

## [SUMMARY]

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 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.
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 outliers were found.
3. Dummy Variables: The dummy variables were created and later on the dummies with null elements were removed.
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 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  $p\text{-value} < 0.05$  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 92.9%, 91.8% and 93.6% respectively.
7. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.4 with accuracy, sensitivity and specificity of 92.9%, 91.8% and 93.6% respectively.

8. Precision – Recall: This method was also used to recheck and a cut off of 0.48 was found with Precision around 89% and recall around 91% on the test data frame.

**It was found that the variables that mattered the most in the potential buyers are (In descending order):**

1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
  - a. Google
  - b. Direct traffic
  - c. Organic search
  - d. Welingak Website
4. When the last activity was:
  - a. SMS
  - b. Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.