

# Lending Club Case Study

Presentation  
by  
Saravanakumar PERUMAL

# Introduction

- ▶ This project work is for a consumer finance company which specializes in lending various types of loans to urban customers.
- ▶ In this project work I have demonstrated through various EDA ( Exploratory Data Analysis), to understand how consumer attributes and loan attributes influence the tendency of loan default.
- ▶ When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

# Business Understanding

- ▶ When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.
- ▶ This finance company should decide whether to approve the loan or not based on the loan application and available data with the company.- Two types of risks are associated with the bank's decision:
  - 1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - 2) If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
- ▶ The data given from the company contains information about past loan applicants and whether they 'defaulted' or not.
- ▶ The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

# Business Understanding

- ▶ When a person applies for a loan, there are two types of decisions that could be taken by the company:

- ▶ **Loan accepted:**

If the company approves the loan, there are 3 possible scenarios

- a) **Fully paid:**

Applicant has fully paid the loan (the principal and the interest rate).

- b) **Current:**

Applicant is in the process of paying the instalments,

i.e. the tenure of the loan is not yet completed.

These candidates are not labelled as 'defaulted'.

- c) **Charged-off:**

Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan.

# Business Understanding

## ► Loan rejected:

The company had rejected the loan

(because the candidate does not meet their requirements etc.).  
Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

## ► Credit loss :

It is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

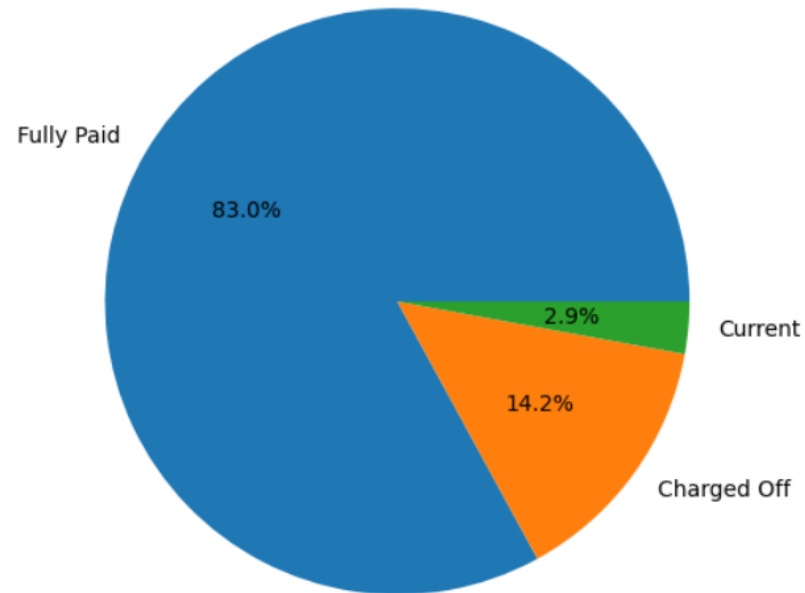
In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

# Targeted Outcome

- ▶ The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- ▶ The company can utilize this knowledge for its portfolio and risk assessment.
- ▶ 'Loan.csv' dataset is used in this project

