Summary

This analysis aims to attract more industry experts to X Education's courses. The basic statistics provided us with a wealth of information about how potential clients visit the site, how long they stay, how they got there, and the conversion rate.

The following are the steps used:

1. **Cleaning data:**
   The data was partially clean, with the exception of a few null values, and the option select had to be replaced with a null value because it provided insufficient information. A few null values were converted to 'not provided' to avoid losing too much data. However, they were eventually eliminated while creating dummies. Because there were many Indians and few outsiders, the elements were altered to 'India', 'Outside India', and 'not provided'.

2. **EDA:**
   A brief EDA was performed to determine the state of our data. It was discovered that many of the items in the categorical variables were irrelevant. The numerical results appear to be accurate, with no outliers discovered.

3. **Dummy Variables:**
   The dummy variables were established, and then the ones with 'not given' elements were removed. For numerical 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:**
   First, RFE was used to identify the top 15 relevant factors. The remaining variables were manually deleted based on their VIF values and p-value (only variables with VIF < 5 and p-value < 0.05 were retained).

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 0.35 with accuracy, sensitivity, and specificity of 80%.

8. **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.

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

1. The total time spent 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, X Education may thrive since they have a very good probability of convincing practically all potential buyers to alter their minds and acquire their courses.