Problem statement

Financial institutions, including banks and credit unions, provide loans to individuals and businesses with the expectation of timely repayment. However, loan default, where borrowers fail to meet the repayment obligations, poses a significant risk to these institutions. High default rates can lead to substantial financial losses.

Business Objectives

When borrower refuses to pay or runs away with the money owed causes credit loss to company. Main goal is to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such loan defaulters early in the loan lifecycle to mitigate risks and reduce financial losses using EDA is the aim of this case study.

Assumptions

There are three types of loan status in data set.

- Current
- Charge off
- Fully paid

Note - Assuming the current loan status will not impact the analysis

Approach

We have followed the below steps for EDA on loan application data set to find the pattern for defaulters

Data Cleaning

- Removed all the columns with null values.
- Remove the duplicate records
- Impute null with most occurring value in column
- Standardize the values of emp_length & interest rate.

Uni-Variant Analysis

- Numerical columns analysis

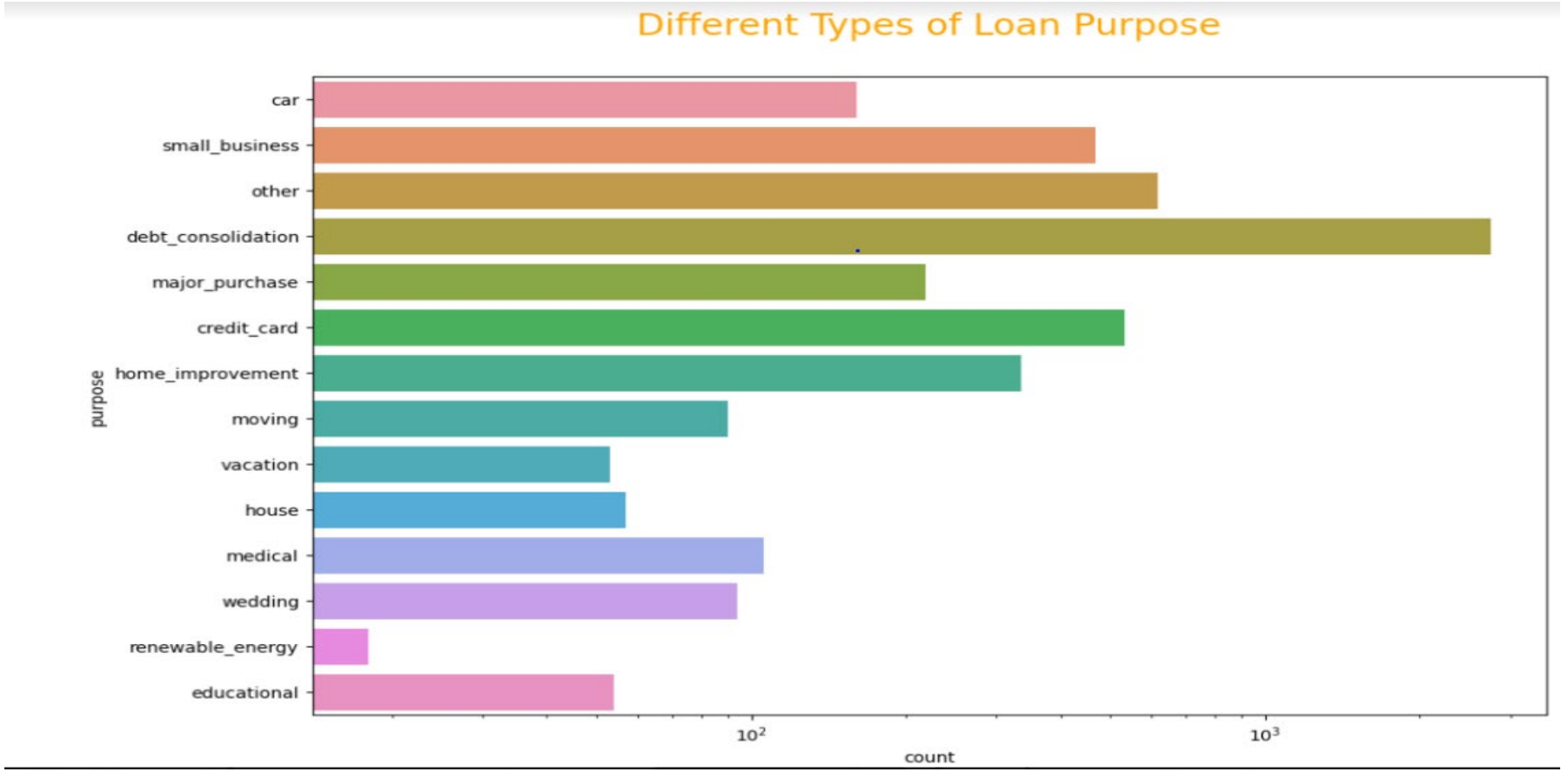
1. Loan amount & funded amount vary from 500 to 35000
2. Maximum borrowers have loan amount in the range 5000 to 110000
3. Maximum borrowers have the 10 years of working experience
4. Interest rate range is 5% to 24%

