

Summary :

This analysis is being conducted for X Education to find ways to bring more industry professionals into the course. The basic data provided gave us a lot of information about how potential customers got to the website, how long they stayed, how they got to the website, and their conversion rates. rice field.

Here are the steps to use:

1. **Cleaning data:** The data was partially clean except for a few null values and didn't give much information, so I had to replace the select options with null values. Changed some null values to "not specified" to avoid losing a lot of data. Although they were later removed in the manufacture of dummies. There were many products from India and few products from overseas, so we changed to products from India and from outside India and did not provide them.

2. **EDA :** A simple EDA was performed to verify the sanity of the data. We found many items in the categorical variables to be irrelevant. The numbers look good and no outliers were found. A dummy variable was created and later the dummy containing the "unprovided" element was removed. For numbers, we used MinMaxScaler.

3. **Test Train Split :** The split was 70% and 30% for training and test data respectively.

4. **Modeling:** First, we ran an RFE to arrive at the 20 most important relevant variables. The remaining variables were then manually removed according to VIF and p-values (variables with $VIF < 5$ and $p\text{-value} < 0.05$ were retained). A confusion matrix was created. The optimal cut-off values (using ROC curves) were then used to determine precision, sensitivity, and specificity of approximately 85%, respectively. Predictions were made using the test data frame and an optimal cutoff of 0.3 with an accuracy, sensitivity and specificity of 85%. Eighth. Precision–Recall method was also used for revalidation and found a cutoff of 0.41 with ~82% precision and ~82% recall on the test data frame.

The below variables play a vital role in choosing the potential buyers :

- What matters most to you in choosing a course.
- Tags will revert after reading the email.
- Last notable activity modified.
- Tags – Other Tags & Ringing

- Last Activity – SMS Sent & Olark Chat Conversation
- Occupation – Working Professional.
- Lead Origin – Lead Import

Highlights :

1. The conversion rate improved after model building to 82.09%.
2. We are successful in targeting the lead conversion rate to be even higher than 80% as demanded by the CEO of X Education.
3. Keeping these in mind X Education can flourish as they have a high chance to get almost all the potential buyers to convince to buy their courses.