LENDING CLUB CASE STUDY

Name : Dhanush M

Anisha Sharma

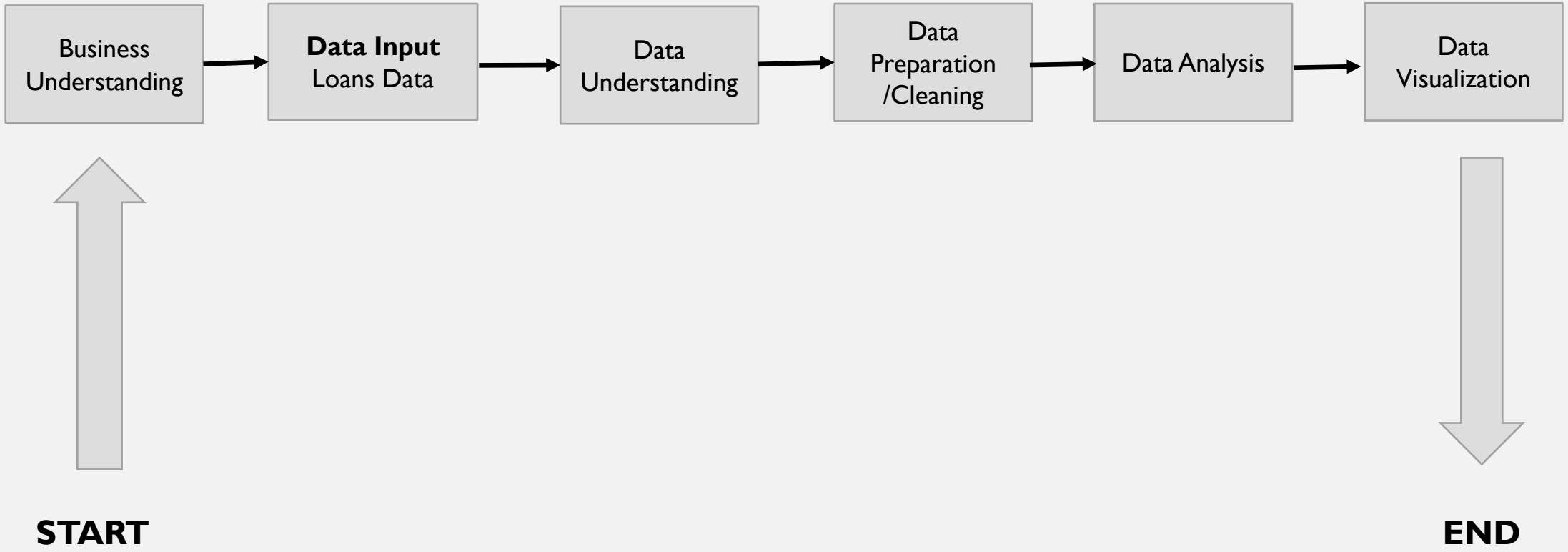
BUSINESS UNDERSTANDING

- The organization is a consumer finance company which specializes in lending loans to urban customers. The company after receiving the application has to make the decision for loan approval based on applicant's profile.
- While taking the decision, two types of risks are associated –
  - ✓ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - ✓ 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 aim is to identify patterns which indicates if a person is likely to default or not.
- 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:
    1. Fully paid
    2. Current
    3. Charged-off
  - ✓ Loan rejected

BUSINESS OBJECTIVE

- Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). 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 the company is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. The aim is to identify such applicants.
- 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 utilize this knowledge for its portfolio and risk assessment.

PROBLEM SOLVING METHODOLOGY



ANALYSIS

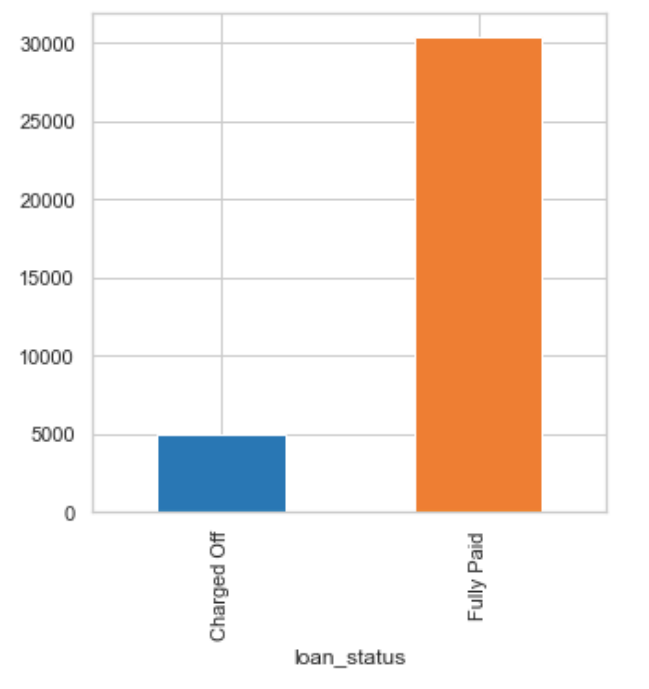
1. UNDERSTANDING THE DATA SET

- The data contains loan applications of multiple customers.
- The data has total **39717** rows.
- The data has total **111** columns.
- We can see the the target variable is '**loan_status**', which we want to compare across other variables.

2. DATA CLEANING

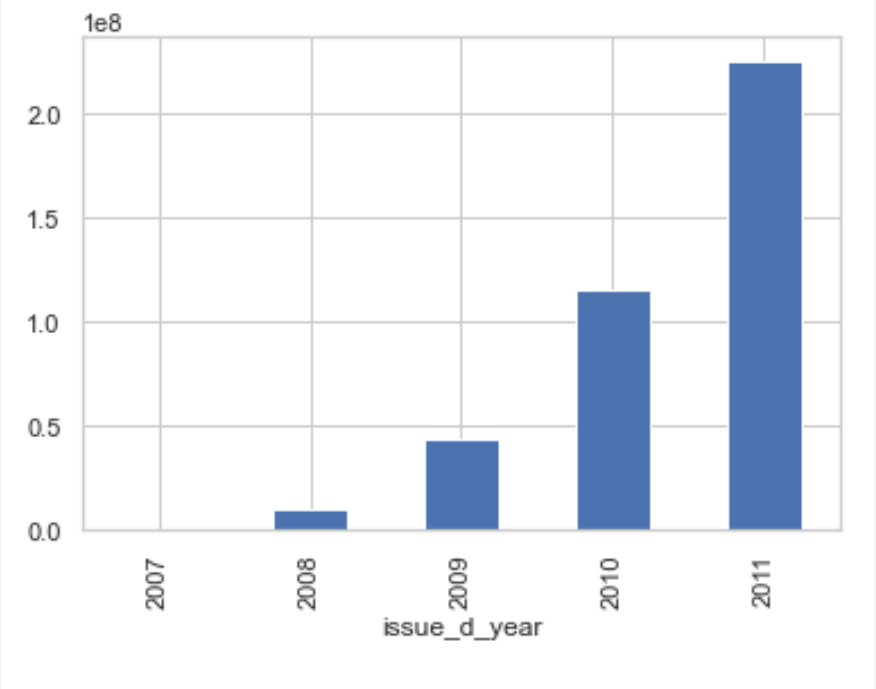
- We can see that many columns have 100% missing values. So, **we dropped all the columns having null values more than 60%.**
- Also, we removed some other columns which will not be required for the analysis according to us. So, we will be using the remaining **38 columns for further analysis.**
- Some columns of object type and date type were converted to the desired(numeric) format like **emp_length, int_rate, issue_d, last_pymnt_d,etc.**
- Some new columns were made for separating the month and year for further analysis.
- After all the above steps, we had **42 columns** on which we will do our further analysis.

