

# LENDING CLUB CASE STUDY

## SUBMISSION

Group Members :

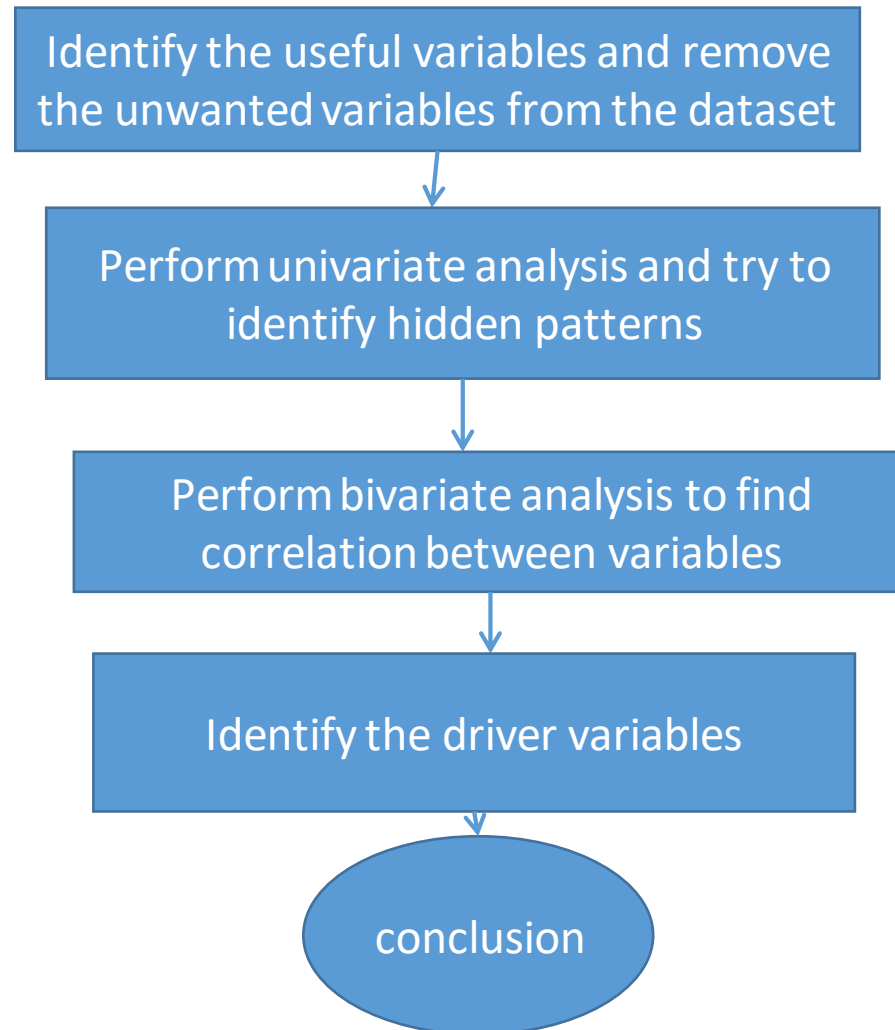
Vinay Kumar Gundappa

Yogesh Bhatt ( Group Facilitator )

## Abstract:

- The company under study is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.
- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss 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'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- In other words, 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 utilise this knowledge for its portfolio and risk assessment.
- Identification of such risky applicants using Exploratory Data Analysis is the aim of this case study.

