Part II – Factors that Determine possibility to Payment of Loans

by Zaburi Frolian Mwachusi

Investigation Overview

In this investigation, I wanted to look at the features that can predict the possibility of the loan to be repaid. The main focus was with Employment Status, Debt To Income Ratio, Property ownership, Income Range, Employment Status Duration, Stated Monthly Income, Borrower State, Available Bankcard Credit, Recommendations, Investors, and Occupation

Dataset Overview

This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others

Distribution of Loan Status

Loan status has long tail distribution with most of loans being Current. LoanStatus is clearly unimodal with one peak at 'Current' status. 56576 loans are in current status followed by 38074 with completed loans
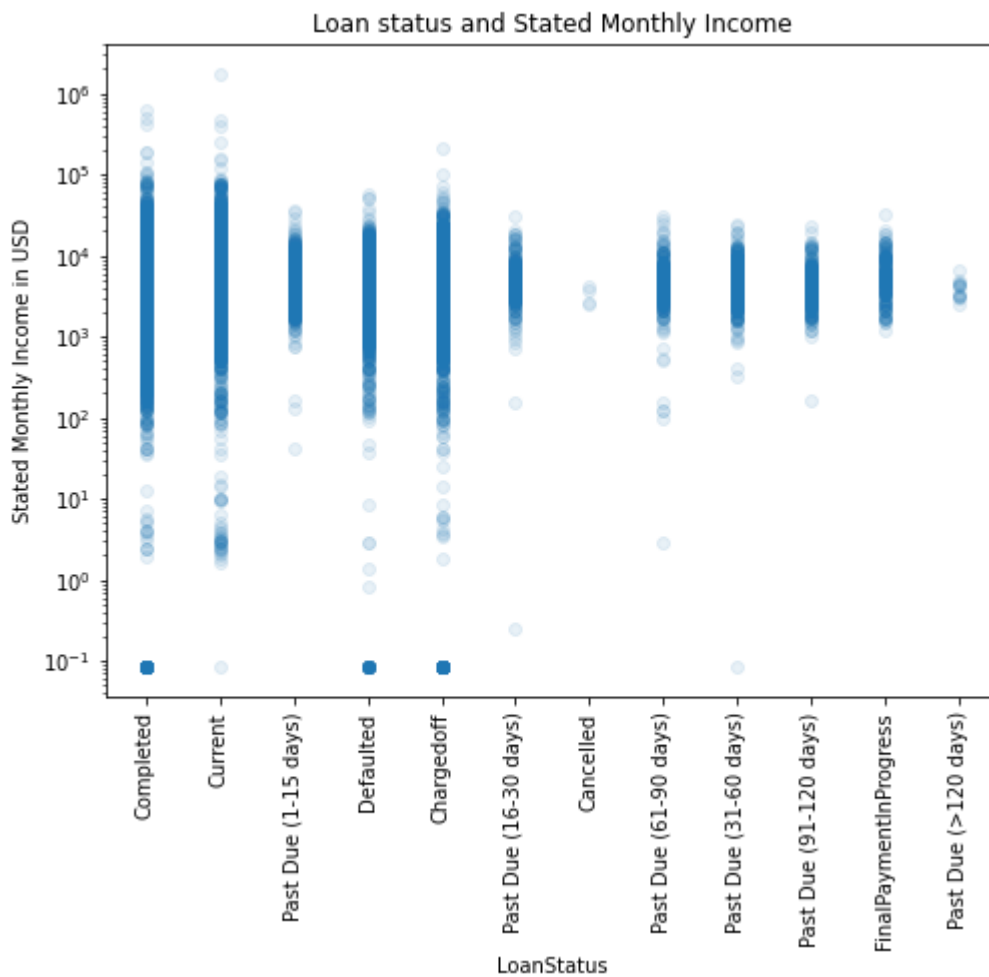
Distribution of Loan Status
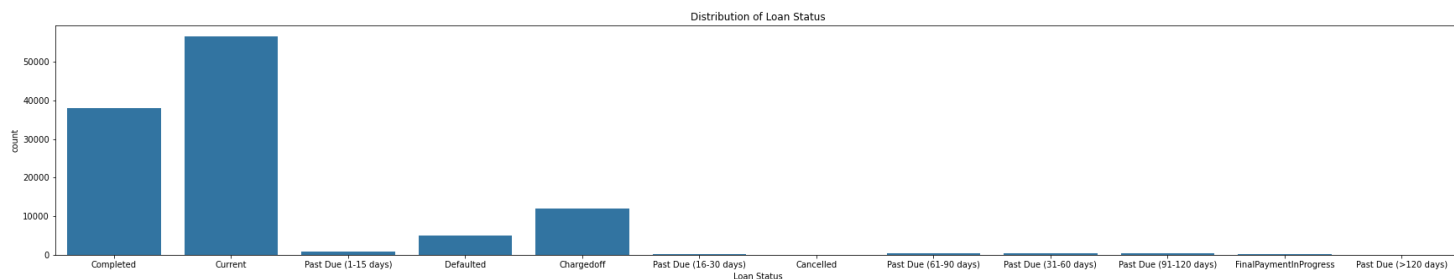
Distribution of Stated Monthly income

