

Lending Club Case Study

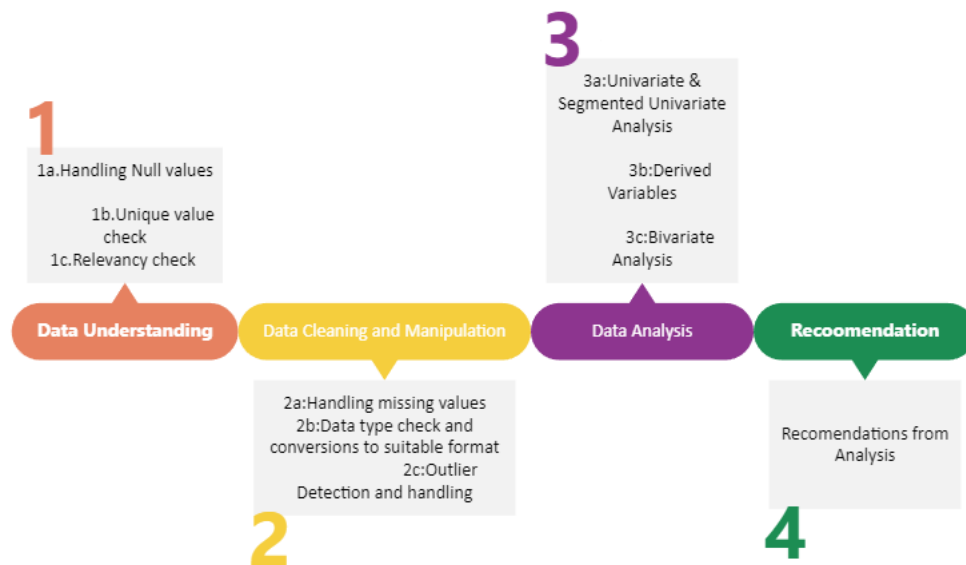
Group members:

~Rakesh Raushan

Business Objective:

- The company is the largest online loan marketplace, facilitating loans. Lending loans to 'risky' applicants is the largest source of financial 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'.
- Identification of such applicants whose loans have a higher risk of getting 'charged-off' and what factors drive to 'charge-off' is the aim of this case study.
- We have to perform Exploratory Data Analysis(EDA) on the provided data and identify such potential drivers for 'charge-off'. company will use this to take decisions regarding which loan applications to approve or reject in future.
- The company will use this to take decisions regarding which loan applications to approve or reject in future.

Steps:

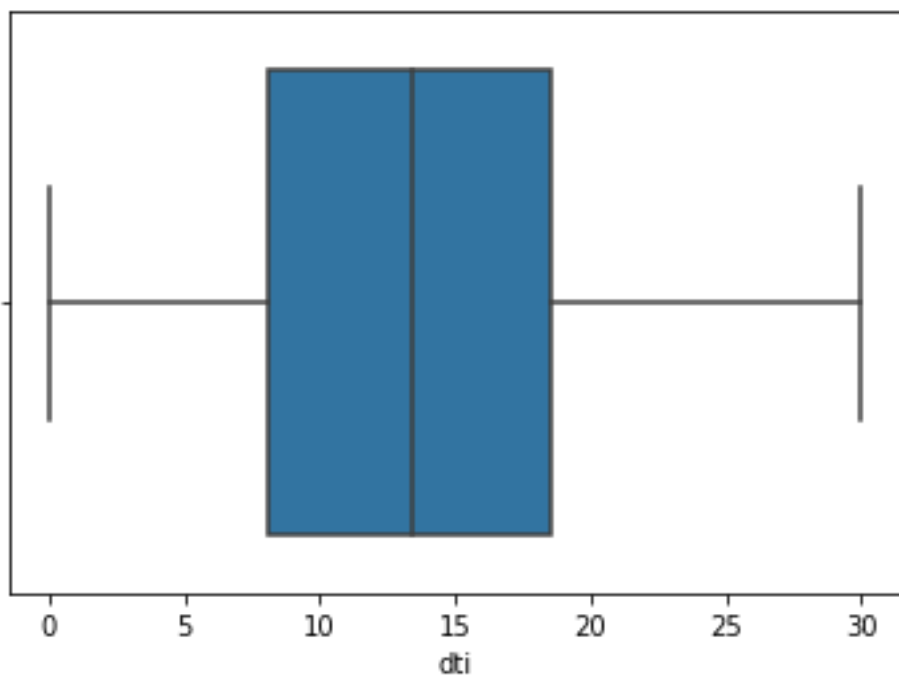


Step1-Data Understanding

1. Check for missing/Null values and drop columns with 100% Null value.
2. Check for columns which have only one unique value and drop as needed.
3. Check for relevancy of the columns whether they are related to pre or post approval of loan and if needed drop them.

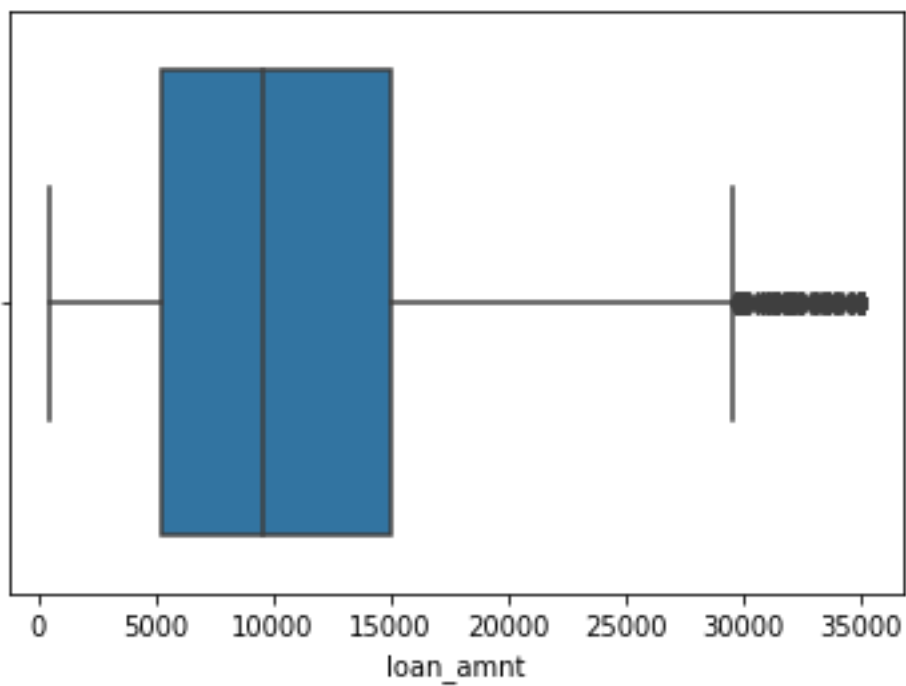
Step2-Data Cleaning and Manipulation

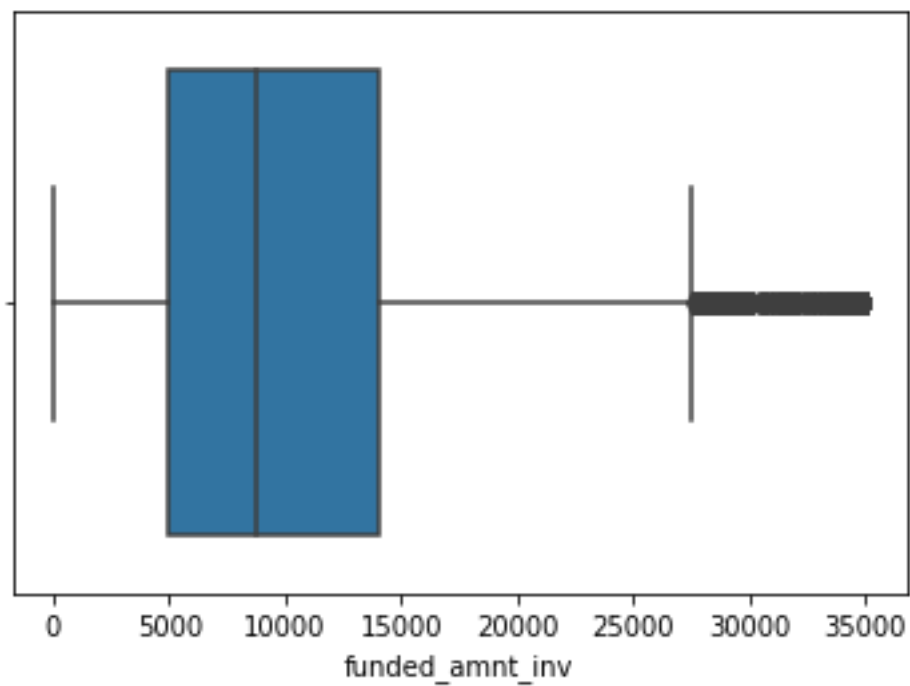
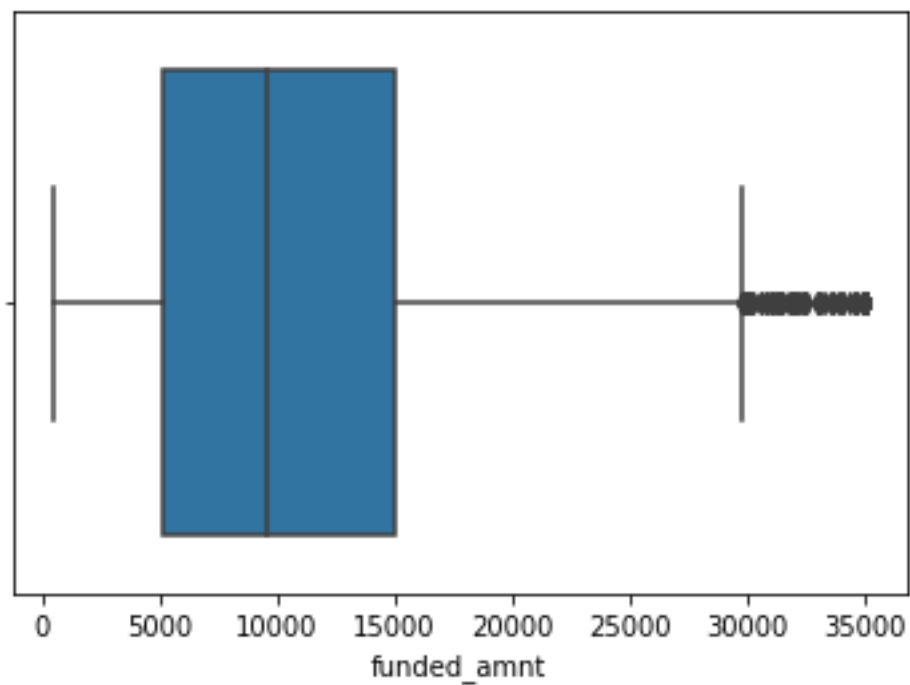
1. Handle missing values by various strategies such as 'mode' on case-to-case basis
2. Check the data type of the columns and convert to suitable format to be used by our libraries such as Pandas, numpy.
3. Detect the **outliers** and treat them as per the case, if it can be retained or dropped with proper explanation.



Observation:

'dti' data looks good and does not contain outliers



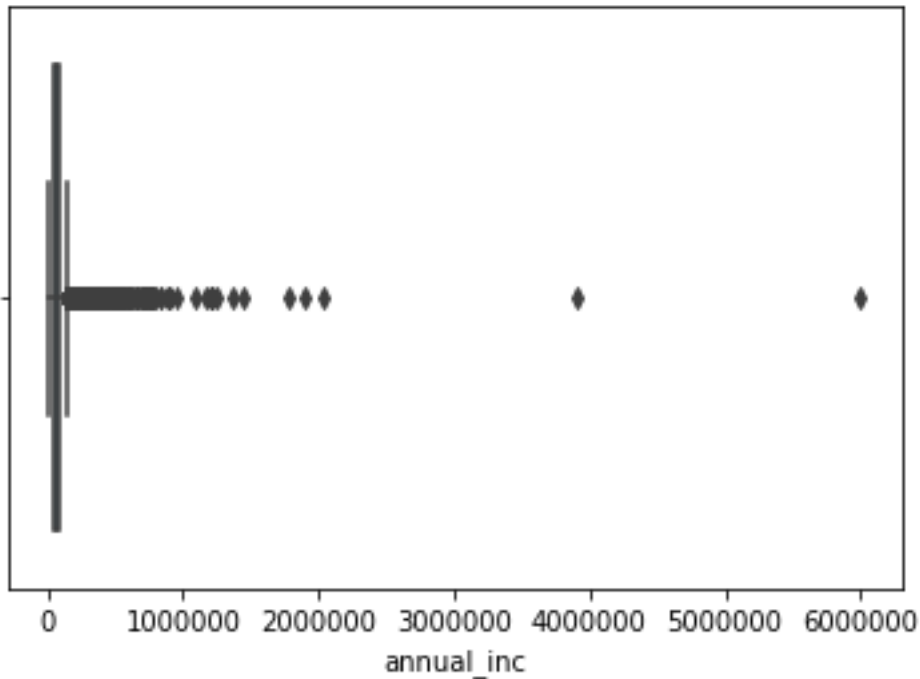


Observation:

1. 'loan_amnt' is the loan amount requested by a borrower; 'funded_amnt', 'funded_amnt_inv' are respectively the amount committed to the loan and amount committed by the investor.

