Lending Club Case Study

Kanika Gupta
Preethi

Problem Statement

When the loan providing company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- 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.

Aim:

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.

The data given contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Steps followed

- Loading Libraries and Data
- Data Cleaning: Handling Missing Values and Outlier Treatment
- Data Manipulation and Filtering
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis

Data Understanding

Given Dataset:

- Aim is to profile loan applicants to predict loan defaulters
- Dataset size: 39717 rows, 111 columns

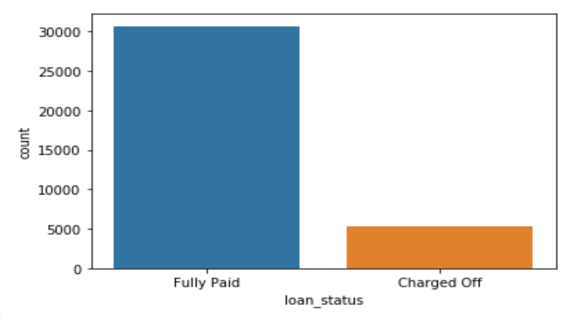
Data Filtering

- Data cleaning involves removal of the following data:
    - Missing values
    - Single-valued columns (containing only 1 value)
- Further, columns irrelevant to loan analysis can be ignored

Data Manipulation

- Data is converted to a usable format. Fields containing symbols are modified.
- Fields containing dates are converted into "date" format.
- New features are created for analysis
- Bins for created for categorizing data

Analysis by Visualization



Conclusion

- 85% of borrowers have paid the loan fully . 14% of borrowers have defaulted

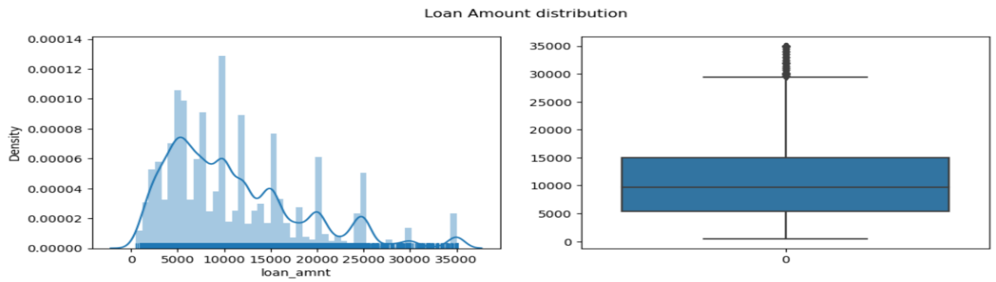
Analysis will be performed in 3 ways:

- **Univariate Analysis**
- **Segmented Univariate Analysis**
- **Bivariate Analysis**

Univariate Analysis

1. Loan amount:

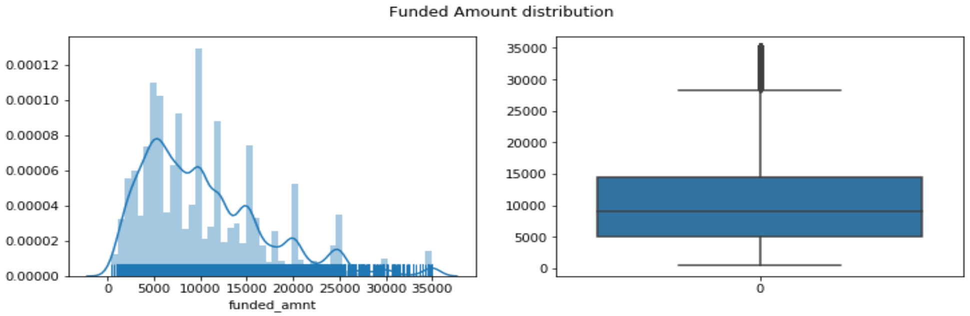
Listed amount of the loan applied by the borrower. Any future changes to the loan amount sanctioned will be reflected here.



Conclusion:

- More number of people have taken a loan of Rs 10000 so median is set as 10000.
- Very few borrowers took a loan amount more than 30000

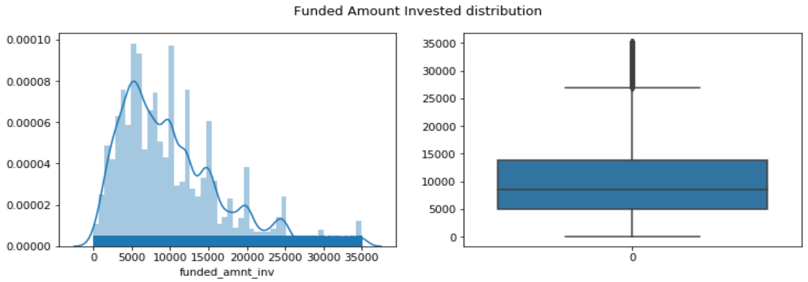
2. **Funded Amount**



Conclusion:

- Funding amount distribution similar to loan amount
- Most of the applied loan amount has been approved

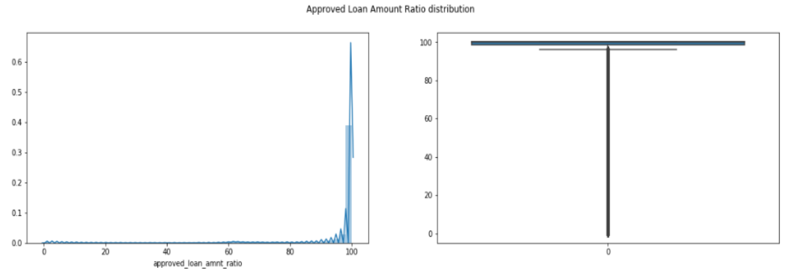
3. Funded Amount Invested



Conclusion:

- Funded amount investment has same distribution as loan amount. Lending club has approved almost all applied loan amount.

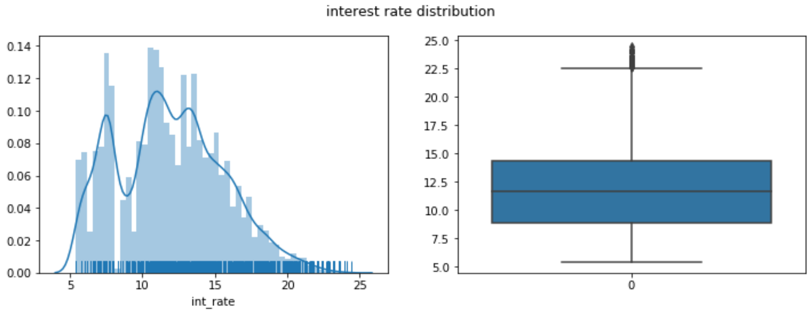
4. **Approved loan amount ratio**



Conclusion:

- 70% of borrowers got 100% of their applied loan amount

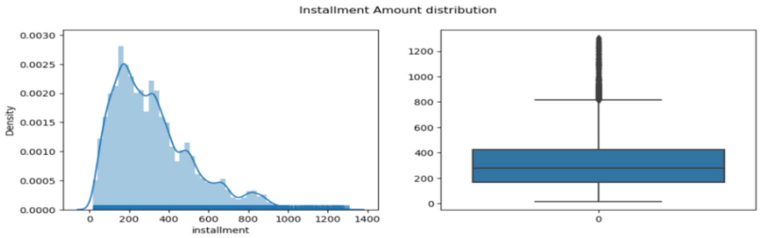
5. **Interest rate distribution:**



Conclusion:

- Most of the interest rates lie between 8% to 14%
- Few people got loans at an interest of 22%

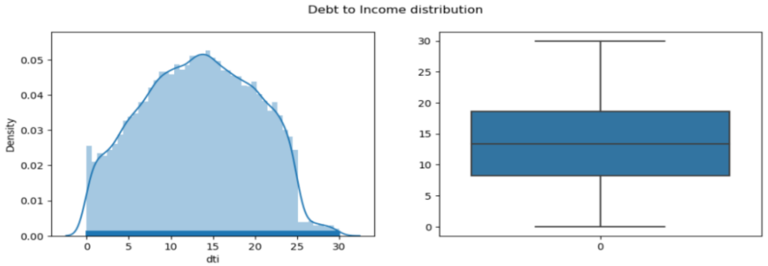
6. **Installment**



Conclusion:

- Most common monthly installment amount is around 280

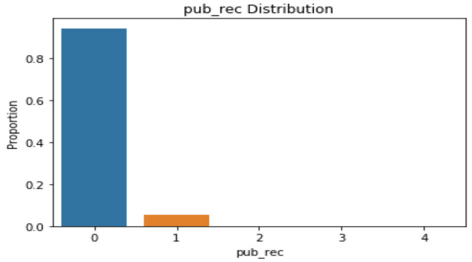
7. **Debt to Income distribution**



Conclusion:

- No outliers. Indicates that the loans have been given to borrowers who have debt to income ratio less than 30

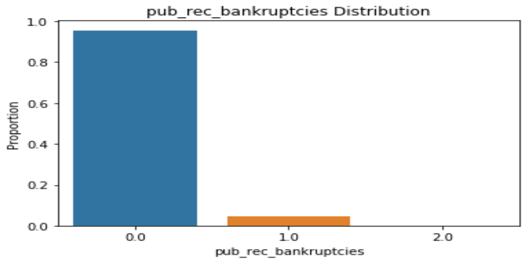
8. Derogatory public records



Conclusion:

- Around 90% of borrowers have no public derogatory records

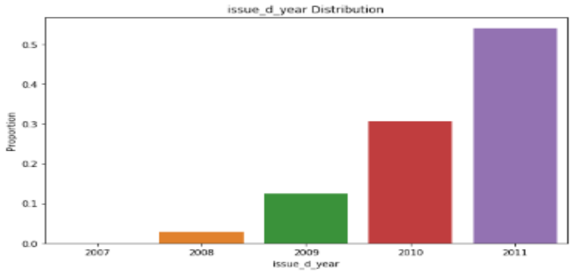
9. **Public record bankruptcies**

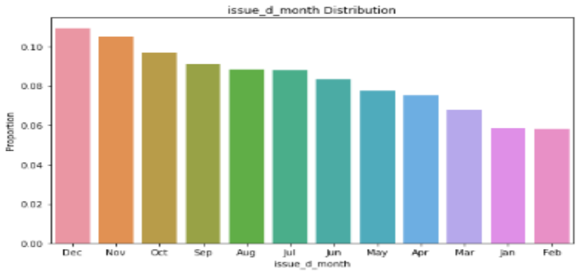


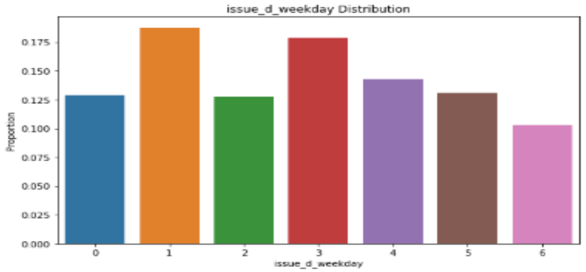
Conclusion:

- 99% of the borrowers have not been bankrupt

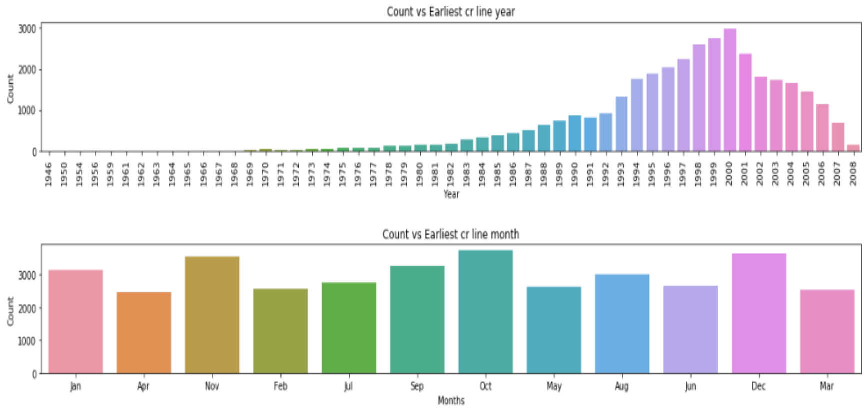
10. **Loan Issued date**







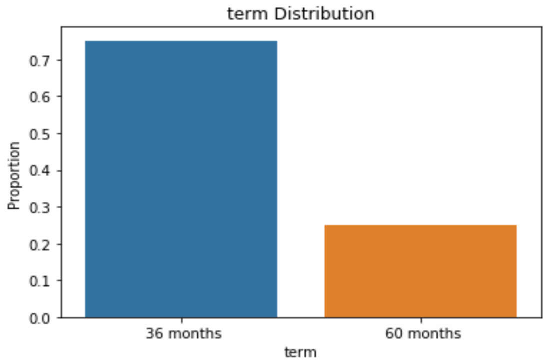
11. **earliest_cr_line**



Conclusion

- Many of Loan borrowers of LC have got earlier credit line in 2000 year, and also most have got earlier credit line on end of the year i.e., Oct, Nov, Dec

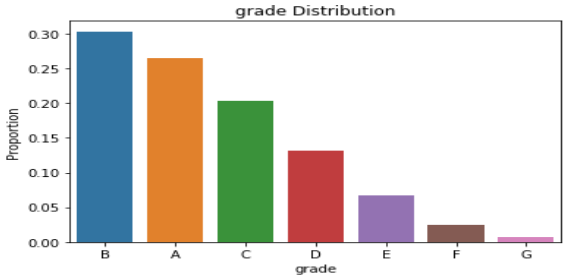
12. **Term**



Conclusion

- Borrowers have taken 36 months tenure for more than 60 months

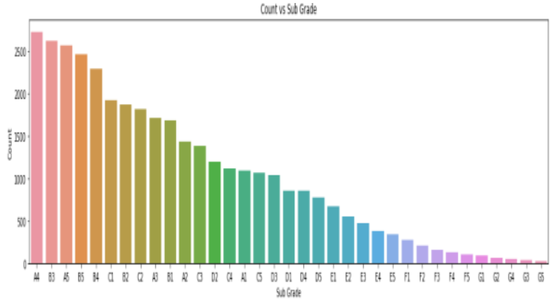
13. **Grade**



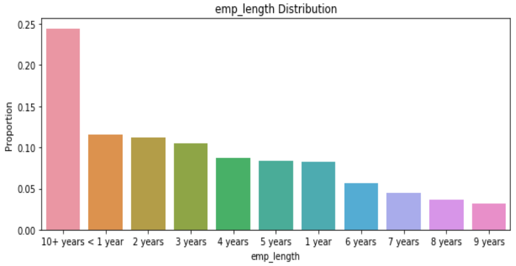
Conclusion

- Most borrowers fall under B and A grades than other grades

14. **Sub Grade**



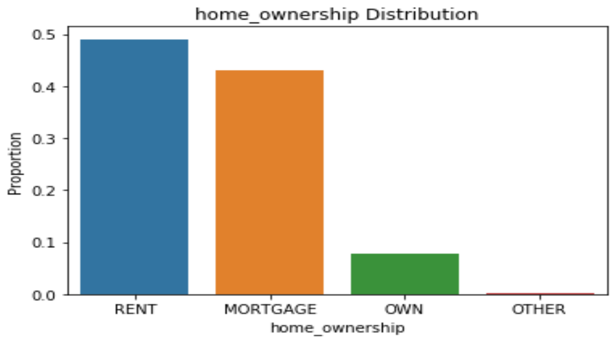
15. Employment length



Conclusion

- Most of the borrowers are having 10+ years employment length

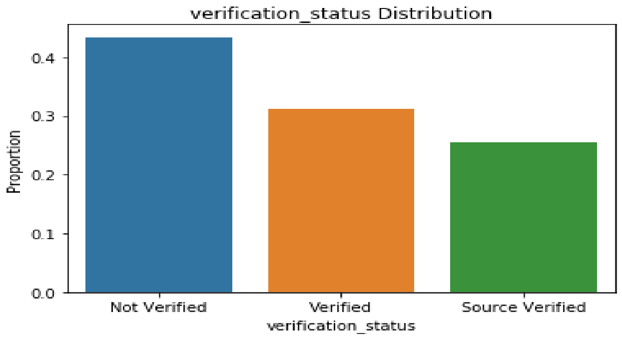
16. Home Ownership



Conclusion:

- The loan borrowers are mostly having rented and mortgage houses

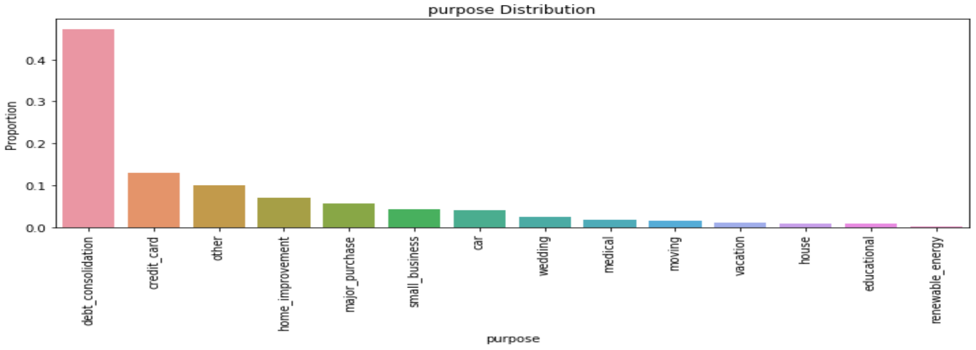
17. Verification Status



Conclusion:

- Majority of loans were given without verification of applicant’s income

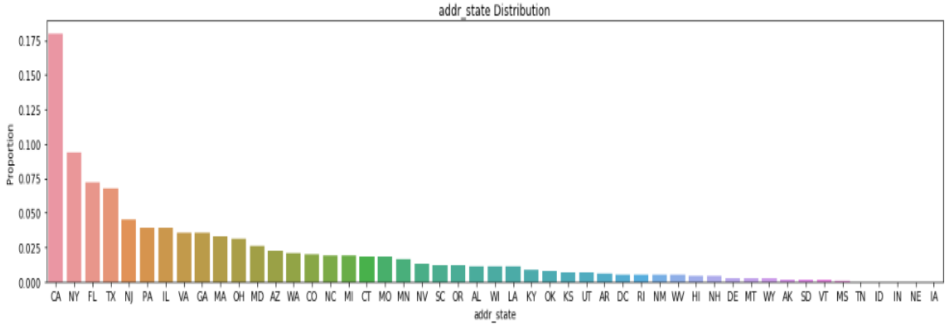
18. **Purpose**



Conclusion:

- Maximum number of loans were taken for debt consolidation and very few people took for renewable energy

19. **Add state**



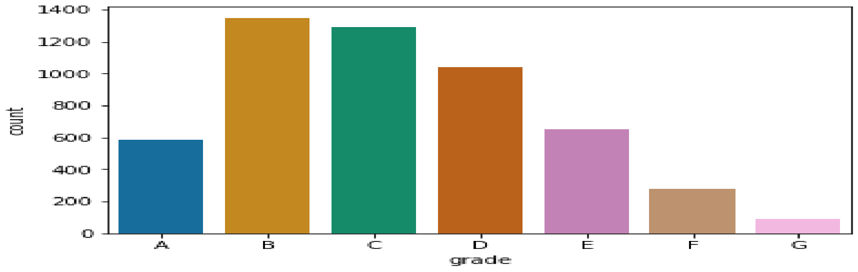
Conclusion:

- Most of the borrowers are from CA followed by NY

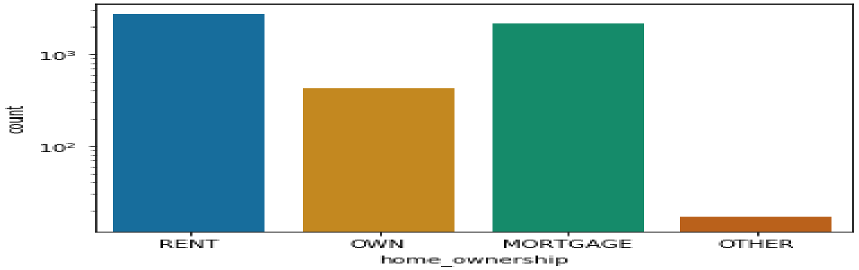
Segmented Univariate Analysis

Analysis wrt "Charged-Off" Loans

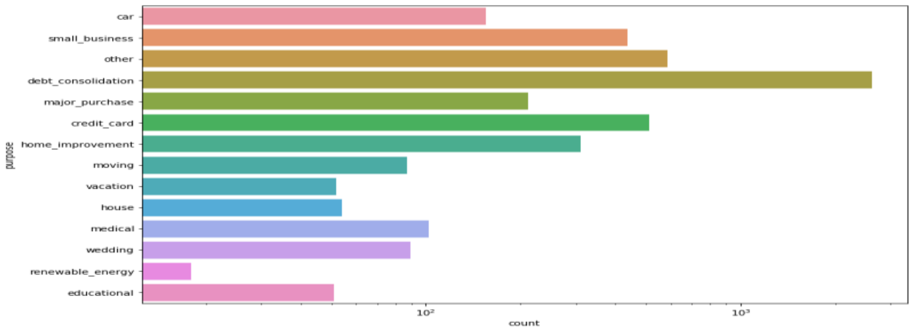
1. **Grade**



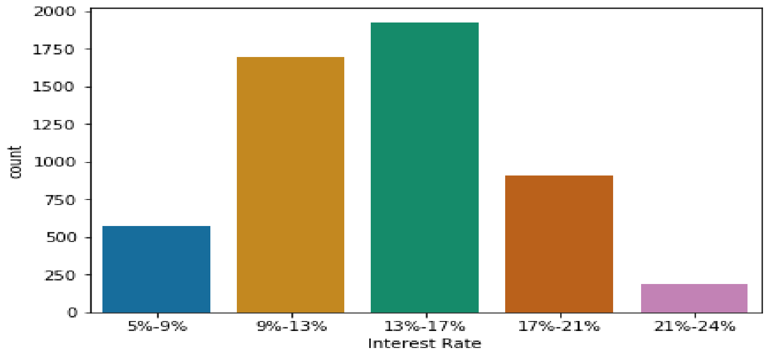
2. **Home ownership**

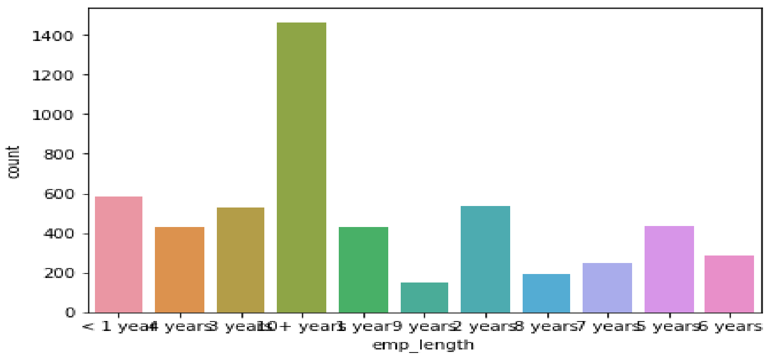


3. **Purpose**

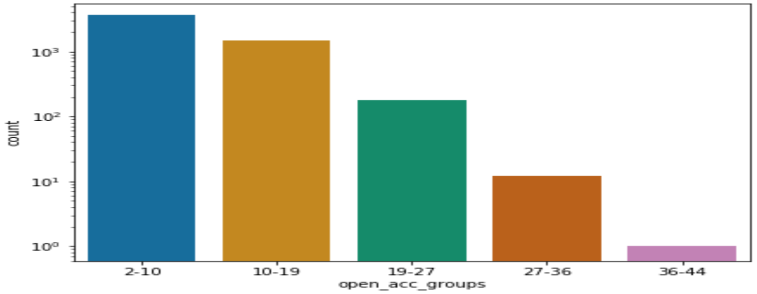


4. Interest rate w.r.t interest rate bins created

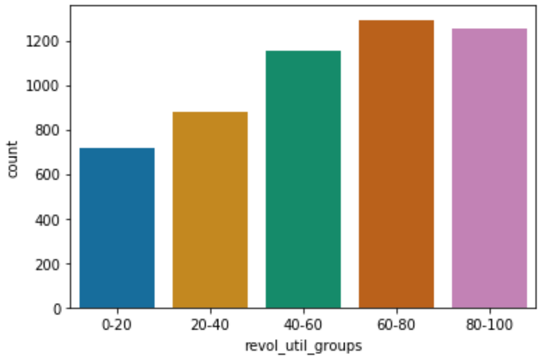




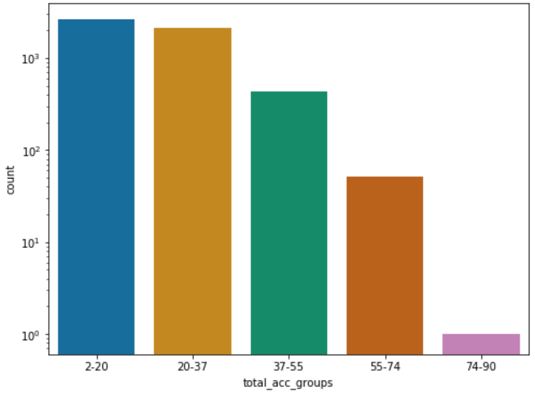
5. open_acc_groups



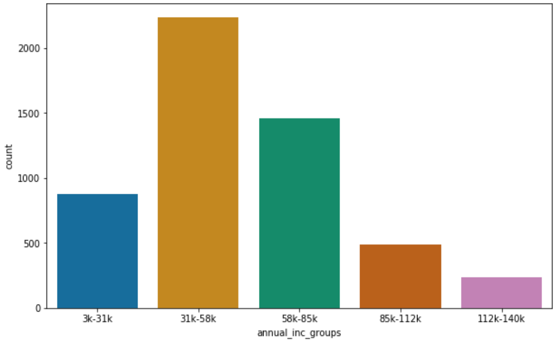
6. revol_util_groups



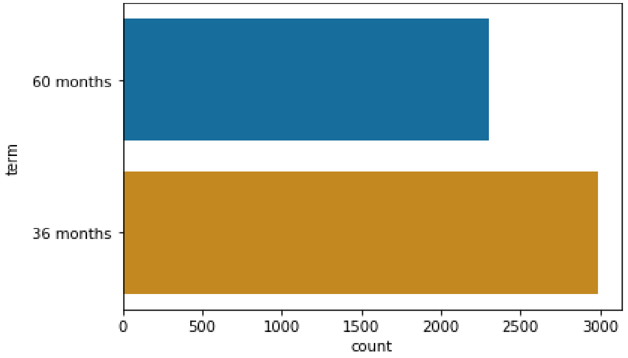
7. total_acc_groups



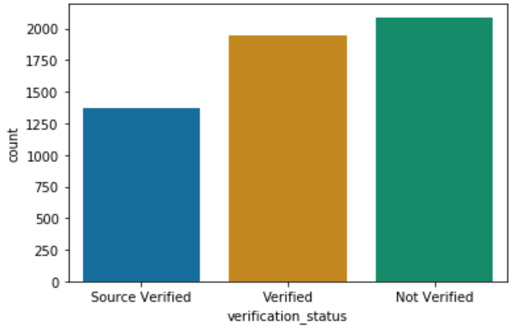
8. annual_inc_groups



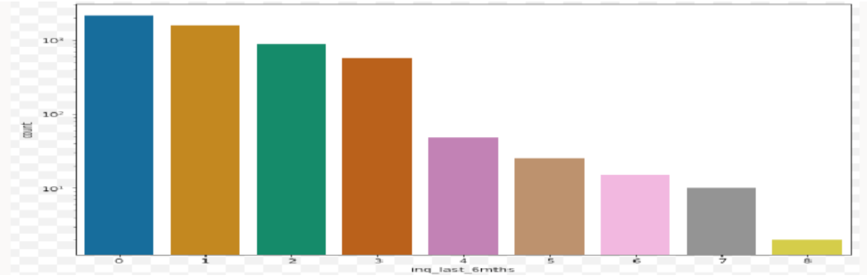
9. term



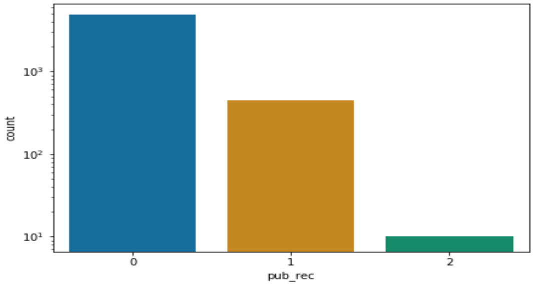
10. Verification Status



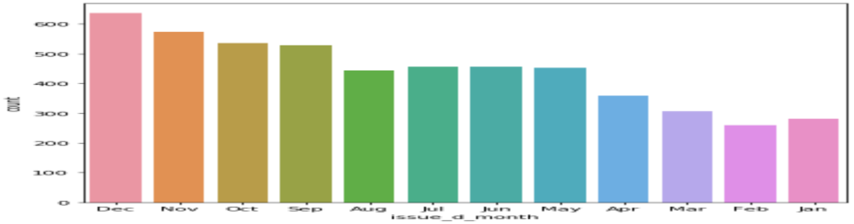
11. Inq_last_6mths

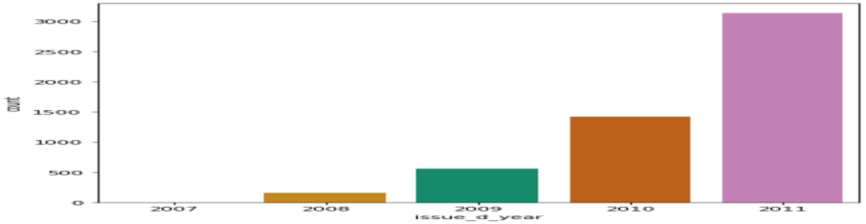


12. pub_rec



13. **Issued month and year**

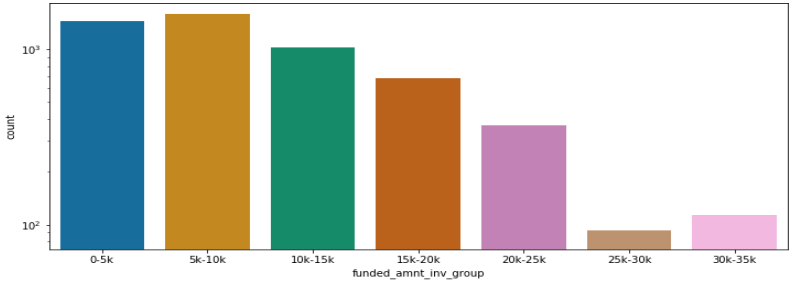




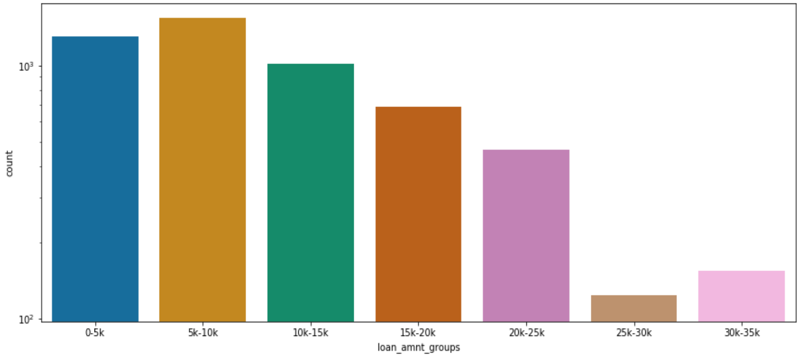
Conclusion:

Maximum number of defaults occured when the loan was issued in December. Loan issued in the year 2011 resulted in defaults as compared to other years.

