Lending Club Case Study
EDA Analysis

Business Objective

Lending Club company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

Approach

Data Cleaning : Dropped columns which has more than 40% null values. Also columns which are having same value all rows are also dropped.

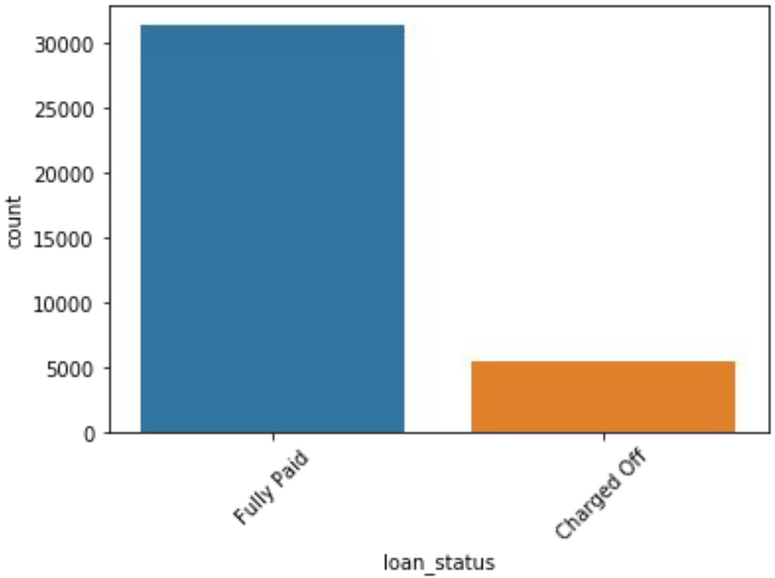
Treat missing values : Either use mode or mean.

Create new columns from existing : created new columns like buckets from existing columns.

Univariate analysis : data distribution single variable.

Bivariate analysis : analysis wrt loan status and other field.

Data distribution of loan status

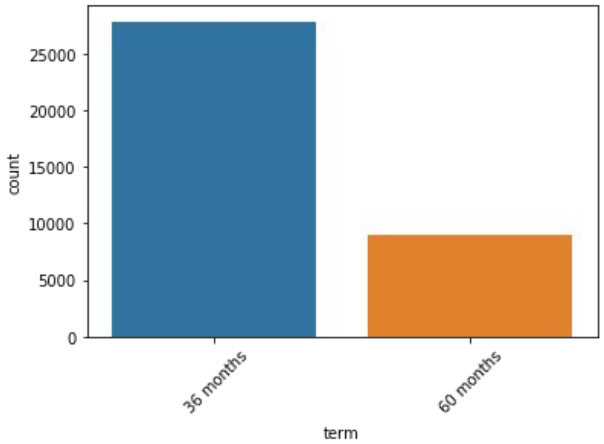


Observation:

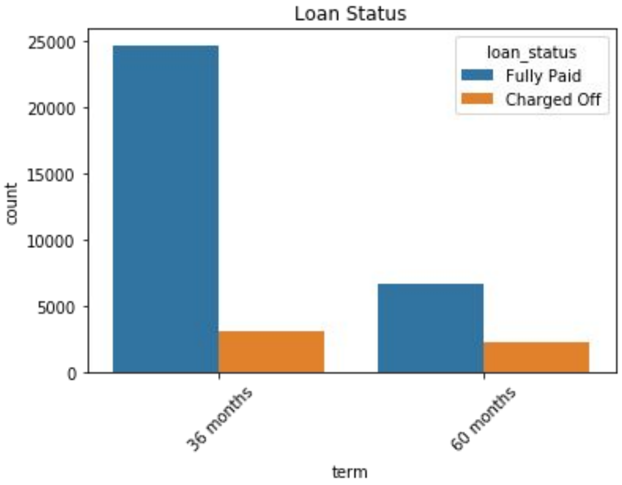
Fully paid applicants 85%

Defaulters 15%

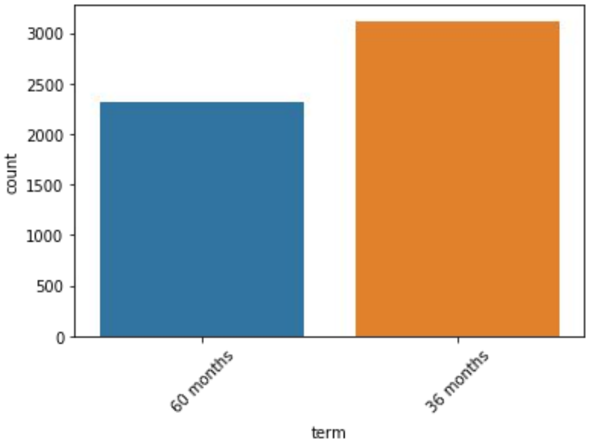
Tenure Distribution in entire data



Tenure Distribution Vs loan status



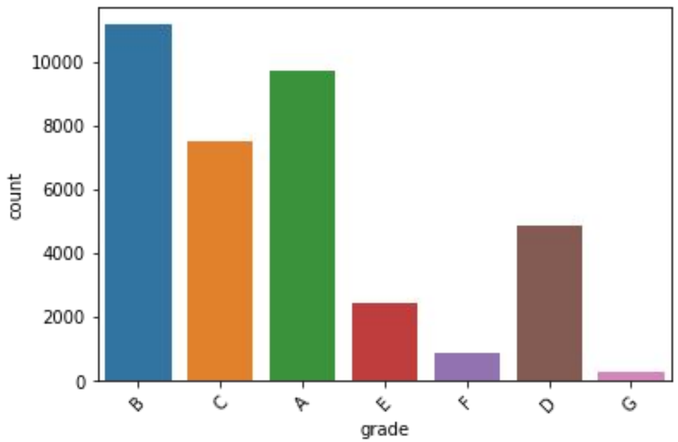
Tenure Distribution in defaulters



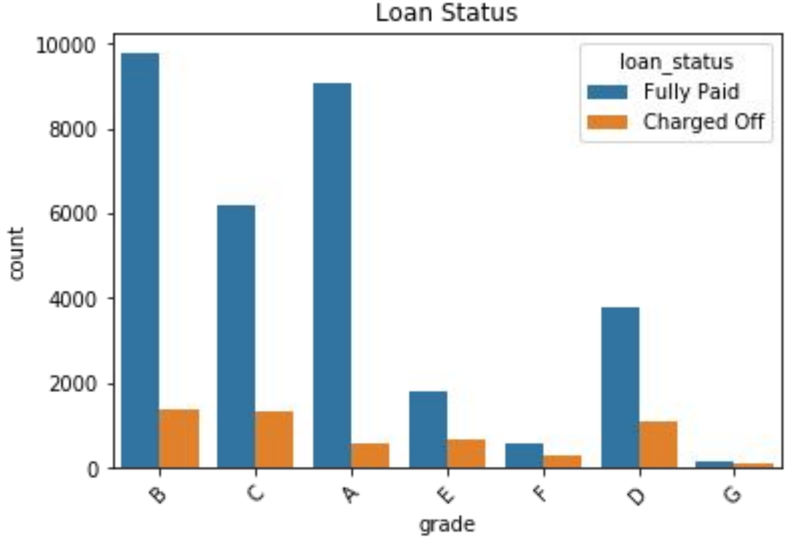
Observation:

