

A vibrant, abstract graphic on the left side of the slide, featuring a diagonal brushstroke of rainbow colors (red, orange, yellow, green, blue, purple) against a white background.

Lending Club Case Study

EXPLORATORY DATA
ANALYSIS

Data analysis report on loan data

- **Problem Statement**

The goal of the analysis is to see who is most likely to default and find the derived factors/variables for that.

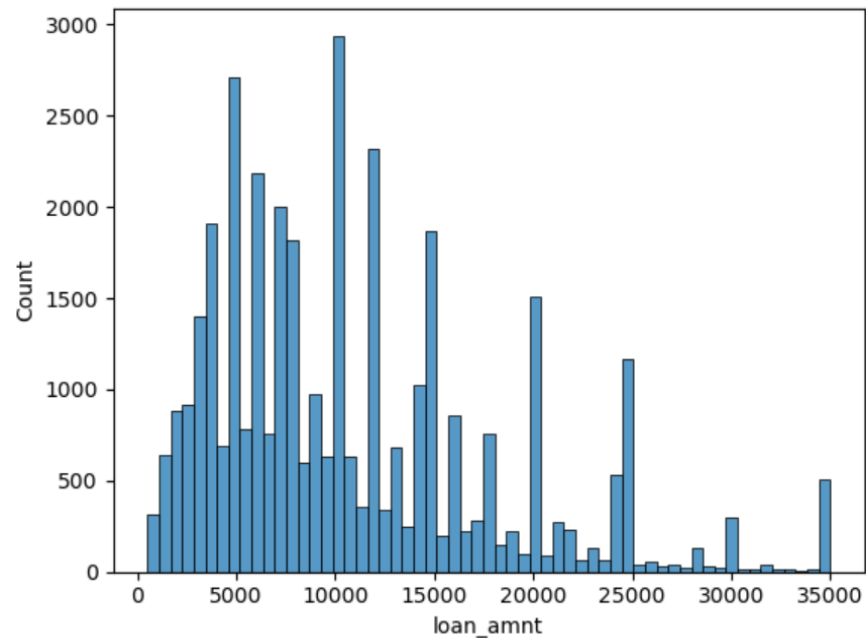
EDA Process

- **Data Loading**
- **Data Cleaning**
 - Remove all missing value
 - Impute values based on Mean, median and mode which is suitable for the column
- **Data Analysis**
 - Standardising Data
 - Univariate Analysis
 - Bivariate Analysis
 - Insights from above analysis
 - Observations