# Analysis of Lending Club Data:

## ► Univariate Analysis on Loan Status - Pie Chart

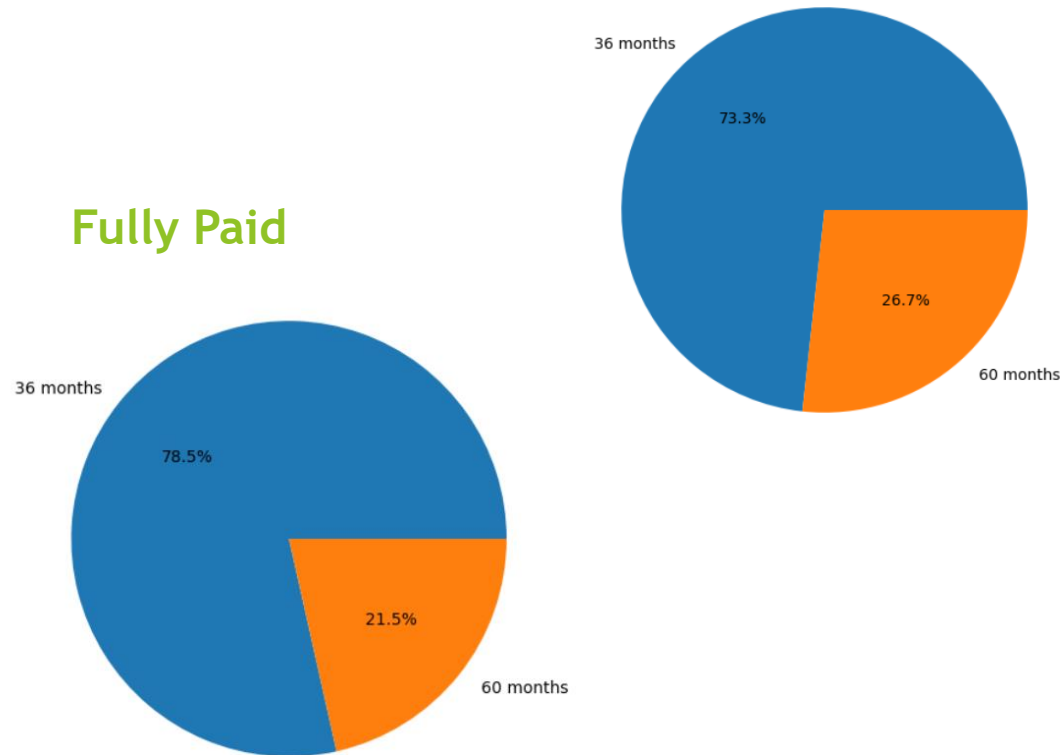


Over all observation on Loan Status :

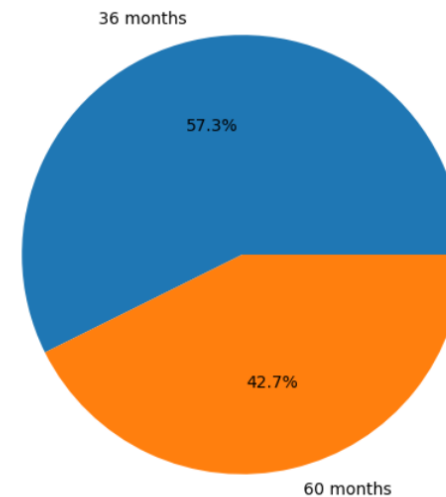
- 83 % of the loans are Fully Paid
- 14.2 % of the loans are Charged Off / Defaulted
- 2.9 % of the loans are currently Active

## ► Univariate Analysis on Loan Term - Pie Chart

### Fully Paid, Current, Charged Off



### Charged Off / Defaulted



## ► Important Observations :

1) Loan term of 60 months (5 Years) are higher chances of default (Mean of 42.7 % Defaulted compared to 21.5 % Fully paid).

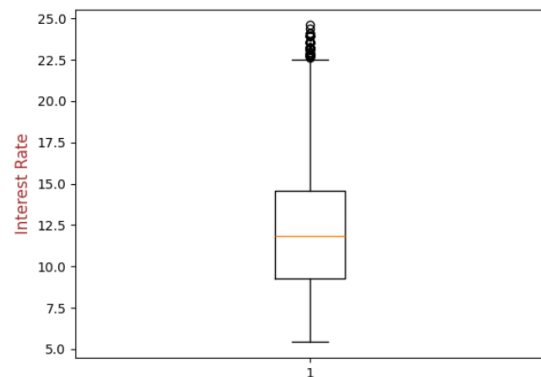
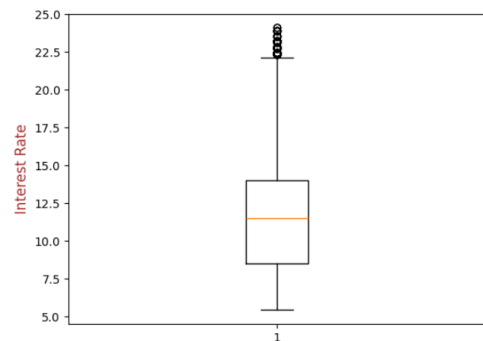
2) Lending company should scrutinize 5 Years loan with stringent norms.



