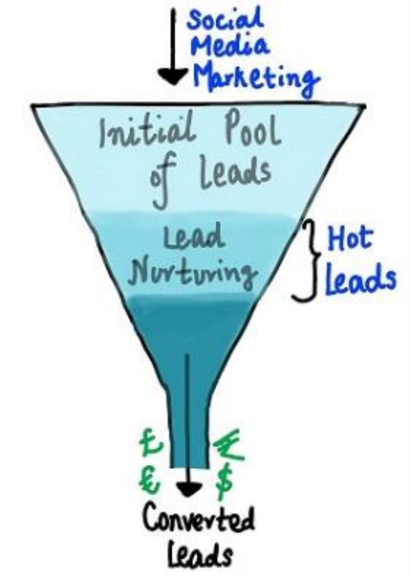


Lead Conversion Process - Demonstrated as a funnel

Lead Scoring Case Study

Presented by – Vivek Ankit and Yatin Bajaj

Business Objectives and Strategy

Problem Statement :

- X education sells the online course to industry professionals.
- The company gets leads through various sources and through past referrals.
- Suggest the company for the Hot leads.
- Suggest the sales team for the potential leads for larger conversion.

Business Objective :

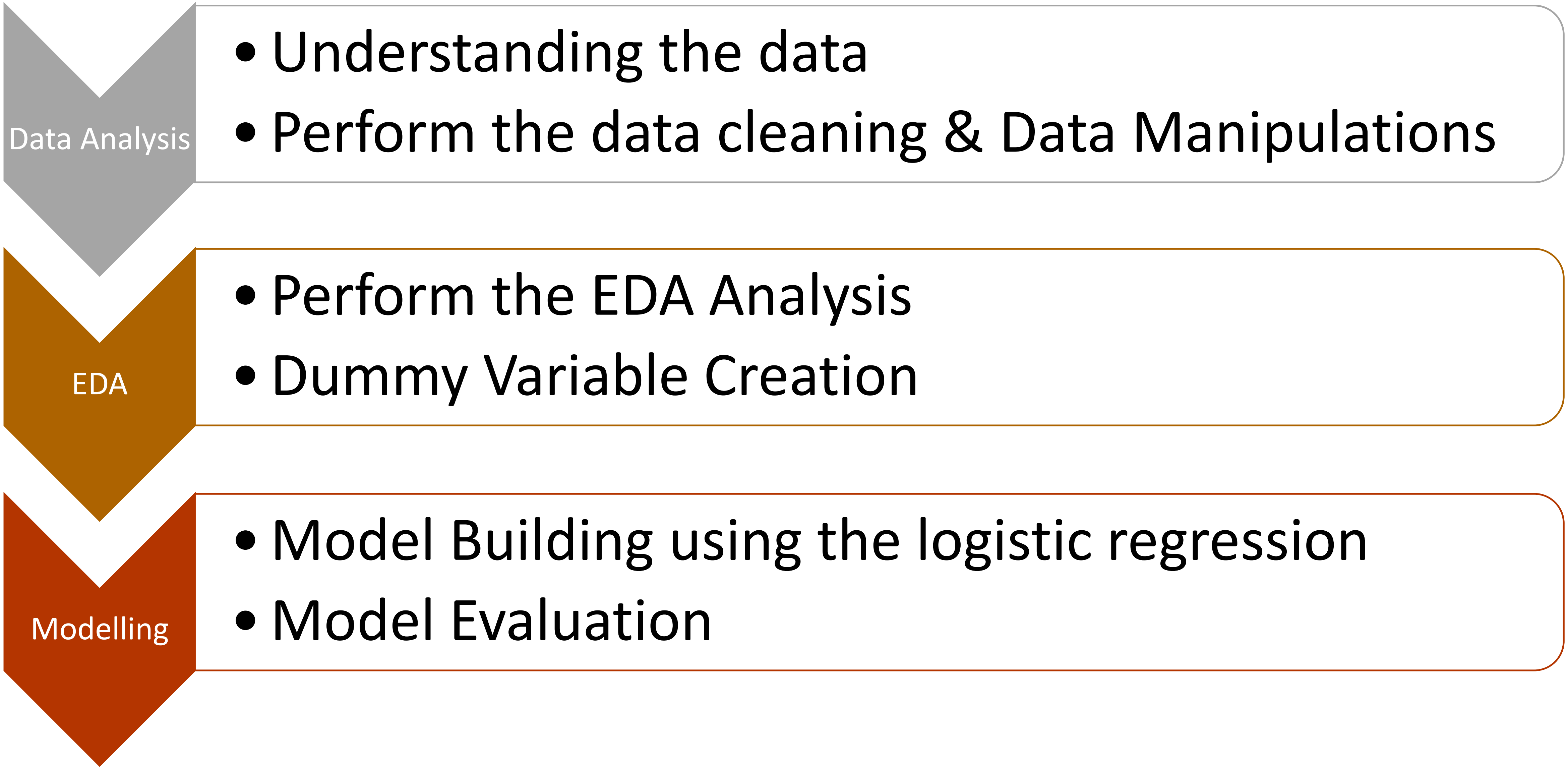
- The CEO has given us ballpark of the target lead conversion rate to be around 80%.

