**Problem Statement**

You are being provided a dataset containing characteristics of a customer/individual and the Delinquency (SeriousDlqin2yrs) which is the Dependent Variable

The challenge is to build a model to classify the customers into Delinquent/Non-Delinquent (SeriousDlqin2yrs = 1 or 0)

Language: Preferably R but Python or SAS would be fine

**Evaluation Criteria:**

* Exploratory data analysis – variable characteristics, skewness, etc [Not for all but for important and significant variables] – accompanied by visualization [Not more than 5 charts]
* Missing Value Imputation
* Feature Engineering
* Prediction:
  + Using Logistic Regression to classify the customers
  + Evaluation Metrices – Choose the best evaluation metrices which explain the model comprehensively
* Advanced Techniques to Improve the results
  + SVM / Random Forest / Boosting – Techniques to improve the model
* The final model will need to run on the Test data provided separately.

**Submission Format:**

* You can submit a Rmd file, or Jupyter Notebook, or a PPT presentation. You can also submit an excel sheet with all the results neatly placed in the file
* Submitting code is mandatory, this will be used for evaluation
* Do not put unnecessary information or redundant graphs and numbers in the result. **All in all, It should not take more than 20 minutes for you to take us through the model and results**
* **The final scoring on test data should have two columns – ID (test data ID), Probability and Class(1/0 for that ID)**

**Data Dictionary : Attached separately**