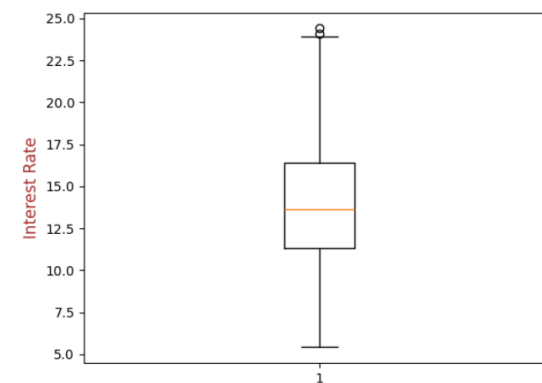
## ► Univariate Analysis on Interest Rate - Box Plot

### Fully Paid, Current, Charged Off

#### Fully Paid



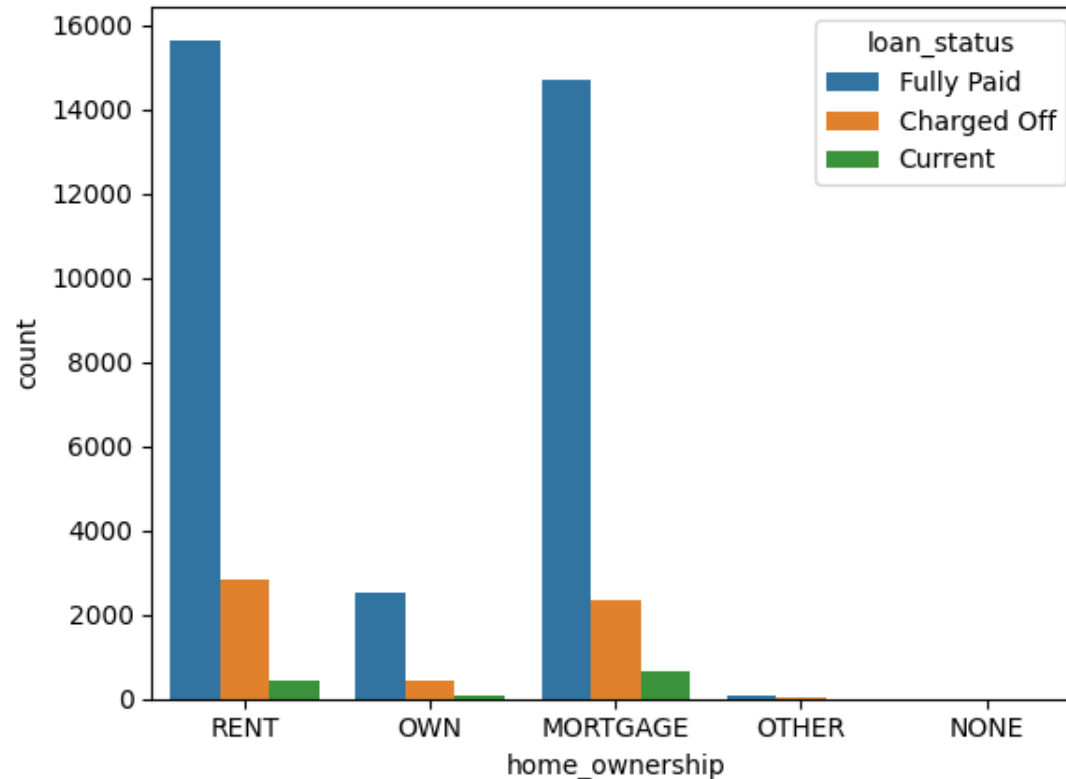
#### Charged Off / Defaulted



### ► Important Observations :

- 1) Loans with high interest rates have higher chances of default (Mean of 13.8 % interest rate for defaulted compared to 11.6 % for fully paid).
- 2) Lending company should analyze better for those customers who are ready to accept higher interest rates which loan applications have higher chances of default loans.

