Summary

This analysis was conducted for X Education to identify strategies to attract more industry professionals to enroll in their courses. Using the provided data, we gained insights into how potential customers interact with the website, including their visit patterns, time spent on the site, traffic sources, and conversion rates.

Steps Followed:

1. **Data Cleaning**: The dataset was partially clean, but it included some null values and irrelevant entries (such as "option select," which was replaced with null for clarity). Columns with more than 30% null values were removed, and other missing values were filled using median and mode values.
2. **Exploratory Data Analysis (EDA)**: A quick EDA helped evaluate the data's quality. We observed that some categorical variables contained irrelevant elements, prompting a univariate and bivariate analysis, along with outlier detection and treatment.
3. **Dummy Variables Creation**: Dummy variables were created for categorical data, followed by dropping redundant columns. The data was then standardized using the StandardScaler.
4. **Train-Test Split**: The dataset was divided into training and test sets, with a split of 70% for training and 30% for testing.
5. **Model Building**: RFE (Recursive Feature Elimination) was initially used to select the top 15 relevant variables. Variables with high VIF or insignificant p-values were then manually removed to improve model performance.
6. **Model Evaluation**: A confusion matrix was created, and an optimal cutoff value was identified to assess the accuracy, sensitivity, and specificity, each approximately 80%.
7. **Prediction**: Predictions were made on the test data using an optimal cutoff of 0.35, yielding an accuracy, sensitivity, and specificity close to 80%.
8. **Precision and Recall**: Precision and recall analysis confirmed an optimal cutoff of 0.35, with approximately 75% precision and 76% recall on the test data.

Key Findings and Recommendations
To enhance lead conversion rates, X Education should focus on the following influential features:

- **Lead Source: Welingak Website**: Leads directed from the Welingak website show a higher conversion rate, suggesting that targeted efforts on this platform could yield more high-potential leads.
- **Lead Origin: Lead Add Form**: Leads who engaged through the 'Lead Add Form' demonstrated higher conversion rates, indicating that emphasizing this entry point may increase the lead pool with a higher chance of conversion.
- **Current Occupation: Working Professional**: Leads identified as working professionals have a higher likelihood of conversion. Focusing on engaging this demographic could enhance overall conversion rates.