- Categorical column analysis

1. Grade have the values A,B,C,D,E,F,G
2. Home ownership have the values rent, mortgage, own,other,none

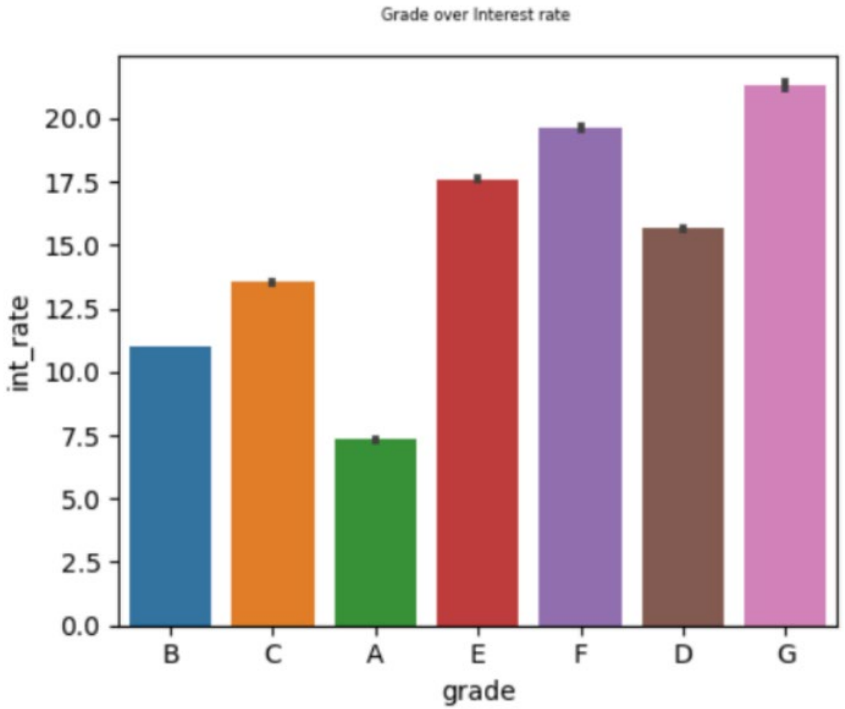
Visualization : Bi-Variant Analysis

Purpose:
Loans are taken mostly for debt consolidation followed by credit card payment.



