Summary

This analysis is done for X Education to find ways to get more industry professionals to join their courses. The basic data provided gave us a lot of information about how the potential customers visit the site, the time they spend on their website, how they reached the site, and the conversion rate.

The following are the steps used:

1) **Reading and Cleaning the data:**
- Read the data and performed the basic checks e.g. info, shape, describe.
- To clean the dataset we removed the redundant columns.
- Dropped the columns having more than 30% of missing values.
- Few of the null values were changed to 'not provided' so as to not lose much data.

2) **EDA:**
- A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seem good.
- Checked the outliers and created bins for them.
- Removed all the redundant and repeated columns

3) **Data Preparation:**
- Changed the multicategory labels into dummy variables and binary variables into '0' and '1'.
- The split was done at 70% and 30% for train and test data respectively.
- After this, we plot a heatmap to check the correlations among the variables.

4) **Model Building:**
- Created model with rfe count 19 and 15 and compared the model and choose our final model with rfe 19 variables as has more stability and accuracy than the other.

- Checked the precision and recall with accuracy, sensitivity, and specificity for our final model and the tradeoffs.
- We found the accuracy and sensitivity score from our final test model and it is in the acceptable range. The Accuracy was 81%, Precision was 74% and Recall 79%.

5) Conclusion:

- Important features responsible for good conversion rate or the ones which contribute more towards the probability of a lead getting converted are :
  - ➢ Last Notable Activity_Had a Phone Conversation
  - ➢ Lead Origin_Lead Add Form
  - ➢ What is your current occupation_Working Professional