# Problem solving methodology



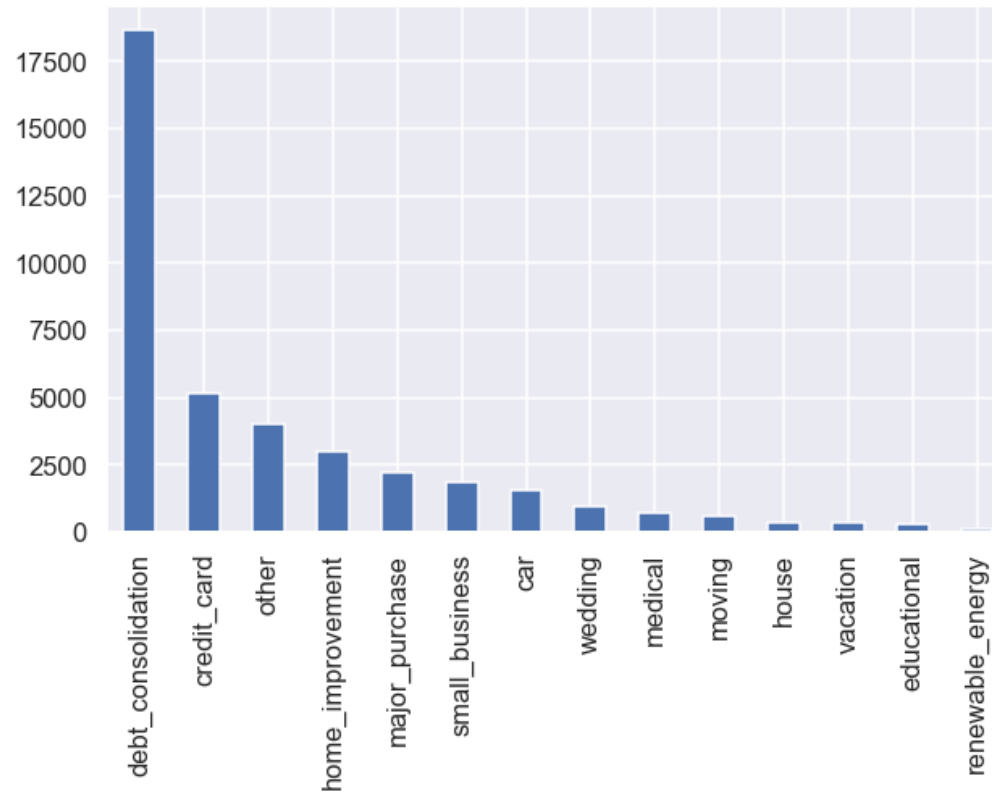
# Analysis

- **Identifying the useful variables :**

- On analysing the data and performing some research following variables were picked for EDA
  - Purpose : A category provided by the borrower for the loan request.
  - loan\_amnt : The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
  - annual\_inc : The annual income of the borrower
  - emp\_length : Employment length of the borrower
  - pub\_rec\_bankruptcies : the number of public record bankruptcies
  - term : the number of payments on the loan. Values are in months and can be either 36 or 60.
  - dti : a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
  - grade : LC assigned loan grade
  - loan\_status : Status of the loan – 'Fully Paid', 'Current' or 'Charged Off'

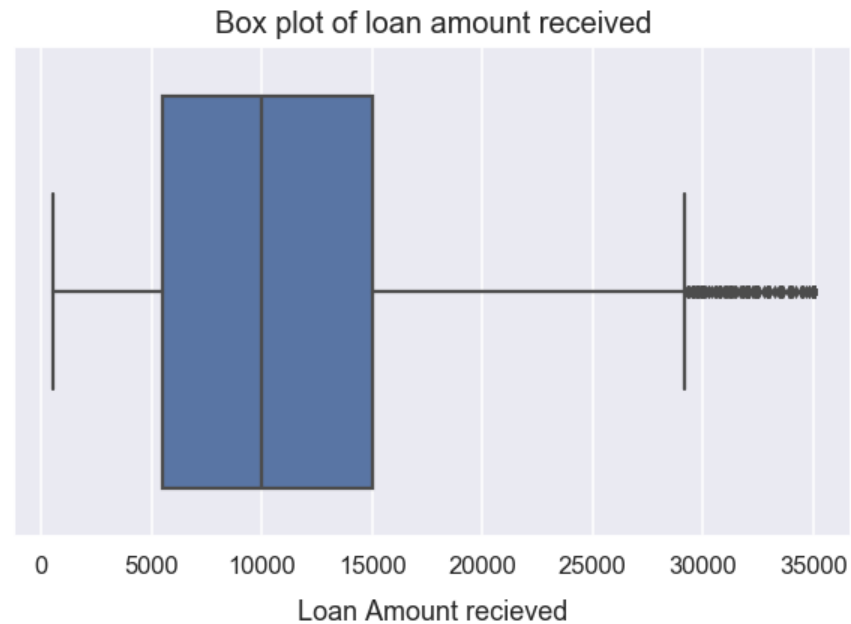
# Analysis

- Univariate Analysis [ variable = purpose ]
  - The plot of purpose against loan\_amount reveals that the highest number of loans are offered to people for debt consolidation . This does not sounds a profitable thing to do from business perspective as offering loan to people already in debt is risky.