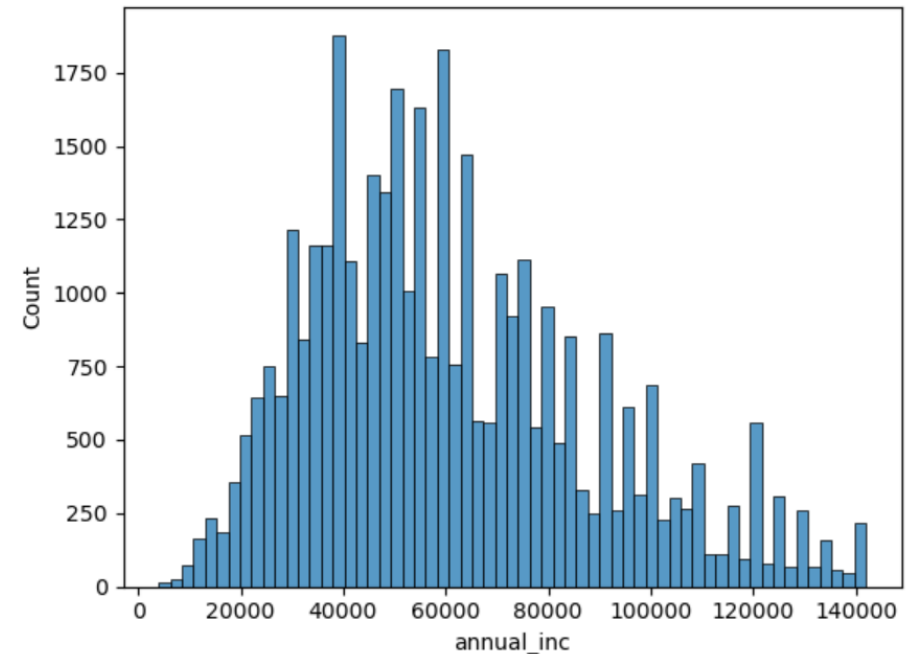
Univariate Analysis

- 1 Numerical columns
- Loan_amnt, annual_inc

Histplot of loan_amnt

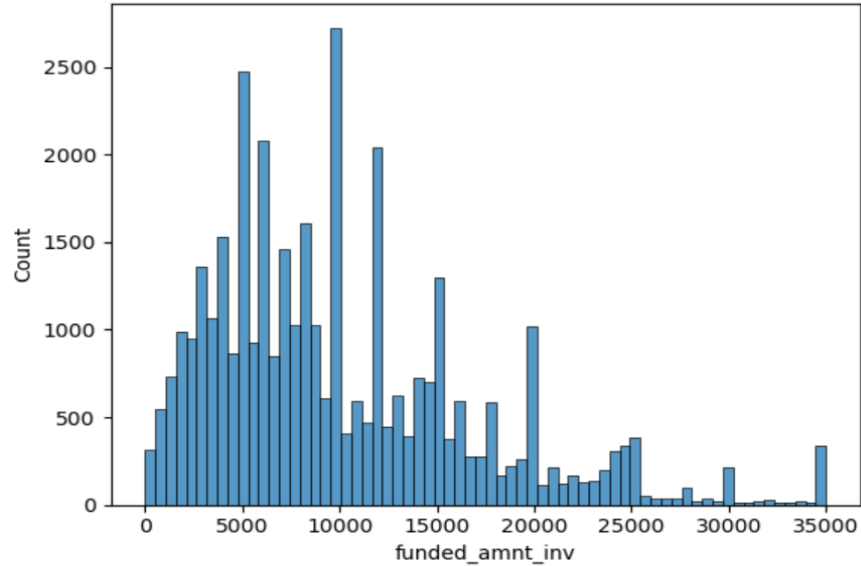


Histplot of annual_inc

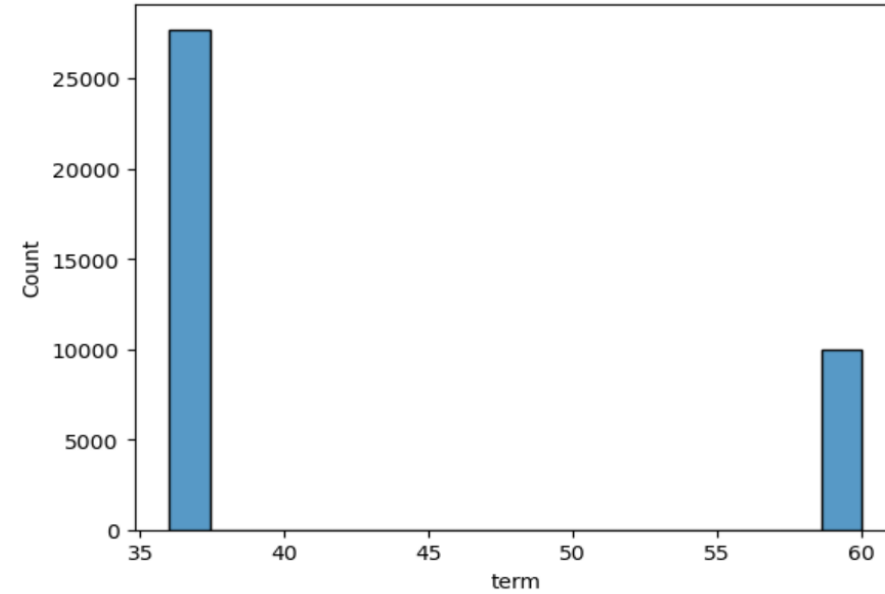


Funded_amnt_inv , term, interest rate and installment

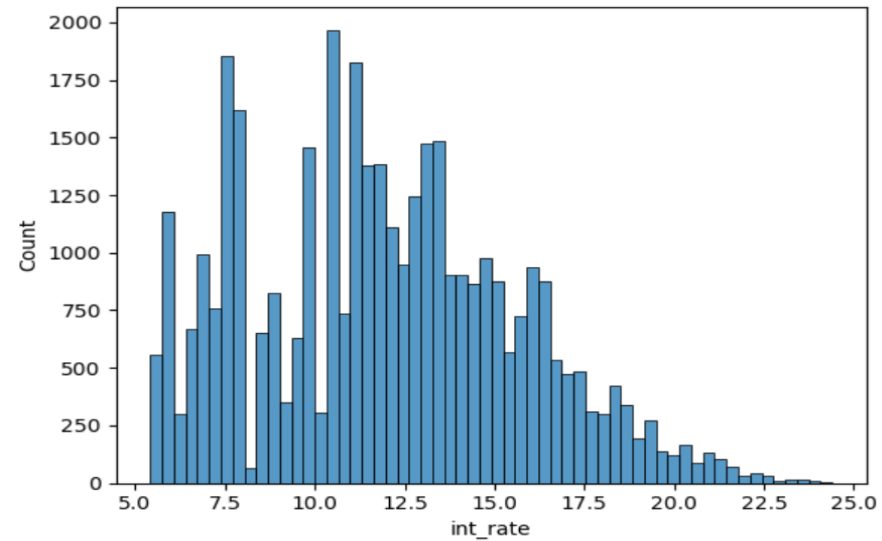
Histplot of funded_amnt_inv



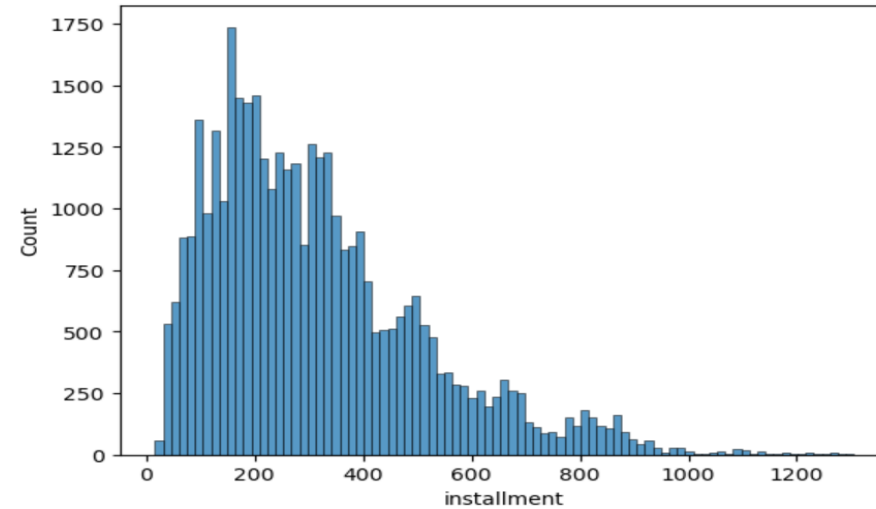
Histplot of term



Histplot of int_rate

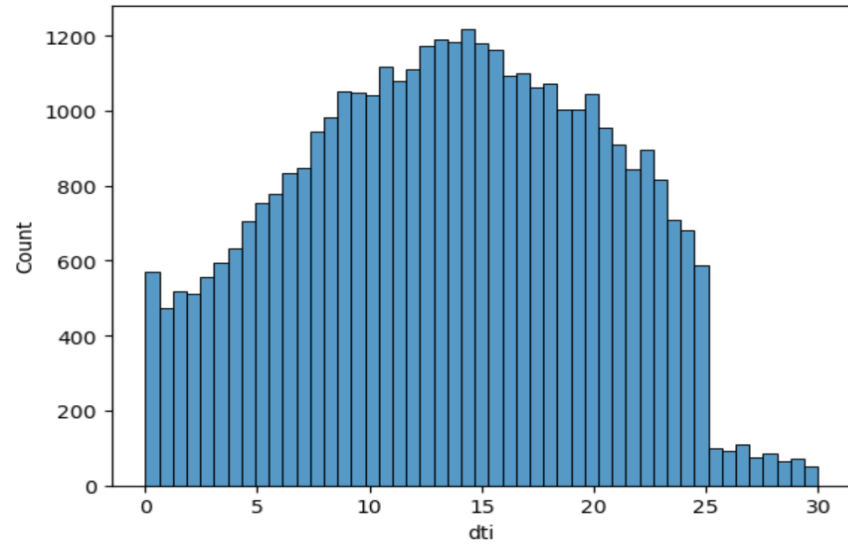


Histplot of installment

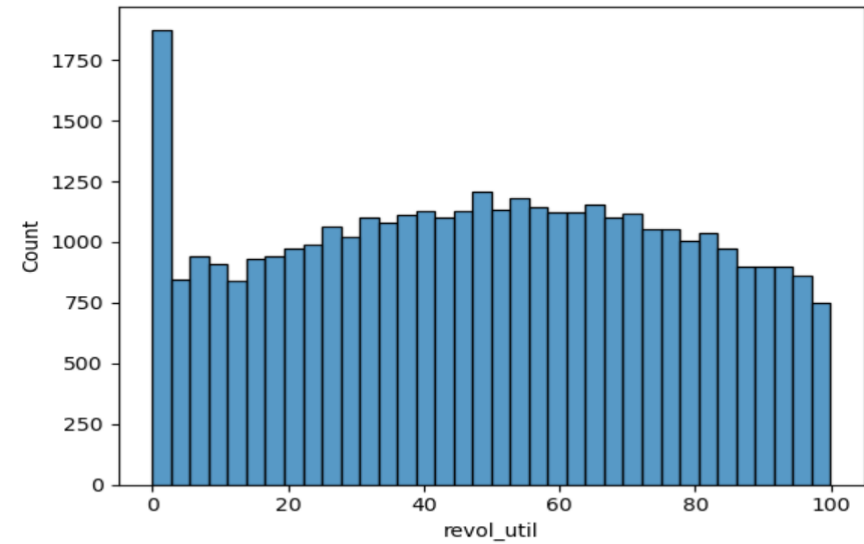


Dti, revol_util, total_acc and total_pymnt

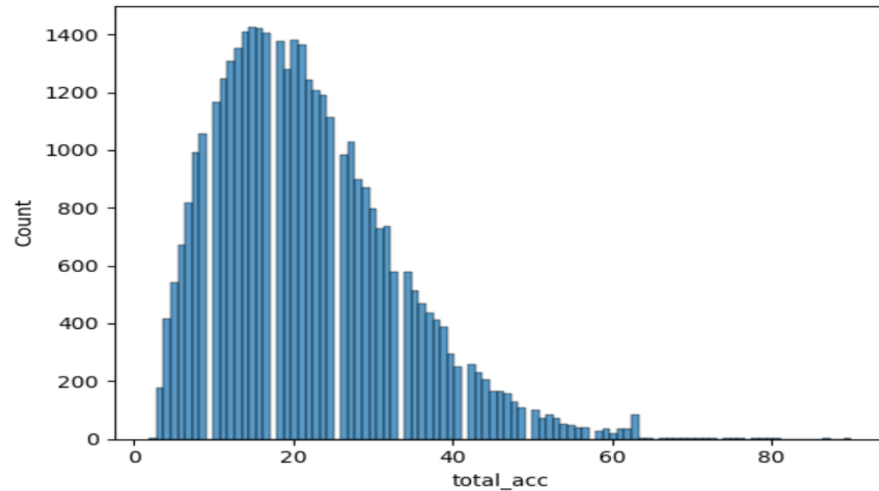
Histplot of dti



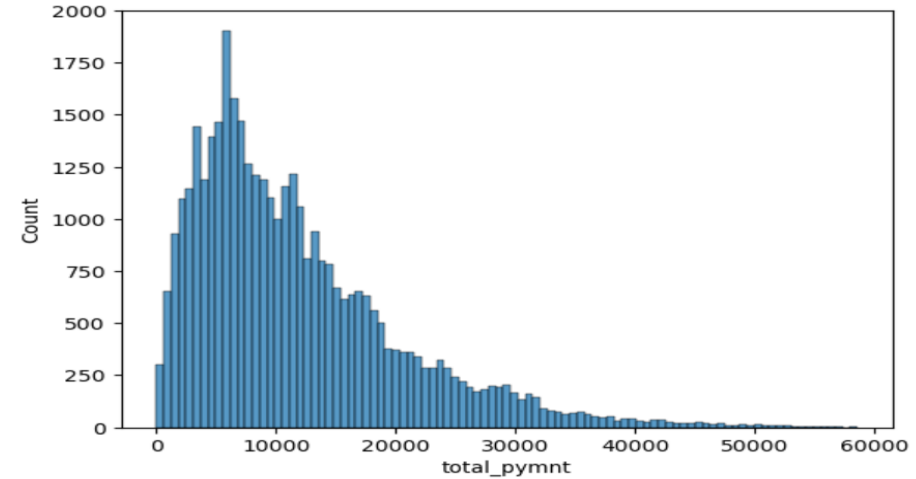
Histplot of revol_util



Histplot of total_acc



Histplot of total_pymnt



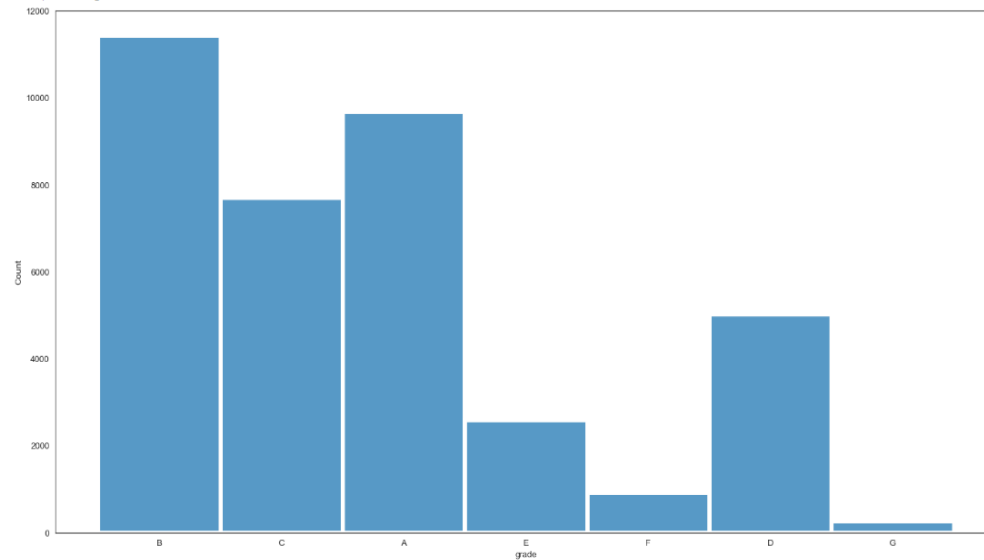
Insights on univariate analysis for numerical data

- Loan amount is varying from 2000 to 14000 but the highest loan amount borrowed is between 4000 to 6000 and 8000 to 12000
- Highest Amount funded by investors at the time ranges between 4000 to 12000
- Maximum terms for the loan is 35 to 36 months
- Maximum Interest rate varying between 6 to 16.5 in that highest rate of interest ranges between 6 to 7.5 and 10.5 to 11.5
- Maximum number of instalments for the loan is less than 400
- Maximum annual income of the borrowers is 30000 to 60000
- Total payment wrt loan amount is less than 10000

