Lending Club Case Study

Idea about how real business problems are solved using EDA.

Group Contributors

1.Lydia Sharon James (Group Leader)

2.Paratyaksh Singh

Problem Statement

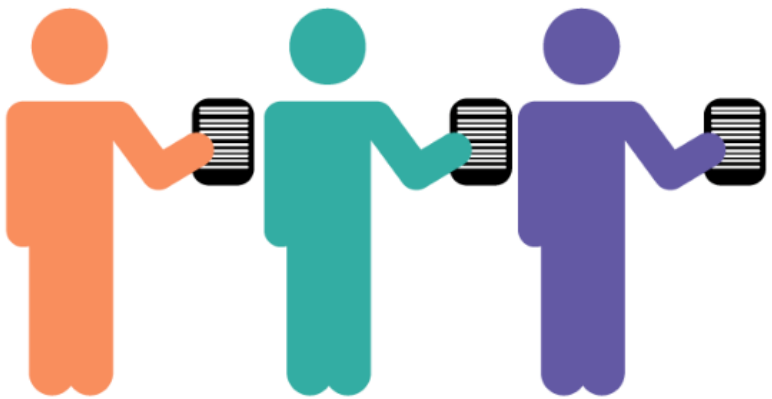
The aim of the project is to identify the patterns in a consumer finance company - called Lending Club's data. It is a consumer finance company which specializes in giving loan to various kind of Urban consumers.

The objective here is to identify the patterns in the data which helps the company to identify less risky customers.

The company has 2 problem statements:

1) Not to lose opportunity by not giving the loan to the right customer
2) Not to lose money by giving the loan to the risky customer

LOAN DATASET



Loan Accepted → Default

Loan Accepted → Non-Default

Loan Rejected

(Not considered in dataset)