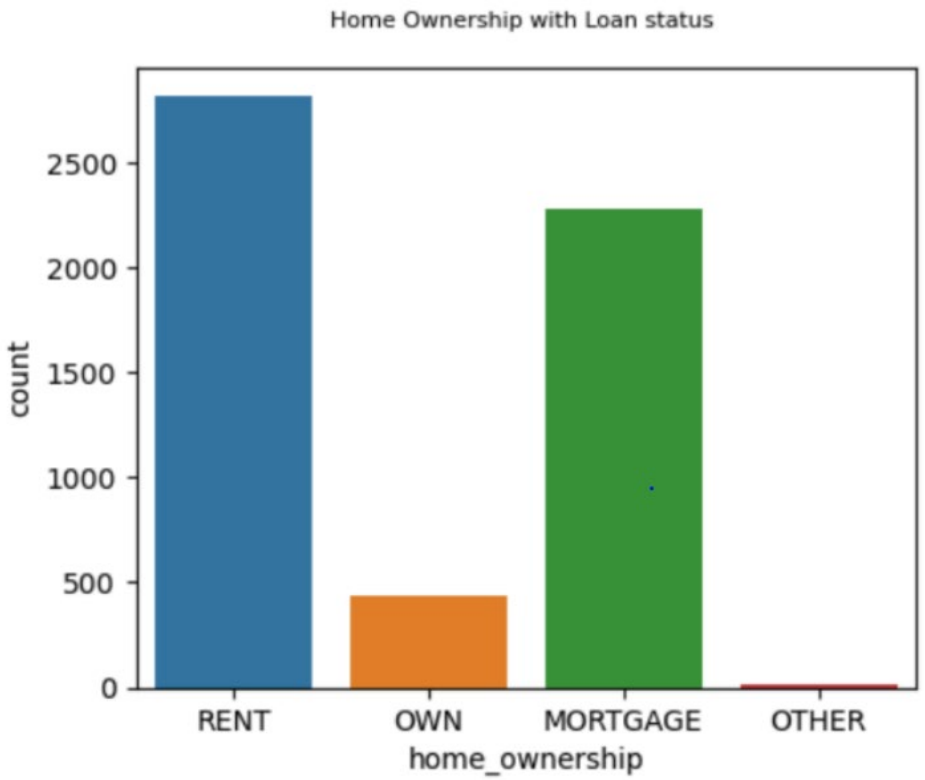
Grade

Grade/Interest Rate



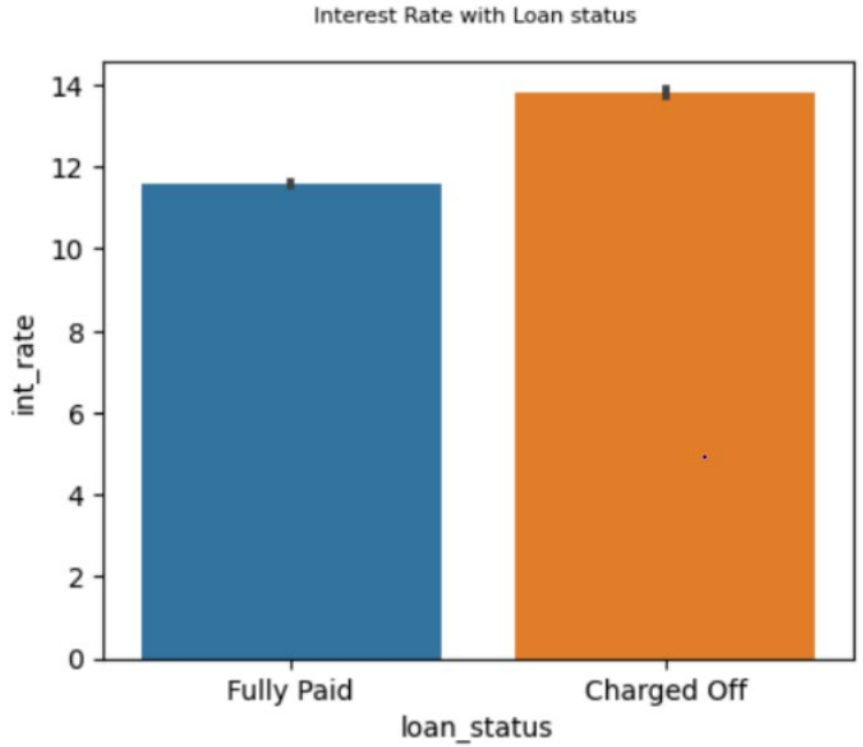
Interest rate increases with increase in Grade Order(A,B,C,D,E,F). Grade A has lowest interest rate while Grade G has highest interest rate.

