

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. Cleaning Data:** The data was partially clean except for a few null values and the option select had to be replaced with a null value since it did not give us much information. Few of the null values were changed to 'not provided' to not lose much data. Although they were later removed while making dummies.
- 2. EDA:** A quick EDA was done to check the condition of our data. It was found that a lot of elements in the categorical variables were irrelevant. The numeric values seems good and no outliers were found.
- 3. Dummy Variables:** The dummy variables were created and later on the dummies with 'not provided' elements were removed.
- 4. Test-Train 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.
- 6. Model Evaluation:** A confusion matrix was made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy of 80%, sensitivity of 70% and specificity which came to be around 88% each.
- 7. Prediction:** Prediction was done on the test data frame and with an optimum cut off as 0.35 with accuracy of 80%, sensitivity of 81% and specificity of 80%.
- 8. Precision – Recall:** This method was also used to recheck and a cut off of 0.41 was found with Accuracy around 81%, Precision around 74% and Recall around 77% on the test data frame.

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

1. Total Visits
2. Total Time Spent on Website
3. Page Views Per Visit

Then, you must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies.

Hold question-answer sessions with leads to extract the right information you need about them. Make further inquiries and appointments with the leads to determine their intention and mentality to join online courses.