# Summary

This analysis is done for X Education and to find ways to get more industry professionals to join their courses. The data provides information about the potential customers , the time they spend on website, their source of information and the conversion rate.

The following are the steps used:

1. Cleaning data:
   The data had few null values and the option select had to be replaced with a null value since it did not give us much information. Data inspection was performed.
2. EDA: EDA was done to check the condition of data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values had outliers and outliers were treated.
3. Dummy Variables:
   The dummy variables were created.
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 relevant variables Then, rest of the variables were checked with respect to 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 ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 80% each.
7. Prediction:
   Prediction was done on the test data frame and with an optimum cut off as 0.30 with sensitivity of 80%.
8. Precision – Recall:
   This method was also used to recheck and a cut off of 0.40 was found with Sensitivity 81% Precision 68% and recall 81% on the test data frame.

It was observed that the variables that can improve the conversion rate are in top down priority are as follows:

1. Lead Origin
    - Lead_ Add Form
2. What is your current occupation
    - Working Professional
3. Last Activity
    - SMS Sent
4. Total time spent on website
5. Last Activity
    - Others
6. Lead Source
    - Olark Chart
7. Last Notable Activity
    - Other_Notable Activity