2.All the three graphs look consistent with each other and related to each other. In business terms it means that certain customers requested for higher loan and same was committed to them.

3.Looking at the consistency of the three variables, in my opinion they don't need outlier treatment.

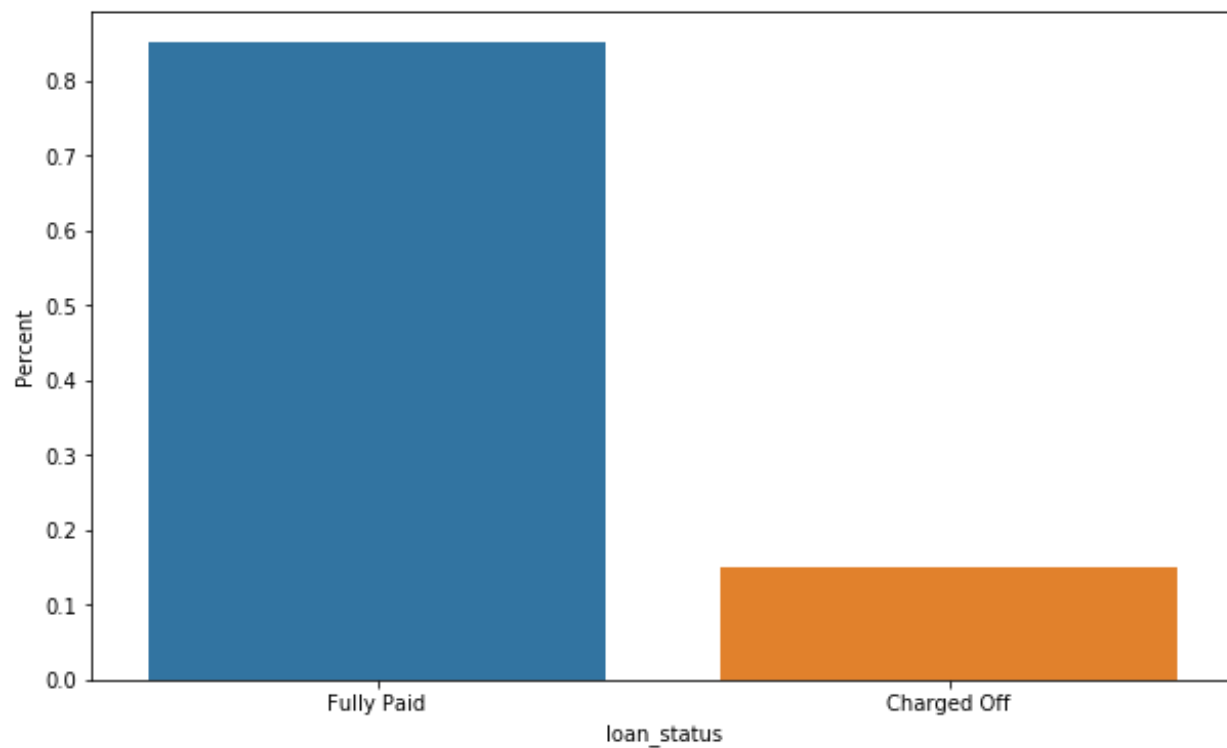


Observation:

There are few outliers which will impact our analysis by skewing it ,so we will remove the outlier beyond quantile 0.9

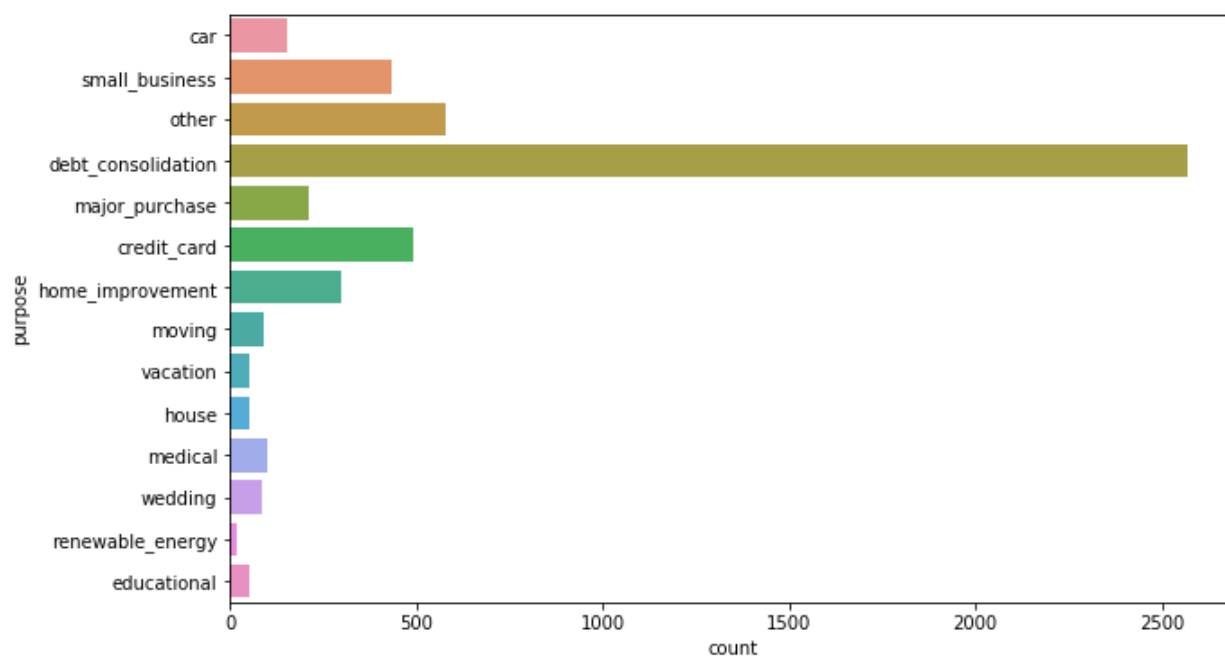
Step3-Data Analysis

3a:Univariate & Segmented Univariate Analysis



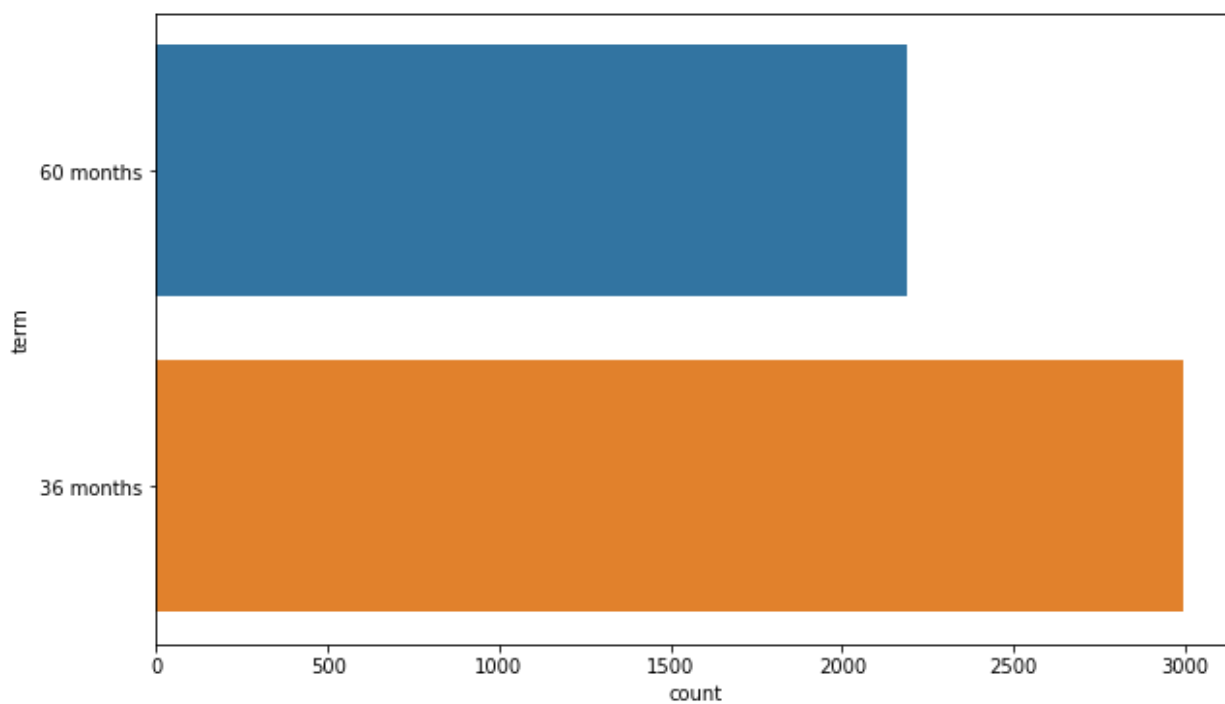
Observation:

85% of the borrowers have fully paid the loan while around 15% is Charged Off



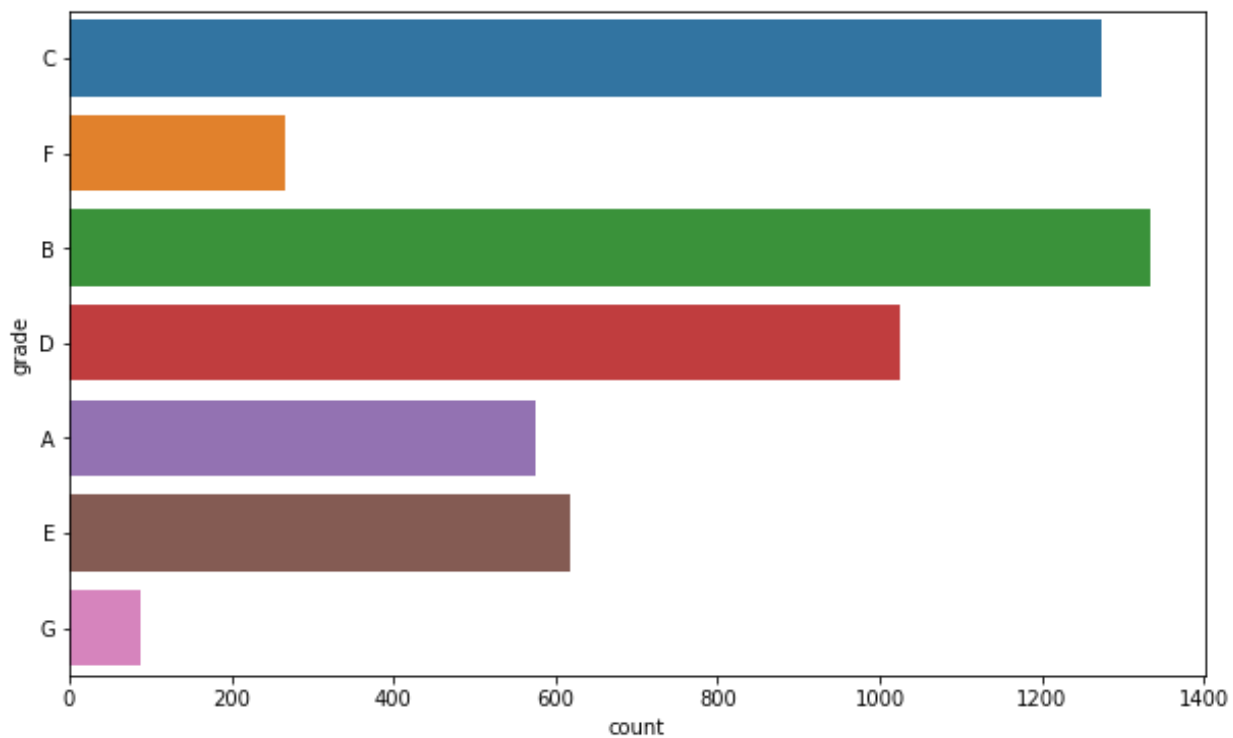
Observation:

