.
3) In order to increase the lead conversion rate, company wants to identify most potential leads, also known as “Hot Leads”
4) If they identify sets of leads, the lead conversion rate should go up as the sales team will be more focusing on communication with the potential leads rather than making calls to everyone.

BUSINESS OBJECTIVE:

1) X-Education wants to know most promising leads.
2) For that they want to build a Model which identify potential leads, also known as Hot Leads.
3) Model Lead Conversion rate should be around 80%.

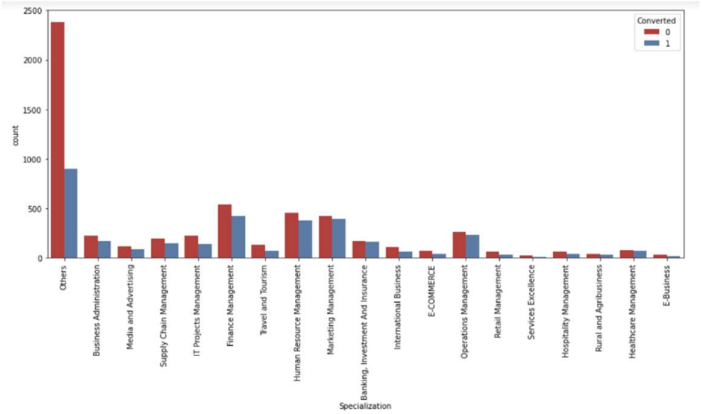
SOLUTION STRATEGY:

1) Importing Data
2) Checking Statistical Summary and Checking missing data percentages.
3) Checking missing values and then dropping them which are more than 40% missing value and prepare the data for further analysis.
4) Checking each column for further analysis.
5) Imputation of the values.
6) Check and handled outliers in the data.
7) Exploratory data Analysis for checking attributes.
8) Prepare the data for model building.

9) Build logistic regression model.
10) Assign lead score to each lead.
11) Test the Model on train set.
12) Evaluate the Model by different measures and metrics.
13) Test the Model on the Test set.
14) Measure the accuracy of the model and other metrics for the evaluation.

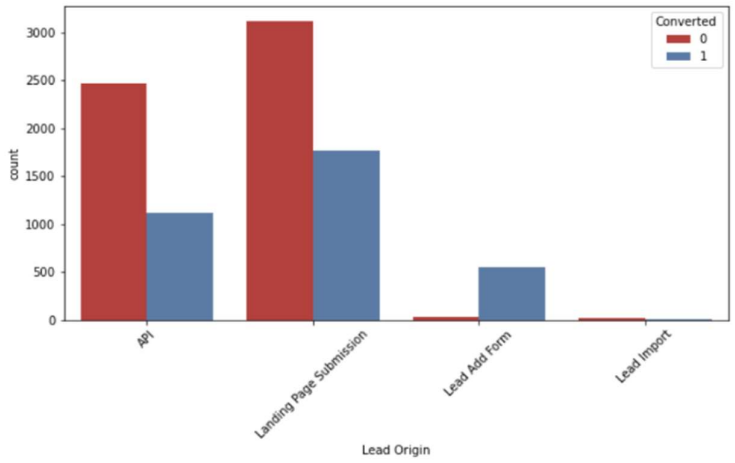
Following insights can be seen during the analysis:

1) Specialization Vs Converted



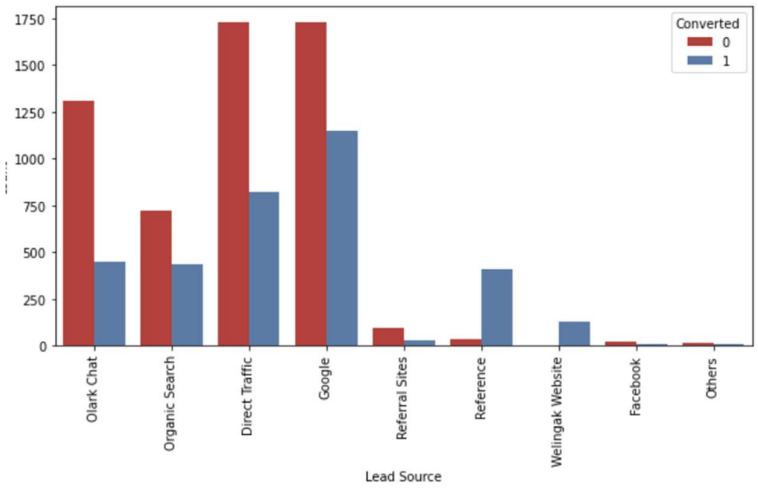
a) As 37% missing values present in the Specialization column.
b) It may be possible that the lead has left this column blank if he may be a student or not having any specialization or his specialization is not there in the options given.
c) So, we can create another category 'Others' for this.

