


LEAD SCORE CASE STUDY

LOGISTIC REGRESSION

Diksha Sharma
Arnav Moitra
Pratyusha Panigrahi

PROBLEM STATEMENT

- ▶ Online courses are offered by X Education for professionals in the sector. The business lists its classes on a number of well-known websites, including Google. In order to turn the most promising leads into paying clients, X Education wants to choose them.
 - ▶ The firm produces a lot of leads, but only a small percentage of them turn into paying clients. This is a lead conversion rate that is too low for the company. There are several ways to get leads, including email, website marketing, Google searches, etc.
 - ▶ The organisation approached those leads who were determined to be interested in taking the course, and as a result, had a 30% conversion rate for all leads that were turned into clients. Implementing qualities that generate leads is ineffective in increasing conversions.
- 

BUSINESS GOAL

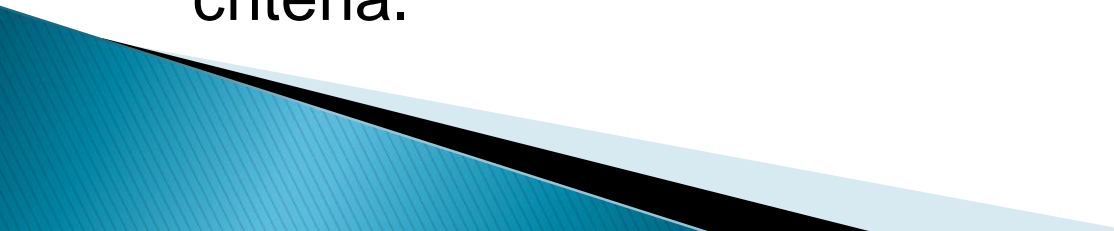
The business needs a model to be develop for choosing the most promising prospects.

Each lead should be assigned a lead score that reflects how promising they may be. The lead's likelihood of conversion increases with lead score; the lower it is, the lower its likelihood of conversion.

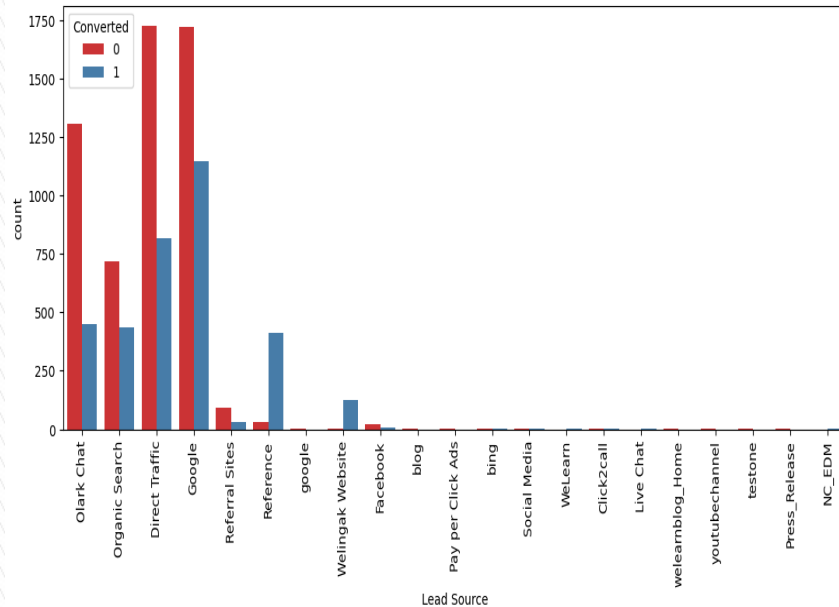
The target lead conversion rate is 78% or above.



