

LENDING CLUB CASE STUDY

Submitted by
Vaishnavi Ramkumar
Cohort:C50, Feb 2023

Lending Club: EDA Case Study

- Problem Statement

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss) for a lending club. The company wants to understand the driving factors (or driver variables) behind loan default

- Business Objective

When the 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

- Outcome

Provide insights on strong driving factors for loan default and the company will use this for its portfolio and risk assessment.

Approach

There are four major parts that are needed to be done for this case study:

- 1. Data understanding
- 2. Data cleaning (cleaning missing values, removing redundant columns etc.)
- 3. Data Analysis
- 4. Recommendations

Data Understanding

Lending club is a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

If the company approves the loan, there are 3 possible scenarios

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.

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

The Area of interest here is the Charged-off applicants. We will be analyzing the patterns for this category in detail to understand the strong driving factors for default

Data cleansing

- Understand the data
- Calculate column's missing percentage
- Remove columns with high NA percentage
- Remove Columns with only one unique values which do not have really help in analysis
- Removing columns which have unique variables like id, member_id, url etc.
- Remove the post loan attributes, these are basically generated post loan variables like collection_recovery_fee, recoveries, total_pymnt, total_pymnt_inv, total_rec_prncp etc.

Variables chosen for Analysis

The below listed data columns for analysis are chosen

loan_amnt - The listed amount of the loan applied for by the borrower.

funded_amnt - The total amount committed to that loan at that point in time.

funded_amnt_inv - The total amount committed by investors for that loan at that point in time.

int_rate - Interest Rate on the loan

installment - The monthly payment owed by the borrower if the loan originates.

annual_inc - The self-reported annual income provided by the borrower during registration.

dti - A ratio calculated using the borrower's total monthly debt payments on the total debt obligations

total_acc - The total number of credit lines currently in the borrower's credit file

term - The number of payments on the loan. Values are in months and can be either 36 or 60.

grade - LC assigned loan grade

sub_grade - LC assigned loan subgrade

emp_length - Employment length in years.

home_ownership - The home ownership status provided by the borrower during registration

verification_status - Indicates if income was verified by LC, not verified, or if the income source was verified

purpose - A category provided by the borrower for the loan request.

addr_state - The state provided by the borrower in the loan application

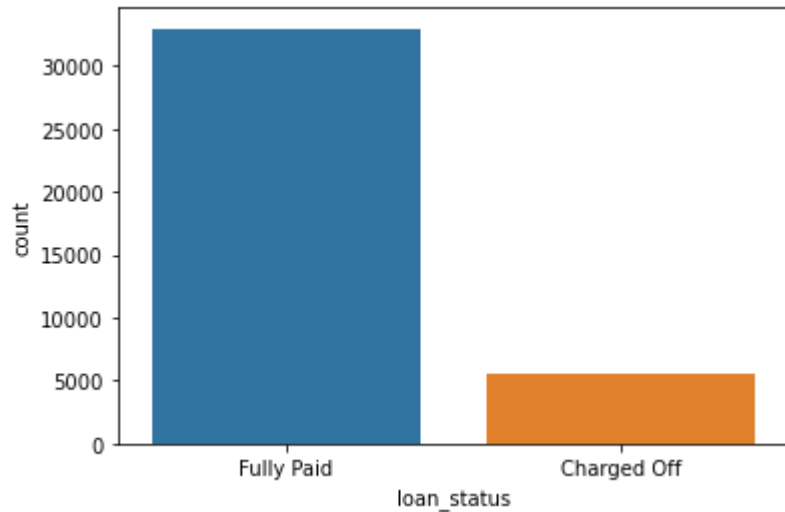
Issue_Month - The month the loan was issued

Issue_Year - The year in which loan was issued

These columns are compared against the loan_status value as charged-off.

