Summary Report

The data of potential customers of X Education was analyzed to assign a lead score for every lead, where a higher number would mean a hot lead.

Below 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. 'Select' was imputed to Nan and then later to relevant catagories.

2. **Explanatory Data Analysis:**
   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 seem good and no outliers were found.

3. **Dummy Variables:**
   The dummy variables were created and later on the dummies were removed.

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 20 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 the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 90% each.

7. **Prediction:**
   Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 90%.

8. **Precision – Recall:**
   This method was also used to recheck and a cut off of 0.41 was found with Precision around 84% and recall around 91% on the test data frame.

- People who spend more time on the website or with more no. of visits convert more.
- When the last activity was SMS or Olark chat conversation the conversion was more.
- Working professionals have a high chance of conversion.
- While Reference are the best form of leads there are significant conversions from Google, Direct Traffic, Organic search & Welingak 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 products.