SUMMARY

This analysis was conducted for X Education to find out how to get more industry professionals to enroll in courses. From the basic data provided, a lot of information about how prospects visited the website, how long they stayed, how they (LEADS] got to the website and what conversion rate was obtained at the end.

The following are the steps used:

1. **Cleaning data:**
   -> The data was partially cleared except for a few null values and the option select needed 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.
   -> Since there were many from India and few from outside, the elements were changed to 'India', 'Outside India' and 'not provided'. To get a good overview of the analysis.

2. **EDA [EXPLORATORY DATA ANALYSIS]:**
   -> A simple EDA was run to check the sanity of the data. Many of his items on the categorical variables were found to be irrelevant. The number looked good at and no outliers were found.

3. **Dummy Variables:**
   -> A dummy variable was created and later the dummy with the "not provided" element was removed. For numbers, we used Min Max Scaler.

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 18 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 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 0.3with accuracy, sensitivity and specificity of 91%.

8. **Precision – Recall:**
   -> Precision recall came out to be 88 and 91 %.

9. **Final Observation Towards the X EDUCATION CASE STUDY:-**
   Training Data:
   Accuracy : 92.14%
   Sensitivity : 91.49%
   Specificity : 92.54%

   Test Data:
   Accuracy : 92.57%
   Sensitivity : 91.18%
   Specificity : 93.46%

   The model seems to be performing well. This can be recommended for further business processing.