

## **SUMMARY**

This analysis is 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.

### **DATA CLEANING:**

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. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies.

### **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.

### **TRAIN\_TEST\_SPLIT:**

The split was done at 70% and 30% for train and test data respectively.

### **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.

### **MODEL EVALUTION:**

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.

### **PREDICTION:**

Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 70% more.

### **PRECISION – RECALL:**

This method was also used to recheck and a cut off of 0.41 was found with Precision around 73% and recall around 75% on the test data frame.