

# **Summary**

This analysis is carried out for X Education in an effort to attract more business professionals to their courses. We learned a lot from the basic data provided about how potential clients visit the site, how long they stay there, how they got there, and rate of conversion.

The procedures are as follows:

1. **Cleaning the data:** Most of the data was clean, save for a few null values, and the option choose had to be changed to a null value because it didn't provide much information. To avoid losing too much data, a few of the null values were changed to "not provided." Nevertheless, they were later taken out while making dummies.
2. **EDA:** To assess the state of our data, a brief EDA was conducted. It was discovered that several of the categorical variables' components were unnecessary. The numerical figures are accurate, and no irregularities were discovered.
3. **Train-Test split:** The split for train and test data was done at 70% and 30%, respectively.
4. **Dummy Variables:** After creating the dummy variables, the dummies that had the phrase "not provided" were subsequently removed. We utilized the MinMaxScaler to scale numerical numbers.
5. **Model Construction:** RFE was first used to identify the top 15 pertinent variables. Later, based on the VIF values and p-value, the remaining variables were manually eliminated (the variables with VIF 5 and p-value 0.05 were retained).
6. **Precision - Recall:** This method was also utilized to retest, and on the test data frame, a cut off of 0.41 was discovered with Precision around 73% and recall around 75%.
7. **Modal Evaluation:** Confusion matrix for the model was created. Later, the accuracy, sensitivity, and specificity were determined using the ROC curve, and they all came to be about 80% each.
8. **Prediction:** Using an optimal cutoff of 0.35 and a prediction accuracy, sensitivity, and specificity of 80%, the test data frame was used.

The lead and customer funnel are 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.

- 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.
- 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.
- 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.
- 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.

Keeping these in mind, X Education can thrive because they have a very high chance of convincing almost all potential buyers to change their minds and purchase their courses.