LENDING CLUB CASE STUDY

UpGrad

ज्ञानमुत्तमम्

LENDING CLUB CASE STUDY

- Trapti Agarwal
- Kushalava Reddy Yanala

Table of Contents

1. Problem Statement
2. Data Cleaning and Manipulation
3. Univariate Analysis
4. Bivariate Analysis
5. Multi Variate Analysis
6. Recommendations

UpGrad

ज्ञानमुत्तमम्

Problem Statement

UpGrad

iit-b ज्ञानमुत्तमम्

Problem Statement:

Lending Club is a Consumer Finance Company, which is specialised in lending the various types of loan to customers. Lending the loan to ‘Risky’ applicants causes the financial loss to the lenders.

The objective of this case study is to identify the Risky applicants and helps in cutting down the amount of credit loss to Organization.

UpGrad

ज्ञानमुत्तमम्

Data Cleaning and Manipulation

UpGrad

iit-b
ज्ञानमुत्तमम्

2. Data Cleaning and Manipulation

Data Cleaning

As part of Fixing rows and columns:

- Columns with NA values were deleted (54 columns)
- Columns with only 1 value for all 39717 rows were deleted (9 columns)
- Text columns with 39717 values were deleted (4 columns)

As part of missing values:

- Columns with very % of values missing were deleted (3 columns)

As part of Standardizing values:

- term, int_rate and revol_util columns were standardized by removing “months” from the first columns and “%” from the remaining 2 columns

As part of Filtering rows and columns:

- Removed the outliers for salary (more than 99 percentile)

2. Data Cleaning and Manipulation

Data Transformation

As part of Data Typecasting:

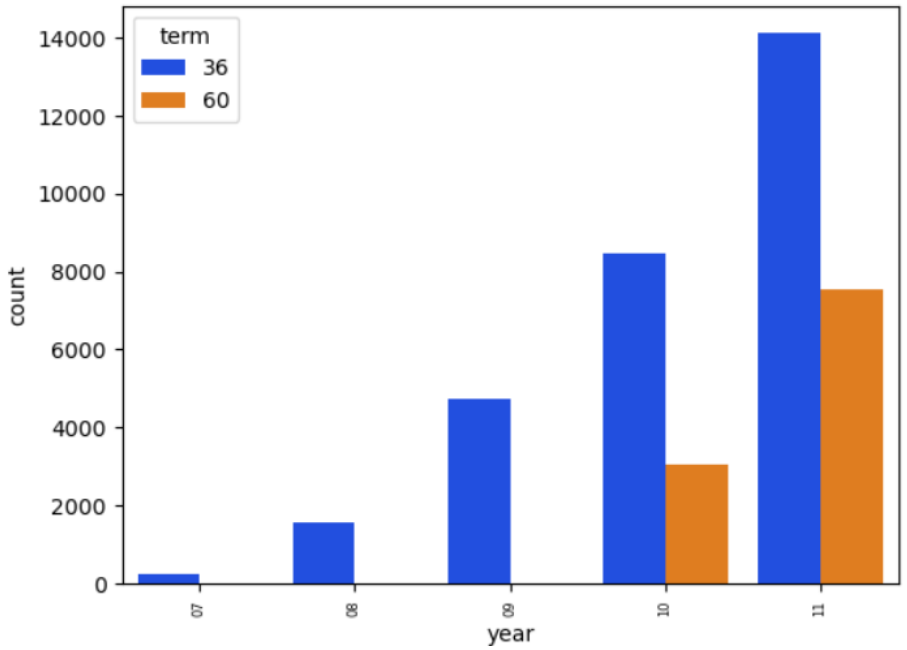
- Converted term column to int
- Converted int_rate and revol_util to float

As part of deriving new columns:

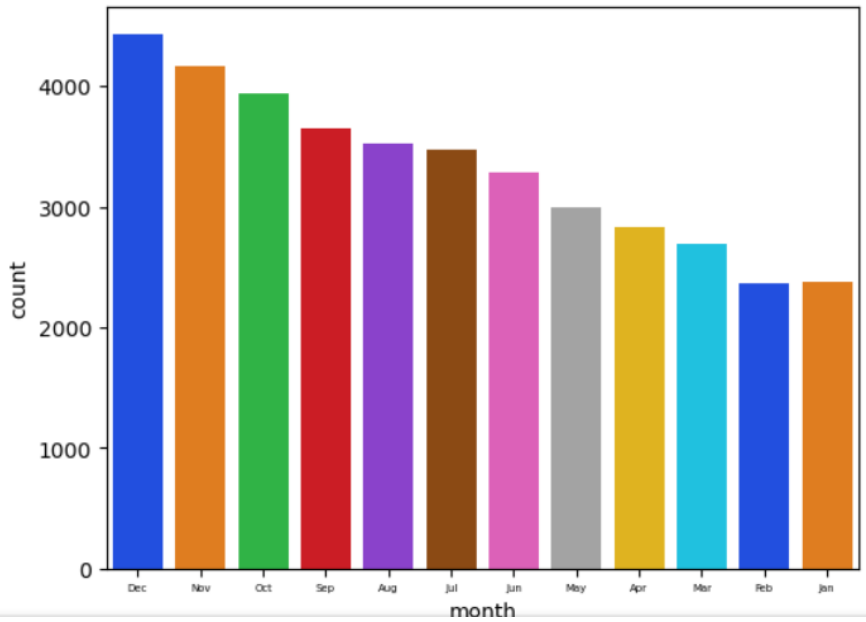
- Derived month and year column from issue_d column
- Derived charged off ratio for multiple parameters using the number of charged of loans/ total no. of loans in the category