RESULT



STATUS OF THE LOAN

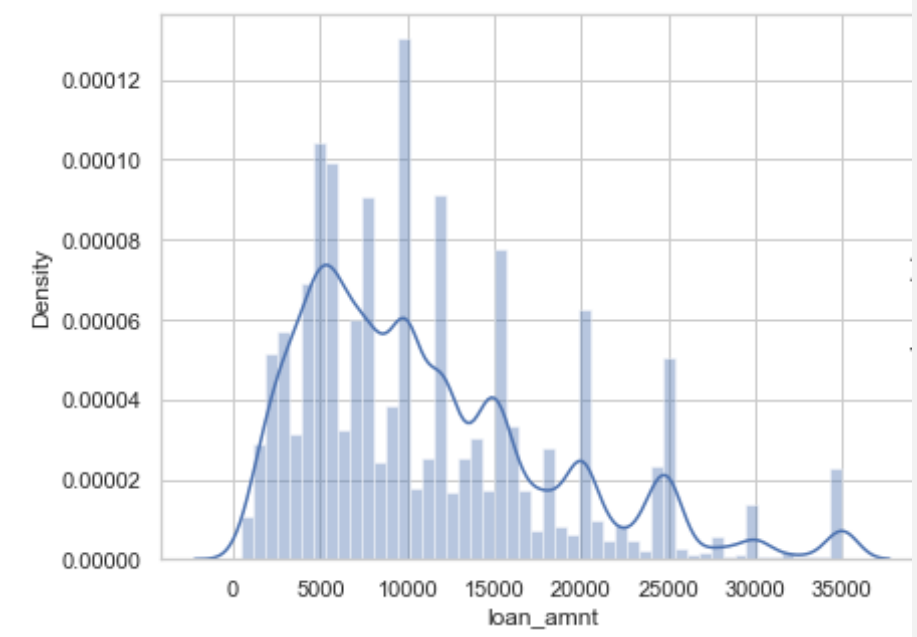
The number of loans fully paid is marginally more as compared to loans charged-off

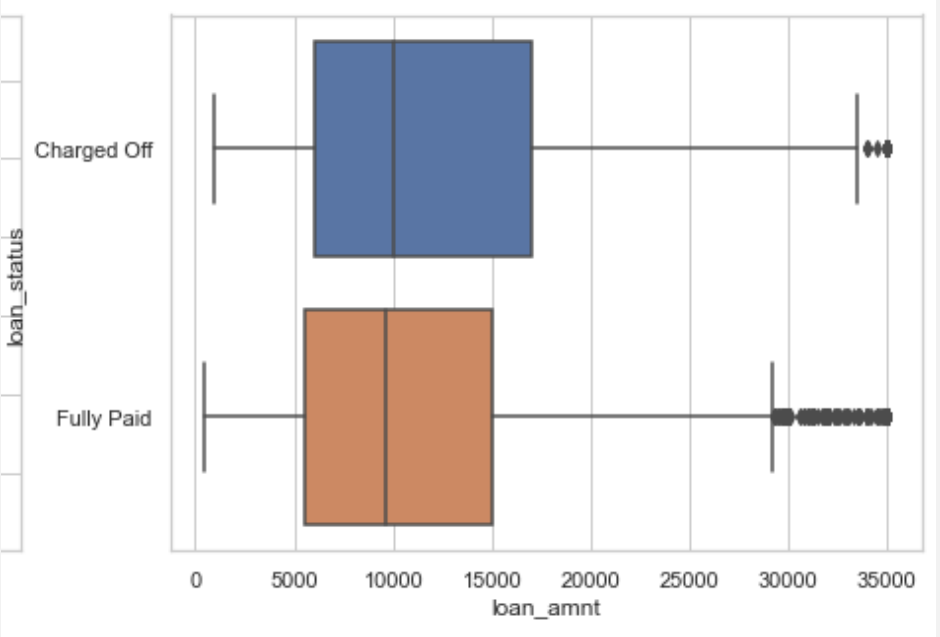


TOTAL LOAN AMOUNT LENT IN EACH YEAR

The maximum number of loans are given in 2011

RESULT





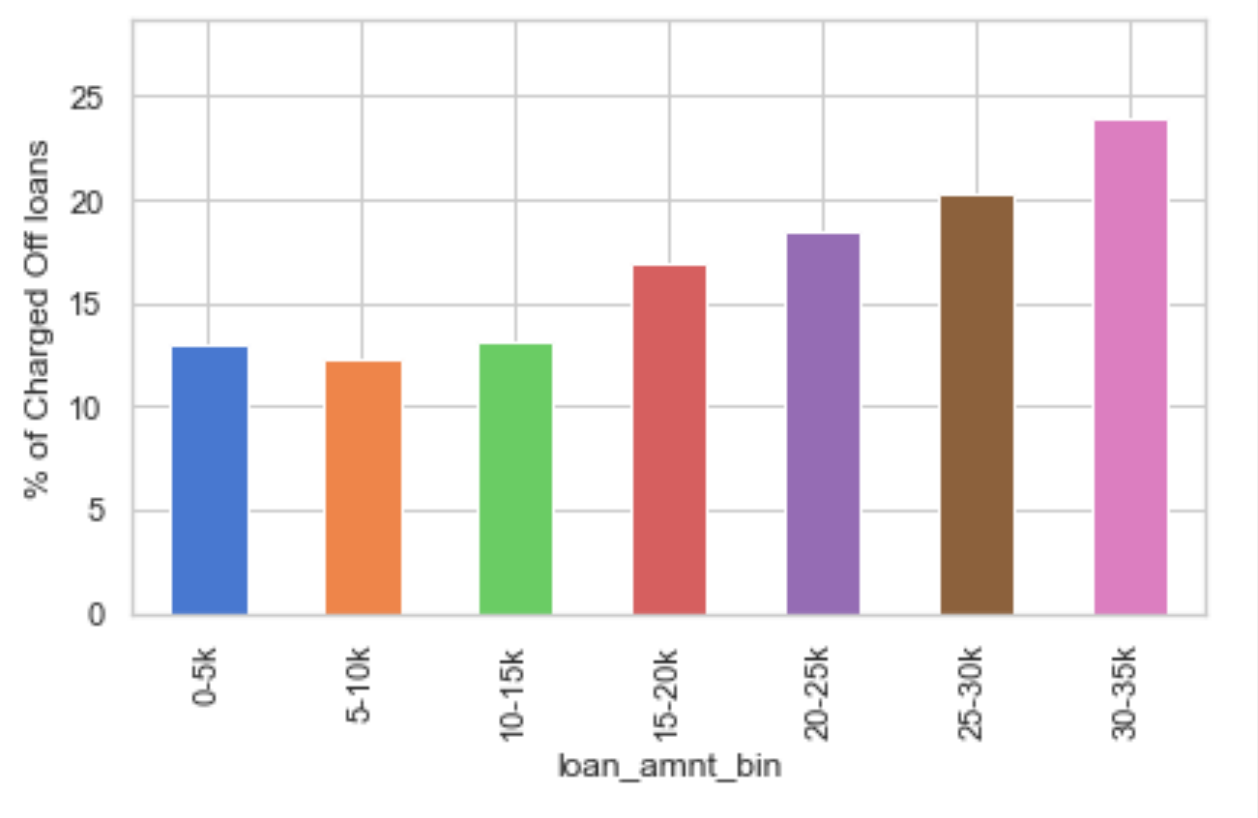
LOAN AMOUNT DISTRIBUTION

Most of the loans granted are below 15000 (75 percentile value)

Funding amounts see a spike around each 5000 boundary.

Charged off loans are shifted towards higher average loan amount request.