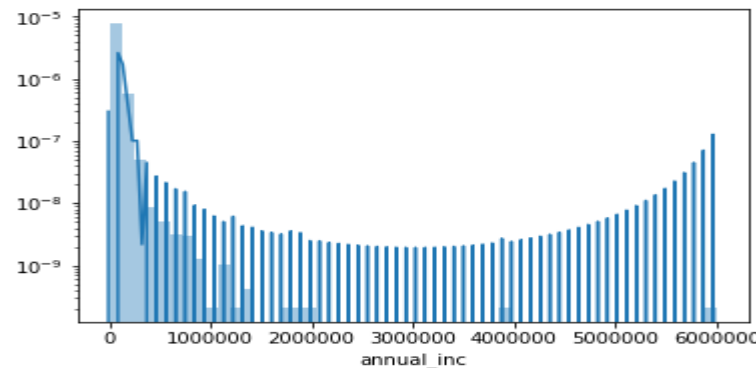
# Analysis

- Univariate Analysis [ variable = loan\_amnt]
  - The boxplot clearly reveals that most of the loans fall between 5000 to 15000. However presence of a considerable number of outliers beyond the maximum limit shows that the risk of offering huge loans is not addressed

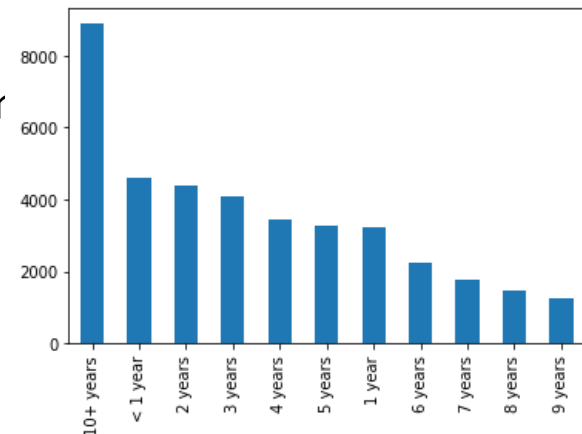


# Analysis

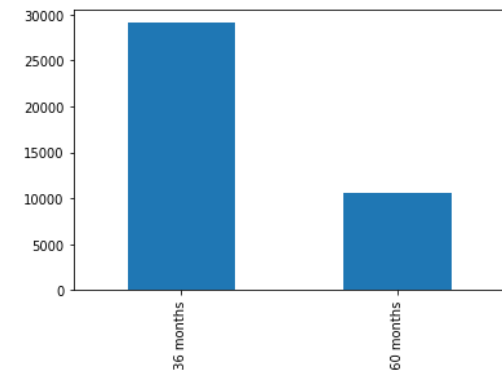
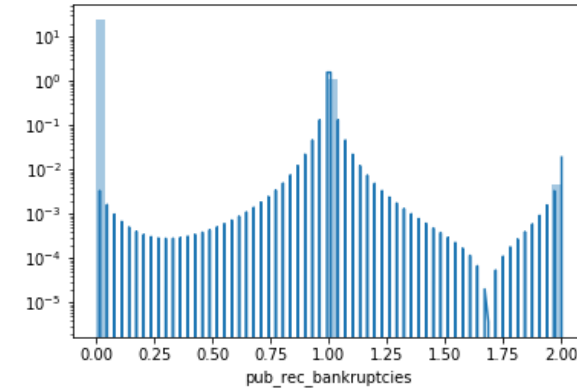
- Univariate Analysis [ variable = annual\_inc ]
  - The plot shows that the maximum number of loans have been given to the people with negligible annual income . This is a risk factor which needs to be kept in mind while predicting loan repayment