Impact of Home Ownership on defaulters



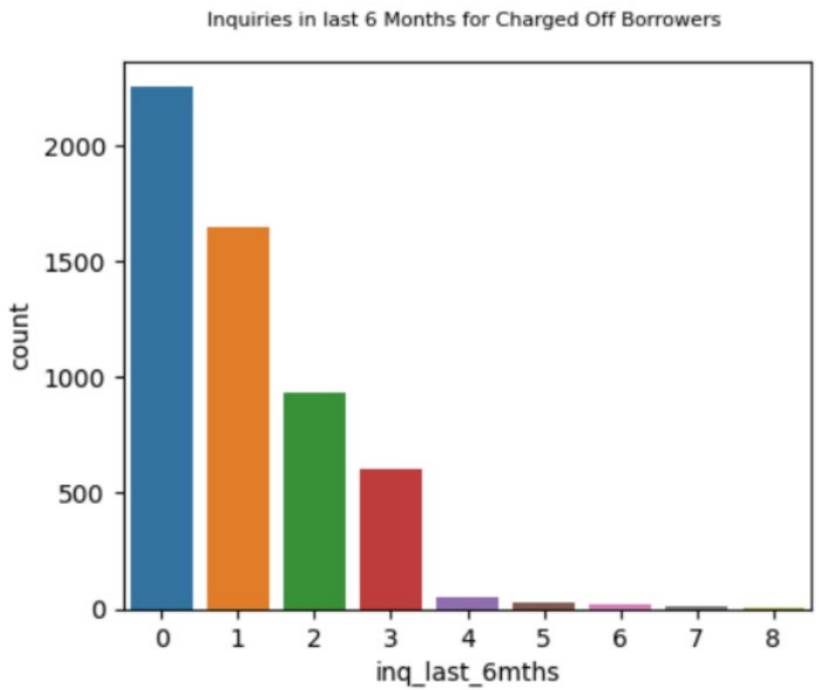
Default loan are more for applicants which don't have their own house and are rented or Mortgage.

Interest Rate with Loan status



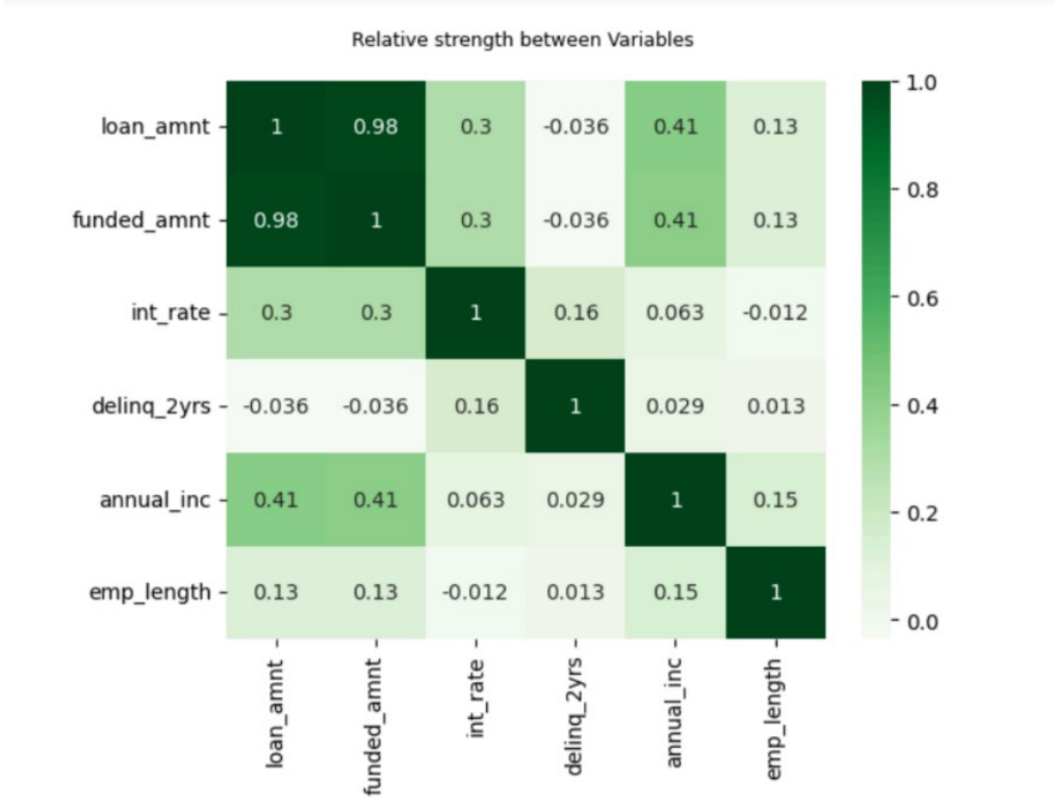
Majority of default customers have high interest rate.