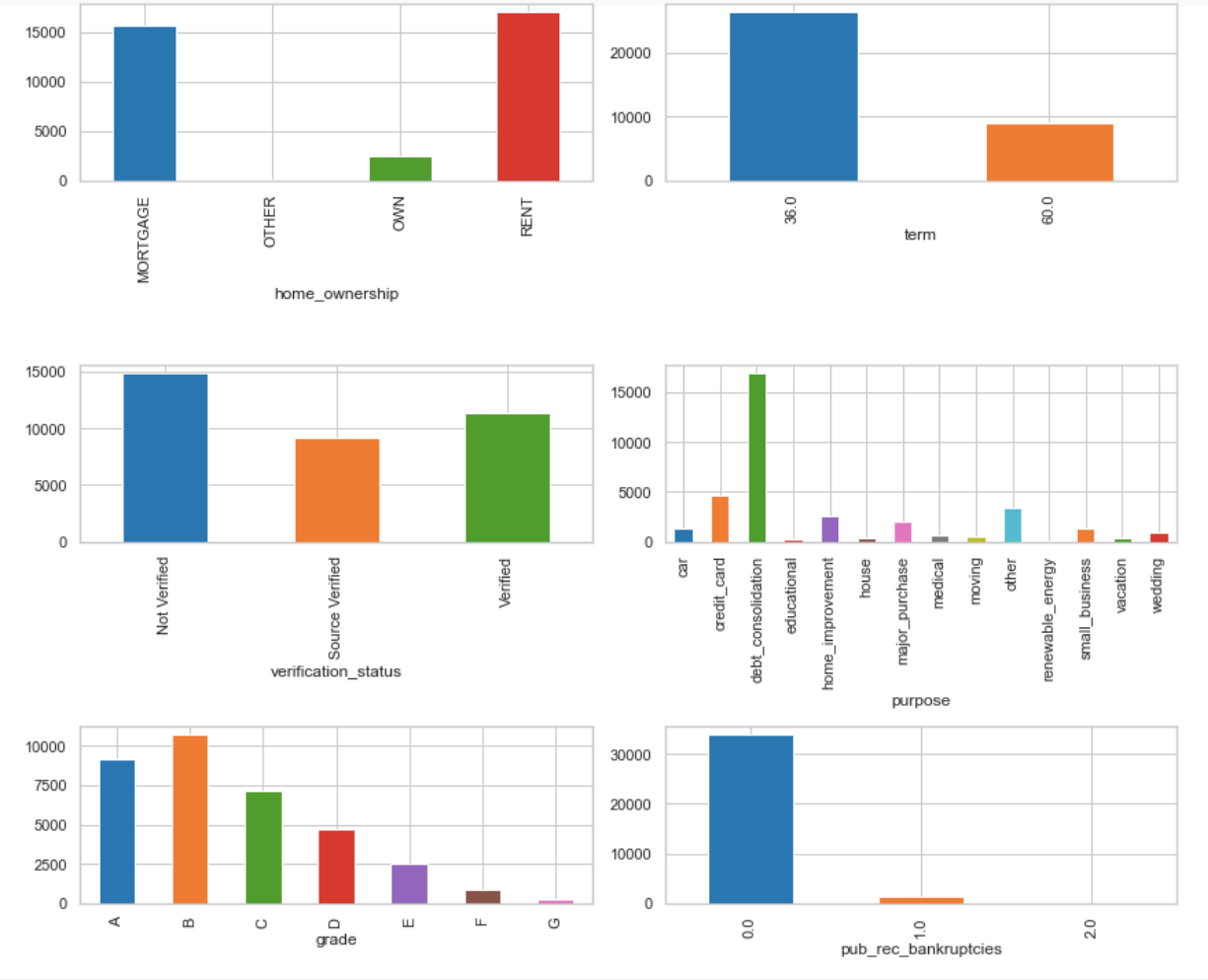
RESULT



DEFAULT RATES ACROSS VARIOUS LOAN AMOUNT

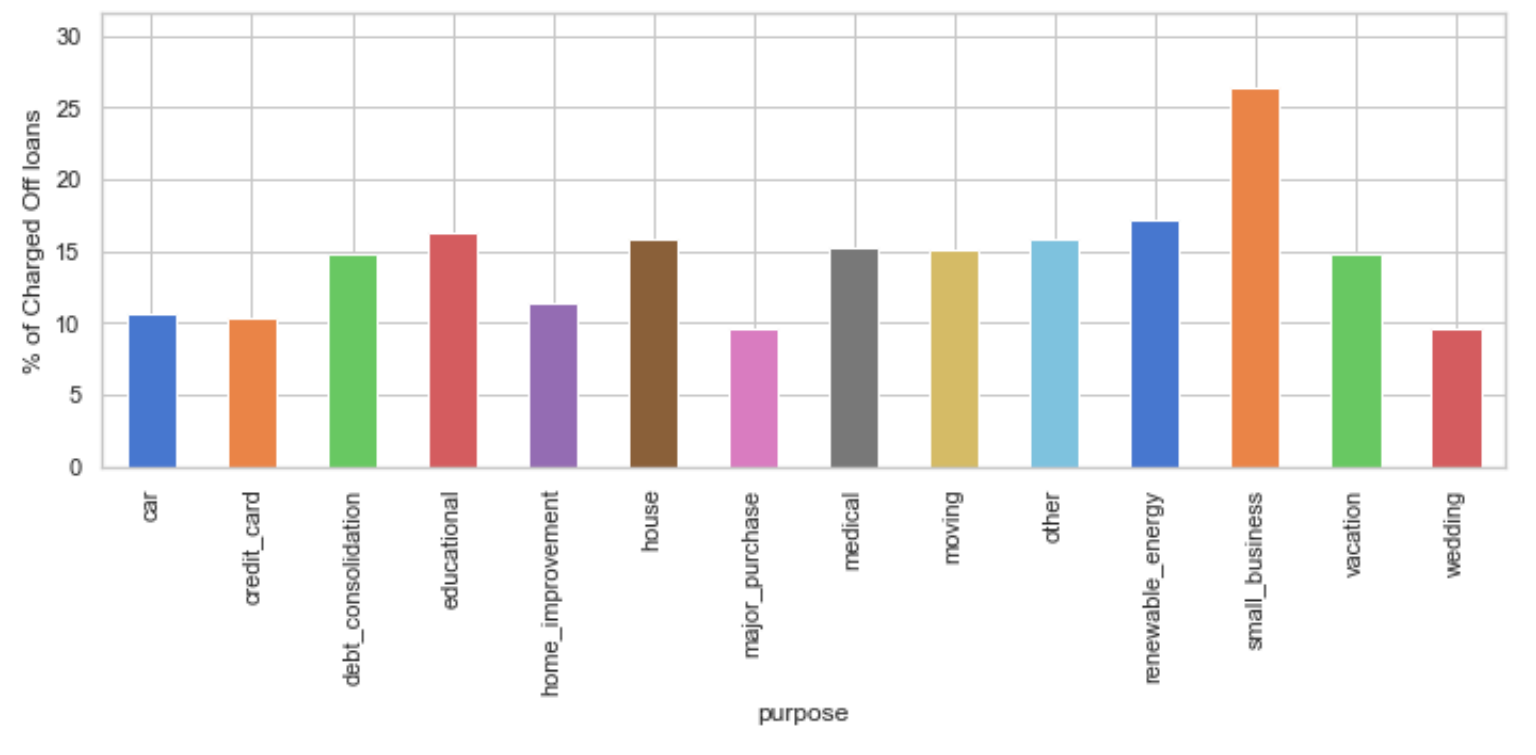
Most loans are below 20000 amount. The higher loans, though lesser in number, but higher risk of default