- Univariate Analysis [ variable = emp\_length ]
  - The plot shows that people with 10+ years of experience are given the maximum Number of loans. An interesting discovery is that the next to follow are the borrowers with less than 1 year of experience . They are more likely to default as they haven't reached a stable position in their career . Thus this is a risk factor



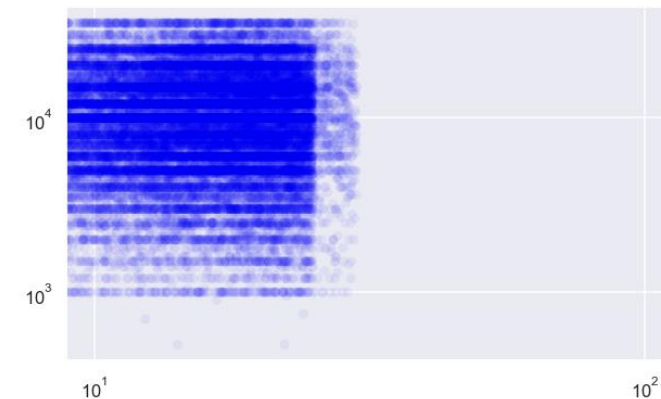
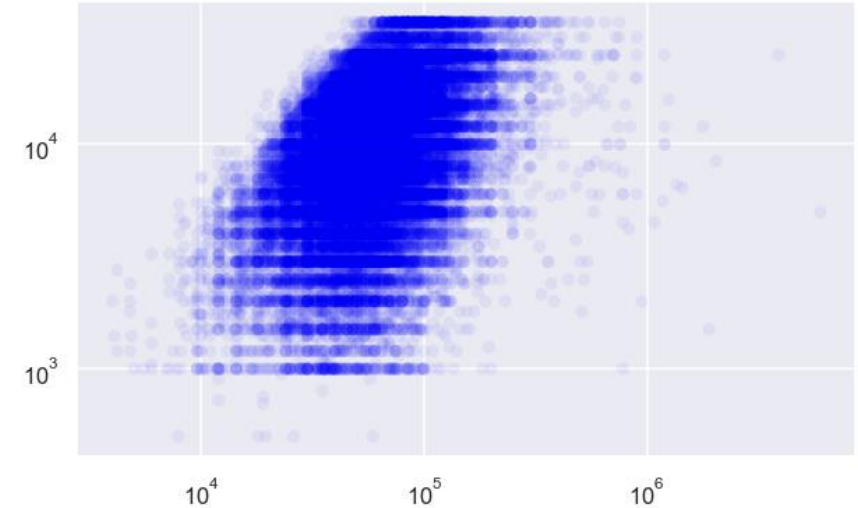
- Univariate Analysis [variable = pub\_rec\_bankruptcies]
  - The plot shows that the maximum number of loans have been given to the people with zero record of bankruptcies which is a good sign .However a pattern that raises eyebrows is the spike at 1 ( i.e high number of loans being offered to people with 1 bankruptcy record in the past ) . This is a risk factor and needs to be addressed by the organisation
- Univariate Analysis [variable = term]
  - The plot shows that the maximum number of loans have been given to the people for 36 months .There are comparatively lesser number of loans for long term which is a good step by the organisation as short term loans are more profitable due to higher interest rates .



