**Summary**

- An education company named X Education sells online courses to industry professionals.
- professionals interested in the courses land on their website and browse for courses on any given day.
- From there they are directed to the X Education form, where they provide information by filling the form.
- This form information will be stored in the X Education database. This information contains the internal parameters and professional information which will be used to determine potential leads.
- The sales team will use that data and make calls to professionals. either professional will be converted into lead or not.
- Based on interactions with professionals, the sales team will update the sales parameters against the original form data.
- Data could be divided into original captured form data and sales team generated data.
- we must use originally captured form data for model training because the prediction will get the same
- The model will predict the potential leads which will be used by the sales team to make calls which will save time and increase the high converted rate.

- **The following are the steps used:**
    - Business understanding
    - Data understanding
    - Data importing:
    - Data Cleaning:
    - Handle the “Select” level that is present in many of the categorical variables.
    - Drop columns that are having a high percentage of missing values. Check all the columns before dropping them.
    - Check the number of unique categories in each categorical column. Here you may need to do something.
    - For the columns with less percentage of missing, use some imputation technique.
    - Finally check the percentage of rows retained in the data cleaning process. Prepare the data for Model Building:
    - EDA
    - Create dummies for all categorical columns. • Perform train-test split. • Perform scaling.
    - **Model Building**
        - Use techniques like RFE to perform variable selection.
        - Build a Logistic Regression model with good sensitivity.
        - Check p-value and VIF.
        - Find the optimal probability cutoff.
        - Check the model performance over the test data.
        - Generate the score variable.
    - **Model Evaluation**
        - All performance metrics, f score Making Predictions on the Test Set

- **Final Observation:** Let us compare the values obtained for Train & Test:
- Train Data: Accuracy: 79.19% Sensitivity: 71.69% Specificity: 83.79% Test Data: Accuracy: 79.63% Sensitivity: 76.33% Specificity: 81.62% The Model seems to predict the Conversion Rate very well and we should be able to give the CEO confidence in making good calls based on this model