Masters of Science

in

Machine Learning & AI

Lending Club Case Study

Contributors:-

Ankush Sharma

Rakesh Kumar Reddy Mudda

Problem Statement :-

Lending Club 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). 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'.

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 described below:

- **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
- **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'.
- **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

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'*.
The core objective of the excercise is to **help the company minimise the credit loss**. There are two potential sources of **credit loss** are:

- Applicant **likely to repay the loan**, such an applicant will bring in profit to the company with interest rates.** Rejecting such applicants will result in loss of business**.
- Applicant **not likely to repay** the loan, i.e. and will potentially default, then approving the loan may lead to a financial loss* for the company

Problem Solving Methodology :-

Data Understanding

↓

Data Cleaning

↓

Univariate Analysis

↓

Bivariate Analysis

↓

Multivariate Analysis

↓

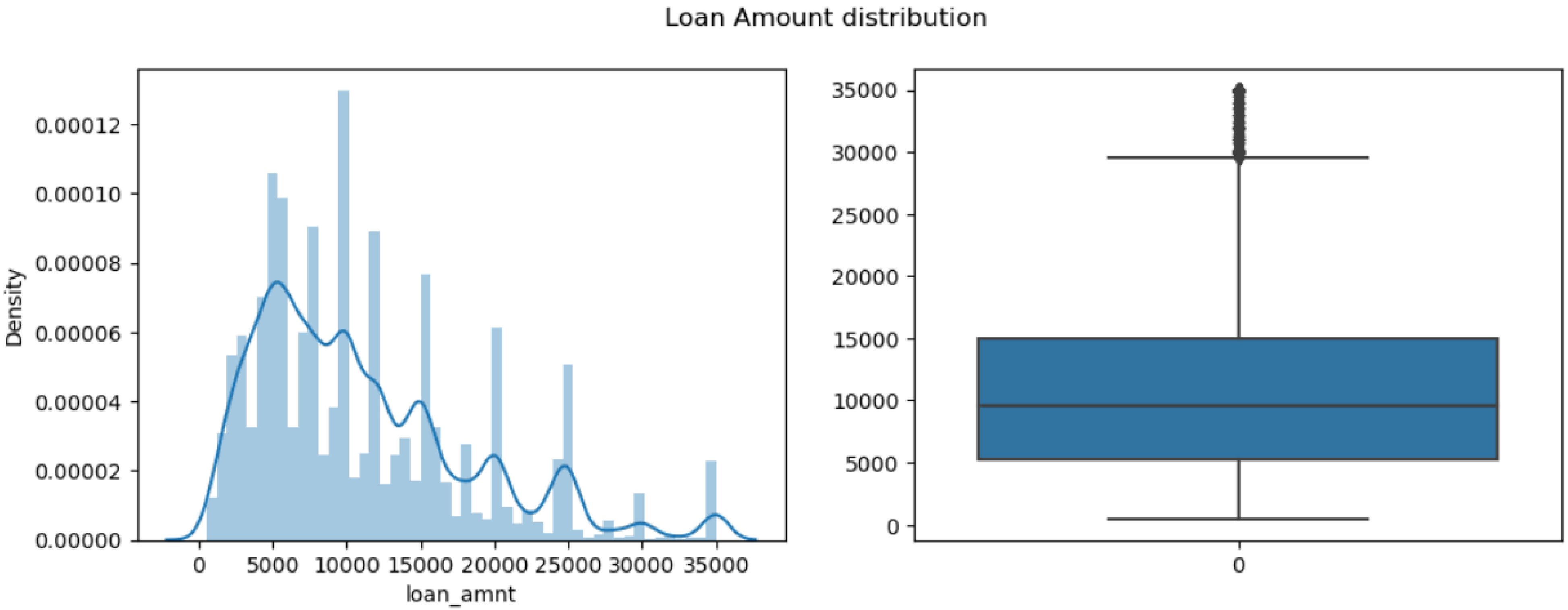
Recommendations

Data Cleaning and Manipulation :-

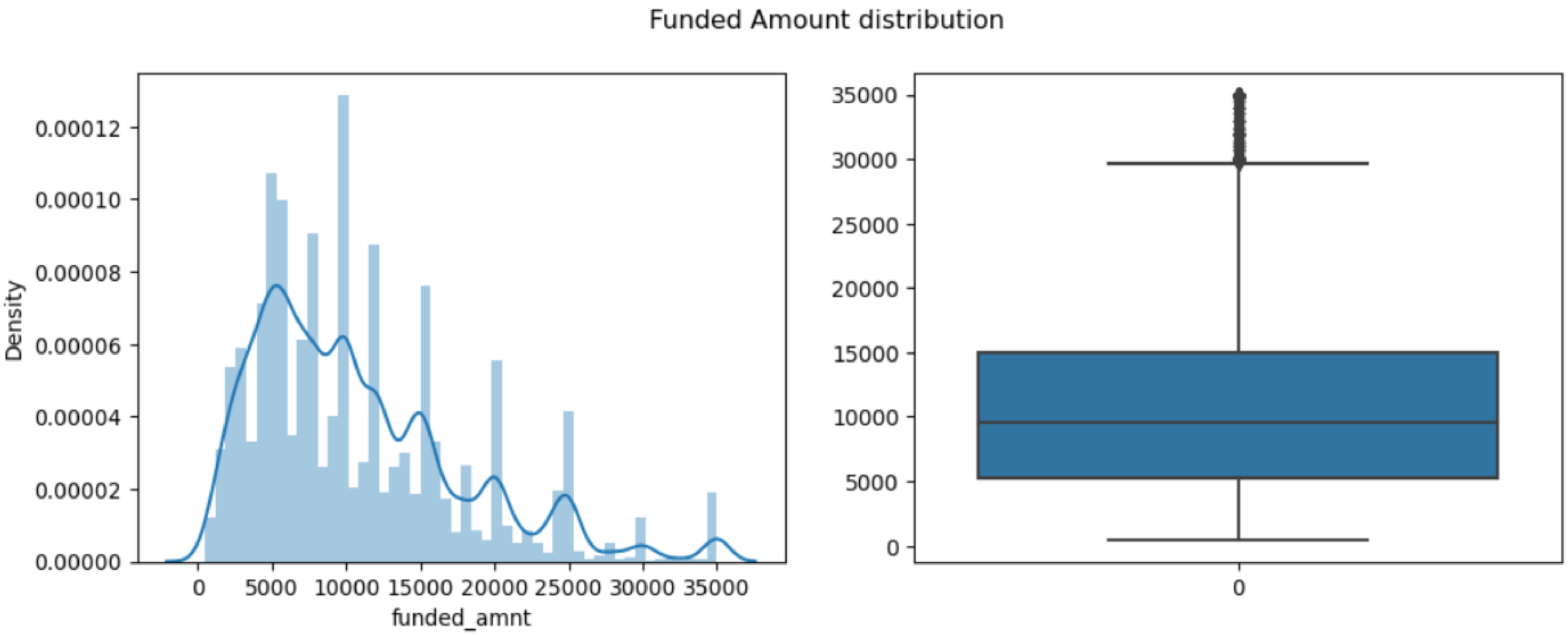
Data Quality Issues can be treated by

- For Missing Values in the Columns:
    - Dropping the columns containing maximum null values
    - Dropping the columns which are not required for the analysis
    - Drop columns with just a single value
    - Drop columns which does not add any value to analysis.
- For Missing Values in the Rows
    - Filter rows to perform targeted analysis
    - Dropping the rows containing the missing values
    - Check for data inconsistencies
- Missing Value Imputation
- Removing Duplicate Entries
- Incorrect Data Types:
    - Clean certain values/Manipulation of strings
    - Clean and convert an entire column to correct data type e.g date, int64 etc.
    - Converting the columns to numeric variables in order to perform more analysis.
- Treating outliers

