# A BRIEF SUMMARY

As a part of the Lead Scoring case study, we have been presented with the details how the company X Education pursues customer leads from various sources and tries to convert them to potential customers.

The current conversion rate is quite low at 30%. So, we have been tasked to analyse the data and come up with a model which can make predictions to the order to 80% Lead conversion.

So, a brief summary of how it was done is as below.

- **Model Building process:**

1. Identifying the columns based on Data Dictionary.
2. Elimination invalid / redundant columns.
3. Removing records with > 70% missing data.
4. Imputing few columns with missing data.
5. Identifying the potential data columns which can factor in for accurate prediction.
6. Identifying the relationship and distribution of column data using graphs.
7. Removing the outliers in numerical variables.
8. Plotting heat map to see the correlations.
9. Unknown values that transformed into columns are dropped from the Dummy columns. The data is then split into training and test data in ratio of 70:30.
10. The training data is fed into a Generalized Logistic Model (GLM). The ineffective variables are eliminated using RFE and VIF.
11. After a bit of fine-tuning, the final model is built and then some metrics like Sensitivity, Specificity etc. were recorded.
12. Finally, ROC Curve, Precision-Recall Trade-off graphs are plotted and the model is finalized.

- ## A Few Insights:

1. Lead scoring case study has been done using Logistic Regression model to meet the constraints as per business requirements.
2. There are a lot of leads in the initial stage but only a few of them are converted into paying customers.
3. The leads are joined course for Better Career Prospects, most of having Specialization from Finance Management. Leads from HR, Finance & marketing management specializations are high probability to convert.
4. Most of leads current occupation is Unemployed, which means the Team should focus more on unemployed leads.