Analysis Overview

This study for X Education aims to identify strategies for attracting more industry professionals to enroll in their courses. Using the provided dataset, we examined customer behavior, including site visits, engagement metrics, and conversion rates. Below are the key steps and findings:

Methodology

1. Data Cleaning:
   - Addressed missing values, replacing irrelevant options (e.g., "Select") with nulls or "Not Provided."
   - Simplified location data into categories: "India," "Outside India," and "Not Provided."
2. Exploratory Data Analysis (EDA):
   - Identified and removed irrelevant categorical elements.
   - Verified that numeric data was clean and free of outliers.
3. Feature Engineering:
   - Created dummy variables; excluded those with "Not Provided" values.
   - Applied MinMaxScaler for numeric values.
4. Train-Test Split:
   - Split the data into 70% training and 30% testing sets.
5. Model Building:
   - Used Recursive Feature Elimination (RFE) to identify the top 15 variables.
   - Removed features with high Variance Inflation Factor (VIF > 5) or insignificant p-values (p > 0.05).
6. Model Evaluation:
   - Used a confusion matrix and ROC curve to determine the optimal cutoff value.
   - Achieved ~80% accuracy, sensitivity, and specificity.
7. Prediction:
   - Applied the model on test data with a cutoff of 0.35, resulting in 75% accuracy, 62% sensitivity, and 82% specificity.
8. Precision-Recall Analysis:
   - Determined an alternative cutoff of 0.41 with precision of 68% and recall of 64%.

Key Insights

The most influential factors for identifying potential buyers (in descending importance):

1. Time spent on the website.
2. Number of site visits.
3. Lead Source: Google, Direct Traffic, Organic Search, Welingak Website.
4. Last Activity: SMS, Olark Chat Conversations.

5. Lead Origin: Lead Add Format.
6. Occupation: Working professionals.

Conclusion
By focusing on these factors, X Education has a strong potential to convert prospective buyers into customers, enhancing their market reach and enrollment success.