STRATEGY

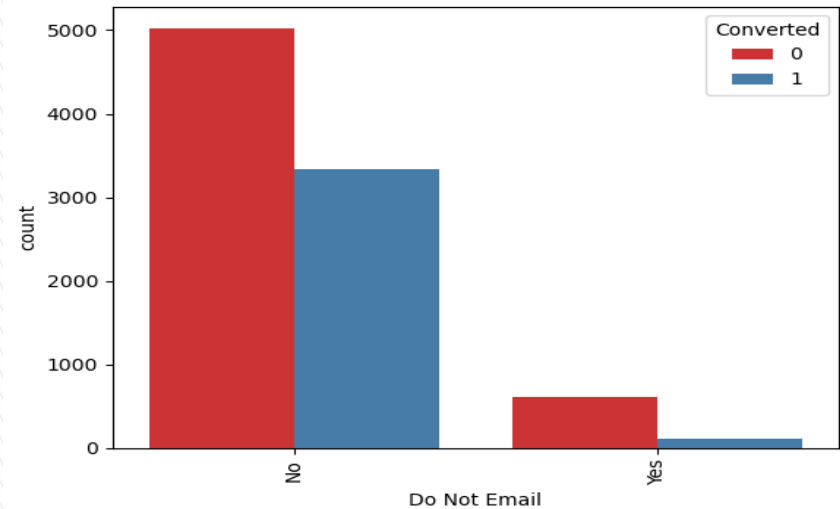
- ▶ importing data.
 - ▶ Clean up and get the data needed for more analysis.
 - ▶ Exploratory data analysis to identify the traits that are most helpful for conversion aspects of scaling.
 - ▶ Get the data ready for model building.
 - ▶ Construct a logistic regression model.
 - ▶ Each leads should be given a score.
 - ▶ On a train set, test the model.
 - ▶ Model evaluation using multiple metricsUse a test set to run the model.
 - ▶ Check the model's accuracy and other assessment criteria.
- 

EXPLORATORY DATA ANALYSIS



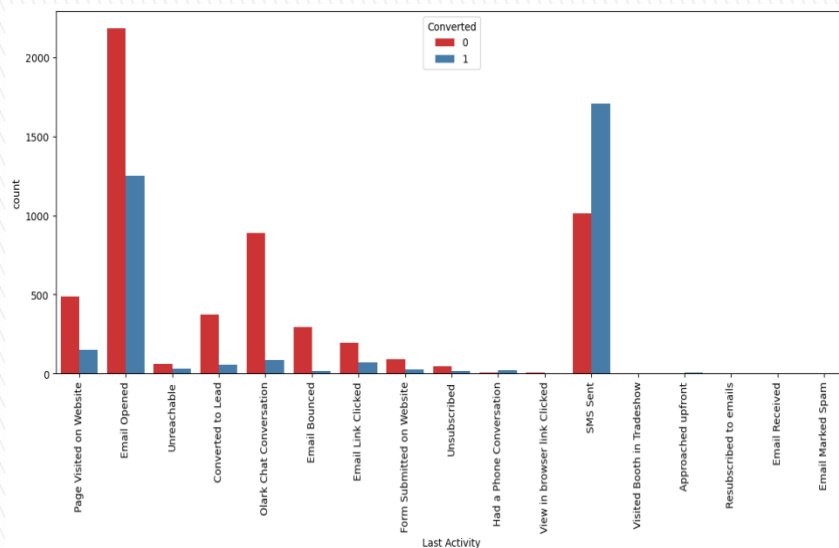
Lead Source Vs Converted

Google and Direct traffic generates maximum number of leads. Conversion Rate of reference leads and leads through welingak website is high.



