SUMMARY

This analysis is carried out for X Education in an effort to attract more business professionals to their courses. We learned a lot from the fundamental data about how potential customers use the website, including how long they stay there, how they got there, and the conversion rate.

The procedures are as follows:

1. **Importing and understanding the data:** As per the information provided, there are 3 integer columns, 4 float valur columns and 30 object type columns which include categorical variables.
2. **Cleaning the data:** Most of the data was clean, save for a few null values, and the option choose had to be changed to a null value because it didn't provide much information. To avoid losing too much data, a few of the null values were changed to "not provided." Nevertheless, they were later taken out while manufacturing dummies. The elements were altered to "India," "Outside India," and "not provided" because there were a lot of people from India and a small number from elsewhere.
3. **EDA:** A brief EDA was performed to assess the state of our data. It was discovered that several of the categorical variables' components were unnecessary. The numerical figures are accurate, and no anomalies were discovered.
4. **Dummy Variables:** After creating the dummy variables, the dummies that had the phrase "not provided" were later deleted. We utilised the MinMaxScaler to scale numerical numbers.
5. **Train-Test Split:** For train and test data, the split was done at 70% and 30%, respectively.
6. **Model Building:** First, the top 15 pertinent factors were determined by RFE. Later, based on the VIF values and p-value, the remaining variables were manually deleted).
7. **Evaluation of the model:** A confusion matrix was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 96% each.
8. **Prediction:** Using an optimal cutoff of 0.3 and a prediction accuracy, sensitivity, and specificity of 96%, the test data frame was used.
9. **Precision – Recall:** As the model's MSE is close to 0, our model accurately predicts the variance between the test and train datasets.

The factors that affected potential purchasers the most were discovered to:

1. The overall amount of time spent on the Website.

2. The overall volume of visits.

3. If the lead came from one of the following:

1. Google

2. Direct traffic

3. Organic search

4. The Welingak website

4. The most recent activity was when:

SMS exchanges and conversations on Olark

5. If the format of the lead add is the lead origin.

6. If they are a working professional at the time.

Conclusion:

With these in mind, X Education can succeed since they have a very good probability of persuading nearly all prospective customers to change their minds and purchase their courses.

- The final model has a sensitivity of 0.932 indicating that it can correctly predict 93% of the positive conversions.
- The final model has a precision of 0.66, which means that 66% of the leads that it predicted as "hot" are actually true hot leads.
- The final model has an Accuracy of 0.80
- The final model has a Precision: 0.66
- Features used in final model are
    - ['Do Not Email', 'Lead Origin_Lead Add Form', 'Lead Source_Welingak Website', 'Last Activity_SMS Sent', 'Tags_Busy', 'Tags_Closed by Horizzon', 'Tags_Lost to EINS', 'Tags_Ringing', 'Tags_Will revert after reading the email', 'Tags_switched off', 'Lead Quality_Not Sure', 'Lead Quality_Worst', 'Last Notable Activity_Modified', 'Last Notable Activity_Olark Chat Conversation']