**Summary for Lead Score Assignment**

**Problem Statement :-**

X Education is an online platform where user register for online courses. The company markets its product through search engine and several websites. Now the company has a very low lead conversion of 30%. So to improve upon this lead conversion, the company wants to identify ‘Hot Leads’ which can be converted easily. The CEO has given a target lead conversion of 80%. So the company essentially wants to know :-

1. What are the 3 variables which contribute the most in terms of leads getting converted
2. Top 3 categorical/dummy variables which should be focused the most in order to increase the probability of lead conversion

**How did we solve the problem?**

At first, we have cleaned the data by removing unwanted Sales generated and Unique data columns, NULL value greater than 45%, removed some highly skewed data and combined the data of skewed column into a single value and imputed the numerical columns.

Then we did EDA, Train-Test and Data Modelling using RFE and manually analyse through GLM and VIF to find the optimal set of data with P values less than 0.05 and VIF score of less than 5.

Then we did Model Valuation after plotting the ROC curve and found out Accuracy, Sensitivity, Specificity, Precision, Recall and F1 Score on Train data set

**Conclusion** :-

The sensitivity score was 77.8% for train and 77.5% for Test data set which means this much percentage of Leads can be considered as hot leads and is convertible.