

SUMMARY

This analysis is done for X Education to find ways to get more industry professionals to join their courses. Data provided gave a lot of information about the potential customers activity, their visiting the website, the time spend on the site, how they reached and what has been the conversion rate meaning how many of them got enrolled in any of the courses.

Below is the step-by-step journey of the analysis:

- 1. Data Cleaning**

This step involved cleaning the data which mainly included removal/treatment of null values, any outlier removals, categorizing columns to gain better insights etc.

- 2. EDA**

Exploratory Data Analysis (EDA) is required to check the data all in all. With it, it was found that the numerical data looked good with no potential outliers found. Some elements in the categorical variables seemed irrelevant.

- 3. Train-Test Split**

Training-Testing split is done for train and test data respectively.

- 4. Model Building**

RFE is done for selection of top variables and later remaining variables were removed depending on p-values.

- 5. Model Evaluation**

ROC curve is built to find accuracy, sensitivity and specificity.

- 6. Prediction**

Prediction was then done on the test data using the model built through train data.

OBSERVATIONS:

It was found that the variables that mattered the most in the potential buyers are as below:

1. The total time spent on website
2. Total number of visits
3. When the lead source was
 - a. Google
 - b. Direct Traffic
 - c. Organic search
4. When the last activity was SMS
5. When their current occupation is working professional

Keeping all this in mind the X Education can flourish given that they have high chances to get almost all potential customers to change their mind and buy their courses.