Approach

Data Understanding

Data Cleaning

Data Analysis

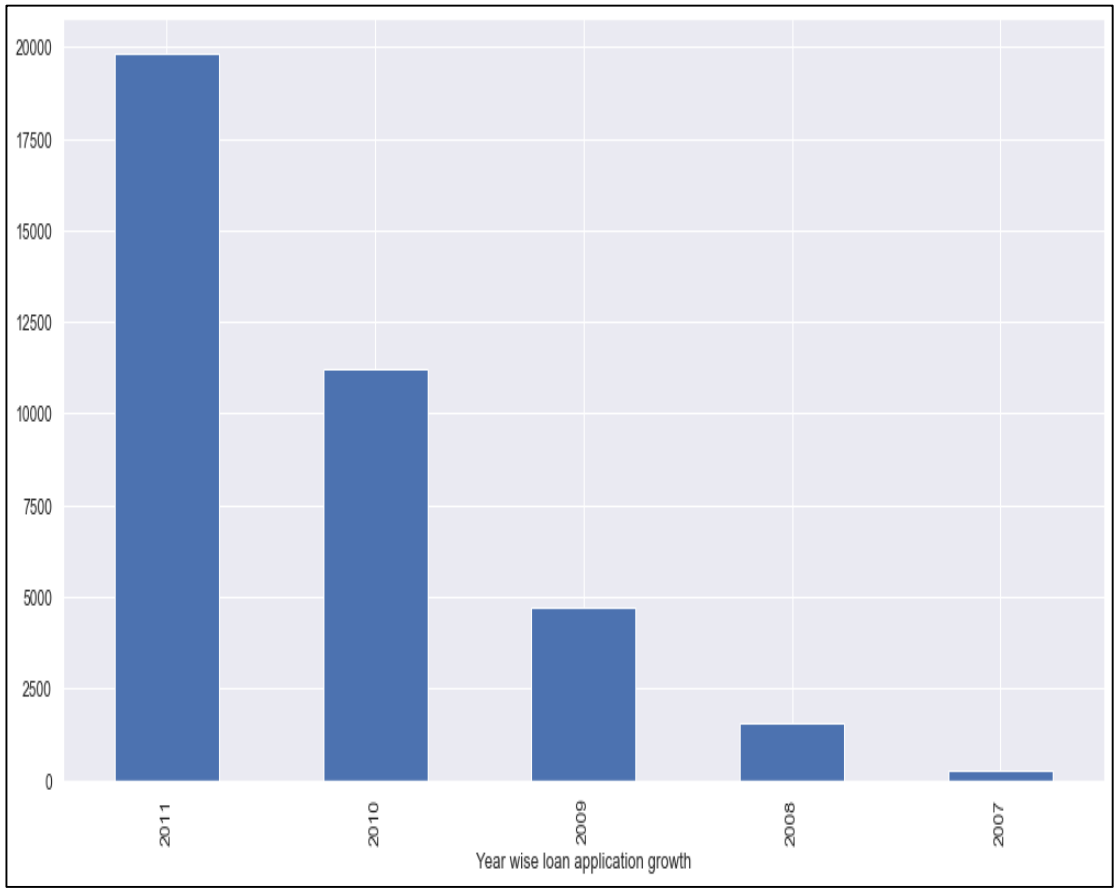
Recommendations

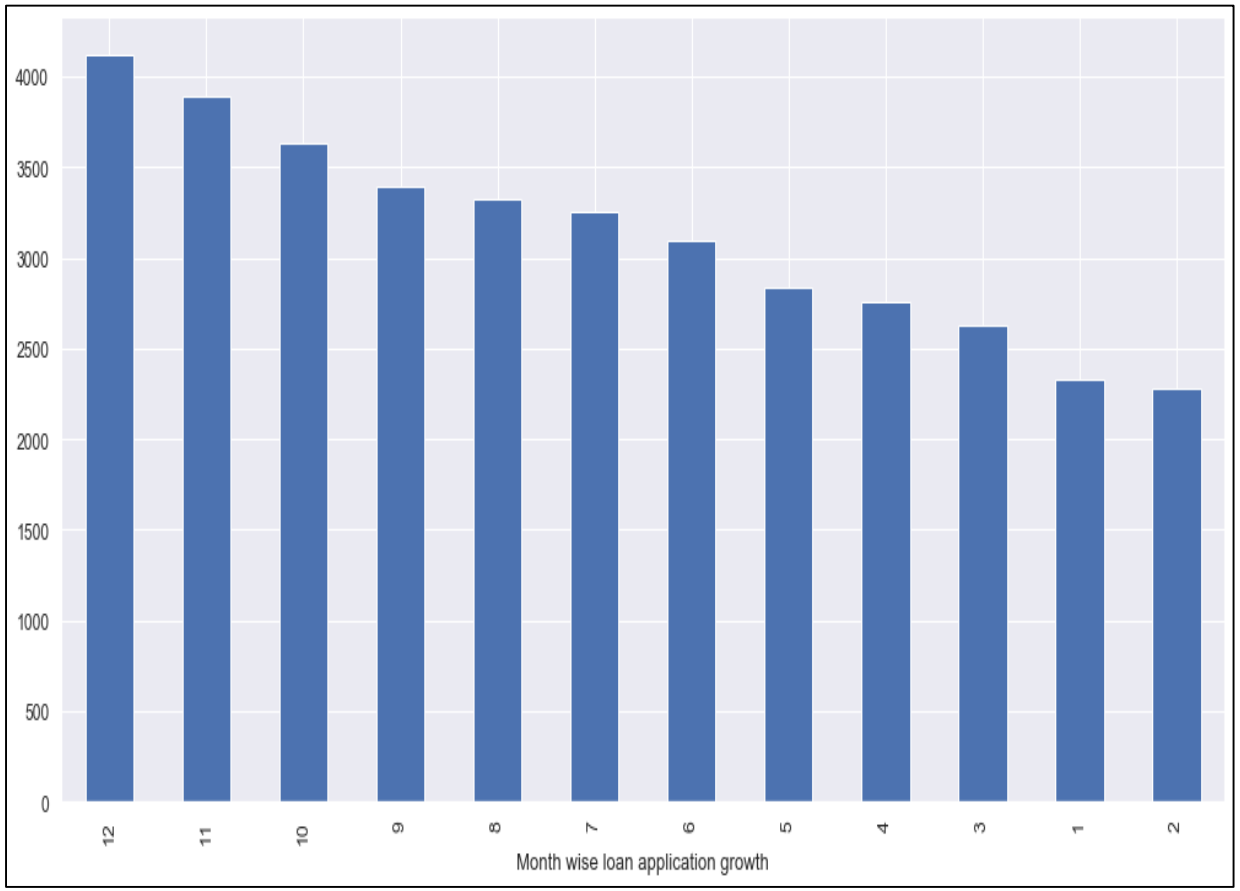
Data Cleaning

- Removing the null valued columns
- Unnecessary variables
- Checking the null value percentage
- Removing the respective rows.
- Any column with a singular value across the rows
- Insignificant number of NaN values
- NaN values in a column higher than 60% of total values
- Columns which were text as well as non-categorical such as description, title, etc
- Columns which were insignificant to the analysis such as id, member_id, etc

