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. **Data Cleaning and manipulation:**

   The Dataset had 37 columns & over 9000 rows initially.
   ❖ We removed columns which had more than 4000 missing values.
   ❖ Removed remaining redundant columns which had more than 90% similar
   ❖ Assigned Blank columns

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.

3. **Dummy Variables:**
   The dummy variables were created and later on the dummies with 'not provided' elements were removed. For numeric values we used the MinMaxScaler.

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) was used to find the accuracy, sensitivity and specificity which came to be above 80% each.

   Train Data Set metrics v/s Test Data Set metrics

   Sensitivity: 89.73 vs Sensitivity: 88.02

   Specificity: 89.18 vs Sensitivity: 88.02
   Precision: 83.59 vs Precision: 83.95

   Recall: 89.73 vs Recall: 88.02
   Accuracy: 89.39 vs Accuracy: 89.17

Top variables and their co-efficients:

```
Tags_Closed by Horizzon                               6.347721
Lead_Origin_Lead Add Form                             5.575817
Tags_Will revert after reading the email              4.320284
Last Notable Activity_SMS Sent                        2.085393
Lead Source_Olark Chat                                1.486492
Total Time Spent on Website                           1.109035
What is your current occupation_Working Professional  0.950998
Specialization_Travel and Tourism                    -1.317662
Do Not Email                                         -1.342537
Last Activity_Olark Chat Conversation                -1.604676
const                                                -1.941104
Tags_Interested in other courses                     -2.427775
Lead Source_Reference                                -2.917966
Tags_Ringing                                         -3.336737
Tags_Already a student                               -3.508195
```

Recommendations

❖ Phone calls must be made to those customers that spend most time on our website as they are seem

more interested in the course.

❖ Horizzon has proved to be a good partner. Therefore, we must provide better funding.

❖ Customers who revert after reading the email should be called ASAP.

❖ Target customers who aren't students as they are least likely to join our program.

❖ Make website more appealing so people are likely to spend more time on our website.

Keeping these in mind the X Education can flourish as they have a very high chance to get almost all the potential buyers to change their mind and buy their courses.

X-----X-----X------X