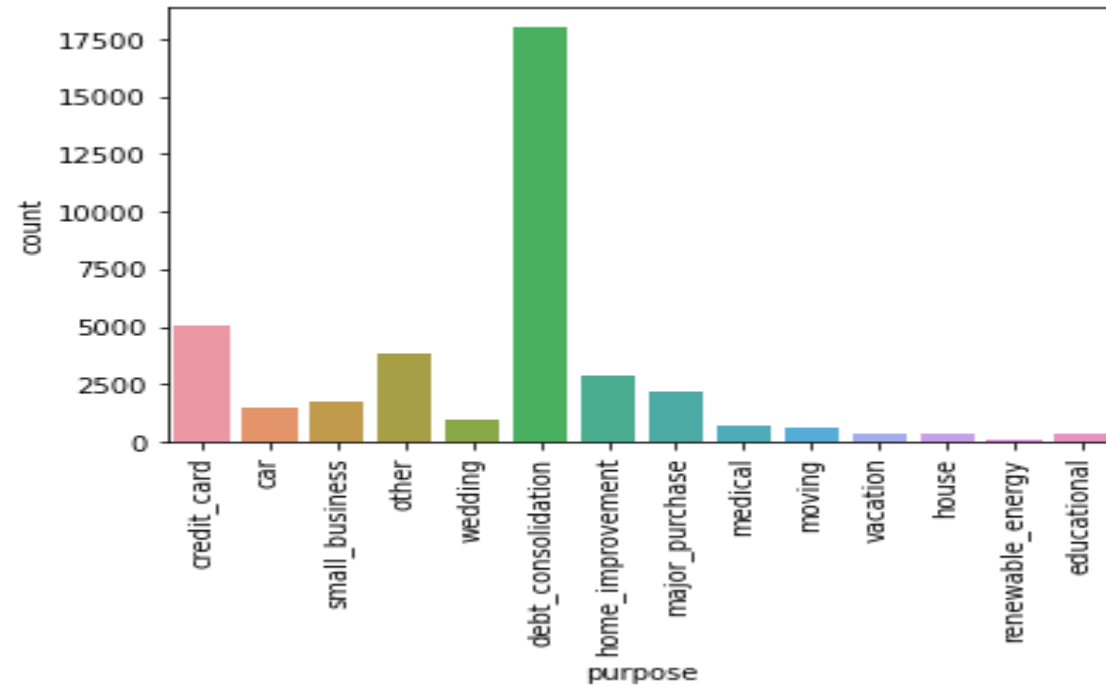
Univariate Analysis

Loan status

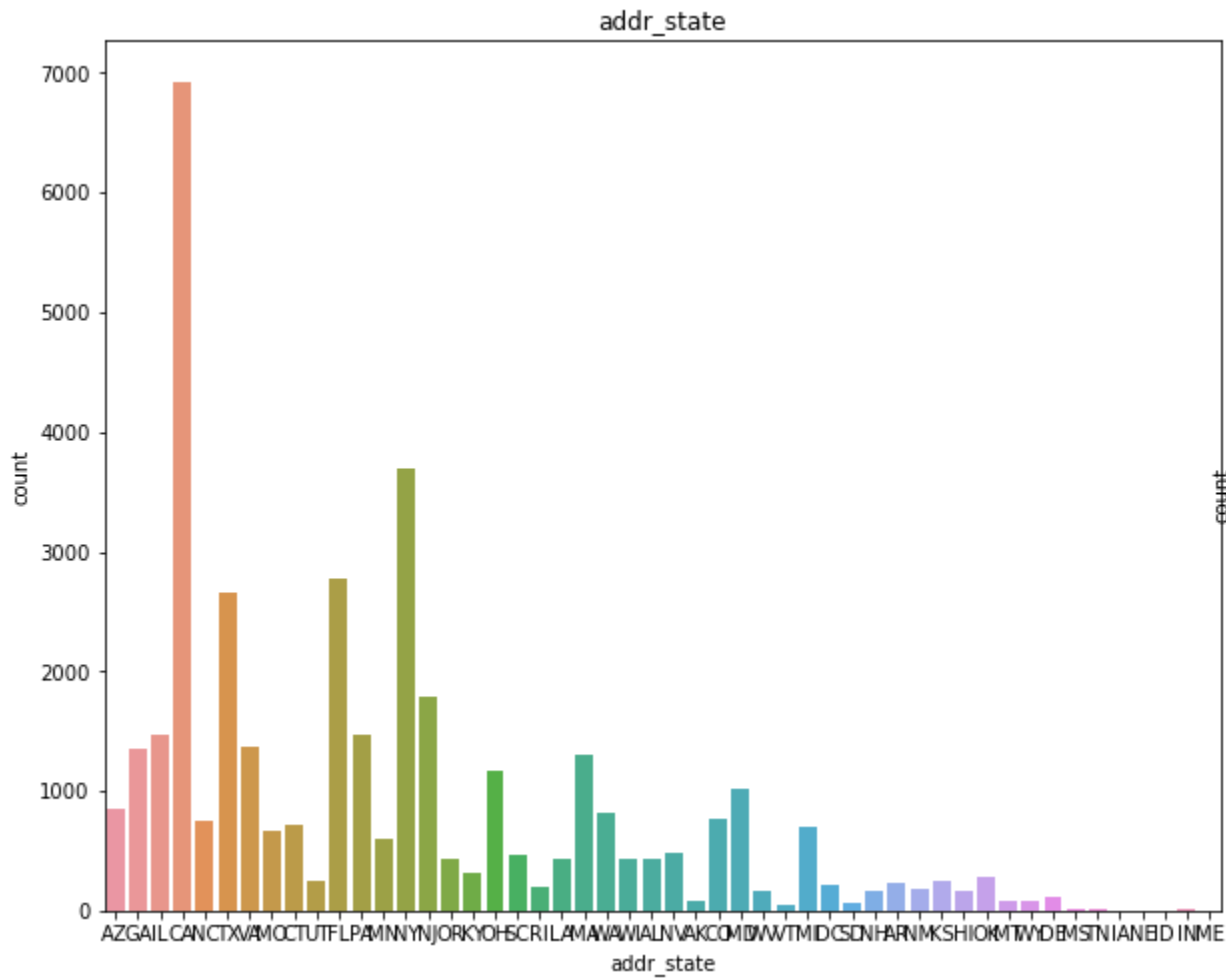


The number of fully paid customers is very high than the charged off customers