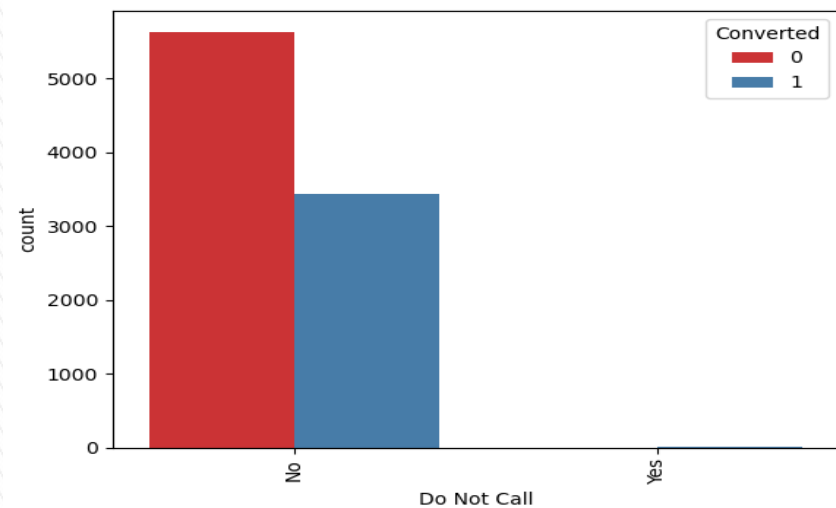
Do Not Email Vs Converted

– Most entries are 'No'. No Inference can be drawn with this parameter.



