Summary

In summary, the purpose of this analysis is to help X Education identify strategies for increasing the number of industry experts enrolled in their courses. We learned a lot about how potential customers visit the website, how long they stay there, how they found it, and the conversion rate from the basic data that was provided. Steps undertaken:

1. Data cleaning:

The option pick had to be changed to a null value because it provided us with little information, and the data was mostly clean aside from a few null values. To avoid losing a lot of data, a small number of the null values were changed to "not provided." However, when creating dummies, they were later taken out. The terms "India," "Outside India," and "not provided" were substituted because there were a lot of people from India and not many from elsewhere.

2. EDA:

To assess the state of our data, a brief EDA was conducted. Numerous components of the category variables were determined to be meaningless. There were no outliers discovered, and the numerical values appear to be good.

3. Dummy Variables:

After creating the dummy variables, the ones that included "not provided" parts were eliminated. We utilized the MinMaxScaler for numerical values.

4. Train-Test split:

The train and test data were separated at 70% and 30%, respectively.

5. Model Building:

To identify the top 17 pertinent variables, RFE was conducted first. Afterwards, the remaining variables were eliminated by hand based on the p-value and VIF values

(the variables with p-value < 0.05 and VIF < 5 were retained).

6. Evaluation of the Model:

A confusion matrix was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve and the optimal cut off value, which turned out to be about 80% each.

7. Projected:
The test data frame was used for prediction, and the best cutoff was 0.30 with 80% accuracy, sensitivity, and specificity.

8. Precision-Recall:

 Using this procedure again, a cutoff of 0.30 was discovered, with precision at about 80%.