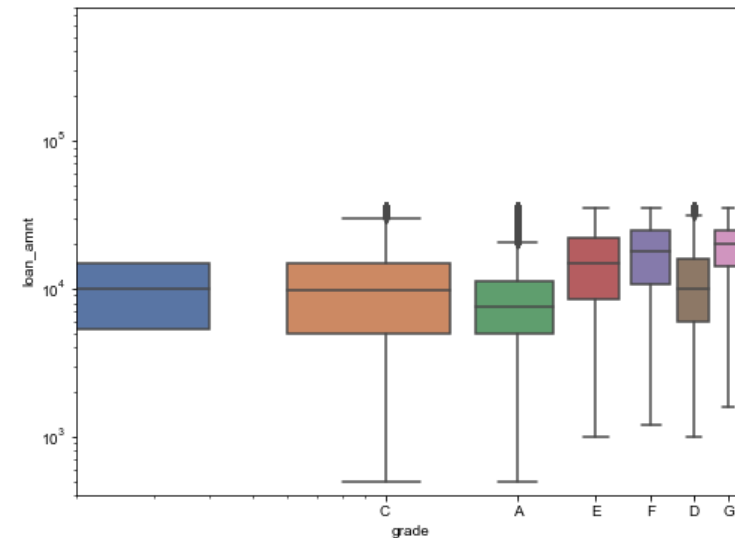


# Analysis

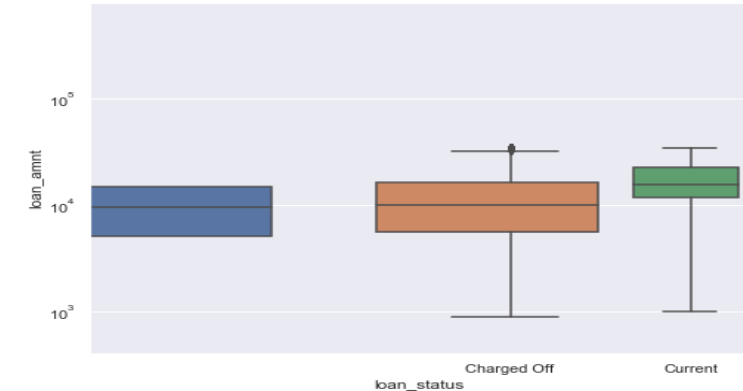
- Bivariate Analysis [loan\_amnt v/s annual\_inc]
  - The plot shows there is a positive correlation between loan amount and annual income of the borrower
- Bivariate Analysis [loan\_amnt v/s dti]
  - The plot shows there is no definite correlation between loan amount and dti. Thus we can not conclude anything by this variable. Not identified as a driver variable



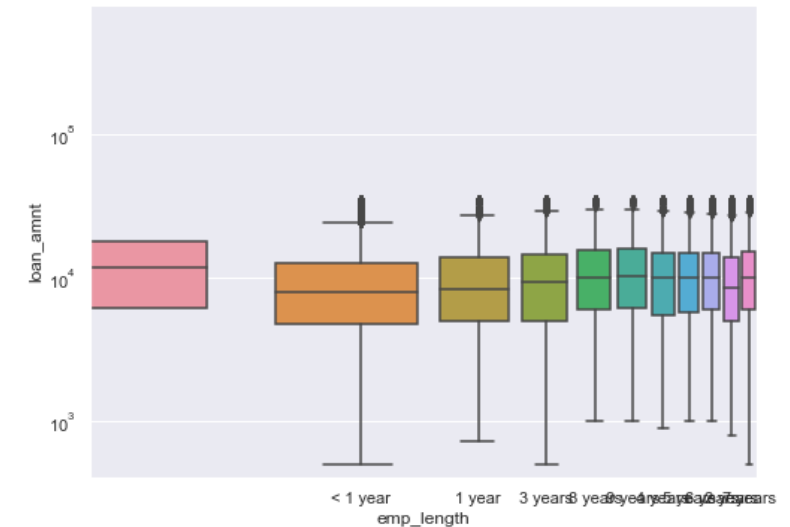
- Bivariate Analysis [loan\_amnt v/s grade]
  - Conclusions :
    1. the median loan amount for grade B and C is almost same and the interquartile range is also almost equal
    2. The median loan amount for grade A is the least and also the interquartile range is the smallest
    3. The grades E and F are offered loans in the same ranges but at times the loan offered exceeds the loan offered to the customers with better grades
    4. The lower quartile of grade G is very close to the upper quartile of grade B and C . That means the customers with poor grades as G are also being offered loans which are close to the highest loan being offered to the customers with better grades



- Bivariate Analysis [loan\_amnt v/s loan\_status]
  - The plot shows that the spread of the box is almost same for Charged Off and Fully paid loans . So there is no correlation between the loan\_amnt and loan\_status

