Summary

- The Problem with which X Education approached us was helping them with increasing their lead conversion rate which was only 30% currently. The CEO anticipated that out of all the leads available, at least 80% would convert.

- The company already advertises its courses on websites and search engines, so many people visit the company website to explore the courses, fill out a contact form, or watch some videos. In order to do so, they must provide contact information like an email address or phone number. They become our leads as a result of this procedure.

- The company's ask from us is to create a model that will assign a score to every lead which will help them to prioritize important and convertible leads over the leads which are less likely to be converted. To do so we have been provided with a dataset having 9000 records having various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc with a target variable 'Converted' which tells us whether these leads are converted or not.

- We first started out the data understanding i.e., EDA. 20% of the columns were having more that 40% of null values so these columns will make 0 value add in our model hence were dropped. We also have handled missing values for categorical variable by mode imputation method

- In data preparation for model building, we created dummy values for categorical variables, created train test split, we also used RFE to reduce number of variables. Total 4 models were built; Final model was stable with P-value < 0.05 and VIF < 5.

- The final model, logm4, had 12 variables, and it was used to make predictions on the train and test sets. The model evaluation involved creating a confusion matrix and selecting a cut-off point of 0.345 based on accuracy, sensitivity, and specificity plot. The lead score was assigned to the train data using the 0.345 cut-off, and the top three features were **Lead Source_Welingak Website, Lead Source_Reference, and Current_occupation_Working Professional.**

- Based on our analysis we recommended to aggressively target **working professionals, aggressively advertise on Welingak, promote referral program more** by increasing referral incentives.

This project helped us to increase our skills on Data cleansing, EDA, Data preparation for model building, model evaluation, finding insights for business problems.