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.

Steps followed to do the analysis:

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, only a small number of the null values were changed to 'not provided'. Nevertheless, they were later taken out while manufacturing dummies. The elements were altered to "India," "Outside India," and "not provided" because there were a lot of people from India and a small number from elsewhere.

2. EDA

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. Creating Dummy Variables

The fake variables were made, then later the fakes with the 'not provided' bits were taken away. We utilised the MinMaxScaler to scale numerical numbers.

4. Training & Testing and splitting the data set

Splitting the data set into 70:30 for training and testing.

5. Building the Model

First, the top 20 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. Evaluating the Model

A matrix of confusion was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 80% each.

7. Predictions made

Summary

On the test data frame, prediction was performed with an optimal cut off of 0.34 and 80% accuracy, sensitivity, and specificity.