



Lending Club Case Study

Mohsin Tariq
Prasad Joshi

Contents

- About Case Study
- Steps
 - Reading and understanding the data
 - Data Cleaning
 - Analysis - Univariate, Bivariate, multivariate
- Conclusion
- References

About Case Study

- A Consumer finance company is into various types of loans to urban customers.
- The company is trying to reach a decision whether to give a loan to a certain individual or not, thus reducing two types of risks:
 - If the person is **likely** to repay the loan, then not approving the loan will result in **loss of business**
 - If the person is **not likely** to repay the loan, then approving the loan will result in **financial loss**
- A company can either choose to accept or reject the loan application. While accepting the loan application, there are various scenarios
 - The customer has **fully paid** the loan
 - There is a current loan **on going**
 - **Charged off:** the customer has missed or delayed some payments for an existing loan
- Based on data analysis, we try and predict which parameters about the customer such as age, monthly income, loan amount may be possible indicators to indicate as to which customer should be given a loan.



Reading the data and understanding

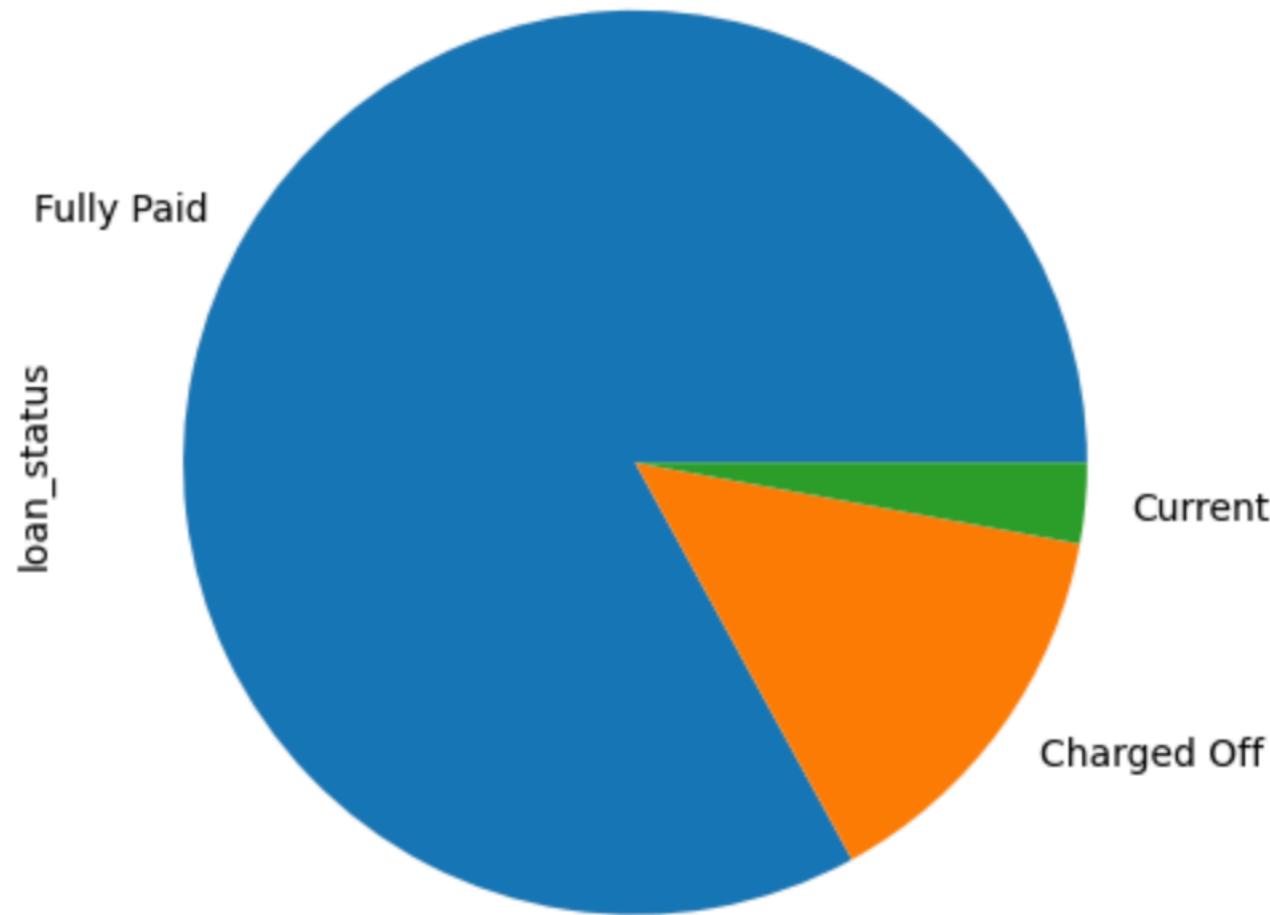
- The Data frame is read and columns are analysed
- Total rows and columns are 39717 and 111 respectively
- There are a mix of columns ranging from categorical, numerical, date etc
- Our target columns is **loan status** which we need to compare and check for defaulters across various parameters.