The loans taken for debt consolidation are Charged Off most often compared to loans taken for other purpose



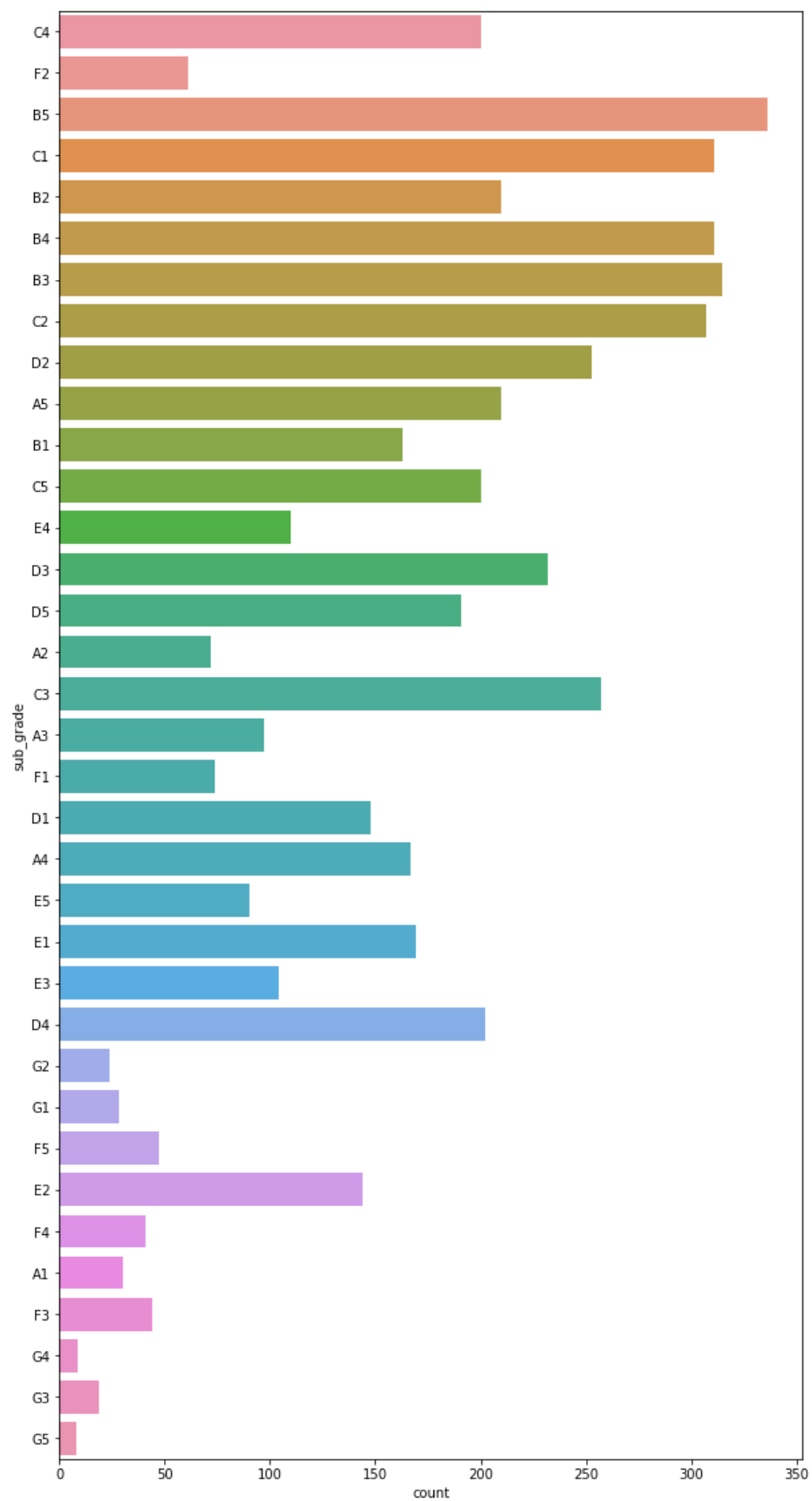
Observation:

36 months term loan are Charged Off more frequently than 60 months



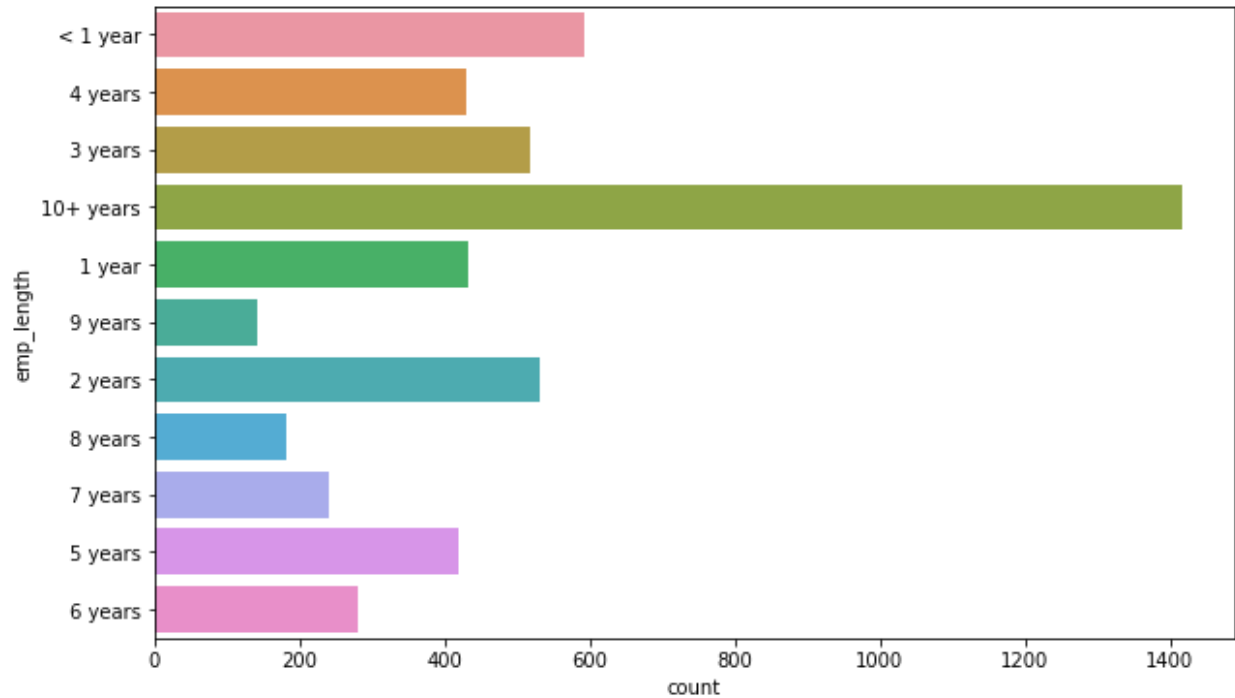
Observation:

B,C followed by D grade loan are Charged Off more frequently compared to other grades.



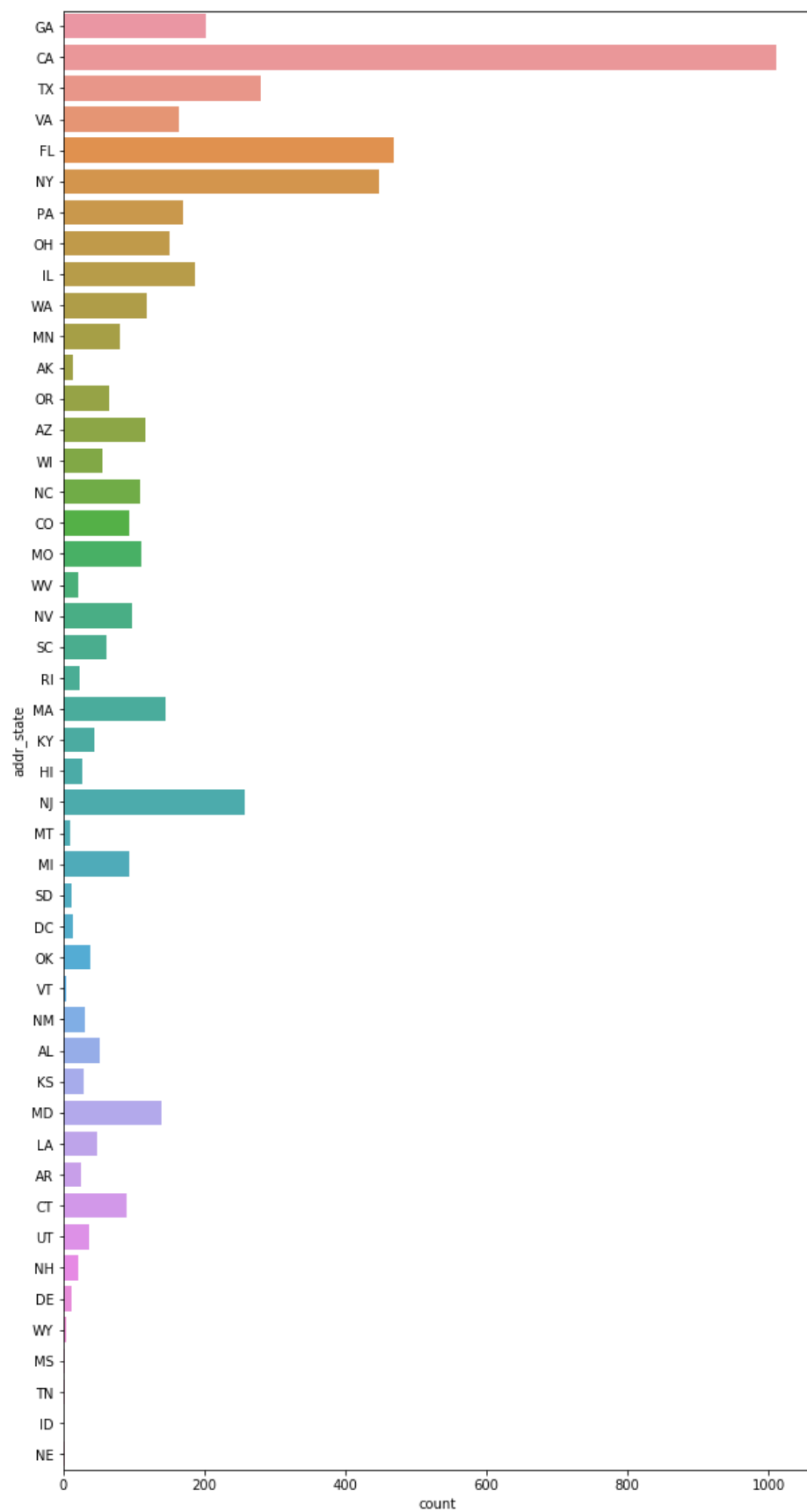
Observation:

B5 followed by B3 sub-grade loan are Charged Off more frequently compared to other sub-grades