The distribution of stated monthly salary is Unimodal with peak between USD2,000 and USD10,000 also slightly left skewed
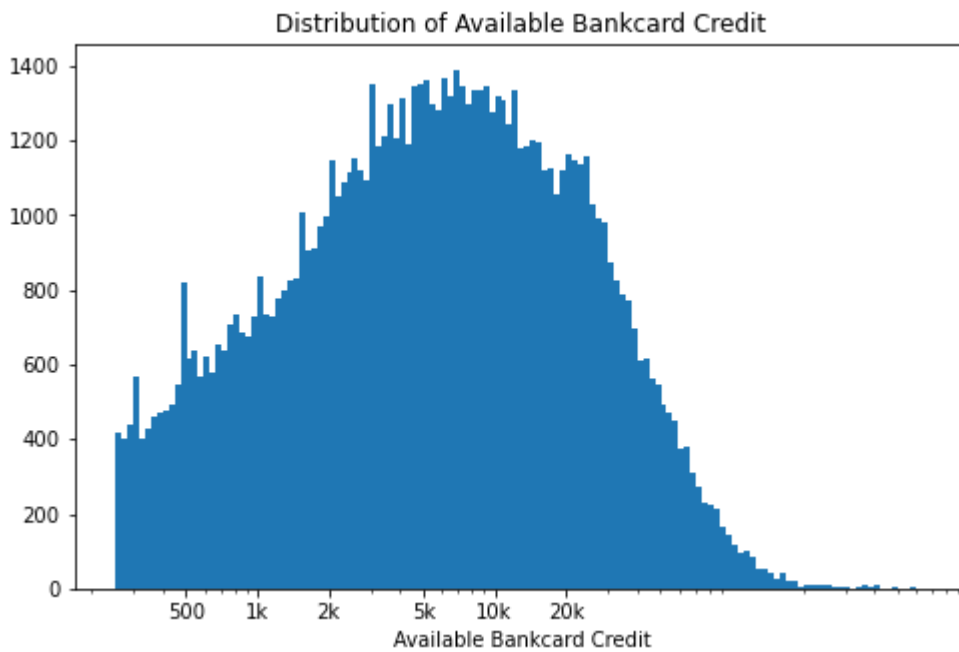