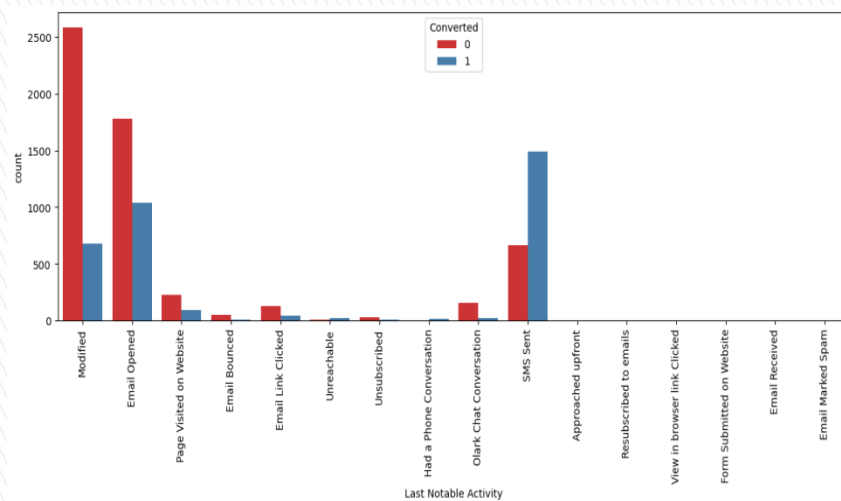
Last Activity Vs Converted

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

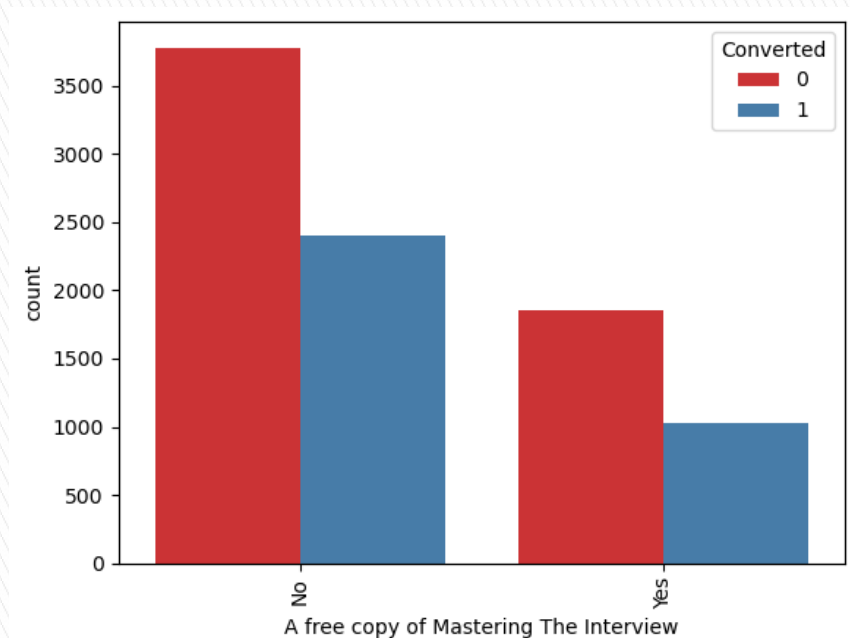


Do Not Call Vs Converted

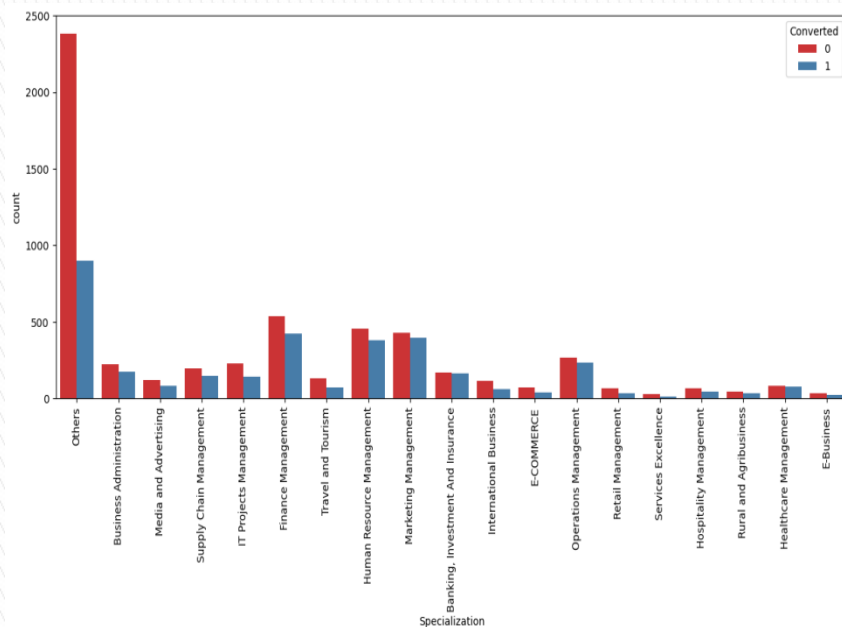
- Most entries are 'No'. No Inference can be drawn with this parameter.



Last Notable Activity Vs
Converted

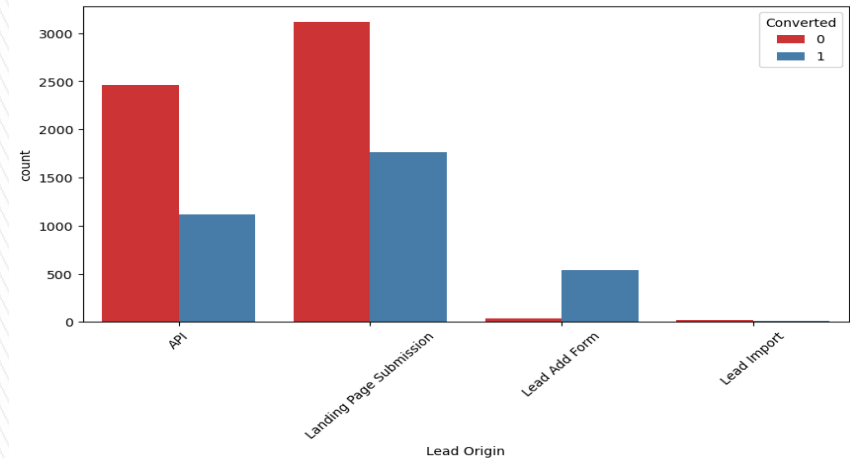


A Free Copy Of Mastering
The Interview Vs Converted



Specialization Vs Converted

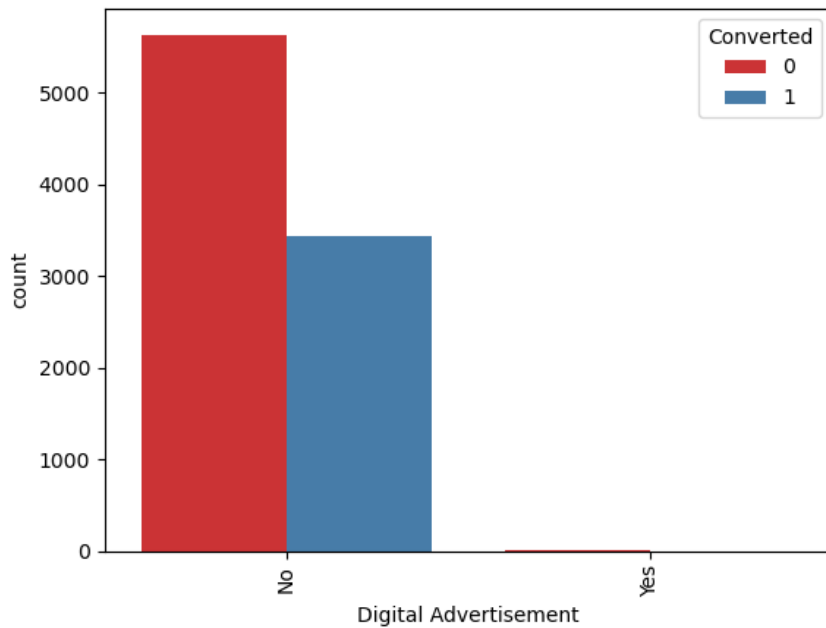
– Focus should be more on the Specialization with high conversion rate.