2) Lead Origin Vs Converted



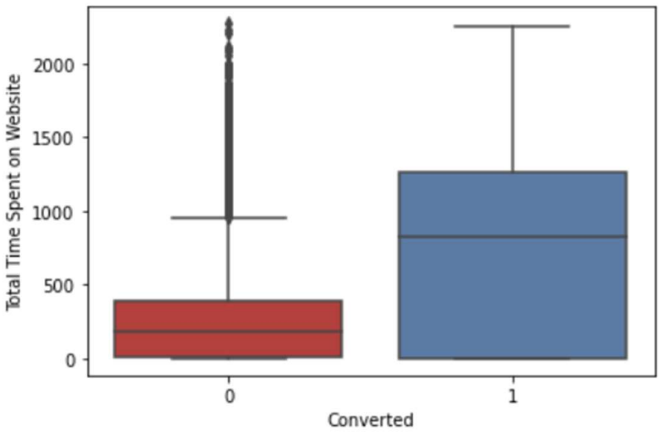
a) API and Landing Page Submission have 30-35% conversion rate but count of lead originated from them are considerable.
b) Lead Add Form has more than 90% conversion rate but count of lead are pretty low.
c) Lead Import are very less in count as well as the conversion rate, thus ca be ignored.
d) To improve overall lead conversion rate, we need to focus more on improving lead conversion of "API" and
e) "Landing Page Submission" origin and generate more leads from "Lead Add Form".

3) Lead Source Vs Converted



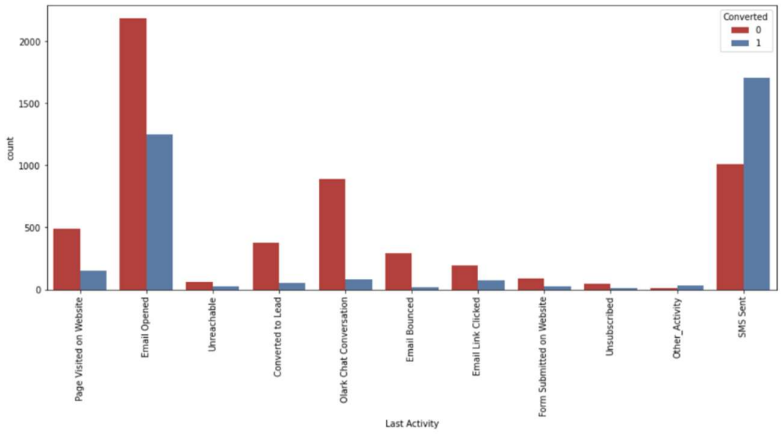
a) Google and Direct traffic generate maximum number of leads.
b) Conversion Rate of reference leads and leads through welingak website is high.
c) To improve overall lead conversion rate, focus should be on improving lead conversion of olark chat, organic search, direct traffic, and google leads and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

4) Total Time Spent on website Vs Converted



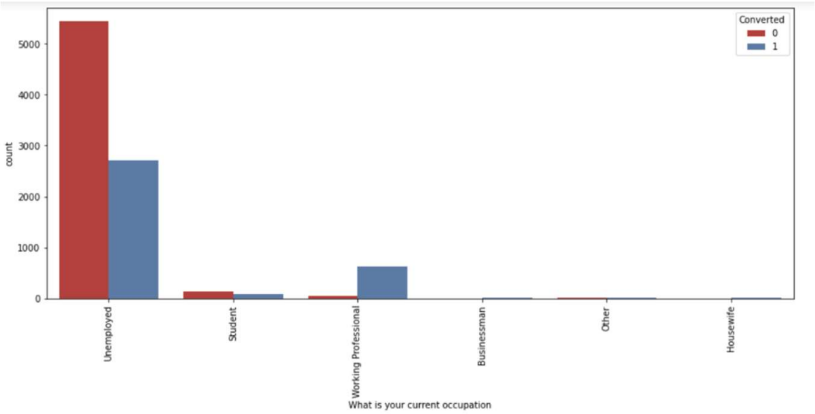
a) Leads spending more time on the website are more likely to be converted.
b) "Website should be made more engaging to make leads spend more time".