Strategy :

- Perform logistic regression and build a 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.

- Evaluate the model on different parameters.

Analysis Approach



Data Analysis

- Understanding the data
- Perform the data cleaning & Data Manipulations

EDA

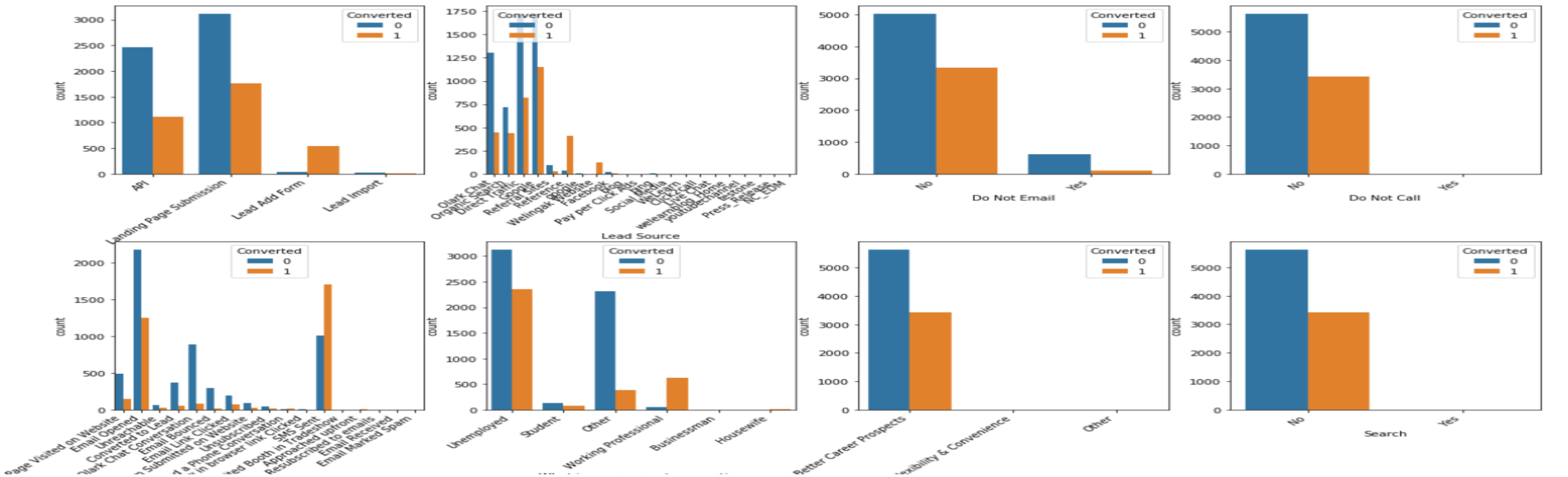
- Perform the EDA Analysis
- Dummy Variable Creation

Modelling

- Model Building using the logistic regression
- Model Evaluation

Univariate Analysis

- We performed the univariate analysis with few of variable present in the dataset.
- We found that person through referral sites , google has a good conversion ratio.
- Person sent with SMS has a higher conversion ratio.
- Unemployed person are converted larger than the others.