Lead Origin Vs Converted

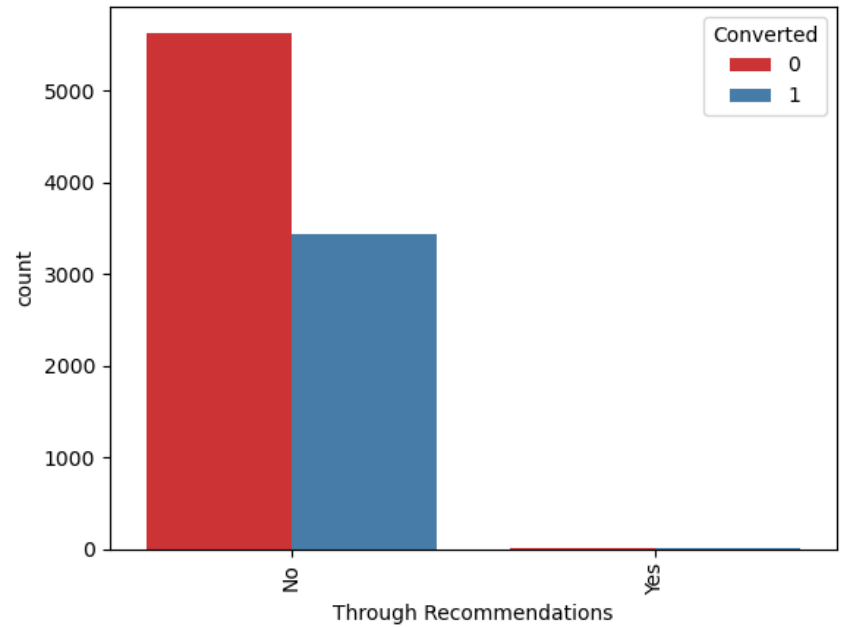
API and Landing Page Submission have 30–35% conversion rate but count of lead originated from them are considerable.

- Lead Add Form has more than 90% conversion rate but count of lead are not very high.
- Lead Import are very less in count.
- To improve overall lead conversion rate, we need to focus more on improving lead conversion of API and Landing Page Submission origin and generate more leads from Lead Add Form.



