

# SUMMARY

The analysis is done for X Education and to find ways to get more professionals to join their courses. The data provided gives us a lot of information about the potential customers visiting the site, how they reached the site, how much time they spend there and the conversion rate.

Followings are the steps used:

Step 1	Cleaning Data	The data was partially clean except for few null values and the option Select had to be replaced with a null value since it did give us much information. Few of the null values were changed to 'not provided' so as to not lose much data. Although they were later removed while making dummies. Since there were many from India and few from outside, the elements were changed to 'India', 'outside India' and 'not provided'.
Step 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 numerical values seem good and no outliers were found.
Step 3	Dummy Variables	Dummy variables were created and later on the dummies with 'not provided' elements were removed. For numeric values we used the MinMaxScaler.
Step 4	Train -Test Split	The split was done at 70% and 30% for train and test data respectively
Step 5	Model Building	Firstly, RFE was done to attain the top 15 relevant variables. Later, rest of the variables were removed manually depending on the VIF values and p-values (the variables with $VIF < 5$ and $p\text{-value} < 0.05$ were kept)
Step 6	Model Evaluation	A confusion matrix was made. Later on the optimum cut off value (using ROC curve) was used to find the accuracy, sensitivity and specificity which came to be around 80% each.
Step 7	Prediction	Prediction was done on the test data frame and with an optimum cut off as 0.45 with accuracy, sensitivity and specificity of 80%
Step 8	Precision - Recall	This method was also used to recheck and a cut off of 0.41 was found with precision around 80% and recall around 74% on the test data frame. And for train data precision is around 78% and recall around 78%.
<ul style="list-style-type: none"> <li>✓ The lead and customer funnel is the right approach to check the outcome. As we can see, there were a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom.</li> <li>✓ In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating etc.) in order to get a higher lead conversion.</li> <li>✓ For the instance of checking the conversion rate, firstly sort out the best prospects from the leads you have generated. 'TotalVisits', 'Total Time Spent on Website', 'Page Views Per Visit' which contribute most towards the probability of a lead getting converted. Then, we must keep a list of leads handy so that you can inform them about new courses, services, job offers and future higher studies.</li> <li>✓ Monitor each lead carefully so that you can tailor the information you send to them. Carefully provide job offerings, information or courses that suits best according to the interest of the leads.</li> </ul>		

- ✓ A proper plan to chart the needs of each lead will go a long way to capture the leads as prospects. Focus on converted leads. 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