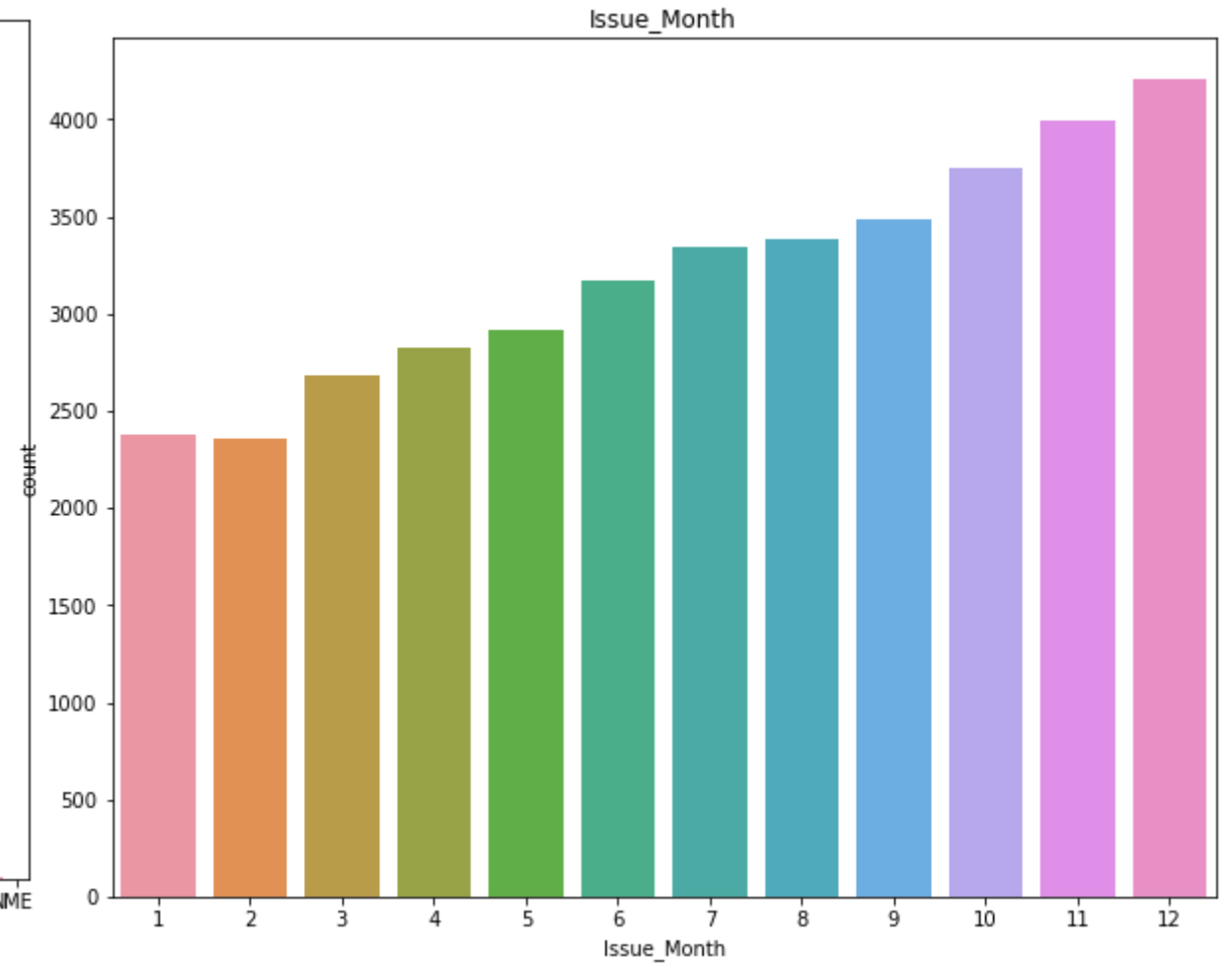
Purpose of loan



The most common purpose of taking a loan is for debt consolidation

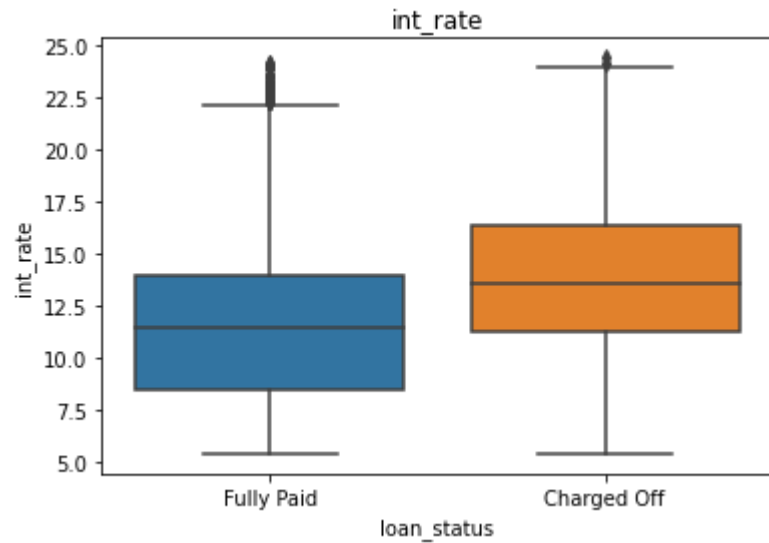


The highest number of loan takers are from the state of CA