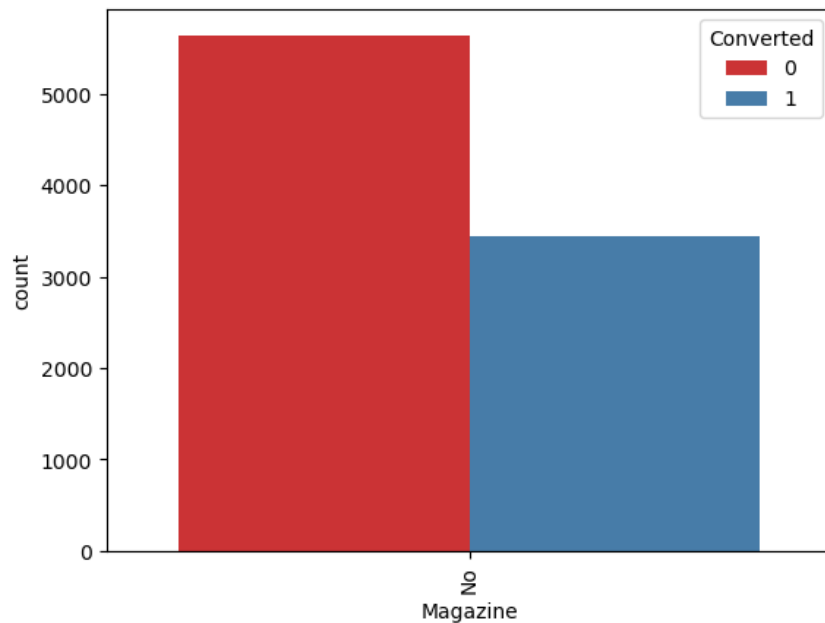
Digital Advertisement Vs Converted

– Most entries are 'No'. No Inference can be drawn with this parameter.



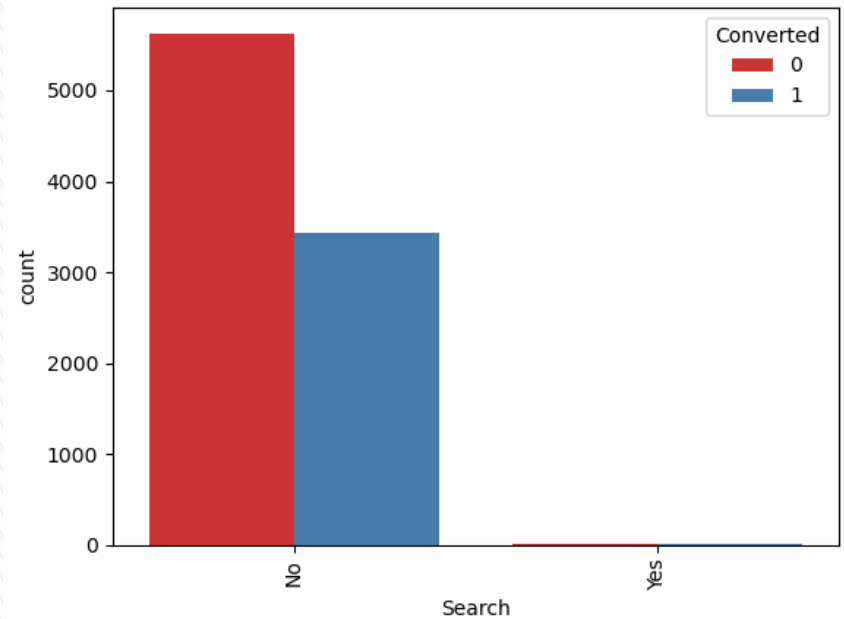
Through Recommendation Vs Converted

– Most entries are 'No'. No Inference can be drawn with this parameter.



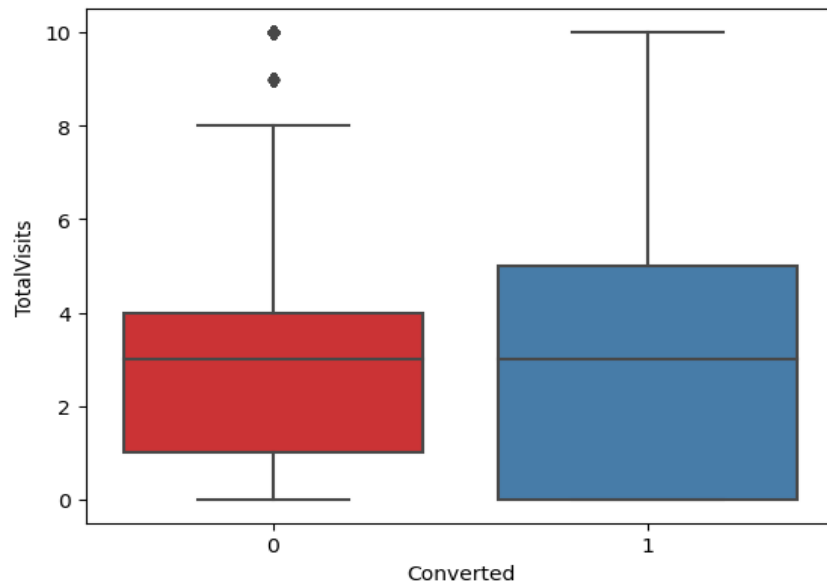
Magazine Vs Converted

- Most entries are 'No'. No Inference can be drawn with this parameter.



