Summary

This analysis is conducted for X Education to identify strategies to attract more industry professionals to their courses. The initial data provided valuable insights into how potential customers visit the site, the time they spend there, how they arrived at the site, and the conversion rate.

Steps Involved:

1. **Data Cleaning:**
   o The data was largely clean except for a few null values. The option select was replaced with a null value due to its lack of informative value. Some null values were changed to 'not provided' to retain as much data as possible, although they were later removed during the creation of dummy variables. The geographic data was categorized as 'India', 'Outside India', and 'not provided'.
2. **Exploratory Data Analysis (EDA):**
   o A preliminary EDA revealed that many categorical variable elements were irrelevant. The numeric values appeared appropriate with no outliers detected.
3. **Dummy Variables:**
   o Dummy variables were created, and those with 'not provided' elements were subsequently removed. MinMaxScaler was applied to the numeric values.
4. **Train-Test Split:**
   o The data was split into 70% for training and 30% for testing.
5. **Model Building:**
   o Recursive Feature Elimination (RFE) was performed to identify the top 15 relevant variables. Additional variables were manually removed based on their Variance Inflation Factor (VIF < 5) and p-value (< 0.05).
6. **Model Evaluation:**
   o A confusion matrix was created. The optimal cutoff value, determined using the ROC curve, was used to calculate the accuracy, sensitivity, and specificity, which were approximately 80% each.
7. **Prediction:**
   o Predictions were made on the test data with an optimal cutoff of 0.35, achieving an accuracy, sensitivity, and specificity of 80%.

Key Influential Variables: In descending order of importance:

1. Total time spent on the website.
2. Total number of visits.
3. Lead source:
   o Google
   o Direct traffic
   o Organic search
   o Welingak website

4. Last activity:
    o SMS
    o Olark chat conversation
5. Lead origin as Lead add format.
6. Current occupation as a working professional.

By focusing on these factors, X Education can significantly increase the likelihood of converting potential buyers into course participants.