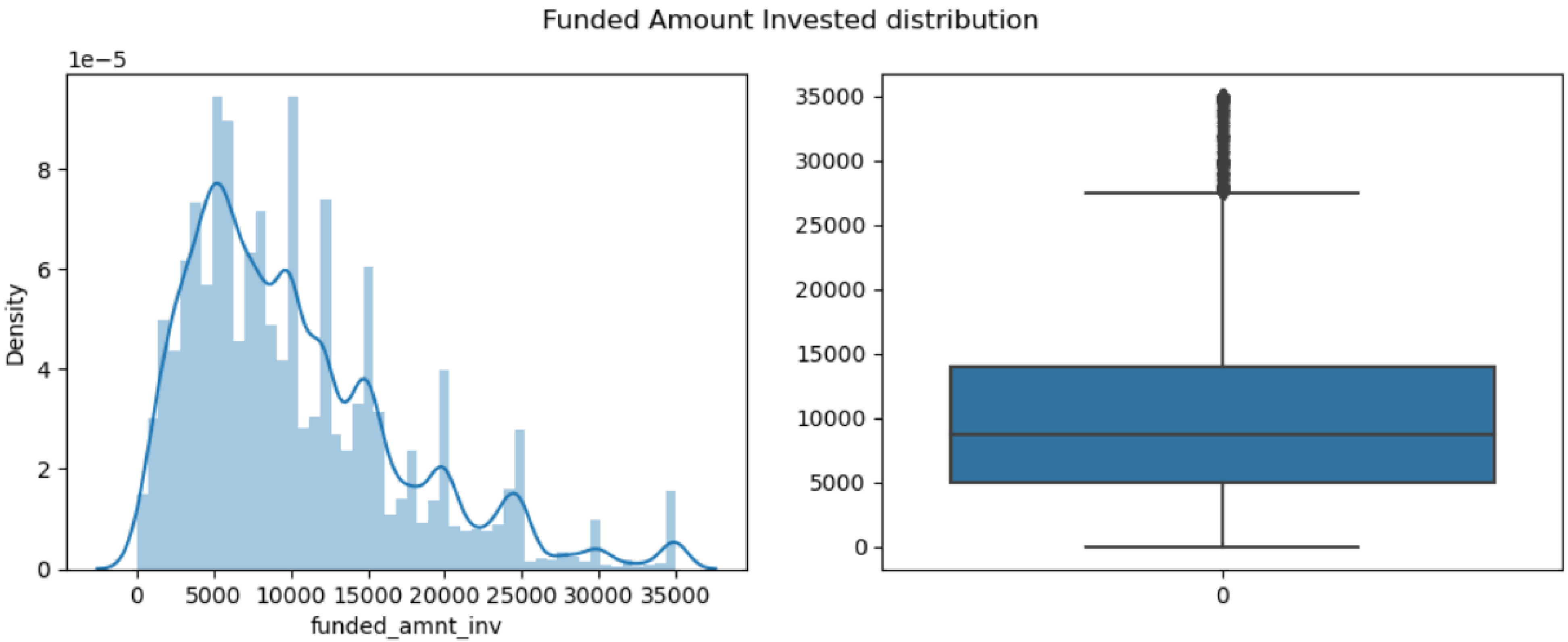
Univariate Analysis:-



Observations:
From the above plots, more number of people took loan amount of 10000, and also median of distribution is 10000. And very few people took more than 30000 loan amount.

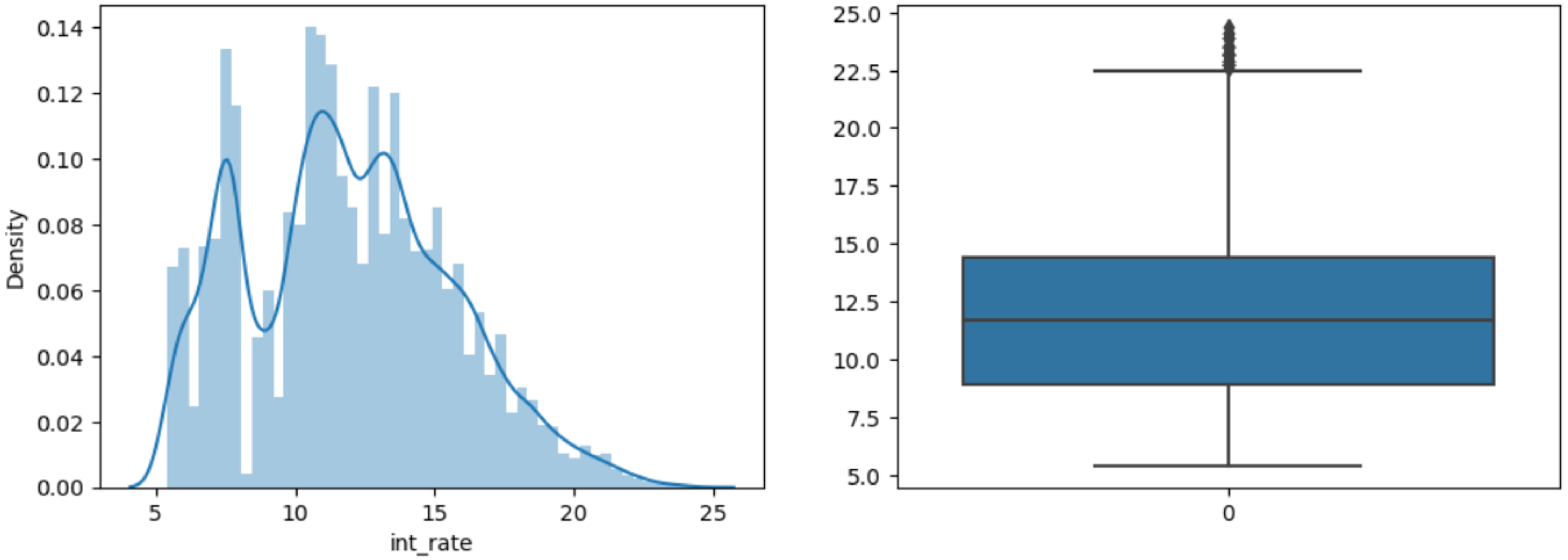


Observations:
Funded amount data behaves similar to loan Amount. It tells us that Lending club approved most of Applied loan amount.



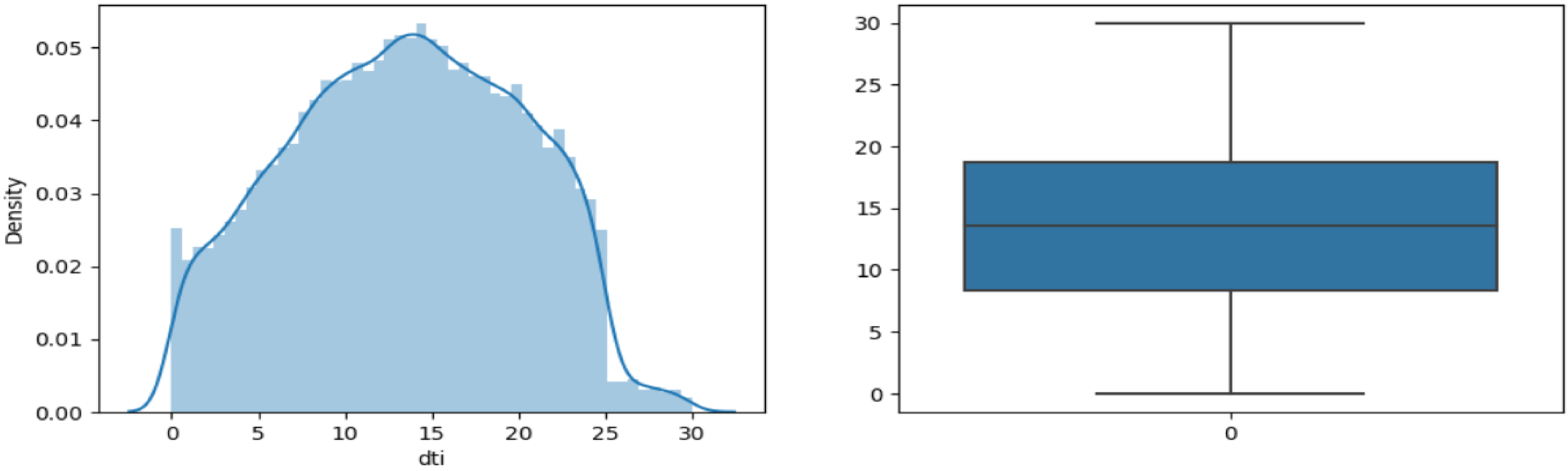
Observations:
Funded amount investment data behaves similar to loan Amount, which tells us that investors in Lending club approved almost of Applied loan amount

Interest Rate Distribution

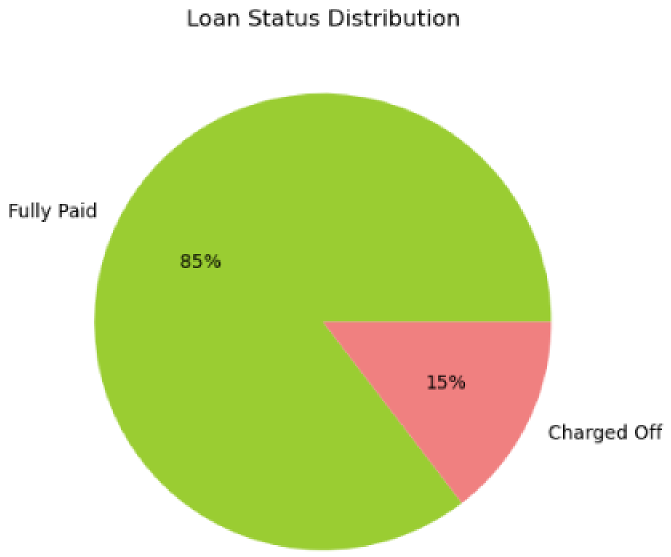


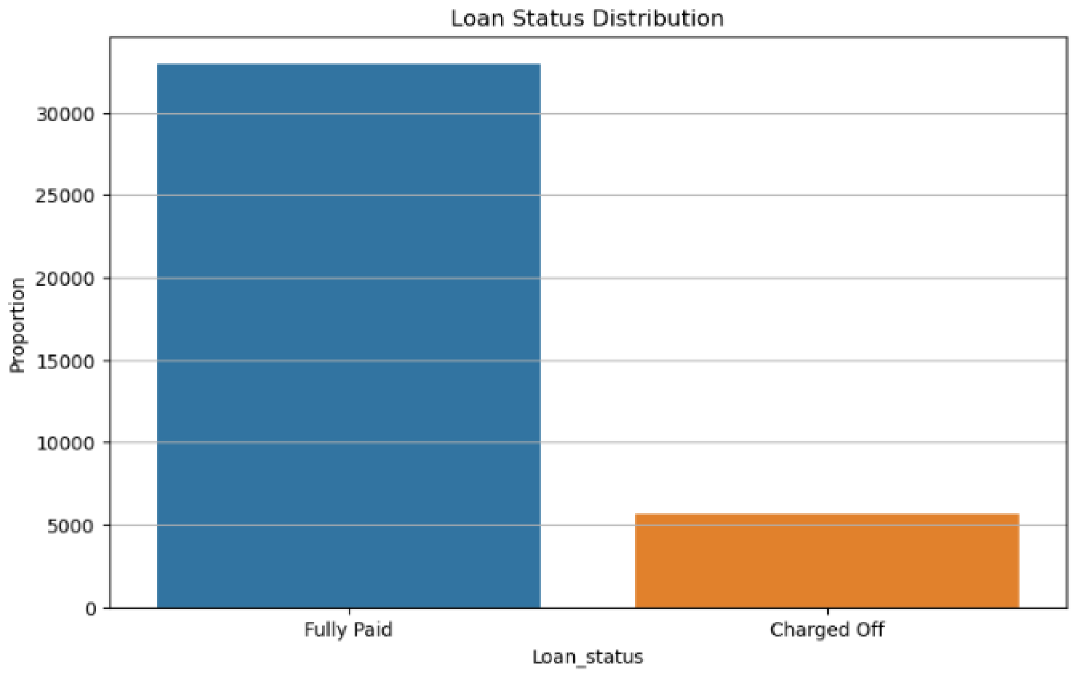
Observations:
From the above 2 plots and statistics of interest rates we can conclude that most of the interest rates lies between 9% to 14.5%. Some people took loan at higher rates of interest i.e., 22.5%

