SUMMARY

Analysis completed for X Education based on dataset provided, data contains enough information about potential customers how many times they visit website, how much time they spend on website, and how they reached the website.

We completed this analysis using following steps:

1) **Data Cleaning:**
   First, we replace 'Select' comment with null values, after that we remove variableswhich have unique values after that we check for Percentage of missing values, those variables have high percent we replace null values with comment 'no info',
   After those rows have missing values we drop that rows, then we combine different country names in single comment 'other'.

2) **EDA(Exploratory data analysis):**
   In the process of EDA we found numerical variables related, two of them have outliers but that doesn't effects mean, but we also found that many categorical variables are not related.

3) **Creating Dummy Variables:**
   Dummy Variables created for variables but after creation we delete variables which contains 'no info' comment.

4) **Train-Test split:**
   We apply MinMaxScaler on numeric variables. After that we split data at 70% train and 30% test data.

5) **Model Building:**
   We attain the top 15 relevant variables, then after that we kept only that variables outof 15 variables those who have VIF<5 and P value<0.05.

6) **Model Evaluation:**
   We create Confusion matrix, and we need the optimum cut off value on which wecan attain accuracy, sensitivity and specificity 80% or nearer.

7) **Prediction:**
   We done prediction on test data at optimum cut off value 0.35 with accuracy, sensitivity and specificity of 80%.

8) **Precision and Recall:**
   We attain Precision 73.2% and Recall 76.6% on optimum cut off value 0.41.

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

1. The total time spend on the Website
2. Total number of visits
3. When the lead source was Welingak website, Olark chat, Organic search and Google.
4. When the last activity was:
   a) SMS
   b) Olark chat conversation

5. When the lead origin was 'Lead add form'.

6. When their current occupation is working professional or unemployed.

Keeping this 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.