DATA ANALYSIS

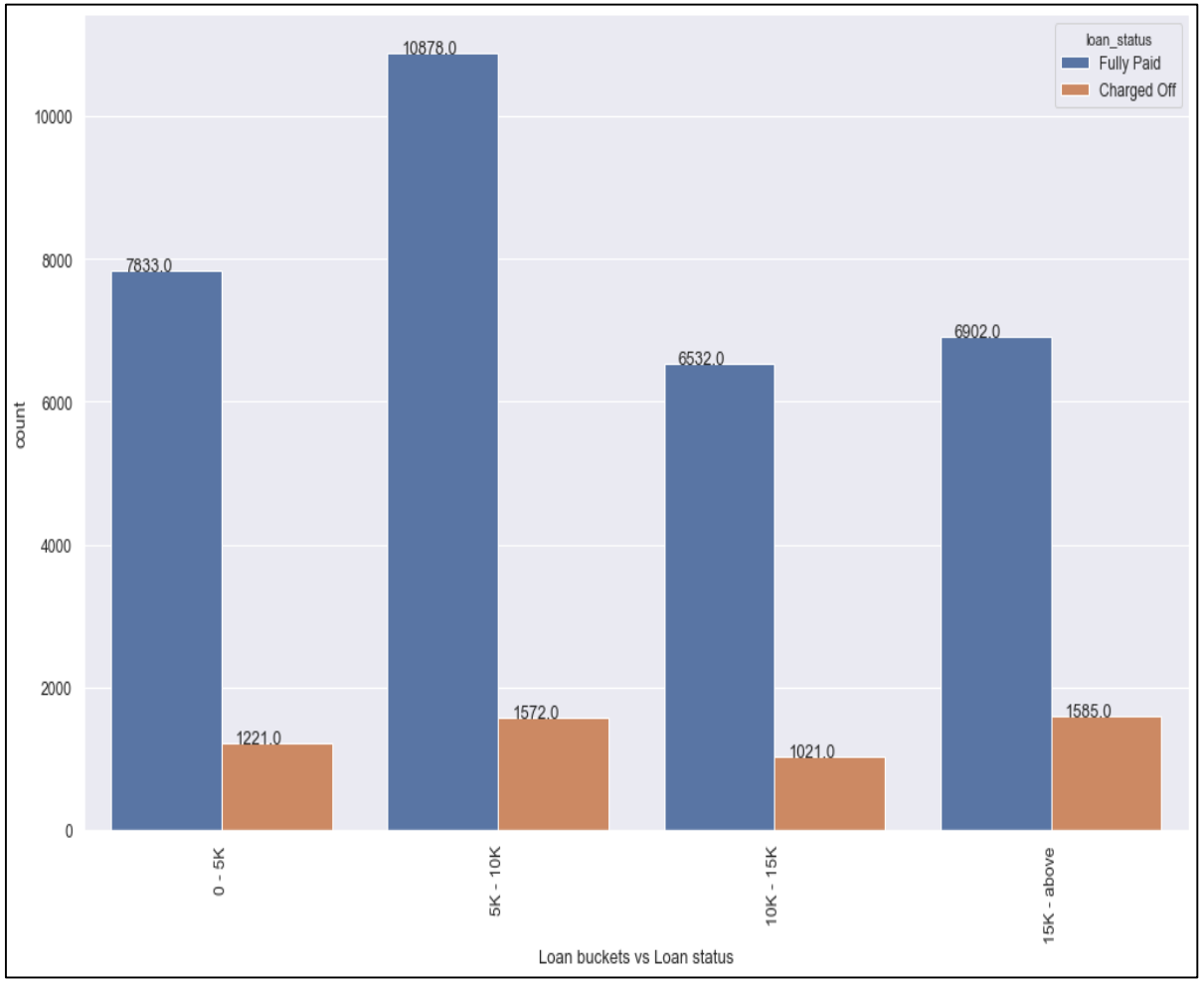
ANALYSIS – COMPANY GROWTH

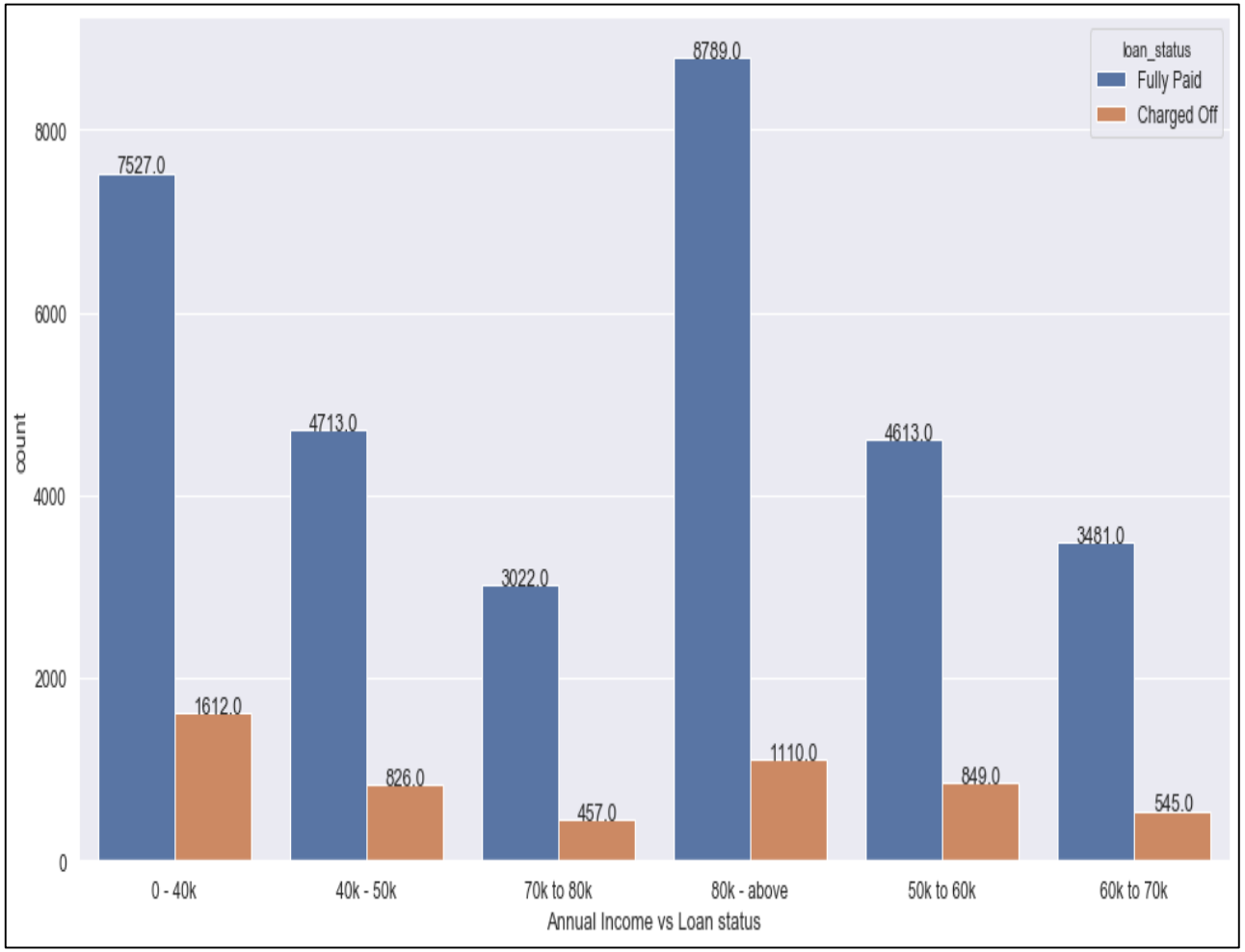




- Number of applications per year are increasing very rapidly year on year
- Monthly all applications are increasing except for February
- Highest number of applications are punched in the Financial year Quarter 3. This may be due to being the festival and holiday season of the year and the demand for consumer products increases at that time, therefore the consumer loan demand may also increase at that time

