

## Summary Report

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 reached the site and conversion rate.

The following steps are used:

1. **Cleaning data:** The data was partially clean except for few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to not provided so as to not lose much.
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. The numeric values seem good and no outlier was found.
3. **Dummy Variable:** The dummy variables are 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:** Finally 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-values.
6. **Model Evaluation:** A confusion matrix was made Later on the optimum cut off value was used to find the accuracy.
7. **Prediction:** Prediction was done on the test data frame and with an optimum cutoff.
8. **Precision-Recall:** This method was also used to recheck and a cut off of 0.41 was found the precision on the test data frame.

It was found out that the variables that mattered the most in the potential buyers are keeping the above mentioned points in mind the X Education can increase all the potential buyers to change their mind and buy their courses.