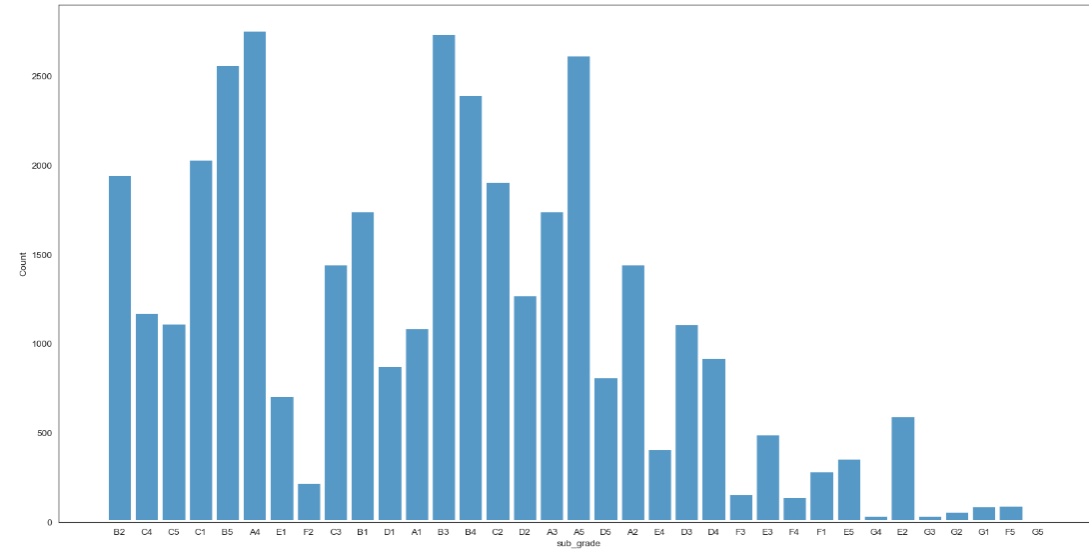
2. Categorical

- grade, subgrade, emp_length

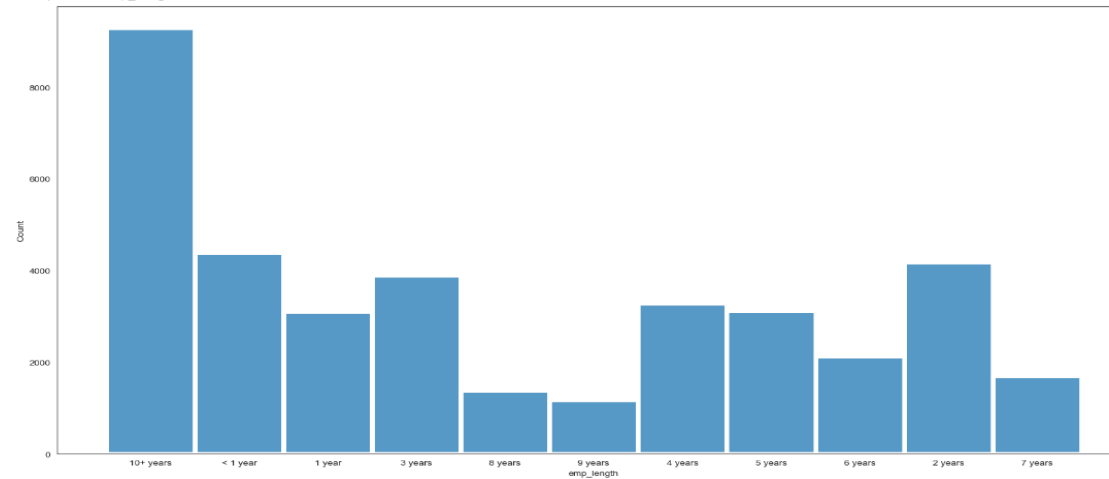
Histplot of grade



Histplot of sub_grade

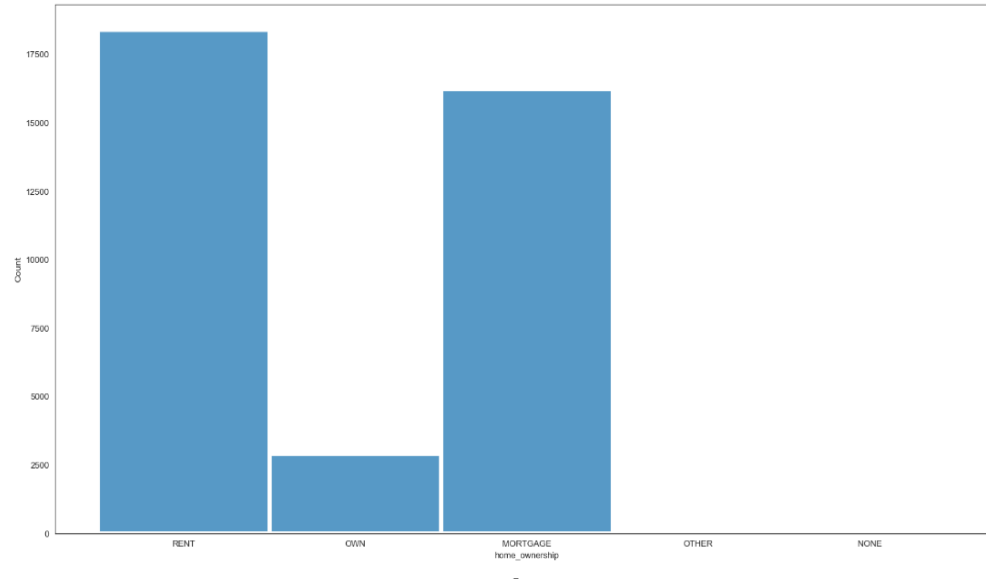


Histplot of emp_length

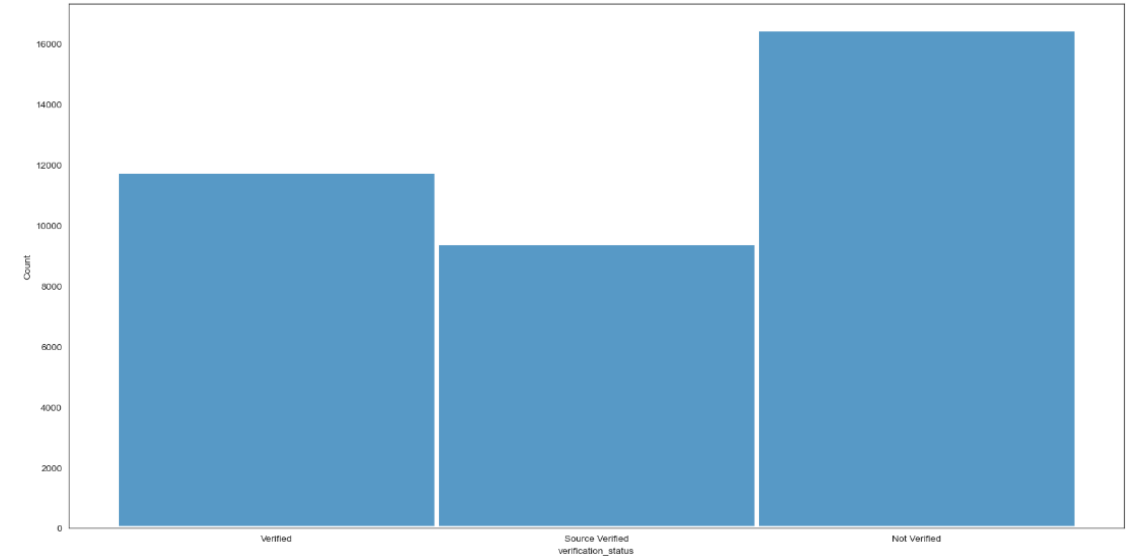


Home_ownership, verification_status, loan_status and purpose

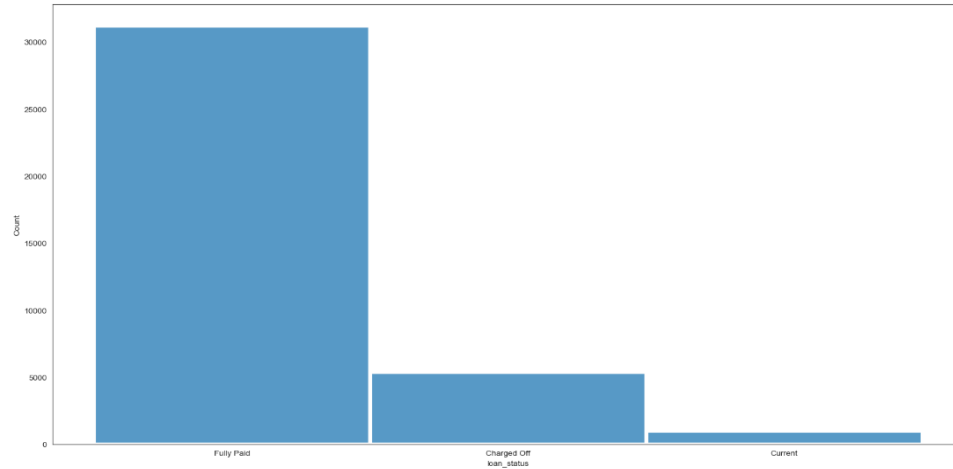
Histplot of home_ownership



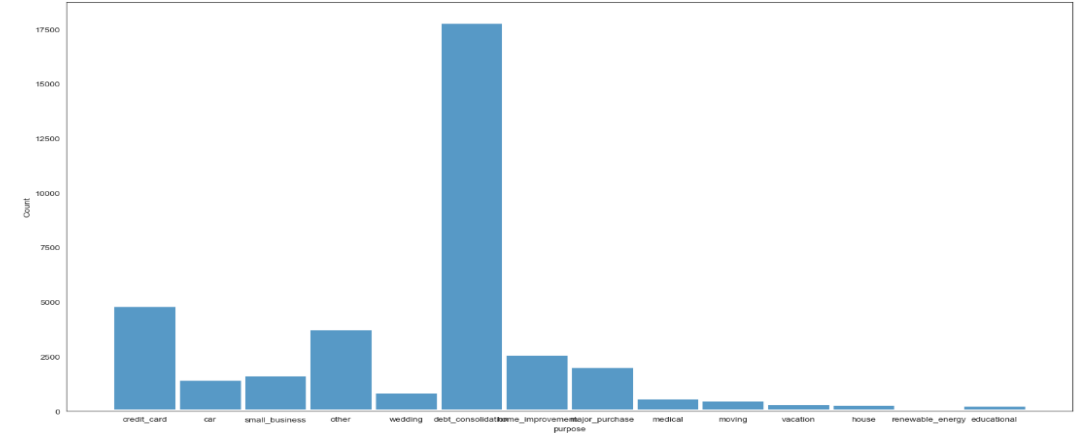
Histplot of verification_status



Histplot of loan_status



Histplot of purpose



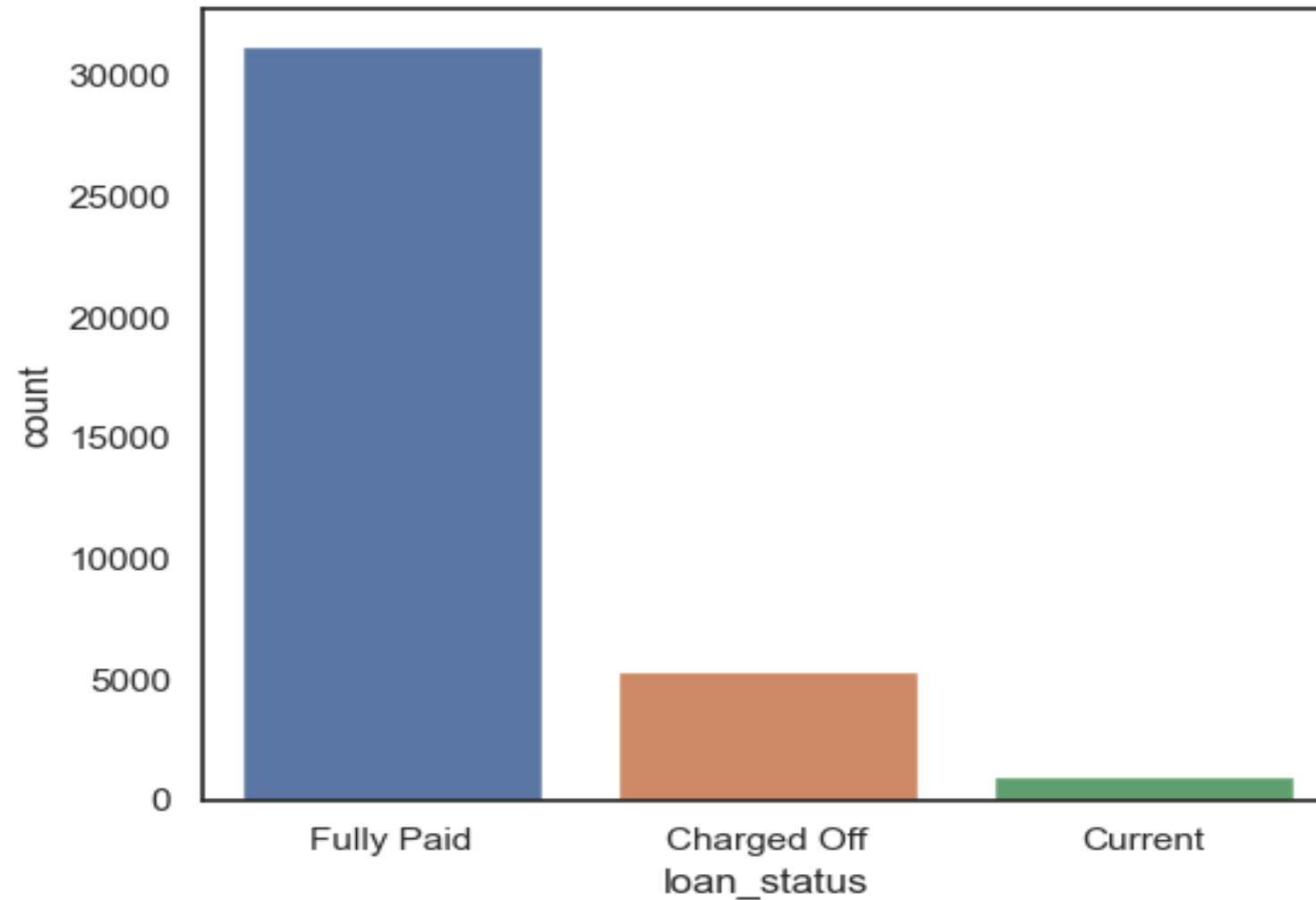
Insights on categorical data

- B Grade borrower's are more than other grades
- A4 subgrade borrower's are more compare to other
- 10+ years is the maximum employment length of borrowers
- Maximum borrower's home ownership is RENT
- Verification status for most of the loan given is not verified
- Most of the borrower's has takenn loan for the purpose of debt_consolidation

Visualizing categorical data with respect to defaulter

Loan Status

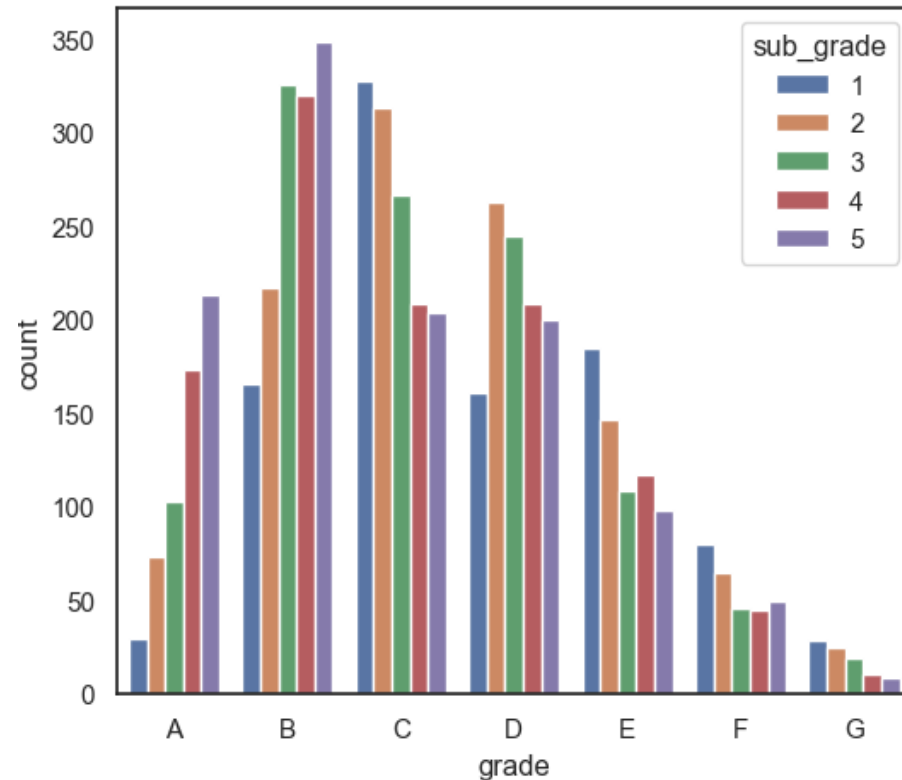
<Axes: xlabel='loan_status', ylabel='count'>



