Summary

An education company named X Education sells online courses to industry professionals. This analysis is done to find ways to get more industry professionals to join their courses. some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%. 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. Through this analysis convert the target lead conversion rate is around 80% or more.

1. **Reading, Understanding, and Cleaning data:**
   Importing the data, checking all the properties, an overview of the data, checked for duplicate data, and there was no duplicated data. removed the columns with more than 40% of missing values. The select values in the column are null values so, we changed those values to null values.

2. **EDA:**
   EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. There are outliers in the numerical variable we cannot deal with them because they are not unreal and they are kind of preferable.

3. **Dummy Variables and Scaling:**
   The dummy variables were created for multiple categorical variables. For numeric variables 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:**
   RFE was done to attain the top 15 relevant variables.

Variables selected by RFE:

1. 'Do Not Email',
2. 'Total Time Spent on Website',
3. 'Lead Origin_Lead Add Form',
4. 'Last Activity_Email Bounced',
5. 'Last Activity_SMS Sent',
6. 'What is your current occupation_Unknown',
7. 'What is your current occupation_Working Professional',
8. 'Tags_Busy',
9. 'Tags_Closed by Horizzon',
10. 'Tags_Lost to EINS',
11. 'Tags_Ringing',
12. 'Tags_Will revert after reading the email',
13. 'Tags_switched off',
14. 'Last Notable Activity_Modified',
15. 'Last Notable Activity_Olark Chat Conversation'

Then, models were created with selected features and checked the summary. The model looked fine all the p values of the features are under 0.05, the all features are a good fit for the model.

6. **Model Evaluation:**
Adding threshold of >0.5 as "1" and <0.5 as "0". A confusion matrix was made. All the checks looked excellent for the train model. Later on the optimum cut-off value (using the ROC curve) was used to find the accuracy, sensitivity, and specificity which came to be around 90% each.

7. **Prediction:**
The prediction was done on the test data frame and with an optimum cut off as 0.4 with accuracy, sensitivity, and specificity of more than 90%.

- Final predicted test model evaluation
- accuracy = 93.14 %
- sensitivity = 92.97 %
- specificity = 93.26 %

The model predicted for the performance form exceptionally on all the evaluation metrics is more than 90%.
Keeping the above-mentioned points in mind X education can increase all potential buyers to change their mind and buy their courses.