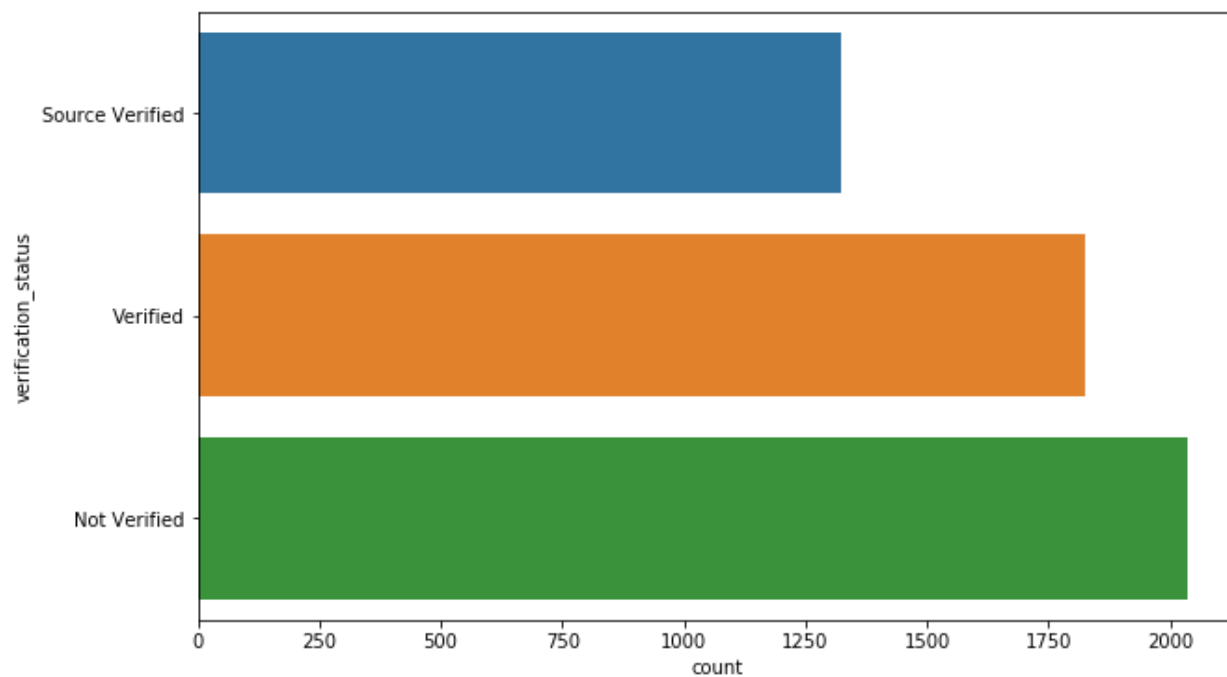
Observation:

Employees who are working for 10+ years have most 'Charged Off' loans



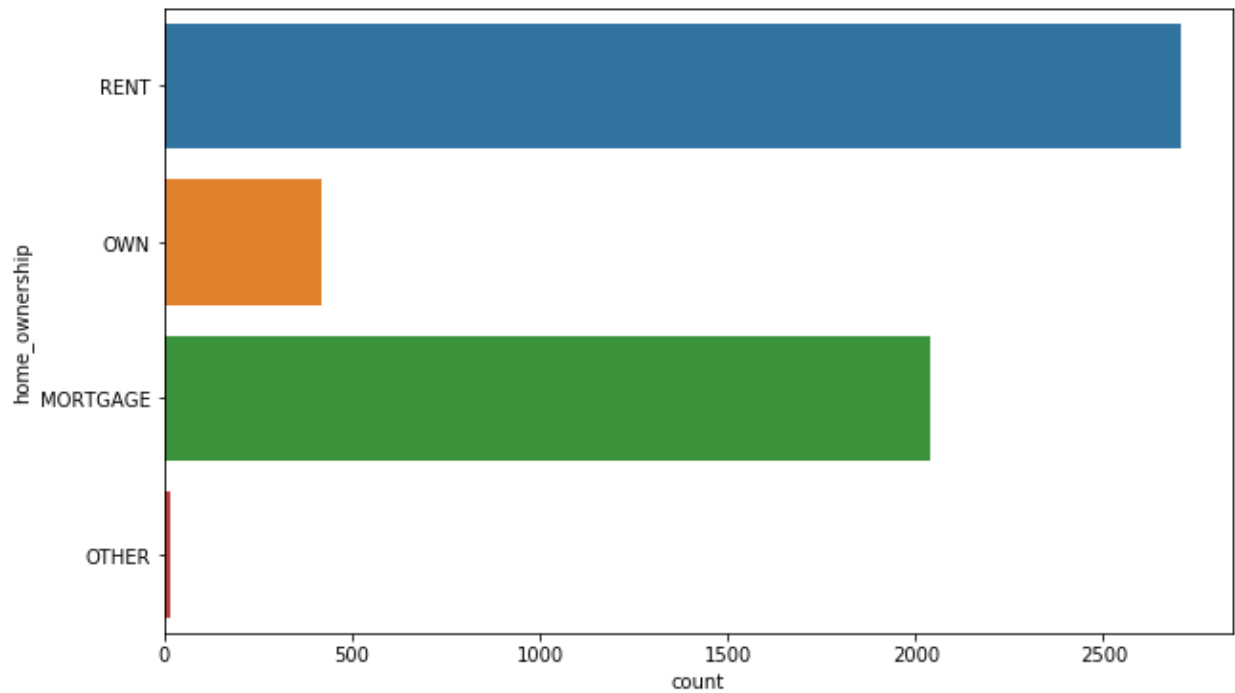
Observation:

Loans in the state of CA(California) are "Charged Off" most often so banks in CA have to be more vigilant before sanctioning loan.



Observation:

Loans with 'verification_status' as 'Not Verified' are 'Charged Off' most often.

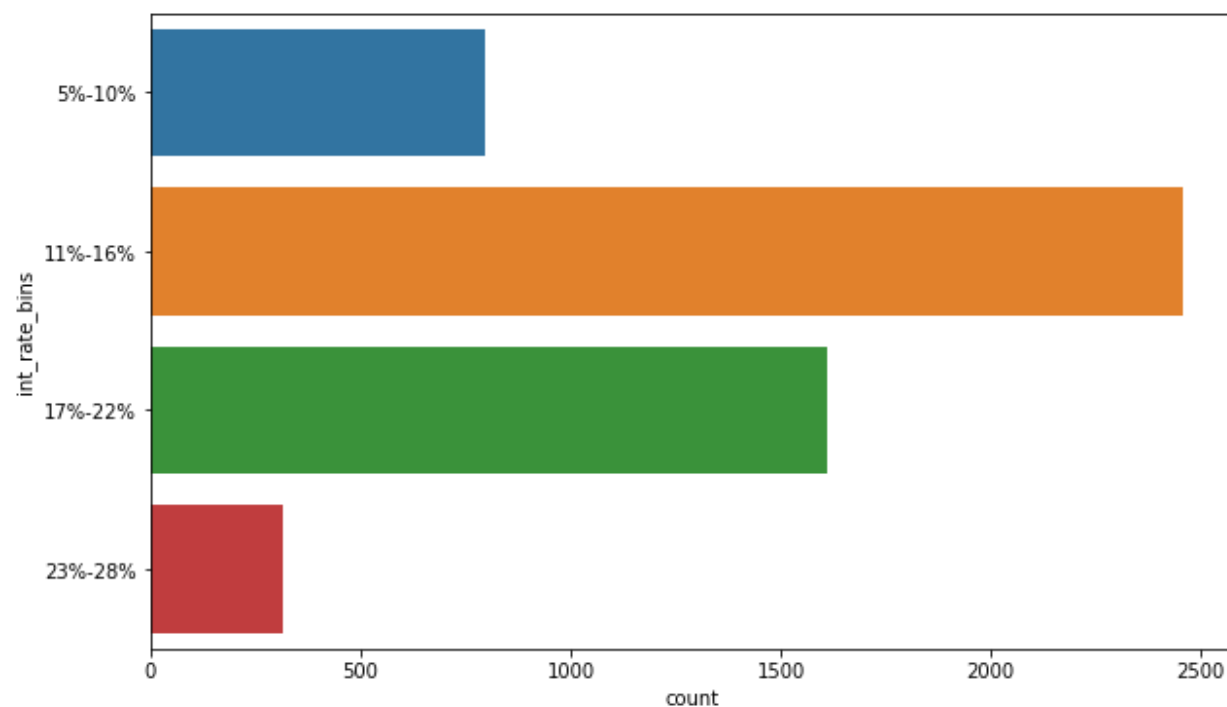


Observation:

People who are in rented home are Charged Off most often

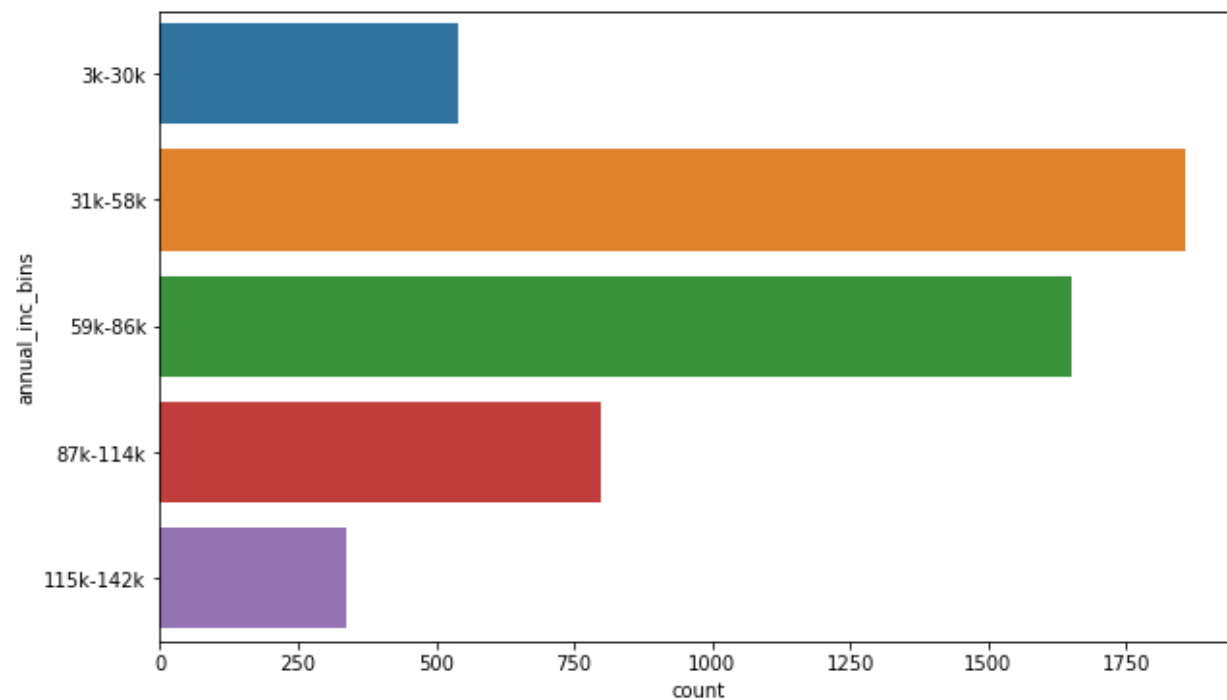
3b:Derived Variables

- 1.Create derived variables which better explain the data.
- 2.Create bins for Categorical variables to be able to plot and better explain the data.



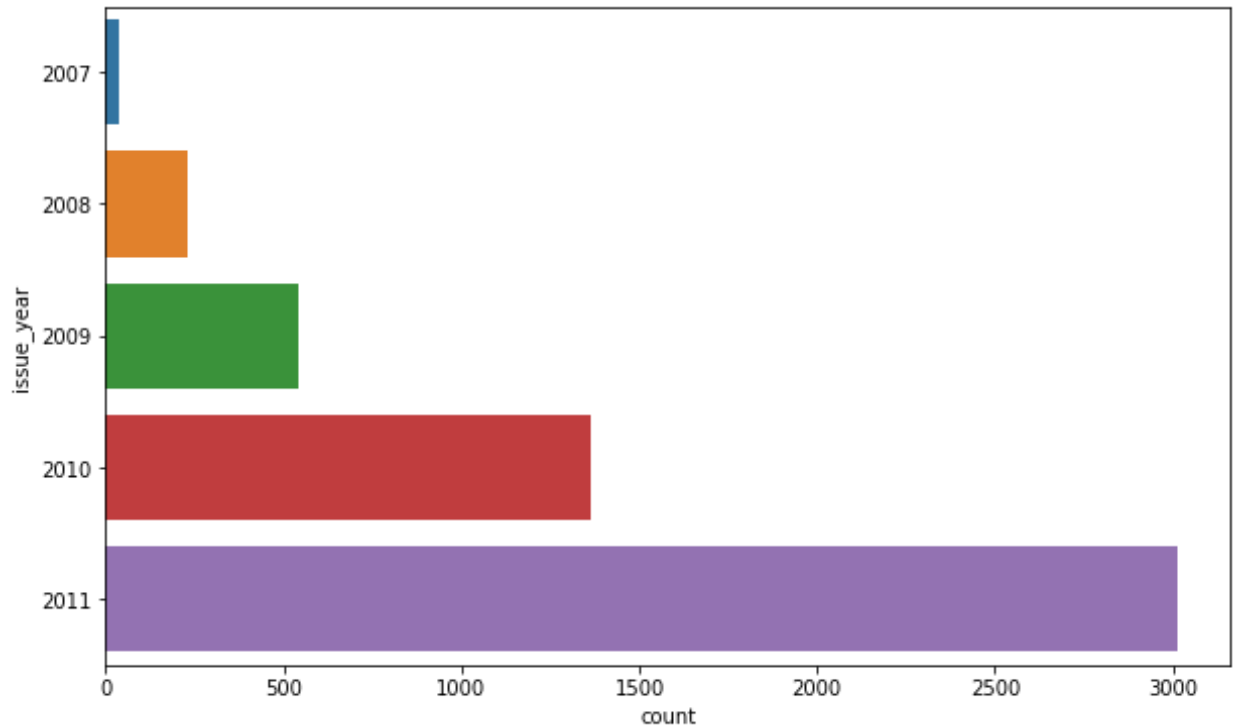
Observation:

Borrowers in interest rate group 11%-16% are most often 'Charged-Off'



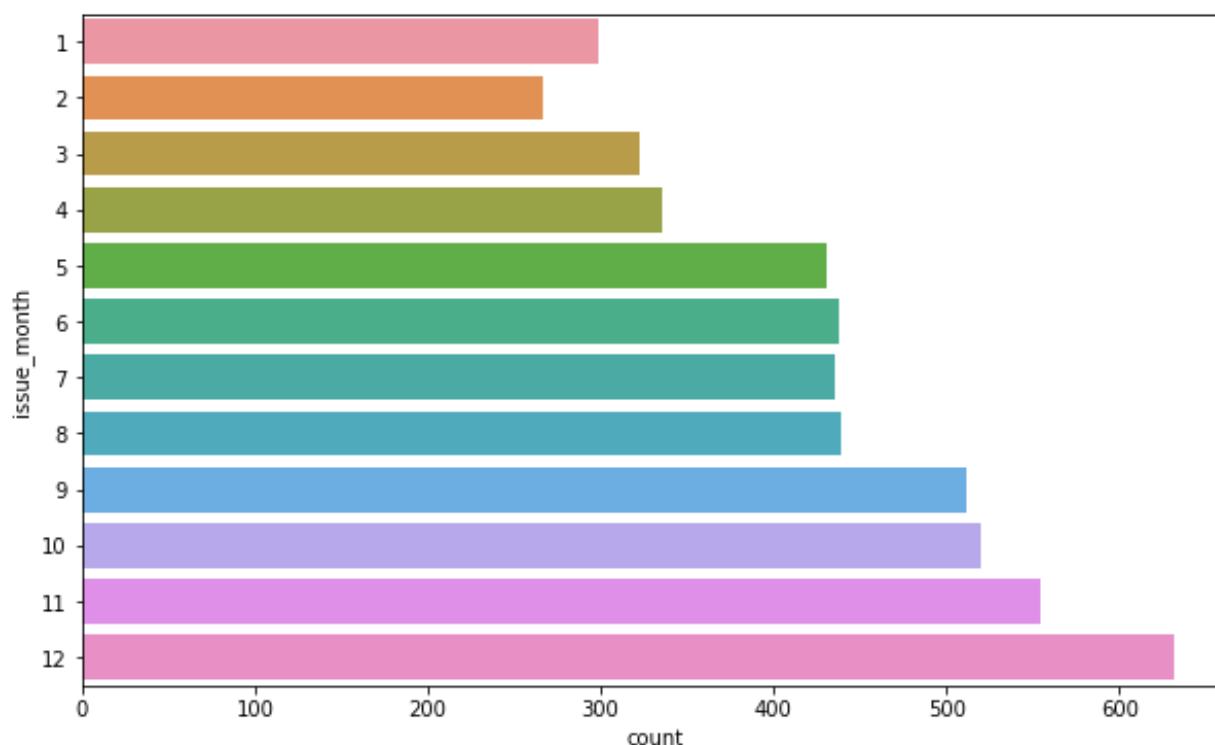
Observation:

Borrowers with annual income in the range 31K-58K are most often 'Charged-Off'



Observation:

Most loans were charged off in the year 2011



Observation:

Most loans were charged off in the 12th month i.e. December

Factors leading to higher Charge-off based on Univariate & Segmented Univariate analysis--

1. 85% of the borrowers have fully paid the loan while around 15% is Charged Off.
2. The loans taken for debt consolidation are Charged Off most often compared to loans taken for other purpose.
3. 36 months term loan are Charged Off more frequently than 60 months.
4. Loans of Grades B,C followed by D are Charged Off more frequently compared to other grades.
5. Loans of sub-grade B5 followed by B3 are Charged Off more frequently compared to other sub-grades.
6. Employees who are working for 10+ years have most 'Charged Off' loans.
7. Loans in the state of CA(California) are 'Charged Off' most often so banks in CA have to be more vigilant.
8. Loans with verification_status as 'Not Verified' are 'Charged Off' most often.
9. People who live in rented home are Charged Off most often.
10. Most commonly 100% of the requested loan amount was funded by investors.
11. Most commonly the annual income of the borrower is 3,4 or 5 times of the loan amount approved by investors.
12. Borrowers in interest rate group 11%-16% are most often 'Charged-Off'.
13. Borrowers with annual income in the range 31K-58K are most often 'Charged-Off'.

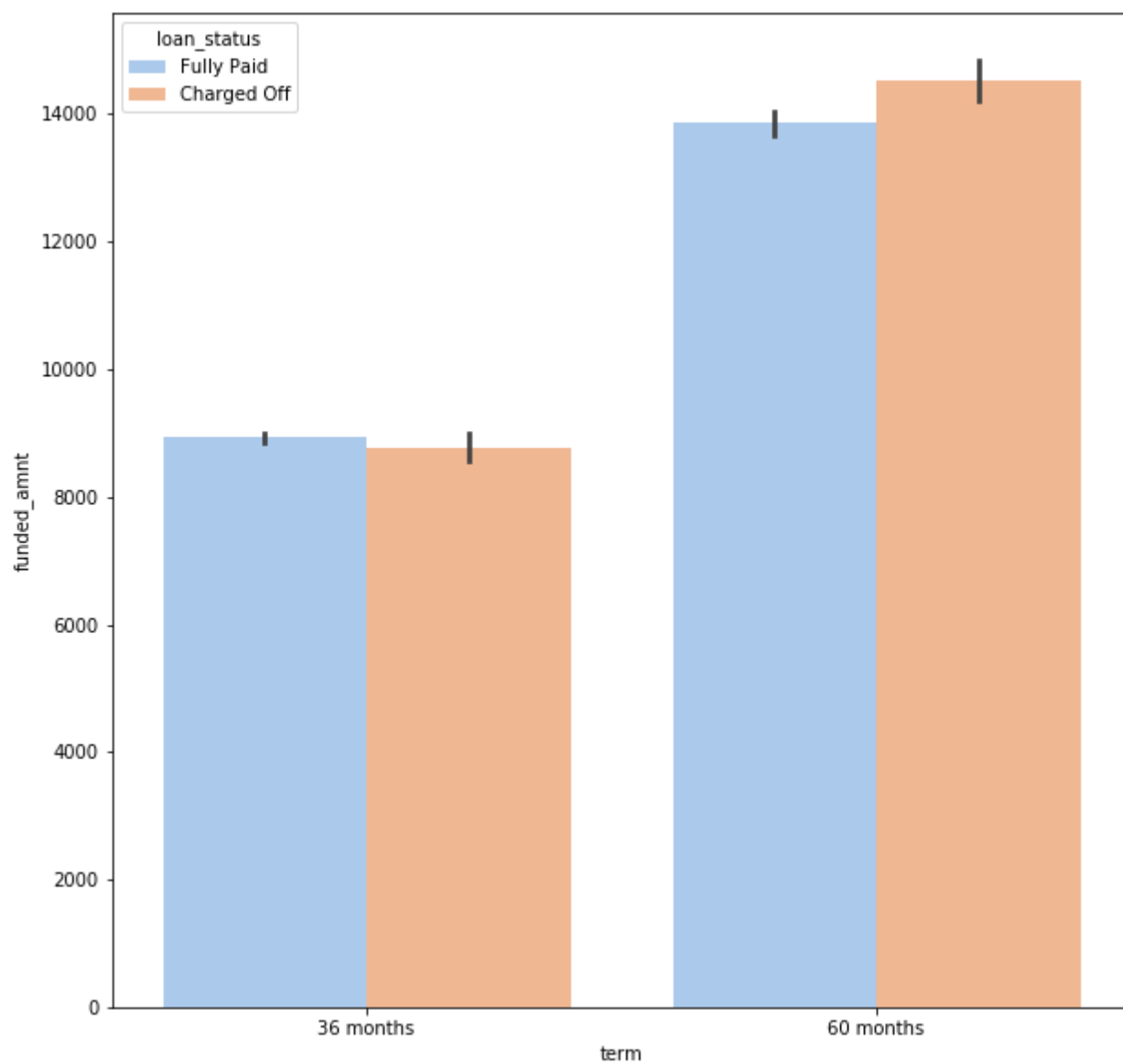
14. Most loans were charged off in the year 2011.
15. Most loans were charged off in the December month.

3c:Bivariate Analysis

Use Correlation Matrix to explain relationships between variables.

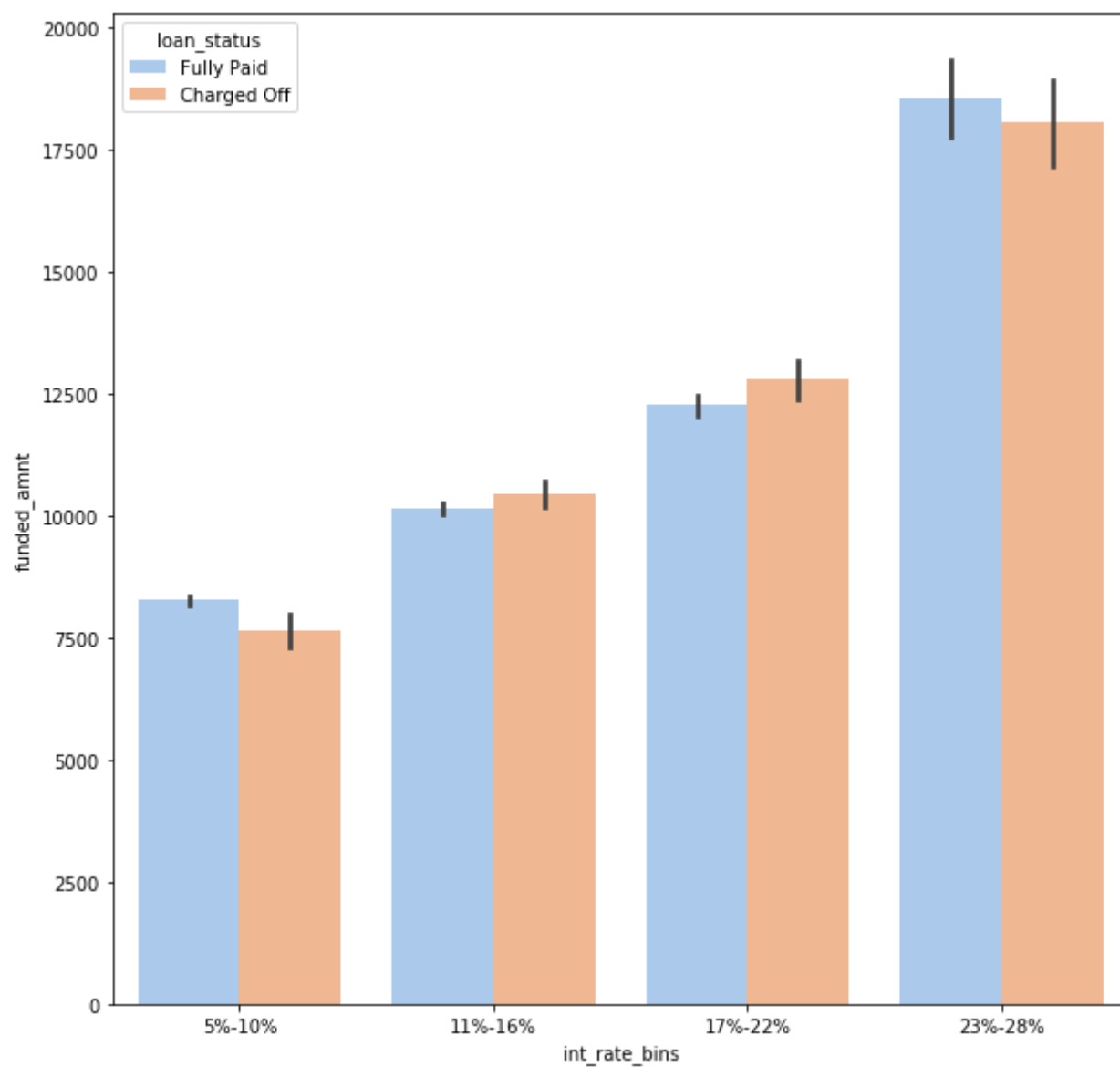
Observations based on Correlation Matrix--