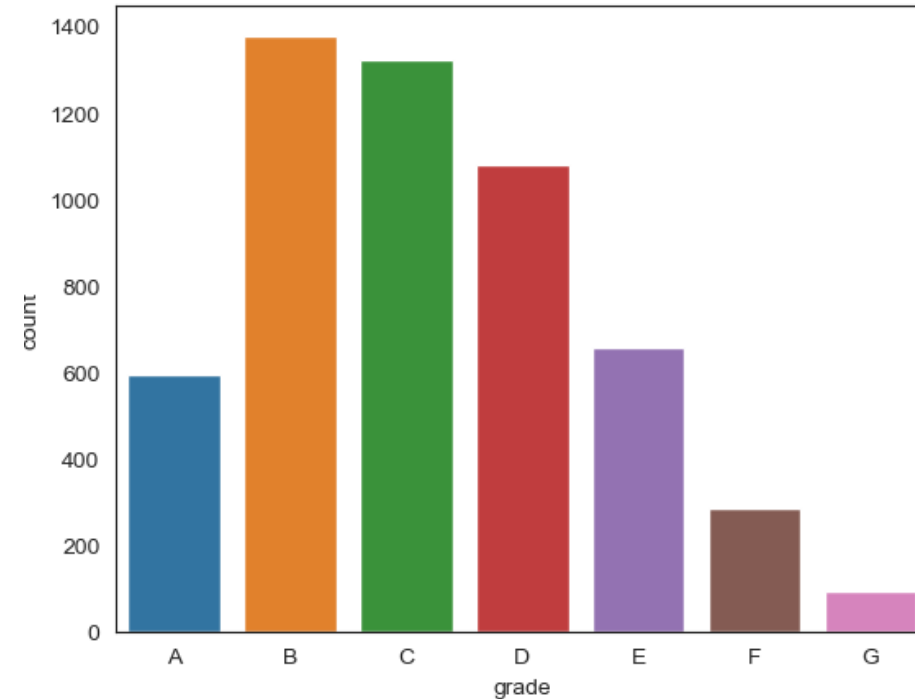
Categorical vs Charged Off

- Grade and subgrade with respect to Charged Off

(Axes: xlabel='grade', ylabel='count')>

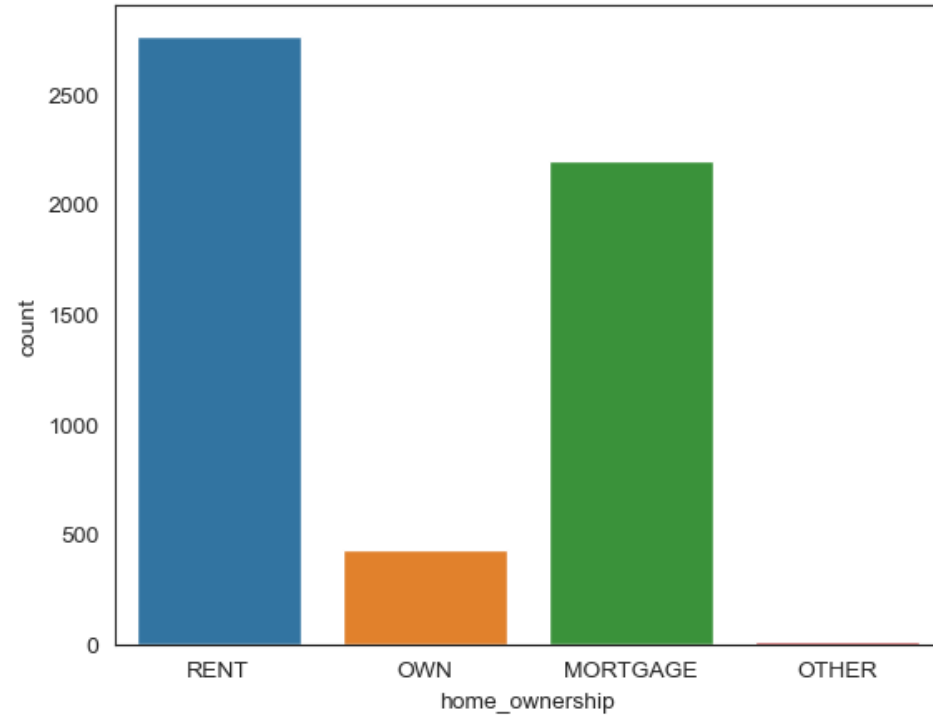


(Axes: xlabel='grade', ylabel='count')>

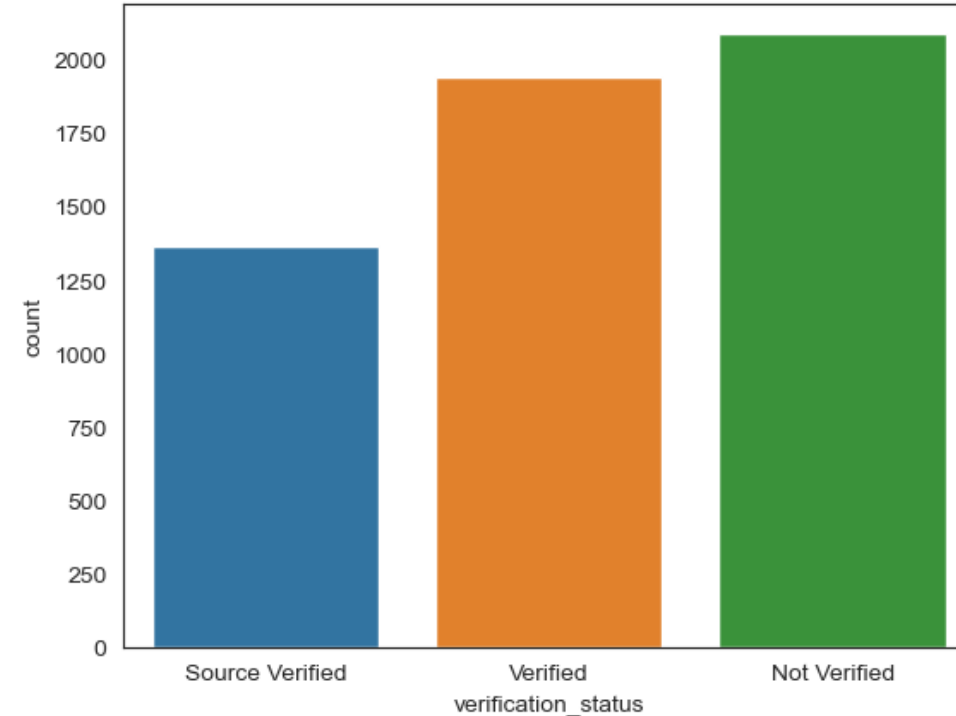


Home ownership, verification_status vs charged off

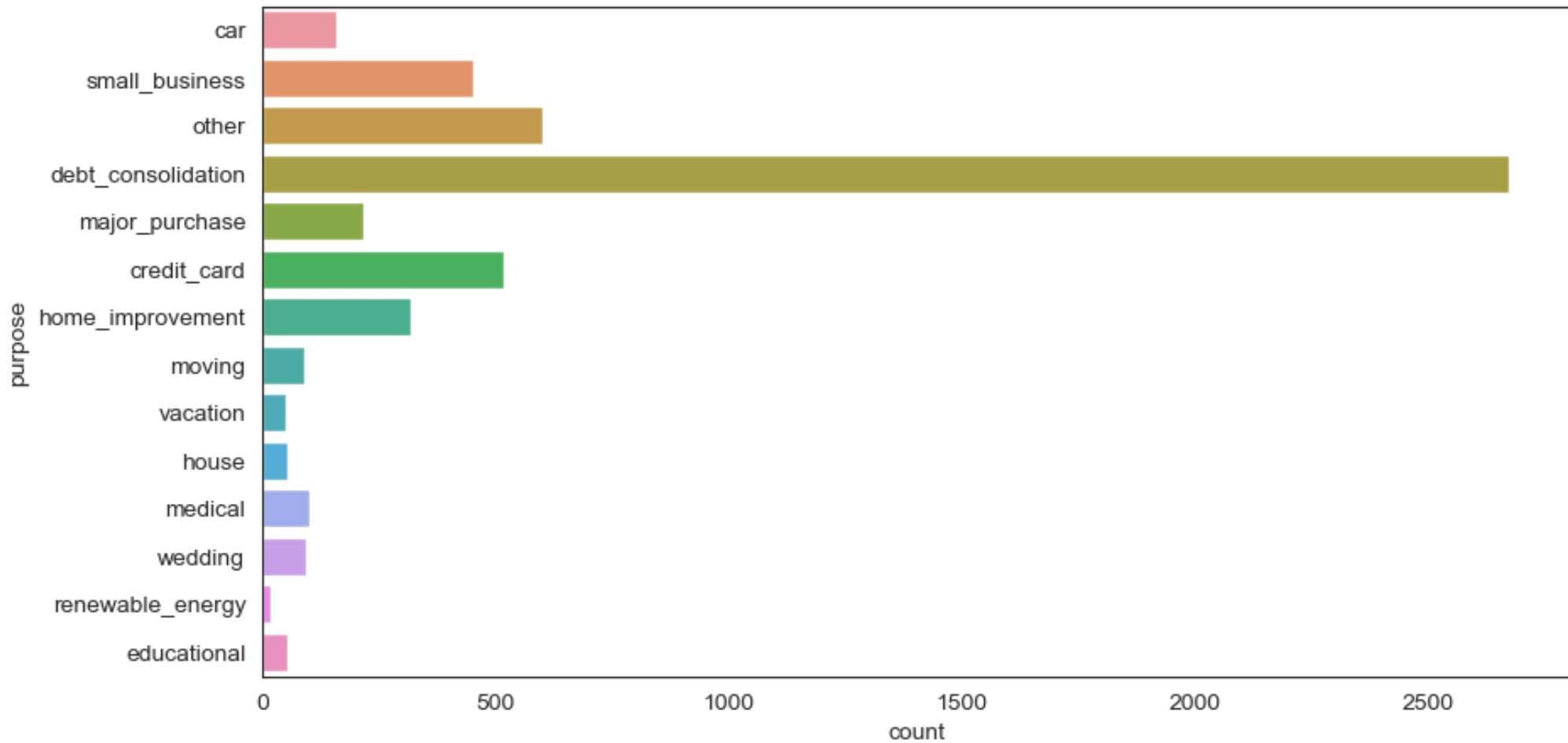
<Axes: xlabel='home_ownership', ylabel='count'>



<Axes: xlabel='verification_status', ylabel='count'>



Purpose vs Charged Off



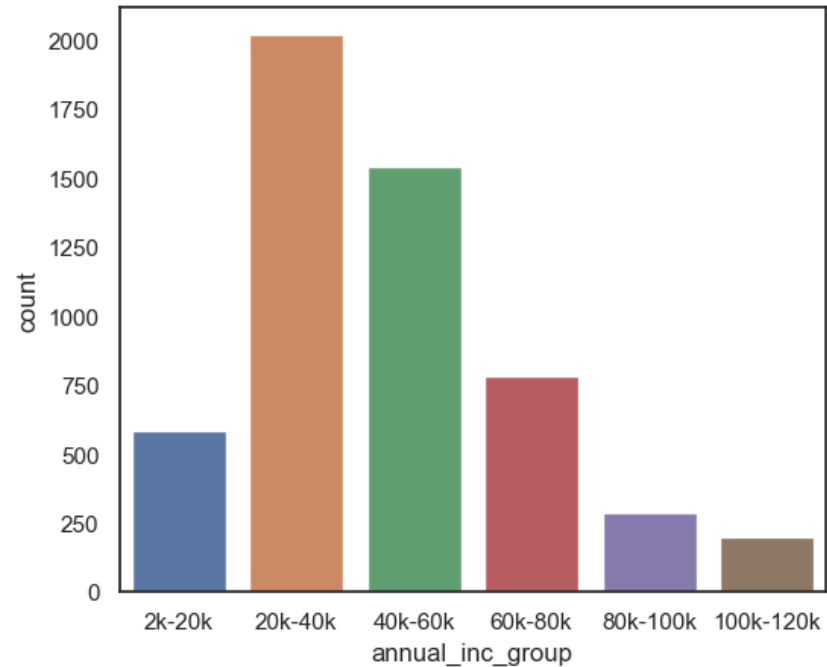
Insights on Categorical Data vs Charged Off

- People under Grade B and subgrade B5 are the most likely to default
- Those who are leaving on Rent are mostly likely to default
- Most people do not have their income sources verified
- debt_consolidation is the foremost purpose for defaulting the loan

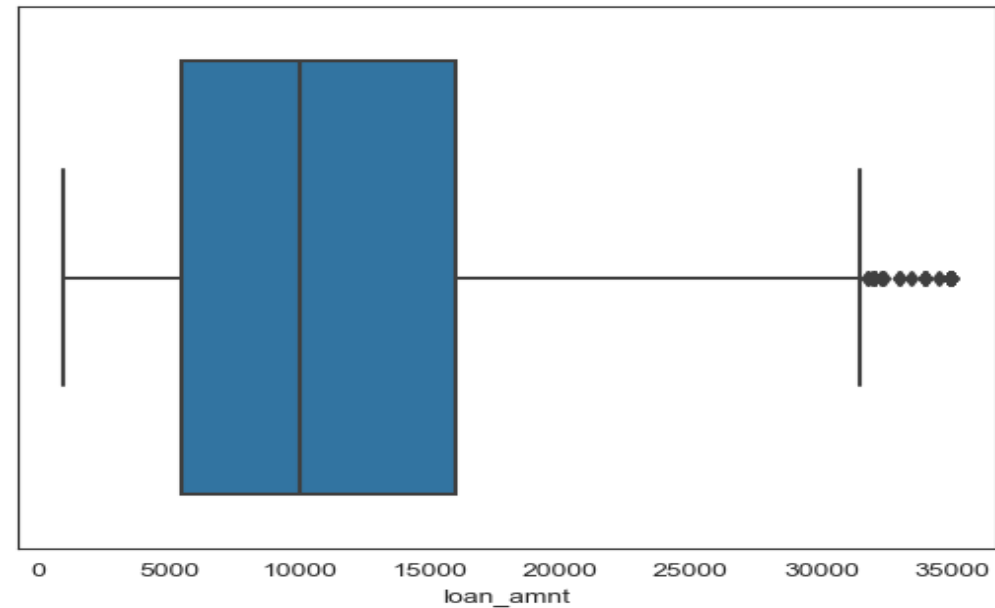
Numerical data with respect to charged off

