# SUMMARY

These are the steps to follow for solving the given problem:

## 1. Importing, Reading, and Understanding the Data:

- The data was imported and read into the environment. Then, we analyzed its data types, statistical information, and overall structure.

## 2. Data Cleaning:

- The data was partially cleaned, addressing a few null values, and instances where the option "select" was replaced with null.
- Variables with a high percentage of null values were evaluated and dropped based on their importance.
- Given that most leads were from India with only a few from outside, we categorized them as 'India,' 'Outside India,' and 'Not Provided.'

## 3. Exploratory Data Analysis (EDA):

- Univariate analysis was conducted on both categorical and numerical variables, revealing various insights.
- Bivariate analysis was also performed, relating the categorical and numerical variables to the target variable ("Converted"), leading to additional insights.
- During this analysis, variables that were redundant or had single values were dropped.

## 4. Creation of Dummy Variables:

- Dummy variables were created for the remaining categorical variables.
- Features were scaled using the Min-Max scaler.

## 5. Train-Test Split:

- The data was split into training and testing sets, with 70% used for training and 30% for testing.

## 6. Building the Model:

- Feature selection was done using RFE (Recursive Feature Elimination), identifying the top 15 important features.
- A logistic regression model was then built.

- Additional variables were manually removed based on VIF values and p-values.

## 7. Model Evaluation:

- A confusion matrix was plotted.
- The optimal cut-off value, determined using the ROC curve, was 0.35. This resulted in accuracy, sensitivity, and specificity of approximately 80% each.

## 8. Model Prediction:

- Predictions were made on the test data, with the optimal cut-off of 0.35, resulting in an accuracy of around 80%, sensitivity around 80%, and specificity around 81%.

## 9. Precision-Recall Analysis:

- A precision-recall analysis was also conducted, yielding an optimal cut-off of 0.41 with precision around 74% and recall around 76%.

Key Insights for X Education:

To enhance lead conversion, X Education should focus on the following factors (in descending order of importance):

- Website Engagement: Leads that spend more time on the website show greater interest. Enhancing user engagement on the website can further increase this interest.
- Website Visits: The number of times a lead visits the website correlates with higher conversion rates.
- Lead Source: Leads coming from Google, direct traffic, organic search, and the Welingak website show varying levels of interest, with Google being the most promising. These sources should be the primary focus for targeting.
- Last Activity: Leads whose last activity involved SMS or Olark chat are more active and engaged.
- Occupation: Leads identified as working professionals are likely seeking career advancement, making them more interested in the courses offered.