Summary

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

Following are the steps involved while preparing the data and building a Logistic Regression model:

1. **Cleaning the Data:** This involved eliminating the null values, imputing them with other significant values, dropping the columns which are not significant for the analysis.
2. **Exploratory Data Analysis:** This involved the visualization of the data and drawing out the inferences from the data available. Various box plots were plotted against the numerical variables in order to investigate the outliers.
3. **Dummy variables:** In order to cater the categorical variables, provision is made to create dummy variables with the use of pandas.dummy_variables method. This made the dataset more effective and useful as now categorical variables could also be potentially considered for the logistic model.
4. **Train-Test Split:** The data dichotomy is made in ratio of 70:30. Since enough train data is available, it made the model quite effective in resulting out the correct outputs.
5. **Model Building:** The model is build using Logistic Regression. Various python libraries and methods were made into use for building the model. Strategies involved RFE (Recursive Feature Elimination), VIF (Variance Inflation Factor). RFE was made to consider 15 columns and for the optimum results, VIF was decided to be less than 5 and p-value to considered below 0.005 only.
6. **Model Evaluation**: The cut-off was decided to be 0.45 as the early assumption. Evaluation in performed via various measures such as: Confusion matrix, sensitivity, specificity, precision, recall. In our Logistic Regression model, we achieved Sensitivity ~ 69.4% Specificity ~ 74.3% Precision ~ 74% Recall ~ 70% In nutshell,

X Education can make use of the following points in order to convert their leads into successful leads:

- approached to convert them into successful leads
- It is observed that those who working professionals are more prone to opt for the courses so business should focus on working professionals for lead
- Those who visits the website and spend considerable amount of time there, can be approached to convert them into successful leads
- Also, those are coming from source such as reference and Welingak website can also be taken into consideration for the successful lead, these applications should be treated as hot leads.
- Business can max the number of leads generated by direct traffic and google
- By closely looking at the data when the last activity is converted to Lead, then there are high chances of them getting converted into successful leads