5) Last Activity Vs Converted



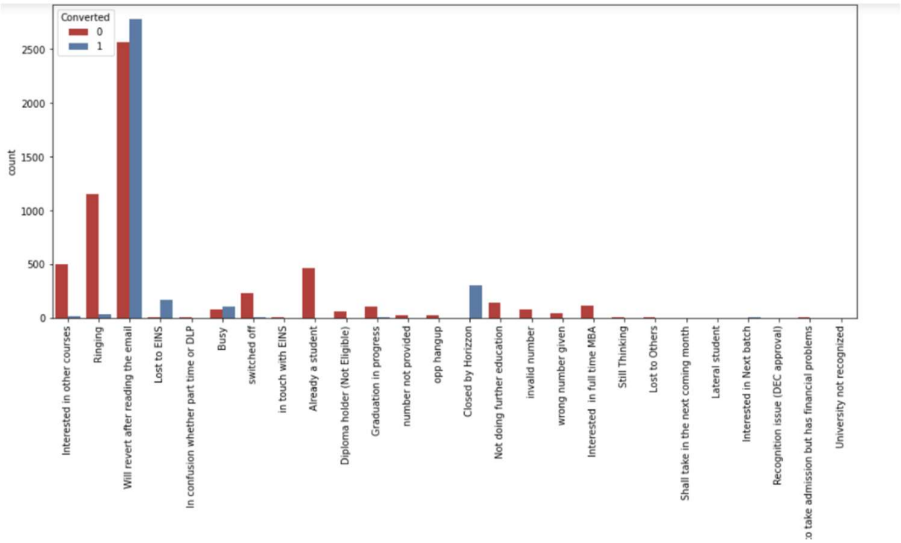
a) Most of the lead have their Email opened as their last activity.
b) Conversion rate for leads with last activity as SMS Sent is almost 60%.

6) Current Occupation Vs Converted



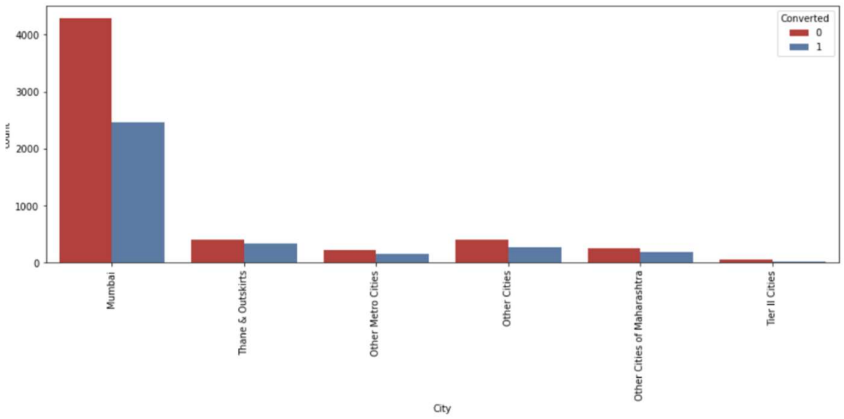
a) Working Professionals are going for the course have high chances of joining it.
b) Unemployed leads are the most in numbers but has around 30-35% conversion rate.

7) Tags Vs Converted



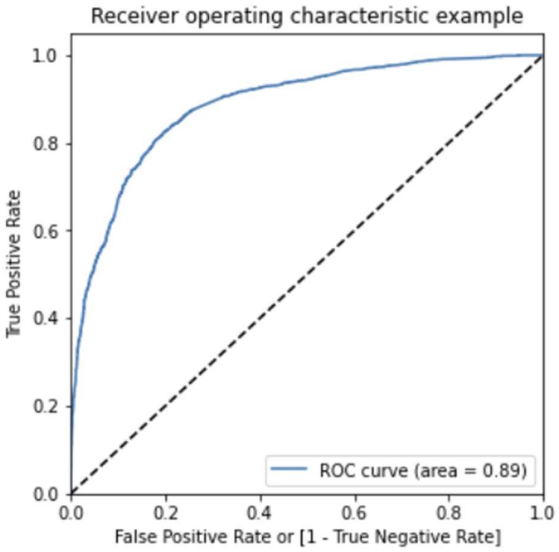
a) Since this is a column which is generated by the sales team for their analysis , so this is not available for model building .
b) So we will need to remove this column before building the model.