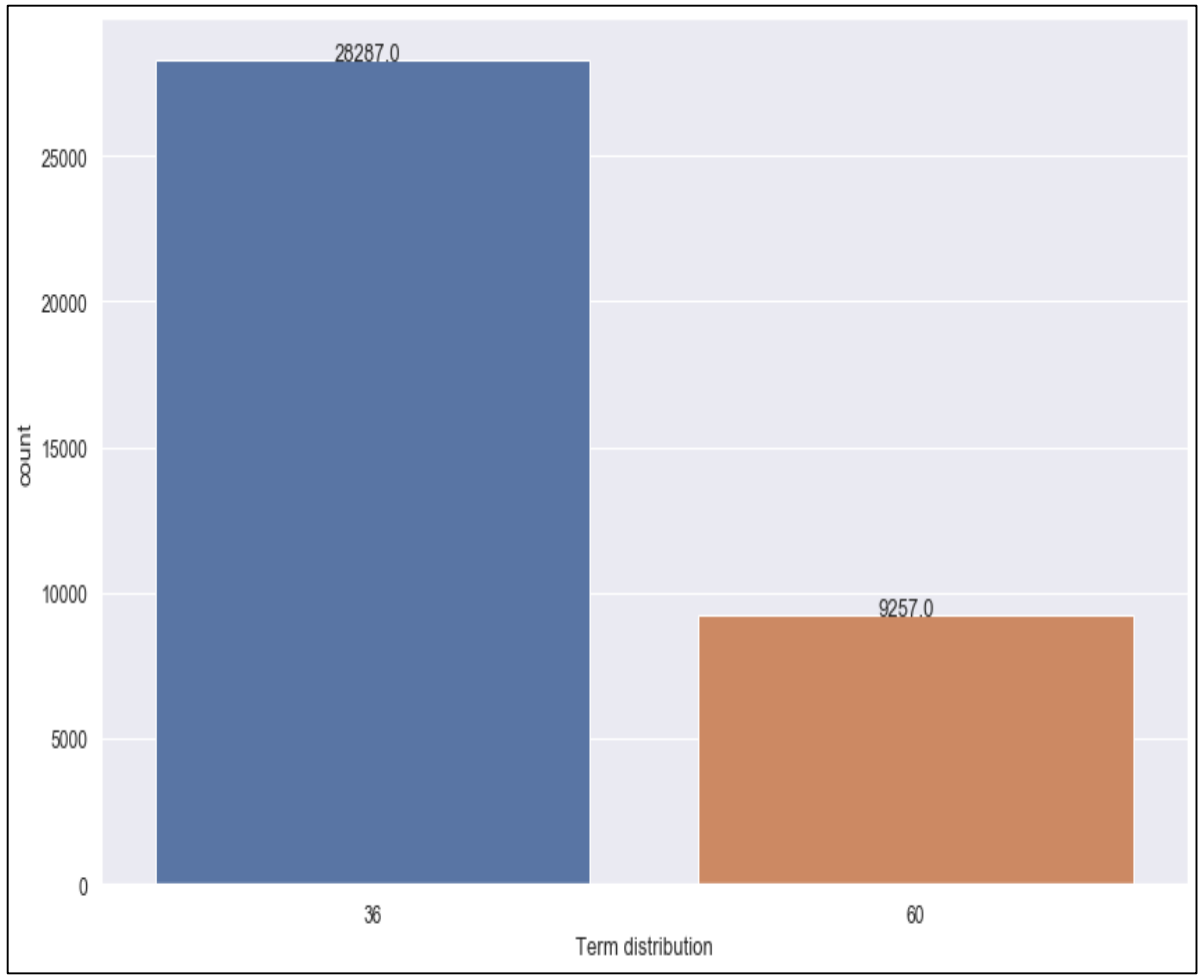
ANALYSIS – LOAN STATUS

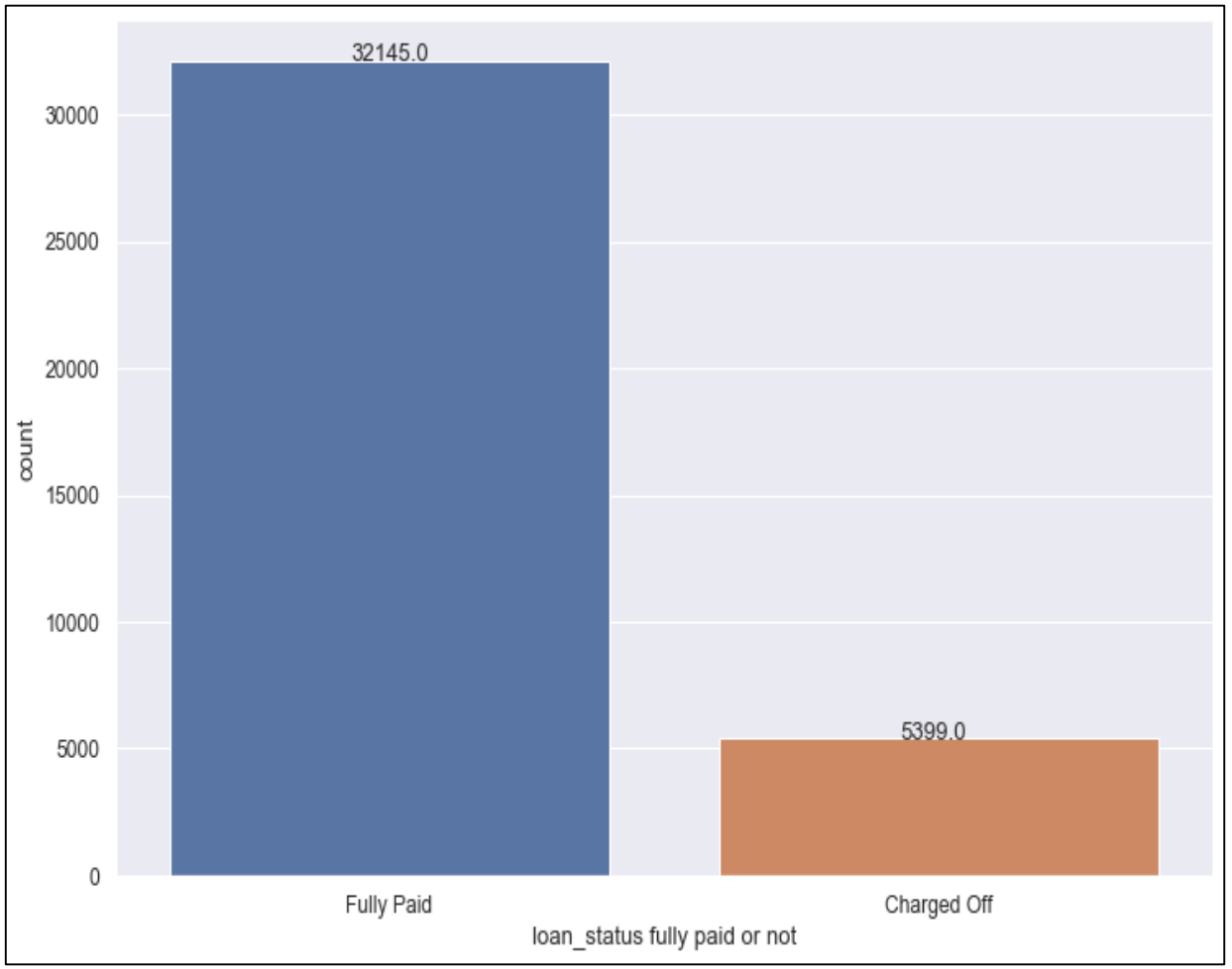




- Maximum Loans have been disbursed in the bucket of ‘5K – 10K’ and it has the highest fully paid status among all other buckets
- The maximum loss is happening in ‘15K – above’ bucket as the number of loans is 40% lesser than ‘5K-10K’ bucket yet the ‘Charged Off’ loans are almost same. This may be due to the fact that higher ticket loans are difficult to re-pay
- ‘80K - above’ have a good ratio of ‘fully paid’ a whereas there is a gradual dip from ‘0-40K’ bucket to ’70K-80K’ bucket.
- ‘0-40K’ bucket have higher loan amount than 40-50K, 50-60K, 60-70K, 70-80K as they may not have enough ‘discretionary income’ as compared to other buckets and need some loan to buy products