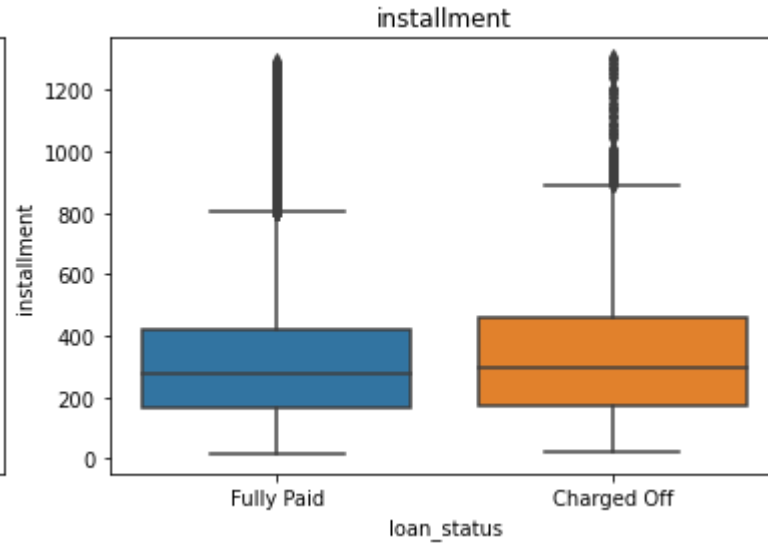


The number loans issued is high in month of December. Vacation planning, festive season could be a reason for high loan applications

