**Lead Scoring Model for X Education**

**Introduction**

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. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

**Business Goals**

The company wishes to identify the most potential leads, also known as ‘Hot Leads’. The company needs a model wherein a lead score is assigned to each of the leads such that the customer with higher lead score have a higher conversion chance and the customer with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark number for the lead conversion rate i.e 80%.

**Overall Approach**

* Data cleaning and imputing missing values
* Exploratory Data Analysis - Univariate, Bivariate and Multivariate Analysis
* Feature Scaling and Dummy Variable Creation
* Logistic Regression Model Building
* Model Evaluation - Sensitivity, Specificity, Precision, Recall
* Conclusion and Recommendation

**Libraries Used**

* Numpy
* Pandas
* Matplotlib
* Seaborn
* StatsModel
* Sklearn