More applicants are choosing for less tenure.
Less tenure applicants like to be more defaulters compare to high tenure.
%wise 60 month tenure segment has more like to be defaulters

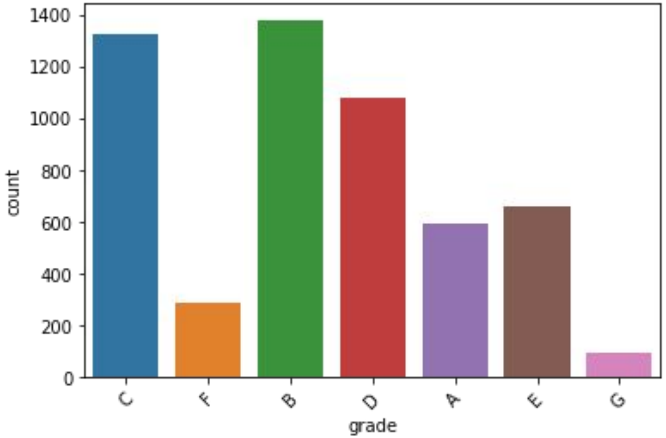
Loan grade distribution in entire data



Loan grade Distribution Vs loan status



Loan grade distribution in defaulters

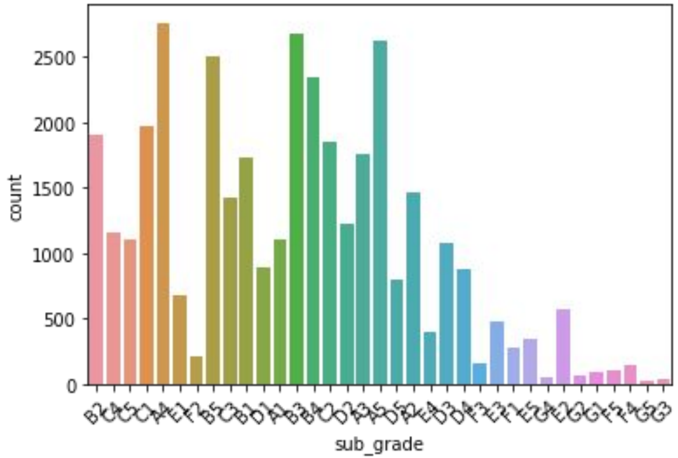


Observation:

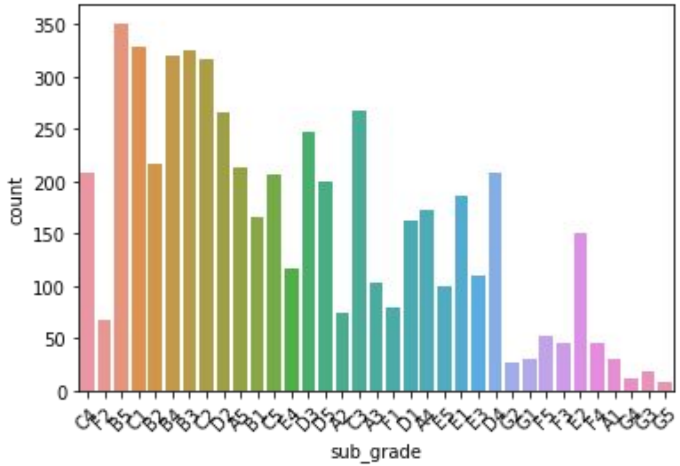
Loan grade B and C are more like to be defaulters compare to other grades