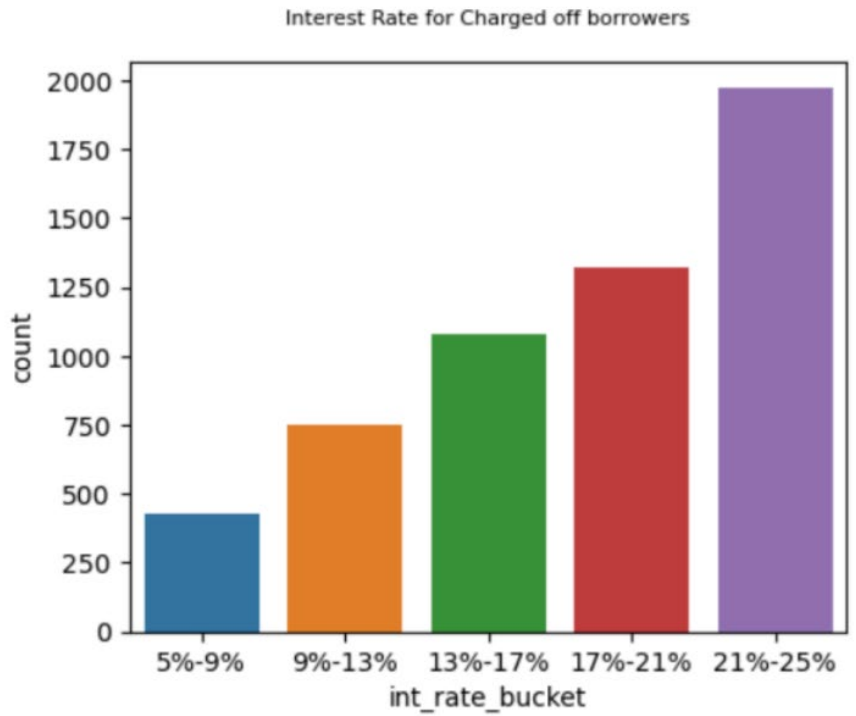
Last 6 Months Inquires



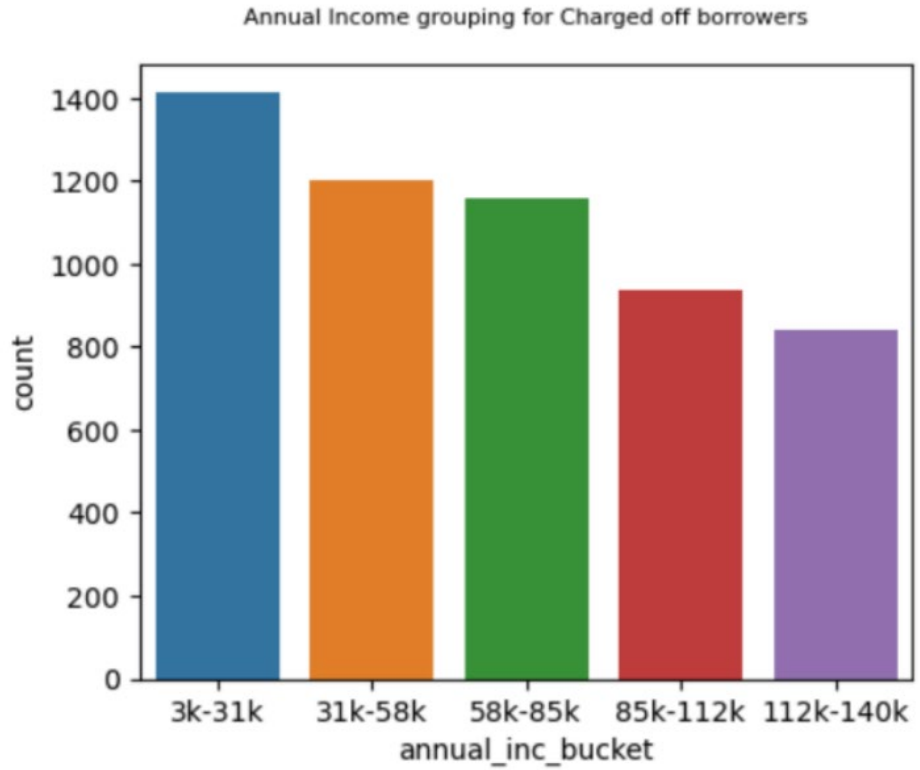
Majority of defaulters have 0 inquries in last 6 months

Relative co-relation between variables.



