

## Summary Report

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 following are the steps:

### 1. Cleaning the data:

The majority of the data was clean, with the exception of a few null values. The option choose had to be changed to a null value because it didn't offer much detail. Only a tiny portion of the null values were changed to 'not provided' in order to prevent losing too much information. However, they were eventually removed while creating dummies. The words "India," "Outside India," and "not provided" were substituted for the original words because the majority of attendees were Indian while only a tiny percentage were from abroad.

### 2. EDA:

To assess the state of our data, a brief EDA was conducted. 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: Dummy variables were made, and those that had 'not provided' items were later eliminated. We utilised the MinMaxScaler to scale numerical numbers.

4. The split for train and test data was done at 70% and 30%, respectively.

### 5. Model Construction:

RFE was first used to identify the top 15 pertinent variables. 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. Confusion matrix for the model 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: Using an optimal cutoff of 0.35 and a prediction accuracy, sensitivity, and specificity of 80%, the test data frame was used.

#### 8. Precision - Recall:

This method was also utilised to retest, and on the test data frame, a cut off of 0.41 was discovered with Precision around 73% and recall around 75%.

The factors that affected potential purchasers the most were discovered to be (in descending order):

1. The overall amount of time spent on the Website.
2. The overall volume of visits.
3. If the lead came from one of the following: 1. Google 2. Direct traffic 3. Organic search 4. The Welingak website
4. The most recent activity was when: SMS exchanges and conversations on Olark
5. If the format of the lead add is the lead origin.
6. If they are a working professional at the time.

Considering them, the X Education can succeed because they have a very high