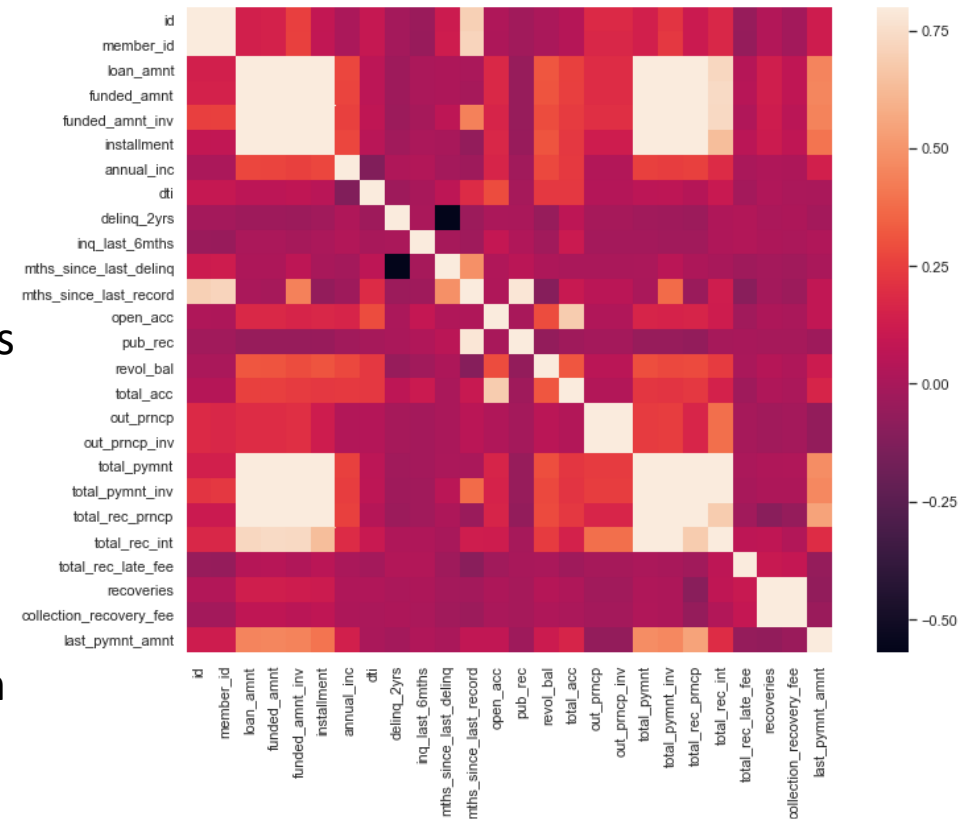


- Bivariate Analysis [loan\_amnt v/s emp\_length]
  - The box plot shows that the spread of loan offered goes on increasing as the employment length increase . The presence of outliers indicates that there is no strict restriction on the amount of loan being offered based on the employment length alone . This needs to be addressed and identifies emp\_length as a potential driving factor