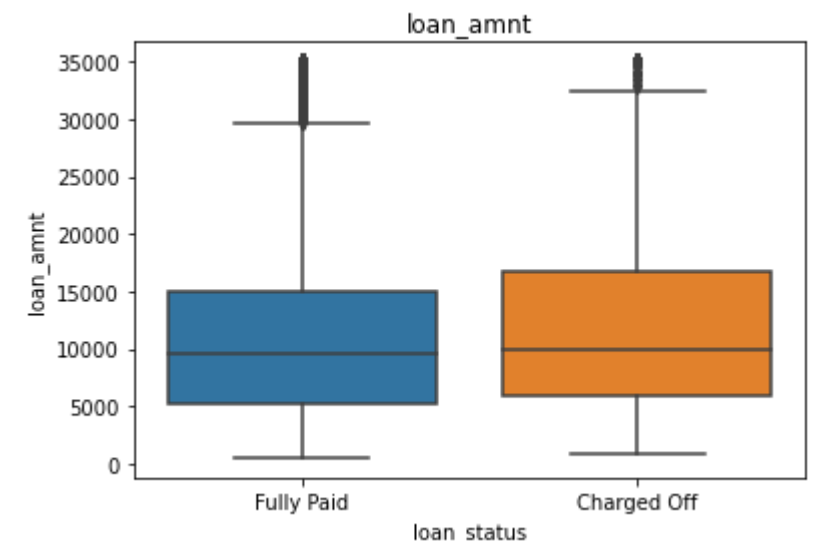
Bivariate analysis



As interest rate is increases, charged off count increases

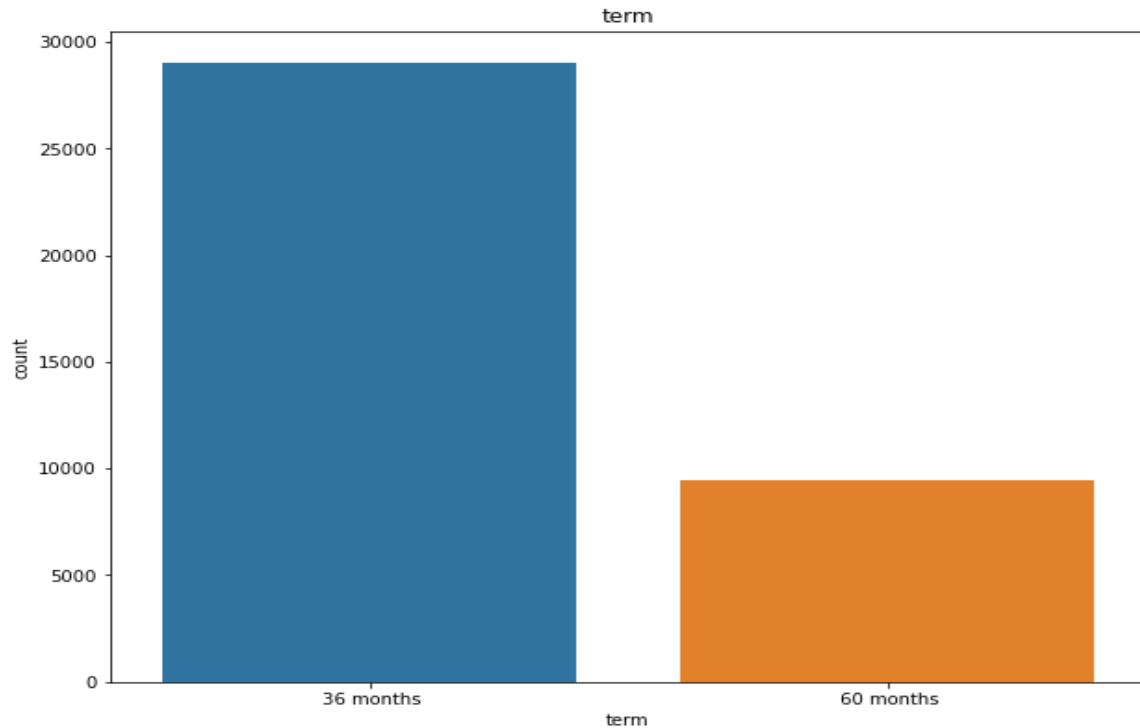


As installment is increases, charged off count increases

