Summary

This case study is basically an analysis done for X Education to increase the number of enrollments in the courses available for industry professionals. The Company has collected a lot of information about the customers, who visited the sites and how much they have spent on the website and the source of customers. On the basis of the same information, the marketing team connects to the customers and converts the leads into final enrollments.

At present the conversion rate is 30% but the company targets to increase this up to 80% using the same data. To get it done, these are the following steps:

1. **Data Preparation and Cleaning:**
   The given data was partially clean with a few null values and 'Select' values. The select values are required to be changed to null as they are not carrying much information. Features having more than 50% null values are dropped and are treated individually as per the importance of the features.

2. **Exploratory Data Analysis:**
   The data after cleaning, went through EDA to analyze the relevance of the data. Here we found a lot of features in categorical variables were irrelevant. The numerical variables were useful and without any outliers.

3. **Dummy Variables:**
   The dummy variables were created and the features without dummy variables were removed. MinMaxScaler method was used for numerical variables.

4. **Train-Test split:**
   The data was splitted at 70:30 for train-test data respectively.

5. **Model Building:**
   RFE was performed to attain the top relevant features. Later on VIF values and p-values were calculated to attain the target.

6. **Model Evaluation:**
   Confusion matrix was drawn. Using ROC curve, the optimum cut off value was used to find the accuracy, sensitivity and specificity with around 80% each.

7. **Prediction:**
   Prediction was done on the test data with an optimum cutoff of 0.40, got by Accuracy, sensitivity and specificity of 80%.