Debt to Income distribution

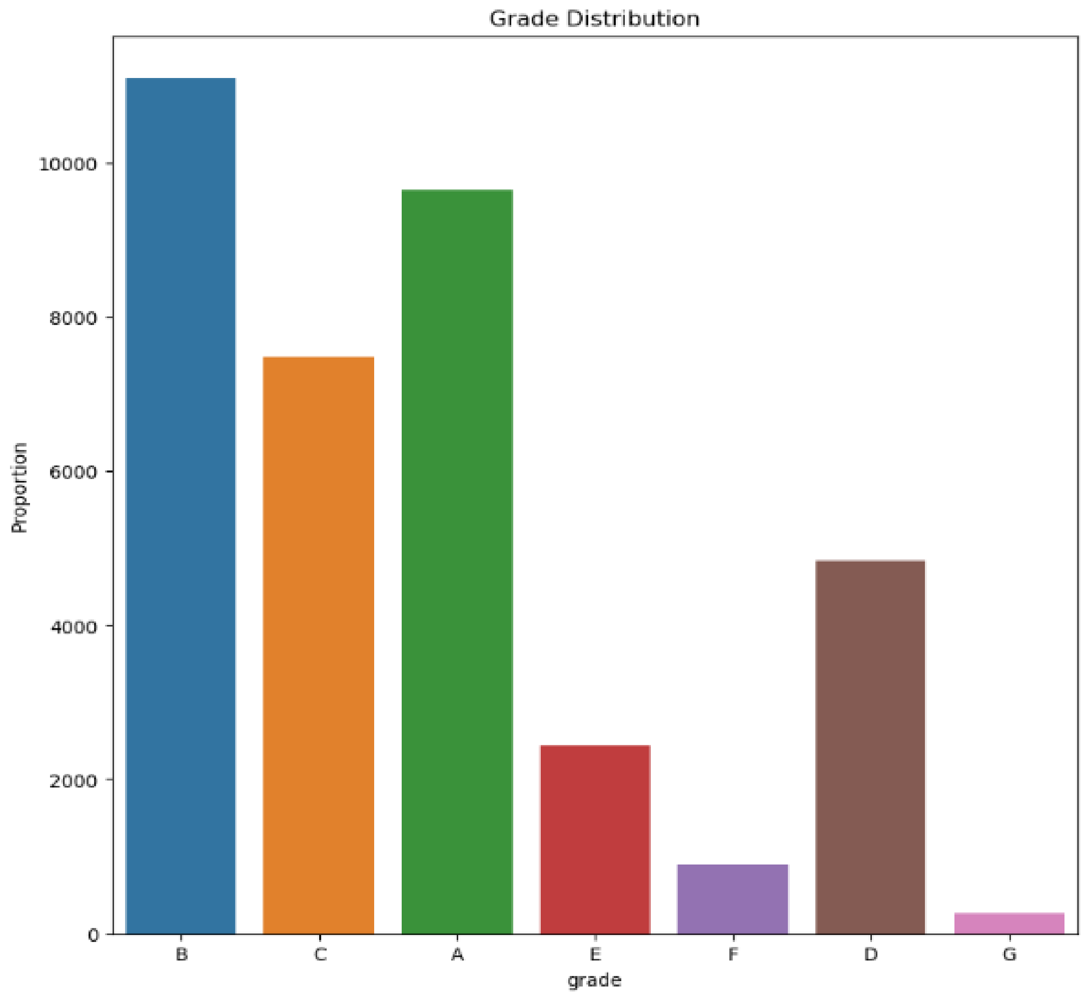


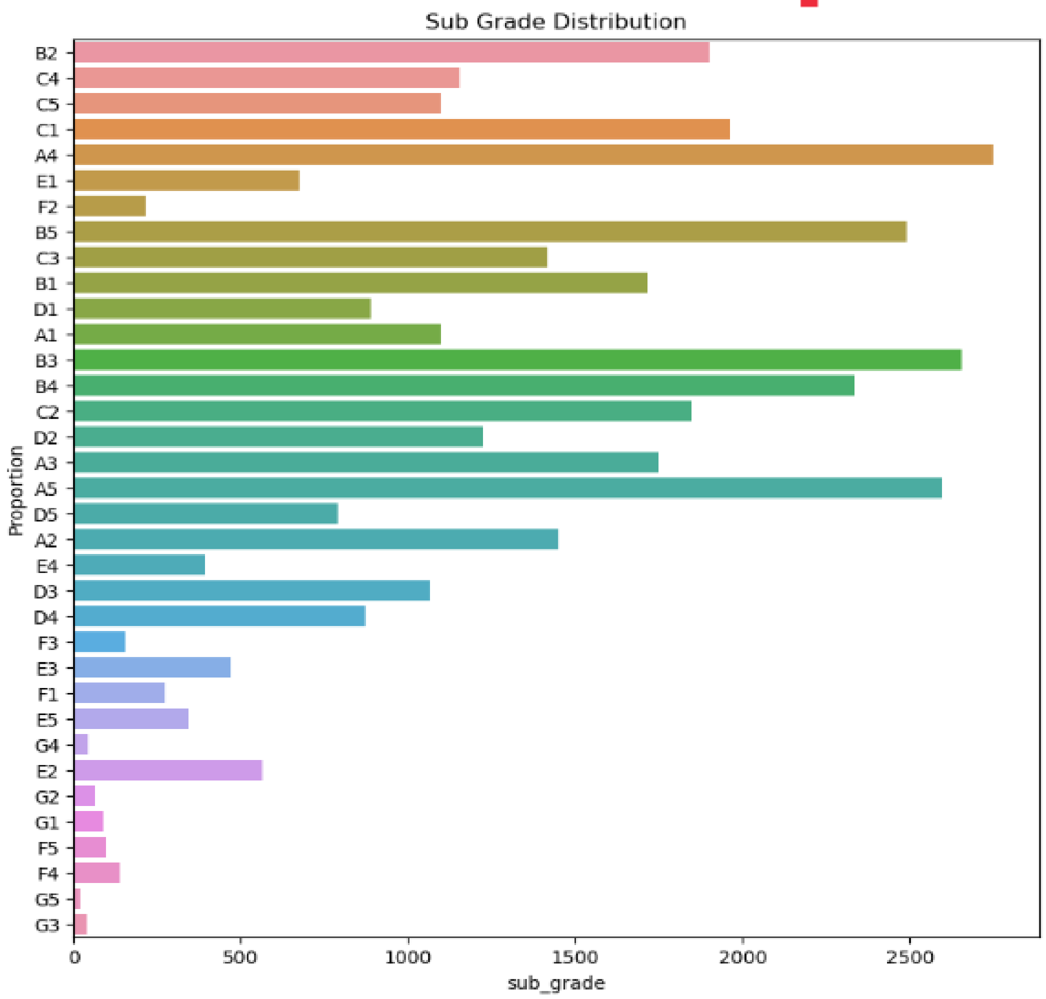
Observations:
There are no outliers, which means that all the loans are given to borrowers who have Debt to Income ration less than 30.





Observations:
85% of borrowers has paid the loan fully. whereas 15% are defaulted the loan.



