Lending Club Case Study

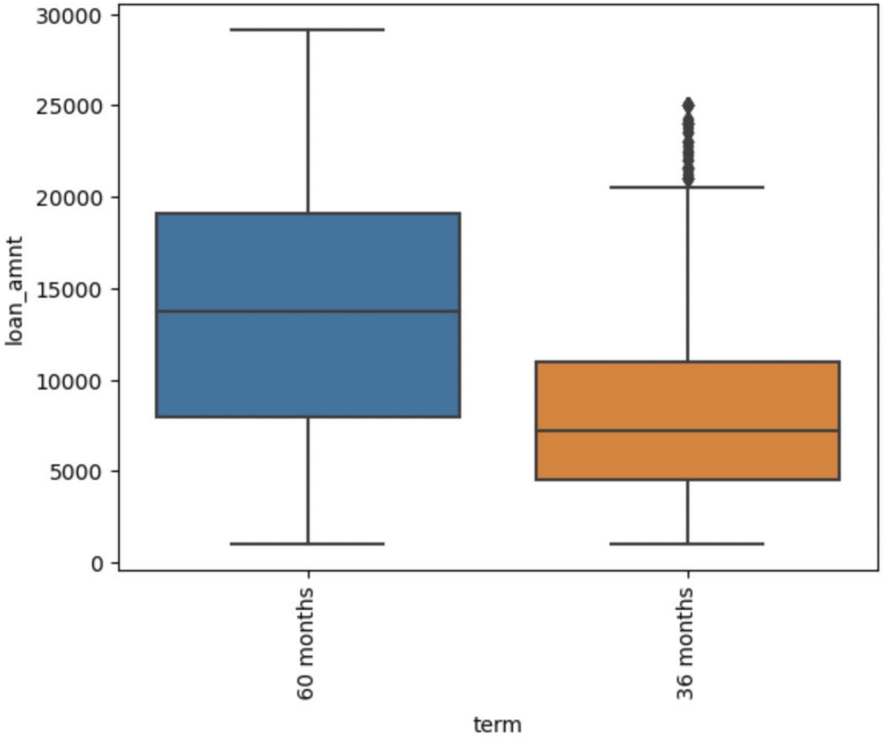
Nityam Rajen Vakil

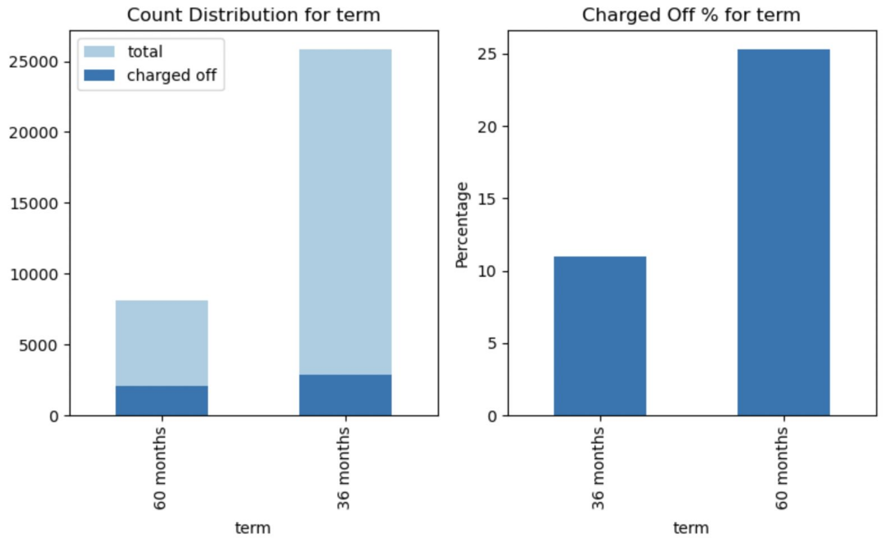
Batch: Oct 31 2023

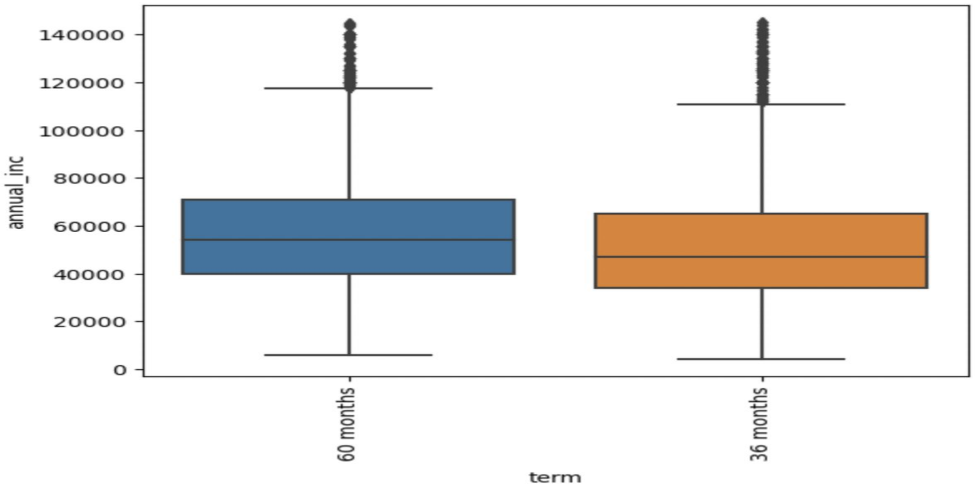
EDA Steps Covered

1. Data Cleaning and Manipulation
    a. Dropping rows, columns, fixing missing values, removing outliers, etc.
2. Segregation
    a. Separate columns into numerical, ordered categorical, unordered categorical.
3. Univariate Analysis
    a. Unordered/Ordered Categorical variables, Quantitative variables.
4. Bivariate Analysis
    a. Numerical vs Numerical
    b. Numerical vs Categorical
5. Multivariate Analysis
    a. Numerical vs Numerical
    b. Categorical vs Categorical

