**Summary**

The purpose of this analysis is to help X Education attract more industry professionals to their courses. We learned a lot about the potential consumers' visitation patterns, length of stay, method of access, and conversion rate from the basic data provided.

The procedures employed are as follows:

**l. Data cleaning:**   
With the exception of a few null values, the data was mostly clean. Because the option select provided little information, it had to be changed to a null value. To save on data loss, a small number of the null values were changed to "not provided." Nevertheless, they were later taken out while creating dummies. The elements were altered to "India," "Outside India," and "not provided" because there were more people from India than from outside.

**2. EDA:**

A brief EDA was conducted to assess the state of our data. Numerous components in the category variables were discovered to be meaningless. The numbers appear to be in range, and there are no anomalies in the data.

**3. Dummy Variables:**

After the dummies with the "not provided" elements were formed, they were eliminated. The MinMaxScaler was utilized for numerical values.

**4. Train-Test split:**

The split percentages for the train and test data were 70% and 30%, respectively.

**5. Building the Model:**

To begin with, RFE was used to identify the top 15 pertinent variables. Afterwards, the remaining variables were eliminated by hand based on the p-value and VIF values (the variables with a p-value of less than 0.05 and a VIF of less than five were retained).

**6. Model Evaluation:**

A matrix of confusion was created. Later, the accuracy, sensitivity, and specificity were determined by utilizing the ROC curve to determine the optimal cut off value; these values were approximately 80% for each.

**7. Forecast:**

Using the test data frame, a prediction was made with an optimal cutoff of 0.35 and 80% accuracy, sensitivity, and specificity.

**8. Precision — Recall:**

This technique was also employed for a recheck, and on the test data frame, a cut off of 0.41 was discovered with precision at roughly 73% and recall at roughly 75%.