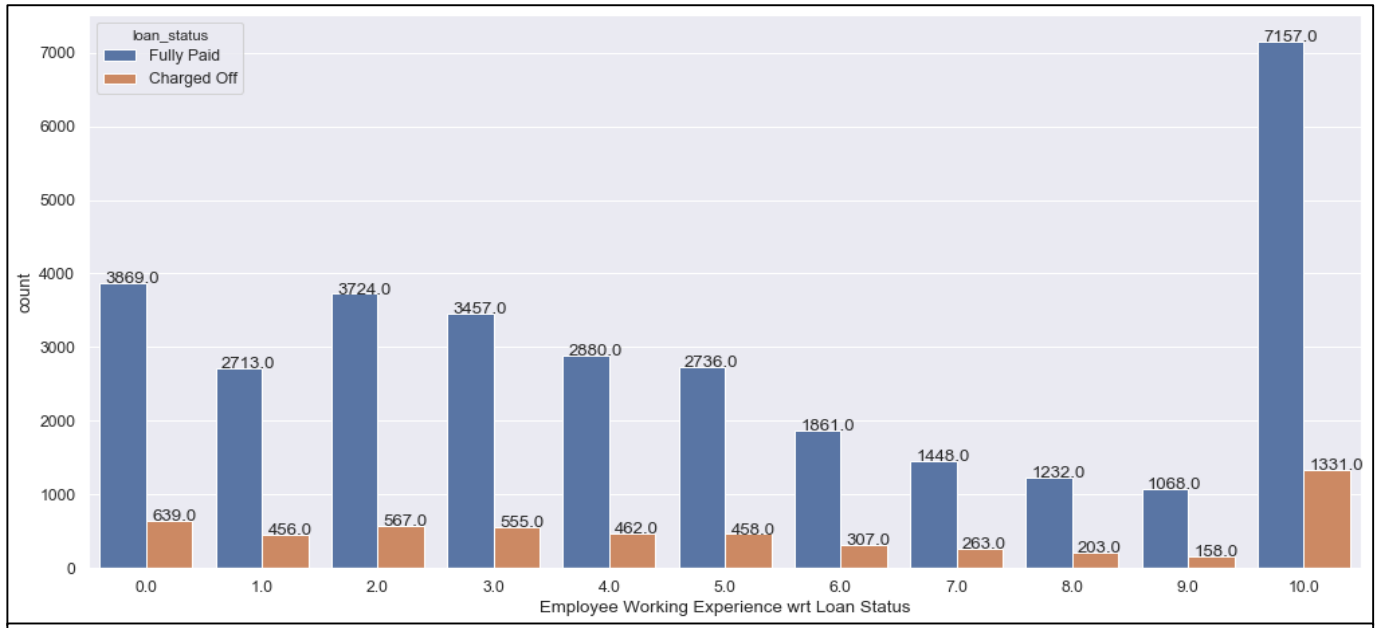
ANALYSIS – LOAN STATUS

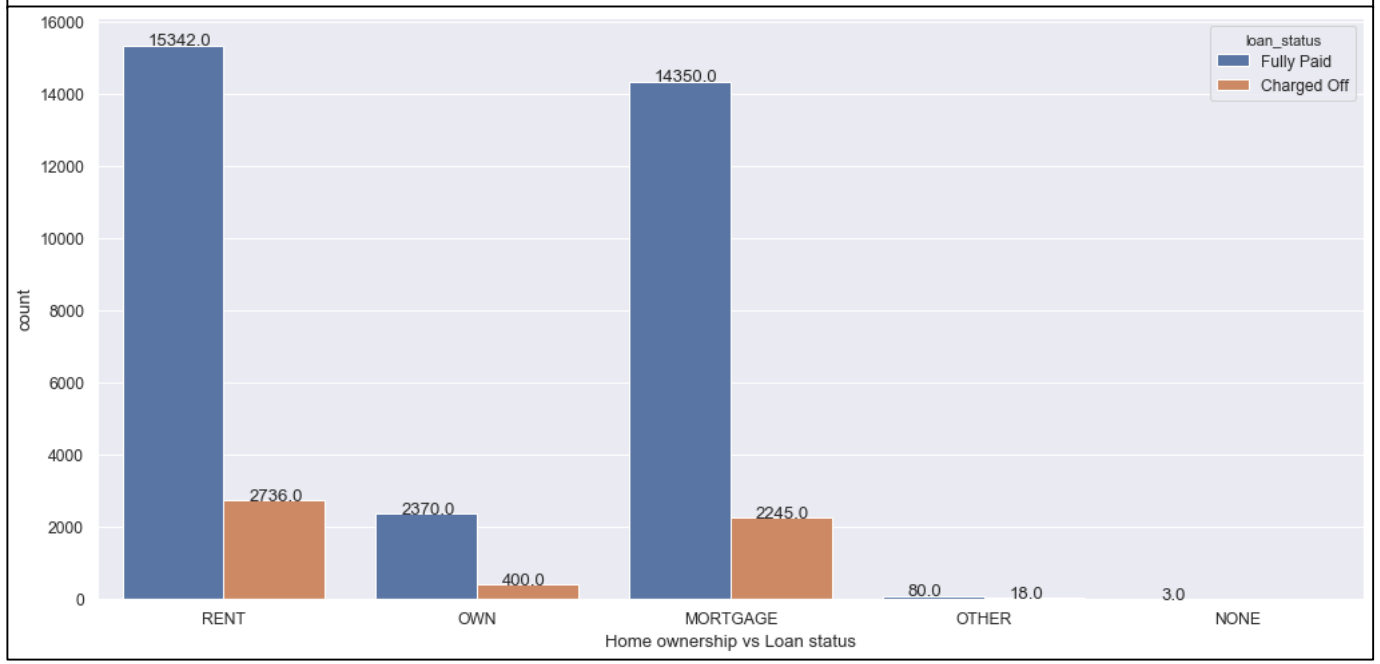




- There are only two loan terms 36 and 60 months
- Almost 75% of the borrowers have taken the loan of 36 months and rest 25% are in the 60 months bucket
- Almost 85% of the loans have been ‘Fully Paid’ whereas 15% are in ‘Charged Off’ State