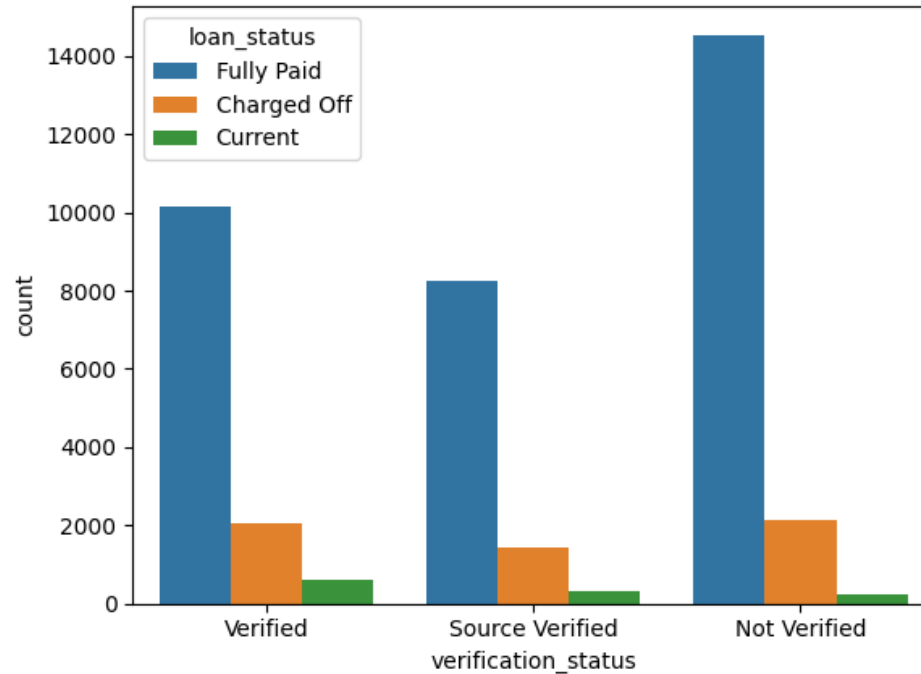
## ► Bivariate Analysis on Home Ownership Vs Loan Status



### ► Important Observations :

- 1) Owner of a house are less likely to default compared to those who are in Rent and Mortgage.
- 2) Lending company should be more cautious while approving the loans from those home status belongs to Rent (or) Mortgage, because those loans are more likely to default compared to customers those who own a house while applying for the loan.

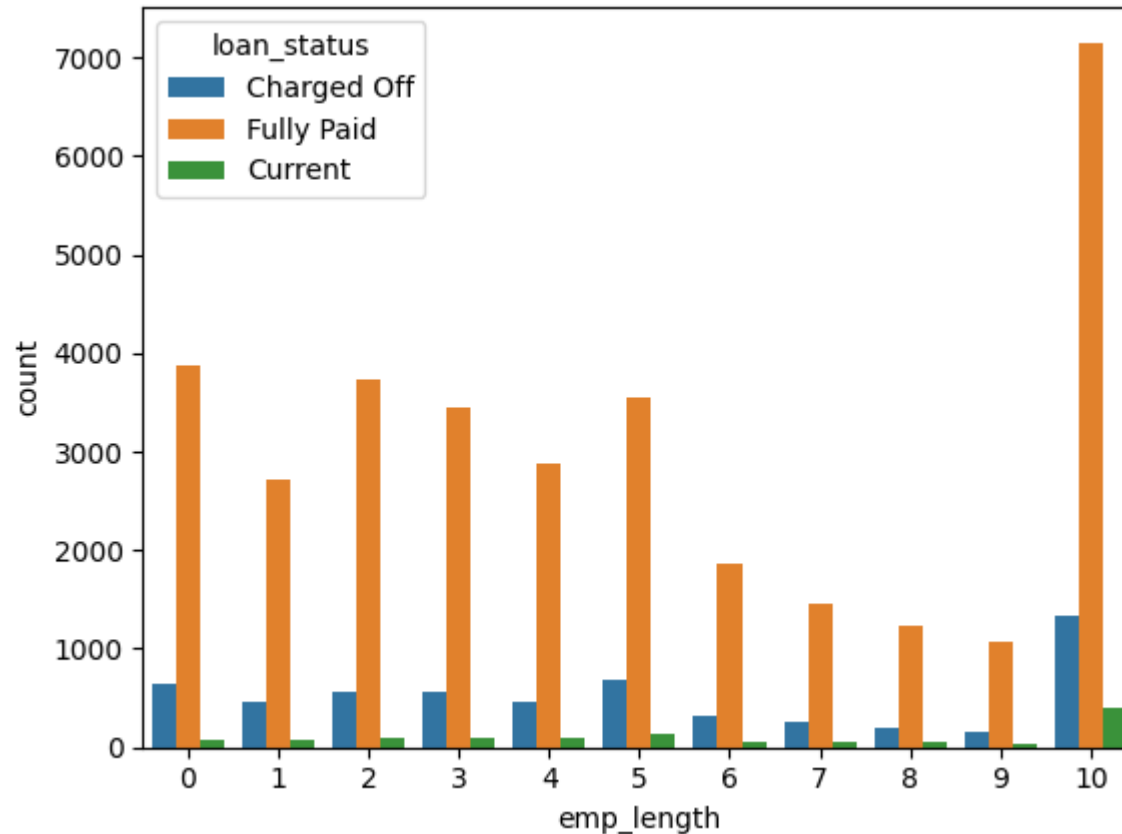
## ► Bivariate Analysis on Verification Status Vs Loan Status



