Summary

This analysis is done for X Education and 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 there, how they reached the site and the conversion rate.

The following are the steps used:

1. **Cleaning data:**

The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed so as to not lose much data. Although they were later removed while making dummies.

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 seems good and outliers were handled for each column individually.

3. **Dummy Variables:**

The dummy variables were created and later on the dummies with ‘not specified elements were removed. For numeric values we used the StandardScaler.

4. **Train-Test split:**

The split was done at 70% and 30% for train and test data respectively.

5. **Model Building:**

Firstly, RFE was done to attain the top 15 relevant variables. Later the rest of the variables were removed manually depending on the VIF values and p-value (The variables with VIF < 5 and p-value < 0.05 were kept).

6. **Model Evaluation:**

A confusion matrix was made. Later on the optimum cut off value (using ROC curve)

The ROC curve has a value of 0.97, which is very good. We have the following values for the Train Data:

- Accuracy : 92.29%
- Sensitivity : 91.70%
- Specificity : 92.66%

7. **Prediction:**

Prediction was done on the test data frame and with an optimum cut off as 0.3 with

- Accuracy : 92.78%
- Sensitivity : 91.98%
- Specificity : 93.26%

8. **Precision – Recall:**

This method was also used and a cut off of 0.3 was found with Precision

around 89% and recall around 91% on the test data frame.

The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model.