Project IDEA 1 : LENDING LOANS:

- **The Problem being solved:**

As an investor it is very important to invest in the people who have the capability to return their money back. This avoids the investors to not incur a loss. As an investor you would want to invest in people who showed a profile of having a high probability of paying you back. I will try to create a model that will help predict this.

- **Clients being focussed here:**

Any type of investors be it Private Banks, Public Banks, Investment Banks, Private Equity firms, Angel Investors need that high assurance that they are going to get their money back and at least they don't go to loss especially if they are investing huge amount of money to their clients. Based on my prediction model using Machine Learning where I used a real life public data, the investors can can find out whether a borrower can repay back the money he's been provided rather than just using excel sheets.

- **Data that I am going to use for my Model:**

For this project I will be exploring publicly available data from LendingClub.com. Lending Club connects people who need money (borrowers) with people who have money (investors). This is the link :https://www.kaggle.com/wendykan/lending-club-loan-data. I will be using the loan data. This data contains complete loan data for all loans issued through the time period stated, including the current loan status (Current, Charged, Fully Paid, etc.) and latest payment information from 2007 to 2011.

- **Brief Approach for solving the problem:**

1. The data provided is a complete real life loan data that has to be cleaned properly as it has many redundant and unimportant columns. I need to eliminate those.
2. I will look for the missing NA entries and remove it from the data.
3. Using Matplotlib and Seaborn modules in Python, I will do some EDA using Visualization. I will do some manipulations with the data if I have to.
4. I will separate independent variables (which will be input data) from the dependent variable ('loan_status'). So finally the input variables will contain columns like Purpose, Interest rate, Installment and the Output will contain the column "loan_status" which will have 0 and 1 values. '0' for the people who did not pay their loans (i.e - loan_status = current, charged) and 1 for those who did pay the loans (i.e - loan_status = Fully Paid)..
5. I will separate the data into training set (80% of the data) and test set (20% of the data).
6. I will use a Machine Learning model like Decision Trees and Random Forest, Support Vector Machine and try to predict the people who paid their loans and also those who don't.
7. A confusion matrix and standard classification matrix will cross check the predictions.