# Summary – Lead Scoring Case Study

The case study is for X Education. X Education provides the courses for students and industry professionals. The objective of the case study is to better the cold calling/sales approach towards the customers, in order to increase the number of customers that purchase the courses.

In order to achieve this objective, we took the dataset provided which gives us insights about the customer (such as time spent on website, last notable activity, etc), and did a series of analysis to make a model that can predict the convertible leads. The steps done in the analysis and model building are explained below.

1. **Data Reading:** Reading the data and understanding the columns and values.

2. **Cleaning Data:** The data provided was partly cleaned. Although some of the data in the columns were required to be cleaned and organized. The 'select' values in the column were changed to NULL as 'select' did not specify anything about the column. The country column was changed to 'India', 'Outside India' and 'Not Provided'. Some of the columns having low percentage of NULL values, the null values were imputed as 'Not Provided'.

3. **Exploratory Data Analysis:** EDA was performed on the data to analyse the data. We found that many of the categories in categorical columns are not significant for the model. No outliers were found in the data.

4. **Data Preparation:** We created the dummy variables and removed the dummy variables with the syntax: 'ColumnName_NotProvided', as 'not provided' categories were not needed.

5. **Train-Test Split:** Train Test split was done on the data, with, 70% train and 30% test data.

6. **Building Model:** The model was built and feature selection was done by RFE. Top 15 columns were taken as output. The columns were drooped on the basis of VIF(VIF>5) and p-values (p-value>0.05.

7. **Model Evaluation:** We made the confusion matrix, and found out the optimal cut-off value using the ROC curve. The accuracy, sensitivity, specificity were found out.

8. **Making Prediction:** Prediction was made on the test data, with the optimal cut-off value of 0.35 and accuracy, sensitivity, specificity of approx. 80%.

9. **Precision-Recall:** With the current cut-off value of 0.35, Precision was found to be around 78% and Recall was found to be around 70%.

After the model is build, following columns were found to increase the probability of lead conversion (in descending order):

1. What is your current occupation_housewife (from What is your current occupation)
2. Last Activity_email marked spam (from Last Activity)
3. Last Activity_email received (from Last Activity)
4. Lead Source_welingak_website (from Lead Source)
5. Total Time spent on the website.
6. Lead Source_reference (from Lead Source)

Using these factors in the columns, X Education can increase the probability of lead conversion.