RESULT



- *Most approved loans are high grade(A,B,C)*
- *Most of the loans are issued for 36 month term*
- *Very few people with prior record of bankruptcy got loan*
- *Most of the people took the loan for debt consolidation*

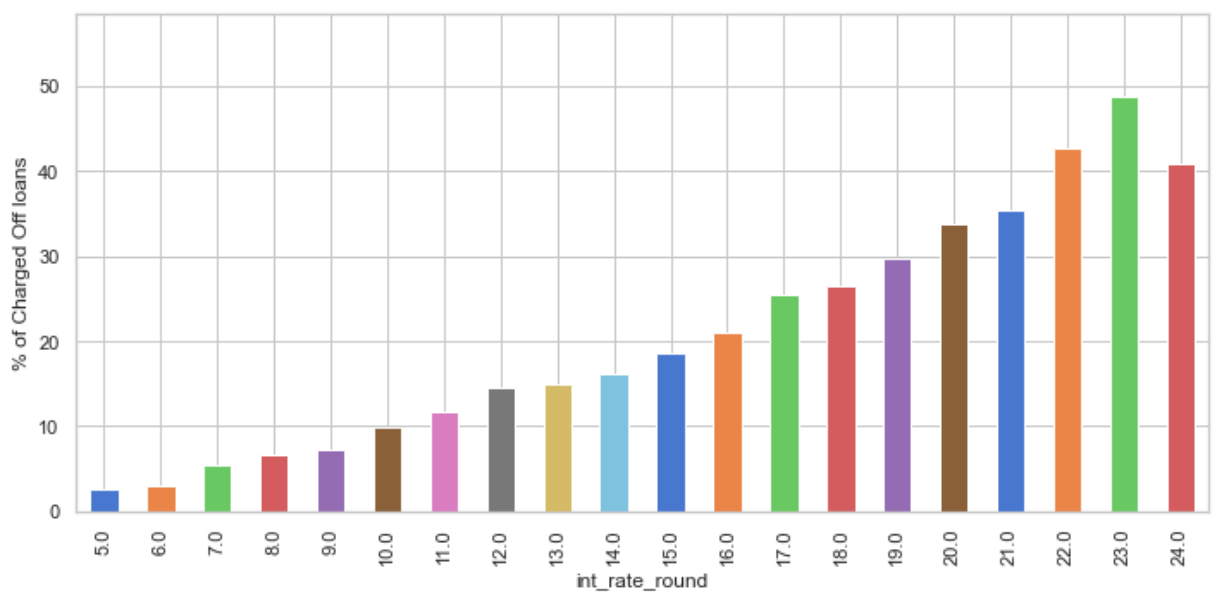
RESULT



DEFAULT RATE ACROSS THE PURPOSE OF LOAN

Small business loans default the most, then renewable energy and education

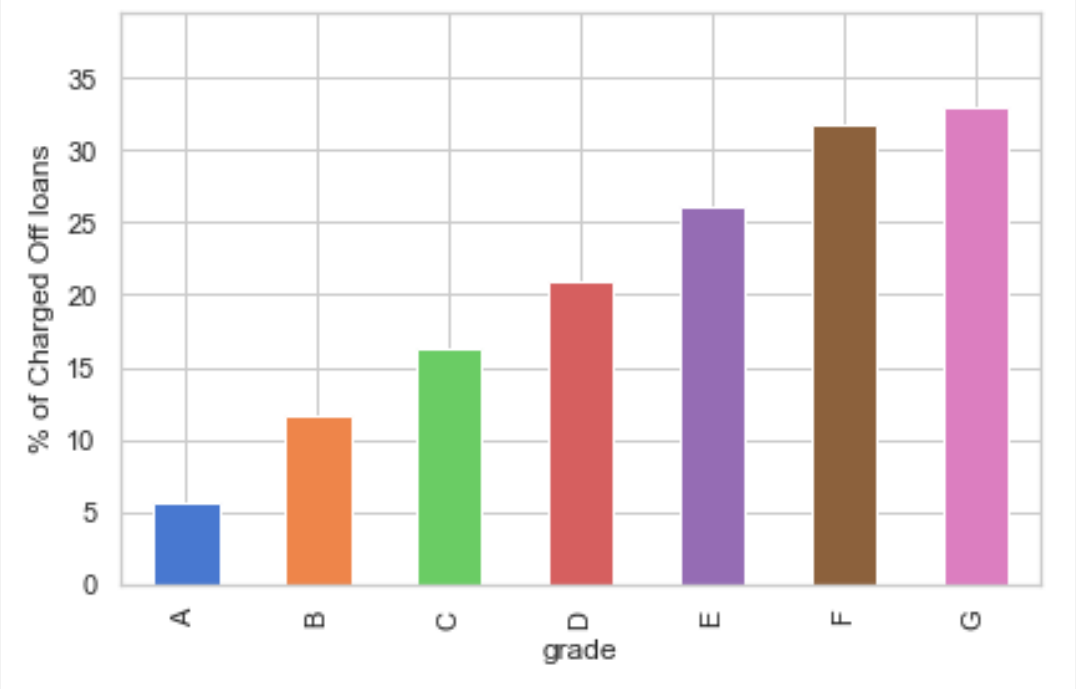
RESULT



DEFAULT RATE ACROSS THE INTEREST RATE

Overall, the interest rate varies from 5.42% to 24.4% with average interest rate of 11.8%.

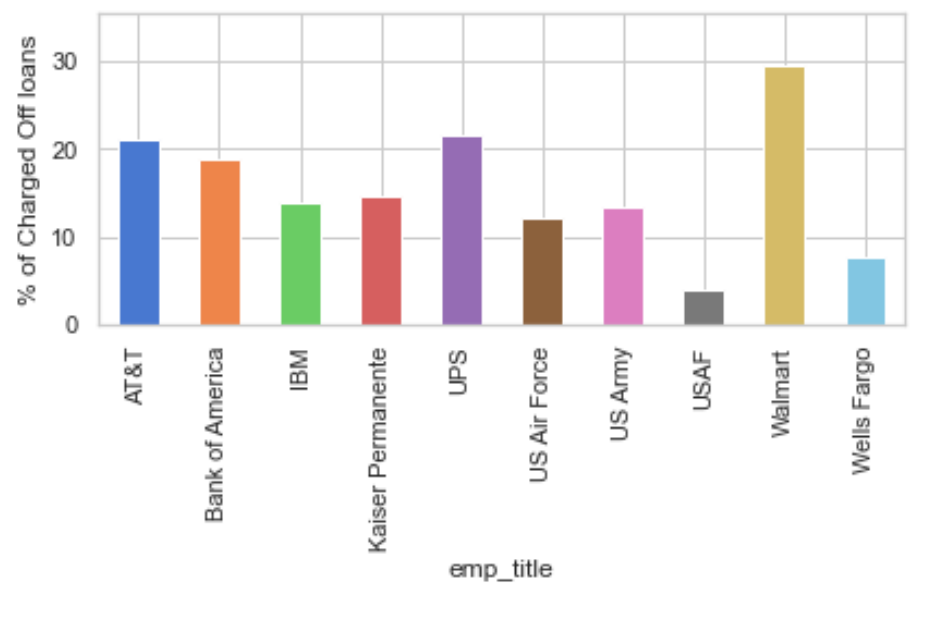
The interest rate for charged off loans appear to be higher than for fully paid.

