# 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 company markets its courses on several websites and search engines like Google. 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%.

Now, although X Education gets a lot of leads, its lead conversion rate is very poor.

For the above problem I have built a logistic regression model for lead scoring.

The code includes:

Data Preparation: The dataset is split into features (X) and the target variable (y), then scaled using StandardScaler.

Feature Selection: Recursive Feature Elimination (RFE) is employed to select the most important features for the model.

Model Building: Logistic regression model is instantiated and fitted using the selected features.

Model Evaluation: Various evaluation metrics are calculated including VIF (Variance Inflation Factor), ROC curve, optimal cutoff point, confusion matrix, accuracy, sensitivity, specificity, precision, and F1 score.

Visualization: ROC curve and confusion matrix are plotted for visual assessment of model performance.

Model Optimization: Some feature elimination is done based on VIF to potentially improve model performance.

Final Model Evaluation: Metrics are recalculated and plotted again to assess the impact of feature elimination.