%wise E,F grade loans are more like to defaulters

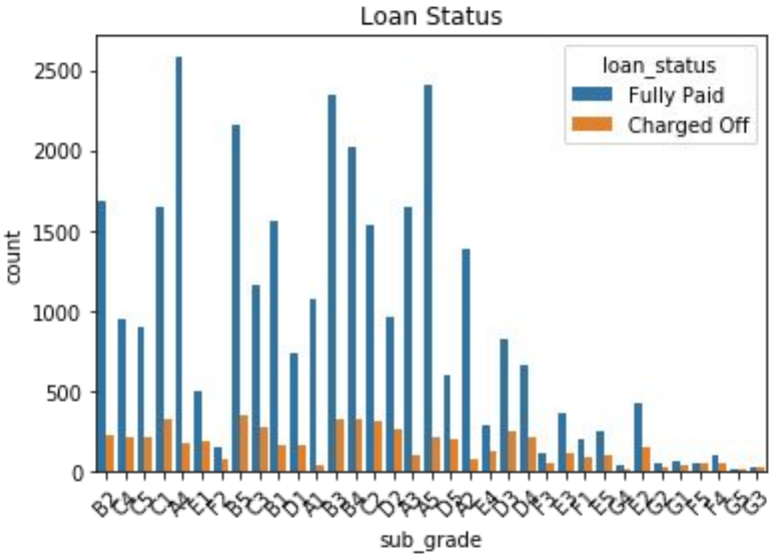
Loan grade distribution in entire data



Loan grade distribution in defaulters



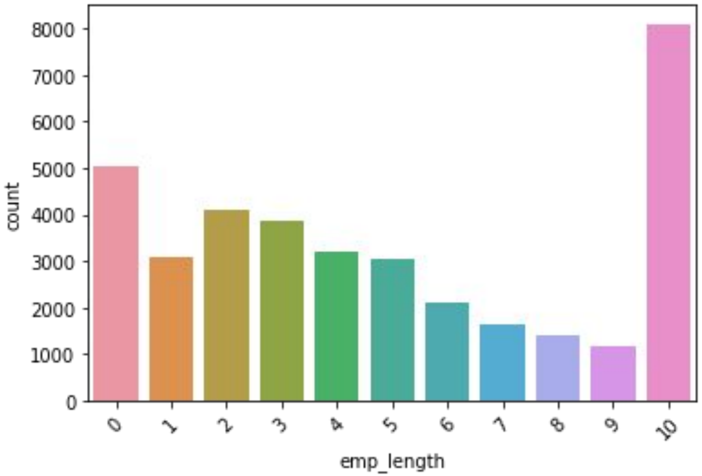
Loan grade Distribution Vs loan status



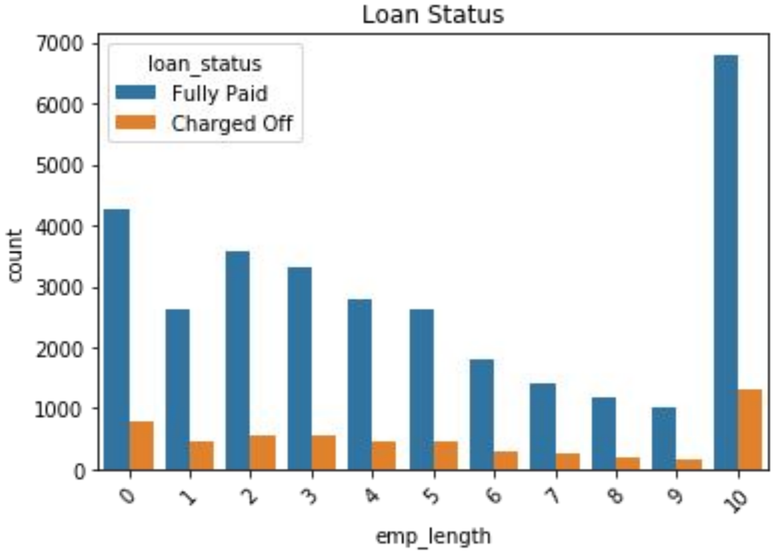
Observation:

In Loan subgrade B3, B5, B5, C1, C2 are more like to be defaulters compare to other subgrades

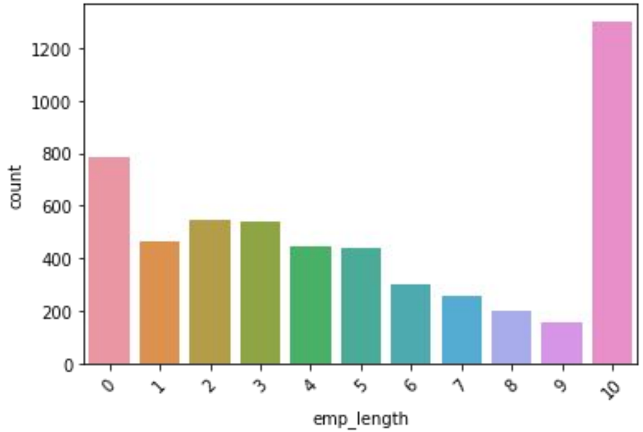
Experience distribution in entire data



Experience Vs loan status distribution



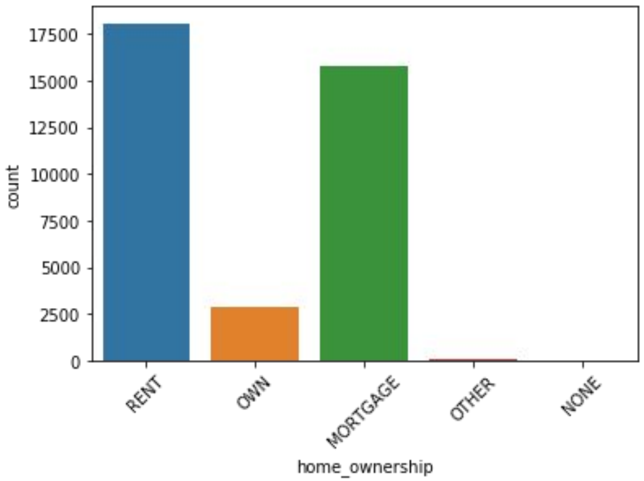
Experience distribution in defaulters



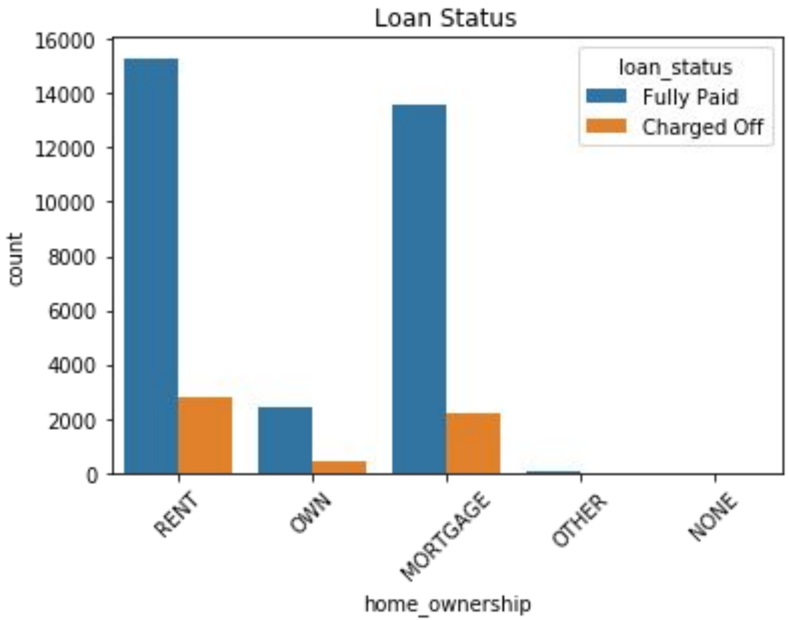
Observation:

Less than 1 year and more than 10 experience are more defaulters.