### ► Important Observations :

- 1) If the income is verified by Lending Company then those loans are less likely defaulted compared to income sources are not verified and only mere verified category
- 2) Lending company should verify the source of income before approving the loan application

## ► Bivariate Analysis on Employment Length Vs Loan Status

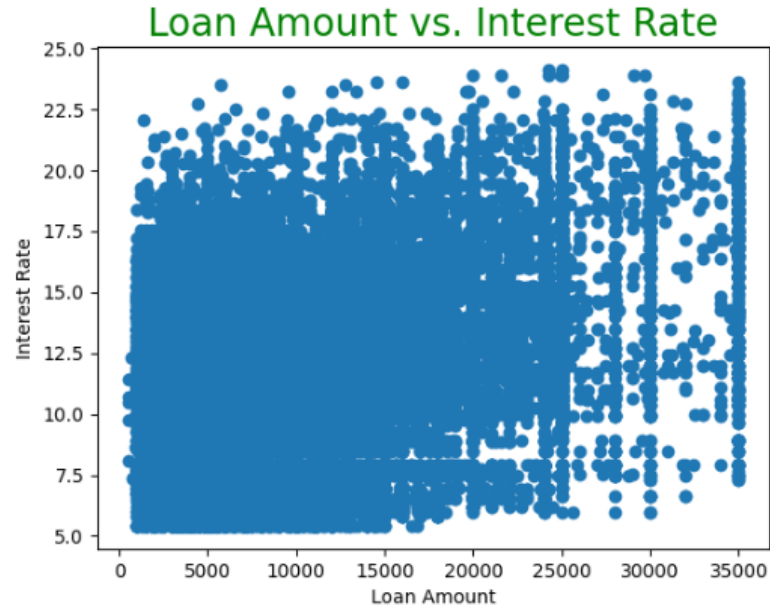


### ► Important Observations :

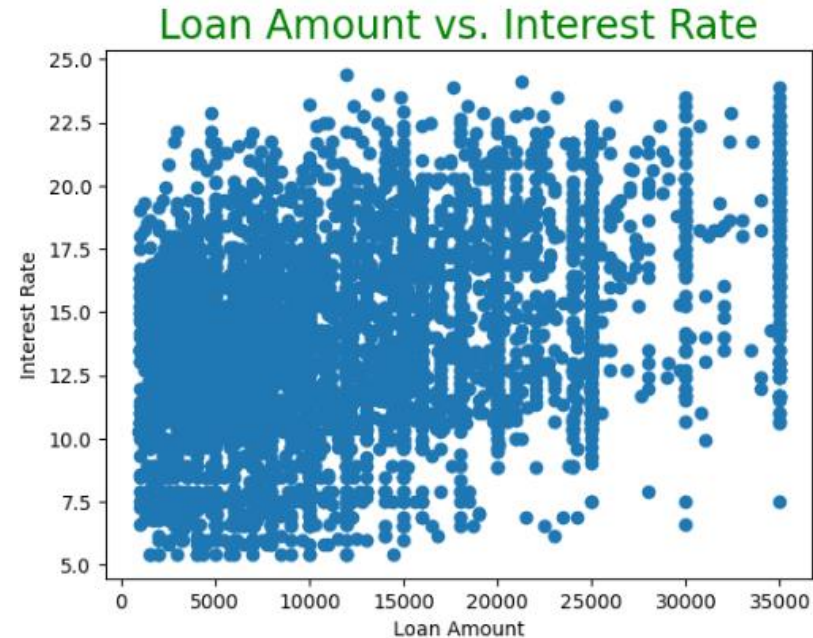
- 1) If the employment length is less than 5 years then those loan applications are higher chances of default
- 2) Lending company should verify the source of income to repay the loan when the employee applies a loan with less than 5 years of experience

