# Summary of Lead Scoring Assignment

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend there, how they reach the site and the conversion rate.

## The following are the steps used:

### 1. Cleaning data:
The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. A few of the null values were changed to 'not provided' to not lose much data. Although they were later removed while making dummies.

### 2. EDA:
A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant.

### 3. Dummy Variables:
The dummy variables were created and later on the dummies with 'not provided' elements were removed. For numeric values, we used the MinMaxScaler.

### 4. Train-Test split:
The split was done at 70% and 30% for train and test data respectively.

### 5. Model Building:
Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with VIF<5 and p-value <0.05 were kept).

### 6. Model Evaluation:
A confusion matrix was made. Later on, the optimum cut-off value (using the ROC curve) was used to find the accuracy (91.58%); sensitivity (85.30%) and specificity (95.31%).

### 7. Prediction:
Prediction was done on the test data frame and with an optimum cut-off with accuracy(91.65%); sensitivity (86.61%) and specificity (94.35%).

### 8. Precision - Recall:
This method was also used to recheck a cut-off. Precision was around 89.14% and recall was around 86.60% on the test data frame.

**Variables that matters the most to the potential buyers are:**

1. SMS sent
2. Email
3. The Total Time Spent on the Website
4. Already a student
5. Lead source
6. Interested in full-time MBA
7. Interested in other courses
8. Lost to EINS
9. Ringing
10. Will revert after reading the email
11. Tags_invalid number
12. Tags_opp hangup