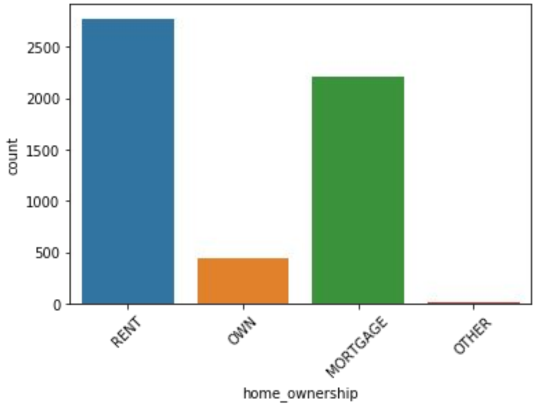
House ownership distribution in entire data



House ownership Vs loan status distribution



House ownership distribution in defaulters

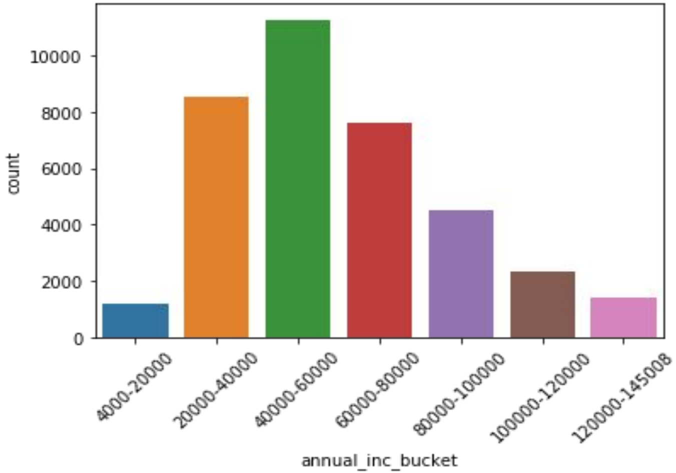


Observation:

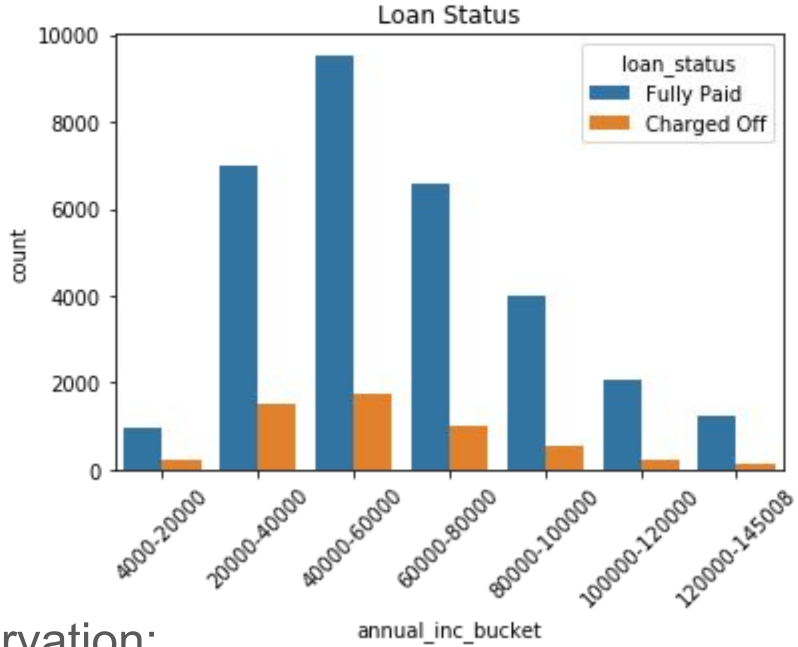
Rented, mortgage applicants are more defaulters compare to own house.

%wise own house loan applicants are like to defaulters

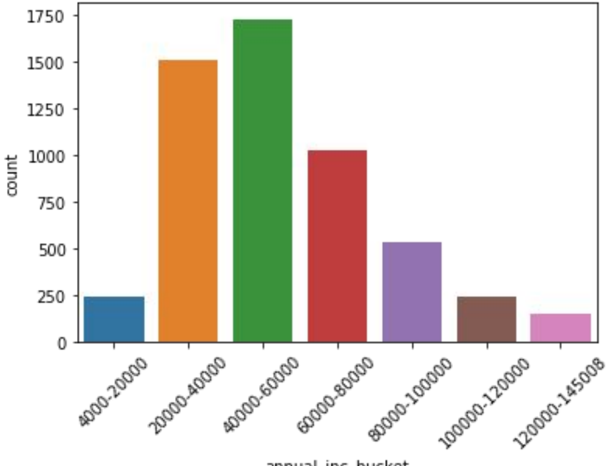
Annual income distribution in entire data



Annual income Vs loan status distribution



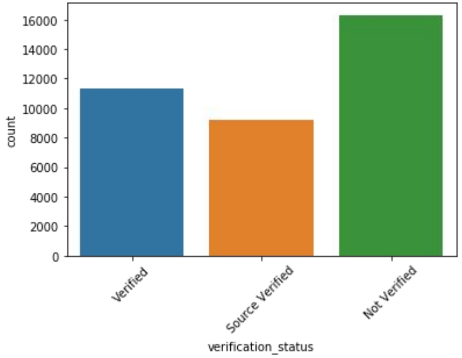
Annual income distribution in defaulters



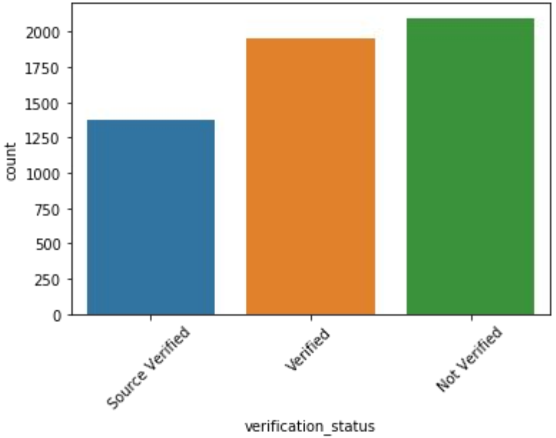
Observation:

% wise, In Annual income 4k-20k segment applicants are more defaulters.

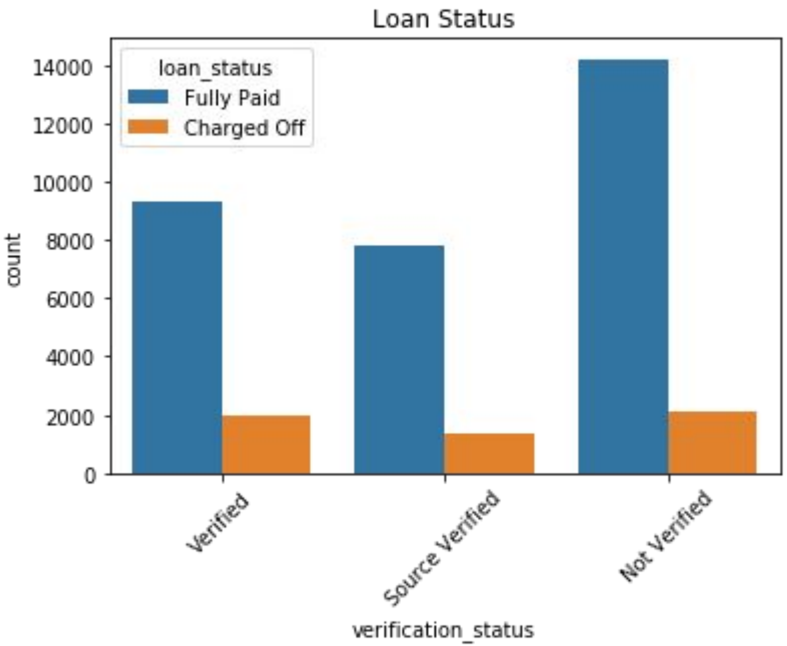
Loan verification distribution in entire data



