Summary

This analysis focuses on lead scoring for X Education, utilizing logistic regression to enhance conversion rates. The goal is to identify the factors that influence potential customers to enroll in courses.

Key Steps in the Analysis :

1. Data Cleaning:
   - The dataset contained some missing values and an option labeled "Select," which did not provide meaningful insights. These were replaced with "Not Provided" where necessary.
   - The country data was categorized into "India," "Outside India," and "Not Provided" for better segmentation.

2. Exploratory Data Analysis (EDA):
   - An initial examination of the data revealed that certain categorical variables had irrelevant entries.
   - The numerical variables were well-structured, with no significant outliers.

3. Feature Engineering:
   - Dummy variables were created for categorical data, and features with "Not Provided" values were excluded.
   - MinMaxScaler was applied to normalize numeric attributes.

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

5. Model Building:
   - Recursive Feature Elimination (RFE) identified the 15 most relevant variables.
   - Features were manually removed based on Variance Inflation Factor (VIF) and p-values (keeping variables with VIF < 5 and p-value < 0.05).

6. Model Evaluation:
   - A confusion matrix was generated, and the optimal cut-off value was determined using the ROC curve.
   - The final model achieved an accuracy, sensitivity, and specificity of approximately 80%.

7. Prediction & Performance Metrics:
   - On test data, an optimal cut-off of 0.35 yielded an accuracy, sensitivity, and specificity of 80%.
   - Precision-Recall analysis suggested an alternate cut-off of 0.41, with precision around 73% and recall at 75%.

Insights & Key Influencing Factors :

The most influential variables in determining lead conversion were:

1. User Engagement Metrics:
   - Total time spent on the website.
   - Total number of visits.

2. Lead Source:
   - Google
   - Direct traffic
   - Organic search
   - Welingak website

3. Last Notable Activity:
   - SMS interactions
   - Olark chat conversations

4. Lead Origin:
   - Leads generated through ad formats had higher conversion rates.

5. Customer Profile:
   - Leads from finance, HR, and marketing specializations had a higher probability of conversion.
   - Most enrolled leads were looking for better career prospects, with finance management being a popular specialization.
   - The majority of leads were unemployed, indicating a potential focus area for targeted marketing strategies.

6. Customer Engagement Strategies:
   - Sending emails and SMS significantly increased the likelihood of conversion.
   - Improved engagement via calls and emails could further boost lead conversion rates.

Conclusion :

By optimizing marketing efforts based on these insights, X Education can significantly improve conversion rates and attract more industry professionals to their courses.