SUMMARY

This analysis focuses on X Education's efforts to attract more industry professionals to their courses.

Here's how we've approached it:

1. **Data Cleaning:** We cleaned the data by handling null values and making necessary adjustments to ensure better analysis. For instance, we replaced 'select' options with null values and categorized locations into 'India', 'Outside India', and 'not provided'.
2. **EDA:** We performed exploratory data analysis to understand our data better. We found some irrelevant elements in categorical variables but no outliers in numeric values.
3. **Dummy Variables:** We created dummy variables and removed those with 'not provided' elements. Numeric values were scaled using MinMaxScaler.
4. **Train-Test Split:** The data was split into training (70%) and testing (30%) sets.
5. **Model Building:** We used Recursive Feature Elimination (RFE) to select the top 15 relevant variables and further removed variables based on VIF and p-values.
6. **Model Evaluation:** Confusion matrix and ROC curve were used to evaluate the model. An optimum cut-off value of 0.35 yielded an accuracy, sensitivity, and specificity of 80%.
7. **Prediction:** We made predictions on the test data using the selected cut-off value.
8. **Precision-Recall:** This method was also used, resulting in a cut-off value of 0.41 with precision and recall around 73% and 75%, respectively.

Overall, variables like total time spent on the website, total number of visits, specific lead sources (Google, direct traffic, organic search), last activity (SMS, Olark chat), lead origin (Lead add format), and occupation (working professional) were found to be significant in identifying potential buyers. By focusing on these aspects, X Education can increase its chances of converting potential leads into customers effectively.