Summary

This analysis is carried out for X Education in an effort to attract more business professionals to their courses. We learned a lot from the fundamental data on how potential customers use the site, how long they stay there, how they got there, and the conversion rate. The procedures are as follows:

1. Data Cleaning: Except for a few null values, the data was mostly clean. However, the option choose had to be changed to a null value because it provided little useful information. To avoid losing too much data, a few of the null values were changed to "not mentioned". Nevertheless, they were later taken out while manufacturing dummies.

2. Exploratory Data Analysis: To quickly assess the state of our data, an EDA was performed. It was discovered that several of the categorical variables' components were unnecessary. The numerical figures are accurate, and no anomalies were discovered.

3. Dummy Variables: After the dummy variables were made, those that had the phrase "not mentioned" were later deleted. We utilised the MinMaxScaler to scale numerical numbers.

4. Train-Test split: For train and test data, the split was done at 70% and 30%, respectively.

5. Model Building: First, the top 15 pertinent factors were determined by RFE. Later, based on the VIF values and p-value, the remaining variables were manually deleted (the variables with VIF < 5 and p-value < 0.05 were retained).

6. Model Evaluation: A confusion matrix was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 90% each.

7. Prediction: On the test data frame, predictions were made using an optimal cutoff of 0.3, with 90% accuracy, sensitivity, and specificity.

8. Precision – Recall: This method was also utilised to perform a second check, and on the test data frame, a cut off of 0.3 was discovered with precision around 88% and recall around 92%.

The factors that affected potential purchasers the most were discovered to be (in descending order):
1. The total time spend on the Website.
2. Total number of visits.
3. When the lead source was:
   - Google
   - Direct traffic
   - Organic search
   - Welingak website
4. When the last activity was:
   - SMS
   - Olark chat conversation
5. When the lead origin is Lead add format.
6. When their current occupation is as a working professional.

With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.