- Annual_inc, loan amount vs Charged Off

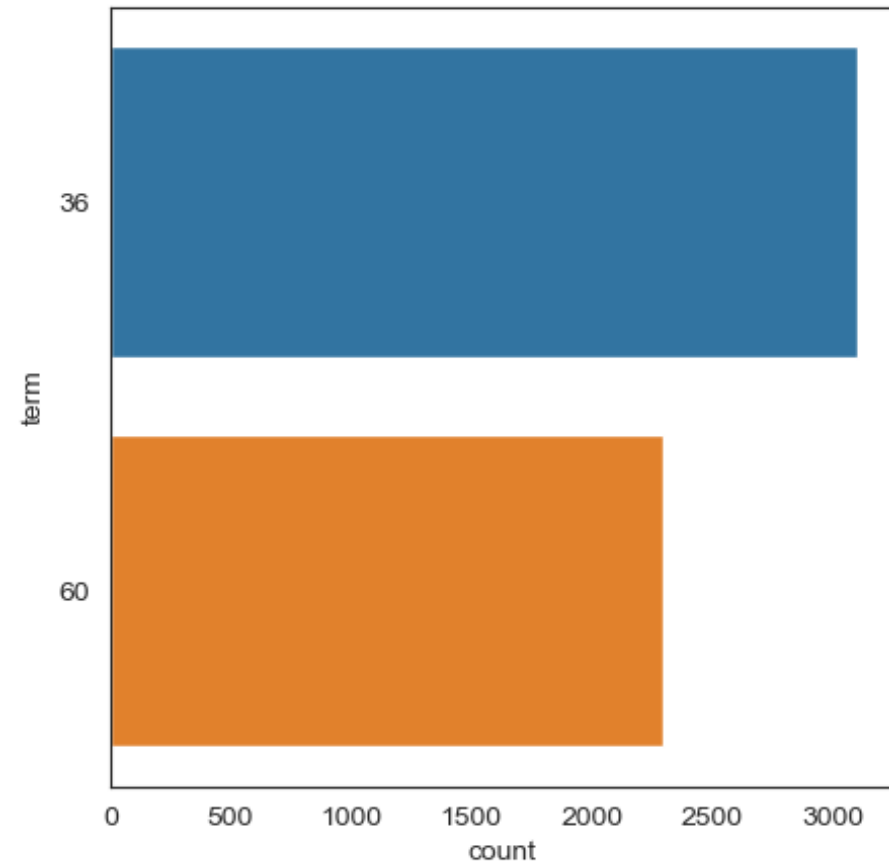
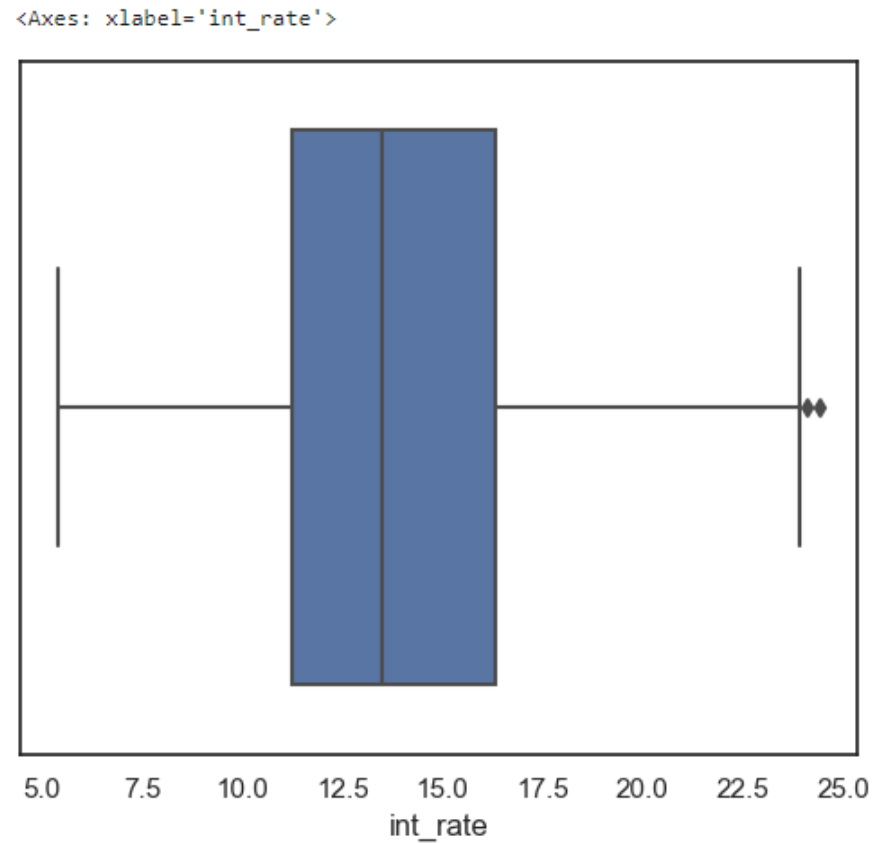
<Axes: xlabel='annual_inc_group', ylabel='count'>



