Summary:

The focus of this endeavour has been to construct and utilize a predictive model for Company X Education, aimed at enhancing the conversion of potential users. The process followed involved comprehensive steps to comprehend and validate the data, ultimately enabling the identification of the appropriate target groups to augment the conversion rate. The key stages are outlined as follows:

1. **Exploratory Data Analysis (EDA):**
   - An initial assessment was conducted to ascertain the percentage of missing values, leading to the exclusion of columns with over 45% missing data.
   - To preserve significant columns, missing values were replaced with 'not provided' since the associated data loss would have been substantial.
   - Recognizing that 'India' was the dominant entry among non-missing values, all 'not provided' entries were imputed with 'India'.
   - Subsequently, given the overwhelming dominance of 'India' entries (about 97% of the dataset), the column was dropped.
   - Focus was also given to numerical variables, including outlier management and the creation of dummy variables.

2. **Train-Test Split & Scaling:**
   - Data was divided into training (70%) and testing (30%) sets.
   - Scaling was employed using min-max scaling on specific variables: ['TotalVisits', 'Page Views Per Visit', 'Total Time Spent on Website'].

3. **Model Building:**
   - Feature selection was executed using Recursive Feature Elimination (RFE).
   - The top 15 relevant variables were identified through RFE.
   - Remaining variables were pruned based on considerations of VIF values and p-values.
   - A confusion matrix was generated to assess overall accuracy, yielding a value of 92.78%.

4. **Model Evaluation:**
   - On the training data:
     - Optimal cutoff value determination involved utilizing the ROC curve, yielding an area under the curve of 0.97.
     - After analysis, an optimum cutoff of 0.5 was identified, resulting in:
       - Accuracy: 92.29%
       - Sensitivity: 91.70%
       - Specificity: 92.66%
   - Prediction on the test data:
     - The outcomes were:
       - Accuracy: 92.78%

- Sensitivity: 91.98%
- Specificity: 93.26%

Conclusion:

The pivotal variables contributing significantly to the conversion rate encompass:

- **Lead Source:**
  - Total Visits
  - Total Time Spent on Website
- **Lead Origin:**
  - Lead Add Form
- **Lead Source:**
  - Direct Traffic
  - Google
  - Welingak Website
  - Organic Search
  - Referral Sites
- **Last Activity:**
  - Do Not Email_Yes
  - Last Activity_Email Bounced
  - Olark Chat Conversation

The model demonstrates a highly effective prediction of conversion rates, instilling a high degree of confidence in the company's decision-making based on the model's insights.