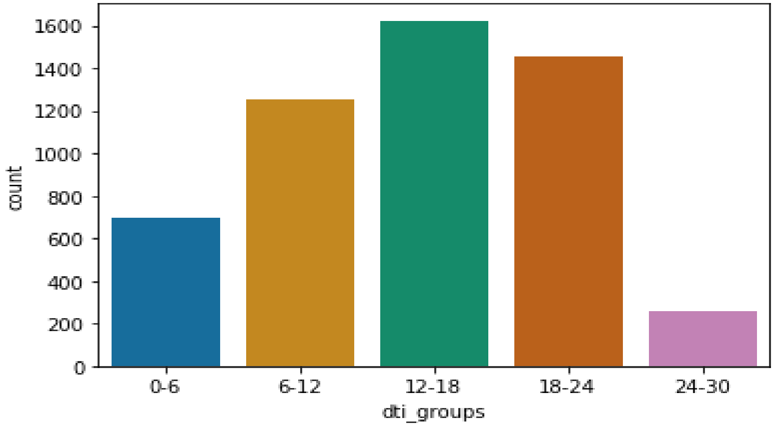
14. **funded_amnt_inv groups**



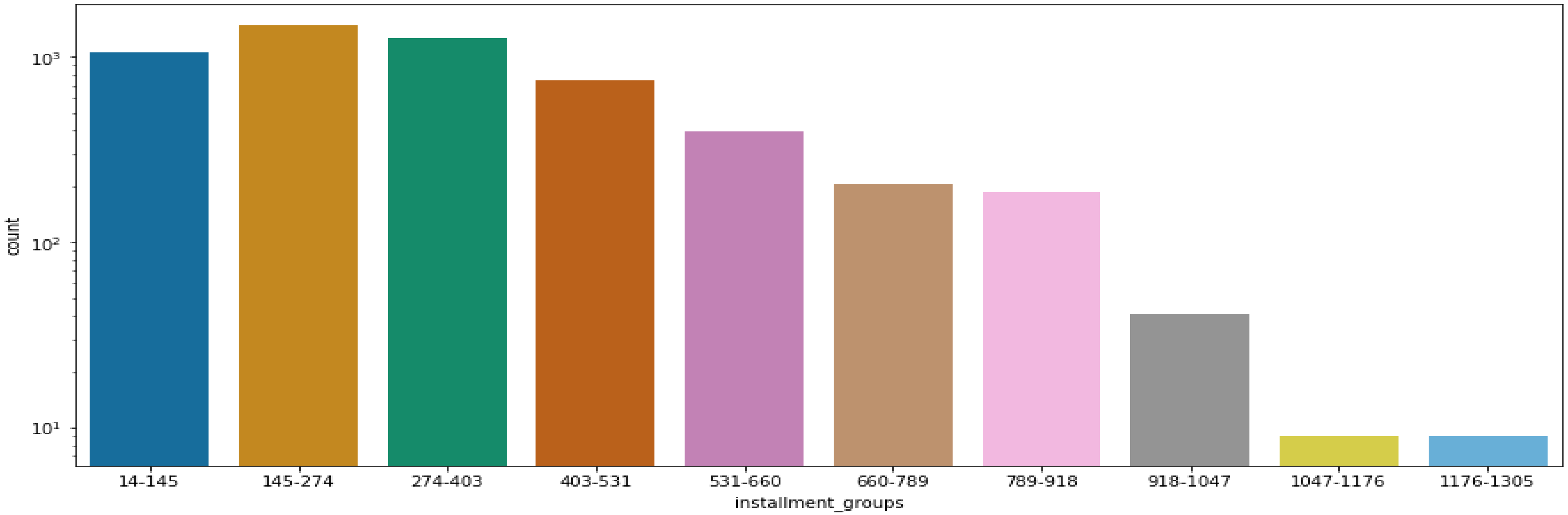
15. **loan_amnt_groups**



16. dti group



17. Installment groups



Conclusions wrt "Charged-off" Loans

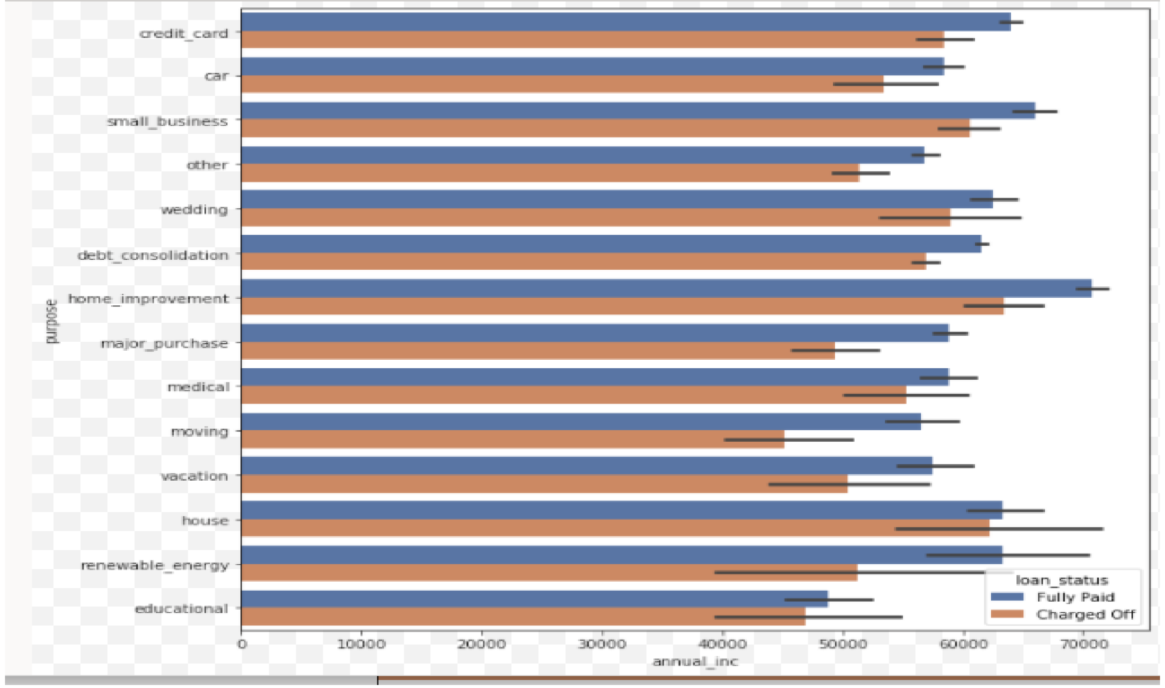
- The above performed analysis with respect to the "charged off" loans for each variable suggests the following that there is a more probability of a loan defaulting :
- When the Grade is 'B'
- When total grade of 'B5' level
- If the applicants having house_ownership as 'RENT'
- When the purpose is 'debt_consolidation'
- For applicants who use the loan to clear other debts
- For applicants who receive interest at the rate of 13-17%
- For applicants who have 20-37 open_acc
- For applicants who have an income of range 31201 - 58402
- For applicants with employement length of 10
- When funded amount by investor is between 5000-10000
- When Loan amount is between 5429 - 10357
- When Dti is between 12-18
- When monthly installments are between 145-274
- For Term of 36 months
- When the loan status is Not verified