Data Cleaning

- **NULL Values:** Several columns such as verification_status_joint, annual_inc_joint etc. consist of only NULL values and are dropped
- **Single values:** initial_list_status, application_type, acc_now_delinq have only one NON Null values and hence these are dropped
- **Non relevant columns:** Columns which may not be useful for our analysis can be dropped as they provide no value
- **Outliers:** We analyse data for outliers to see if certain values may affect our overall interpretation of a certain column
- **Derived Columns:** Some columns such as issue_d may be further used to find year wise or month wise data which may be more relevant in scenarios.
- We only consider data for loan status not equal to current which means we are only checking data for those loans which are not currently ongoing.

Analysis

- Checking Categorical variable loan_status to see number of charged off customers:

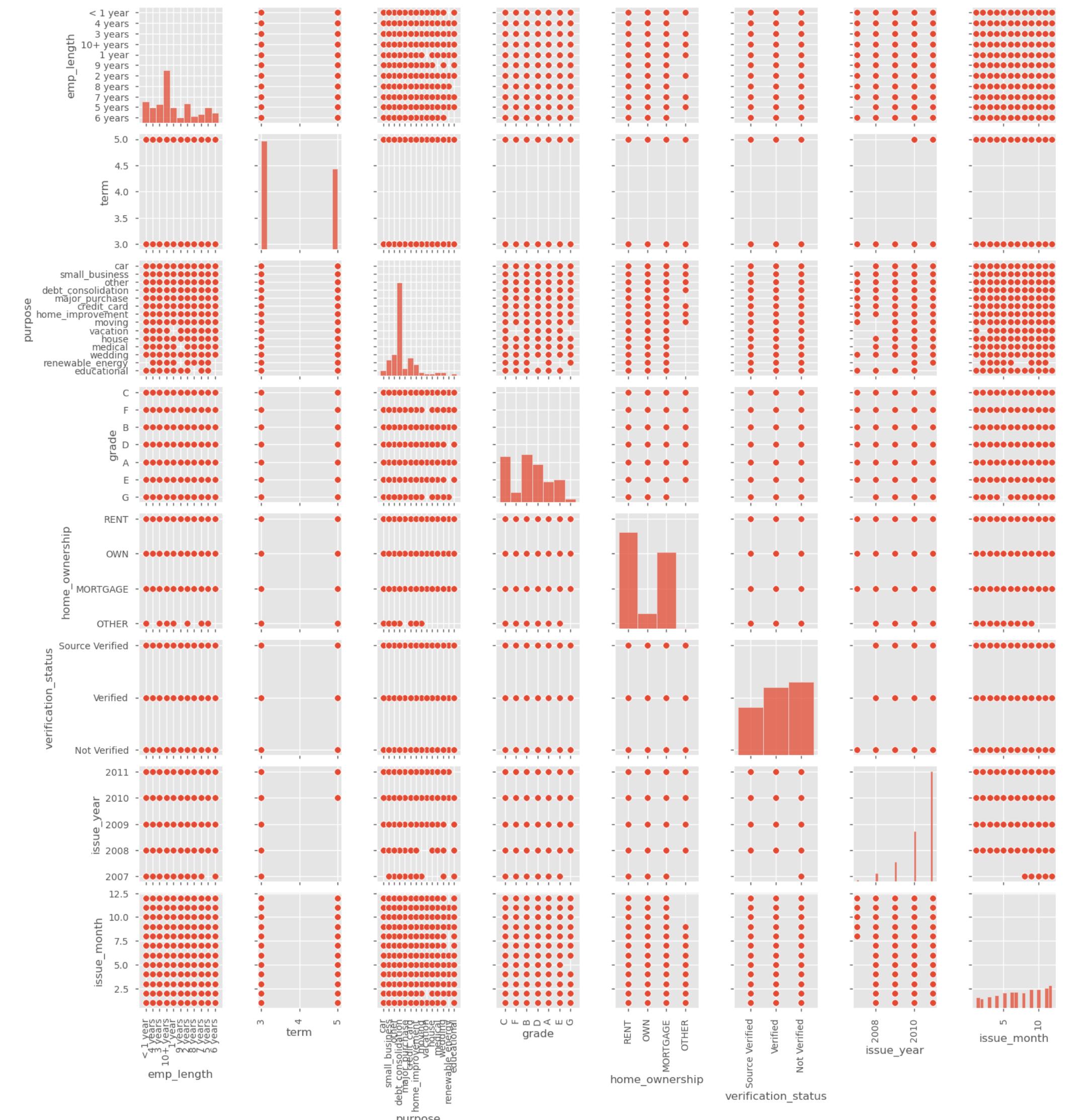


- We consider only charged off data

Analysis Contd..

Though most of the data look similar below are some observations for bivariate Analysis

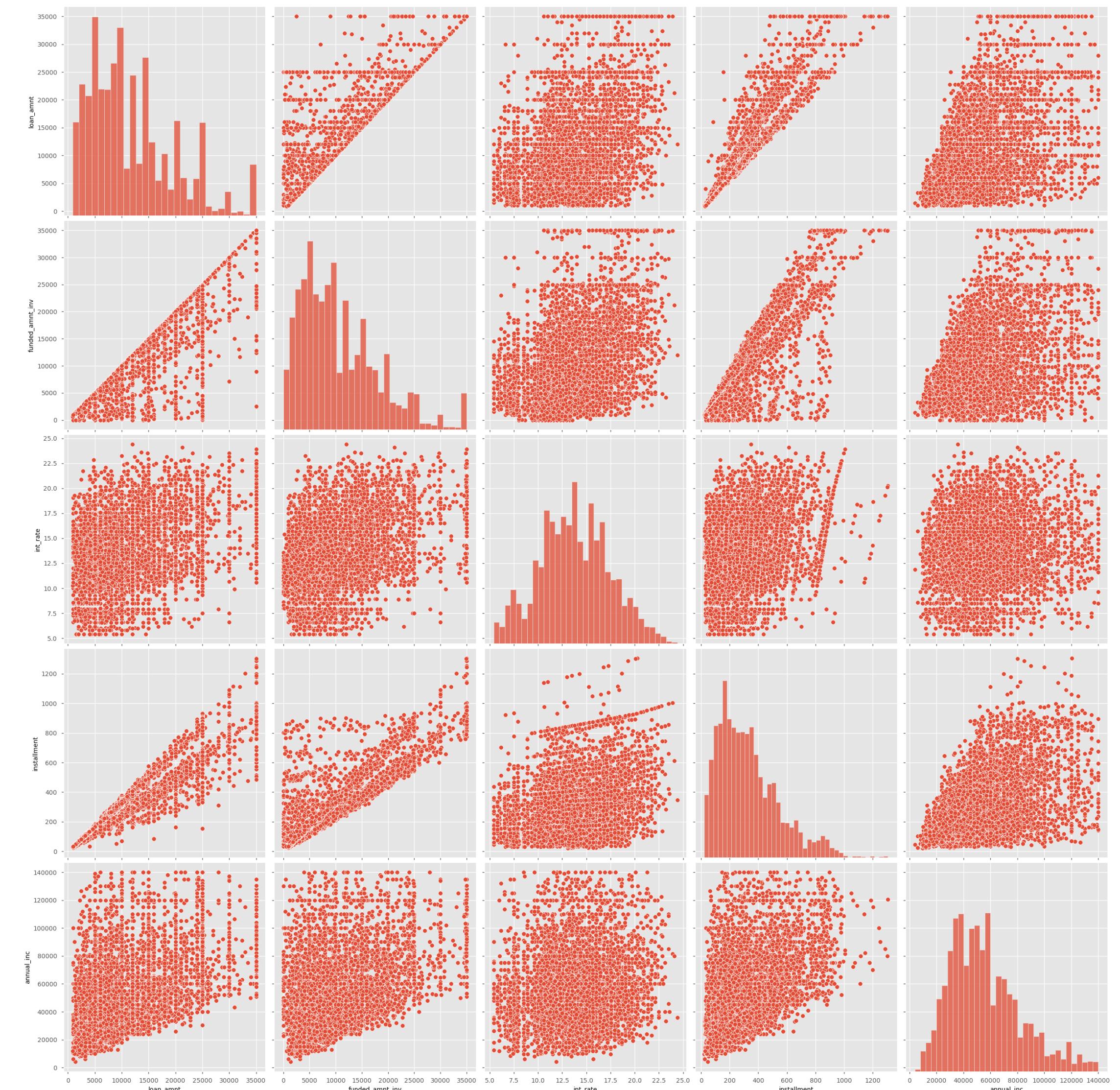
- With every issue_year funded_amnt_inv, int_rate and installment increases
- Term 3 years
- The loan amount was common in most of the year, it increases by approximately 5k in 2011
- Annual income for year 2007 and 2008 looks similar, but a bit of hike in 2009 and them similar till 2011