## ► Bivariate Analysis -Scatter Plot - Loan Amount Vs Interest Rate

### Fully Paid



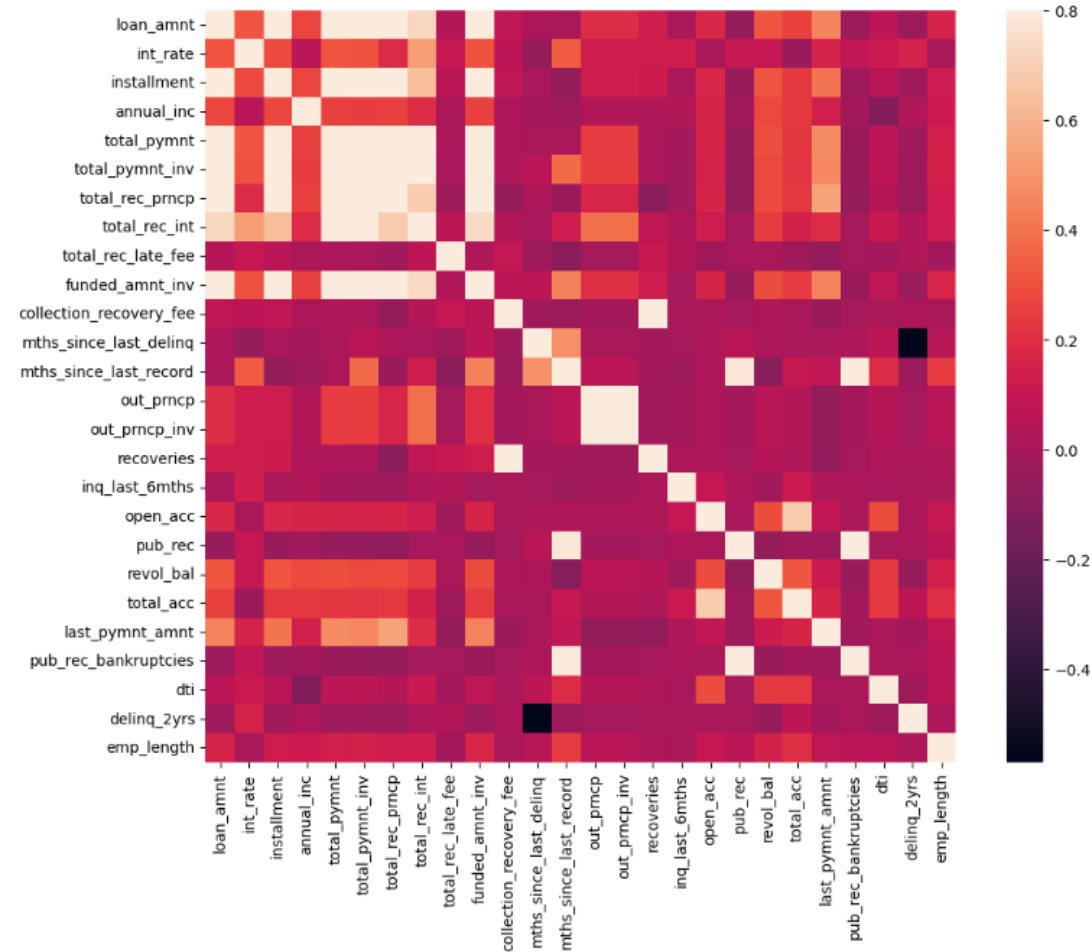
### Charged Off / Defaulted



## ► Important Observations from this scatter plot :

- 1) The density of the loans is high with the lower interest rates for the fully paid loans.
- 2) In defaulted case the density of the loans are high in higher interest loan region which shows that if loan application accepted at higher interest rates can be more likely defaulted
- 3) So the lending company should verify for when customer is ready to accept higher interest rates.

## ► Multi Variate Analysis - Heat Map



### ► Important Observations :

- 1) **pub\_rec\_bankruptcies** - Correlated with almost with all features and hence it is better to avoid lending on those applications with of public record bankruptcies.
- 2) **pub\_rec** - Number of derogatory public records is highly correlated with almost with all features and hence it is better to avoid lending on those applications with derogatory public records.