Loan verification distribution in defaulters



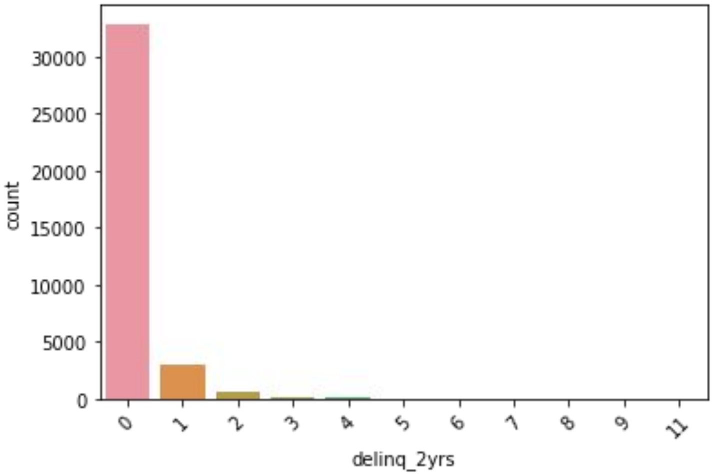
Loan verification Vs loan status distribution



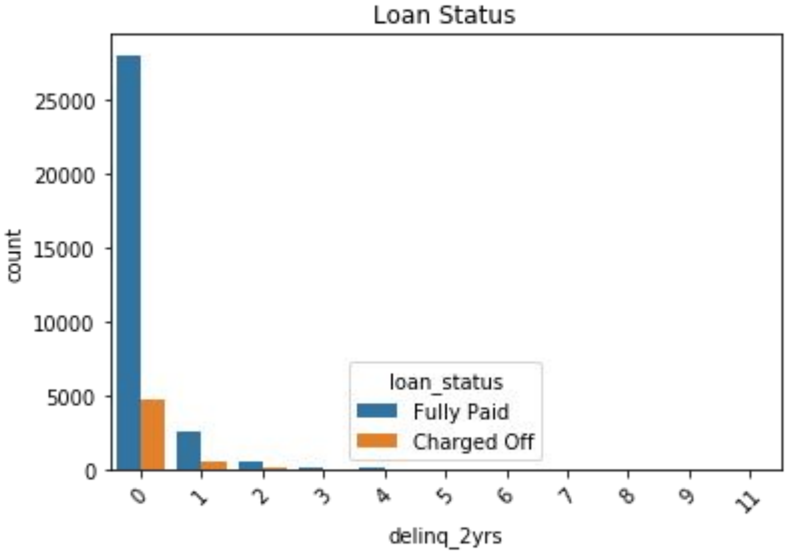
Observation:

Non verified applicants are more likely to be defaulters.

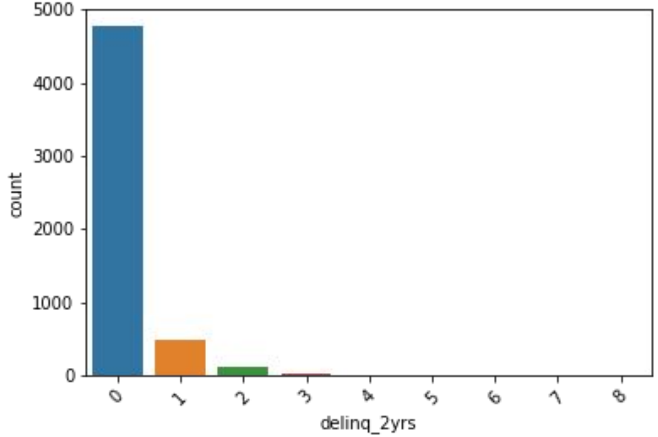
30+ days past-due distribution in entire data



30+ days past-due Vs loan status distribution



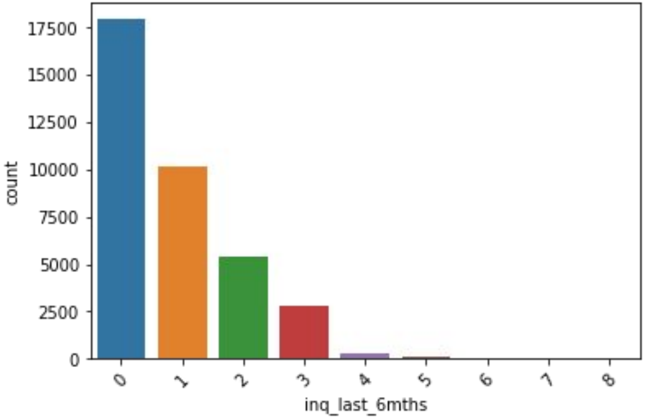
30+ days past-due distribution in defaulters



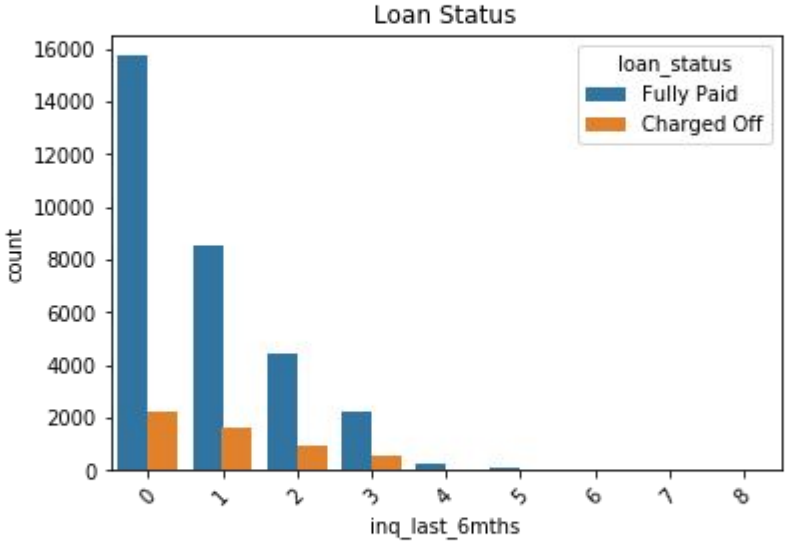
Observation:

More %of defaulters in past-due applicants.