Search Vs Converted

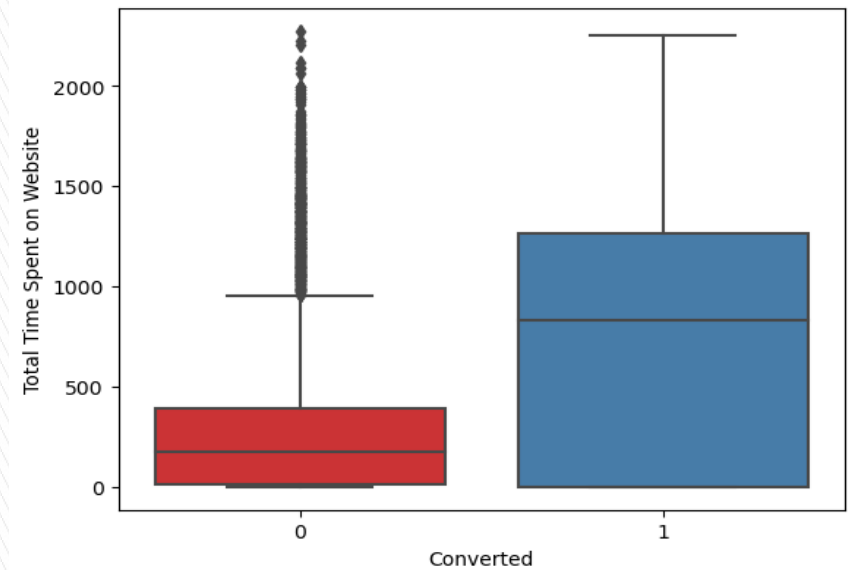
- Most entries are 'No'. No Inference can be drawn with this parameter.



Total Visits Vs Converted

Median for converted and not converted leads are the same.

– Nothing can be concluded on the basis of Total Visits.



Total Time Spent On Website Vs Converted

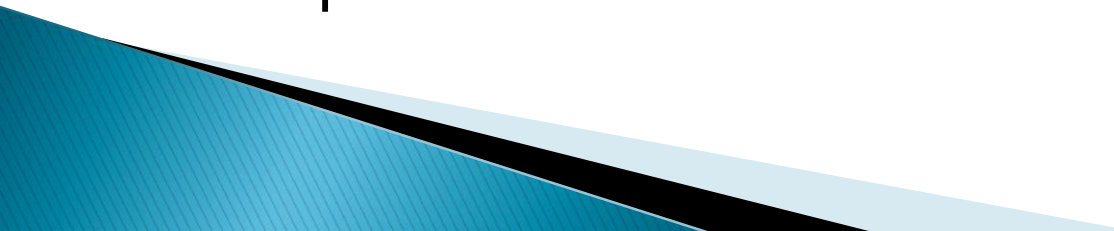
Leads spending more time on the website are more likely to be converted.

– Website should be made more engaging to make leads spend more time.