Distribution of Stated Monthly income
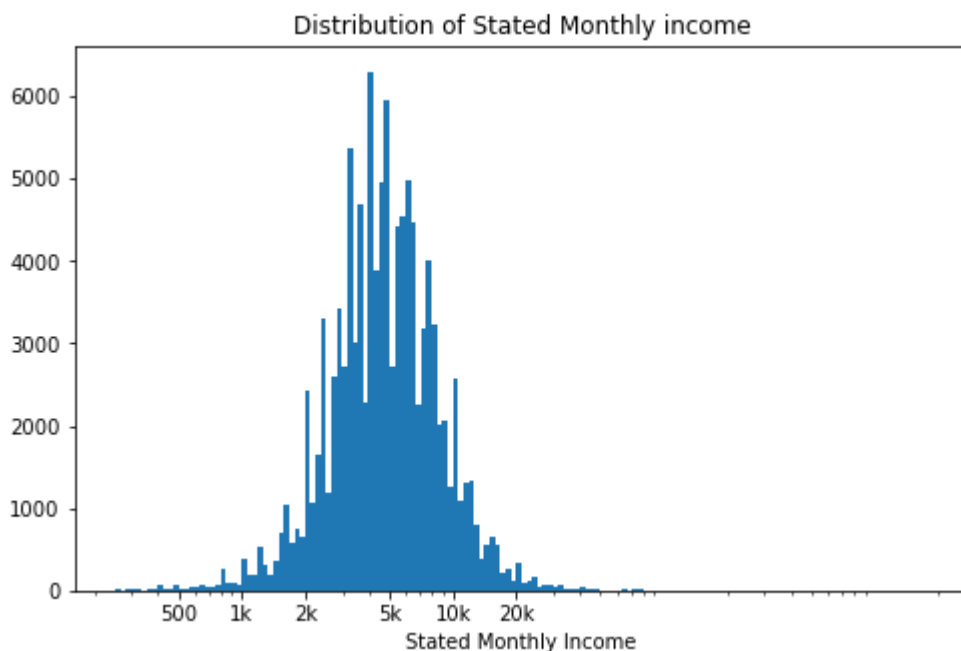
Stated Monthly Income Vs Loan Status

There is high correlation between stated monthly income with the loan status especially the completed and current loans As the stated monthly income increase also the correlation increase thus showing for people with high stated monthly income has high number of completed loans together with current loans

Loan status and Stated Monthly Income

Distribution of Available Bankcard Credit

The distribution have flat top ranging from USD2,000 to USD20,000 and slightly
right skewed. There is high accumulation of loans where by individuals have
bank card credits USD2,000-20,000

Distribution of Available Bankcard Credit

Available Bankcard Credit Vs Loan Status

There is high correlation between Available Bankcard Credit with the loan status especially the completed and current loans As the Available Bankcard Credit increase also the correlation increase thus showing for people with high Available Bankcard Credit has high number of completed loans together with current loans

Loan status and Available Bankcard Credit
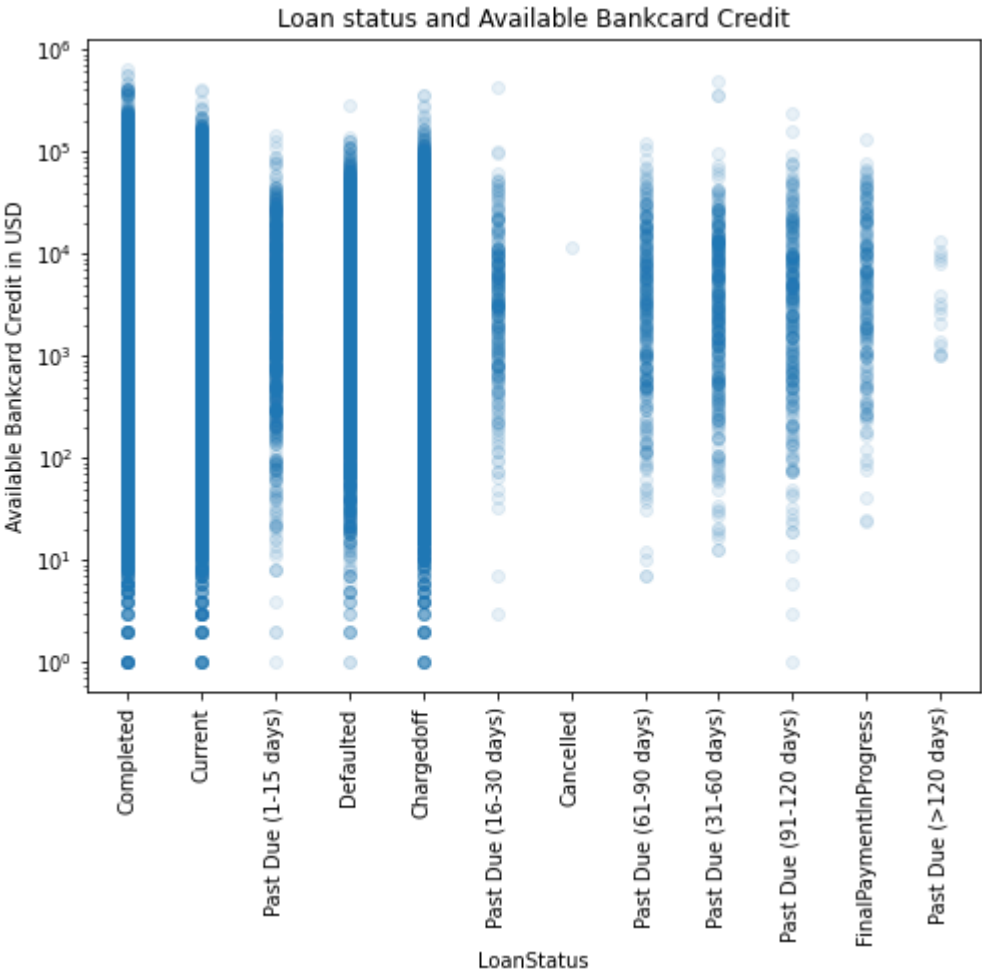
Loan Status and other numeric variables