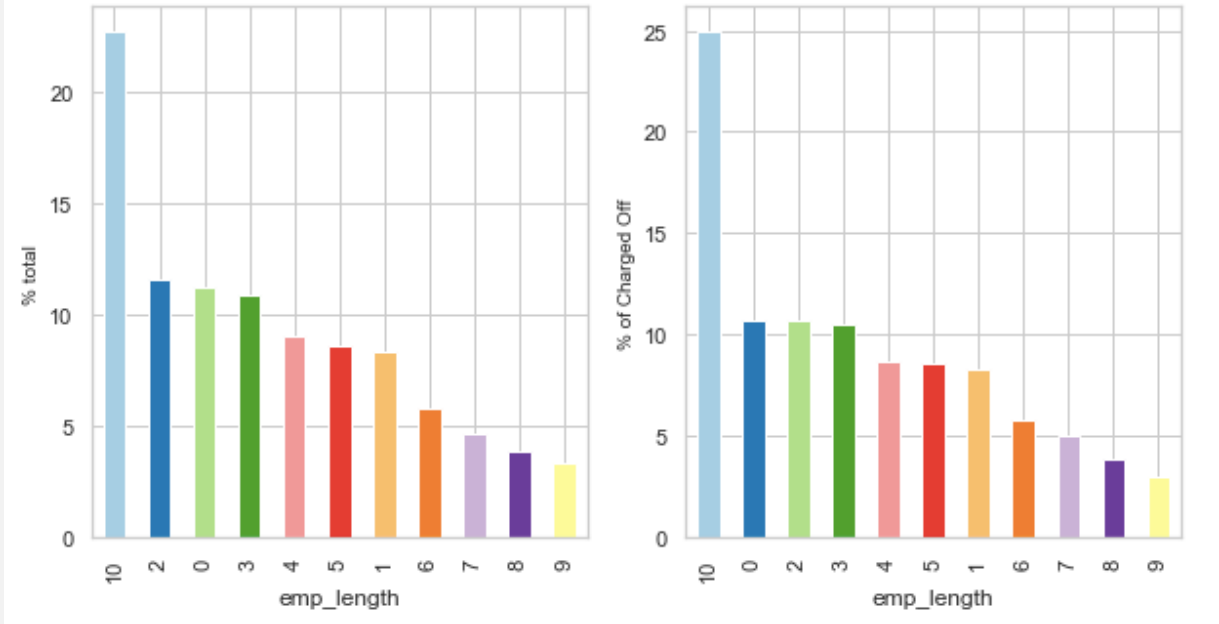


DEFAULT RATE ACROSS VARIOUS GRADES

Grades D,E,F,G are risky loans

RESULT





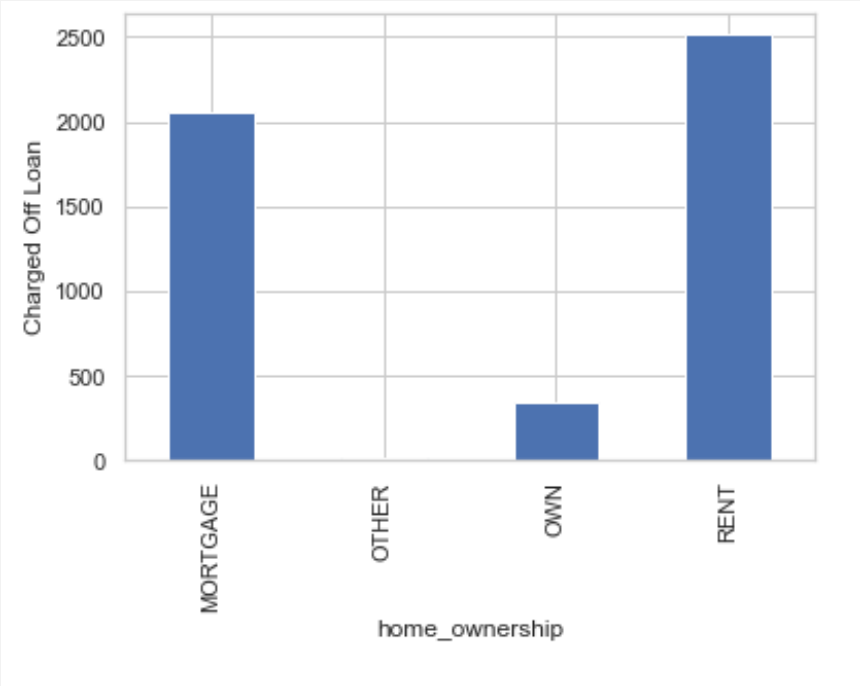
VARIABLE emp_title

- The most frequent loan applicants are US army , Bank of America, IBM
- The Walmart have highest charged off % among the top ten emp_title

VARIABLE emp_length (TOTAL YEARS OF EXPERIENCE)

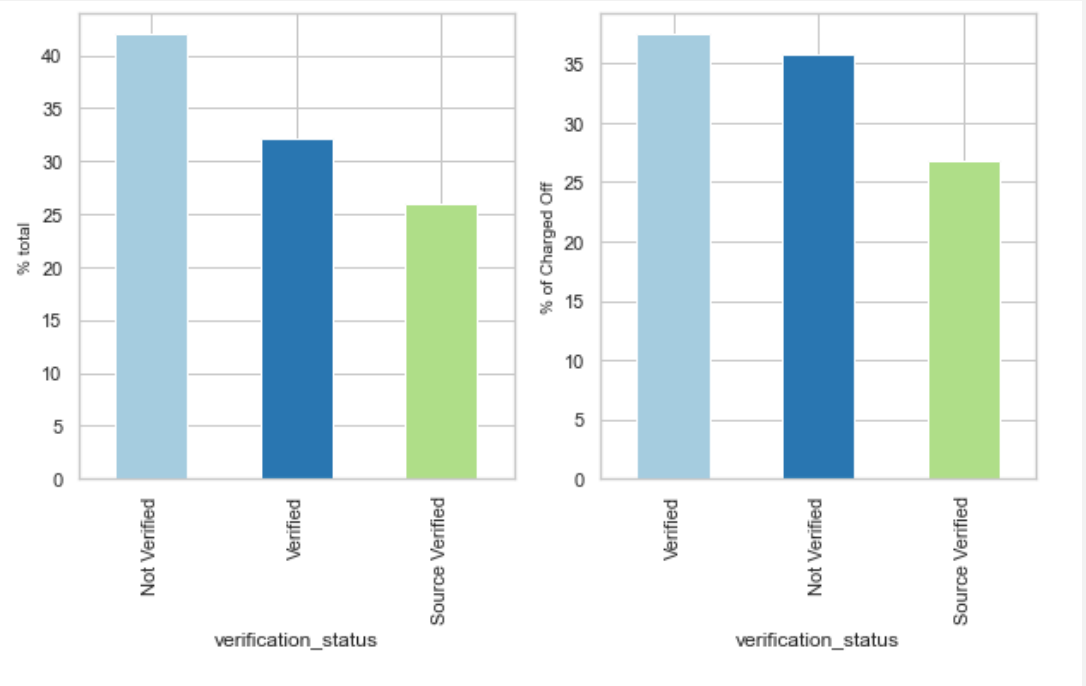
- The employment length doesn't make much difference in charge off loan status.
- Maximum applicants are of 10 years emp_length overall about 22.69