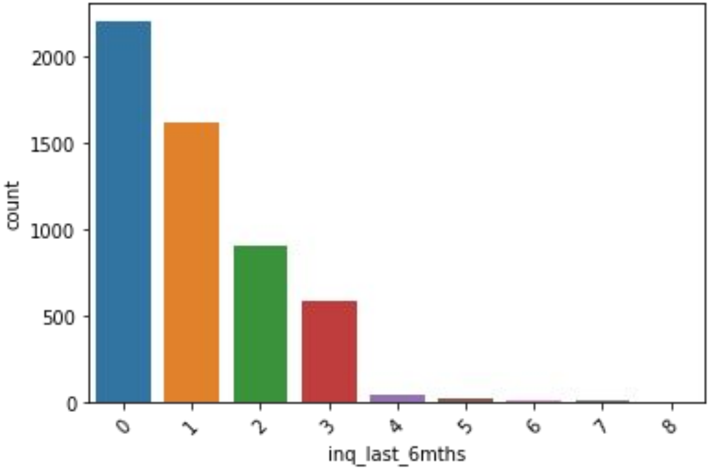
Credit history inquiry distribution in entire data



Credit history inquiry Vs loan status distribution



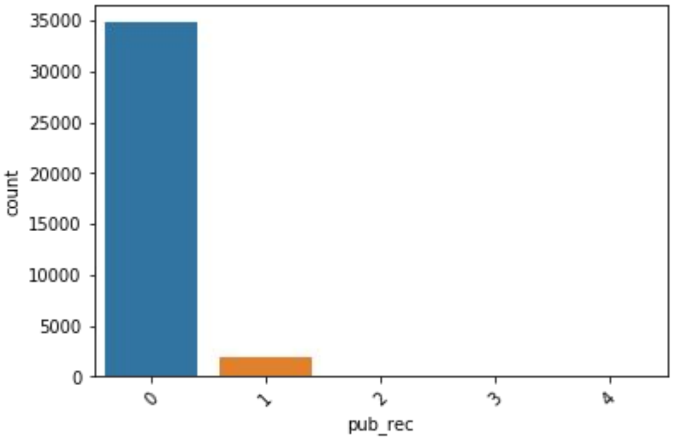
Credit history inquiry distribution in defaulters



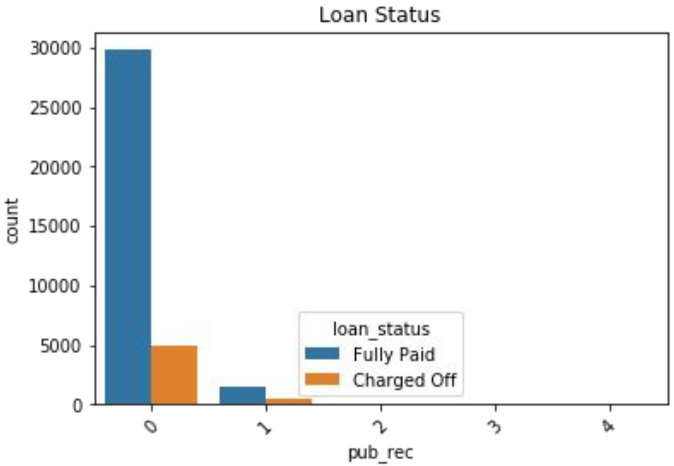
Observation:

More % of applicants are defaulters when more credit history inquiry happened compare to 0 inquiries.

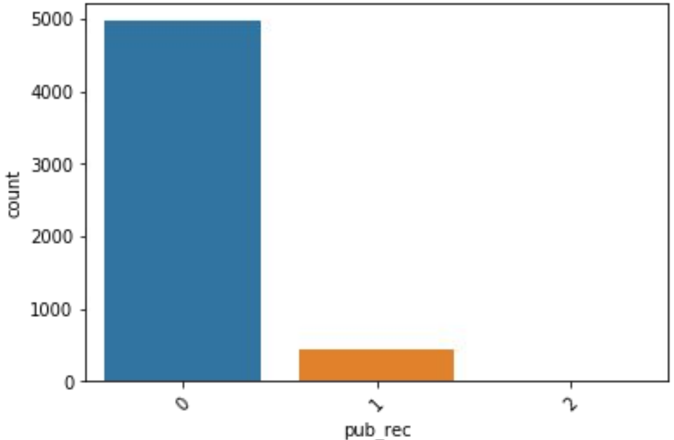
derogatory public records distribution in entire data



derogatory public records Vs loan status distribution



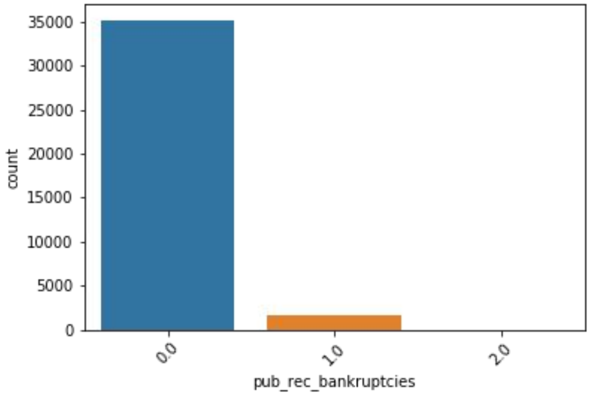
derogatory public records distribution in defaulters



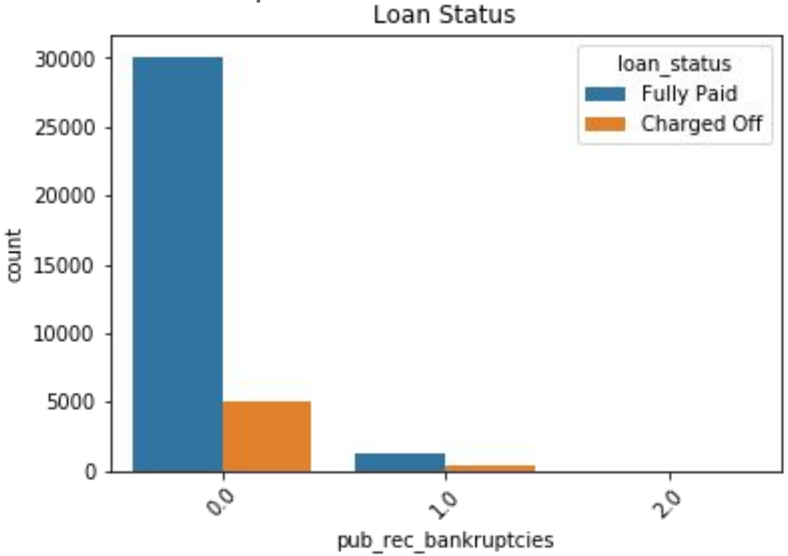
Observation:

Out of applicants who has derogatory public records in that applicants % of defaulters are more.