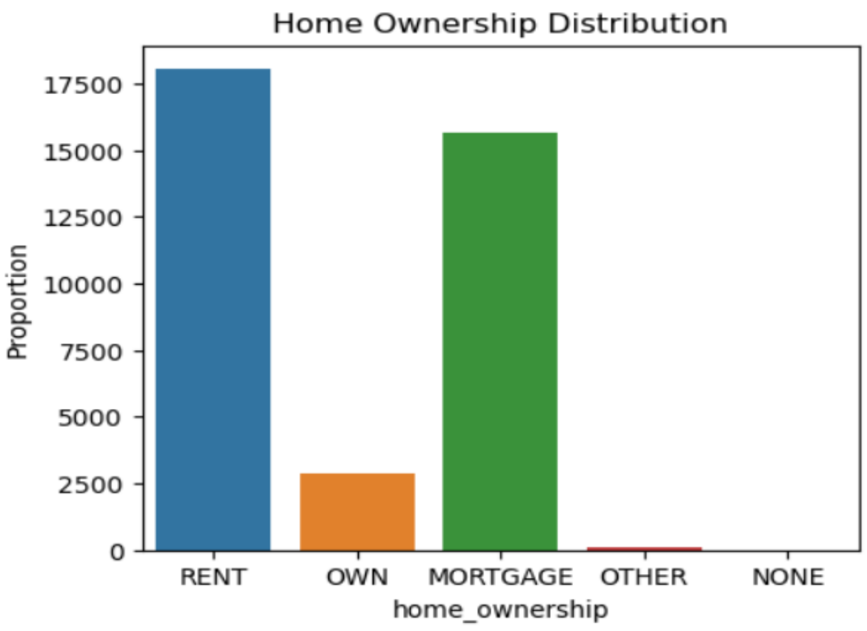


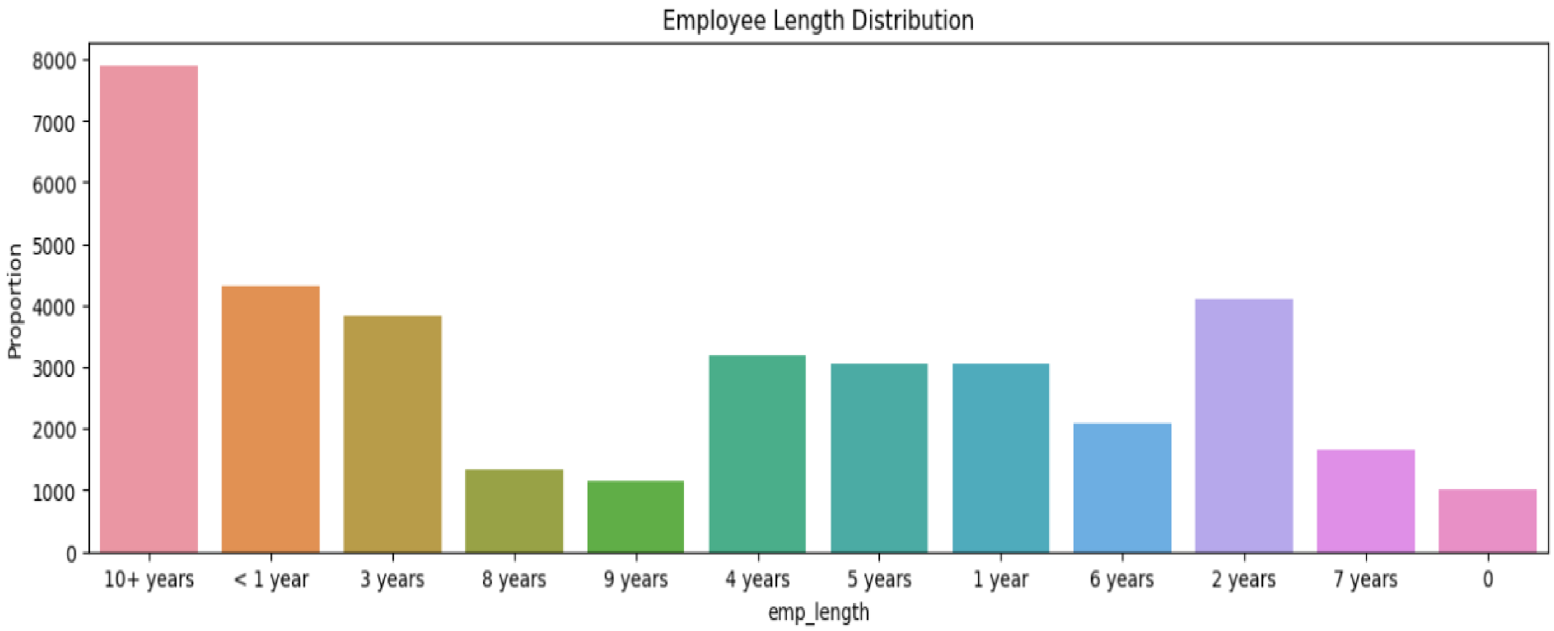
Observations:

Most of the Borrowers lie in A & B grades.

As grades are decreasing, charged off proportion is increasing.

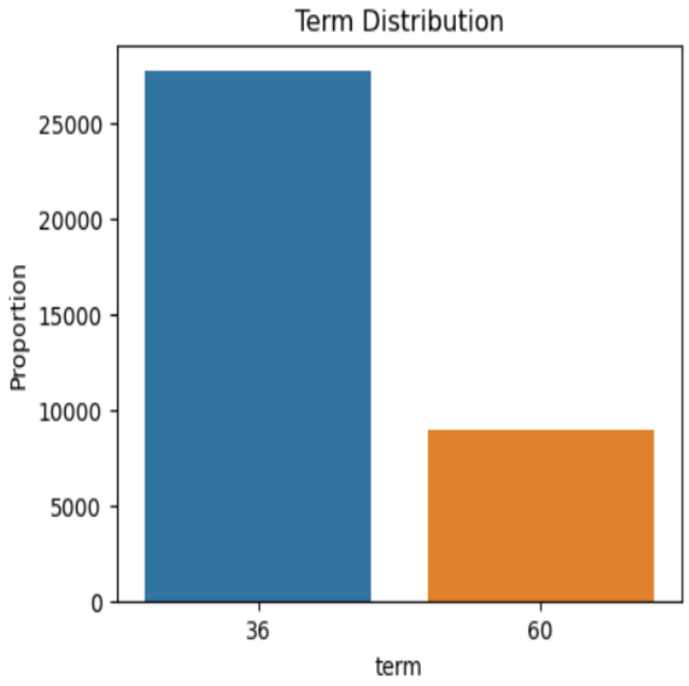
So, we can say that lower the grade more is the chance of getting defaulted due to high loan amount and high interest.

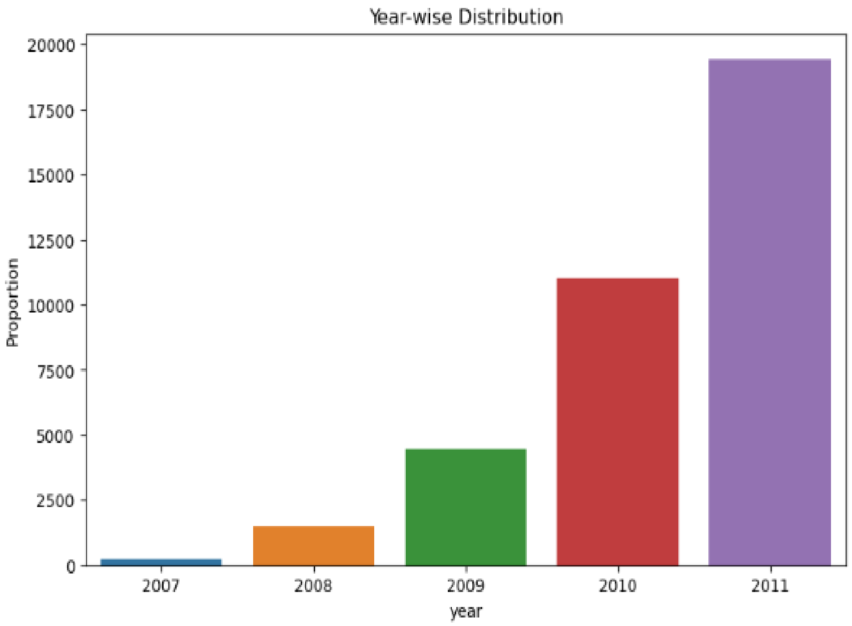


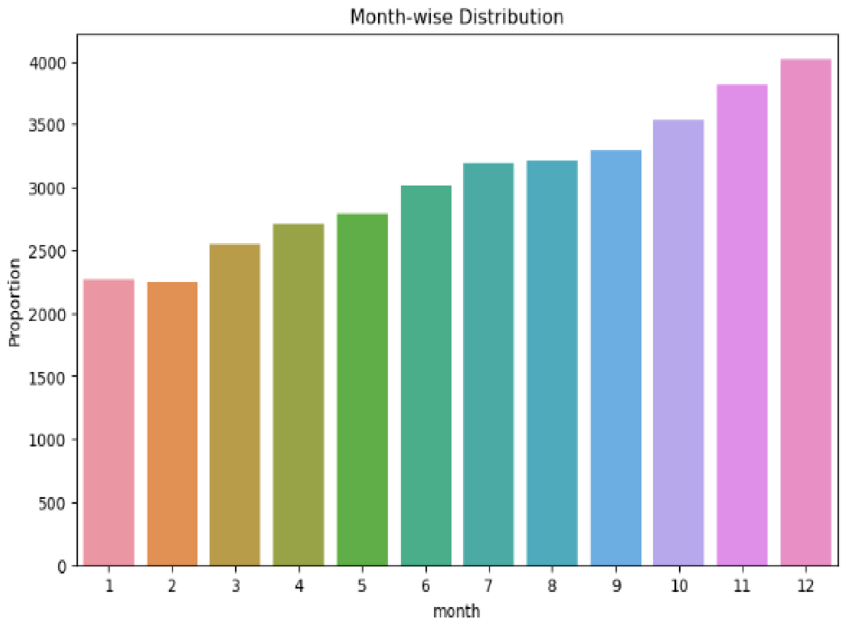


Observations:
The loan borrowers are more having rented and mortgage houses than owning their own house.

Observations:
Borrowers are mostly 10+ year employment length.