RESULT



VARIABLE home_ownership

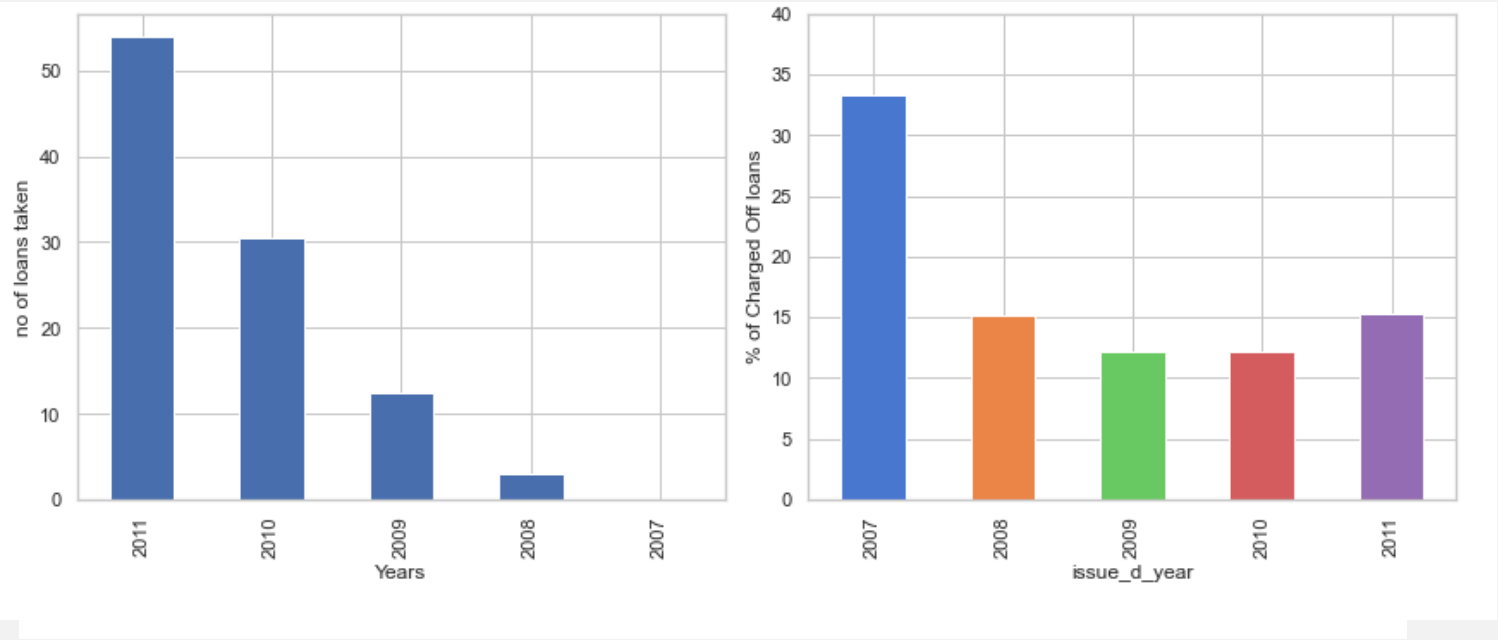
- Most of the Loan applicants has house_ownership as Rent, Mortgage
- Highest charged off of loan is also from Rent and Mortgage

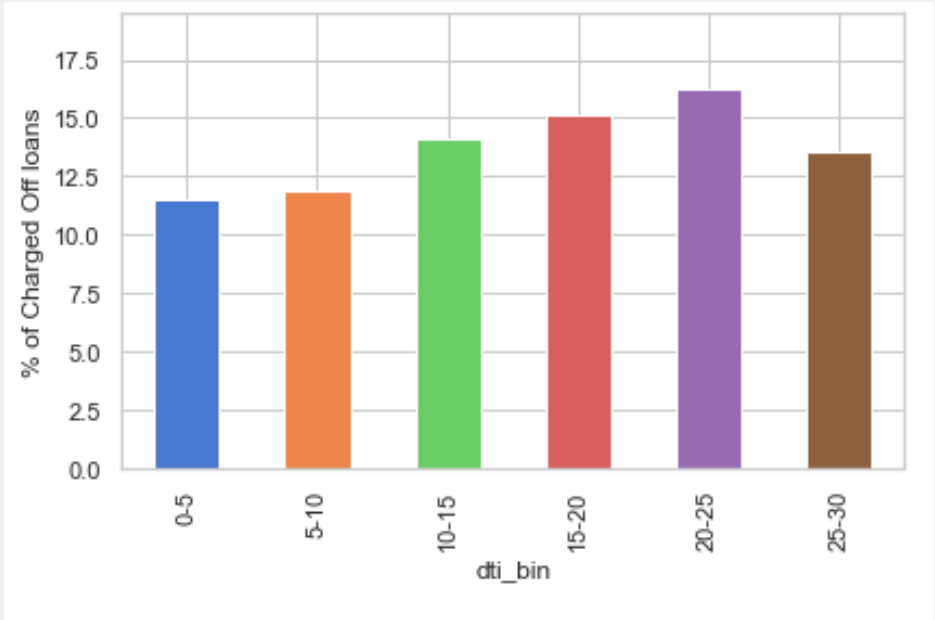


VARIABLE verification_status

- Maximum loans are not verified
- When we consider only charged off loans both verified and not verified are almost same in numbers which difference of 1.2%

RESULT





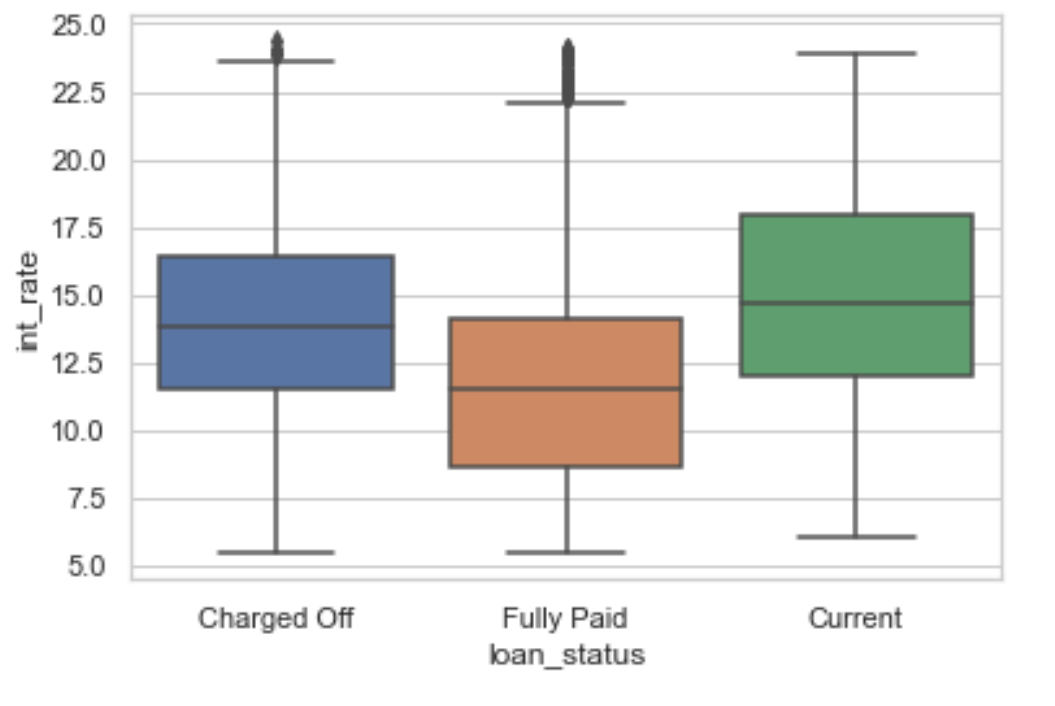
VARIABLE issue_d_year

- Maximum loans are issued in the year 2011
- maximum loans are defaulted in the year 2007

