# SUMMARY

This analysis was conducted for X Education to increase enrollment of industry professionals in their courses. The initial data provided valuable insights into visitor behavior on the site, including visit duration, source of traffic, and conversion rates. The following steps were taken:

## 1. Data Cleaning:

- Null values were addressed, and 'select' options were replaced with null to eliminate irrelevant information.

- Some null values were replaced with 'not provided' to retain data, though they were later removed when creating dummy variables.

- Country values were categorized as 'India', 'Outside India', and 'not provided'.

## 2. Exploratory Data Analysis (EDA):

- A quick examination of the data revealed irrelevant elements in categorical variables, while numeric values appeared sound with no outliers.

## 3. Dummy Variables:

- Dummy variables were created and those with 'not provided' elements were removed.

- MinMaxScaler was used for numeric values.

## 4. Train-Test Split:

   - Data was split into 70% for training and 30% for testing.

## 5. Model Building:

   - Recursive Feature Elimination (RFE) was employed to identify the top 15 relevant variables.

   - Remaining variables were removed based on VIF values and p-values (keeping variables with VIF < 5 and p-value < 0.05).

## 6. Model Evaluation:

   - A confusion matrix was generated.

   - Optimal cutoff value (determined using ROC curve) was used to calculate accuracy, sensitivity, and specificity, all around 80%.

## 7. Prediction:

   - Predictions were made on the test data frame using an optimal cutoff of 0.35, achieving 80% accuracy, sensitivity, and specificity.

## 8. Precision-Recall:

   - This method was also applied, resulting in a cutoff of 0.41 with precision around 73% and recall around 75% on the test data frame.

Key variables influencing potential buyers, in descending order, were found to be:

1. Total time spent on the website.

2. Total number of visits.

3. Lead source:

   - Google

   - Direct traffic

   - Organic search

   - Welingak website

4. Last activity:

   - SMS

   - Olark chat conversation

5. Lead origin as Lead add format.

6. Current occupation as a working professional.

Utilizing these insights, X Education can significantly increase their enrollment by targeting potential buyers effectively.