**SUMMARY**

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. The basic data provided gave us a lot of information about how the potential customers visit the sites, how they reach the site, conversion rate and the time they spend there.

For this we used supervised learning technique to generate a regression model which would help us to get the lead score. Logistic regression is the one used to build the model. There are 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.

The following are the steps used :-

1) Cleaning data

2) EDA

3) Creation of dummy variables

4) Train-Test Split

5) Model building

6) Model evaluation

7) Prediction

8) Precision –Recall

All the categorical variables are converted into dummy variables and the numerical variables are scaled. Then after train and testing of the model is done before building the model. Based on P value and VIF variables are eliminated until they are optimum. Prediction is based on accuracy, sensitivity, specificity, precision, recall etc. To identify the predictive power ROC curve is been used and using these optimum cut-off point is determined.

It was found that the top variables in our model that contributes towards the result are Total time spend on websites, lead source with elements Google, total visits and page views per visit.