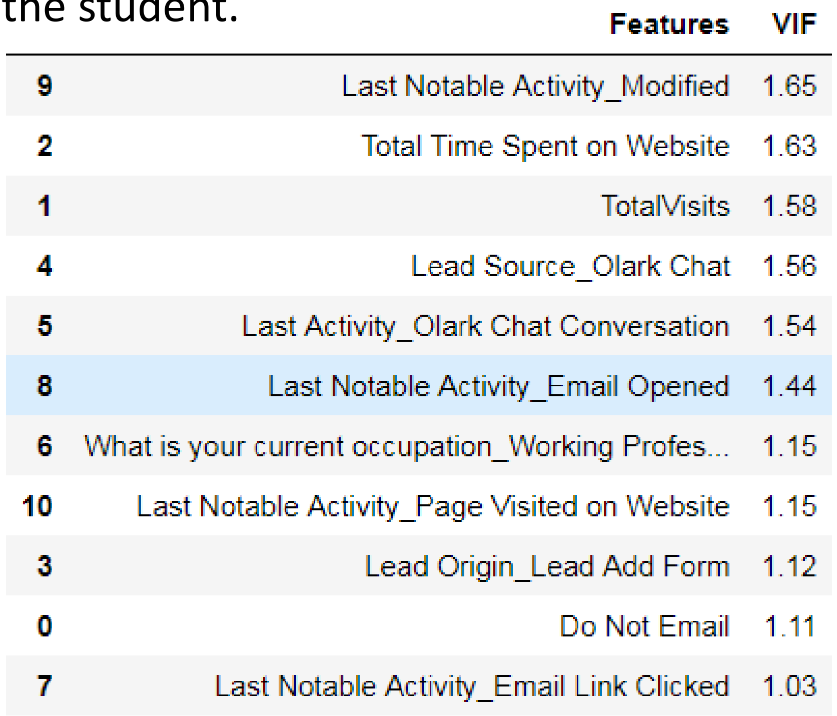
Model Building and Logistic Regression

- On the basis of our model creation, top three variables which contribute the most towards probability of lead getting converted are:

  Total Visits

  Total Time Spent on a Website

  Lead Origin
- After logistic regression we have found below features are required for the lead conversion.
- We have found that we need to focus on the below features for the lead conversion.
- Features which are hot leads are Last notable activity performed by the student.
- Total visit of the customer has the good lead score.

