Case Study Solution Summary

This analysis was conducted to help X Education attract more industry professionals to enroll in their courses. The provided data offered valuable insights into potential customers' behavior, website visits, time spent, referral sources, and conversion rates.

The key steps taken in this analysis are as follows:

1. Data Cleaning: The data was partially clean, but some null values and irrelevant options needed attention. Null values were replaced with 'not provided' to retain crucial information, and categorical variables were appropriately categorized. Numeric values were also checked for outliers.

2. Exploratory Data Analysis (EDA): A quick EDA helped identify irrelevant elements in categorical variables while confirming the suitability of numeric values with no outliers.

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

4. Train-Test Split: The data was divided into 70% for training and 30% for testing purposes.

5. Model Building and Feature Selection: Recursive Feature Elimination (RFE) was employed to select the top 15 relevant variables. Additional variables were manually removed based on VIF values and p-values (keeping variables with VIF < 5 and p-value < 0.05).

6. Model Evaluation: A confusion matrix was used to evaluate the model. The optimal cut-off value, obtained using the ROC curve, resulted in an accuracy, sensitivity, and specificity of around 80%.

7. Prediction: The model was used to predict outcomes on the test dataset, using an optimal cut-off of 0.35, with accuracy, sensitivity, and specificity of 80%.

8. Precision-Recall: An additional evaluation using Precision-Recall was performed, identifying a cut-off of 0.41 with precision around 73% and recall around 75% on the test dataset.

Key variables impacting potential buyers were identified in descending order of importance, including website time spent, total visits, lead source (Google, Direct traffic, Organic search, and Welingak website), last activity (SMS and Olark chat conversation), lead origin as Lead add format, and current occupation as a working professional.

Leveraging this information, X Education can optimize its marketing strategies, catering to potential buyers and significantly increasing the likelihood of successful course enrollments.