

# **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. Cleaning data:**

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 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 results seem accurate, and no outliers were discovered.

## **3. Dummy Variables:**

After the dummy variables were made, those that had the phrase "not provided" 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 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. Prediction:**

On the test data frame, predictions were made using an optimal cut-off of 0.35 and had 80% accuracy, sensitivity, and specificity

## **8. Precision – Recall:**

On the test data frame, this method was also used to recheck, and a cut off, of 0.41 was discovered with precision and recall averaging around 73% and 75%, respectively.

According to research, the following factors affected potential purchasers the most (in descending order):

- 1) The length of time spent on the website overall.
- 2) The overall volume of visits.
- 3) If the lead came from one of the following:
  - A) Google
  - B) Direct traffic
  - C) Organic search
  - D) The Welingak website
- 4) The previous activity was:
  - A) an Olark chat session
  - B) an SMS.
- 5) If the format of the lead add is the lead origin.
- 6) If they are a working professional at the time.

With these in mind, X Education can succeed since they have a very good probability of convincing practically all prospective students to change their minds and enlist in their courses.