APPLICANT CREDIBILITY PREDICTION FOR LOAN APPROVAL

## PROBLEM STATEMENT:

The goal is to create a classifier that can predict whether an individual should be approved for loan or not

the tasks involved are the following:

i) Download and pre-process the data.

ii) Gain insights about the data, engineer features (If necessary) and make it ready

for the machine learning algorithms.

iii) Train the data and test for evaluation metrics.

iv) Fine tune the hyperparameters.

v) Make predictions on the testing set.

In [finance](https://en.wikipedia.org/wiki/Finance), a **loan** is the lending of [money](https://en.wikipedia.org/wiki/Money) by one or more individuals, organizations, or other entities to other individuals, organizations, etc. The recipient (i.e. the borrower) incurs a debt, and is usually liable to pay [interest](https://en.wikipedia.org/wiki/Interest) on that debt until it is repaid, and also to repay the principal amount borrowed. The whole process of ascertaining if a borrower would pay back loans might be tedious hence the need to automate the procedure.

### Problem Description:

A finance company wants to automate the loan eligibility process based on customer detail provided while filling its online application form.

These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others.

To automate this process, they have given a problem to identify the customers who are eligible for a loan amount so that they can specifically target these customers.

The output schema required is as follows (in a csv file):

Loan\_ID, Loan\_Status

Loan Amount term had only 10 unique values in the form of the number of years. Therefore, it was label encoded.

EDA of Property Area showed that individuals living in semi-urban areas were more likely to get their loans approved. Therefore, most weight was given to the applicants coming from semiurban areas.

Finally, the last feature i.e., Co-applicant Income. This feature provided some pretty interesting insights as mentioned in the EDA part. Because the increase in co-applicant income did not play a big role in increasing the chances of approval of an applicant combined with the fact that 44% of the values were exactly 0, we can engineer a new feature and call it “co-applicant income exists” which is 1 if the co-applicant income is non-zero, 0 otherwise.