# Dataset info

For the final project will be using the lending loan club dataset, located over [here](https://www.kaggle.com/wordsforthewise/lending-club). Data will be filtered to use loans issued from 2015 to 2018, to minimize the dataset size. Only approved loans dataset will be in scope for this project. Currently I am working with start up with which deals with personal loans, so wanted to select related dataset, hence I have chosen this dataset.

The Lending Club is a peer-to-peer lending service (it lends money to customers by matching lenders to borrowers), based in the United States. This company enables borrowers to create loan listings on its website by supplying details about themselves and the loans that they would like to request. Based on the borrower’s credit score, credit history, desired loan amount and the borrower’s debt-to-income ratio, Lending Club determines whether the borrower is credit worthy and approved the loan.

Dataset contains various attributes based on which the credit lending decision takes place, this includes Debt to income ratio, annual income, home ownership type (Rented, Mortgage, Owned), employment details. Dataset also contains geographical information like state and zip code. It also has issued loan details like funded Loan Amount, issued month and year, loan Term (36 or 60 months) and interest rate. Post loan has been issued, loan performance needs to be tracked, and this dataset also contains those details loan Status (Current, Paid Off, Charged Off, Delinquent), loan Grades and delinquency details.

# Qualitative Research Question

1. Build a model which will predict whether a loan will default.  Qualitative response variable would be loanDefaulted, values would be "Yes" or "No".
2. Build a model which will predict whether a loan will get charged off. In this case a qualitative response variable would be loanChargedOff, values would be "Yes" or "No"

# Quantitative Research Question

1. For given set of credit attribute for the borrower like fico, debt to income ratio, annual income, home ownership, applicant type, average current balance of all accounts, employment length, identify the what would be approved/funded loan amount and associated interest rate. In this case the response variable would approvedLoanAmount.

# Unsupervised Research questions

1. Identify association between loan funded amount, annual income, dti, fico, loan grade and loan status.
2. For all the charged off loans identify the correlation or association between annual income, dti, last available fico and fico recorded during the loan application.
3. For all delinquent loans identify correlations with credit attributes fico, debt to income ratio, annual income.