Summary

This analysis was conducted for X Education to explore strategies for attracting more industry professionals to their courses. The initial data provided valuable insights into potential customers' site visits, the duration of their stay, their entry points to the site, and conversion rates.

The steps undertaken were as follows:

1. Data Cleaning:
   - The data was mostly clean, with a few null values. Options selected with limited information were replaced with null values, and some null values were changed to 'not provided' to retain data. Elements were categorized into 'India', 'Outside India', and 'not provided' for better clarity.

2. Exploratory Data Analysis (EDA):
   - A preliminary EDA revealed that many elements in categorical variables were irrelevant. Numeric values were satisfactory, with no outliers detected.

3. Dummy Variables:
   - Dummy variables were created, and those with 'not provided' elements were removed. Numeric values were scaled using MinMaxScaler.

4. Train-Test Split:
   - The dataset was split into 70% training data and 30% testing data.

5. Model Building:
   - Recursive Feature Elimination (RFE) identified the top 15 relevant variables. Additional variables were manually removed based on VIF values and p-values, retaining those with VIF < 5 and p-value < 0.05.

6. Model Evaluation:
   - A confusion matrix was created, and the optimal cutoff value (determined using the ROC curve) provided accuracy, sensitivity, and specificity around 80%.

7. Prediction:
   - Predictions were made on the test dataset with an optimal cutoff of 0.35, achieving accuracy, sensitivity, and specificity of 80%.

8. Precision-Recall:
   - Precision-Recall analysis was performed, identifying an optimal cutoff of 0.41 with a precision of approximately 73% and a recall of 75% on the test dataset.

The key variables influencing potential buyers, in descending order of importance, are:

1. Total time spent on the website.

2. Total number of visits.
3. Lead source, specifically:
   - Google
   - Direct traffic
   - Organic search
   - Welingak website
4. Last activity, specifically:
   - SMS
   - Olark chat conversation
5. Lead origin as a lead ad format.
6. Current occupation as a working professional.

With these insights, X Education has a strong opportunity to convert a significant number of potential buyers into customers for their courses.