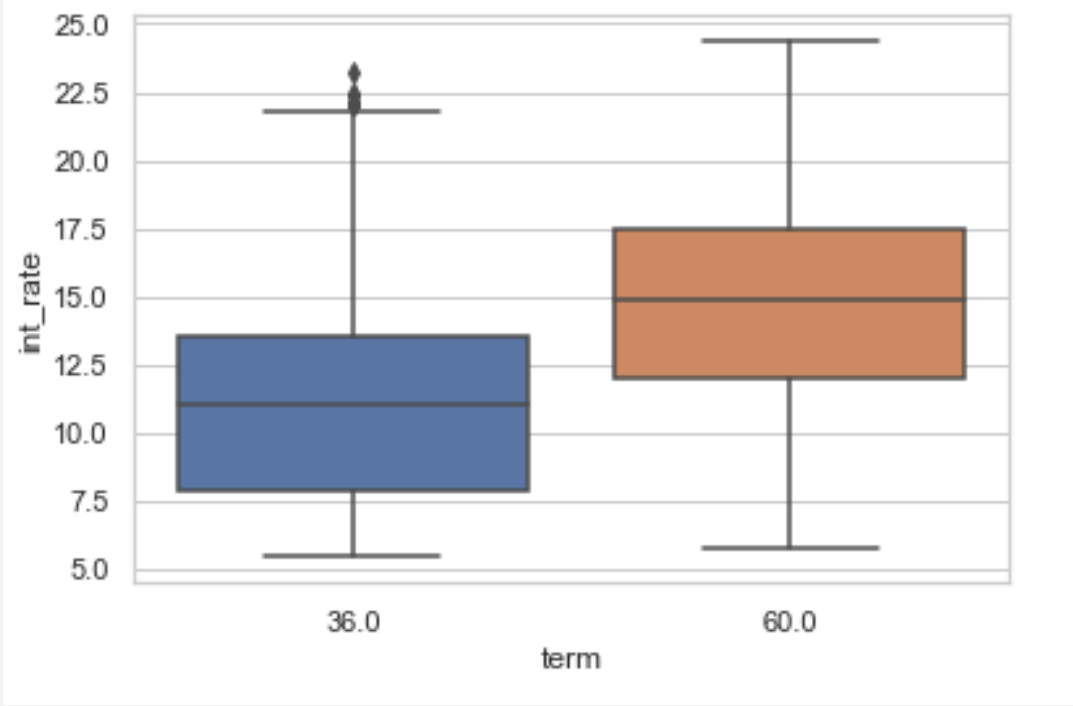
VARIABLE dti

- dti is directly proportional to the number of defaults or charged Off
- the dti bin 20 above are most likely to be default

RESULT

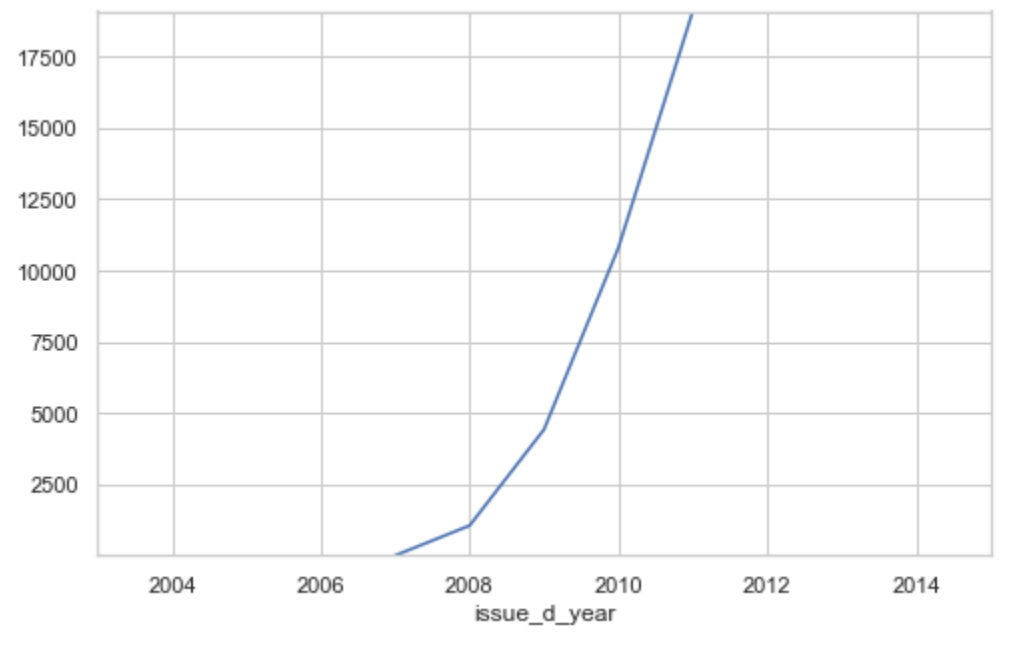


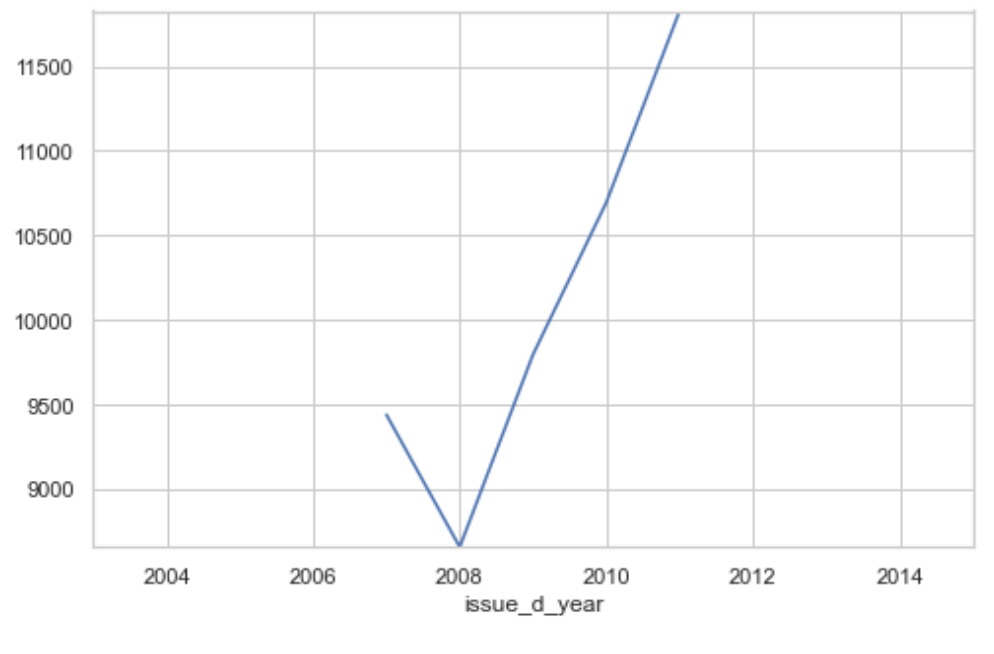
Interest rates are more for the charged off loans compared to fully paid loan_status, so more the interest rates more likely to be the defaults



Interest rates for 60 months term is more compared to 36 months term, which indicates that for higher amount of loan associated with long term the interest rates are more

