Summary

Objective:

- Increase enrollment of industry professionals in X Education courses.
- Identify 'Hot Leads' with higher conversion probability.

Data Analysis Process:

1. **Data Cleaning**:
    - Replaced 'select' with null.
    - Changed null values to 'not provided'.
    - Categorized locations into 'India', 'Outside India', and 'Not Provided'.
2. **Exploratory Data Analysis (EDA)**:
    - Identified irrelevant elements in categorical variables.
    - Numeric values had no major outliers.
3. **Dummy Variables & Scaling**:
    - Created dummy variables and removed 'not provided' elements.
    - Applied MinMaxScaler for numeric values.
4. **Train-Test Split**:
    - Split data into 70% training and 30% testing.
5. **Model Building**:
    - Used RFE for top 15 variables.
    - Removed variables manually based on VIF (<5) and p-value (<0.05).
6. **Model Evaluation**:
    - Confusion matrix showed 80% accuracy, sensitivity, and specificity.
    - Optimal cut-off value (using ROC curve) found at 0.35.
7. **Prediction**:
    - Test data predictions with 0.35 cut-off: 80% accuracy, sensitivity, specificity.
8. **Precision – Recall**:
    - Optimal cut-off of 0.41: Precision 73%, Recall 75%.

Key Variables Impacting Conversion:

1. Total time spent on the website.
2. Total number of visits.
3. Lead Source:
    - Google
    - Direct traffic
    - Organic search
    - Welingak website
4. Last Activity:
    - SMS
    - Olark chat conversation
5. Lead Origin: Lead add format.
6. Current Occupation: Working professional.

Conclusion: By focusing on these key variables, X Education can significantly increase their lead conversion rates, targeting potential buyers more effectively and efficiently.