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. Importing and Inspecting the Dataset:
The data was clean except for a few null values and the option select had to be replaced with a null value. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies.

2. Data Cleaning:
The missing values column-wise were checked and values with 30% missing values were dropped. The 'Lead Source', 'Total Visits', 'Page Views Per Visit', 'Last Activity', 'What matters most to you in choosing a course' and 'Lead Profiles' are variables checked.

3. Data Preparation:
Some binary variables were changed to 0/1. The values were checked for outliers.

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. Rescaling the features with MinMaxScaler:
 The converted rate is 39%.

7. Correlations Matrix, Model Building and Finding Optimal Cutoff Point:

The optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 80% each.

8. Prediction:
Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy, sensitivity and specificity of 80%.

9. Precision – Recall and many more steps:
This method was also used to recheck and a cut off of 0.41 was found with Precision around 80% and recall around 80% on the test data frame.

Summary: It was found that the variables that mattered the most in the potential buyers are (In descending order):

1. The total time spend on the Website.

2. Total number of visits.

3. Lead source was: Google, Direct traffic and Organic search
4. When the last activity was: Olark chat conversation
6. When their current occupation is as a working professional.

The X Education has a very high chance to get almost all the potential buyers to change their mind and buy their courses.