RESULT





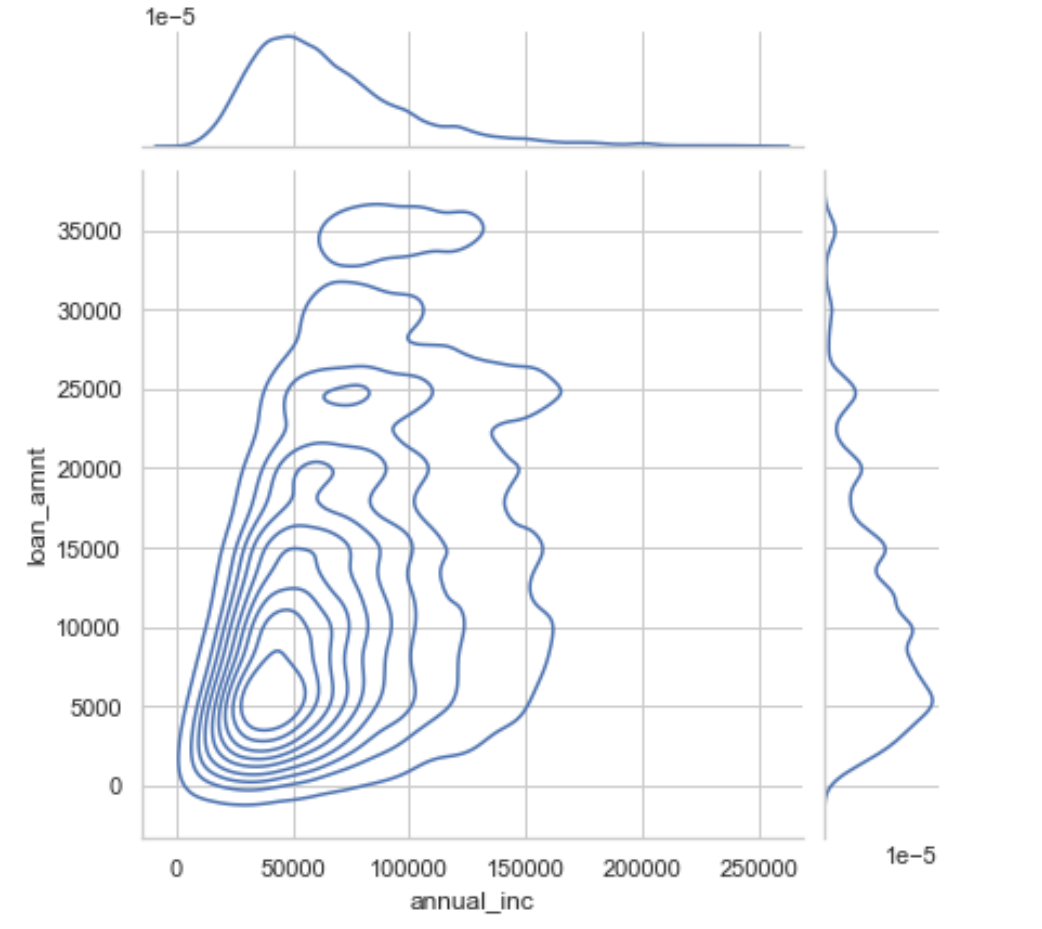
AVERAGE loan_amnt ISSUED IN THAT PARTICULAR YEAR

Number of loans lent to the customers increasing year to year

TOTAL SUM OF loan_amnt ISSUED IN THAT PARTICULAR YEAR

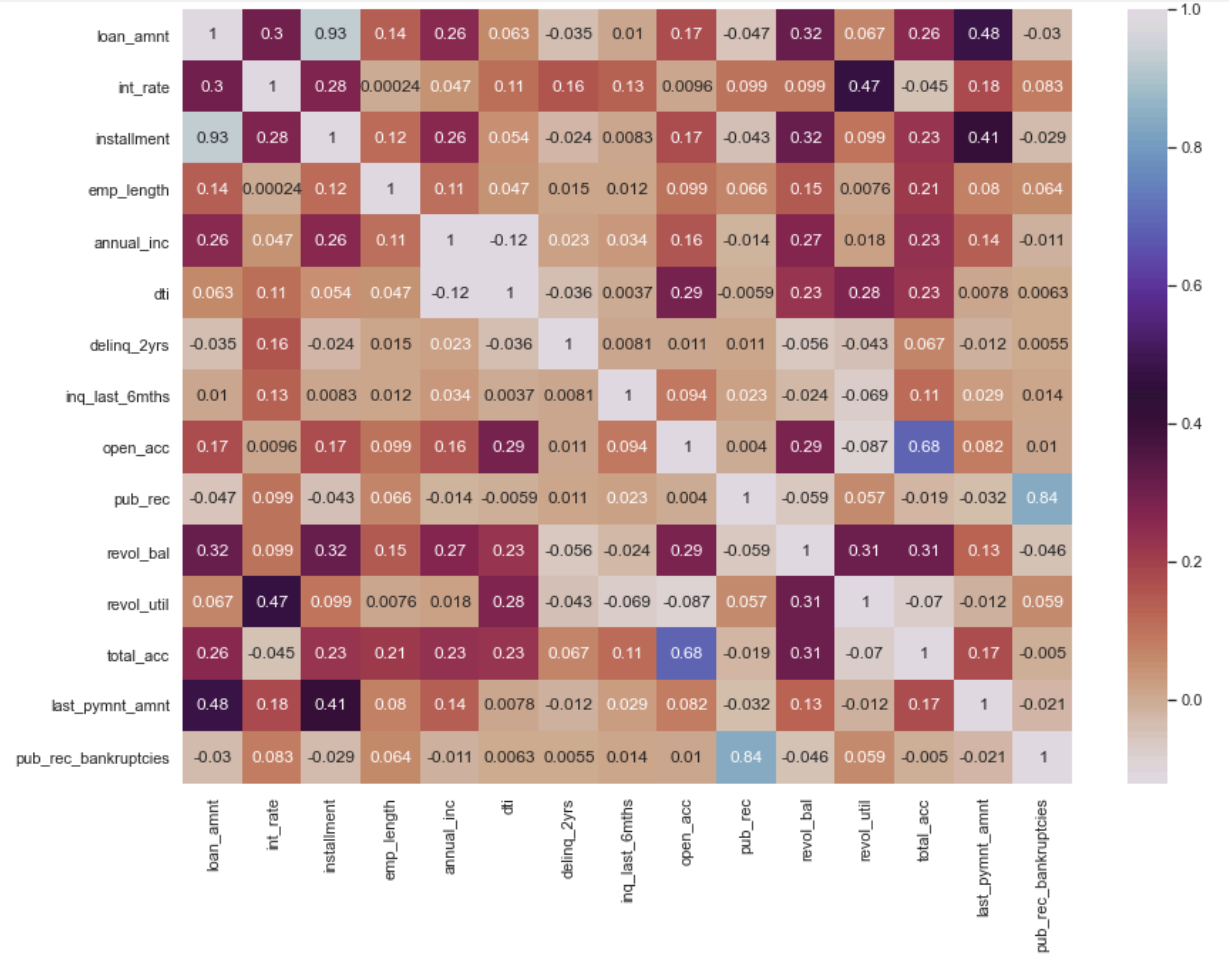
In the Year 2008 very less amount of loans are issued

RESULT



Most of the loan given to the applicants who have very less annual income i.e. (income below 5000 and taken loan more than 25000),those kind of loans are high risk

RESULT



THE loan_amnt AND ITS CORRELATION WITH THE FOLLOWING VARIABLES

- Revol_bal has 0.319781 with loan_amnt
- Int_rate has 0.299997 with loan_amnt
- Annual_inc has 0.258984 with loan_amnt
- Last_pymnt_amnt has 0.477646 with loan_amnt
- Total_acc has 0.255773 with loan_amnt
- The interest rate has higher correlation i.E(0.47) with revol_util. Lending club charges higher interest from risky customers.
- The loan amount given to the customers without considering the annual_income(where few loan applicants have taken the loan amount 50% more than there annual income).This scenario has to be taken care.

CONCLUSIONS

- Below can be the investment strategy for Spark Funds considering all the business understandings –

1. Grades should be considered before giving loans. Grades A and B should be considered as grades D,E,F,G are risky loans.
2. The purpose is also and important factor that the company should consider. Small business purpose for loan defaults the most. So, they should avoid considering such customers.
3. People with own house ownership should be considered over those with rent and mortgage ownership.
4. Interest rates are more for the charged off loans compared to fully paid loan_status, so more the interest rates more likely to be the defaults.
5. Interest rates for 60 months term is more compared to 36 months term, which indicates that for higher amount of loan associated with long term the interest rates are more