36 Term vs 60 Term







36 Term vs 60 Term

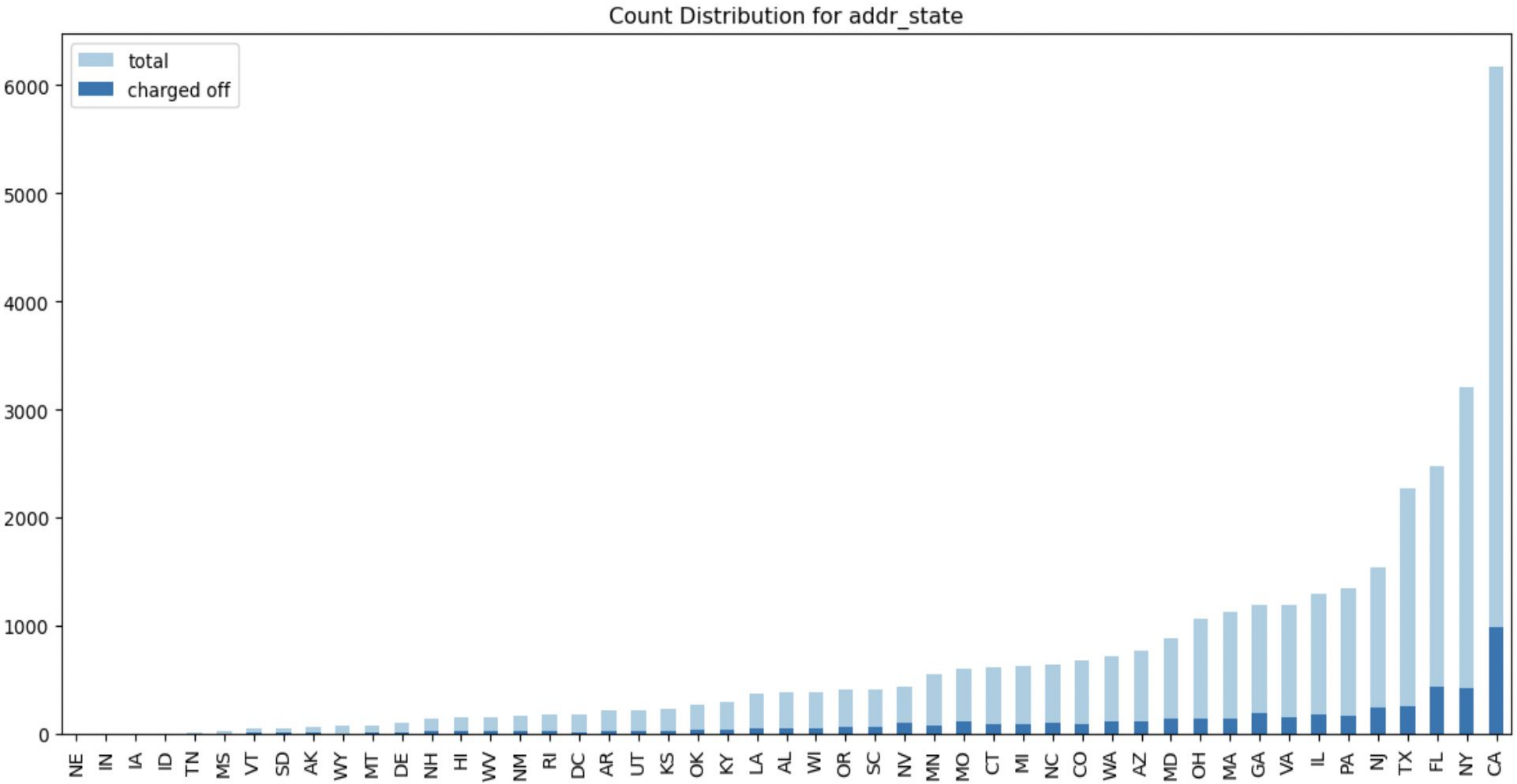
Conclusions:

1. There are almost thrice more loan takers for the term of 36 months with 15% less defaulters than 60 months term.
2. 60 term have issued very high loan amounts compared to 36 term. While annual income of both categories is almost similar.

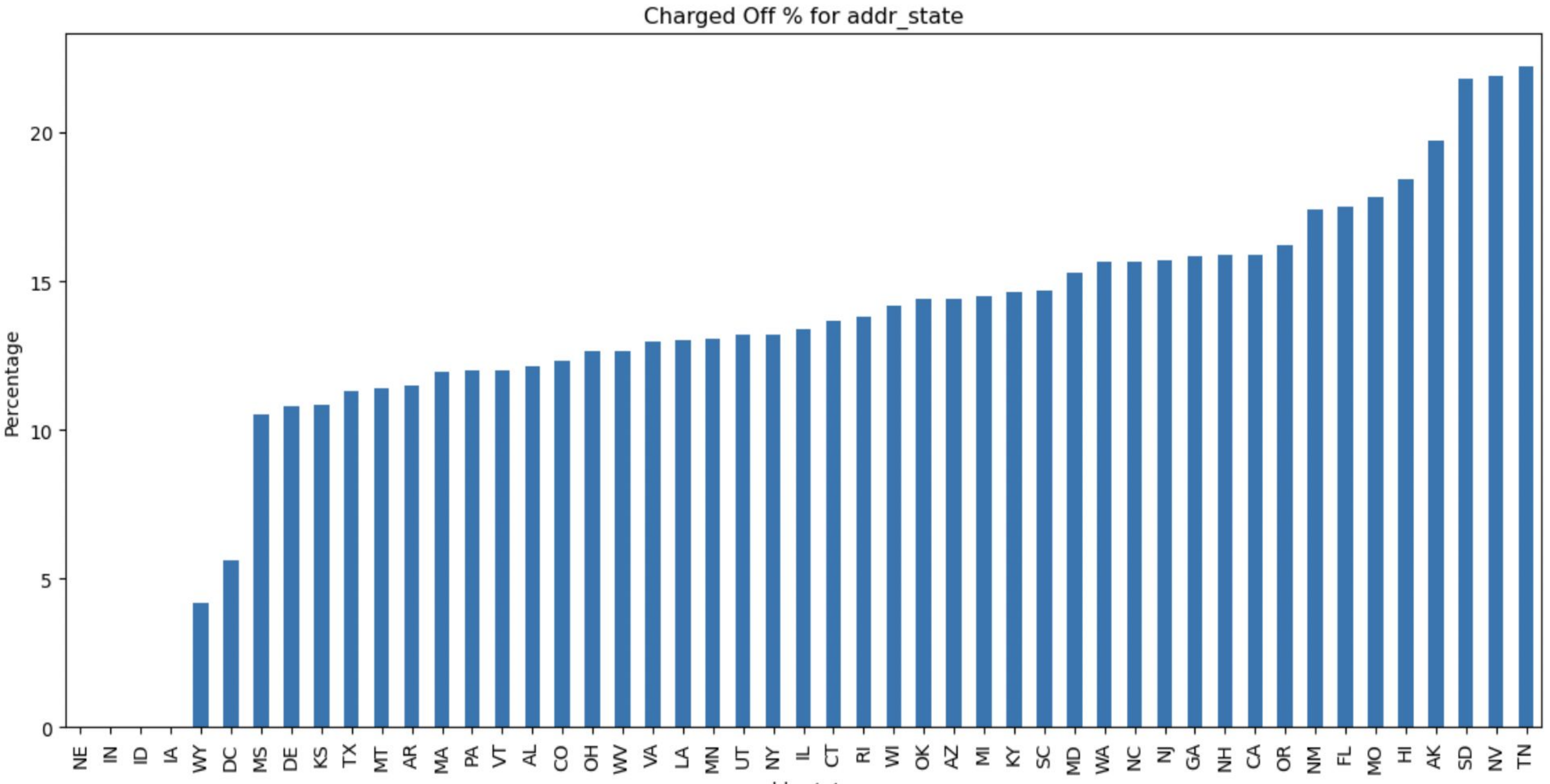
Actions:

1. Do not issue loans for 60 term when income is similar to 36 term.

State Address



State Address



State Address

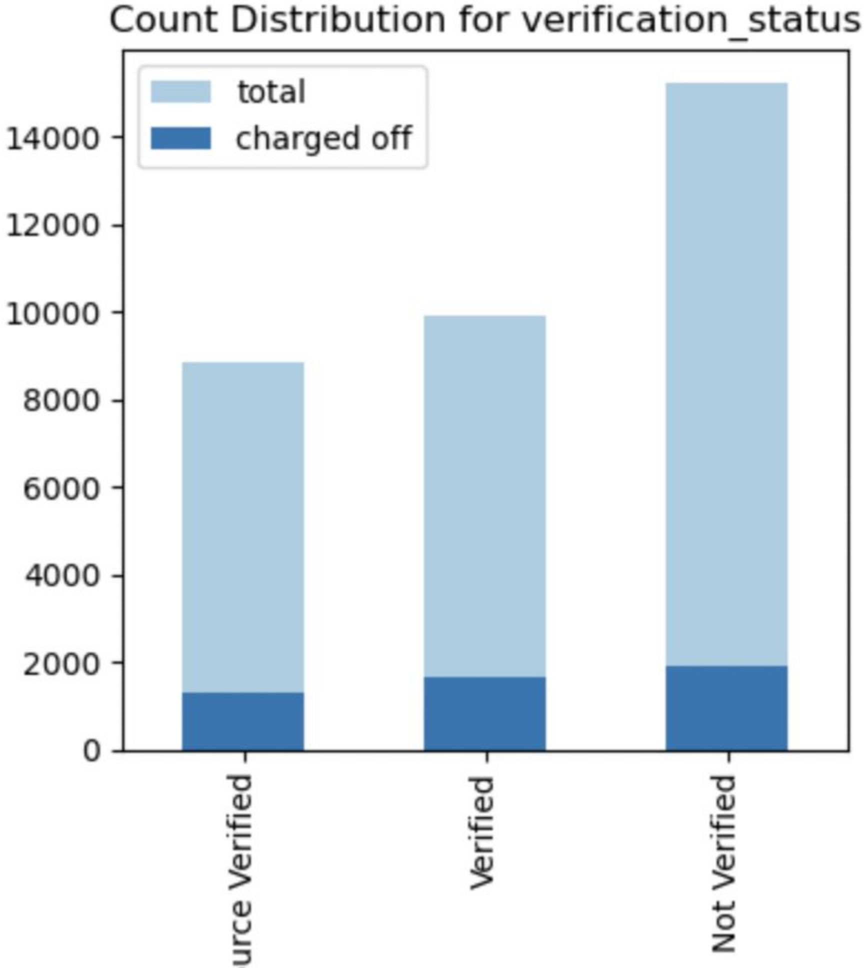
Conclusions:

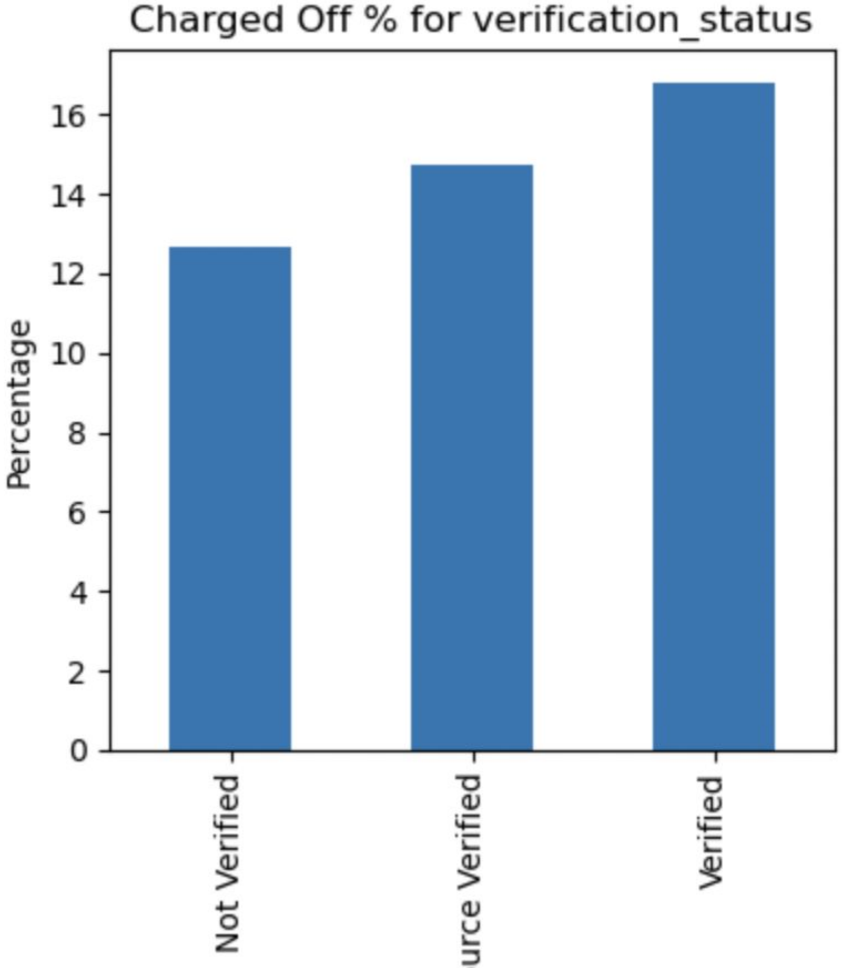
1. Interest rate for NY is less than 15% while CA goes above 15%.
2. NY is a better state for profitability than CA.
3. AK, NV, SD, TN states should be avoided to issue loans due to very high default rate.