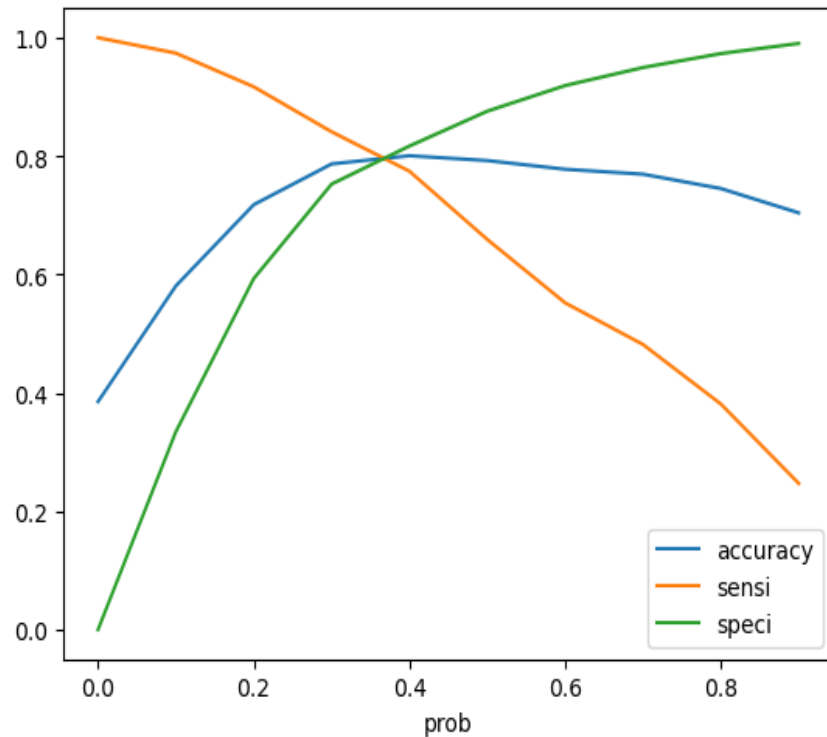
MODEL BUILDING

The following steps are taken:

dividing the data into train and test sets; scaling the variables in the train set; building the first model; using RE to remove less relevant variables; building the next model; removing variables based on high p-values; checking the value of the VIF for all the existing columns; predicting using the train set; assessing accuracy and other metrics; predicting using the test set; and performing precision and recall analysis on test predictions.

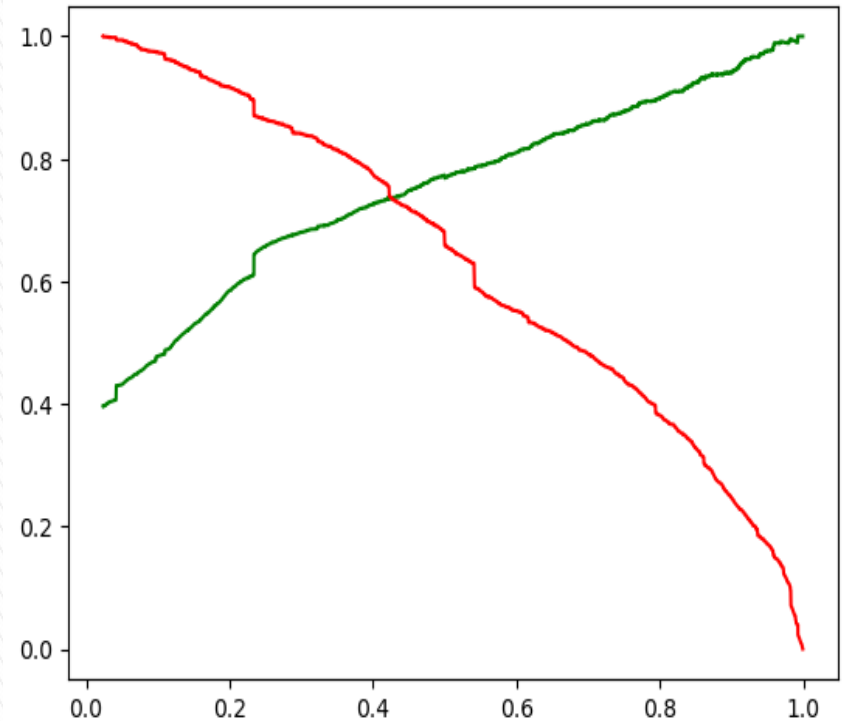


MODEL EVALUATION (TRAIN)



ACCURACY SENSITIVITY & SPECIFICITY

From the curve above, 0.34 is the optimum point to take it as a cutoff probability.



PRECISION AND RECALL

The above graph shows the trade-off between the Precision and Recall

- Making predictions on the test set
- Scaling the test data

RESULT FROM TRAIN & TEST

Comparing the values obtained for Train & Test:

- ▶ **Train Data:**

- Accuracy : 81.0 %
- Sensitivity : 81.7 %
- Specificity : 80.6 %

- ▶ **Test Data:**

- Accuracy : 80.4 %
- Sensitivity : 80.4 %
- 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%.

CONCLUSION

- ▶ 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.
- 