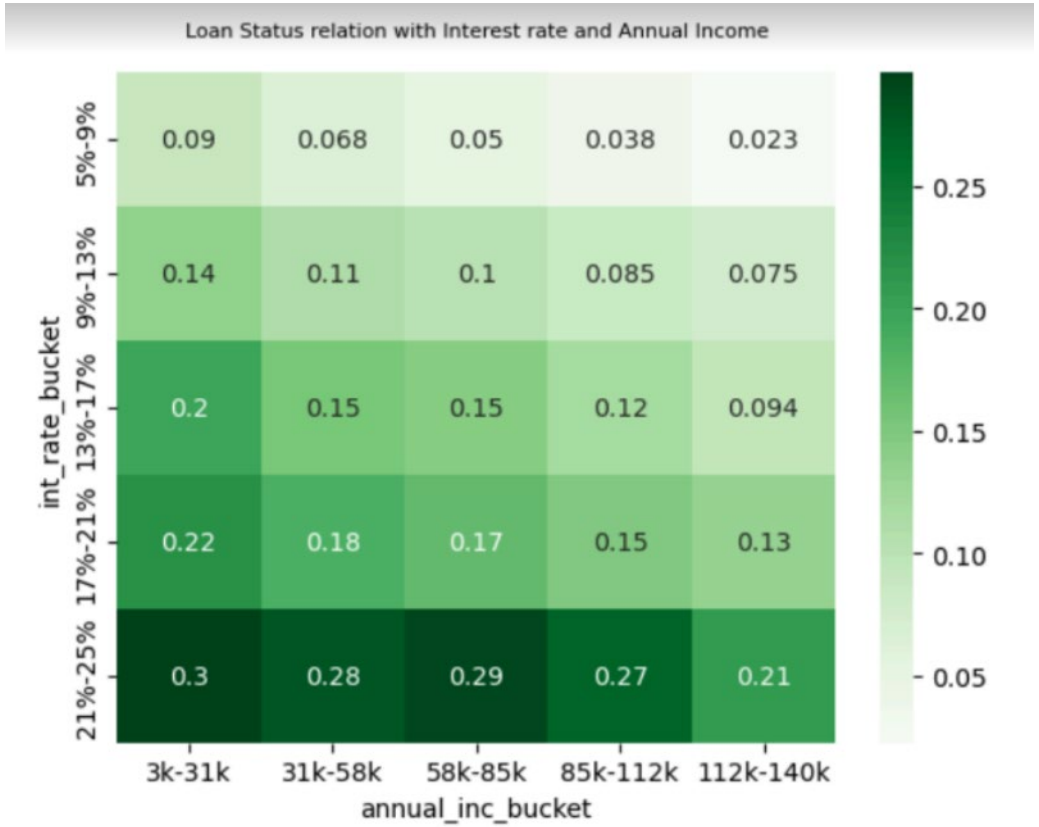


Majority of charged off customers are in high interest rate range 21% to 25%.

