**Summary**

This analysis is done for X Education to identify potential leads that could be “Hot Leads”. Based on the data provided, we should aim to reach a lead conversion percentage of close to 80%+

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, what they specialize in, 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 to ‘not sure’ so as to not lose much data. The categorical binary data were converted to 1/0. The categorical data were later converted to dummy variables with one hot encoding. Since there were many from India and few from outside, the elements were mostly changed to ‘India’

**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 were good and no outliers were found.

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). The final set had around 12 variables that were significant

6. **Model Evaluation:**

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

was found

7. **Prediction:**

Prediction was done on the test data frame and with an optimum cut off as 0.2 with accuracy (91%), sensitivity (86%) and specificity (94%).

8. **Precision – Recall:**

This method was also used to recheck and a cut off of 0.2 was found

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

1. When the lead origin is Lead add format.
2. When the lead source was:
   1. Google
   2. Direct traffic
   3. Organic search
   4. Welingak website
3. When their current occupation is as a working professional
4. The total time spend on the Website
5. When the last activity was:
   1. SMS
   2. Olark chat conversation
6. Total number of visits.

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.

===============