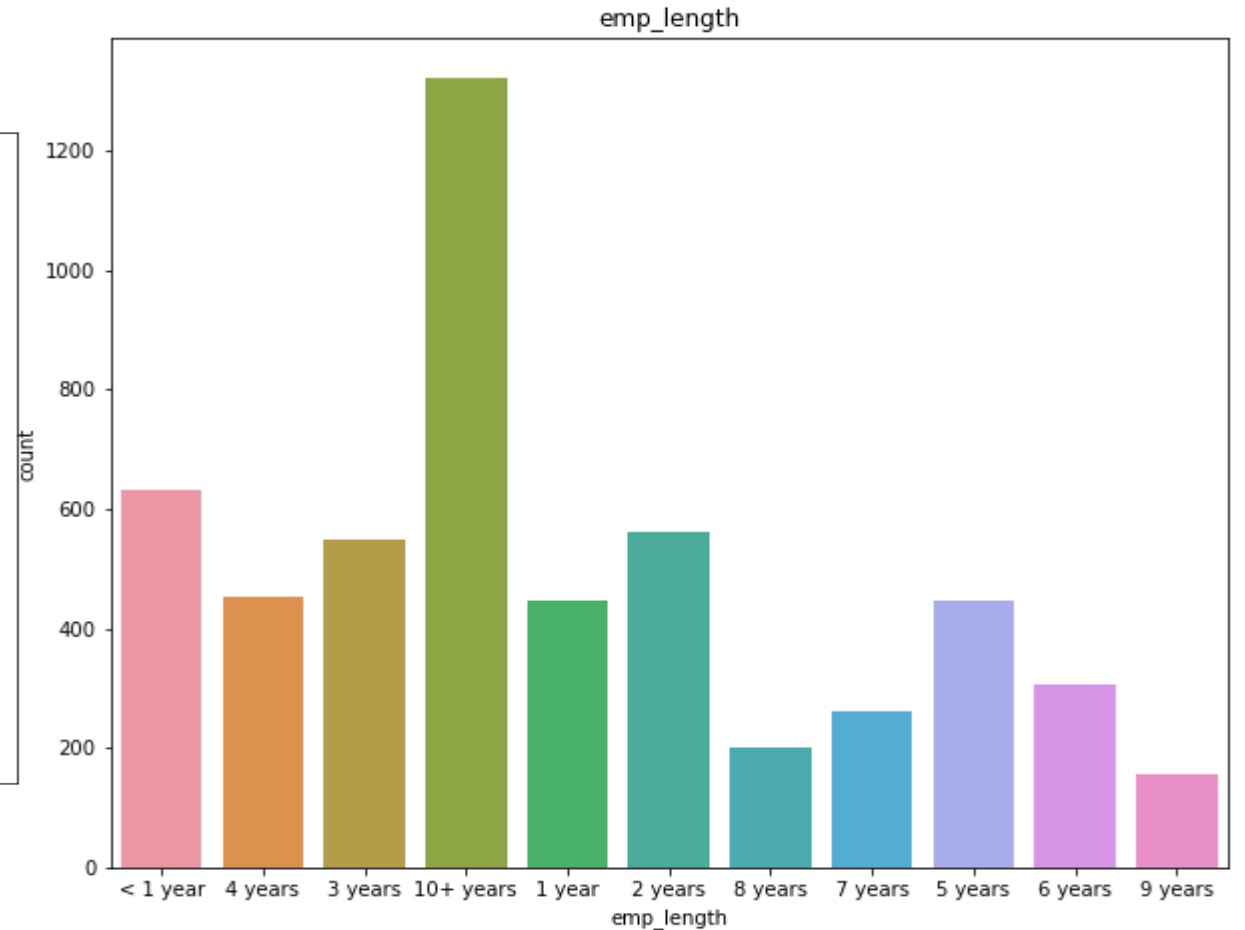


The 75% of loan defaulters have a higher amount of loan amount

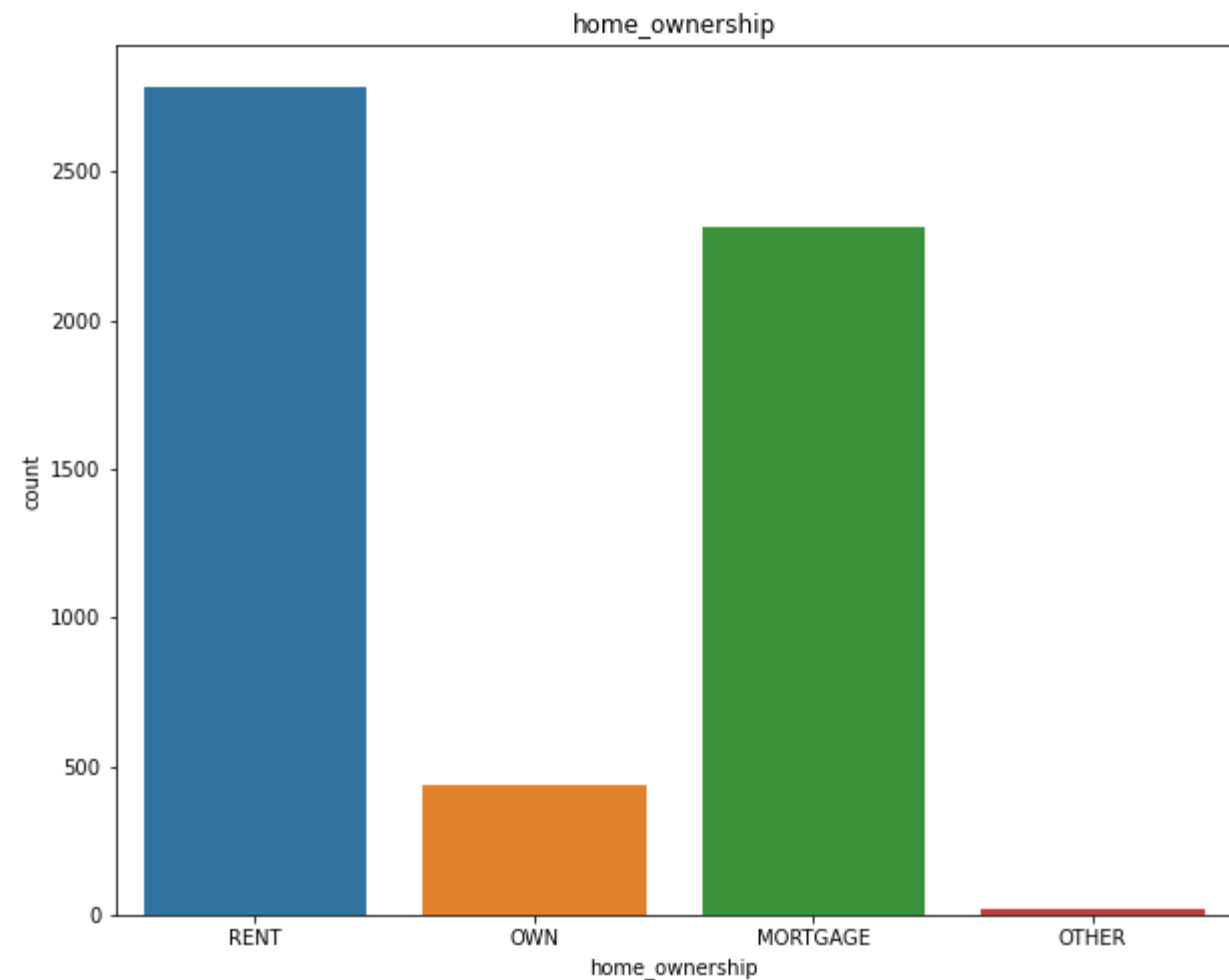
Segmented Analysis – Analysing the trend of variables for loan status as charged off