- When the no of enquiries in last 6 months is 0
- When the number of derogatory public records is 0

A conclusion derived from date issued is that during end of a year there is a higher possibility of loan defaulting.

- The high number of loan defaults in 2011 could be due to the financial crisis in USA (Assuming the data is of US origin)

Bivariate Analysis

1. **Annual income vs loan purpose**



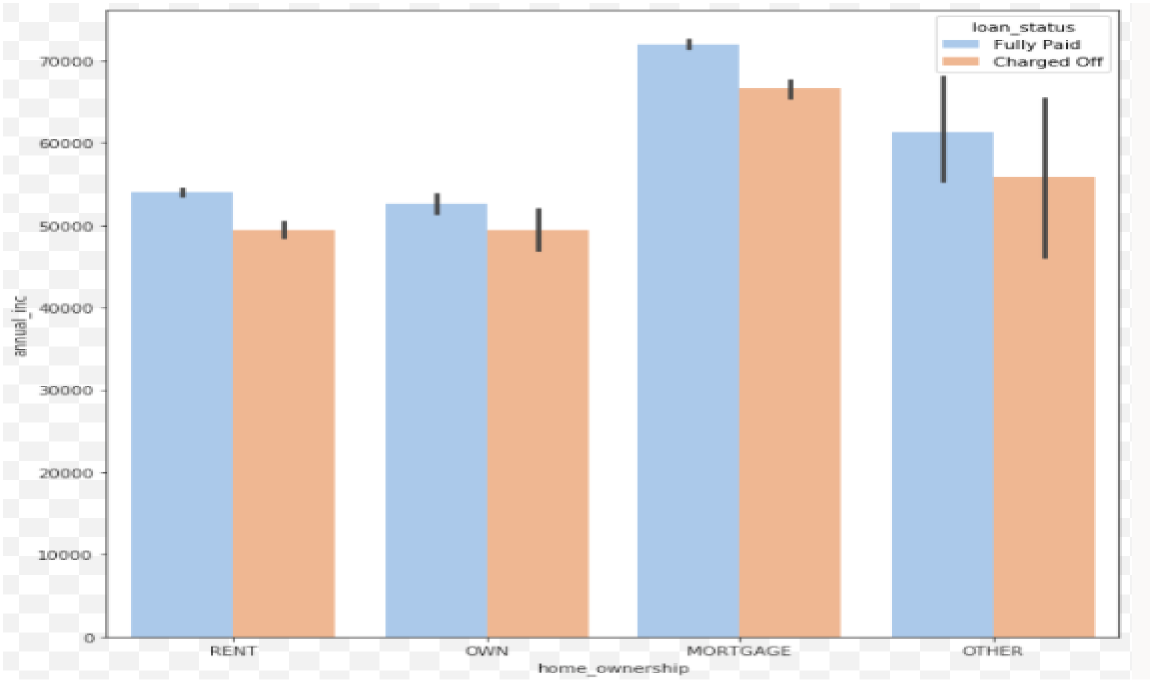
Conclusion:

- We can see though that the maximum number of loans applied and defaulted are "debt_consolation" but the annual income of those who applied isn't the highest.
- Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses"

Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants taking loan for 'home improvement' and have income of 60k -70k

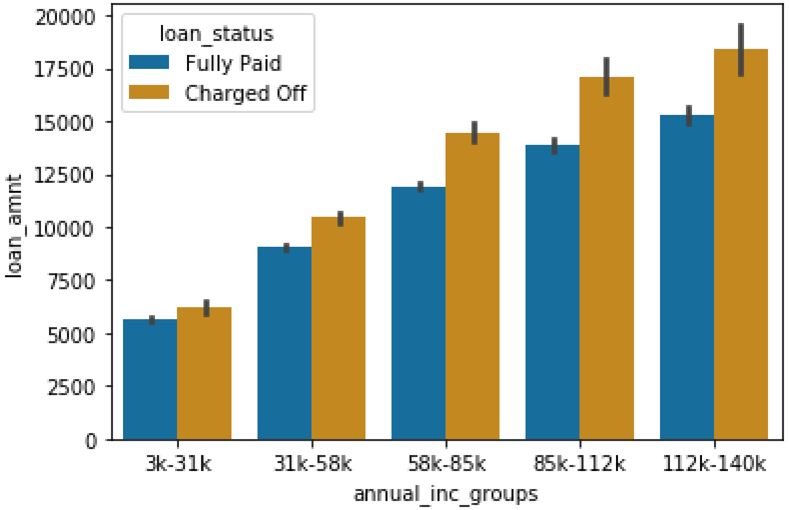
2. Annual income vs Home ownership



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants whose home ownership is 'MORTGAGE and have income of 60-70k.

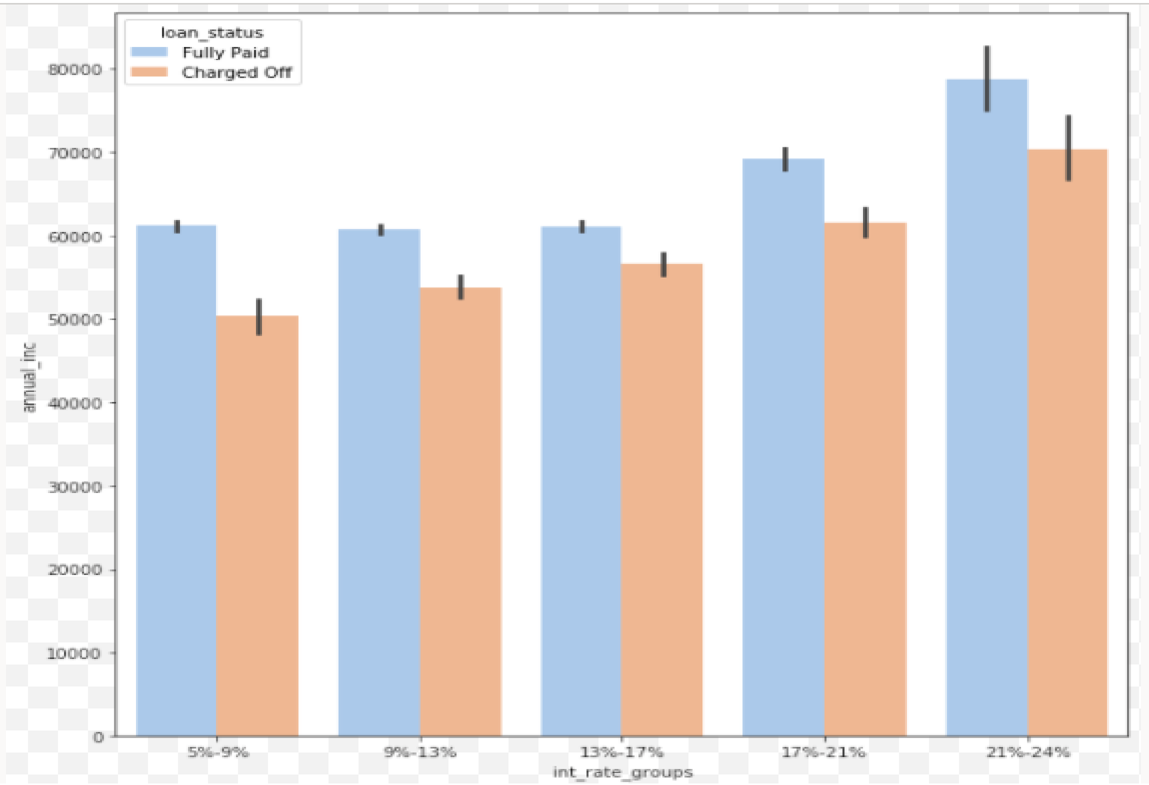
3. **Annual income vs loan amount**



Conclusion:

- We can see that across all income groups, the loan_amount is higher for people who defaulted

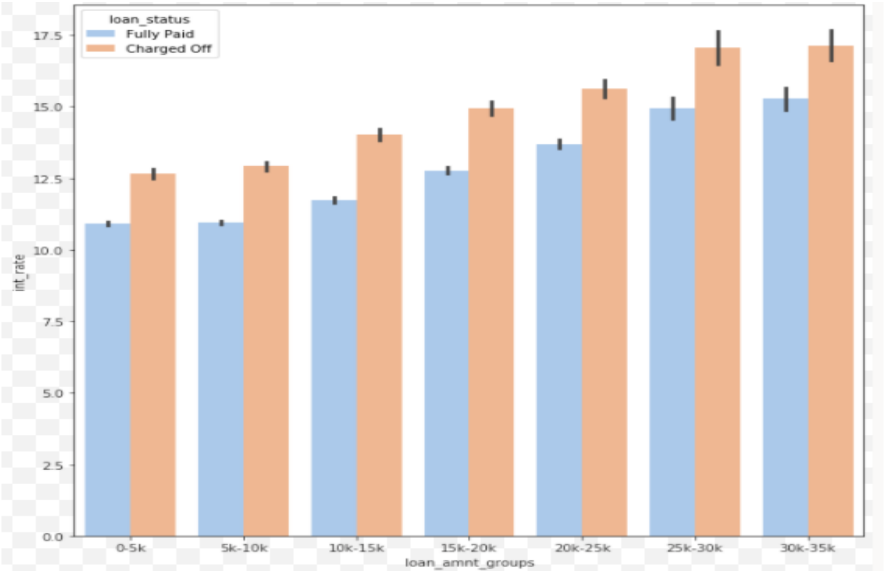
4. **Annual income vs int_rate**



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants who receive interest at the rate of 21-24% and have an income of 70k-80k.

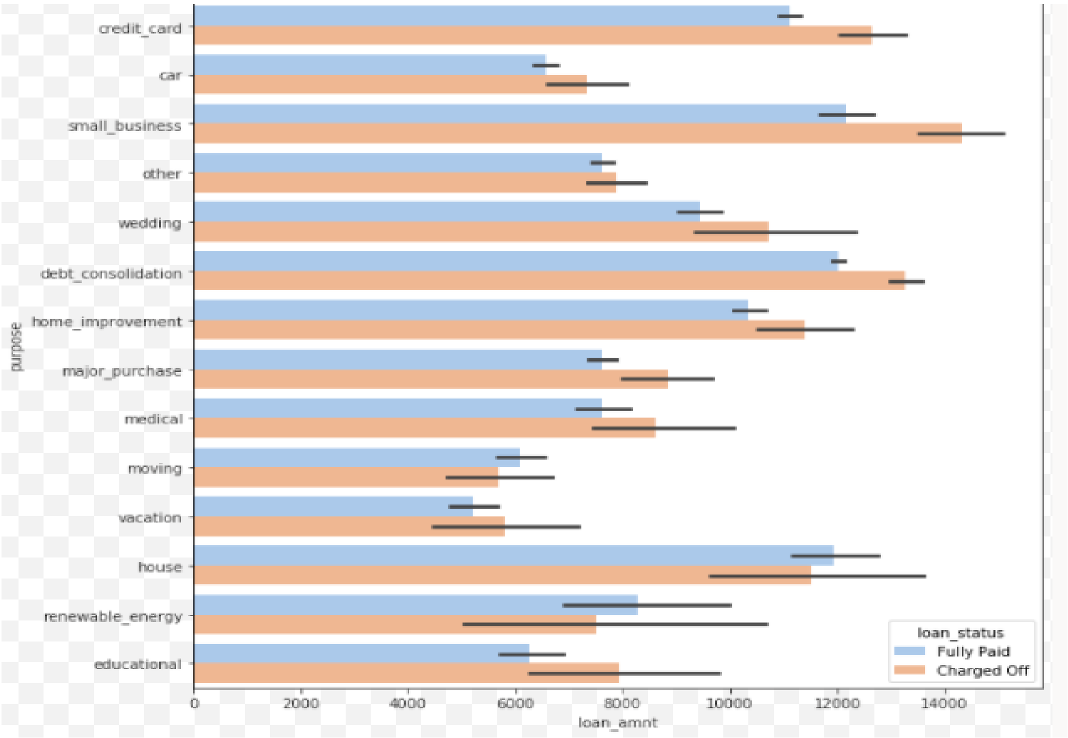
5. Loan amount vs int_rate



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants who have taken a loan in the range 25k - 35k and are charged interest rate of 15-17.5 %.

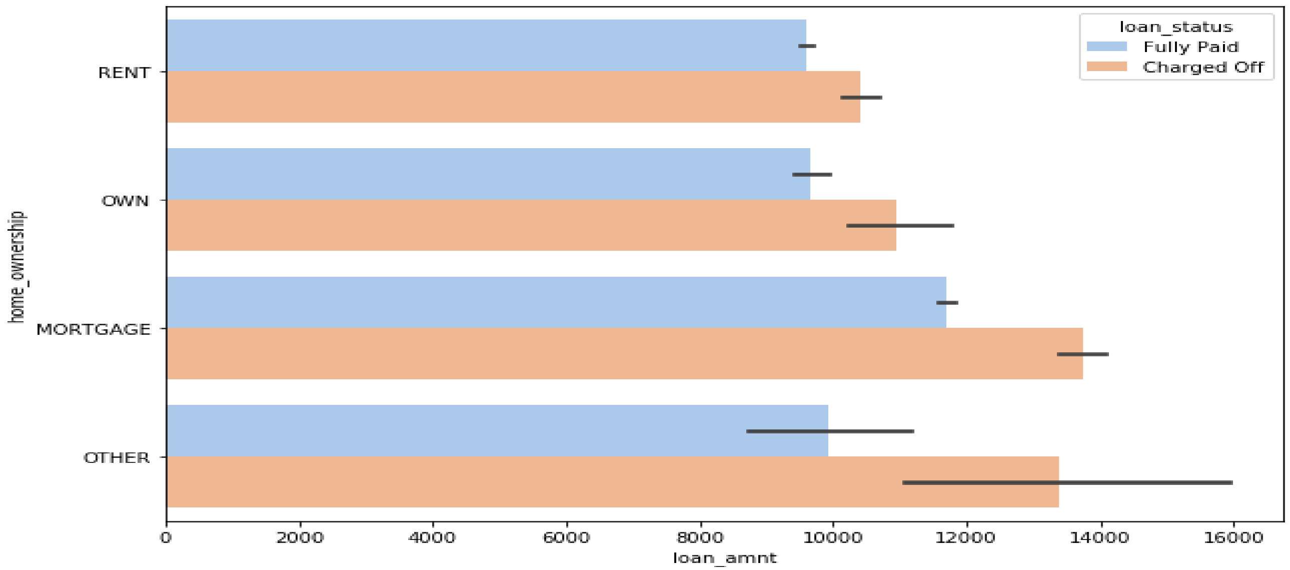
6. **Loan vs Loan purpose**



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants who have taken a loan for small business and the loan amount is close to 14k or above.