8) City Vs Converted



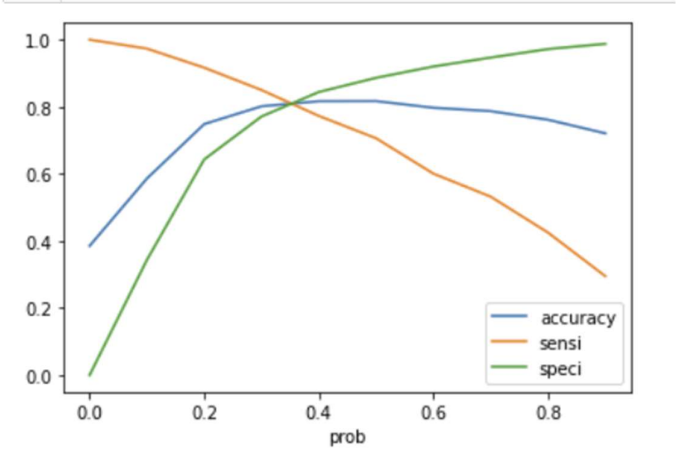
a) Most leads are from Mumbai with around 50% conversion rate.

9) ROC



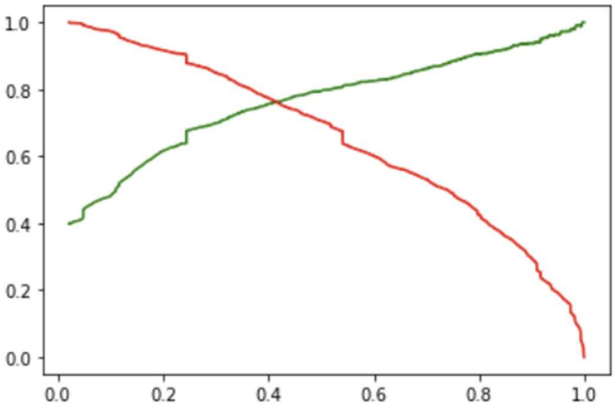
Since we have higher (0.89) area under the ROC curve, therefore our model is a good one.

10) Cut Off Probability



a) from the plot above we can observe, 0.34 is the optimum point to take it as a cutoff probability**.**

11) Precision Vs Recall



a) Precision: 0.7954
b) Recall: 0.7060

Compared the values obtained for Train & Test:

Train Data:

a) Accuracy : 81.0 %

b) Sensitivity : 81.7 %

c) Specificity : 80.6 %

Test Data:

a) Accuracy : 80.4 %

b) Sensitivity : 80.4 %

c) 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%.

#2) Finding out the leads which should be contacted

 # The customers which should be contacted are the customers whose "Lead Score" is equal to or greater than 85.

 # They can be termed as 'Hot Leads'.

 # So there are 368 leads which can be contacted and have a high chance of getting converted.

FINAL RECOMMENDATION

FINAL RECOMMENDATION ARE:

The company should make calls to the leads coming from the lead sources "Welingak Websites" and "Reference" as these are more likely to get converted.

The company should make calls to the leads who are the "working professionals" as they are more likely to get converted.

The company should make calls to the leads who spent "more time on the websites" as these are more likely to get converted.

The company should make calls to the leads coming from the lead sources "Olark Chat" as these are more likely to get converted.

The company should make calls to the leads whose last activity was SMS Sent as they are more likely to get converted.

The company should not make calls to the leads whose last activity was "Olark Chat Conversation" as they are not likely to get converted.

The company should not make calls to the leads whose lead origin is "Landing Page Submission" as they are not likely to get converted.

The company should not make calls to the leads whose Specialization was "Others" as they are not likely to get converted.

The company should not make calls to the leads who chose the option of "Do not Email" as "yes" as they are not likely to get converted.

CONCLUSION OF LOGISTIC REGRESSION MODEL:

1) The model shows high almost 80% accuracy.
2) The threshold has been selected from Accuracy, Sensitivity, Specificity measures and Precision, Recall Curve.
3) The Model shows 80% sensitivity and 80 % specificity.
4) The Model finds the correct promising leads that have high chances of getting converted.
5) The Model also specifies area where initiative can be taken to increase the chances getting converted.
6) Overall this Model proves to be accurate.