1. loan_amnt(amount for which loan is applied) is strongly correlated with 'funded_amnt','funded_amnt_inv' and 'installment'.
2. 'loan_amnt','funded_amnt','funded_amnt_inv' and 'installment'are negatively correlated to 'pub_rec','pub_rec_bankruptcies'.
3. The number of open credit lines('open_acc') is negatively correlated to interest rate('int_rate').
4. 'revol_bal'(revolving balance) and 'revol_util'(revolving utilization rate) are negatively correlated to 'delinq_2yrs'(30+ days past due incidences) ,'inq_last_6mths'(inquiry in last 6 months)
5. 'pub_rec_bankruptcies'(number of public record bankruptcies) are negatively correlated to 'last_pymnt_amnt'(last payment)
6. Annual income('annual_inc') is negatively correlated to Debt to income ratio('dti')



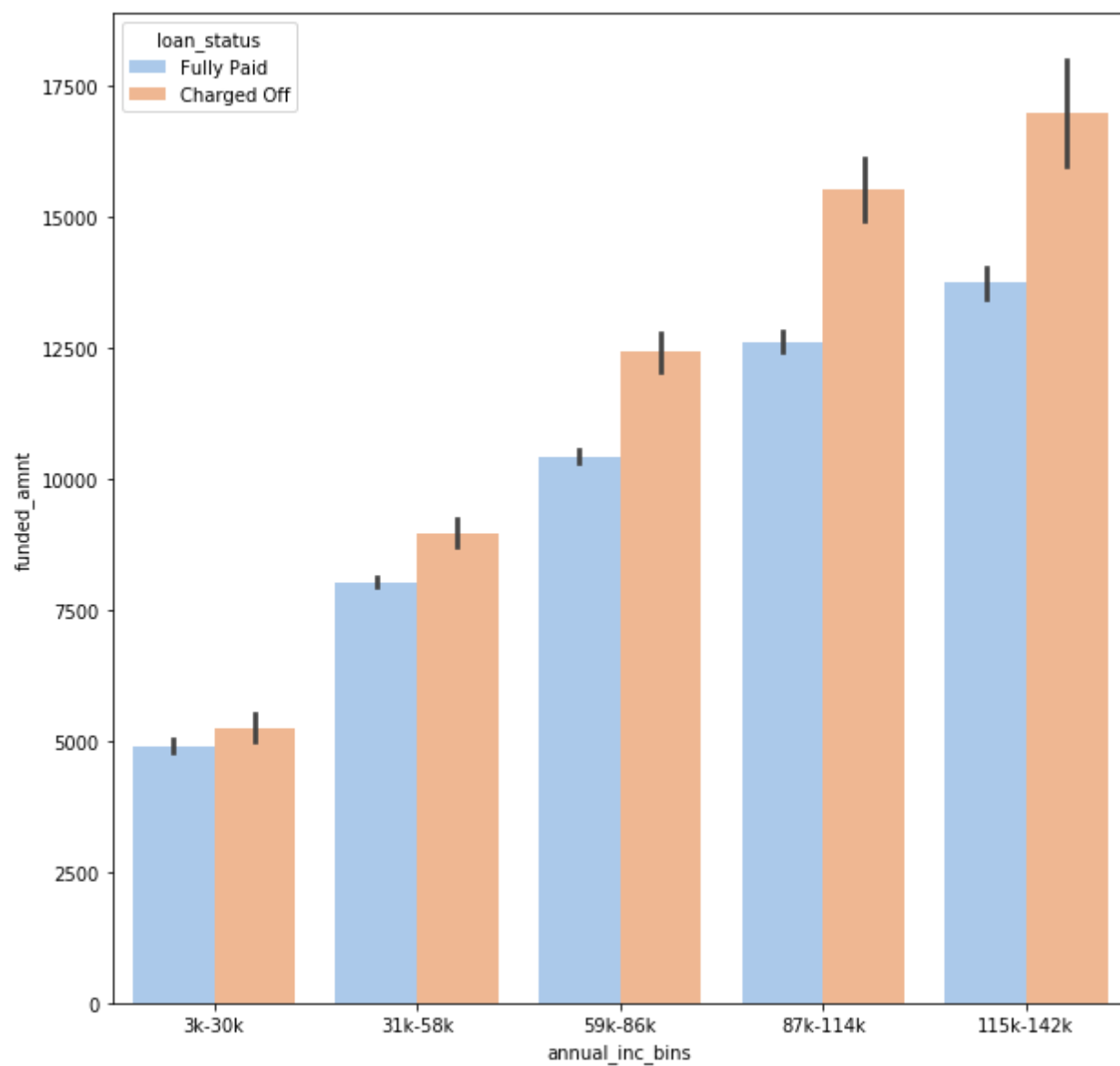
Observation:

60 months term loan have higher funding amount than 36 months and accordingly both Fully Paid and Charged Off is higher for 60 months



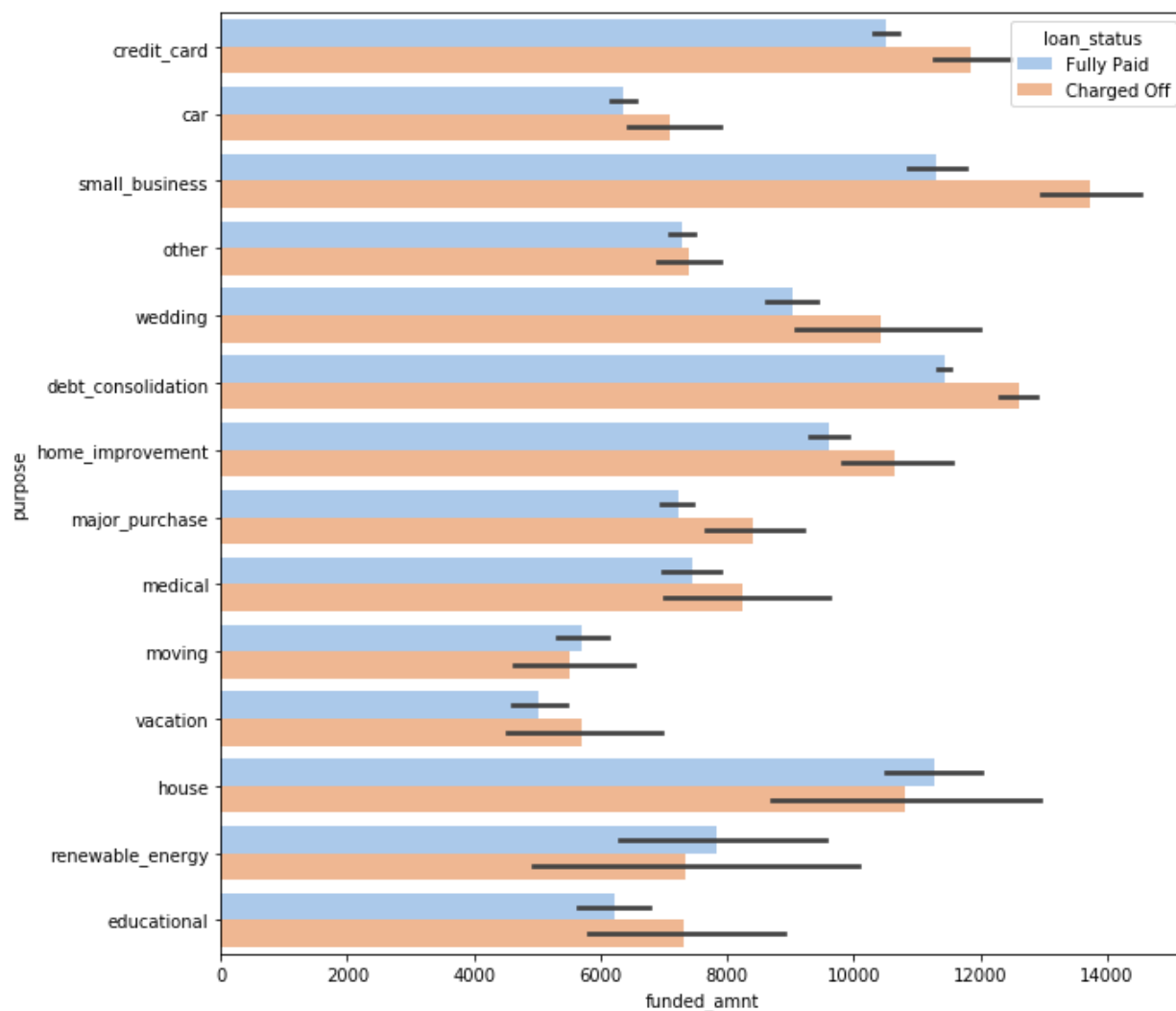
Observation:

Higher value loans are funded at higher interest rate, which is good source of business for bank.



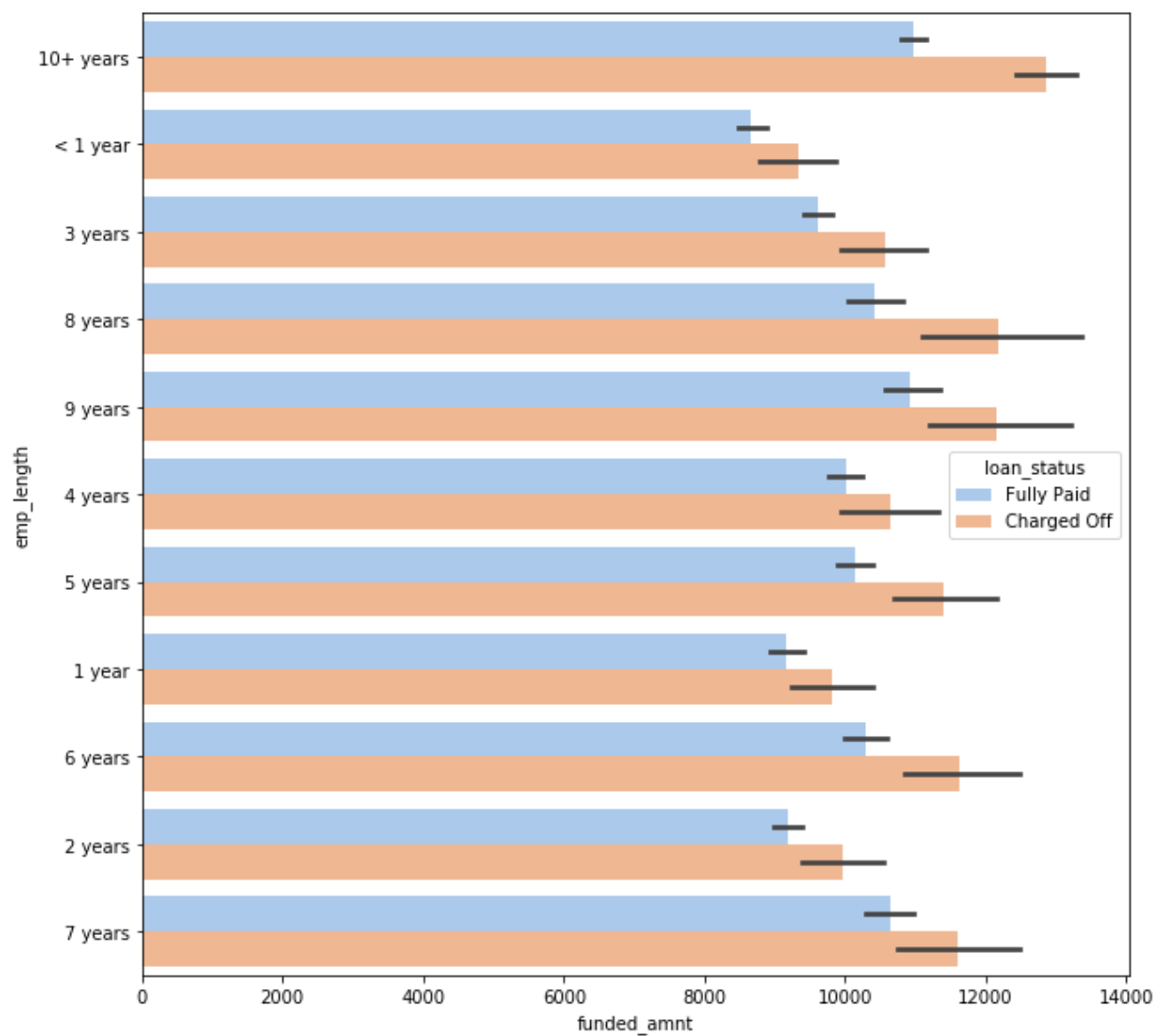
Observation:

Higher value loans are funded to high income group borrowers



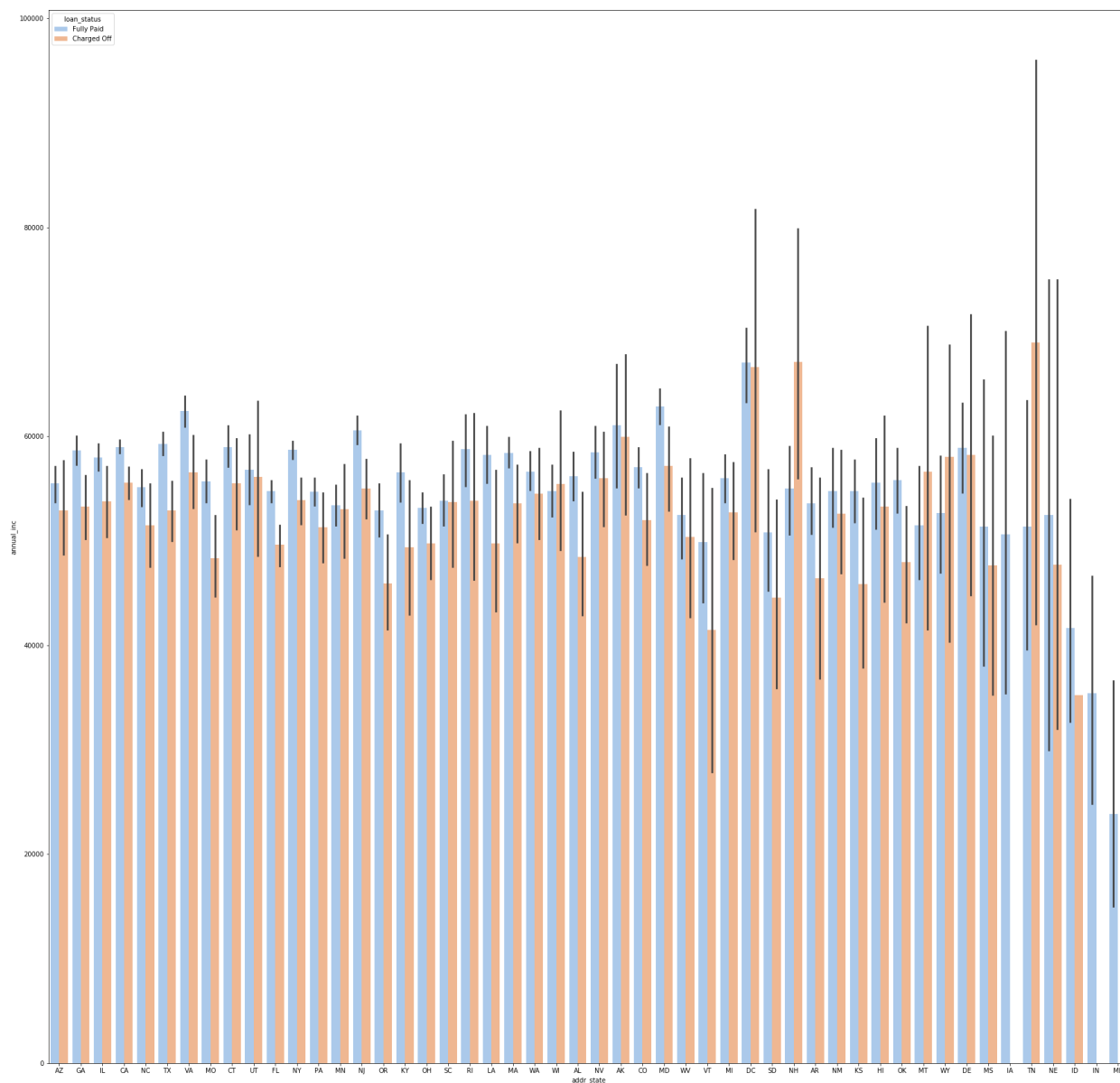
Observation:

1. Most high value loans are funded for the purpose of Small business, Debt consolidation and Credit card.
2. If it's a high value loan and funded to Small Business Charge Off is more.



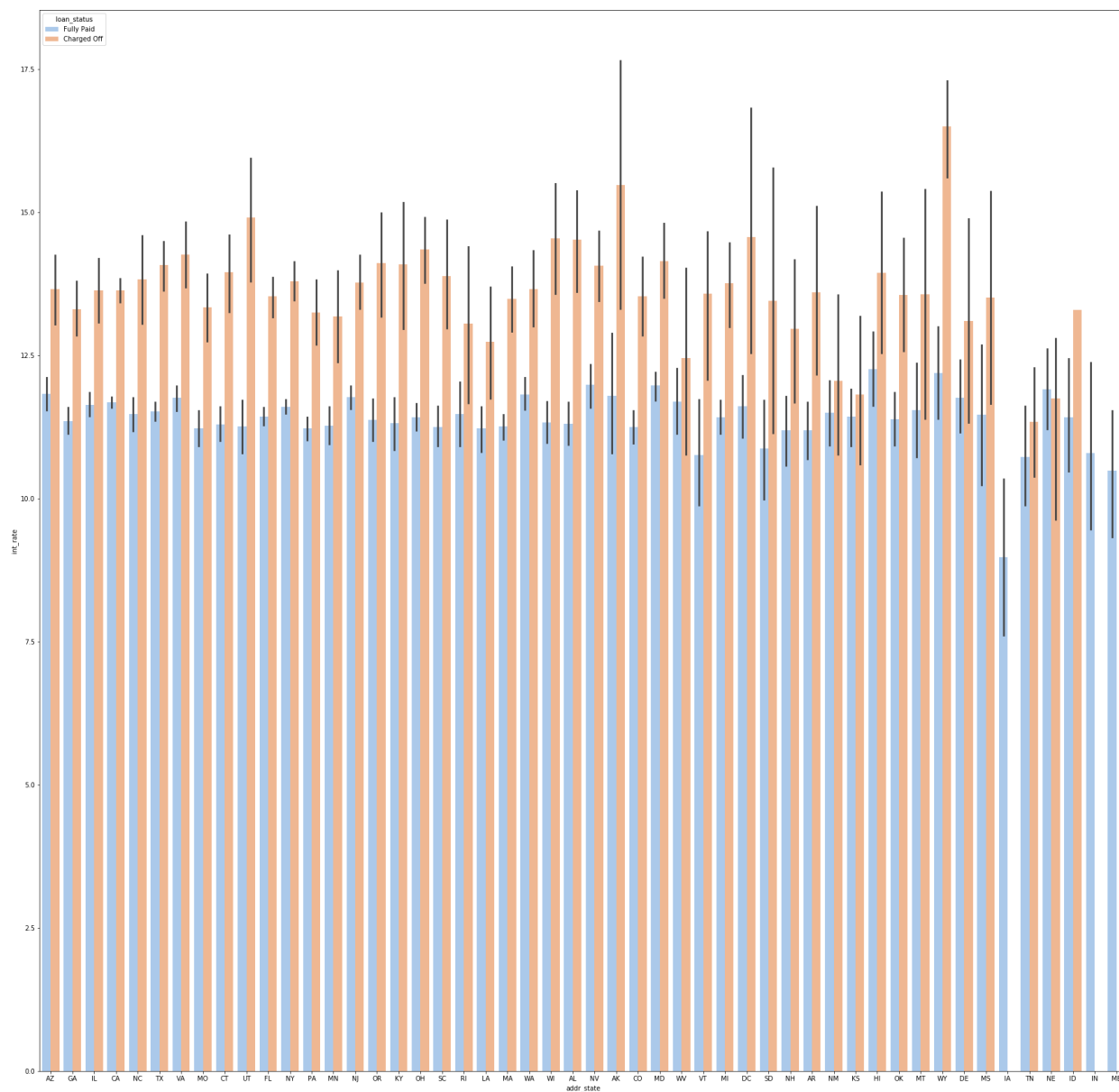
Observation:

Most high value loans are Charged Off for borrowers with 10+ years of experience.



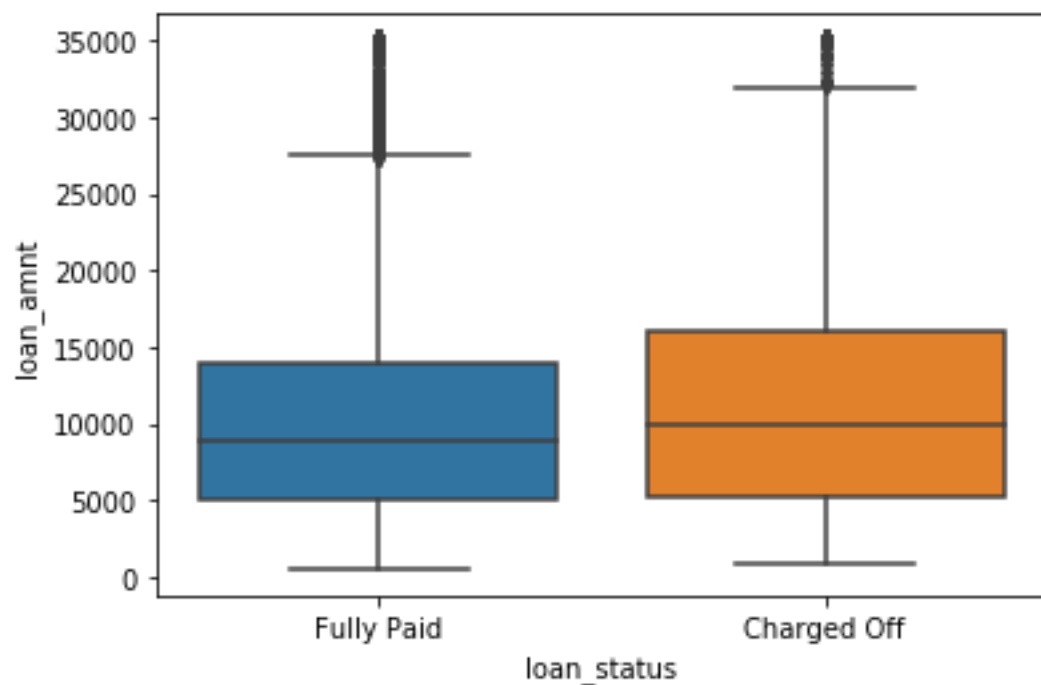
Observation:

- 1.If the borrower has high annual income and is from the state of 'TN' they are more likely to Charge Off
- 2.If the borrower has high annual income and is from the state of 'DC' they are more like to fully pay their loans.