<Axes: xlabel='loan_amnt'>

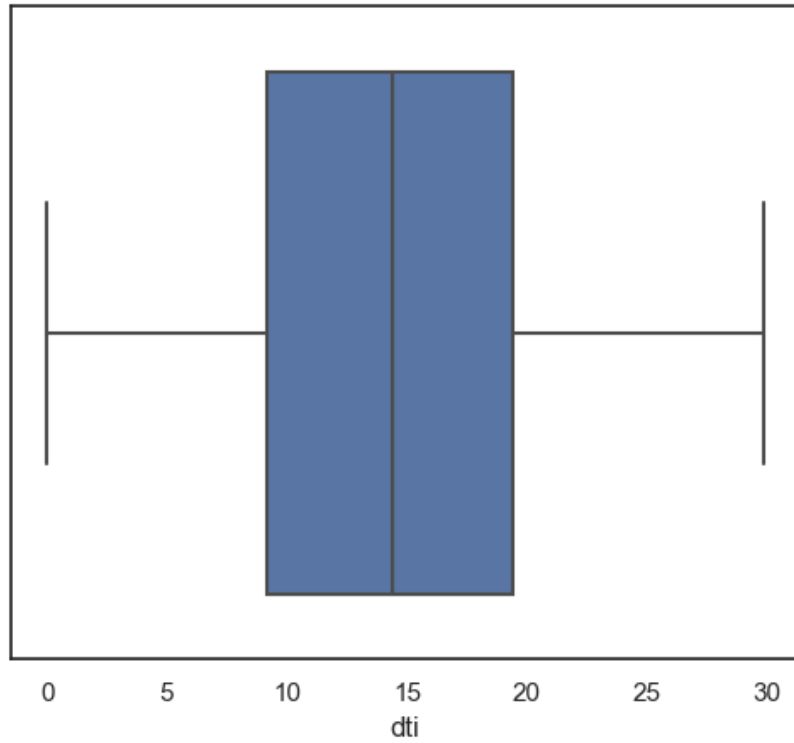


Interest rate, term vs Charged Off

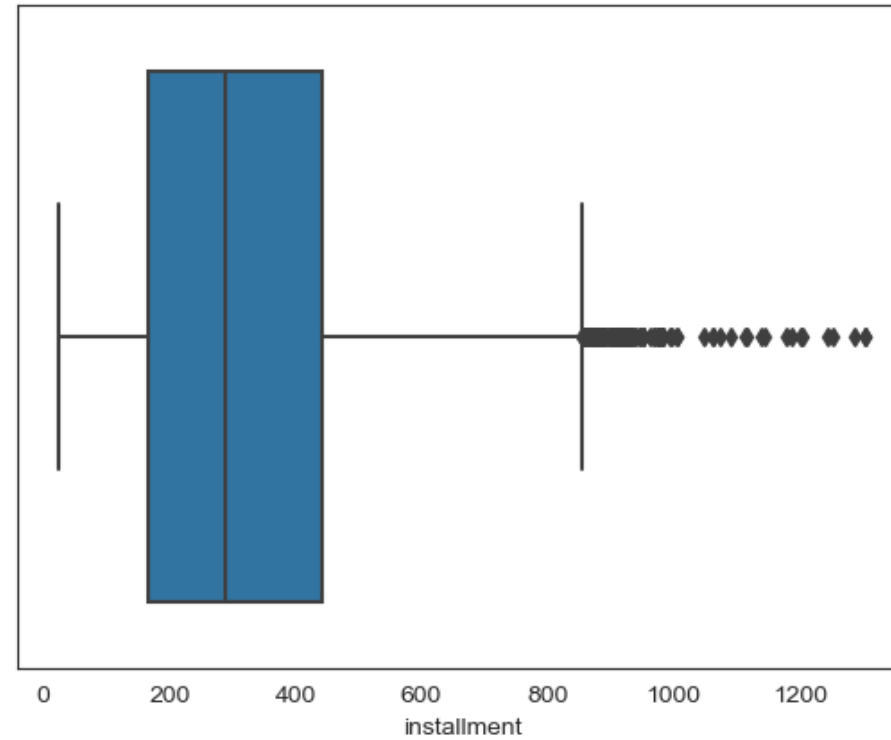


Dti, installments vs Charged Off

<Axes: xlabel='dti'>

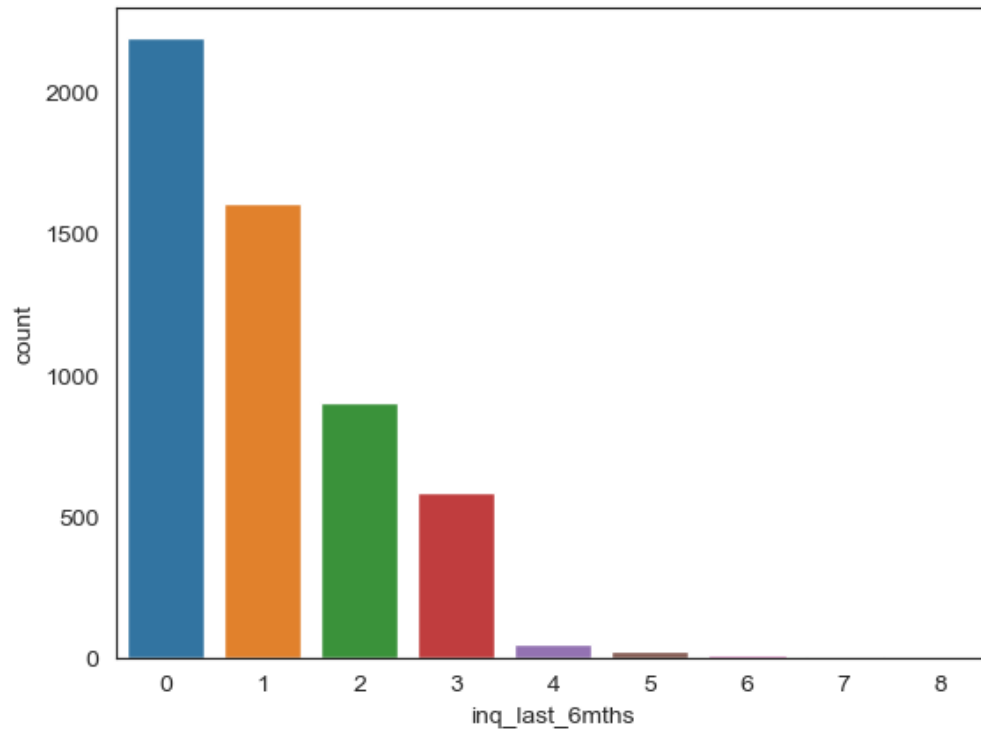


<Axes: xlabel='installment'>

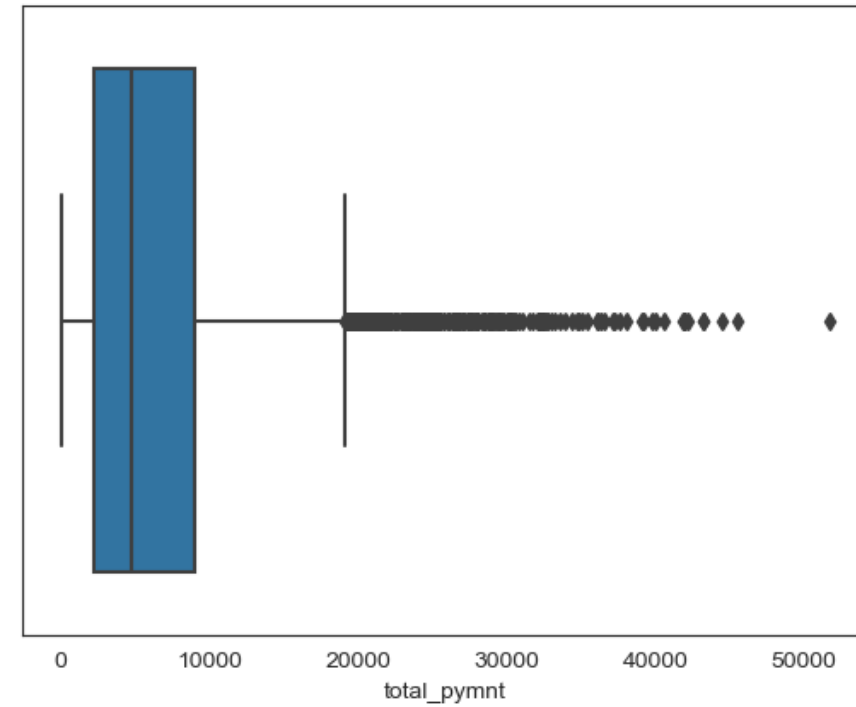


Inq_last_6months, total_pymnt vs Charged Off

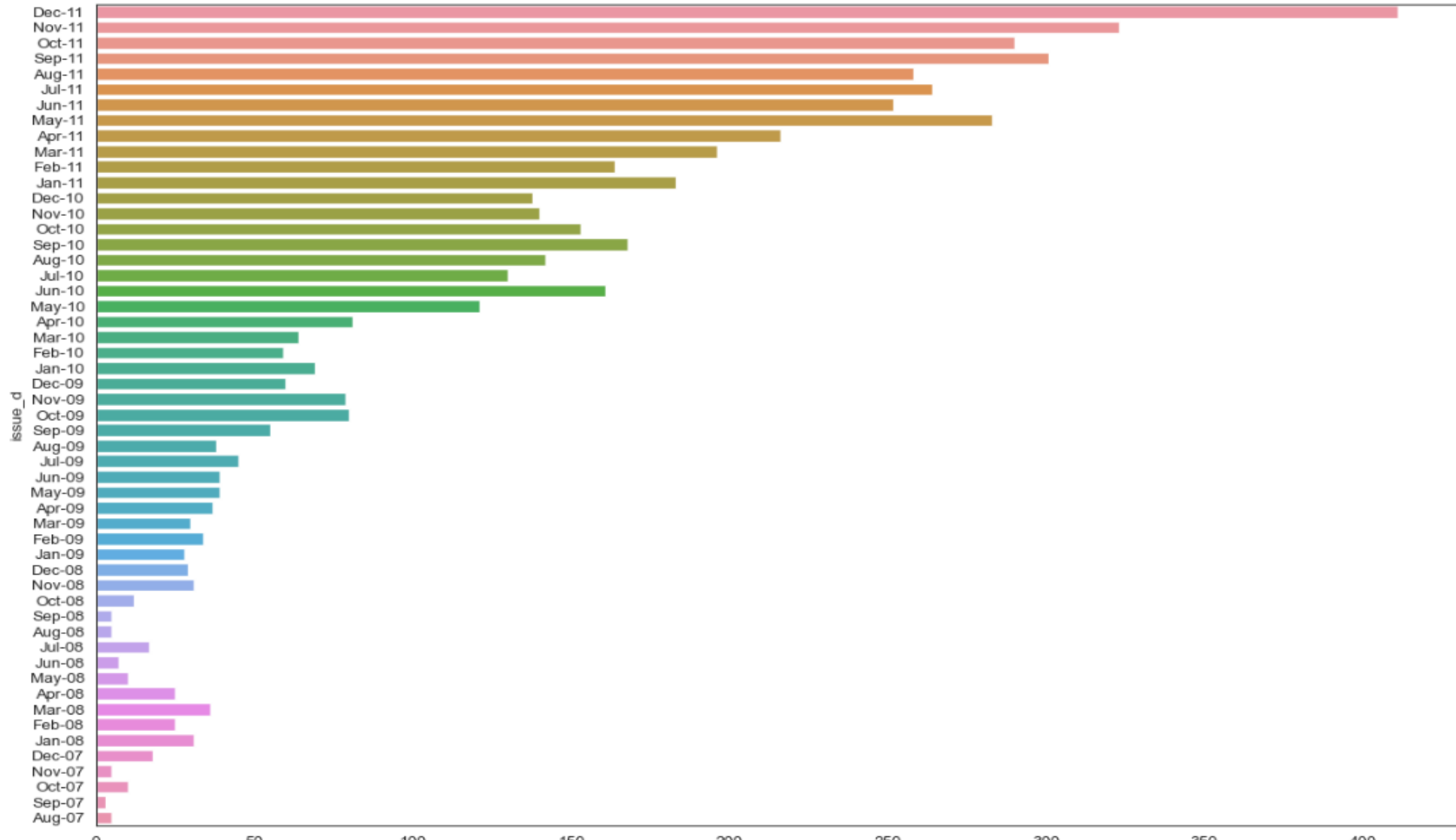
<Axes: xlabel='inq_last_6mths', ylabel='count'>



<Axes: xlabel='total_pymnt'>



Issued date vs Charged Off

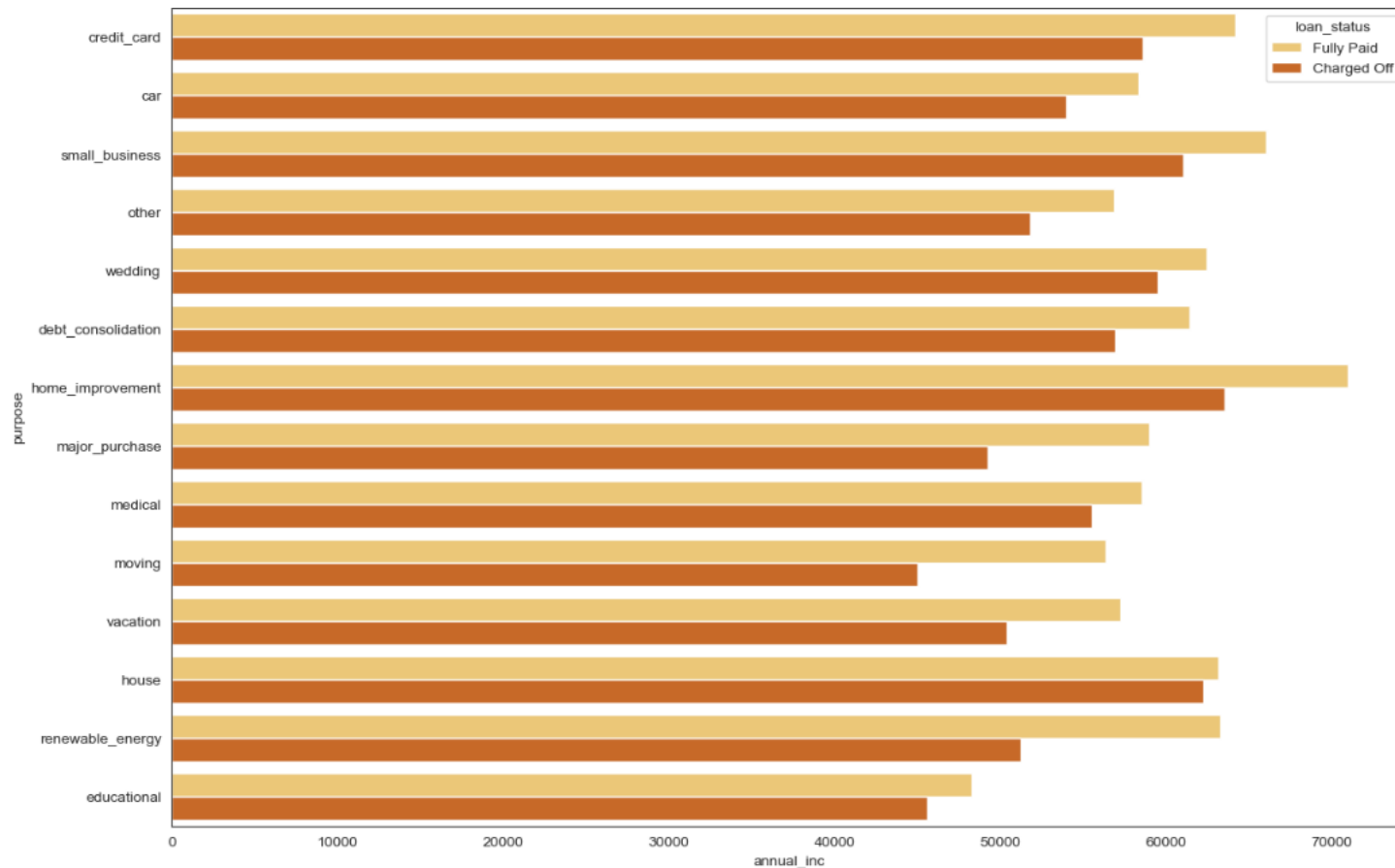


Insights of Numerical Vs Charged Off

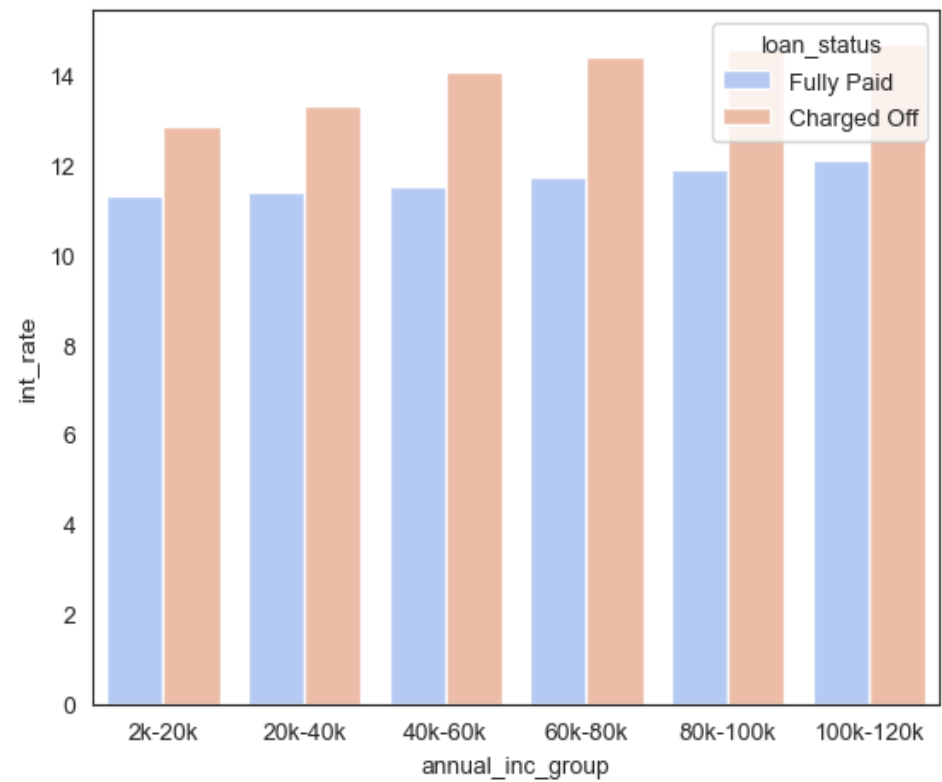
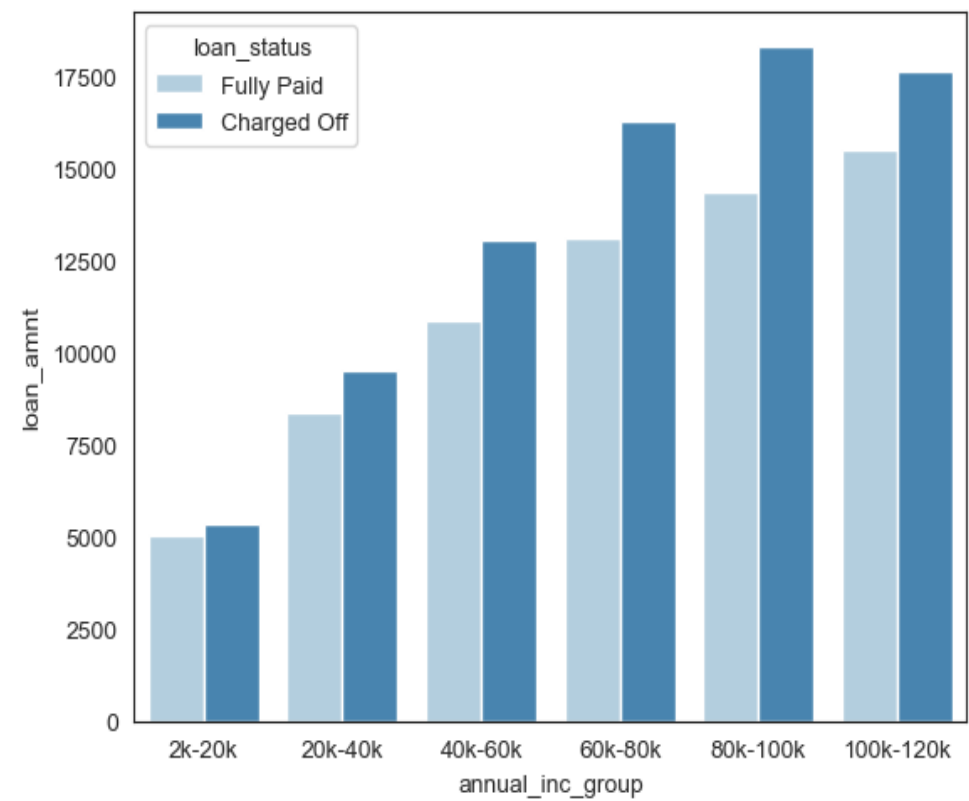
- Most people have annual income between 20k to 40k
- 5k to 15k is the loan amount given to the majority people who are defaulting
- interest rate is 11 to 17 for those under charged off category
- Maximum people got the term of 36 months
- dti ratio lies between 10 to 20 in this scenario
- instalment amount is rupees 200 to 400 for the defaulted ones
- The trend of inquiry last 6 months is 0 for most and is constantly dropping
- 0 to 10k is the range of total payments that defaulted borrowers are having
- Maximum number of defaults occurred when the loan was issued in Dec in the year 2011