There is some negative relationships for the features like debt to income ratio and Recommendations. When the value of these features increases the number of completed loans and current loans tend to decrease these features are worthwhile exploring before reaching the final conclusion for the financial institution

```
<Figure size 720x720 with 0 Axes>
```


Distribution of loan status to other numeric variables

Distribution of loan status to other numeric variables
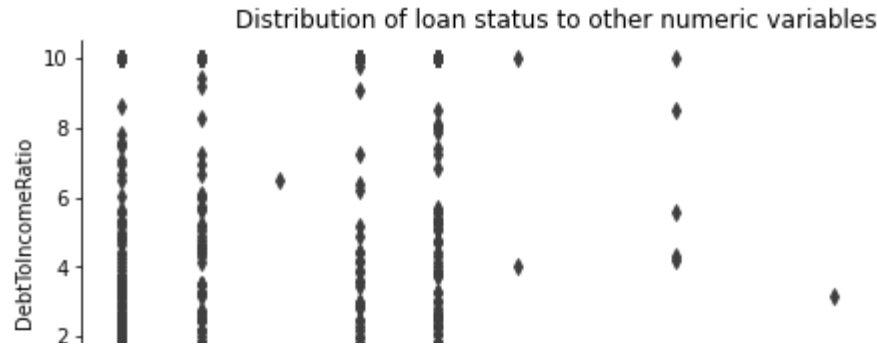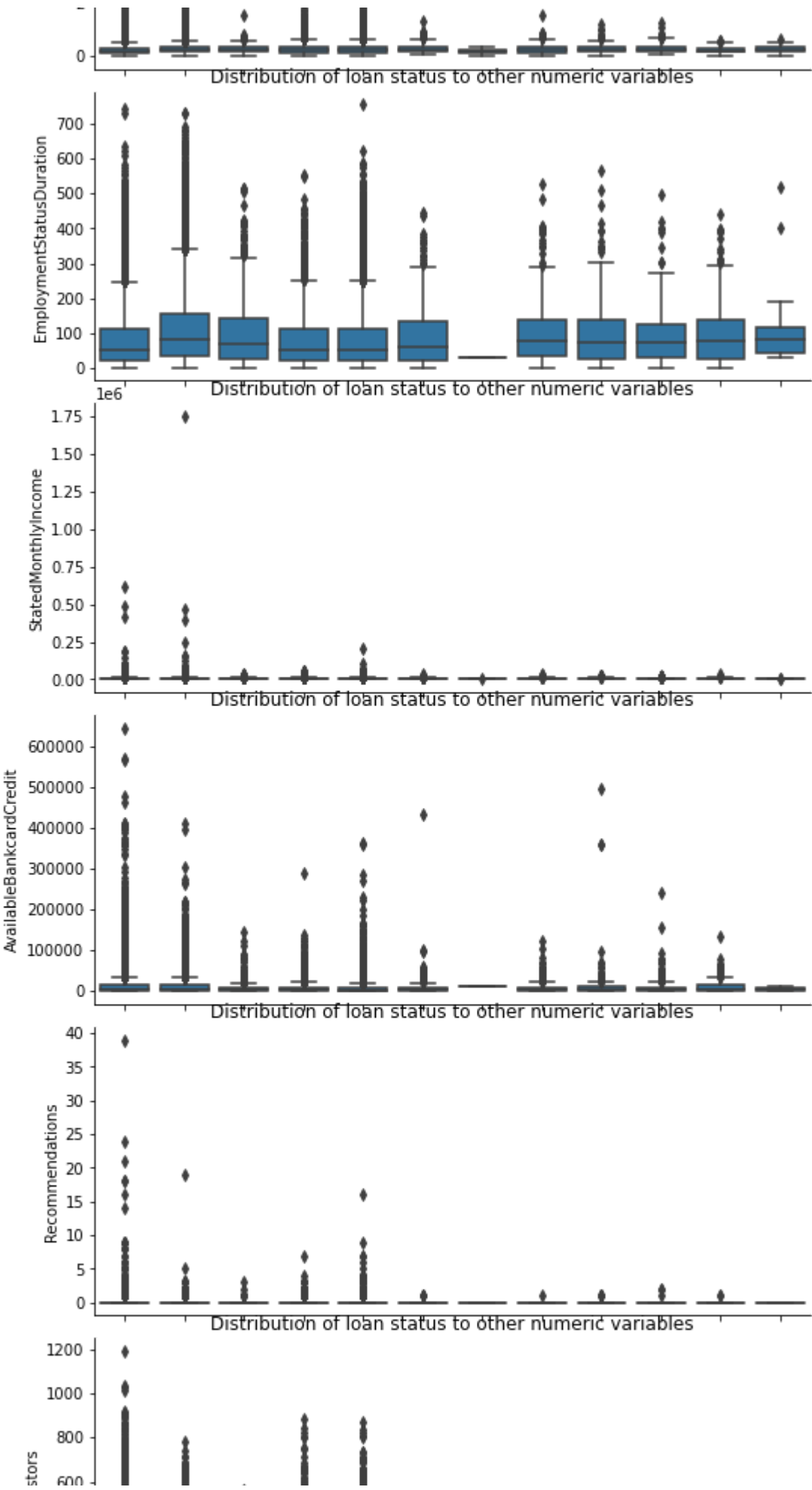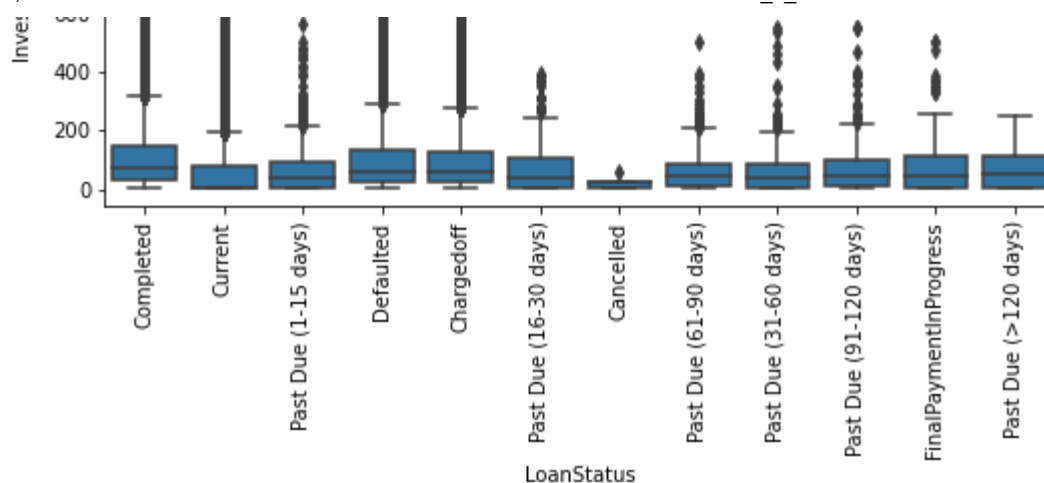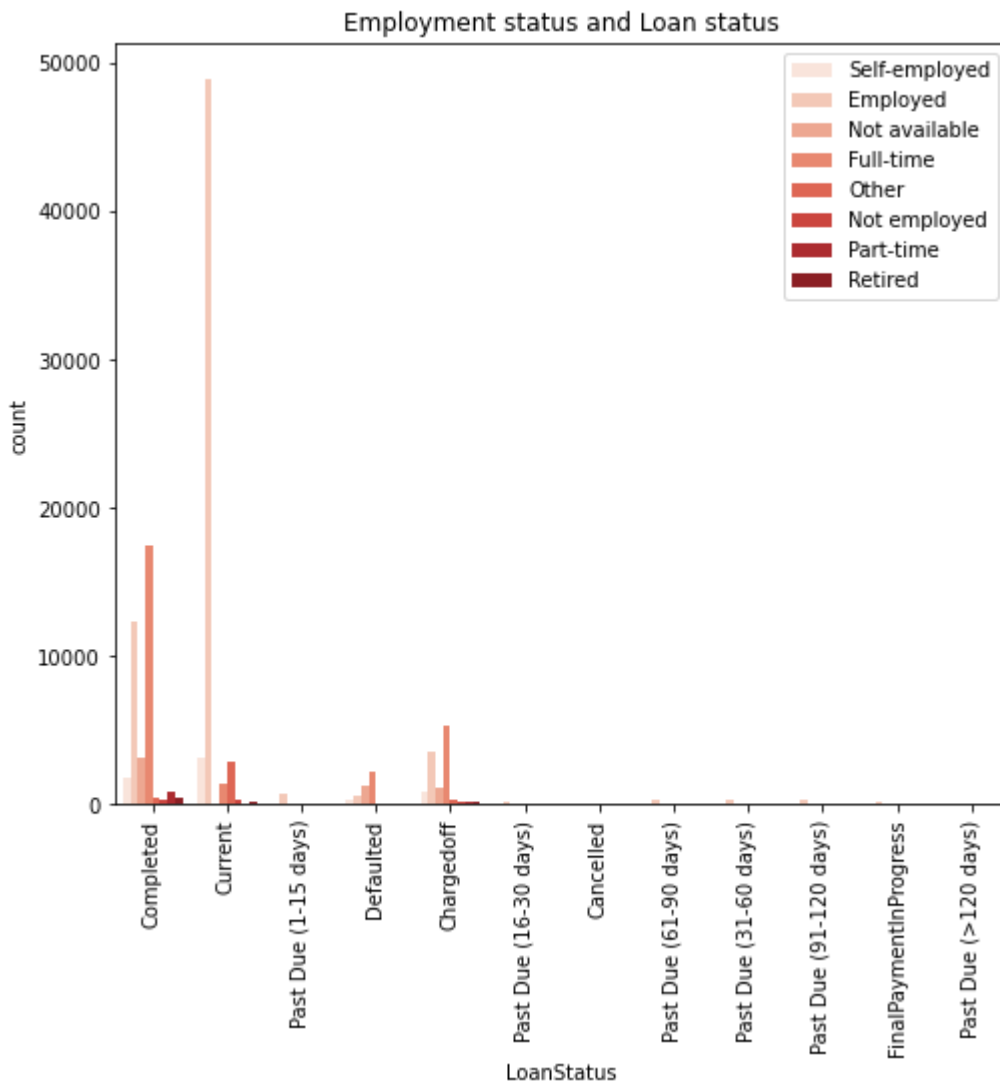
Loan Status with Categorical Variables

For employment status employed and full-time showed high trend for completed and current loans the trend decrease as you approach not employed and loans from retired individuals. Guess what? For the property Ownership Trend went so different from our perspective loans that are not home owners had a lot of completed loans. But for the current loans home owners beat the number. We expected property ownership to be one of the key features for loans completion as most of financial institutions use houses and other properties as collaterals to offer access to loans. But data shows people with no houses are the one who completed there loans at high number compared to those with property ownership

This suppose to be taken to much of consideration. Financial institutions must modify there rules and procedures to make sure those with considerable cashflow are prioritized

For the state of borrowers trend suggests completed and current loans are mostly from Colorado, Minnesota and trend decrease as you approach North Dakota. Also California and Newyork also have considerable amount of loans that are completed

Income range is vital key to predict cashflow, For most loans that are completed or current income range of USD25,000-49,999 and USD50,000 - 74,999 has the most number of loans the trend decrease as you approach to unemployed.

Occupation also show the trend with high number of completed and current loans are from professionals, skilled labor and executives. The trend decrease as you approach to student-College Freshman.


```
<function matplotlib.pyplot.show(close=None, block=None)>
```

Property ownership and Loan status

Loan status and Income Range

Loan status and Borrower State

Loan status and Occupation