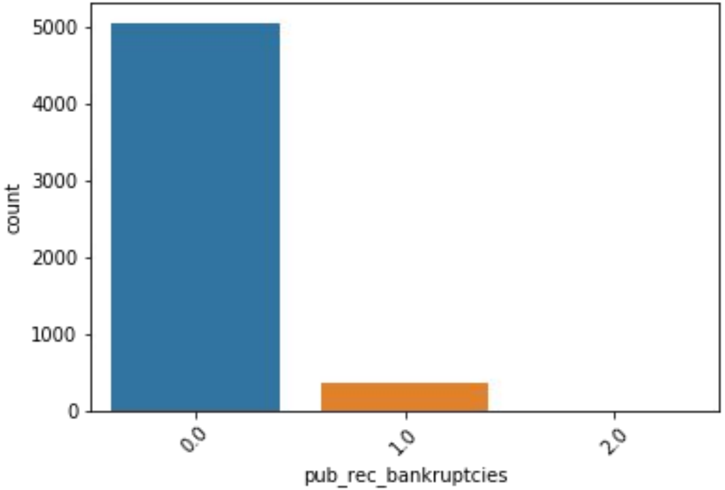
public record bankruptcies distribution in entire data



public record bankruptcies Vs loan status distribution



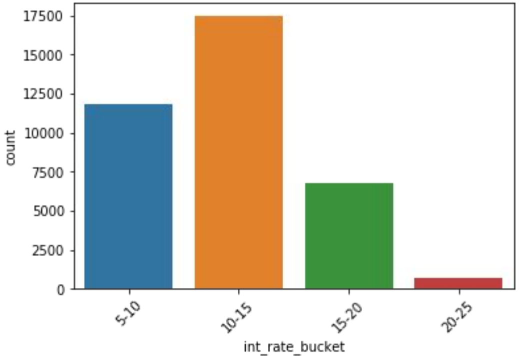
public record bankruptcies distribution in defaulters



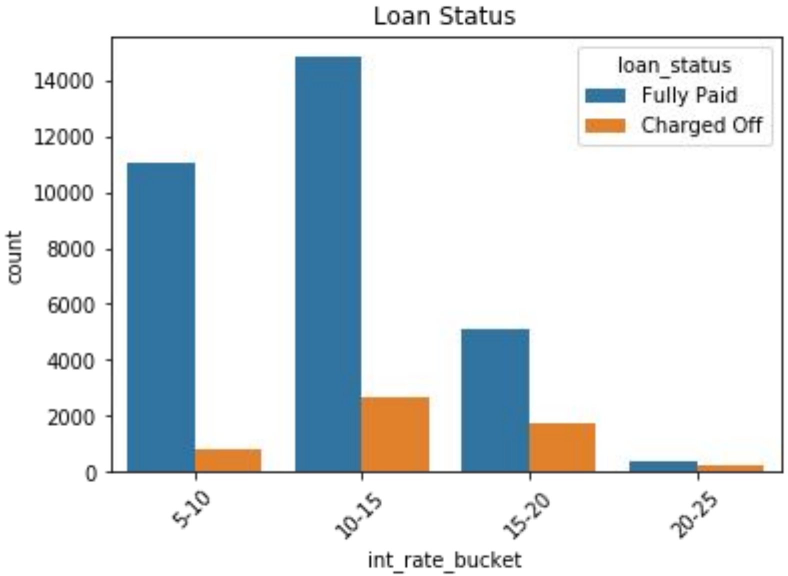
Observation:

Out of applicants who has public record bankruptcies in that applicants, % of defaulters are more.

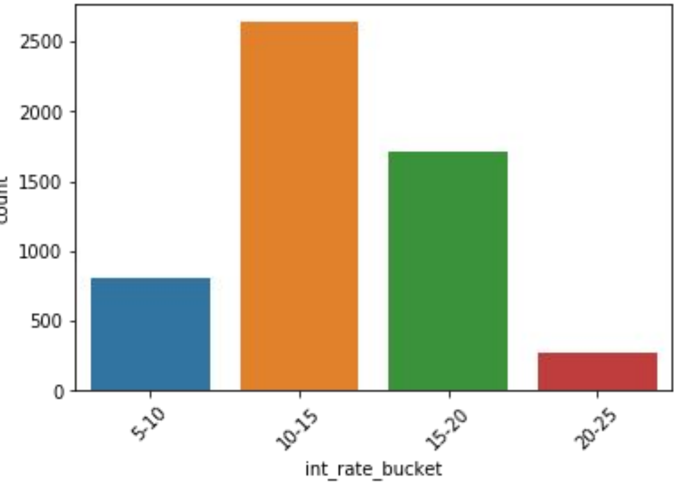
Interest rate distribution in entire data



Interest rates Vs loan status distribution



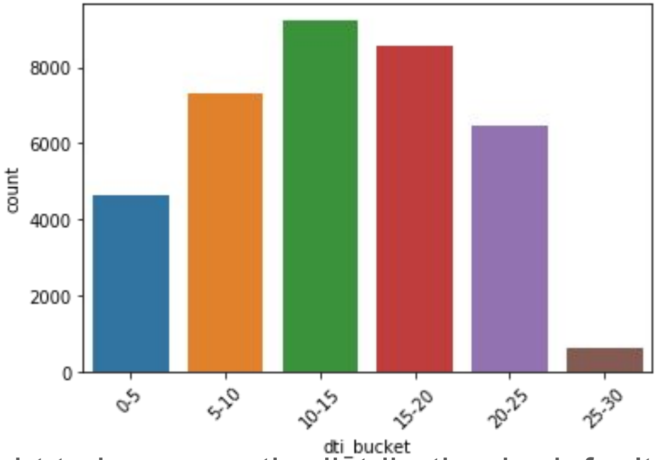
Interest rates distribution in defaulters



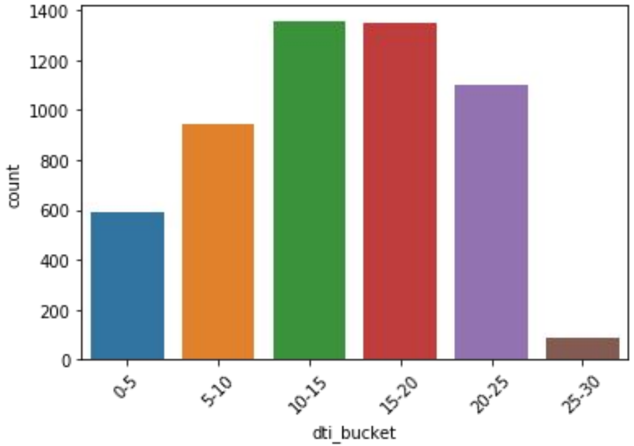
Observation:

Out of all high interest rates loan approvals, like to be more % defaulters.
In 15-20 % interest segment around 40% defaulters out of applicants in that segments.