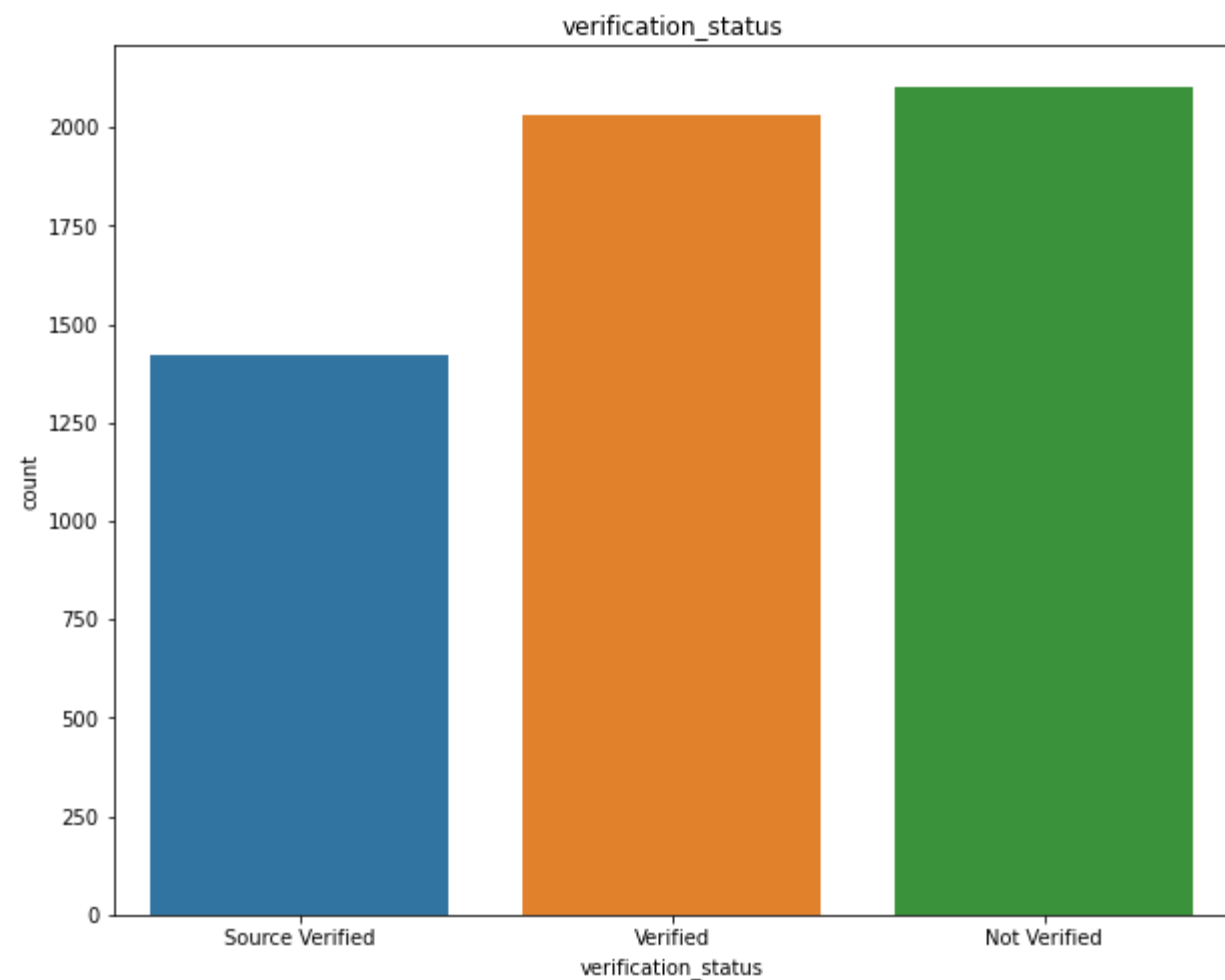
Defaulting is higher when the term is less. This might be due to high installments when term is less



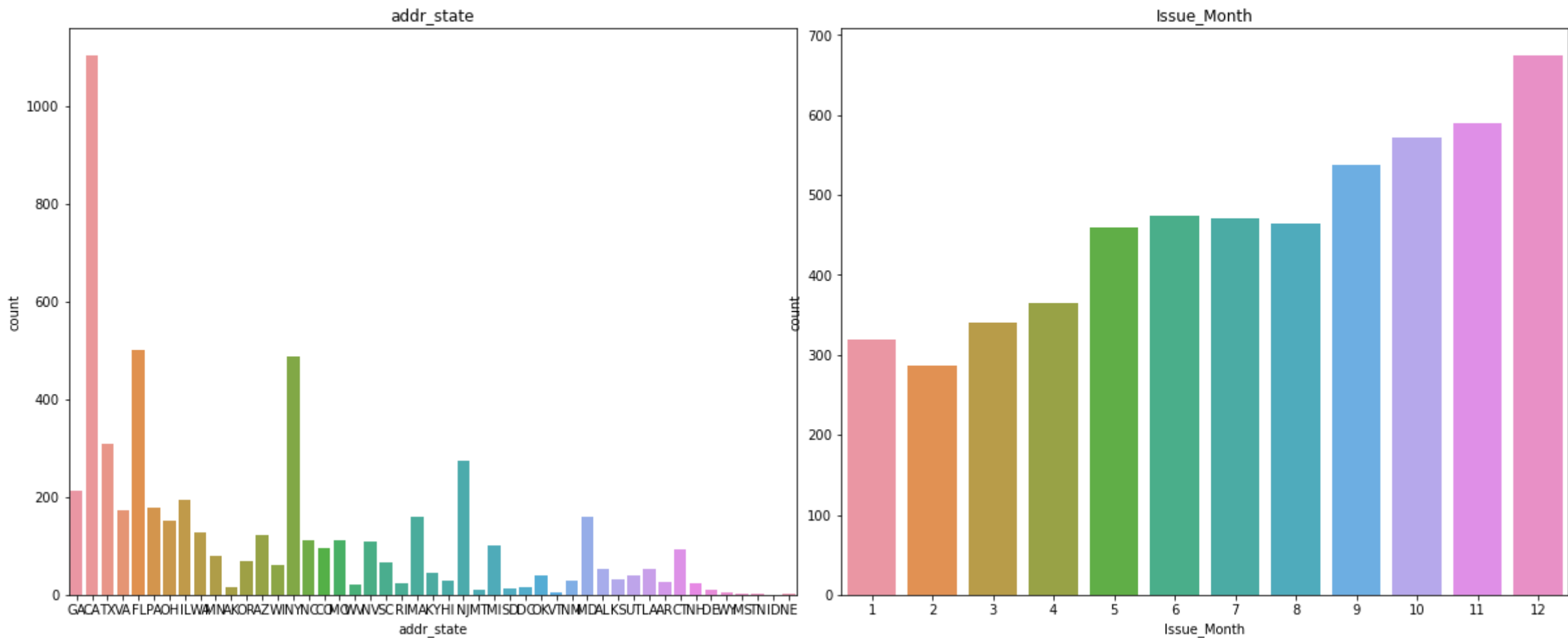
The charged-off customers are high in 10+ years employment category. Need to verify the age and occupation of the customers to understand their repayment capability



