**Loan Data Analysis of dataset from LendingClub.com**

**Dataset Description**

This dataset is about loan data from LendingClub.com that publish for commercial in Kaggle. LendingClub.com is platform that connects people who need money (borrower) and who have money (investor). **As an investor you would want to invest in people who showed a profile of having a high probability of paying you back**. The analysis is done to determine good criteria of the borrower that have high probability of paying back the loan to the investors.

**Problem Statement**

Analyse the LendingClub.com data from 2007-2010 and try to find out some good criteria of the borrower that have high probability of paying back the loan.

**Tools & Libraries**

• Python • Jupyter Notebook • Pandas • Numpy • Seaborn • Matplotlib • Plotly & Cufflinks

**Data Description**

The dataset contains the following Columns:

* **credit.policy:** **1** if the customer meets the credit underwriting criteria of LendingClub.com, and 0 otherwise
* **purpose:** The purpose of the loan (takes values "credit\_card", "debt\_consolidation", "educational", "major\_purchase", "small\_business", and "all\_other")
* **int.rate:** The interest rate of the loan, as a proportion (a rate of 11% would be stored as 0.11). Borrowers judged by LendingClub.com to be more risky are assigned higher interest rates.
* **instalment:** The monthly instalments owned by the borrower if the loan is funded.
* **log.annual.inc:** The natural log of the self-reported annual income of the borrower.
* **dti:** The debt-to-income ratio of the borrower (amount of debt divided by annual income)
* **fico:** The FICO credit score of the borrower.
* **days.with.cr.line:** The number of days the borrower has had a credit line.
* **revol.bal:** The borrower's revolving balance (amount unpaid at the end of the credit card billing cycle).
* **revol.util:** The borrower's revolving line utilization rate (the amount of the credit line used relative to total credit available).
* **inq.last.6mths:** The borrower's number of inquiries by creditors in the last 6 months.
* **delinq.2yrs:** The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
* **pub.rec:** The borrower's number of derogatory public records (bankruptcy filings, tax liens, or judgments).
* **not.fully.paid :** The borrower can't paid their loan.

**Data Cleaning**

I made the following change in the following variables:

• Renamed the all columns in the dataset. In dataset, all the columns are having (.), it is replaced with the (\_)

**EDA**

By analysing the different trends of the data and below are a few highlights of the analysis;

1. The parameters are having skewness, mostly right skewed
2. Most of the outliers are at the right side
3. The interest rate is between 0.06 (the lower risk) and 0.2164 (the highest risk)
4. Lowest instalment is 15.67 and the highest is 940.14
5. Debt divide by annual income mean is 12.607 point
6. The borrower days line's to pay their loan is usually between 2820 - 5730 days
7. The borrower's remaining payment usually between 3187-18250
8. Some of borrower use more than available total credit amount more than 100 % i.e. 119%
9. High positive correlation is found between FICO & Credit policy
10. Interest rate shows high positive correlation with the borrower's revolving line utilization rate (revol\_util)
11. Instalment and annual borrowers income is highly correlated (positive)
12. Derogatory public records of borrowers (pub\_rec) shows negative correlation with credit policy & FICO
13. From 9578 borrower, 7710 borrower (80%) had approval for criteria policy from Lending Club and 1868 borrower (20%) had not approved for criteria policy from Lending Club
14. 1533 customer i.e. around 16% have not paid their loan and 8045 customer i.e. around 84% have paid their loan
15. Maximum loans (around 41%) are taken for pay off other liabilities and consumer debts i.e. debt consolidation
16. Somewhat positive correlation is seen between instalment and the annual income of the borrower
17. Many borrower who have paid the loan have a good record (don't have public derogatory records)
18. Out of 9578 borrower, 5695 showed a profile of having a high probability of paying you back the loan amount, so the investor can invest their money on the list of above borrower's
19. From the above analysis, we can say, good criteria of the borrower that have high probability of paying back the loan is that minimal fico score must be 627 and better if it is around 722 (median)