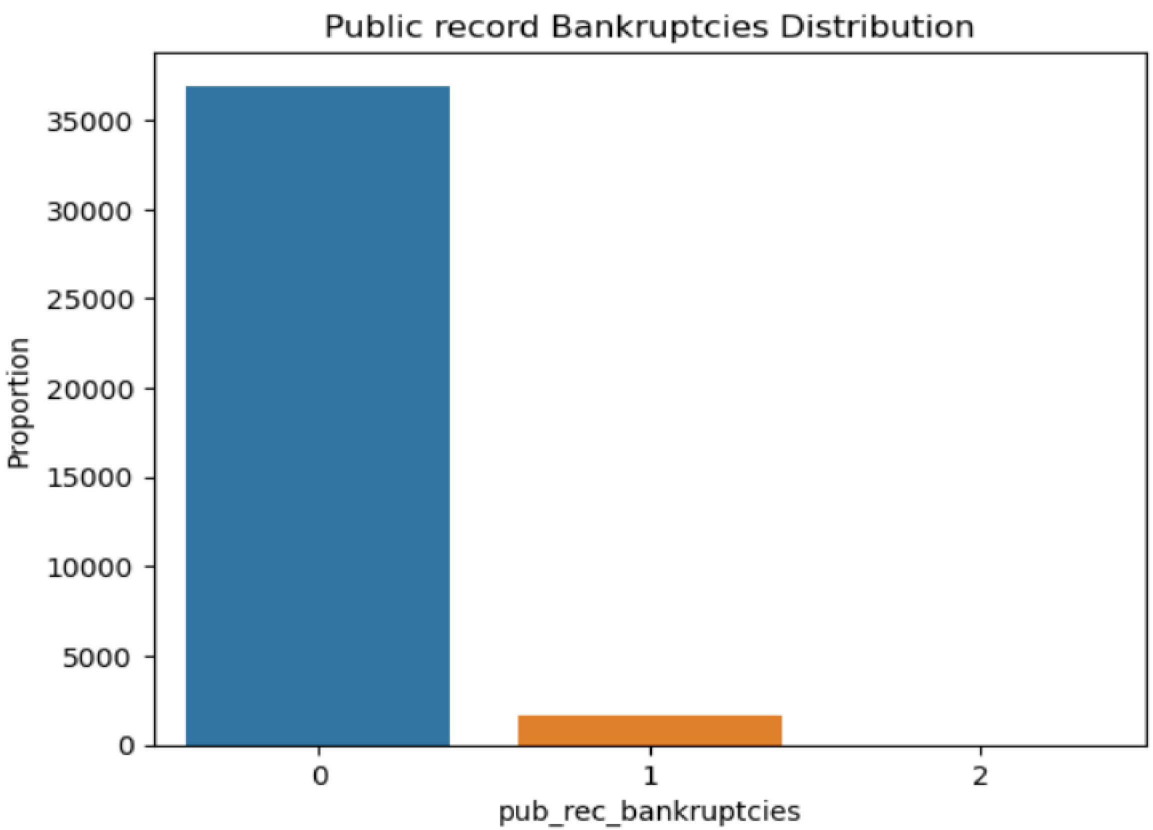


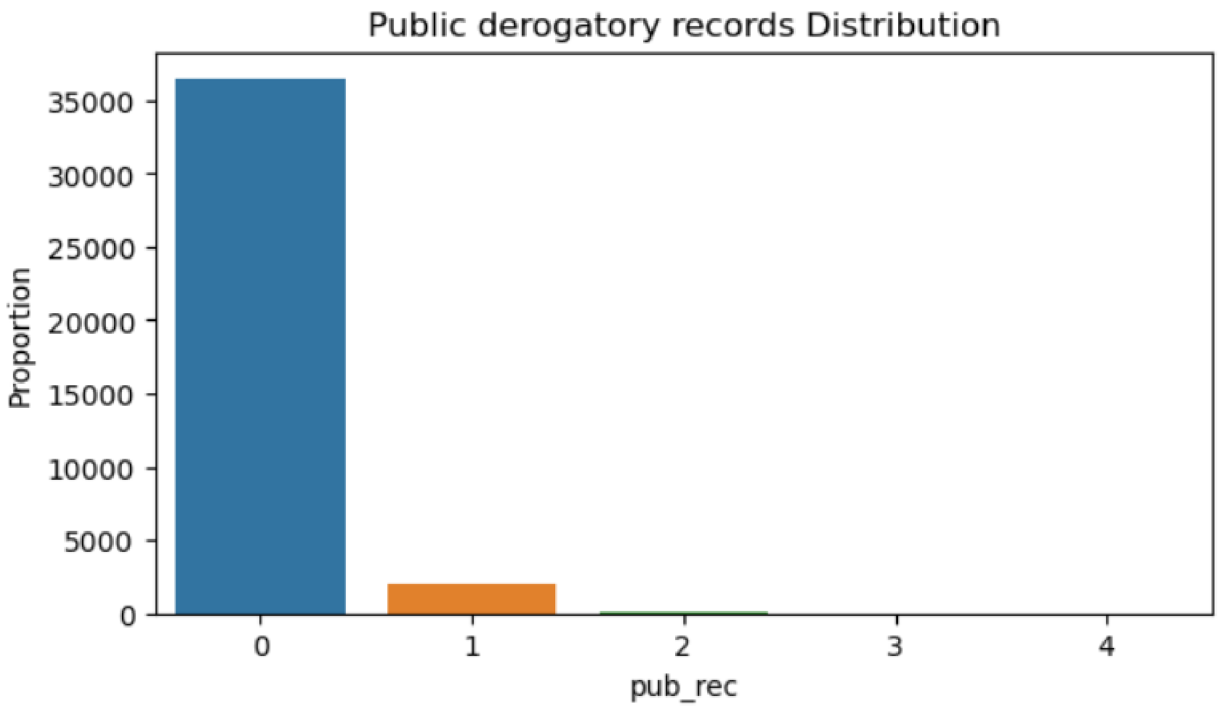




Observations:
There are only two loan terms 36 and 60 months. Around 75% borrowers took loans with 36 months term.

Observations:
The Lending Club has doubling up loan issue every year.
Loan issue gradually increases month-wise. There are more issues of loan in last 3 months every end of the year.

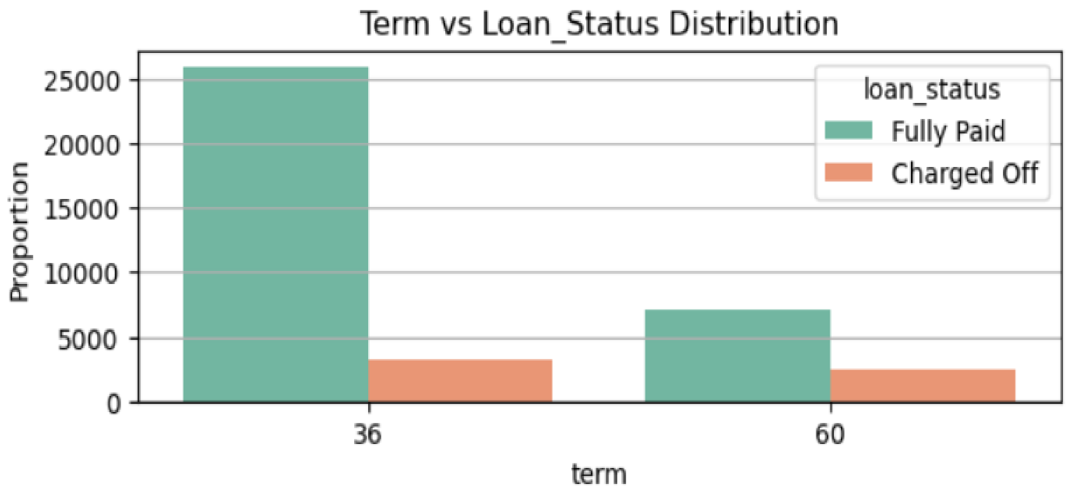


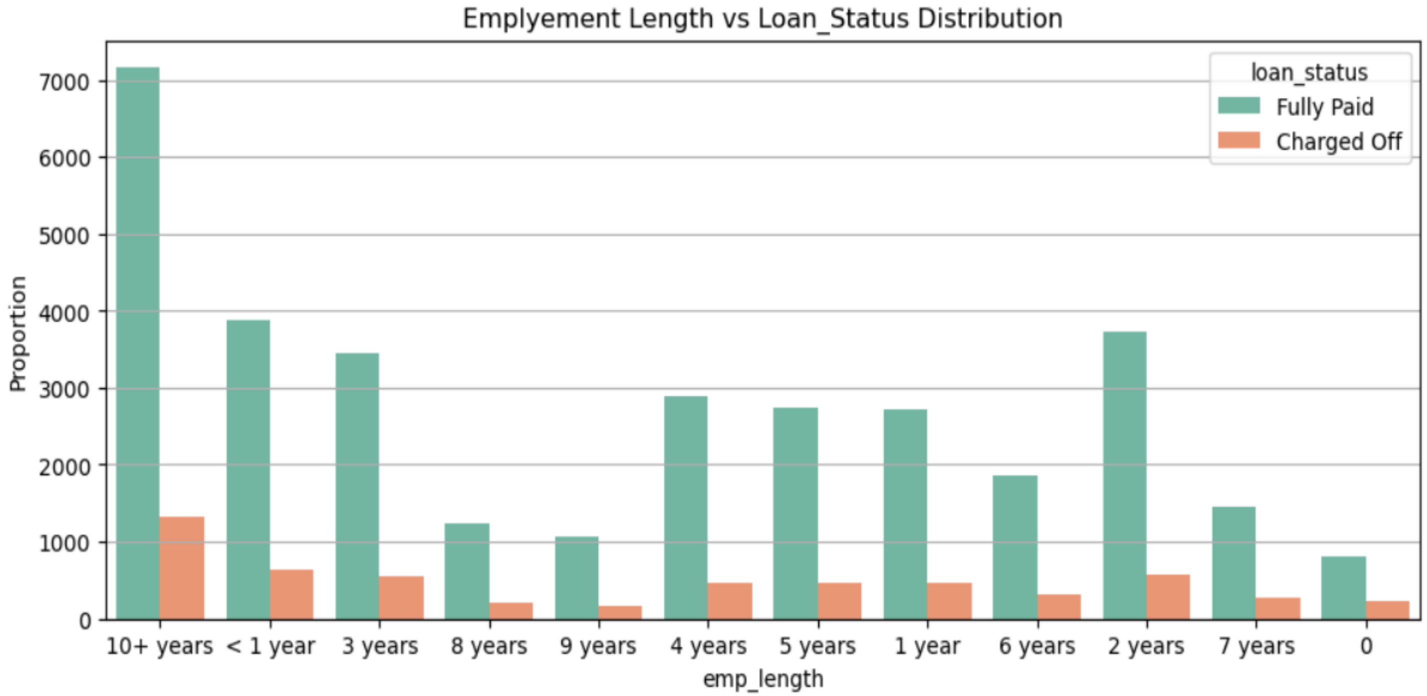


Observations:
Around 99% doesn't have Public Record Bankruptcies.