ANALYSIS - BIVARIATE



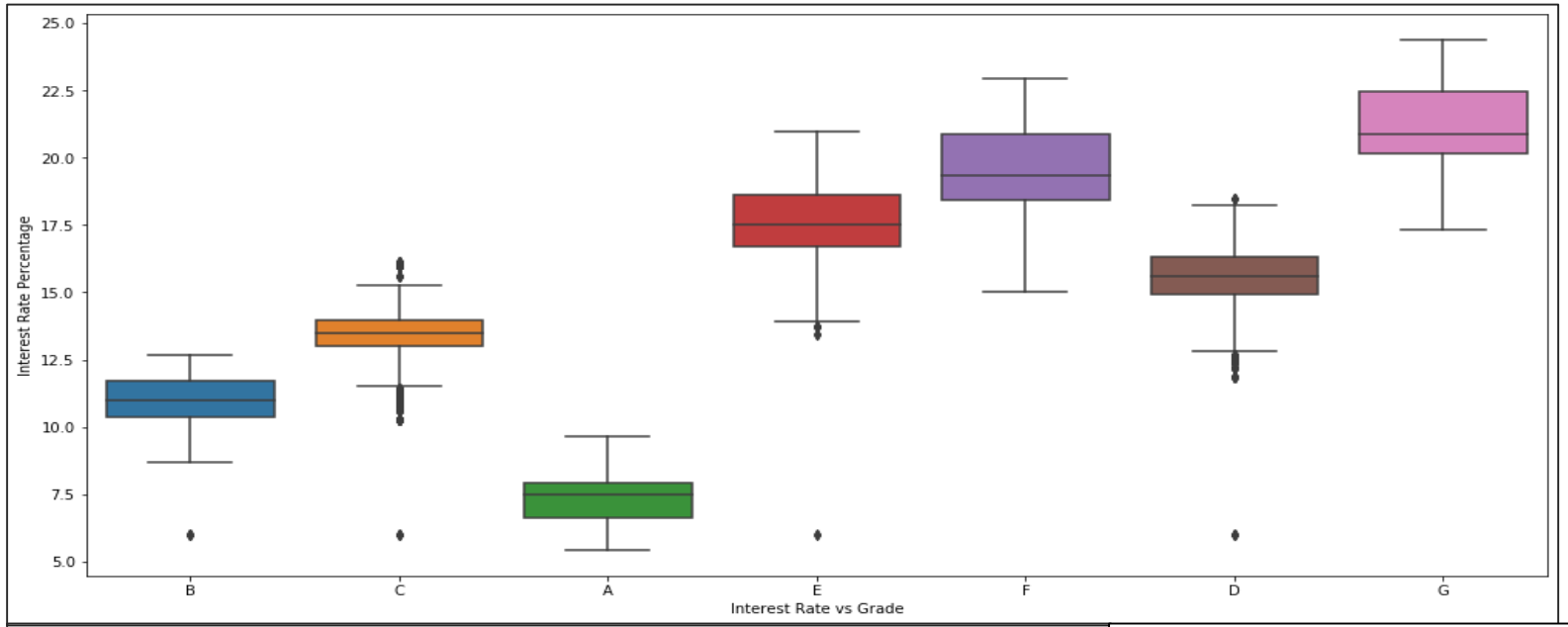


- "Fully Paid" category is almost increasing as the Employee work experience is increasing although there is a certain dip between 0 to 1 year experience which can be further investigated with more data

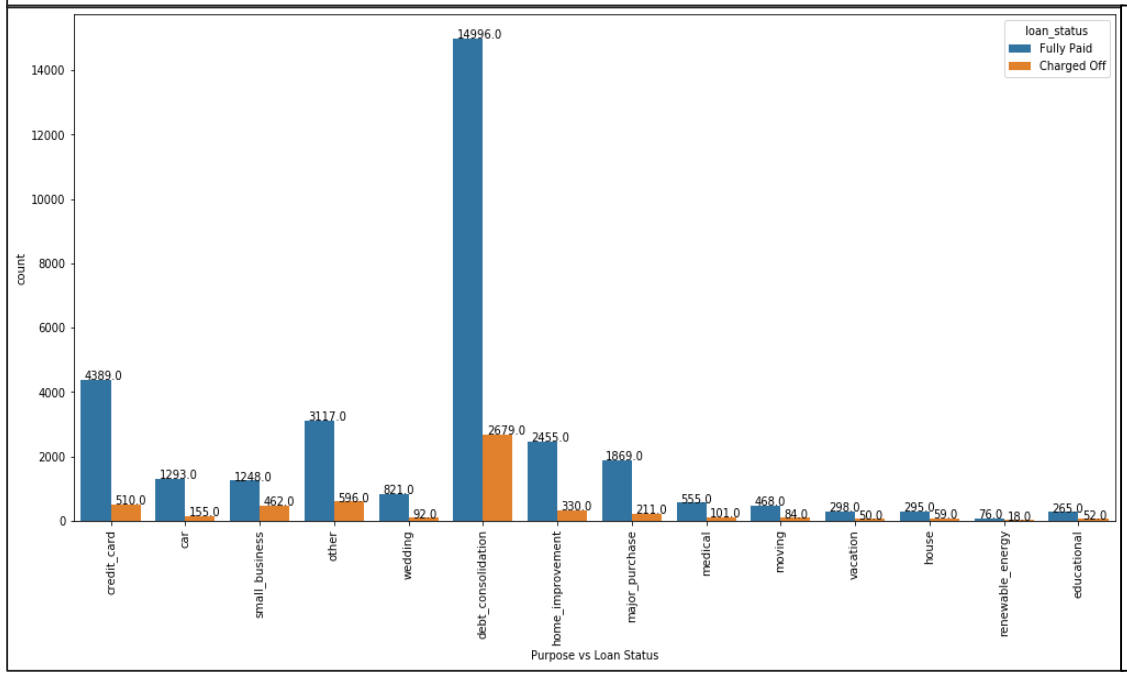
- The people with 10+ years of work ex have the highest number of loan applications as well as highest "Fully Paid" number

