Summary

X Education is using this analysis to entice more business people to enrol in its courses. We learned a lot from the basic data on how potential customers use the site, how long they stay there, how they got there, and the conversion rate.
The following are the steps:

Data cleaning:

With the notable exception of a few null numbers, the data was primarily clean. However, the choice had to be changed to a null value because it wouldn't have offered much useful information. To avoid losing too much data, some of the null values were changed to "not provided"; nonetheless, they were later eliminated when creating dummies. There were numerous people from India and few from elsewhere, therefore the elements were altered to "India,""Outside of India," and "not provided."

2. Explicit Data Analysis (EDA):

We conducted a quick EDA to assess the state of our data. It was found that some of the descriptive analysis's elements weren't necessary. The numerical numbers are accurate, and no contradictions were discovered.

3. Imaginary variables:

After the dummy variables were made, the ones that said "not provided" were taken out. We used the MinMax-Scaler to scale numerical numbers.
4.. Train-Test split:

For train and test data, the split was done at 70% and 30%, respectively.

5. Model construction:

First, RFE determined the top 20 pertinent parameters. The remaining variables were then manually removed according to their VIF values and p-values (the variables with VIF 5 and p-value 0.05 were kept).

6. The outcome was a perplexing matrix. Later, the accuracy, sensitivity, and specificity were calculated using the ROC curve, and they all turned out to be about 80% each.

7. Prediction:

Predictions on the test data frame reached 80% accuracy, sensitivity, and specificity using an ideal threshold of 0.35.

8. Accuracy - Memory

For the test data frame, a cut off of 0.41 was discovered using this method, with a precision and recall of roughly 72% and 80%, respectively.
According to research, the following people (in descending order) influenced prospective customers the most:
1. TotalVisits
2. Google as the Lead Source
3. Direct Lead Source Traffic
4. What is your present employment status? Unemployed
5. Total Length of Visits
6. Most Recent Significant Activity_Modified
7. Lead Source: Natural Search
8. Last Chat Conversation with Olark
9. Email Opened as the Latest Notable Activity
10.Lead Origin_Lead Add Form, page 10
11.The final noteworthy Olark chat conversation
12. What do you do for a living now? Working professional
13. Website Lead Source Welingak 13.
14. Do Not Email_Yes 14.
15. The Website Page Visited Most Recently
16. Sites with Lead Sources and Referrals
17. What are your present duties as a student