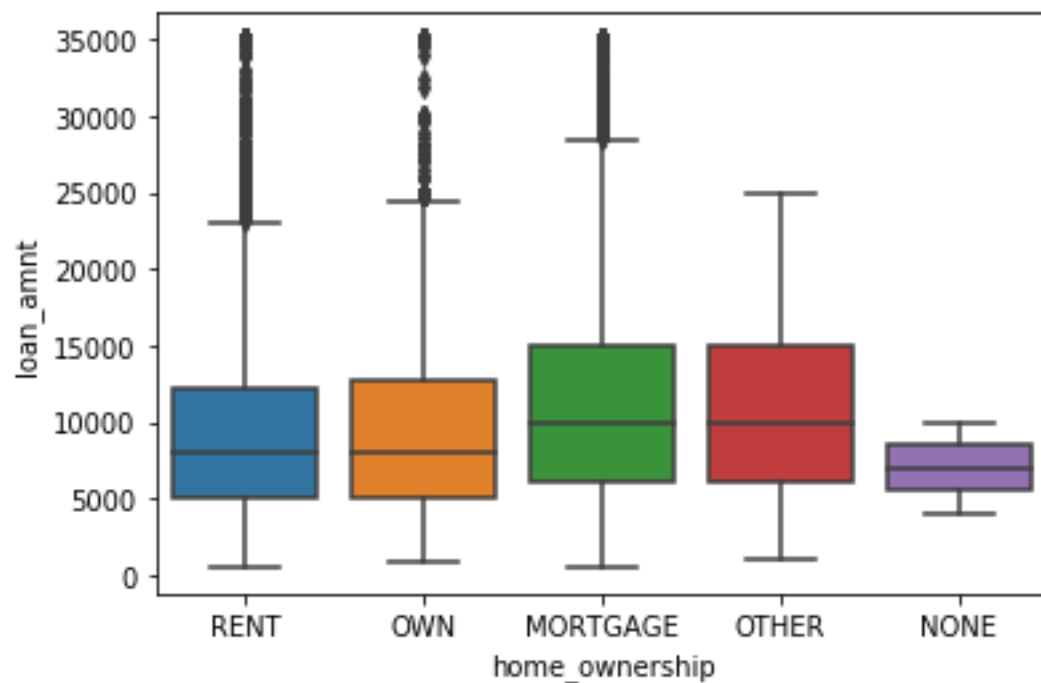
Observation:

State of 'WY' has highest interest rate both in Charged Off and Fully Paid loans



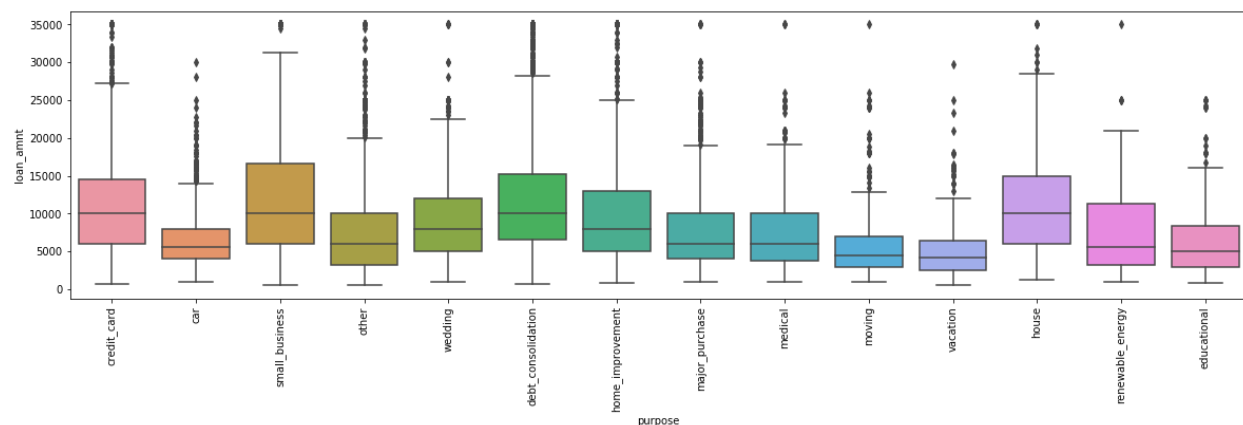
Observation:

Charged Off loans have higher loan amount than Fully Paid loans



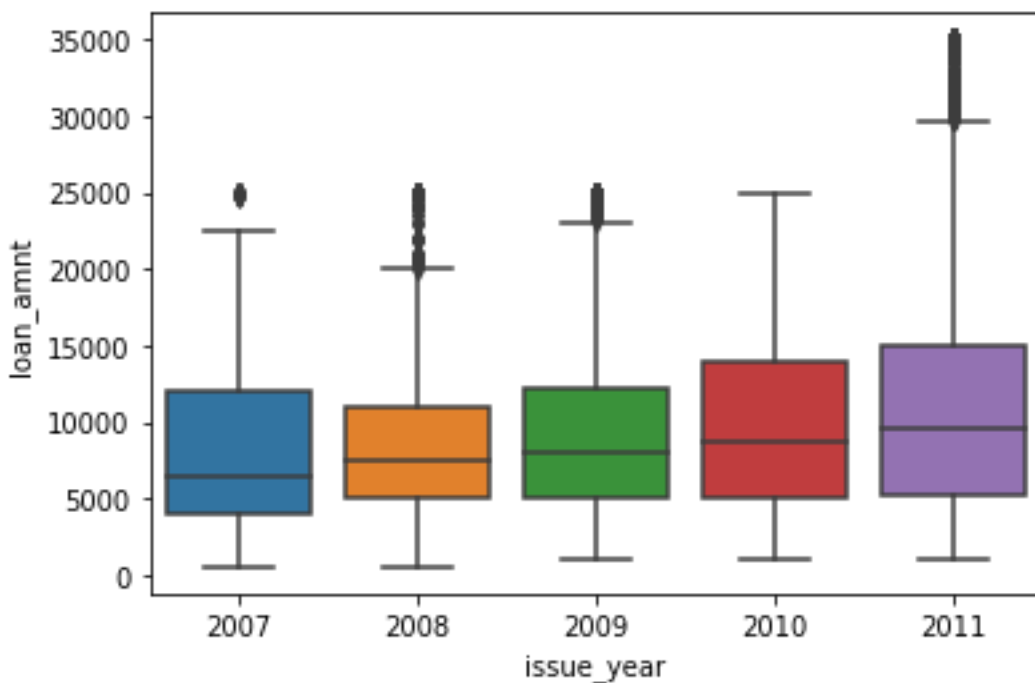
Observation:

Borrowers who live in mortgaged home have requested for higher loan amount



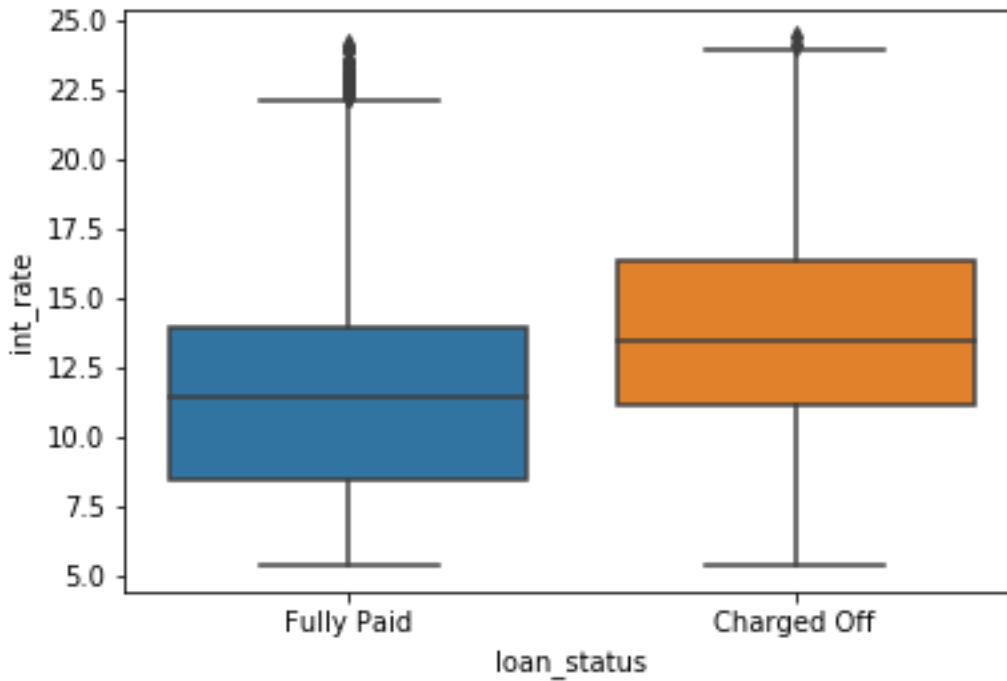
Observation:

Higher loan amount is requested for Small business and Debt reconciliation



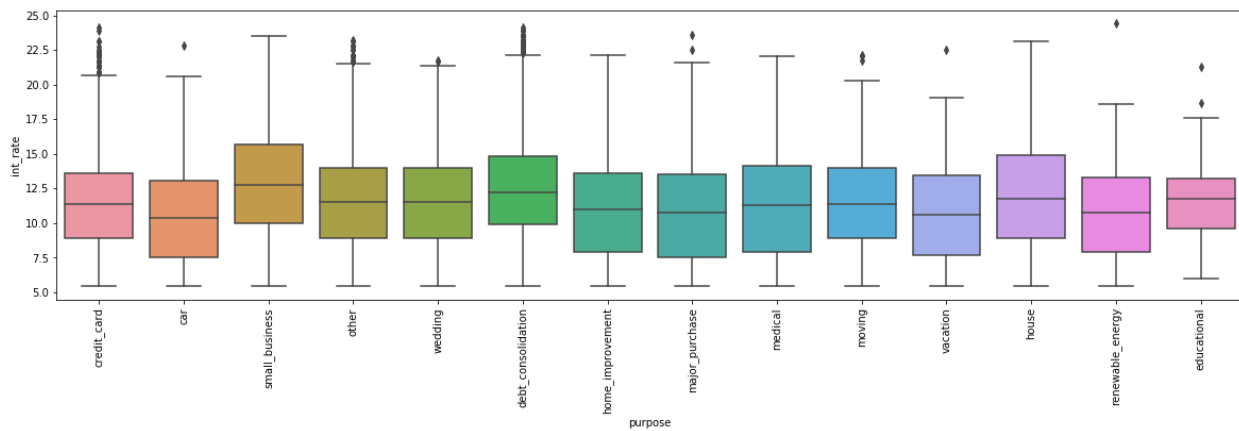
Observation:

The median value for requested loan amount is almost similar each year.



Observation:

Interest rates for Charged Off loans were higher compared to Fully Paid loans.



Observation:

Interest rates are higher for loans for the purpose of Debt Consolidation, Small Business and House.

Observations based on Bivariate analysis:

1. 60 months term loan have higher funding amount than 36 months and accordingly both Fully Paid and Charged Off loan is higher for 60 months.
2. Higher value loans are funded at higher interest rate, which is good source of business for bank.
3. Higher value loans are funded to high income group borrowers.
4. Most high value loans are funded for the purpose of small business, Debt consolidation and Credit card.
5. If it's a high value loan and funded to Small Business Charge Off is more.
6. Most high value loans are Charged Off for borrowers with 10+ years of experience.
7. If the borrower has high annual income and is from the state of 'TN' they are more likely to Charge Off.
8. If the borrower has high annual income and is from the state of 'DC' they are more like to fully pay their loans.
9. State of 'WY' has highest interest rate both in Charged Off and Fully Paid loans.
10. Charged Off loans have higher loan amount than Fully Paid loans.
11. Borrowers who live in mortgaged home have requested for higher loan amount.
12. Higher loan amount is requested for Small business and Debt reconciliation.
13. The median value for requested loan amount is almost similar each year.
14. Interest rates for Charged Off loans were higher compared to Fully Paid loans.
15. Interest rates are higher for loans for the purpose of Debt Consolidation, Small Business and House

Step4-Recommendations

Recommendations:

1. Be cautious while sanctioning loans taken for the purpose of Debt consolidation as they are Charged Off most often.
2. Loans of grade B,C are more Charged Off so need to be more cautious with verification and number of loans offered in these grades.
3. Loans of sub-grades B5 followed by B3 are more Charged Off, so verify the loan application more cautiously.
4. In general, the state of 'CA' has highest Charged Off loan, but among the high income group of borrowers the state of 'TN' has highest Charged Off loan.
5. The bank should verify the loan applications thoroughly since Charge Off is highest among loans with status 'Unverified'.
6. People who live in Rented homes have higher Charged Off loans, so bank needs to be cautious of such applicants.
7. The loans in the interest range 11-16% are most often charged so bank need to set the right interest rate.

8. Borrowers with annual income in the range of 31K-58K are most likely to have Charged Off loan, so bank should target borrowers in other income group.
9. Bank should continue charging higher interest rate for high value loan as it's good business value.
10. Bank should continue funding high value loan to high income group because they can repay it.
11. Bank should strictly avoid\limit funding high value loan to 'Small business' because Charge Off is high there. Small business doesn't seem to handle high value loan very well.
12. Prefer offering high value loan in the state of 'DC'(where completely paying loan is highest) and try to limit as much possible in the state of 'TN'.