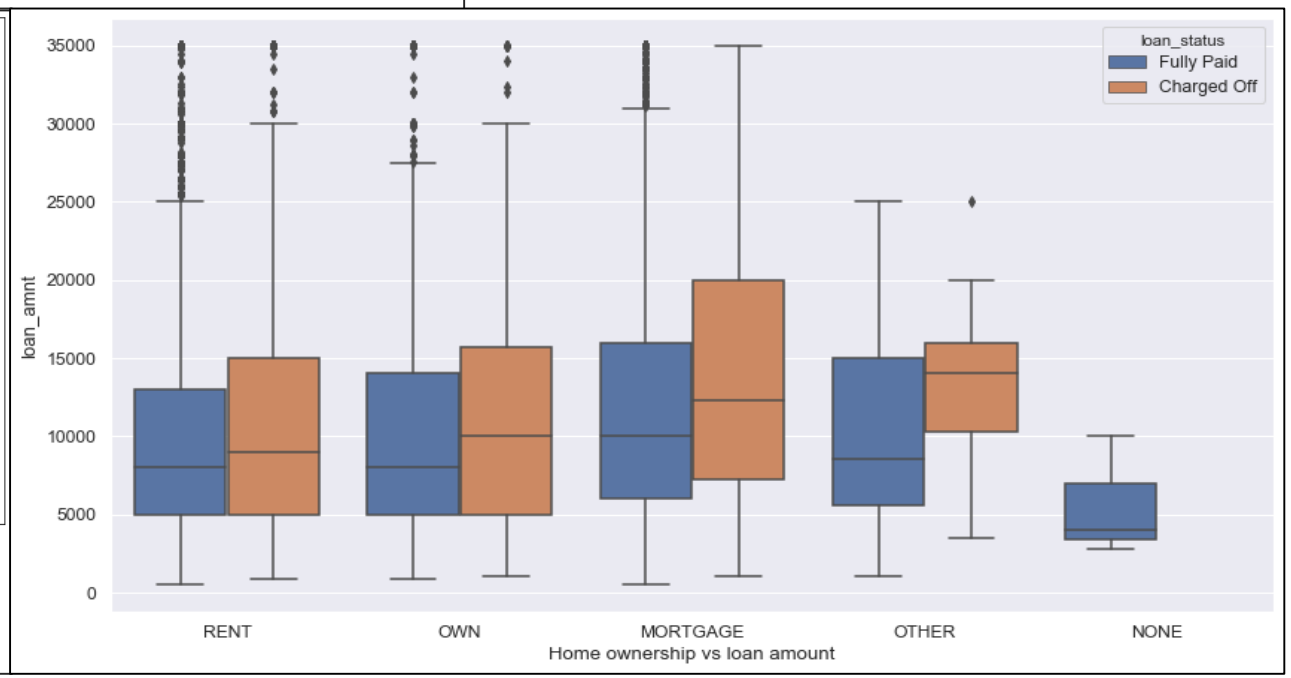
- People who own homes have taken the least number of loans
- Whereas the ratio of "Fully paid" to "Charged Off" is almost similar to Rented as well as mortgaged customers