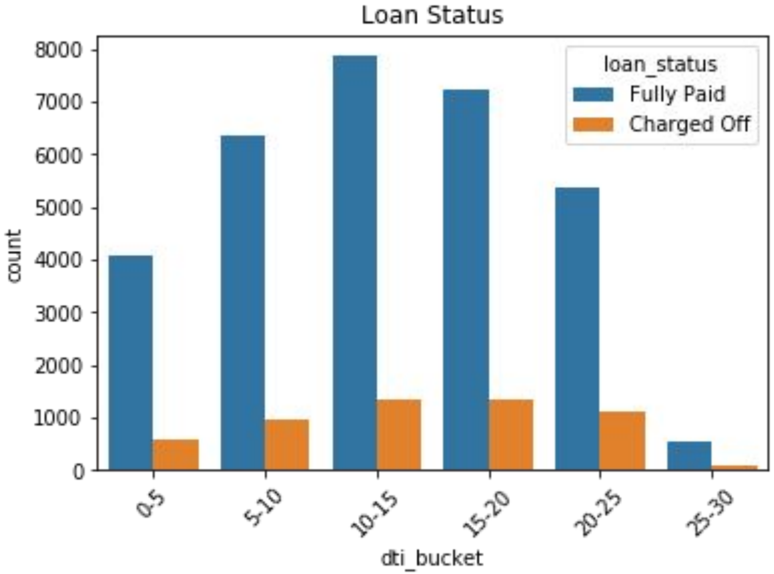
Debt to income ratio distribution in entire data



Debt to income ratio distribution in defaulters



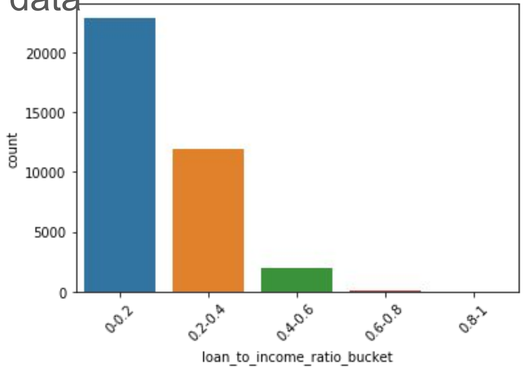
Debt to income ratio Vs loan status distribution



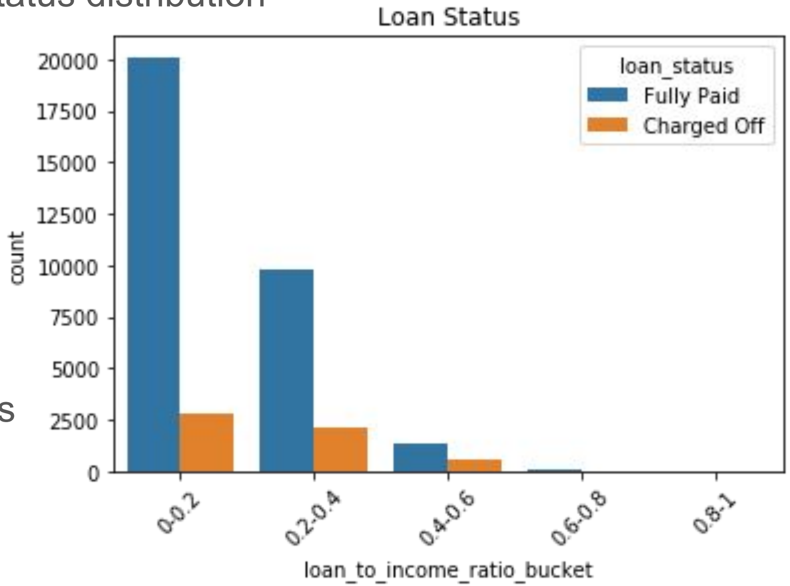
Observation:

In 25-30 ratio segment, like to be more % defaulters.

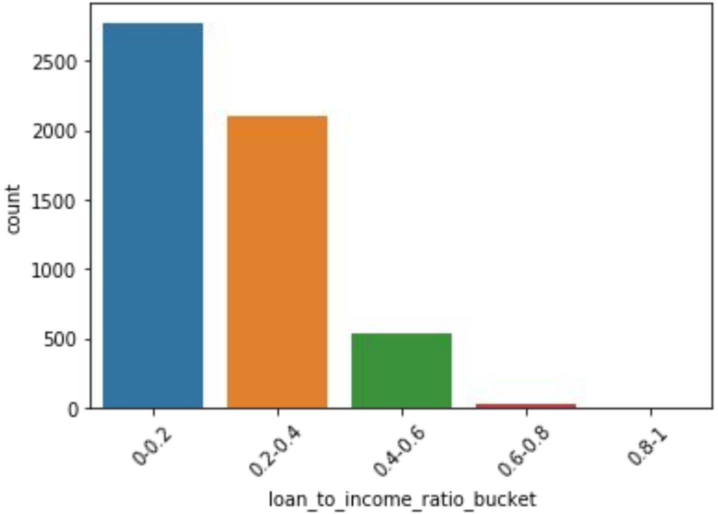
Loan amount to annual income ratio distribution in entire data



Loan amount to annual income ratio ratio Vs loan status distribution



Loan amount to annual income ratio distribution in defaulters



Observation:

In 0.4-0.6 ratio segment, like to be more % defaulters.

Summary

Below parameters customers can be double checked before issuing loan to reduce risk.

1. Less than 1 year
2. Annual income 4k- 20k segment
3. Non verified
4. Debt to income ratio above 25
5. Loan amount to annual income ratio above 0.4
6. derogatory public records

Thank you