**Loan Predictions**

# Problem Description:

Microfinance, also called micro credit, is a type of banking service that is provided to unemployed or low-income individuals or groups who otherwise have no other access to financial services. Given the nature of lending, microfinance is a high risk industry where the loan default rates can be high. Worried about their profits, micro-finance companies have decided to use machine learning techniques (since they read about it in the news) to help them decide how much loan to be given to each lead. In their search to solve the problem, they reach out to the top data scientist in the field, **you**, to help them solve their problem.

Your mission, should you choose to accept it, involves predicting the amount of loan to be given to leads/borrowers. The following information has been given to you by the microfinance company:

**leads\_interview.csv**: contains list of people who have applied for a loan to the finance company along with their attributes. Leads who have been given loans become borrowers and those borrowers who have repaid their loans are listed in **repayments\_interview.csv**. The company has a unique model for loan recovery. For each transaction that the borrower does (as listed in repayments\_interview.csv), the company recover a fixed percentage of the transaction. The percentage recovered is also listed in the csv file. *Cust code* is the unique identifier for each borrower.

Below are the description for the variables.

Volume\_of\_Txn\_Month4 – total transaction amount for 4th month; previous months details.

Approved: 1 - loan disbursed; only these people should be present in repayment file, ‘0’ – not disbursed

Count\_of\_Txn\_Month5 – number of counts of transactions for 5th month; previous months details.

Unique\_Count\_of\_Txn\_Month5 - Unique transaction counts (specified to banking domain)

Vintage (In Months) - Association between customer and finance corporation in months.

*You need to solve the problem by predicting the amount of loan to be disbursed to each new lead and provide detailed documentation on the steps done to solve the problem. Please include code snippets, references in the documentation and graphs to help explain your solution. Your documentation should be self-contained and must be comprehensive and understandable by a non-machine learning engineer. Please send your commented code and the accompanying document in a zip file.*

***Note****: you can program in any language you are comfortable with*