# Lead Scoring Case Study Summary

## Problem Statement:

X Education sells online courses to industry professionals. X Education needs help in selecting the most promising leads, i.e., the leads that are most likely to convert into paying customers.

The company needs a model wherein you a lead score is assigned 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%.

Solution Summary:

Step1: Reading and Understanding Data:

Read and inspected the data.

step2: Data Cleaning:

a. First step to clean the dataset we chose was to drop the variables having unique values.

b. Then, there were few columns with value 'Select' which means the leads did not choose any given option. We changed those values to Null values.

c. We dropped the columns having NULL values greater than 35%.

d. Next, we removed the imbalanced and redundant variables. This step also included imputing the missing values as and where required with median values in case of numerical variables and creation of new classification variables in case of categorical variables. The outliers were identified and removed. Also, in one column was having identical label in different cases (first letter small and capital respectively). We fixed this issue by converting the label with first letter in small case to upper case.

e. All sales team generated variables were removed to avoid any ambiguity in final solution.

Step3: Data Transformation:

Changed the binary variables into '0' and '1'

Step4: Dummy Variables Creation:

a. We created dummy variables for the categorical variables.

b. Removed all the repeated and redundant variables

Step5:

Test Train Split: The next step was to divide the data set into test and train sections with a proportion of 70- 30% values.

Step6: Feature Rescaling:

a. We used the Min Max Scaling to scale the original numerical variables.

b. The, we plot the a heatmap to check the correlations among the variables.

c. Dropped the highly correlated dummy variables.

Step 8: Conclusion:

• The lead score calculated in the test set of data shows the conversion rate of 83% on the final predicted model which clearly meets the expectation of CEO has given a ballpark of the target lead conversion rate to be around 80%.

• Good value of sensitivity of our model will help to select the most promising leads.

• Features which contribute more towards the probability of a lead getting converted are:

i. Lead Origin_Lead Add Form
ii. What is your current occupation_Working Professional
iii. Total Time Spent on Website.