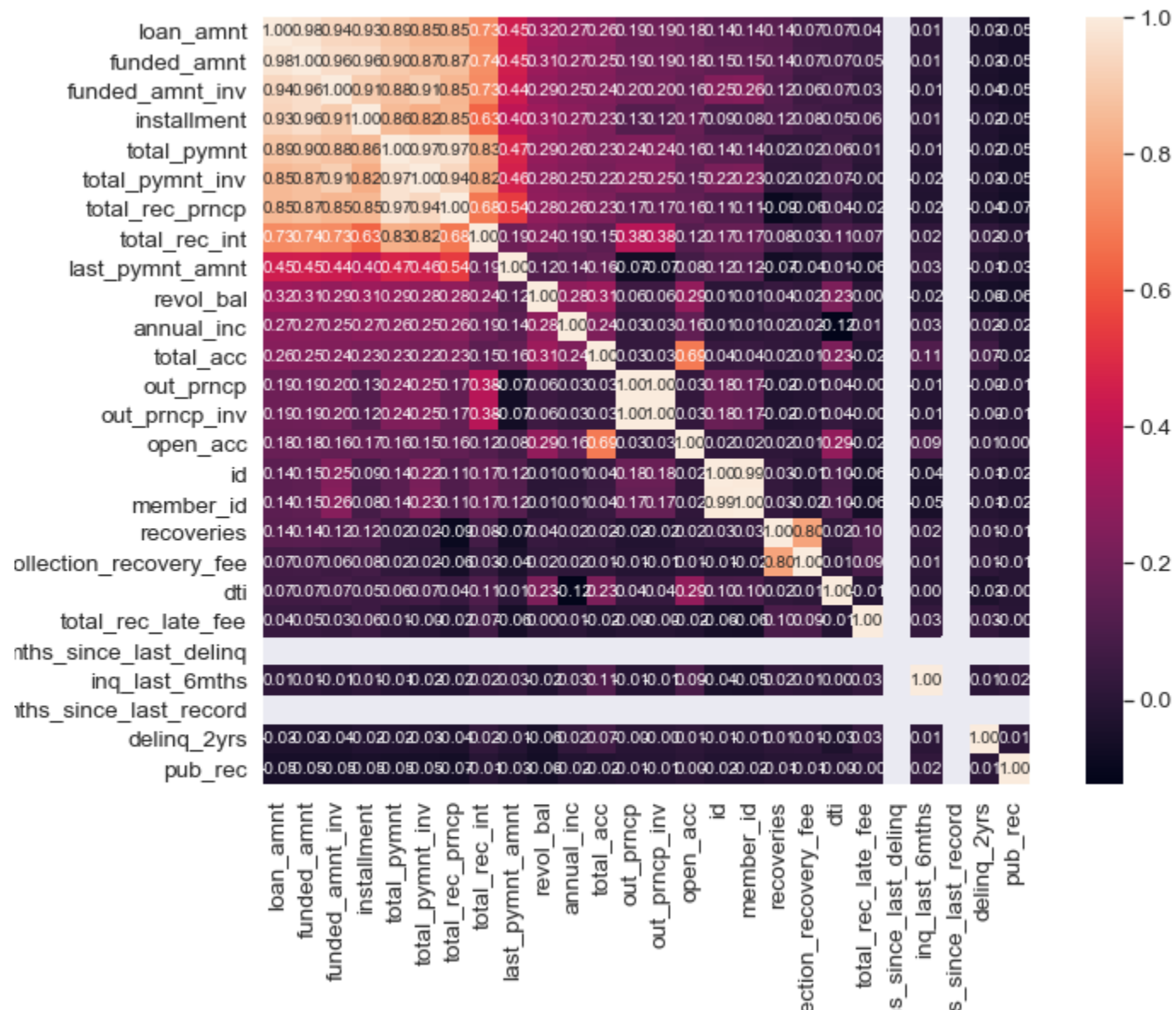


## • Conclusions from Correlation Matrix

1. annual income is positively correlated with loan\_amnt, funded\_amnt, instalment
2. open\_acc, which is the number of credit lines opened in borrowers credit profile is negatively correlated with recoveries. Thus the customers with high open\_acc should be checked
3. pub\_rec shows a similar trend as open\_acc, thus borrowers with a high value of pub\_rec should be watched.
4. total\_pymnt\_inv shows a negative correlation with the pub\_rec field which is good from the lending club perspective as the investors are not investing much on the borrowers with high pub\_rec values
5. months\_since\_last\_delinquent is negatively correlated with almost all variables and thus borrowers with higher values for the aforesaid column should be well analysed before offering any loan

