<u>Summary</u>

To draw more business people to their courses, X Education is conducting this analysis. The basic information on how prospective clients use the site, how long they remain there, how they got there, and the conversion rate taught us a lot.

Following steps were used to perform the analysis:

1. Data Cleaning

The data have a few null values. The chosen option, however, had to be modified to a null value because it didn't offer much helpful information. A handful of the null values were converted to "NaN" to prevent too much data loss. Handled missing values by dropping the columns that have more 45% of missing values and imputing wherever necessary.

2. Exploratory Data Analysis

Assessed the data using EDA via Univariate and Bivariate Analysis on Numeric as well as Categorical Variables in the data

frame. It was found that some of the components of the categorical variables were unnecessary. There were no outliers, and the numerical data are correct.

3. Dummy Variables

The dummy variables were added, and there after those variables that had Null values were removed from the dataset. To scale numerical numbers, we used the 'MinMaxScaler'.

4. Train-Test Split

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

5. Building the Model

RFE first identified the most important 15 relevant parameters. The other variables were later manually eliminated based on the VIF values as well as p-values (The variables with $VIF < 5$ and p-value < 0.05 were kept).

6. Model Evaluation

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

7. Model Prediction

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

8. Precision – Recall

On the test data frame, a cut off of 0.4 was determined for the test data frame, with a Precision and Recall of about 73% and 75%, respectively.

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

a. The total time spent on the Website
b. Total number of visits
c. When the lead source was:
   i.   Google
   ii.   Direct traffic
   iii. Organic search
   iv.   Welingak website

d. When the last activity was:
   i.   SMS
   ii.   Olark chat conversation

e. When the lead origin is Lead add format
f. When their current occupation is as a working professional

Keeping these in mind the X Education can grow their business as they have a very high chance to get most of the potential buyers to change their mind and buy their courses.