Analysis on annual_inc with other columns to get more insights

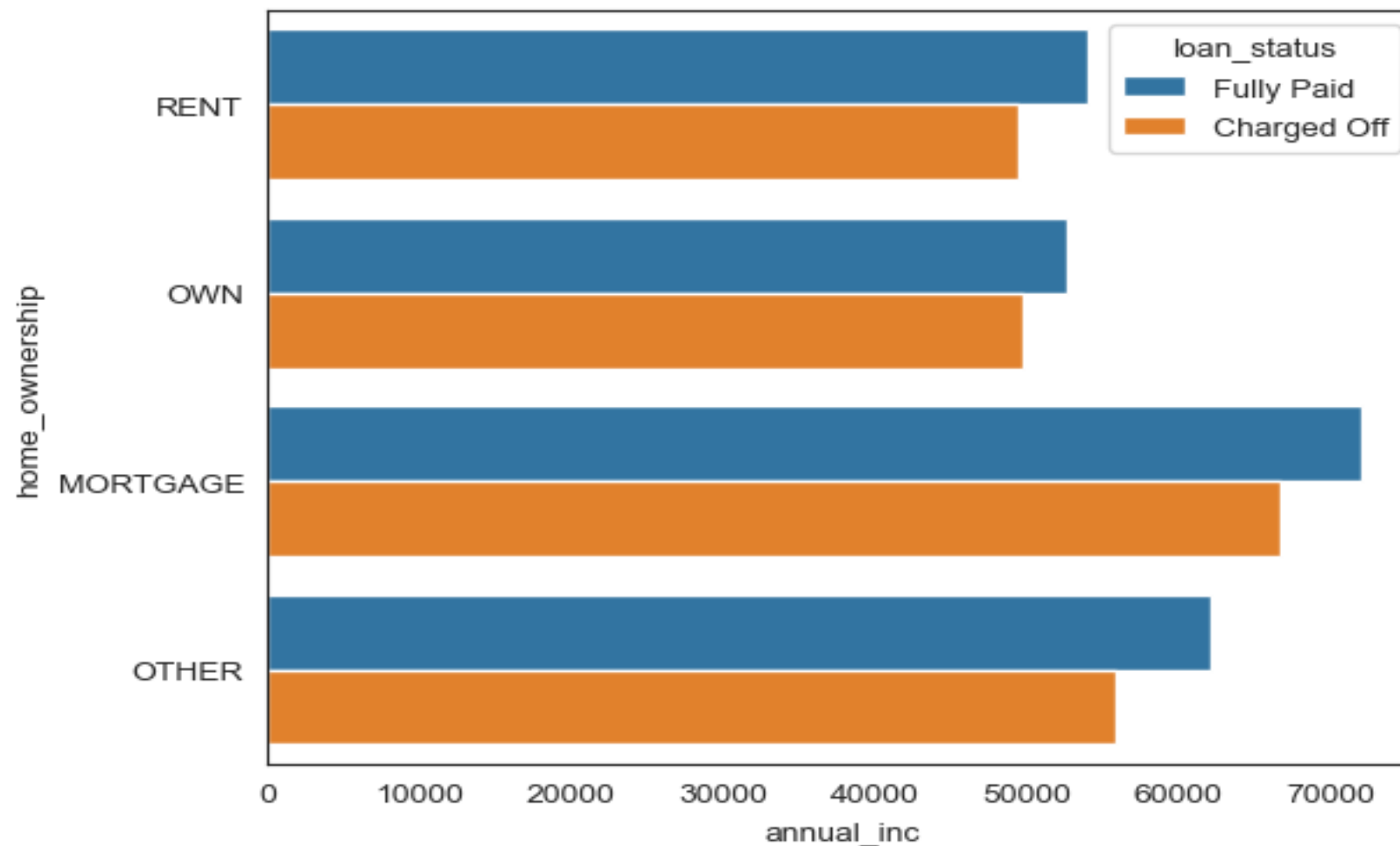
- Annual income vs purpose



Annual income vs Loan Amount and Annual income vs interest Rate



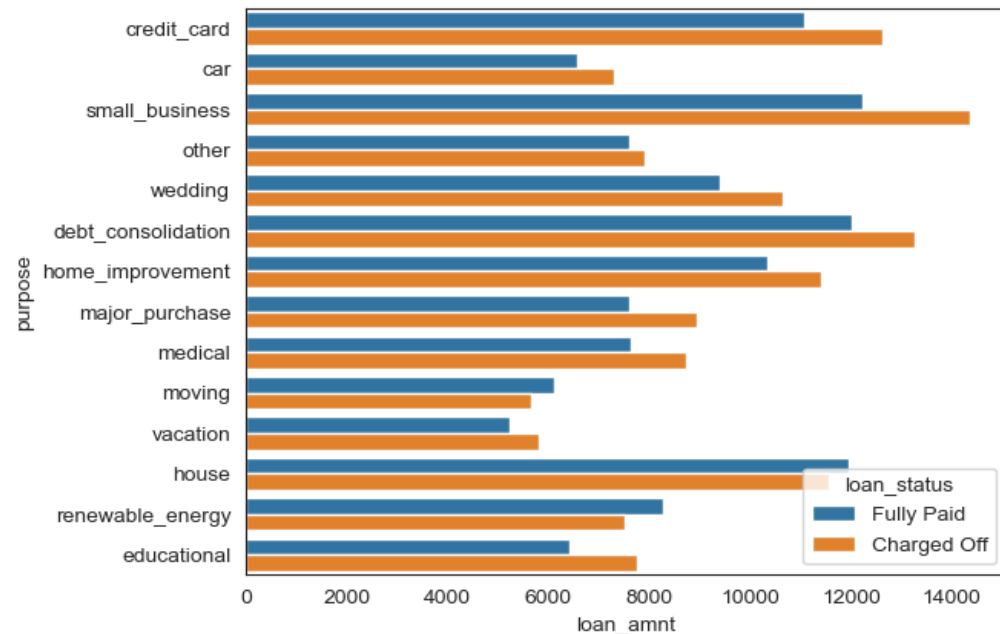
Annual income vs Home ownership



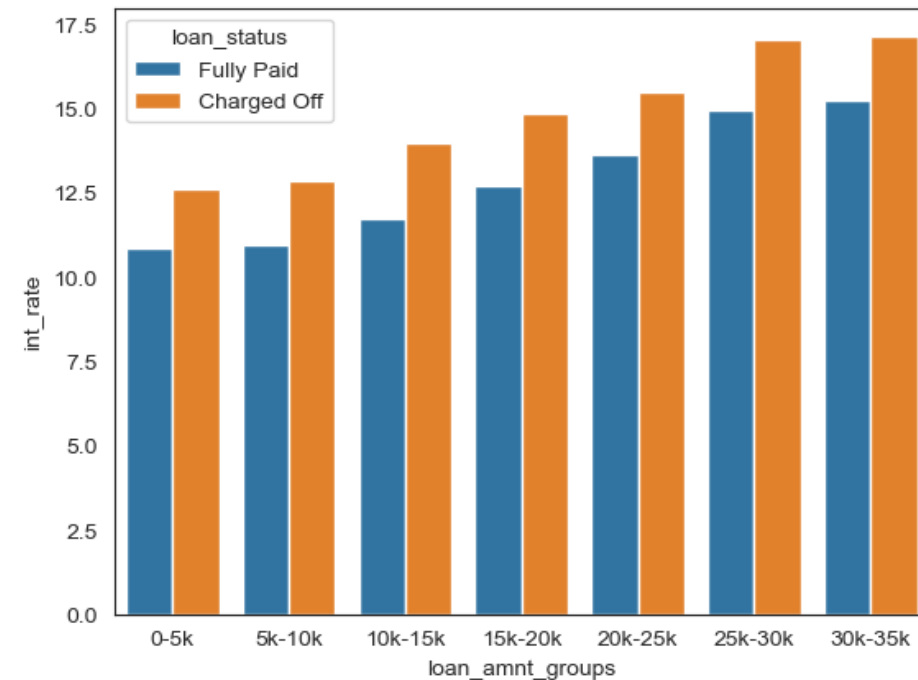
Analysis on loan amount with other columns to get more insights