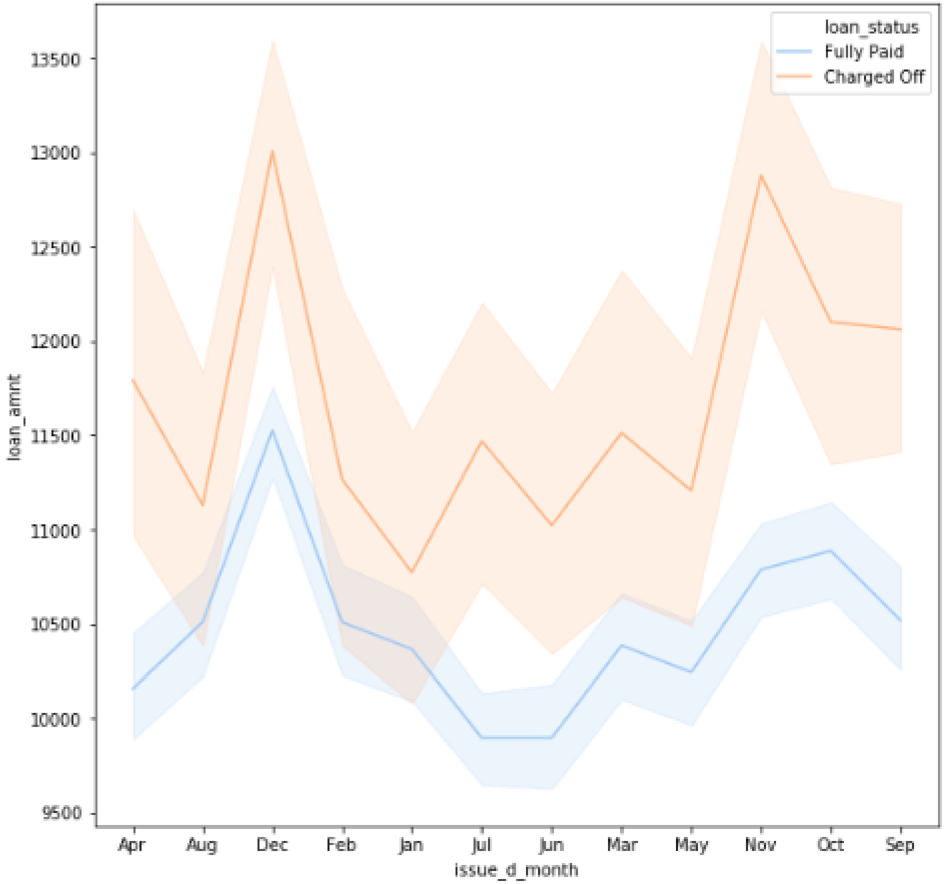
7. Loan vs house ownership

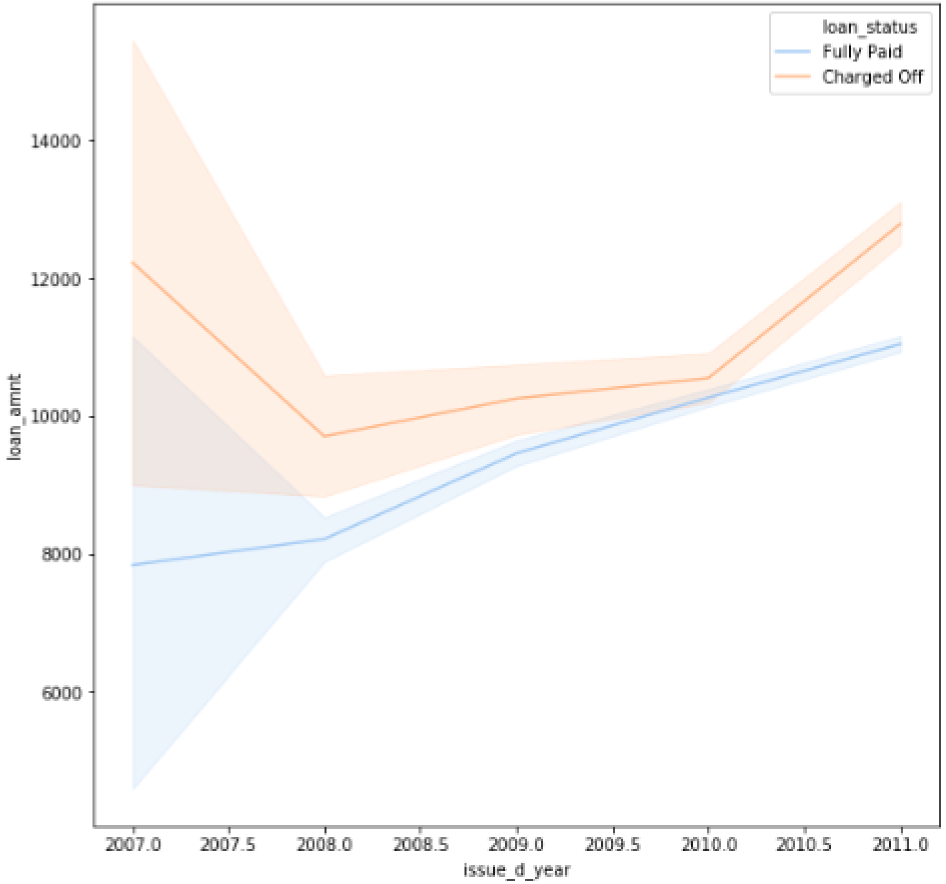


Conclusion wrt "charged-off":

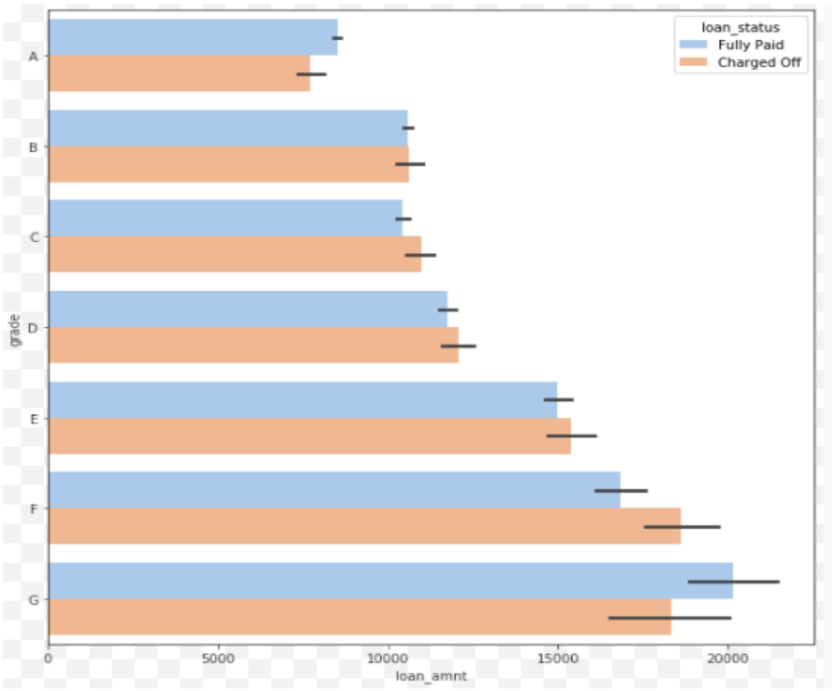
- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when applicants whose home ownership is 'MORTGAGE and have loan of 13k+.

8. Loan amount vs month issued, and year issued





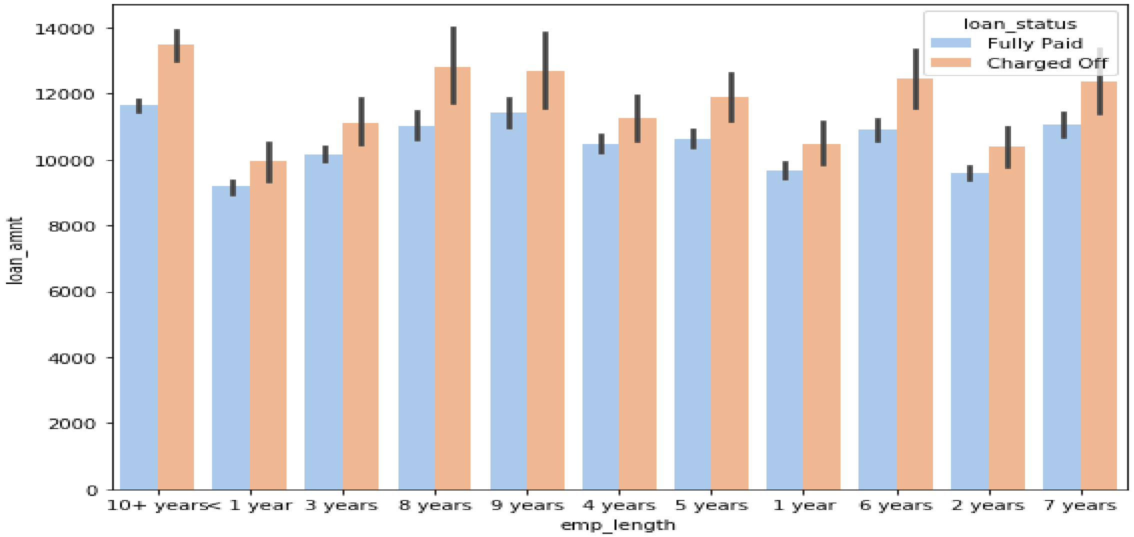
9. Loan amount vs grade



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when the grade is F and loan amount is between 15k-20k.