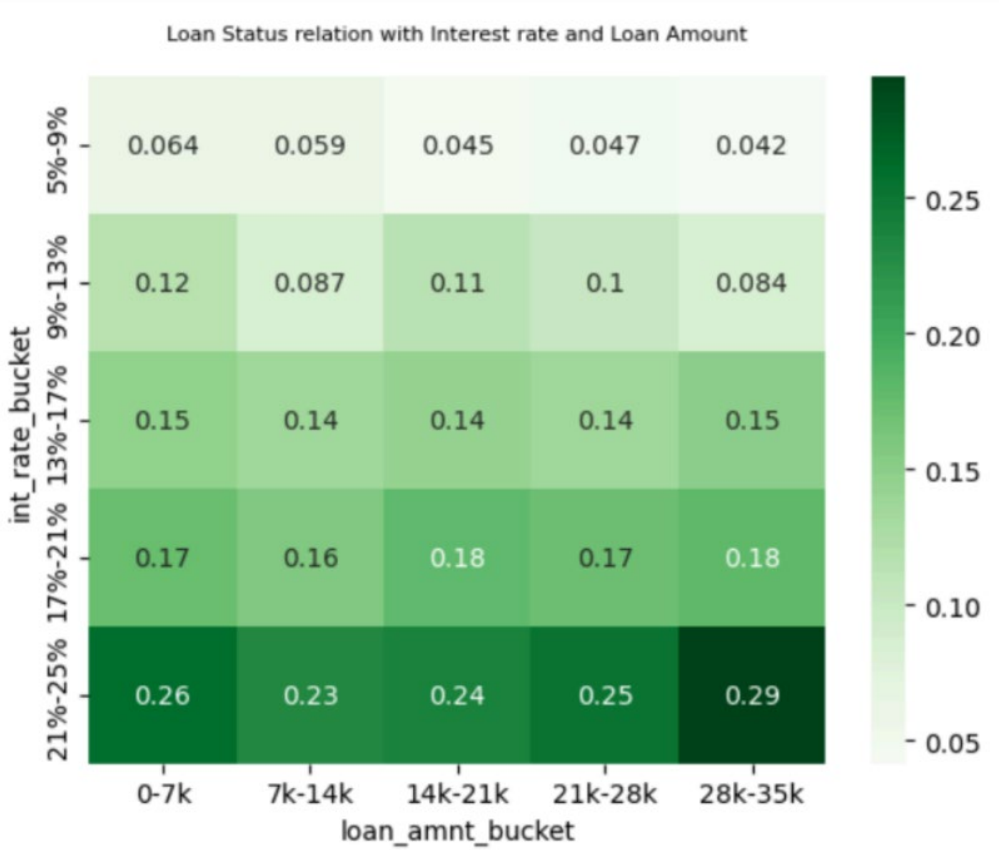


Most of defaulters have low annual income in range 3k to 31k

Visualization : Multivariate Analysis

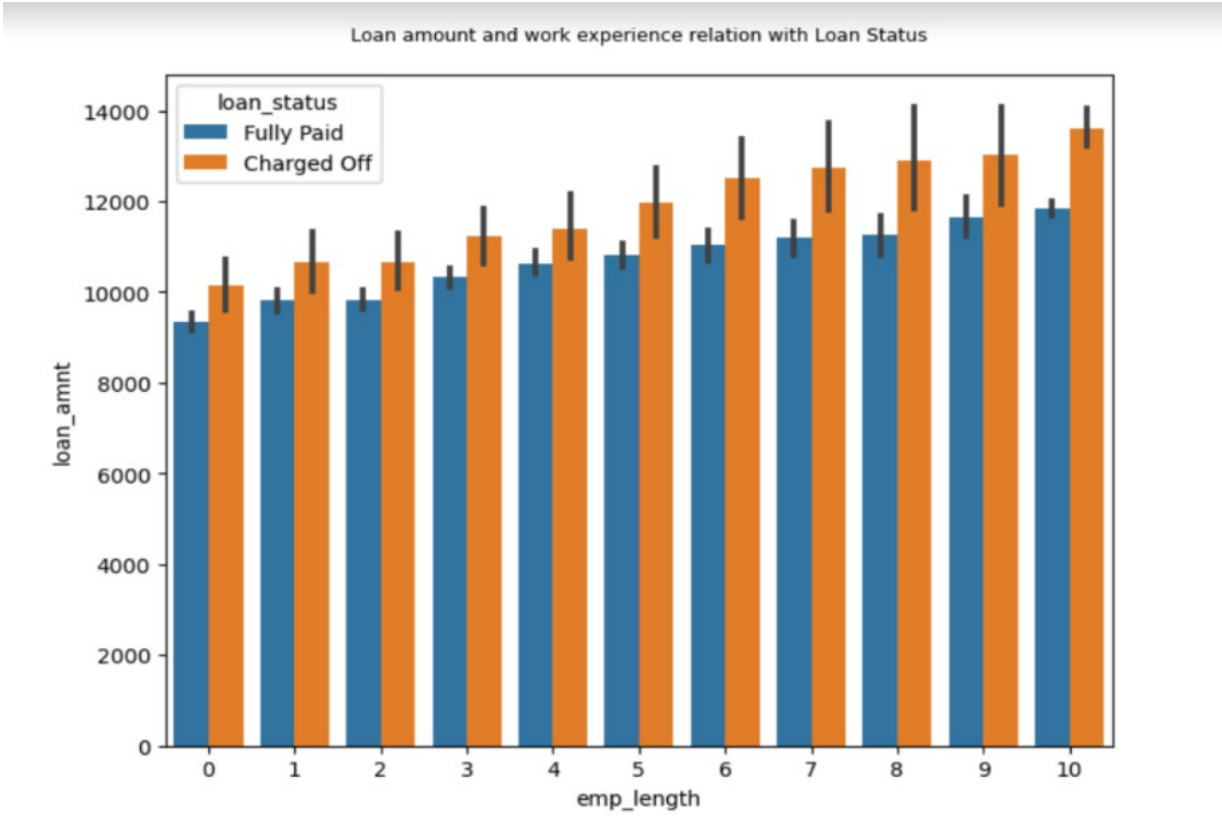


Applicants with high interest rate and low Annual income have high chance of defaulters.



Applicants with high loan amount and high interest rate have high chance of defaulters.

Applicants with annual income range 3k to 31k and loan amount 28k to 35k are more likey to deafulters.
Large amount of loan have high chance of default.



Loan amount slightly increases with increase in Employment length. Applicants with 10 years of employment length have high amount of loan and For defaulters it is more compare to fully paid.

Final Result

- Major Driving factors which can be used to find out or predict default customers and avoiding Credit Loss.

  - Grade
  - Annual Income
  - Verification Status

- Home Ownership, Interest rate ,Term, Delt 2 years also impact analysis of default customers.

- Consideration for default customers are:

  - Customers with low annual income(3k to 31k) and high interest rate(21% to 25%) with loan amount of 28k to
    35k
  - Customers with low Grade like F,G and high interest rate 20% to 24%.
  - Customers not having their own property and are rented or mortgage and having 10 years of working
  - experience.