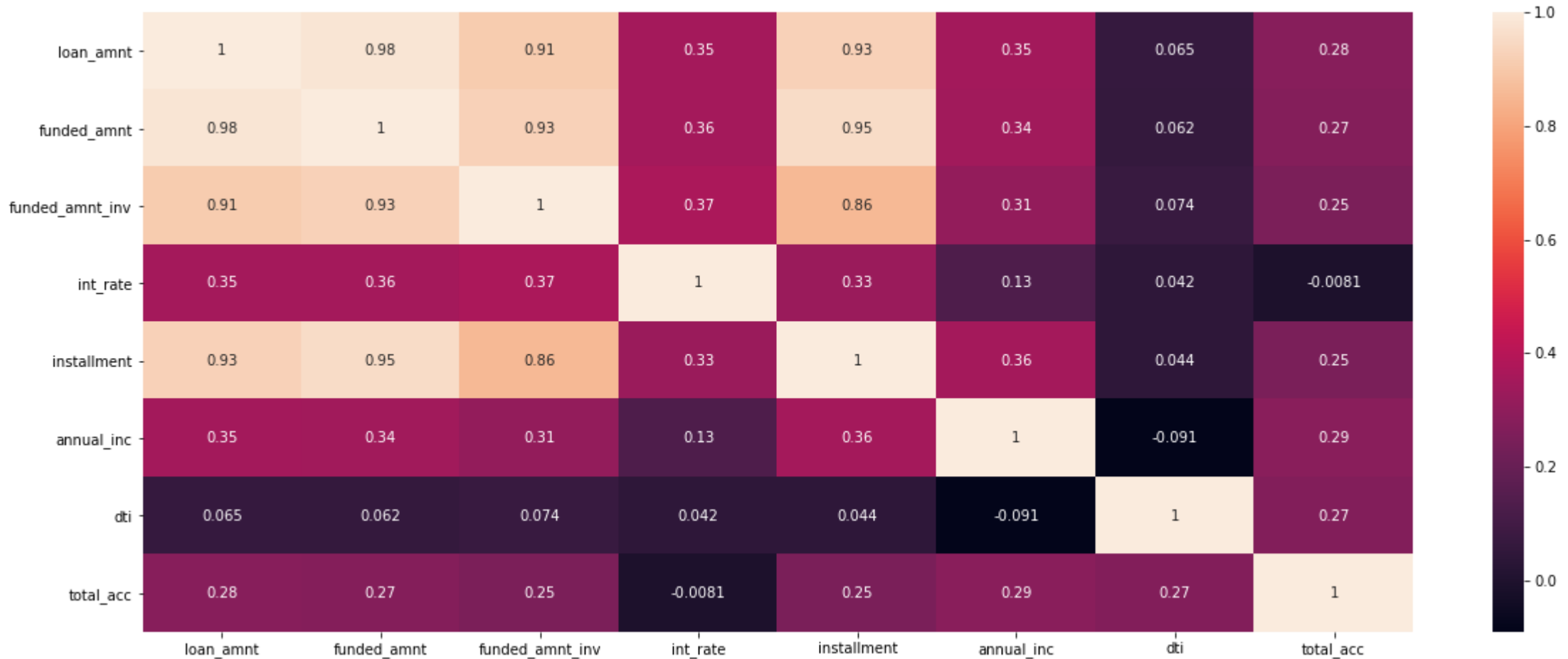
Customers who have Own home are less likely to default.



Most of the charged-off customers do not have their income verified. This should be an important category for loan consideration



Multivariate analysis



This heatmap is the correlation of various factors for Charged off customers. dti is poorly correlated. The loan_amnt and installment are highly correlated. However, the annual_inc is not correlated with either loan_amnt or installment. This could mean the customers would not be able to pay off the loans smoothly

Key Driving factors

- Verification status(Verification_status)
- Annual Income(annual_inc)
- Purpose of the loan (purpose)
- Interest Rate for loan (int_rate / interest_rate_range)
- Term for loan (term)
- Location(address_state)

Key Recommendations

- Applicant's source of income needs to be verified and this needs to be an action item
- Loan amount and the annual income amount should be proportional and needs caution on loan amount being sanctioned
- Customers who opt for lower terms might have high installments. customers with high installments tend to default.
- Customers with 10+ years of employment needs to be verified on the age and occupation to understand their repayment capability