SUMMARY

This analysis is carried out for X Education to find the ways to get or attract more business professionals to join their courses. The basic data provides us a lot of information about how the potential customers visit the site, the time they spent on the website, how they reached to the site, are they willing to open the emails and their conversion rate.

The following are the steps used to provide the final insights:

1. **Cleaning the data:**
   The data was partially clean except for a few null values. However, the option chooses had to be changed to a null value because it provided little useful information and the option select had to be replaced with the null values. Since, it didn't give us much information.
   Merge the data field of few columns to "Others" for which having the very few counts as compared to another fields.
   To avoid losing too much data, only a small number of the null values were changed to "Not Provided". Although, they were later removed while making dummies. Since, there were many from India and very few from outside as compared to India, the elements were changed to "India", "Outside India" and "Not Provided"

2. **EDA (Exploratory Data Analysis):**
   A quick EDA was done to check the condition of our data. It was discovered that lots of elements in the categorical variables were irrelevant. The numerical values seem to be good, no anomalies were discovered, and no outliers were found. If there was any then handled the outliers.

3. **Creating Dummy Variables:**
   The dummy variables were created and later the dummies with "Not Provided" elements were taken away. For numeric values, used the Min-Max Scaler to scale numerical numbers

4. **Train – Test Split dataset:**

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

5. **Model Building:**
   First, RFE was done to attain the top 20 potential variables. Later the rest of the variables were removed manually depending on the VIF values and p-value also (The variables with VIF<5 and p-value <0.05 were retained)

6. **Model Evaluation:**
   A confusion matrix was created. Later, the optimum cut-off was determined using the ROC curve as well as 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 0.34 with accuracy, sensitivity and specificity of 80% each.

8. **Precision & Recall:**
   This method was used to recheck and cut off of 0.34 was found with Precision around 79.5% and recall around 70.2% on the test data frame

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

1. What matters most to you in choosing a course
2. Tags will revert after reading the email
3. Last Notable Activity Modified
4. Tags
   - Tags other tags
   - Tags Ringing
5. Last Activity
   - Last Activity SMS sent
   - Last Activity Olark Chat conversation
6. Occupation Working professional
7. Lead Origin Lead Import

Hence, it’s evident that our model worked well.

The conversion rate before model building was 38.53% while after model building it went up to 80%. We were successful in target lead conversion rate to be even higher than 80% as demanded by the CEO of X education.

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.