SUMMARY

This analysis was conducted for X Education to identify strategies for attracting more industry professionals to their courses. The provided data offered insights into how potential customers interact with the site, including their visit patterns, time spent, sources of traffic, and conversion rates. The steps taken in the analysis are as follows:

1. **Data Cleaning:** The data was mostly clean, with a few null values that required attention. The "option select" was replaced with null values due to its lack of informational value. Some null values were changed to 'Other' or filled with the mode to retain as much data as possible. However, these were later removed during the creation of dummy variables. The geographic distribution was heavily skewed towards India, with fewer entries from other regions.

2. **Exploratory Data Analysis (EDA):** A quick EDA revealed that several elements within the categorical variables were irrelevant. The numeric values appeared well-distributed, with no significant outliers.

3. **Dummy Variables:** Dummy variables were created for categorical features, and numeric values were scaled using MinMaxScaler.

4. **Train-Test Split:** The data was split into training and testing sets with a 70-30 ratio.

5. **Model Building:** Recursive Feature Elimination (RFE) was performed to identify the top 20 relevant variables. Additional variables were removed manually based on Variance Inflation Factor (VIF) and p-values, retaining those with VIF < 5 and p-value < 0.05.

6. **Model Evaluation:** A confusion matrix was constructed, and the optimal cutoff value was determined using the ROC curve. The model achieved an accuracy, sensitivity, and specificity of around 90%.

7. **Prediction:** Predictions were made on the test data, with an optimal cutoff of 0.35, yielding accuracy, sensitivity, and specificity of approximately 80%.

8. **Precision-Recall:** The Precision-Recall method was also employed, identifying a cutoff of 0.41, with a precision of around 79% and recall of 70% on the test data.

The analysis identified key variables influencing potential buyers, including:

1. Total time spent on the website.

2. Total number of visits.

3. Lead sources such as:

   o   Referrals

   o   Phone Conversation

   o   Welingkar website

4. Last activities like:

- o SMS

- o Phone conversation

- o Unsubscribing

- o Phone Conversation

5. Current occupation as a working professional.

By focusing on these factors, X Education has a strong opportunity to convert a significant portion of potential buyers into actual customers, thereby boosting their course enrollments.