# Summary

Company X wants to increase the number of leads to join the courses. There are enough data points available in the dataset based on which the regression model is built which will assist the company in deciding the factors effective in converting the leads into successful leads.

The following steps were undertaking during data preparation and logistic model development:

1. Data Cleaning:
    Eliminating NULL values
    Imputation with significant alternatives
    Removal of non essential columns

2. EDA:
    Visualization & drawing inference
    Including box plots against numerical variable to identify outliers

3. Dummy Variables:
    Dummy variables were created using pandas.dummy_variables method enhancing the dataset's effectiveness for the categorical variables

4. Train Test Split:
    A 70:30 data split ration was applied leveraging sufficient training data to enhance model resulting in correct output

5. Model Building:
    Logistic regression was used along with various other python libraries for building the model incorporating RFE (Recursive Feature Elimination), VIF (Variance Inflation Factor).
    RFE considered 15 columns while VIF    aimed for values < 5 & p-value below .005 only.

6. Model Evaluation:
    A cutoff of 0.45 was set initially. Evaluation metrics included confusion matrix,  sensitivity, specificity, precision, and recall. Achieved results for the Logistic        Regression model were:
    Sensitivity ~ 74%
    Specificity ~ 83%
    Precision ~ 77%
    Recall ~ 79%

To conclude, X Education can employ below strategies to enhance lead conversion:
- Target working professionals, since they have higher inclination to enroll in courses.
- Consider individuals who have actively engaged with emails by opening/clicking them for successful leads.
- Approach website visitors who spend a substantial amount of time on the site for potential conversions.