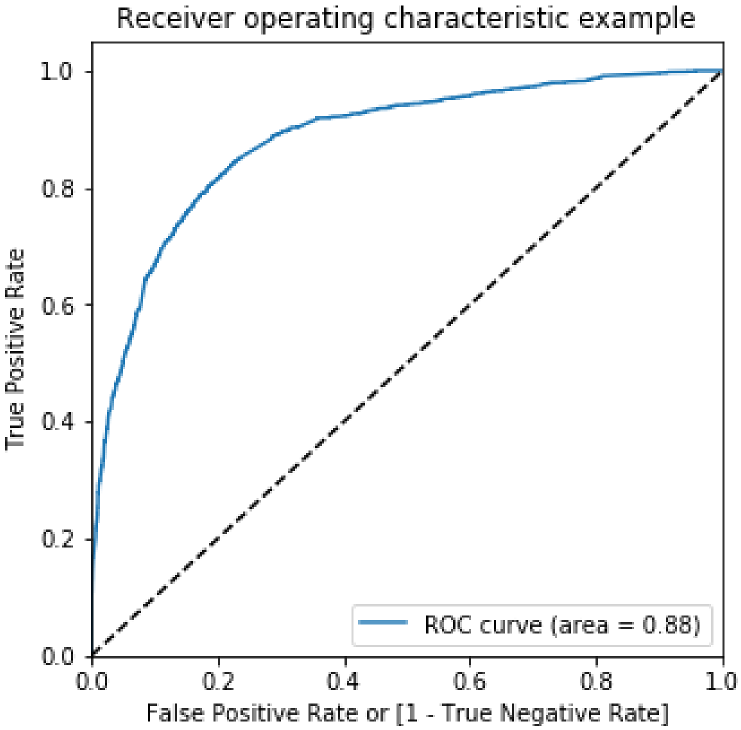


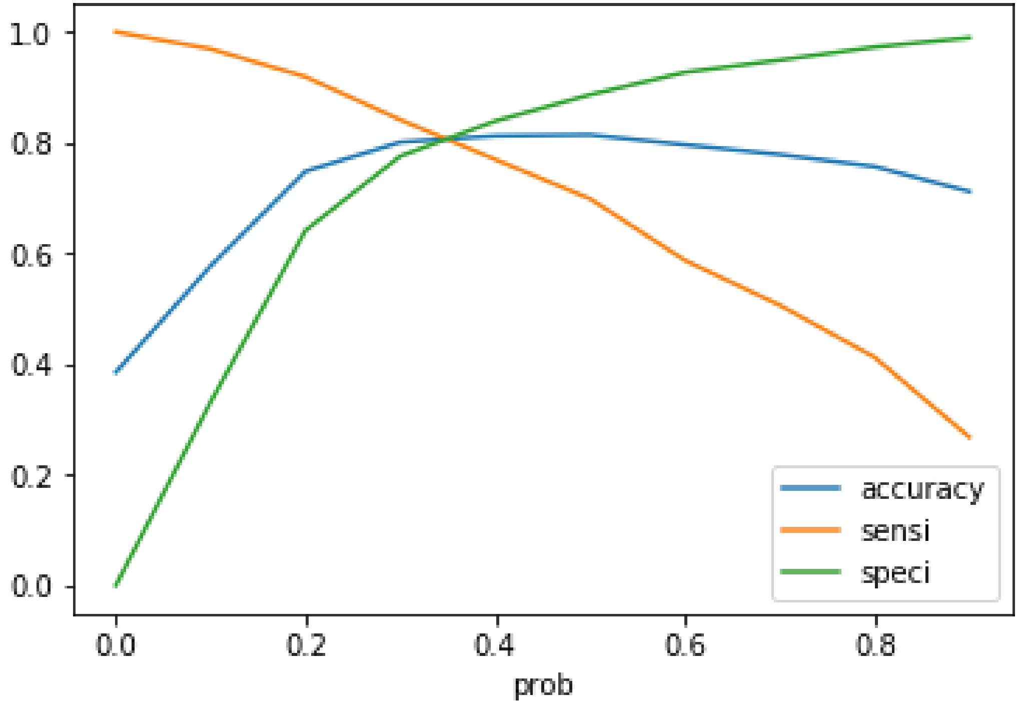
	Features	VIF
9	Last Notable Activity_Modified	1.65
2	Total Time Spent on Website	1.63
1	TotalVisits	1.58
4	Lead Source_Olark Chat	1.56
5	Last Activity_Olark Chat Conversation	1.54
8	Last Notable Activity_Email Opened	1.44
6	What is your current occupation_Working Profes...	1.15
10	Last Notable Activity_Page Visited on Website	1.15
3	Lead Origin_Lead Add Form	1.12
0	Do Not Email	1.11
7	Last Notable Activity_Email Link Clicked	1.03

Model Evaluation

ROC Curve & Identifying Optimal Point

- ROC Curve is plotted between True Positive Rate and False Positive Rate.
- It helps in understanding the overall accuracy of the model.
- Our aim is to maximize the True positive rate and minimize the false positive Rate
- Our model has area under the curve as 0.88, that means our model is quite good and stable.
- We plotted Accuracy, Sensitivity, Specificity against different probabilities.
- Optimal probability is the one at which all three curves meet.
- From our graph, we observe that 0.38 is the optimal point.





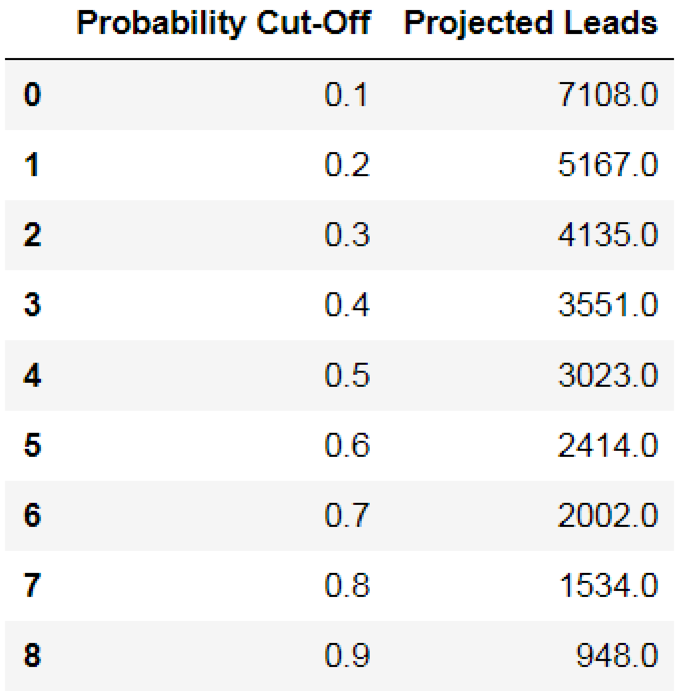
Prediction

- On the basis of our model creation, top three variables which contribute the most towards probability of lead getting converted are:
  - Total Visits
  - Total Time Spent on a Website
  - Lead Origin