Actions:

1. Reduce loans to AK, NV, SD, TN to prevent defaults.
2. Invest more in NY than CA.

Verification Status





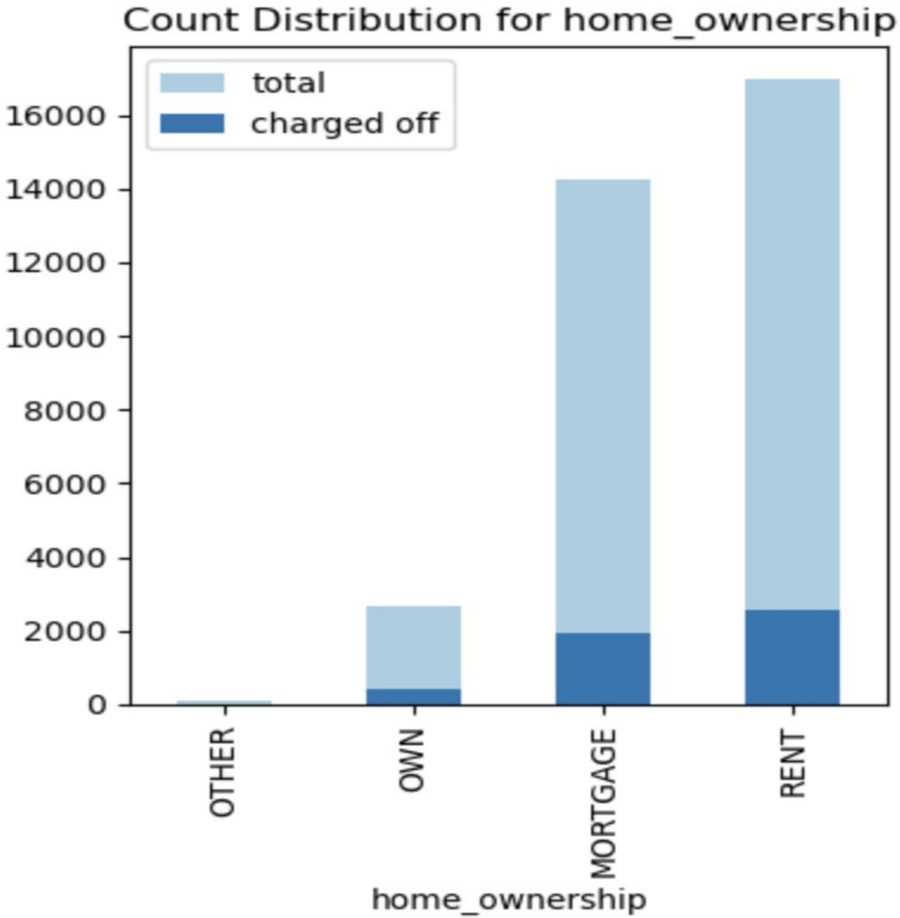
Verification Status

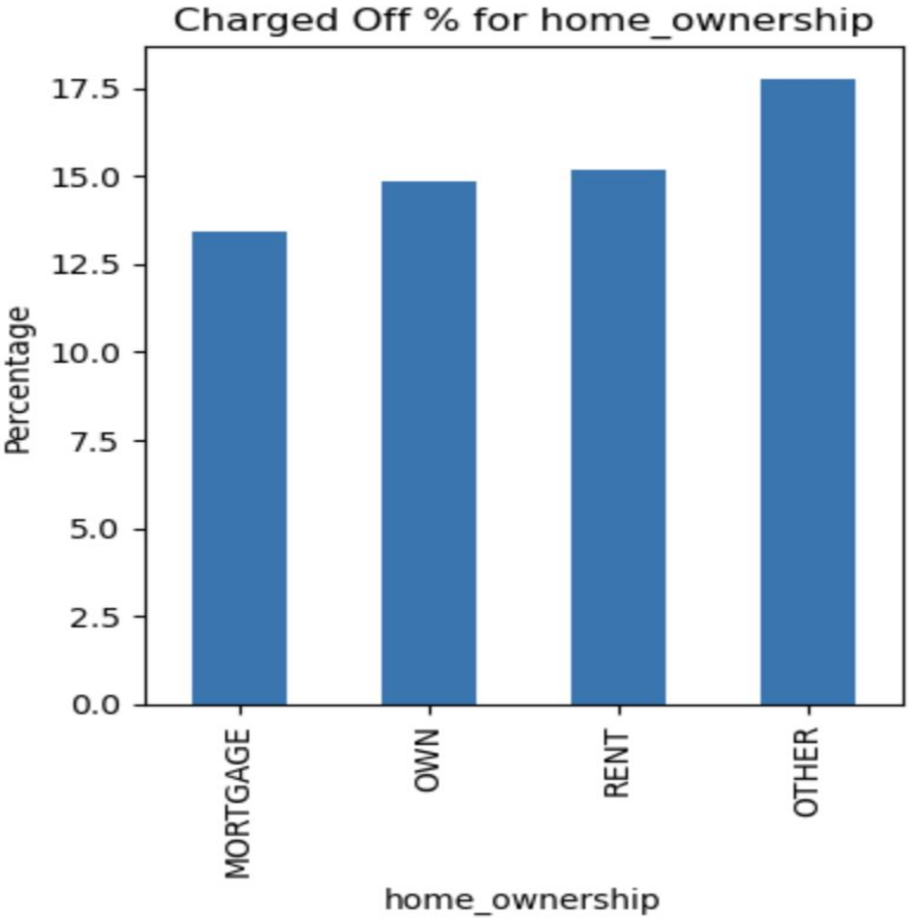
Percentage of defaulters are lower for Not Verified while having the most customer base. This is very counter intuitive.

Actions:

1. Verification does not necessarily lead to less defaults, instead Non Verified customers default less.

Home Ownership





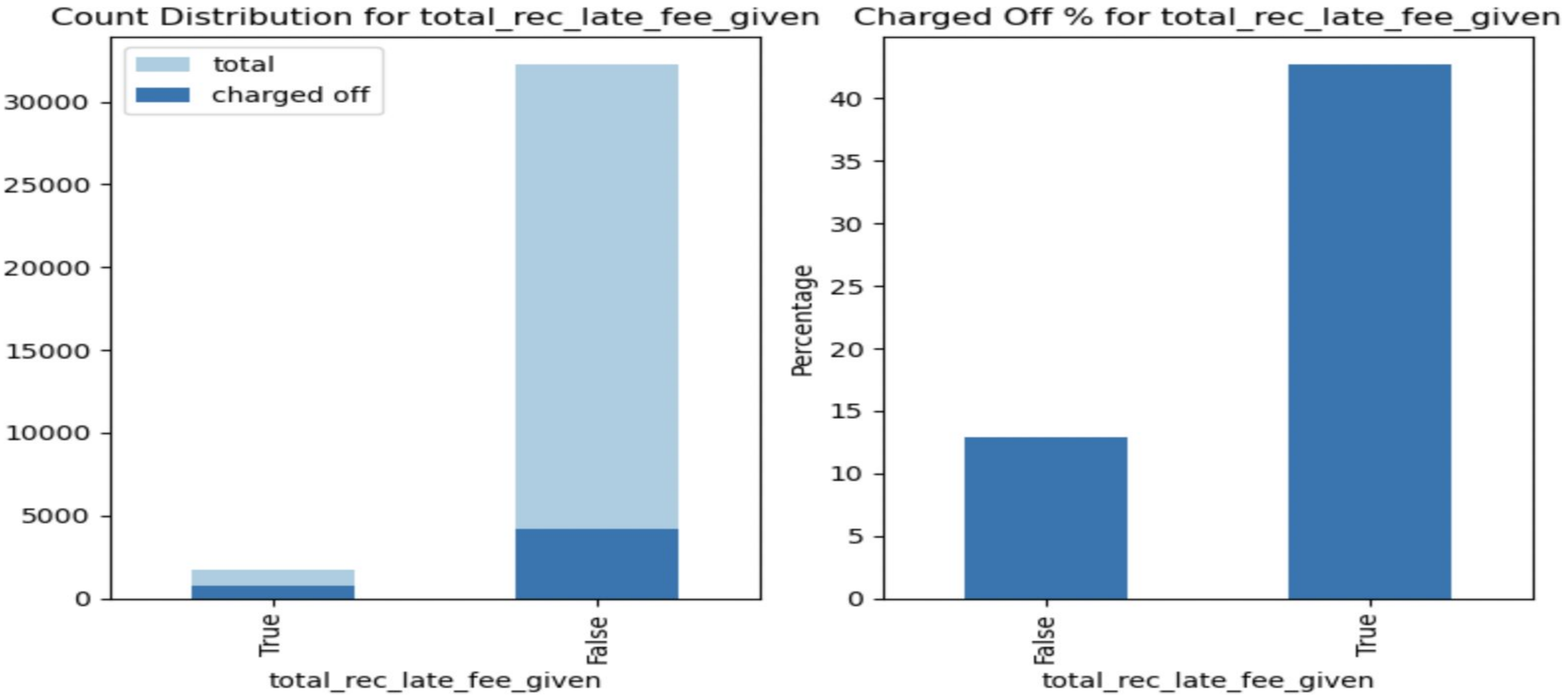
Home Ownership

Rent and Mortgage give almost similar business but Mortgage has less default percentage.
Own and Others has very high default rate
while business is very less.

Actions:

1. Invest more in Mortgage
2. Avoid Own and Other

Late Fee Received

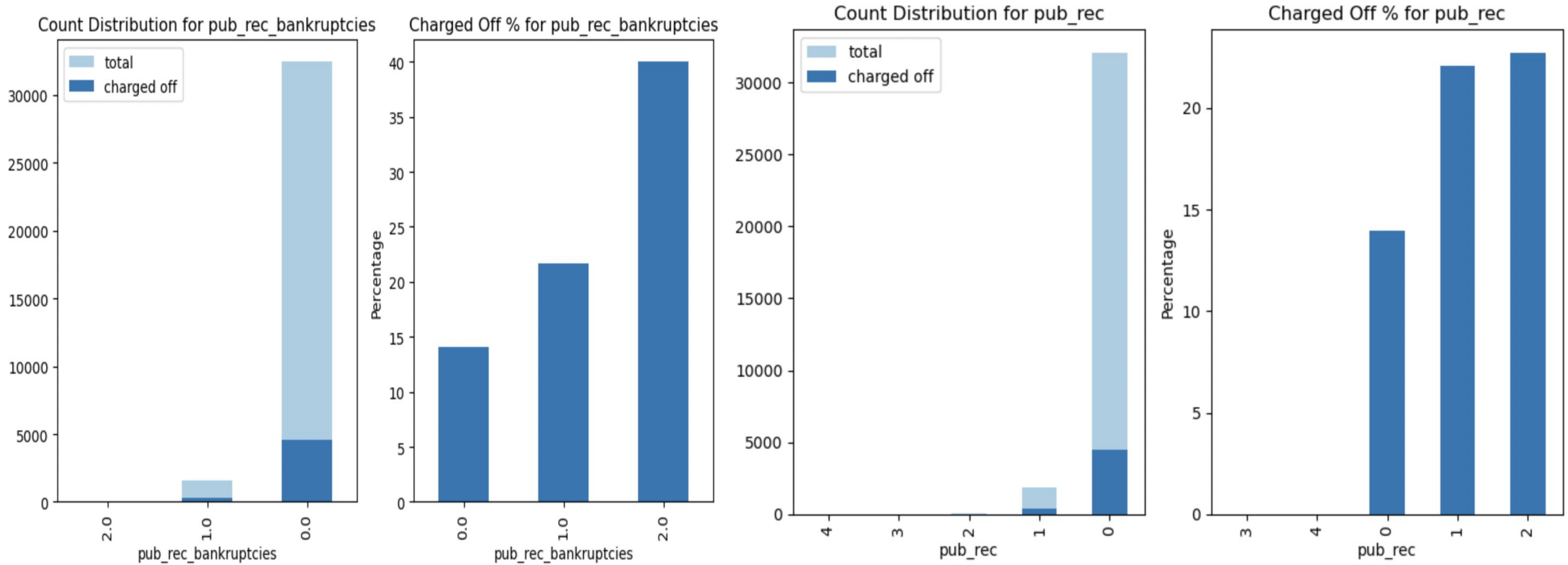


Late fee payers have over 40% chance of defaulting.

Actions:

1. Avoid issuing loans to those who pay late fees in any other instrument.

Public Record and Bankruptcies

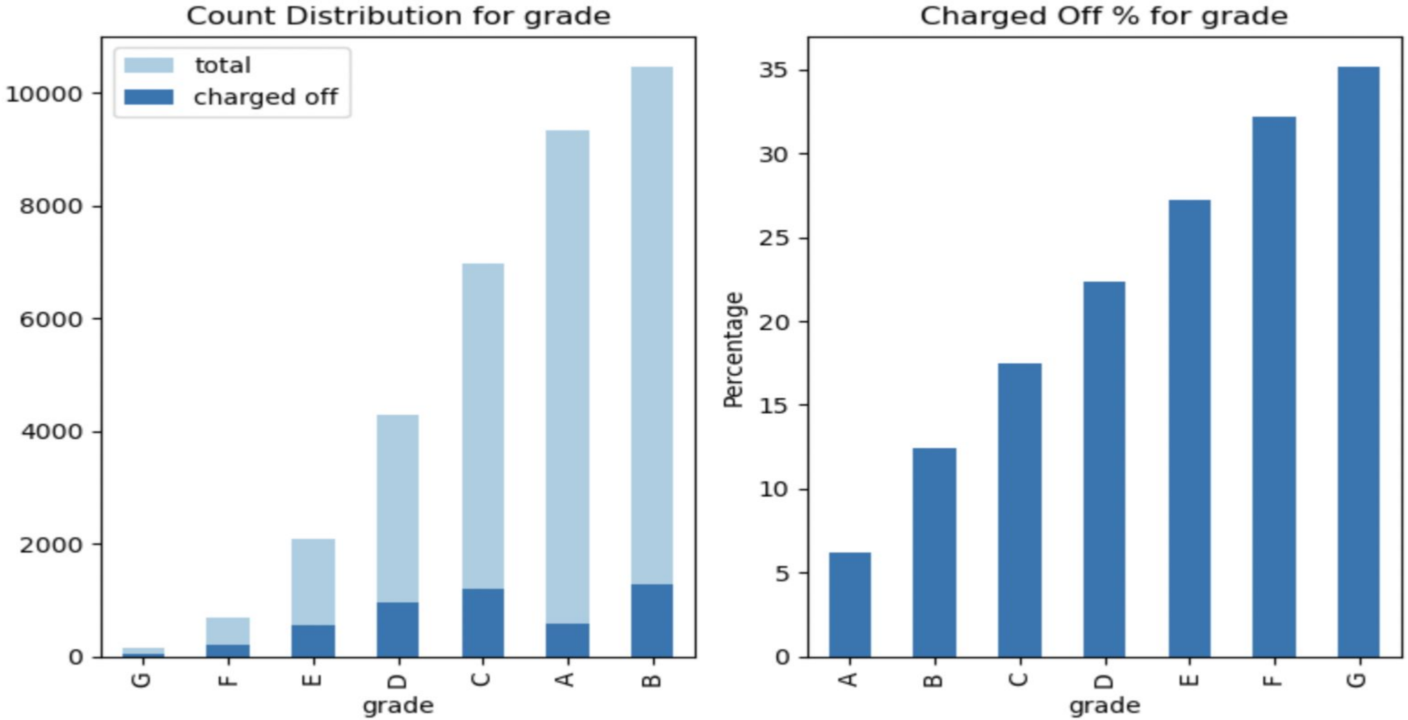


As grade of loan increases, percentage of defaulters go higher.

Actions:

1. Do not issue loans of D E F G categories, they have over 20% default rate.

Loan Grade

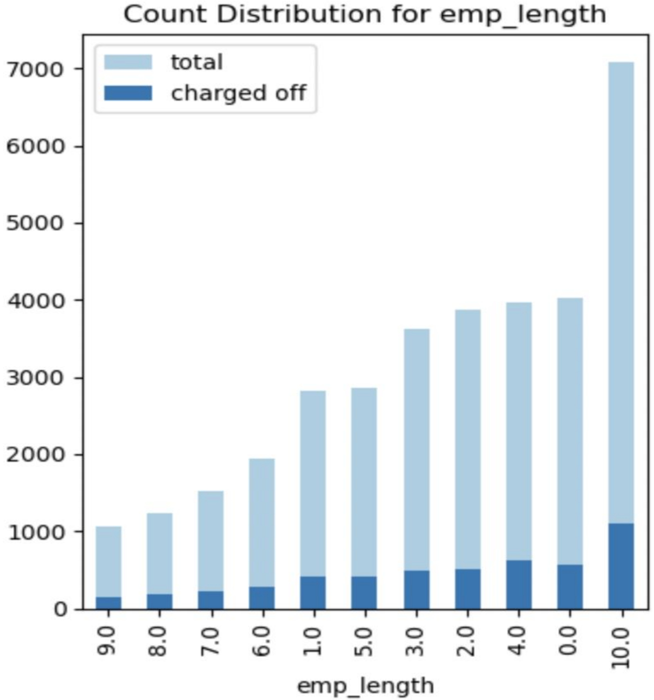


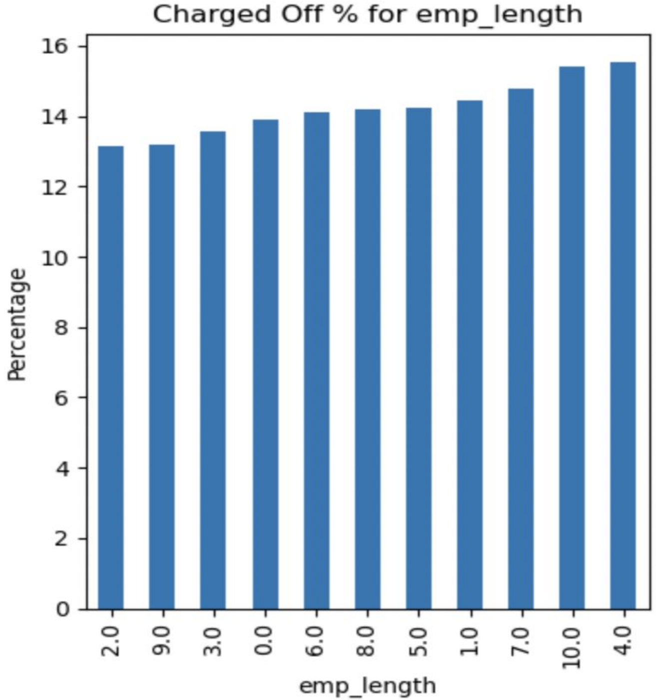
Almost 25% default rate where there is at least 1 public record or recorded bankruptcies.

Actions:

1. Issue loans only to those with no public records or bankruptcies.

Employment Length

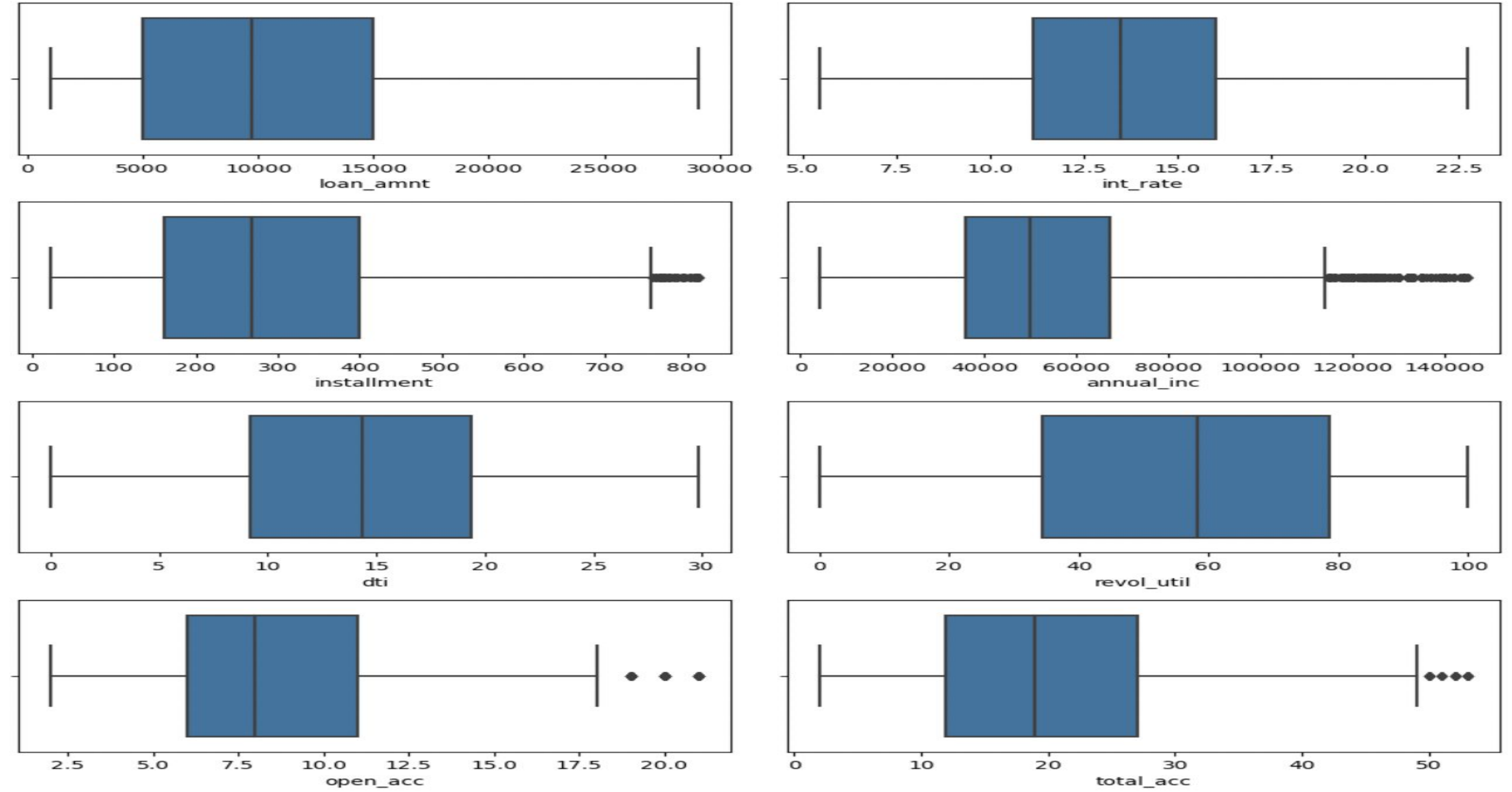




Highest default rate with length of 4 years.
Actions:

1. This would be due to high expenses in other factors in life such as marriage, house buying, car etc. Avoid 4 year emp length issuance

Quantitative Variables Analysis



Quantitative Variables Analysis

Majority of defaulters:

1. issue loan between 5-15k
2. have interest rate of 11-16 with median at 14
3. have annual income between 3.5L to 7L
4. have dti between 10-20
5. have revolving line utilization rate between 10-30
6. have open credit line between 6-11

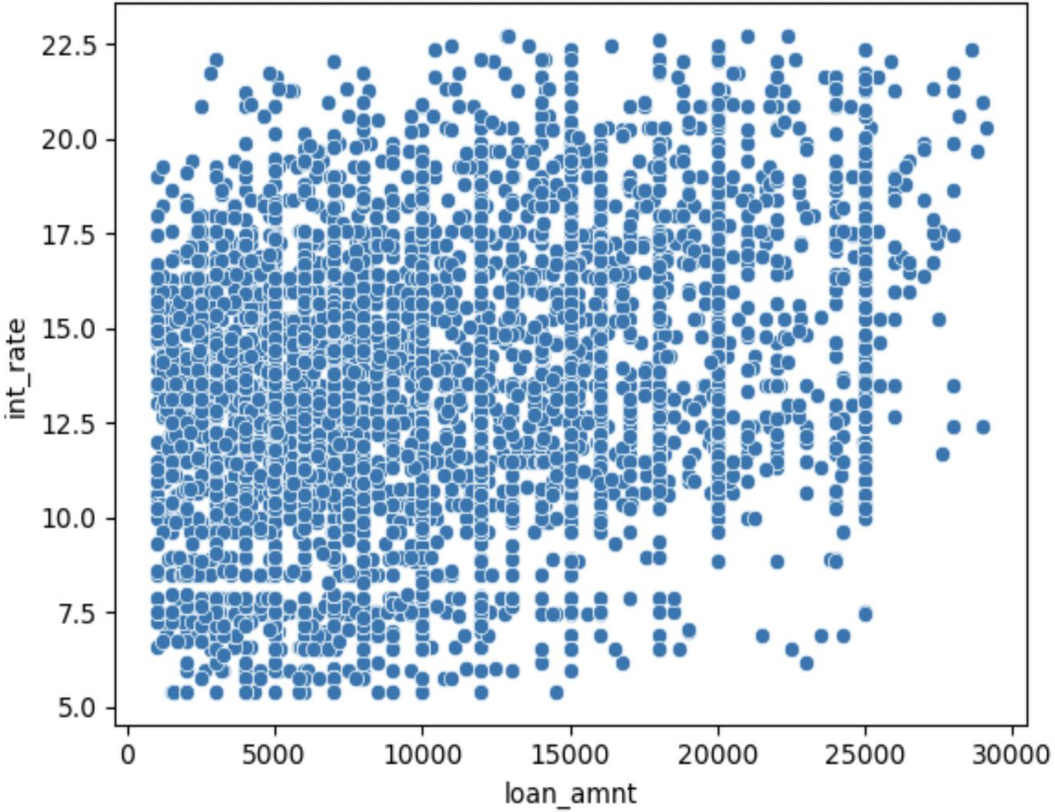
Avoid all these.

Bivariate Analysis: Numerical vs Numerical

The density of defaulters is high when loan amount is between 2k-7k with interest rates ranging from 11-15%.

Actions:

1. For smaller loan amounts interest rates should be kept very high to avoid defaults.

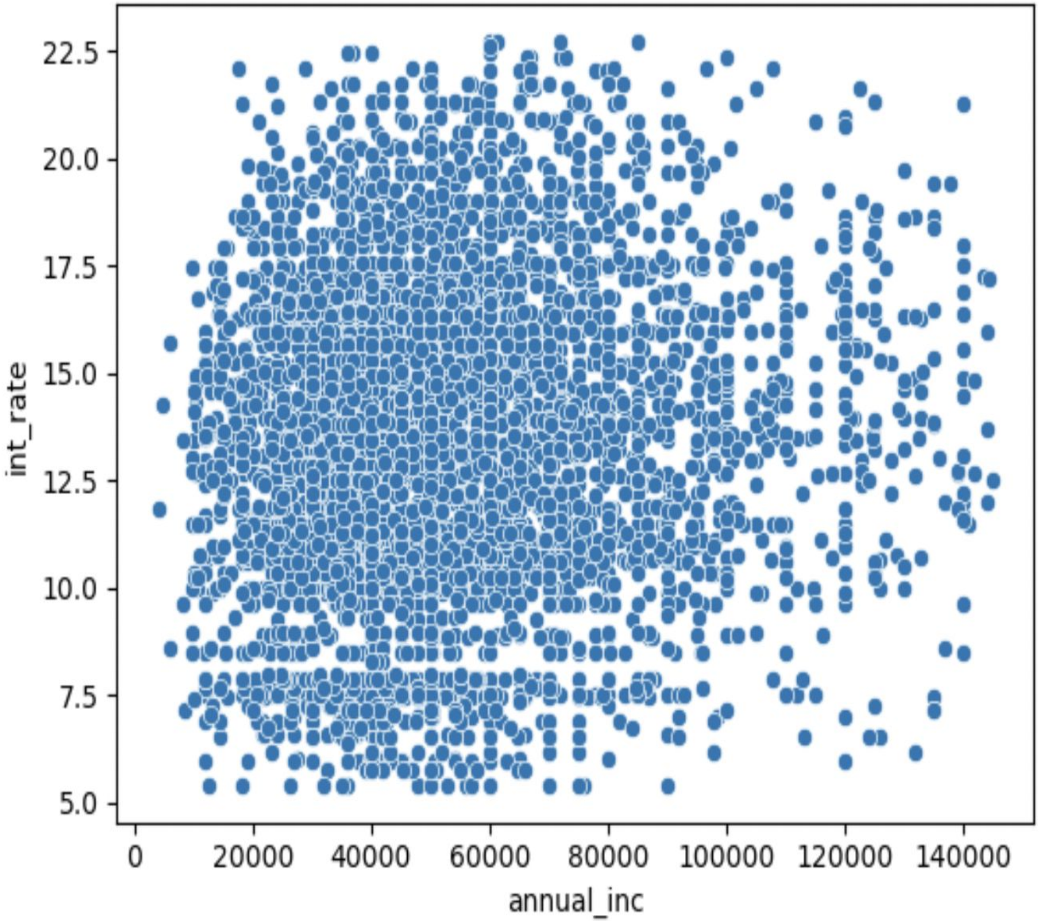


Bivariate Analysis: Numerical vs Numerical

There is high density of defaulters for mid range interest rate and lower side annual income.

Actions:

1. For lower annual income customers, keep interest rate lower.

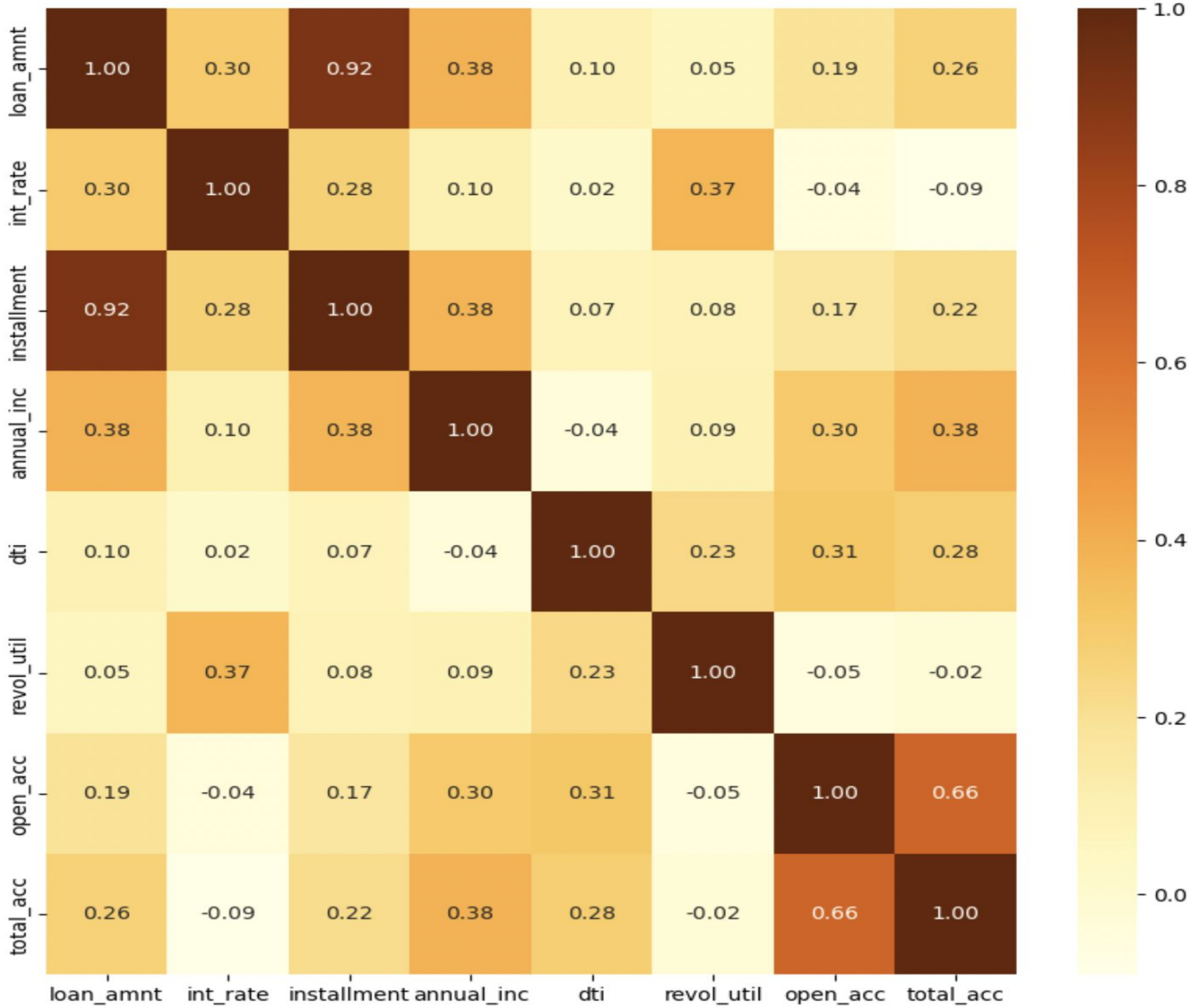


Bivariate Analysis: Numerical vs Categorical (Refer this section for graphs)

Other actions based on Multivariate Analysis:

Actions:
- Do not issue loans for 60 term when income is similar to 36 term.
- Do not issue loans to customers with higher employment duration yet having lower income (less than 6).
- Increase interest rates for Home Owner with “Other” category.
- Do not issue E, F, G loans for customers having income less than 6.

Multivariate Analysis



Multivariate Analysis

Actions:

1. Loan amount and Interest rate have correlation of only 0.3 which suggests interest rate can be increased for higher loan amounts.
2. Annual income and Loan Amount have 0.38 correlation. So give higher loans only for very high annual income customers.