Lending Club Case Study

## **Problem Statement:**

Assume you are working as a consumer finance company which provides various types of loans to urban customers. When a loan application is received by company then company must decide whether the loan application will be accepted or rejected. Although this decision has two types of risks:

* Approving the loan applications of the customers who are likely to be a defaulter, leads to a financial loss.
* Rejecting the loan application of the customers who are not likely to be a defaulter, leads to a business loss.

So you are given a dataset and with the help of the dataset you must determine that what are all the factors that are responsible for the person to be a defaulter?

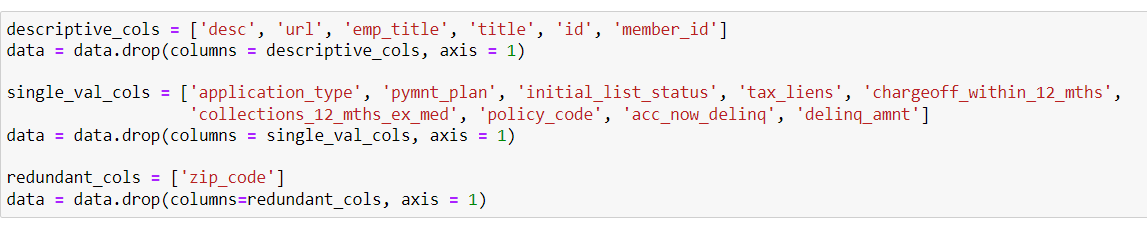
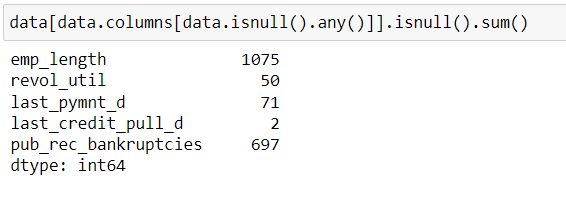
## **Dataset Description:**

We have a dataset name as ‘loan.csv’ which contain records of all the loan applicants. After importing the dataset in jupyter notebook we have 39717 rows and 111 columns.

## **Approach Description:**

### **Dataset Cleaning:**

To clean the whole dataset we performed the following steps:

* We identify all the columns which have more than 50% of empty records and we deleted those columns.
* Some columns have descriptive values, some has only single value and some are redundant columns. We identify those columns and delete all of them. 
* We identify all the rows which has missing value in a column, and we deleted those rows. 
* We remove outliers from our dataset column ‘*annual\_inc*’ which denotes the annual income of the loan applicants.
* We fix the value and datatype of our dataset column ‘*int\_rate*’ which denotes the loan interest rate.

After this data cleaning activity, we have 37075 rows and 38 columns for our data analysis.

### **Variables Identification:**

We select some of the variables from the dataset and divide them in two categories,

1. Categorical Variables
2. Numeric or Quantitative Variables

Following image shows the various categorical and numerical variables



## **Business Results from Univariate Analysis:**

* Nearly 14% of loan applicants has ‘charged off’ or ‘defaulter’ status on their loan application.
* Around 17K+ loan applications have a loan purpose as ‘debt consolidation’.
* Nearly 73% of loan applicants select ’36 months’ of loan term plan.
* Around 8K+ loan applicants have 10+ years of employment length.
* Around 17K+ loan applicants reside in rented house.
* From state CA, the highest number of loan applications received.
* For majority of loan applicants, the loan amount varies in between 5K to 15K.
* For majority of loan applicants, the loan interest rate varies in between 8% to 15%.
* For majority of loan applicants, their annual income varies in between 40K to 80K.

## **Business Results from Derived Matrices:**

* **Derive Year from Loan Issue Date:** After deriving year from loan issue date, we can say that the loan applications are increase over the year from 2007 to 2011.
* **Derive Annual Income Categories from Annual Income:** We divided the annual income of loan applicants into the following income categories
  + Below 25K – Has nearly 2K loan applicants
  + Range between 25 to 50K – Has nearly 12K loan applicants
  + Range between 50 to 75K – Has nearly 11K loan applicants
  + Range between 75 to 100K – Has nearly 6K loan applicants
  + Above 100K – Has nearly 5K+ loan applicants

## **Business Results from Segmented Univariate Analysis:**

* People who earn between 0-25K per year are most likely to be defaulter or charged off.
* People who earn 100K+ per year are less likely to defaulter or charged off.
* People who are taking loan for small business establishment purpose are most likely to be defaulter or charged off.
* People who are taking loan for their wedding purpose are less likely to be defaulter or charged off.
* Loan grade F and G are most likely to be defaulter or charged off.
* Loan grade A is less likely to be a defaulter or charged off.
* People who belong from state NV, TN and AK are most likely to be defaulter or charged off.
* People who belong from state IN, ID and IA are less likely to defaulter or charged off.

## **Business Results from Bivariate Analysis:**

* Loan applicants who are charged off are drawing less salary than the loan applicants who fully paid their loans.
* Loan applicants who are charged off are paying loan with higher interest rate than the loan applicants who fully paid their loans.
* Loan applicants who are currently paying their loans have highest loan interest rate and monthly installment among other applicants.
* Majority of loan amount is taken either for small businesses or for house.
* 60 months loan term amount is greater than 36 months loan term amount.
* Loan applicants who are taking highest amount of loan have more than 10 years of service or employment length.
* Mortgage people are taking highest amount of loan.
* People having 10+ years of service length, have mostly recorded their bankruptcy in any public forum.

## **Conclusion & Advice to the Business:**

* Business should accept the loan application of a person having 100K+ annual salary as he is least likely to be a defaulter.
* Business should control to give loans to the people who are earning below 25K+ salary.
* Business should examine more on loan grades before approving the loan application, as grade A & B have less defaulters while F & G have more.
* Business should control their loan in state NV, TN and AK.
* Business should accept more loan application from IN, ID and IA.
* Business should be careful while approving the business loans and house loans.
* Business should avoid to give loans to bankrupt.