- Loan amount vs purpose and Loan amount vs interest rate

<Axes: xlabel='loan_amnt', ylabel='purpose'>

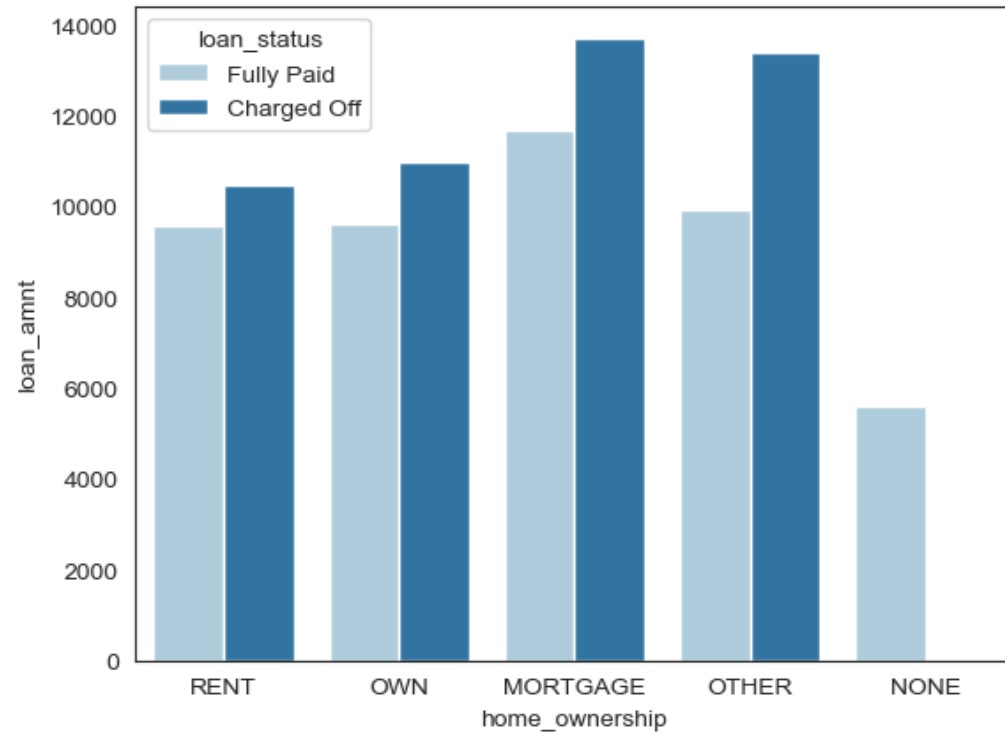


<Axes: xlabel='loan_amnt_groups', ylabel='int_rate'>

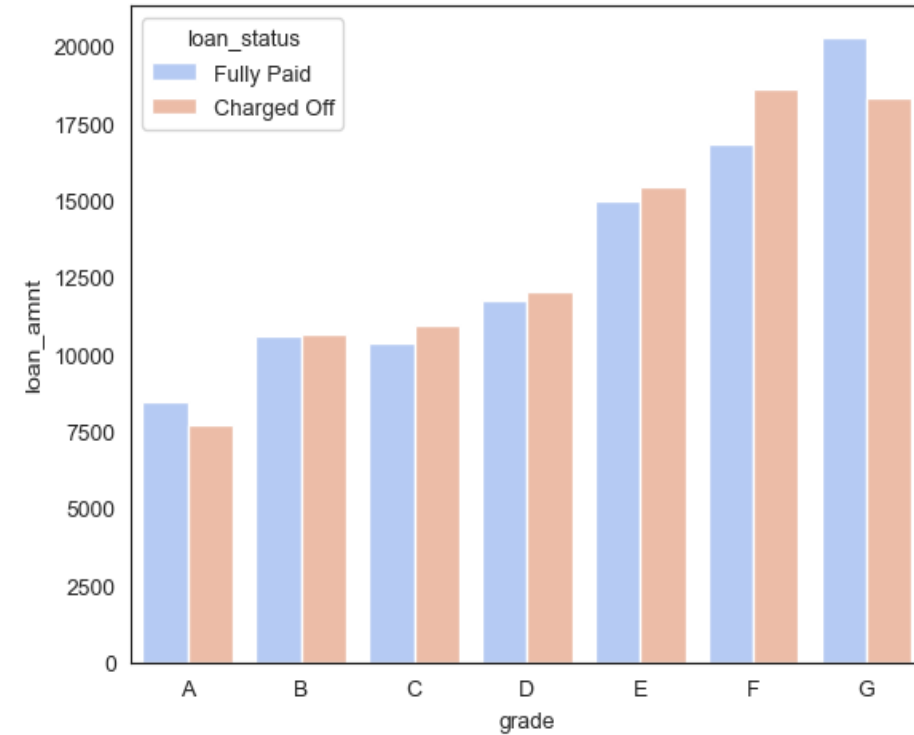


Loan amount vs home ownership and Loan amount vs grade

<Axes: xlabel='home_ownership', ylabel='loan_amnt'>

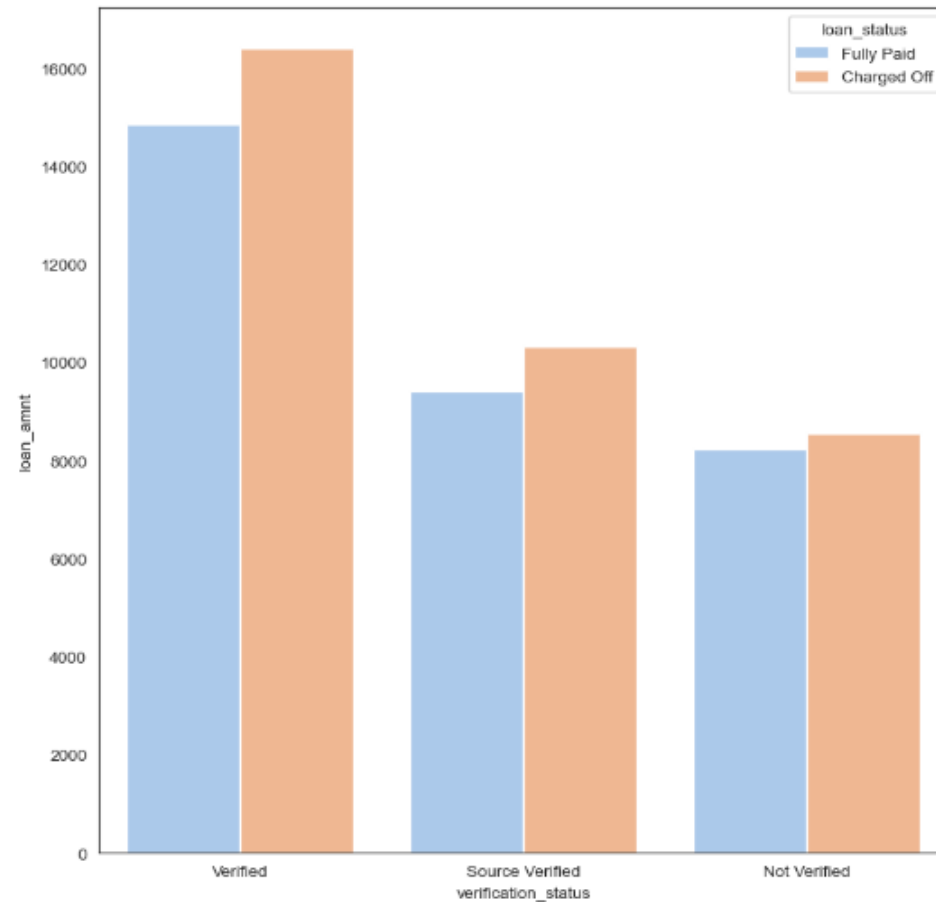
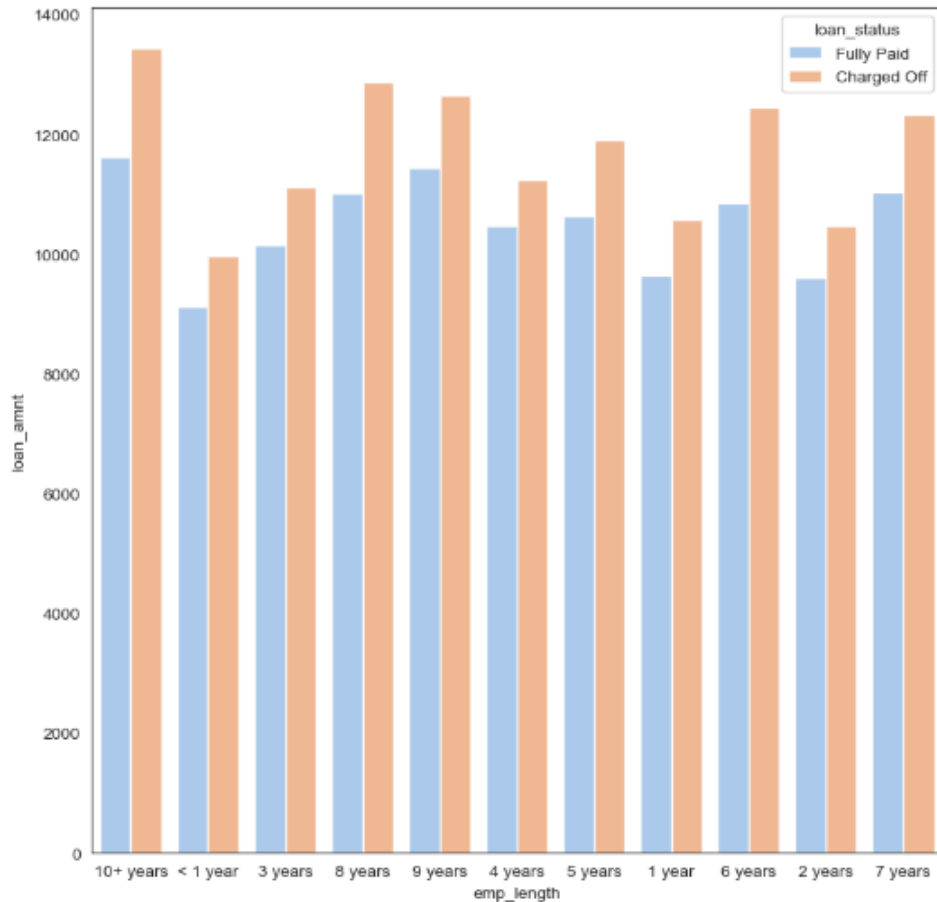


<Axes: xlabel='grade', ylabel='loan_amnt'>



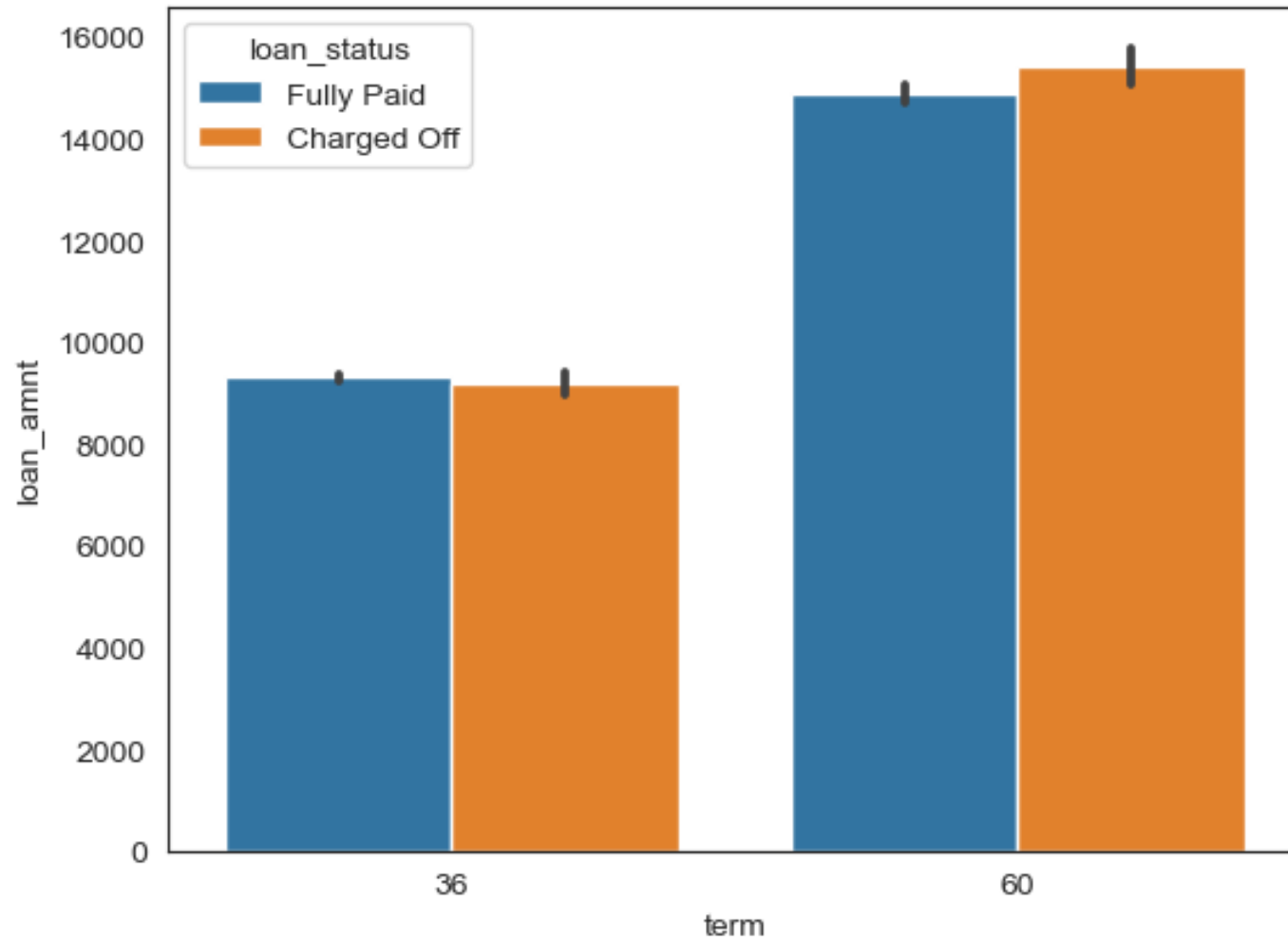
Loan amount vs employment length and verification status

<Axes: xlabel='verification_status', ylabel='loan_amnt'>



Loan amount vs term

<Axes: xlabel='term', ylabel='loan_amnt'>



Observations

- People who has taken loan for home improvement and have income of 60k to 70k
- People whose home ownership is 'MORTGAGE' and have income 60k to 70k
- People whose interest rate is greater than 12% and have income of 60k to 120k
- Who defaulted the most has taken loan amount of 12k to 14k for small business purpose
- loan amount of 25k to 30k with greater than 15% of interest rate
- People whose home ownership is 'MORTGAGE' has loan amount of 10k to 12k
- When grade is 'F' with loan amount of 15k to 17.5k
- Whose work experience is '10+ years' and have loan amount of 10k to 12k
- Verified only for higher loan amount with 14k to 16k

Thank you