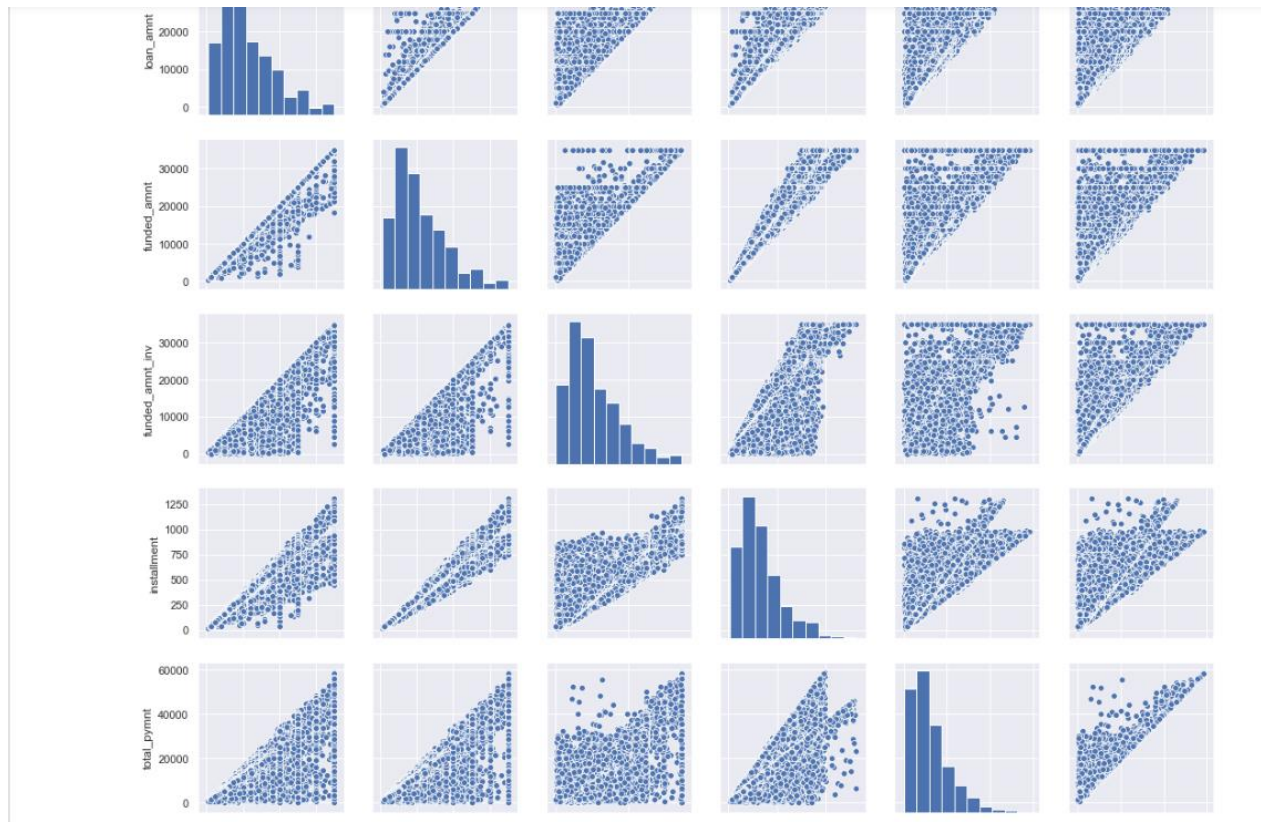


# Analysis



Some of the strongly correlated variables with loan\_amnt field are funded\_amnt , funded\_amnt\_inv , installment , total\_pymnt , total\_pymnt\_inv. Let us plot and analyse scatter pair plots for these

It is evident from the plot that there is a positive correlation between the variables



# Conclusions

Following recommendations are made on the basis of Exploratory Data Analysis so as to minimise the default rate

- A. Driving variable : annual\_inc: maximum number of loans have been given to the people with negligible annual income . This is a risk factor which needs to be kept in mind while predicting loan repayment
- B. Driving variable : purpose : the highest number of loans are offered to people for debt consolidation . This does not sounds a profitable thing to do from business perspective as offering loan to people already in debt is risky.
- C. Driving variable : emp\_length : An interesting discovery is the spike for the borrowers with less than 1 year of experience . They are more likely to default as they haven't reached a stable position in their career . Thus this is a risk factor
- D. Driving variable : pub\_rec\_bankruptcies : A pattern that raises eyebrows is the spike at 1 ( i.e high number of loans being offered to people with 1 bankruptcy record in the past ) . This is a risk factor and needs to be addressed by the organisation
- E. Driving variable : annual\_inc : There is a positive correlation between loan amount and annual income of the borrower. However this should not be the case as annual income should not be the only criteria for offering loan .
- F. Driving variable : grade : customers with poor grades as G are also being offered loans which are close to the highest loan being offered to the customers with better grades

Along with these the insights shared by observing the correlation matrix should also be kept in mind so as to minimise the default rate