Observations:
Around 95% borrower's are having no public derogatory records.

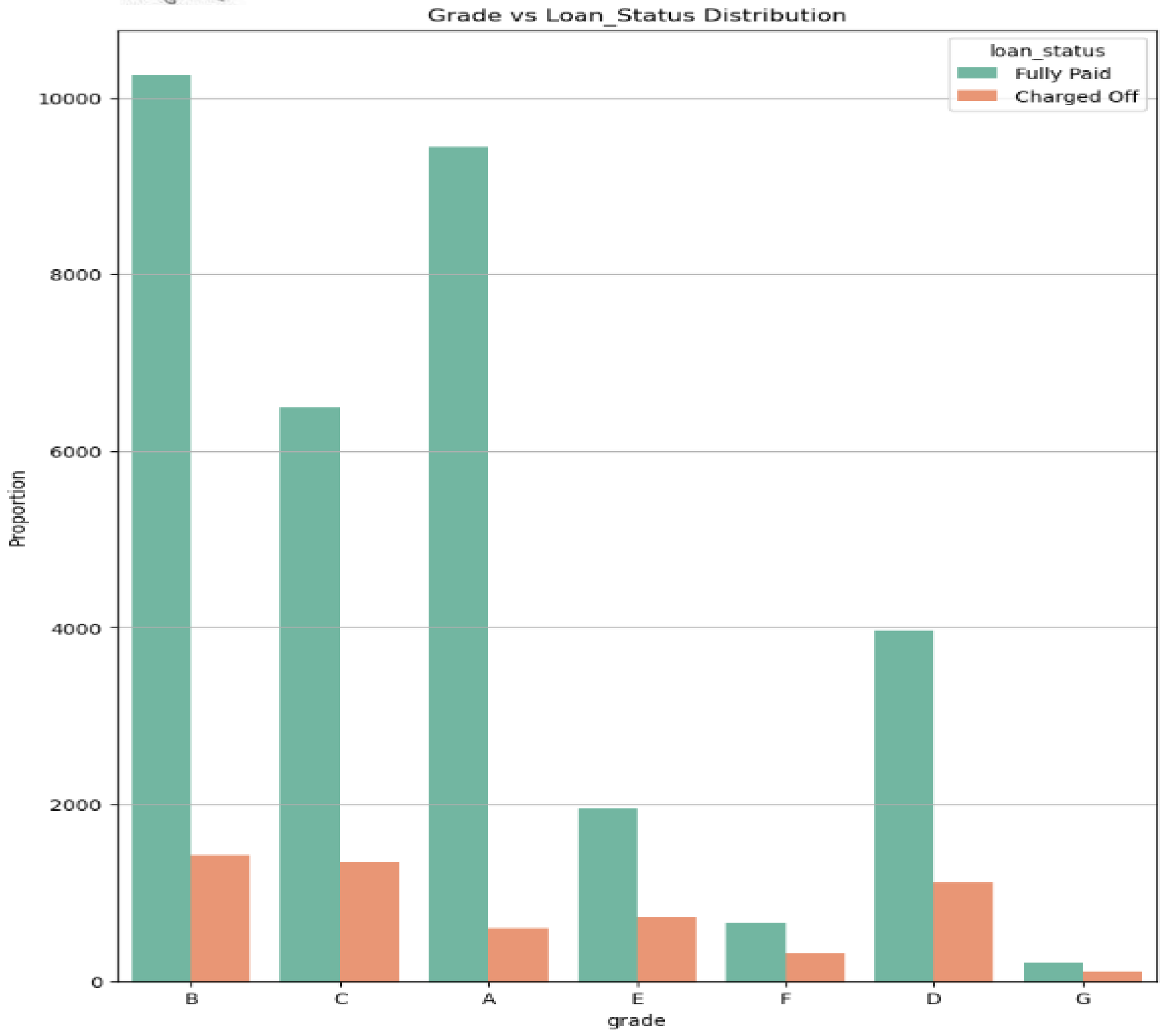
Bivariate Analysis :-

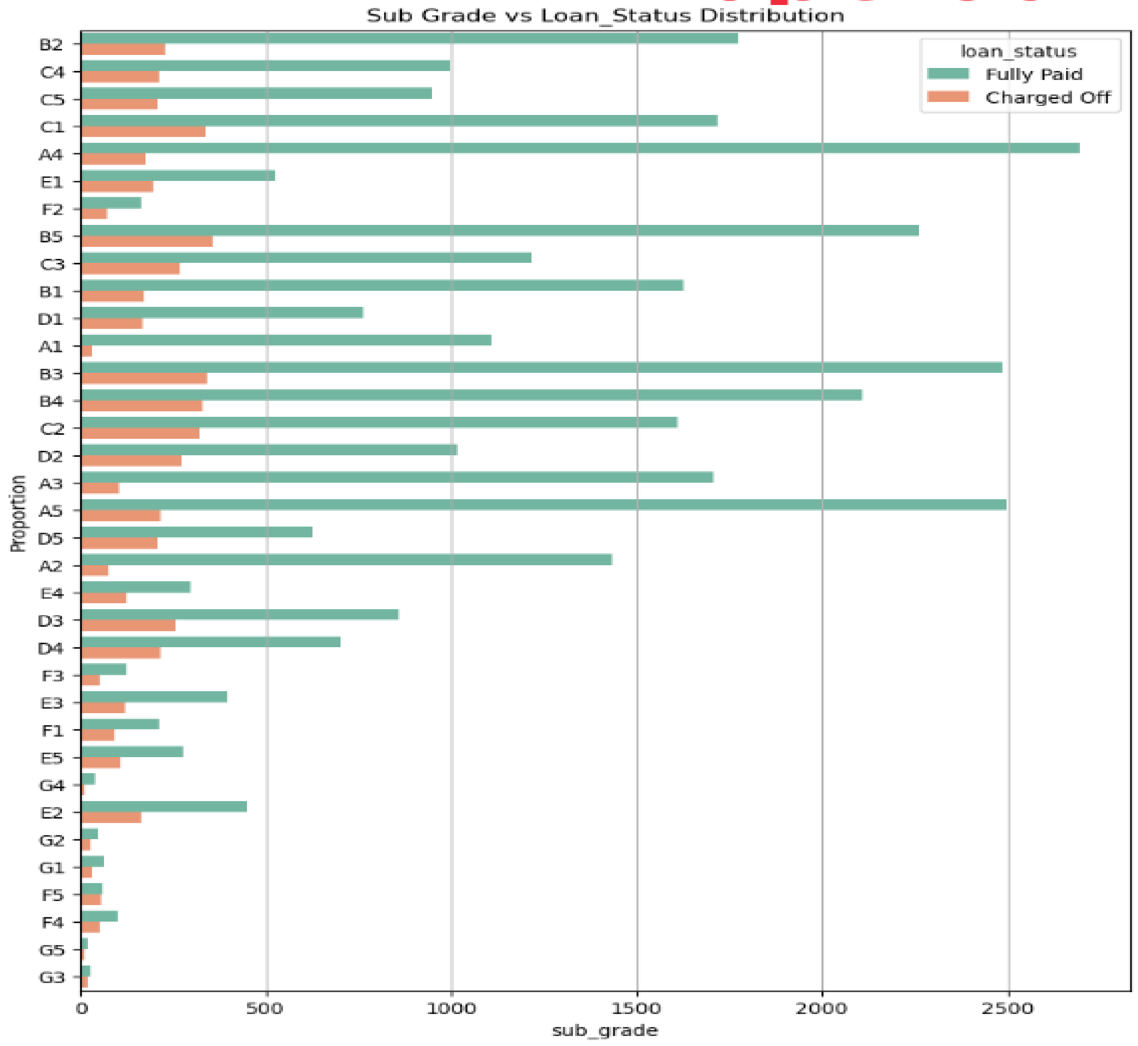




Observations:
36 month loan default is more compared to 60 month.