ANALYSIS - BIVARIATE

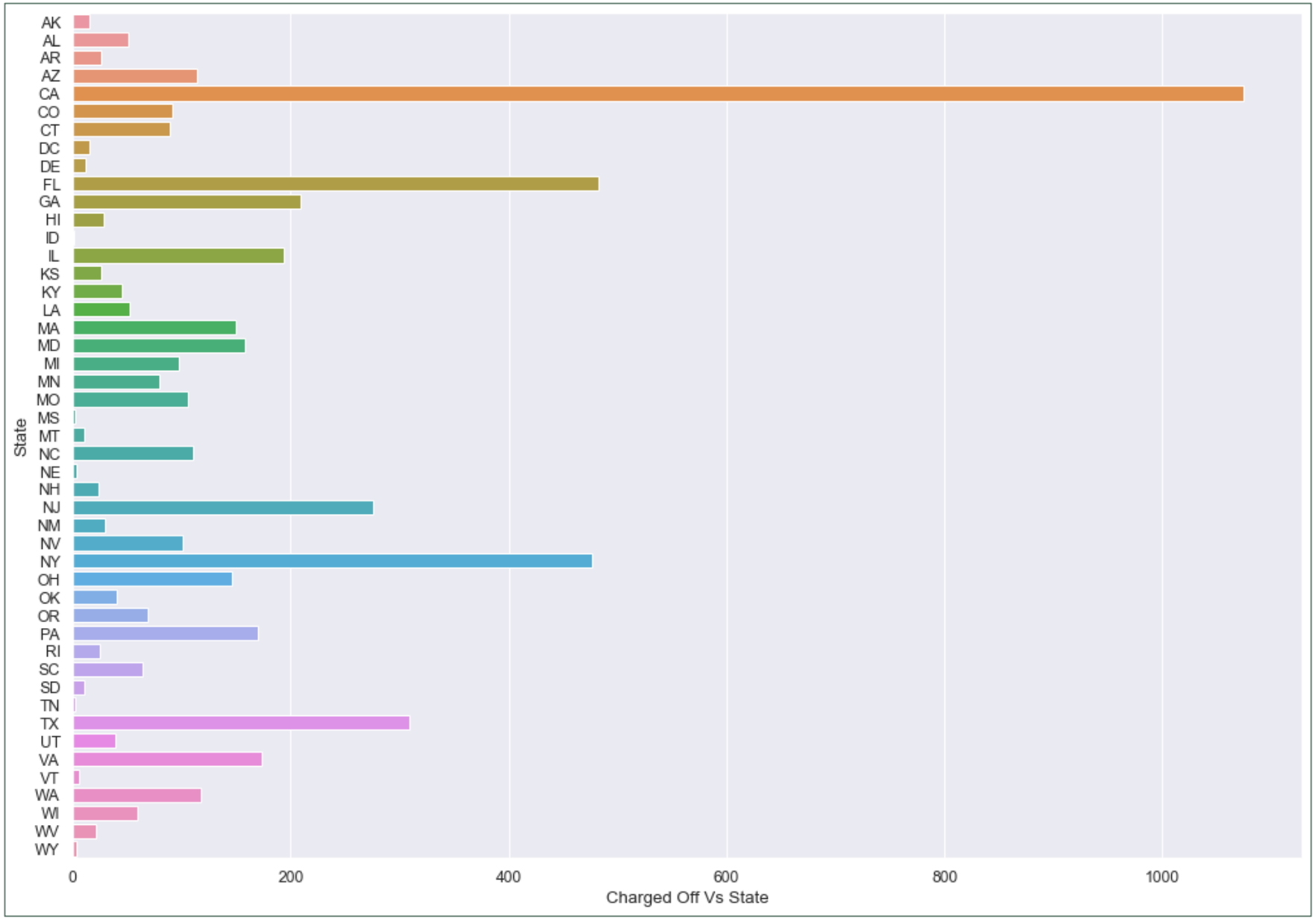


- Higher Grade attracts more interest rate and there is a steady increase in the medians from ‘A to G’ Grade
- Majority of the loans are taken for debt consolidation
- Lower grades such as A, B are also getting interest rate at higher percentage which might be a reason for default





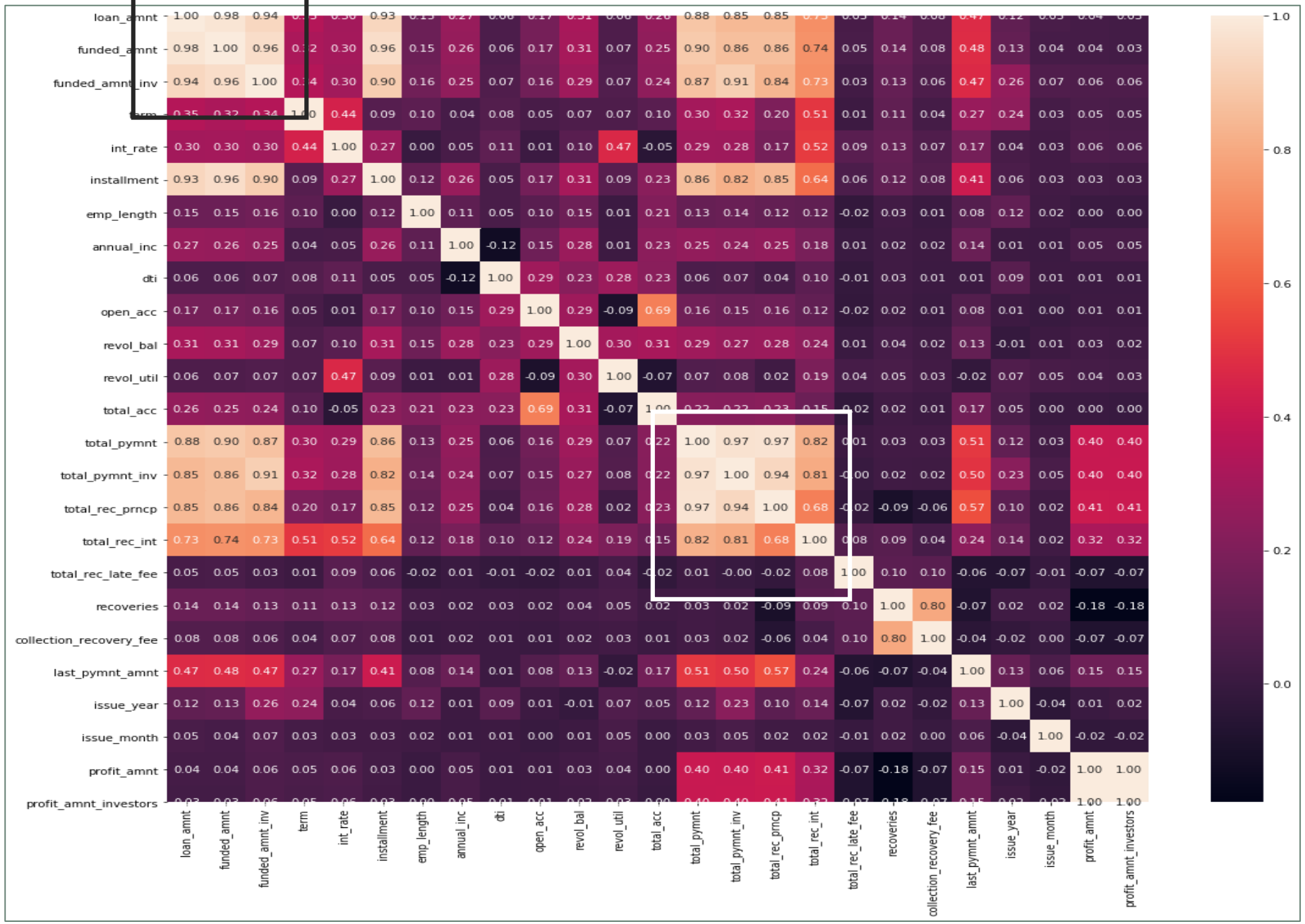
ANALYSIS – STATE WISE VS ‘CHARGED OFF’ BUCKET



- The data shows the highest number of ‘Charged Off’ cases per state
- As per the data California is the state for which thorough verification should be done before issuing a loan

- Same goes for Florida, New York and Newark

ANALYSIS - CORRELATION



- As per the correlation function we can see in the white square that total_pymnt, total_pymnt_inv, total_rec_prncp and total received interest are highly correlated therefore we can take any one of these

- As visible from the black square the loan amount, funded amount and funded amount invested are highly correlated so we have taken loan amount for every analysis

- Revolving utility and revolving balance are positively correlated with interest rate and dti

RECOMMENDATIONS

- For the customers coming for a loan for vacation, house or small business should be either charged higher interest or some collateral should be taken as their default percentage is almost 17%.
- The above mentioned loans are generally not small ticket loans as well
- Also for small business and house loans term should be lessened as almost 50% of the cases are of “Charged Off” and have 60 months as term
- Lending club should also consider not giving 60 month term loans to other categories as this term leads to higher “Charged Off” cases
- Grades are a good predictor of loan default and some associated metrics should be taken into account for calculating the interest rate by the Lending Club
- As per the data California is the state for which thorough verification should be done before issuing a loan
- Same goes for Florida, New York and Newark