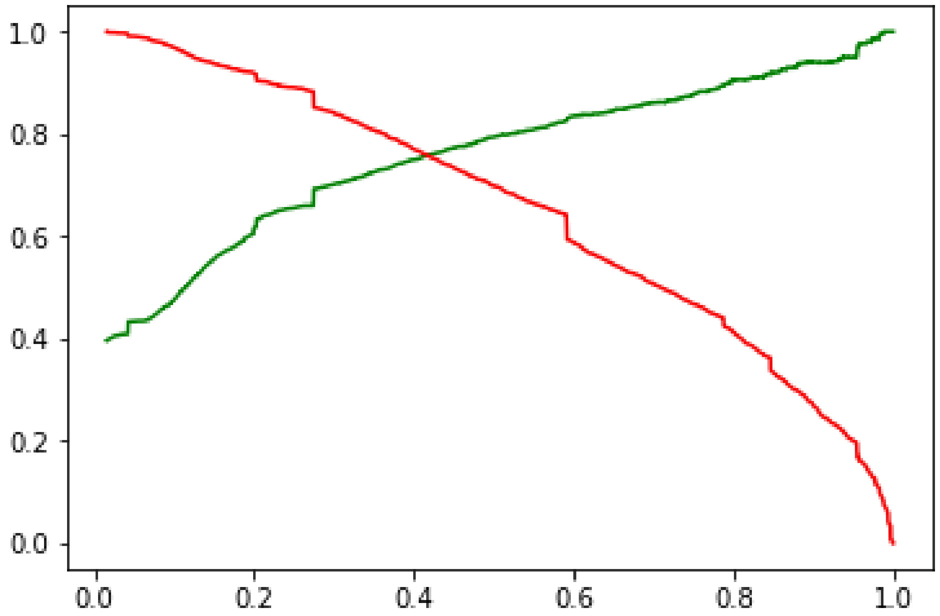
- On the basis of our model creation, top three categorical/dummy variables which contribute the most towards probability of lead getting converted are:
  - Lead Origin_Lead Add form
  - What is your current occupation_Working Professional
  - Lead Source_Olark Chat

Probability Cut off vs Projected Leads

- As per our model creation we observe that the total number of the projected leads is inversely propositional to probability cut off score.
- We can see from the graph that probability cut off for 0.1 is 7108
- Similarly for 0.2 is 5167 and so on
- **Precision** can be seen as a measure of exactness or quality is 73%.
- **Recall** is a measure of completeness or quantity or high **recall** means that an algorithm returned most of the relevant results – Our model has the recall value as 76%.



	Probability Cut-Off	Projected Leads
0	0.1	7108.0
1	0.2	5167.0
2	0.3	4135.0
3	0.4	3551.0
4	0.5	3023.0
5	0.6	2414.0
6	0.7	2002.0
7	0.8	1534.0
8	0.9	948.0



Recommendations for Requirements change

- **Strategy to use interns to increase leads conversion rate aggressively-**

- If we want almost all the potential leads to be converted, we can DECREASE the cut-off value from 0.4 to 0.2
- Currently, with 0.4 cut-off, the number of potential leads is 3023.
- If we lower down the cut-off value to 0.2, the number of potential leads will be 5167.
- That means an increase in 2144 leads, which is quite a good number. Hence, we can make phone calls to as many of such people as possible.

- **Strategy to use phone calls when the target is reached-**

- if we want reduce the number of potential leads to be converted, we should INCREASE the cut-off value from 0.4 to 0.8
- Currently, with 0.4 cut-off, the number of potential leads is 3023.
- If we lower down the cut-off value to 0.8, the number of potential leads will be 1534.
- That means a decrease in 1489 leads, which is quite a good number. Hence, we can avoid making phone calls unless it’s extremely necessary.

Data Science IIITB

Vivek Ankit & Yatin Bajaj