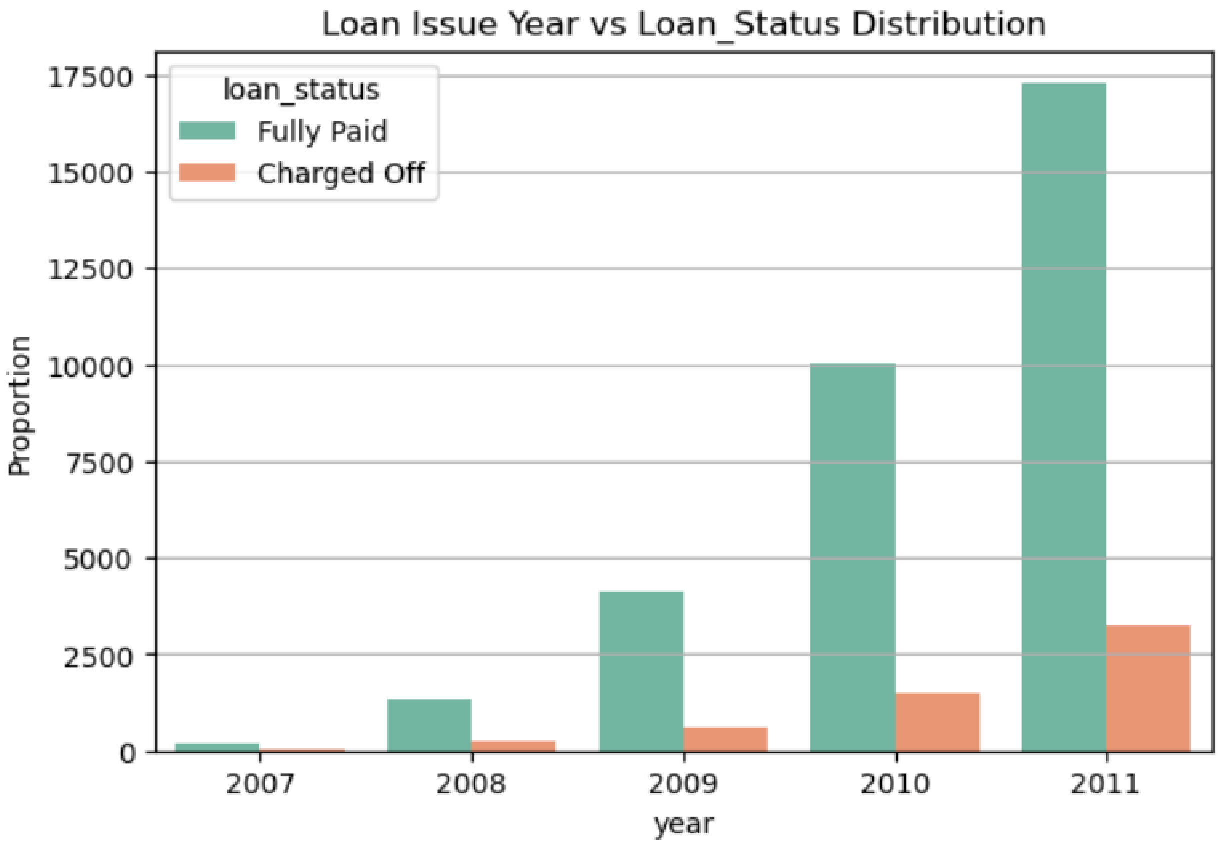
Observations:
Maximum loans are issued for category 10+ years and there are the maximum defaulters.

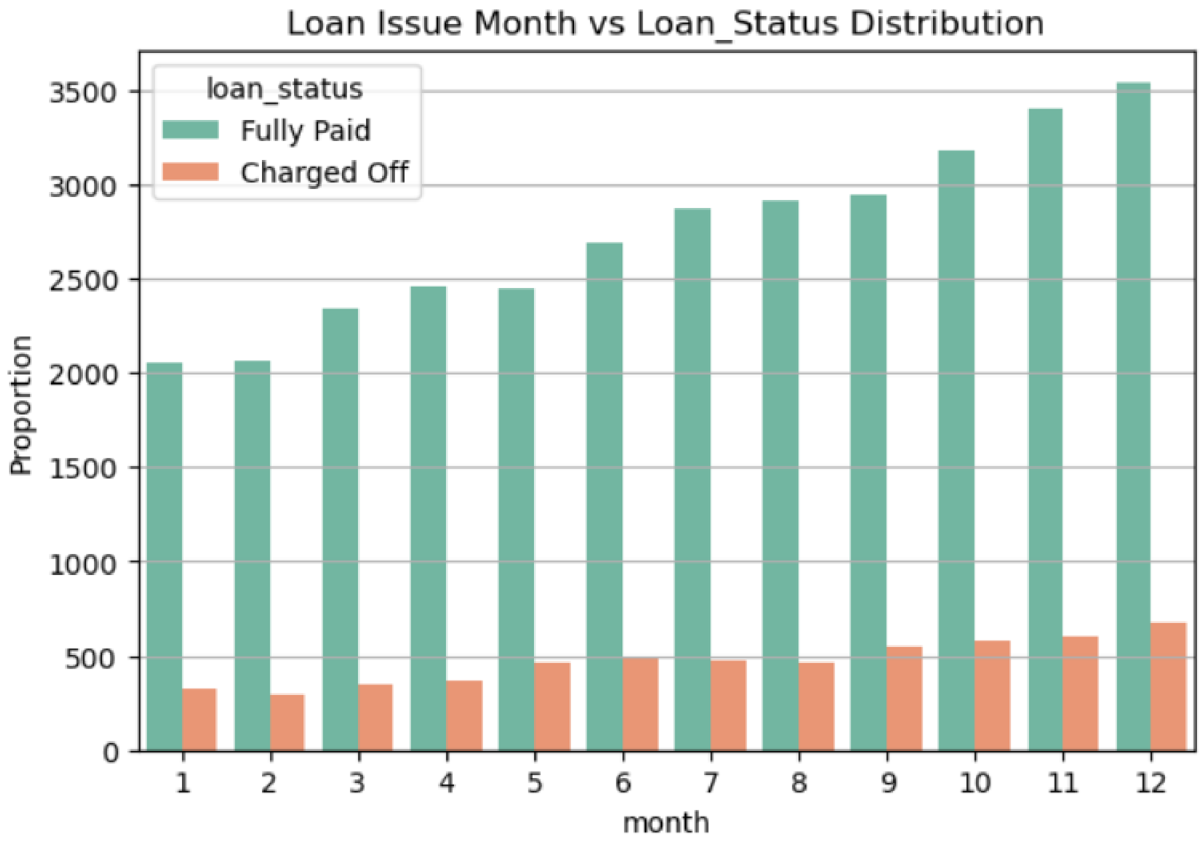




Observation:

- Based on the counts, Grade B, C and D are top three in Charged Off
- Based on the counts, Grade B3,B4,B5, C1,C2, D3, D4 top sub grades in Charged Off



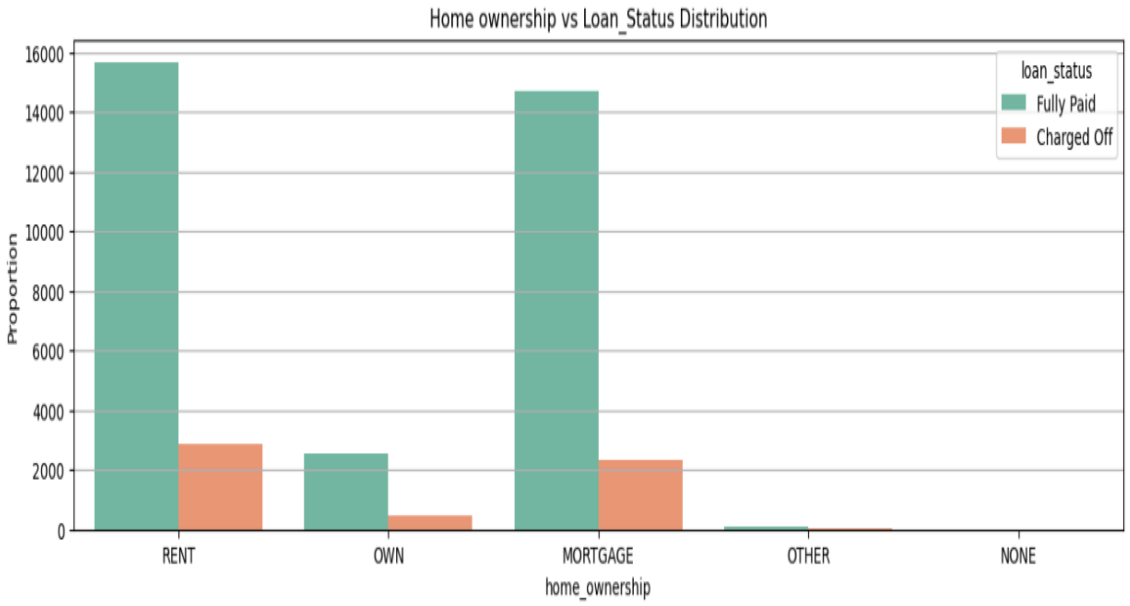


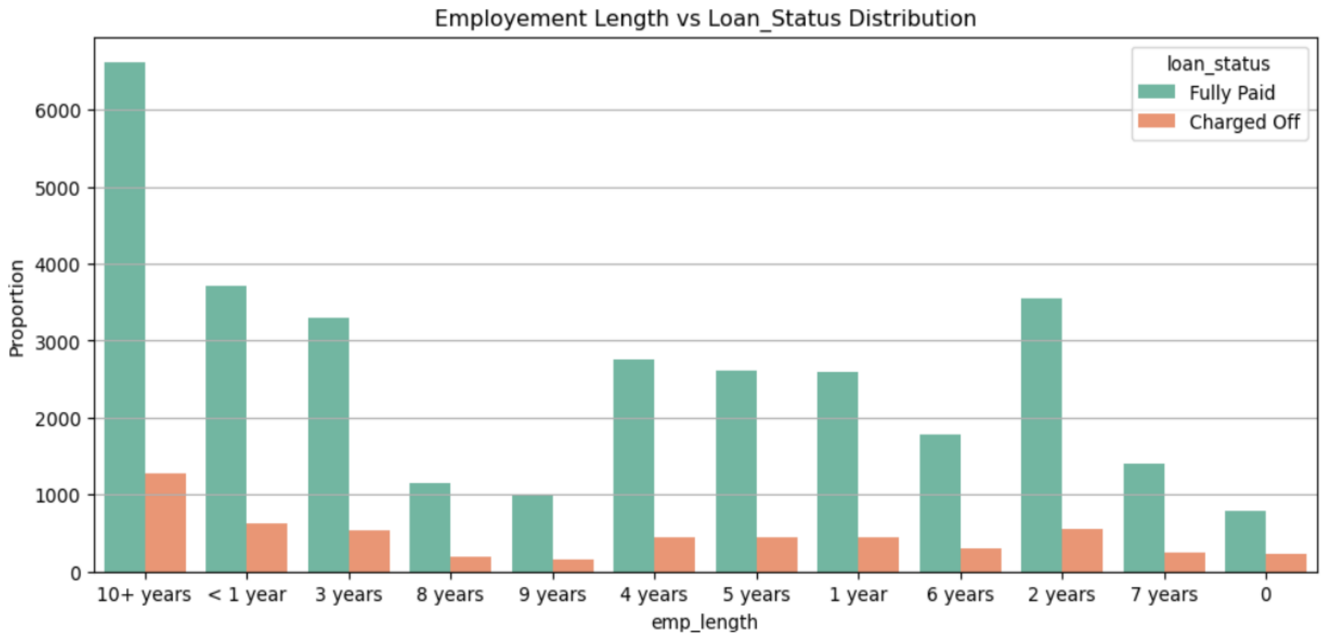
Observations:

- Plot of loan issue year shows maximum loans were taken in the year 2011
- Also high loans are being Charged Off in 2011

Observations:

- Plot of the loan issue month shows maximum loans were given in the month of Oct, Nov, Dec.
- Also high loans are being Charged Off for the loans issued in Sep - Dec months



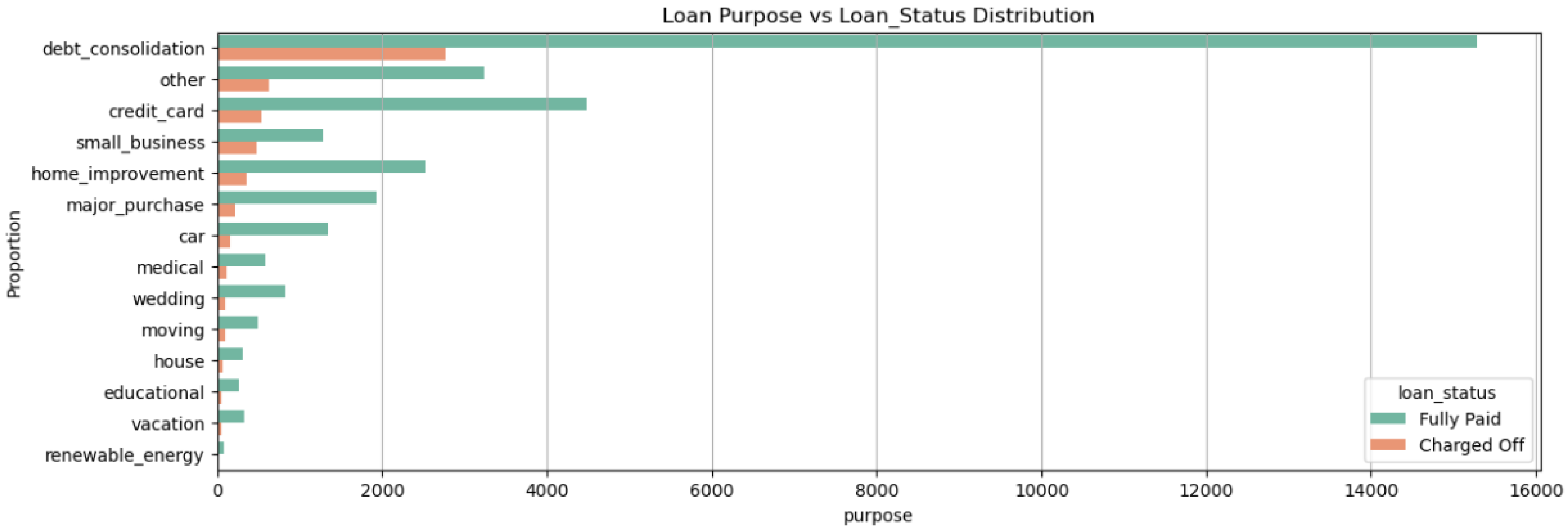


Observations:

People who are in Rent or Mortgage have failed to pay the loan compared with people in Own house

Observations:

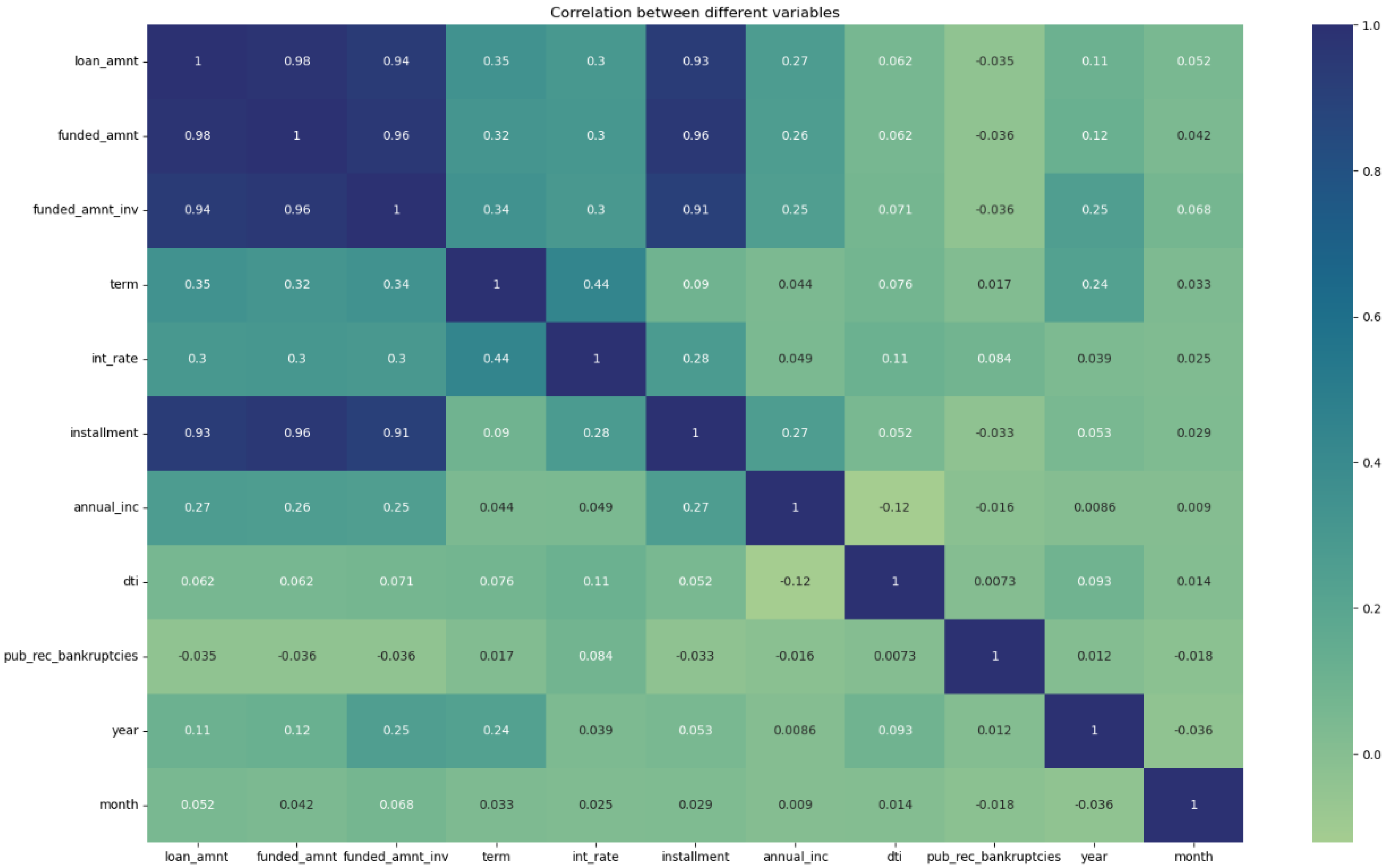
Employment length of 10 years got more loans and they are the maximum defaulters



Observations:

- Loans with purpose debt consolidation, other, credit card and home improvement categories have failed to pay the loan compared with education / renewable energy
- Also debt consolidation is the category where maximum loans are issued.

Data Cleaning and Manipulation :-



Observations:

- Correlation graph shows variables annual_inc, installment, month, loan_amt, funded_amt, year have negative impact on the loan_status variable

Negative Correlation

- loan_amnt has negative correlation with pub_rec_bankrupticies
- annual income has a negative correlation with dti

Strong Correlation

- term has a strong correlation with loan amount
- term has a strong correlation with interest rate
- annual income has a strong correlation with loan_amount

Weak Correlation

- pub_rec_bankruptcies has weak correlation with most of the fields

Recommendations

- Lending club should reduce the high interest loans for 60 months tenure, they are prone to loan defaulted.
- Grades are good metric for detecting defaulters. Lending club should examine more information from borrowers before issuing loans to Low grade (G to A).
- Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans. Lending club should stop giving loans to this category when loan amount requested is more than 12000.
- People with more number of public derogatory records are having more chance of filing a bankruptcy.
- When the purpose is debt consolidation check applicant thoroughly as it has high tendency todefault.
- Lending Club before approving loan should take a look at the DTI(Debt to Income ratio) of the applicants. As DTI of borrowers increases their tendency of defaults also increases.
- Applicants with more revolving utilization rate tends to default more.
- Lending club should be more alert when approving loans for applicants in the lower annual income segment.
- High loan amounts has more chances of Charge-offs. Lending club should properly analyze the applicants demanding higher loan.