Analysis Contd..

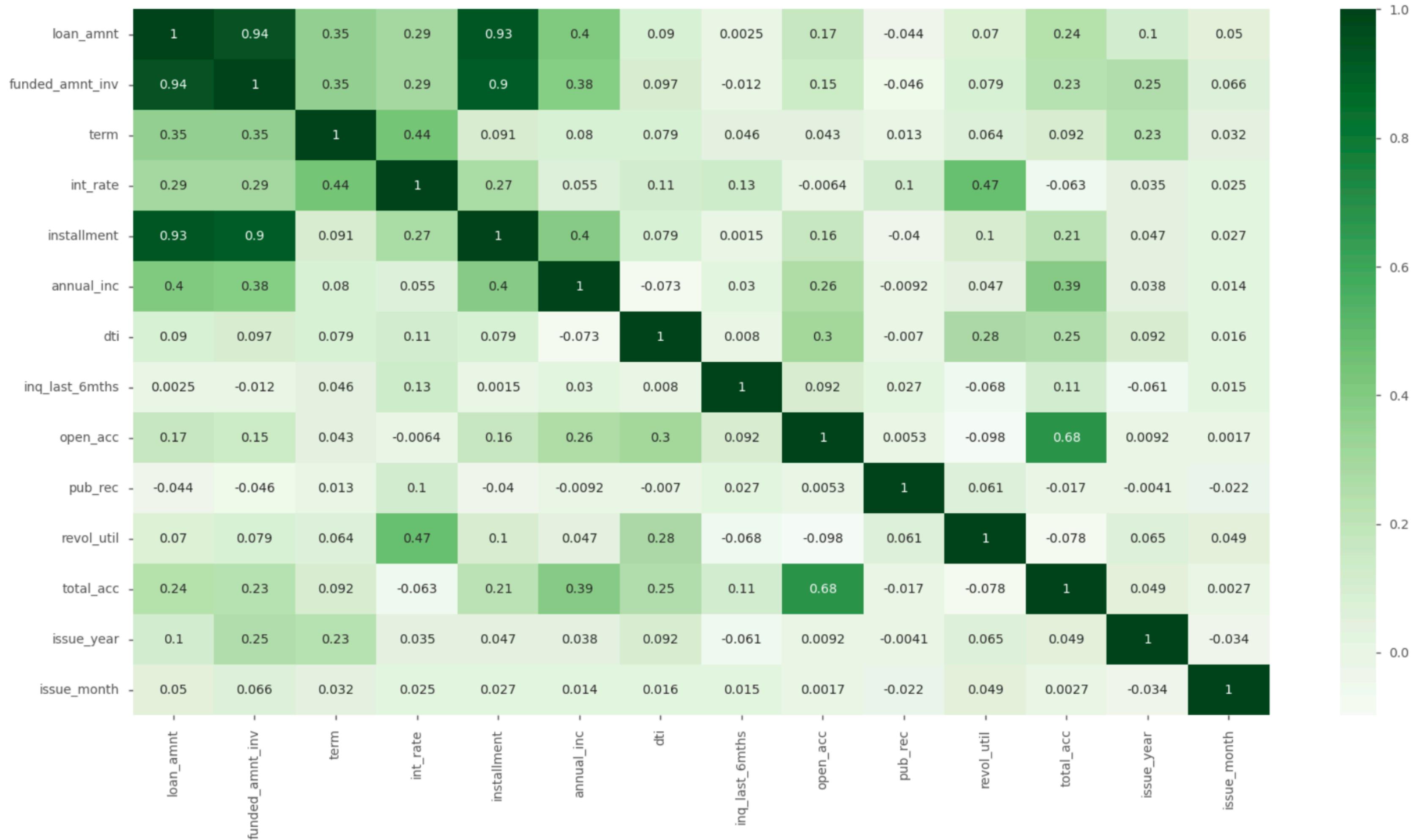
Analysis of above data to be defaulted

- People with 6k-8k of loan amount
 - People with 6k-8k of the amount committed by investors for that loan (funded_amnt_inv)
 - People with rate of interest between 11 to 13
 - People with installed 945 or near by
 - People with dti in between 12 to 17
 - People with open_acc between 6 to 9
 - People with revol_util 65 to 85
 - People with year of issue 2011 or month of issue December
-
- With an increase in funded_amnt_inv the loan amount decreases
 - With an increase in funded_amnt_inv number of installments also increases
 - The number of installments increases with the increase in loan amount



Analysis Contd..

- A heat map is drawn to check correlation between various columns in the dataset:



Conclusion

- People with more than **10 years of experience** (employee length 10+ years)
- People with **3** years of **term**
- People with **Grade B**
- People with house ownership as **RENT** or **Mortgage**
- People **without verification**
- People with **6k-8k** of loan amount
- People with **rate of interest** between **11 to 13**
- People with year of issue **2011** or month of issue **December**
- With an increase in **funded_amnt_inv** the loan **amount decreases**
- With an increase in funded_amnt_inv number of **installments** also increases
- The number of installments increases with the increase in loan amount
- For home ownership for MORTGAGE for annual_inc in year 2008
- For home ownership for MORTGAGE for loan_amnt in year 2011
- For home ownership for OTHER for loan_amnt in year 2011
- For home ownership for MORTGAGE funded_amnt_inv, installment, int_rate, loan_amnt, revol_utils increased rapidly every year. So increase in any of these columns will directly impact others and will have more changes to be defaulted
- For Home ownership **MORTGAGE** for verified and for source verified for annual income
- For home ownership **OTHER** for verified and source **verified** for annual income
- For medical purpose for loan status fully paid for loan_amnt in year 2007
- For Major Purchas for status charged off for loan_amnt in 2007
- With every issue_year funded_amnt_inv, int_rate and installment increases
- The loan amount was common in most of the year, it increases by approximately 5k in 2011

References

- Google images
- <https://www.youtube.com/watch?v=BVmL3VQCmEM>
- https://youtu.be/fiQhxn9RjEQ?si=BCJXn_9PzShUvkNj
- Lending Club Live Session conducted on 2nd Sep 2023