Summary

This case study is for X Education and to find ways to convert prospective leads to join their courses. The basic data provided has information about how the potential customers visit the site, the time they spend there, how they reached the site and theconversion rate.

The following are the steps used:

1. **Cleaning data:**
   The data was partially clean except for few null values and the option select was replaced with a null value since it did not give us much information. Few missing values were changed using mode as they would not have added much to the data. Few variables like Lead Profile, How did you hear about X Education and Asymmetric predict and score were dropped.

2. **EDA:**
   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 no outliers were found.

3. **Dummy Variables:**
   The dummy variables were created using One-Hot Encoding. For numeric values we used the Standard scaler.

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

5. **Model Building:**
   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 (Thevariables with VIF < 8 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)was used to find the accuracy, sensitivity and specificity which came to be around ~85-90%.

7. **Prediction:**
   Prediction was done on the test data frame and with an optimum cut off as 0.2 withaccuracy, sensitivity and specificity between ~85 % - 90%.

8. **Precision – Recall:**
   This method was also used to recheck and a cut off of 0.2 was found with Precisionaround 93% and recall around 85% on the test data frame.

Conclusion:

It was found that the variables that mattered the most in the potential buyers are

1. Tags_Lost to EINS

2. Tags_Closed by Horizzon

3. Tags_Will revert after reading the email

4. Tags_Busy

5. Lead Source_Welingak Website

6. Last Notable Activity_SMS Sent

7. Lead Origin_Lead Add Form

X education has high chance of converting prospective leads to buy their courses if they focus on the above variables with "tags" variable being the top 4 variables to be utilized most.