10. **Loan amount vs Emp length**



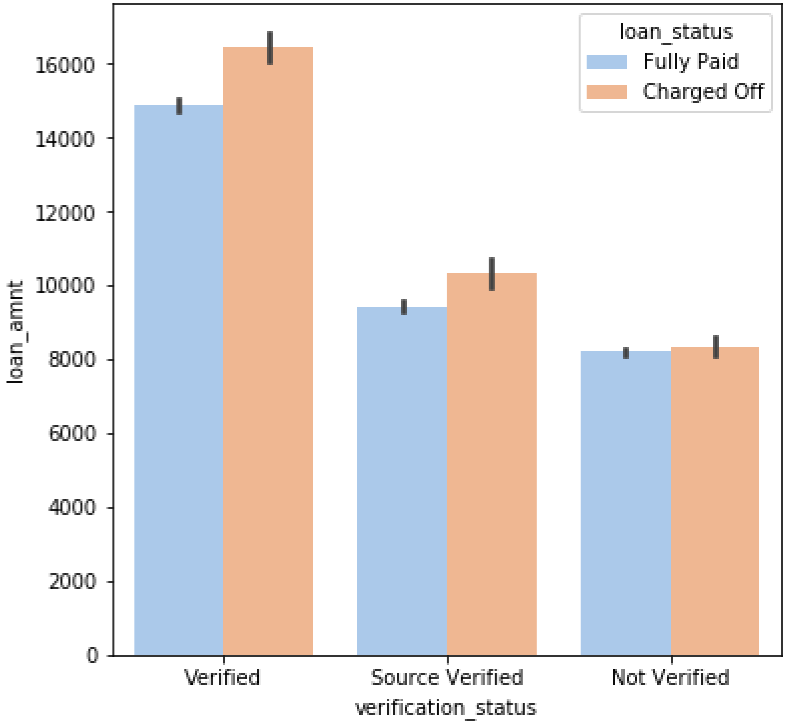
Conclusion:

- We can see that employees with longer working history got the loan approved for a higher amount.

Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when the employment length is more than 7yrs and loan amount is 12k-14k

11. **Loan amount vs verification status**



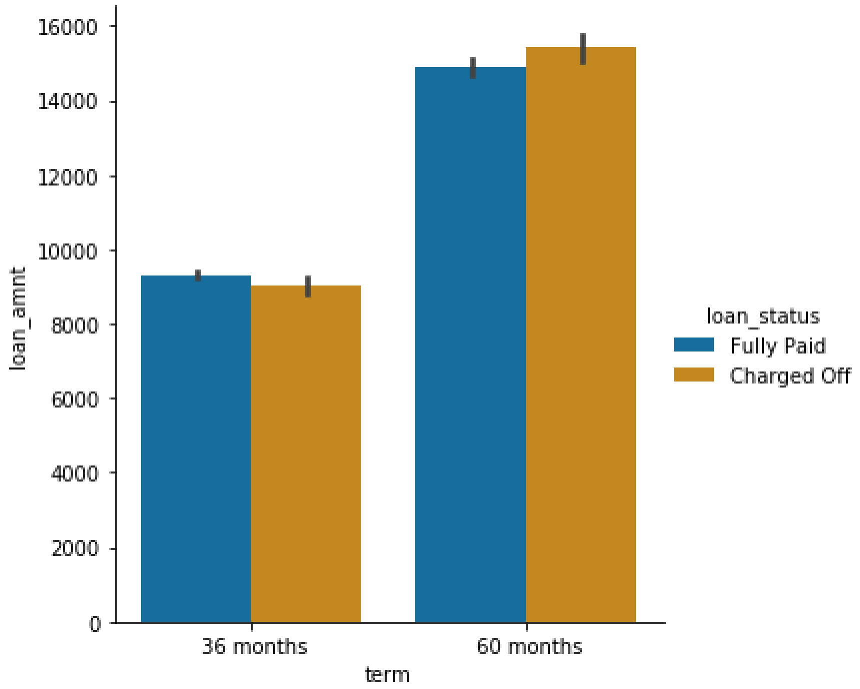
Conclusion:

- Looking at the verification status data, we can say verified loan applications tend to have higher loan amount. This might indicate that the loan banks are first verifying the loans with higher values.

Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting when the loan is verified and loan amount is above 15k.

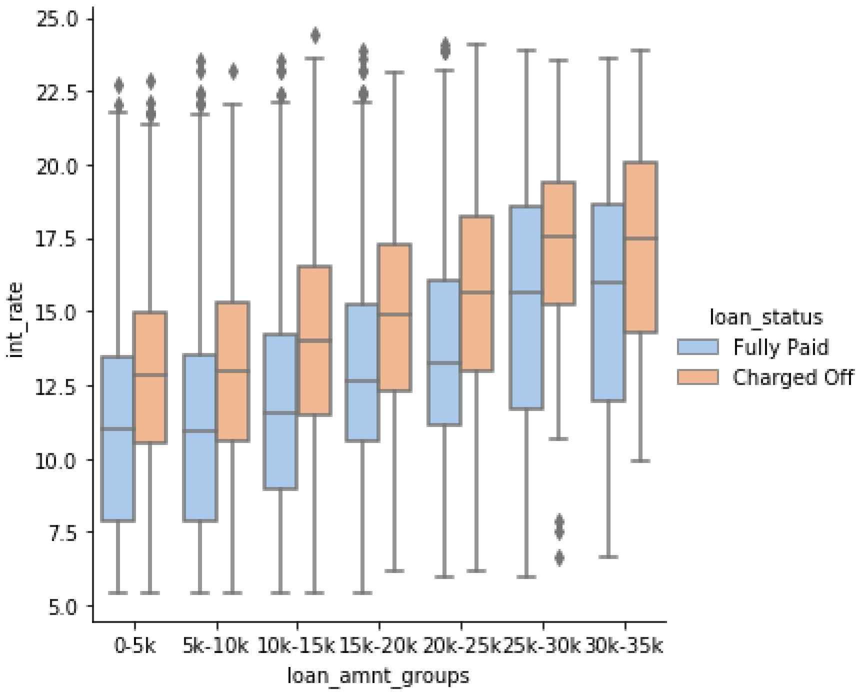
12. **Term vs Loan amount**



Conclusion:

- We see that the applicants who paid and defaulted have no significant difference in loan_amounts
- We can only say that applicants applying for long term loans have taken loan for a higher loan amount

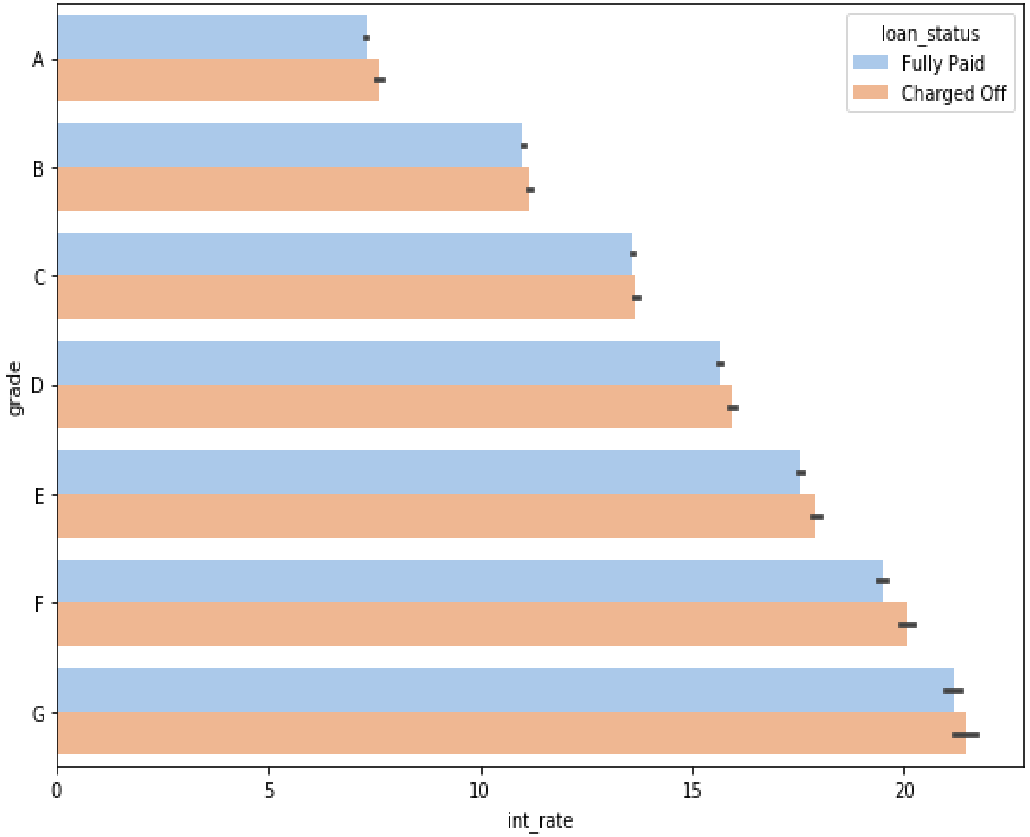
13. Loan amount group vs interest rate



Conclusion:

- We see that the interest rate for charged off loans is pretty high than that of fully paid loans in all the loan_amount groups.
- We suspect this could be a pretty strong driving factor for loan defaulting

14. **Grade vs interest rate**



Conclusion wrt "charged-off":

- With respect to "charged-off" loans, we can say that there is a more probability of defaulting for grade G with interest rate above 20%.