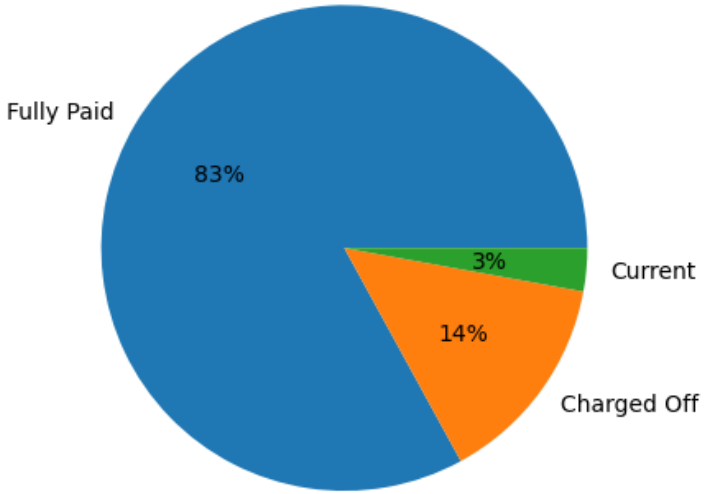
Univariate Analysis

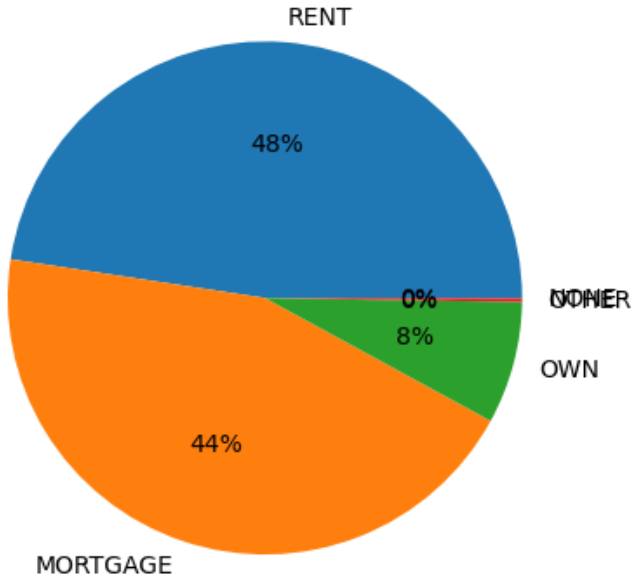
Countplot of Year & Term



Countplot of month



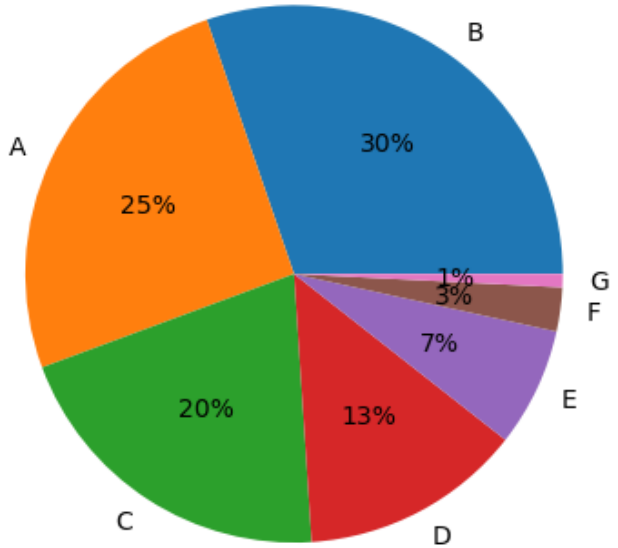


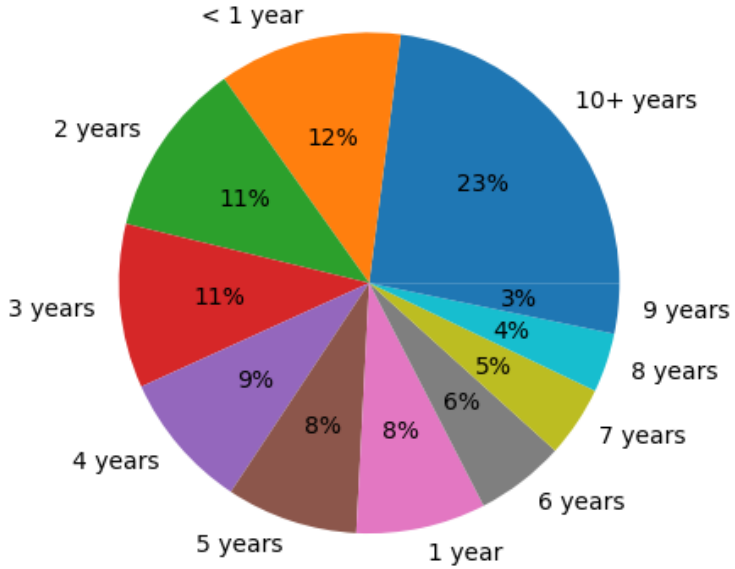


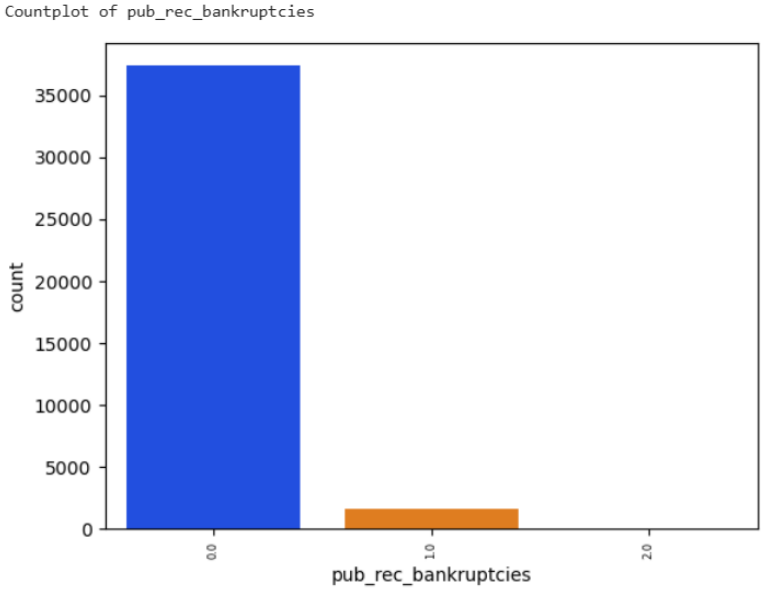
Observations:

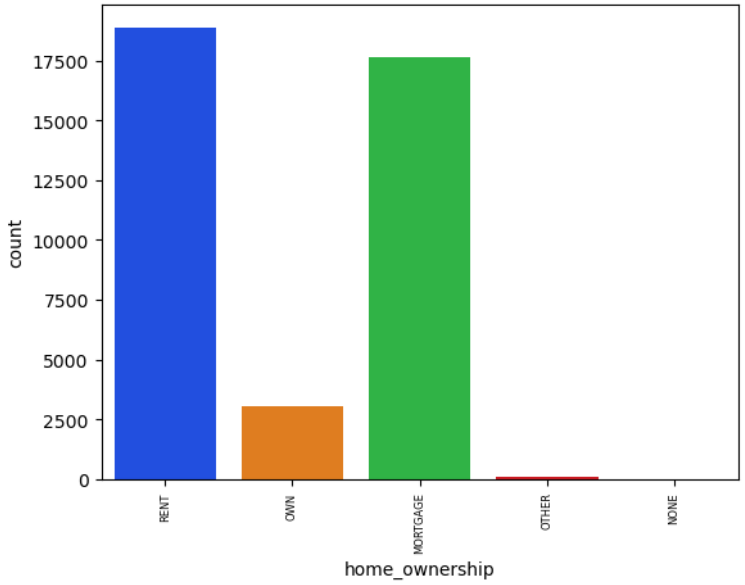
- Number of loans issued have increased every year from 2007 to 2011
- Most of the loans are given in the last quarter
- 83% of the loans are Fully paid customer
- 92% of the loan customers are either rented or mortgaged homes

Majority loans are given for grade A, B and C



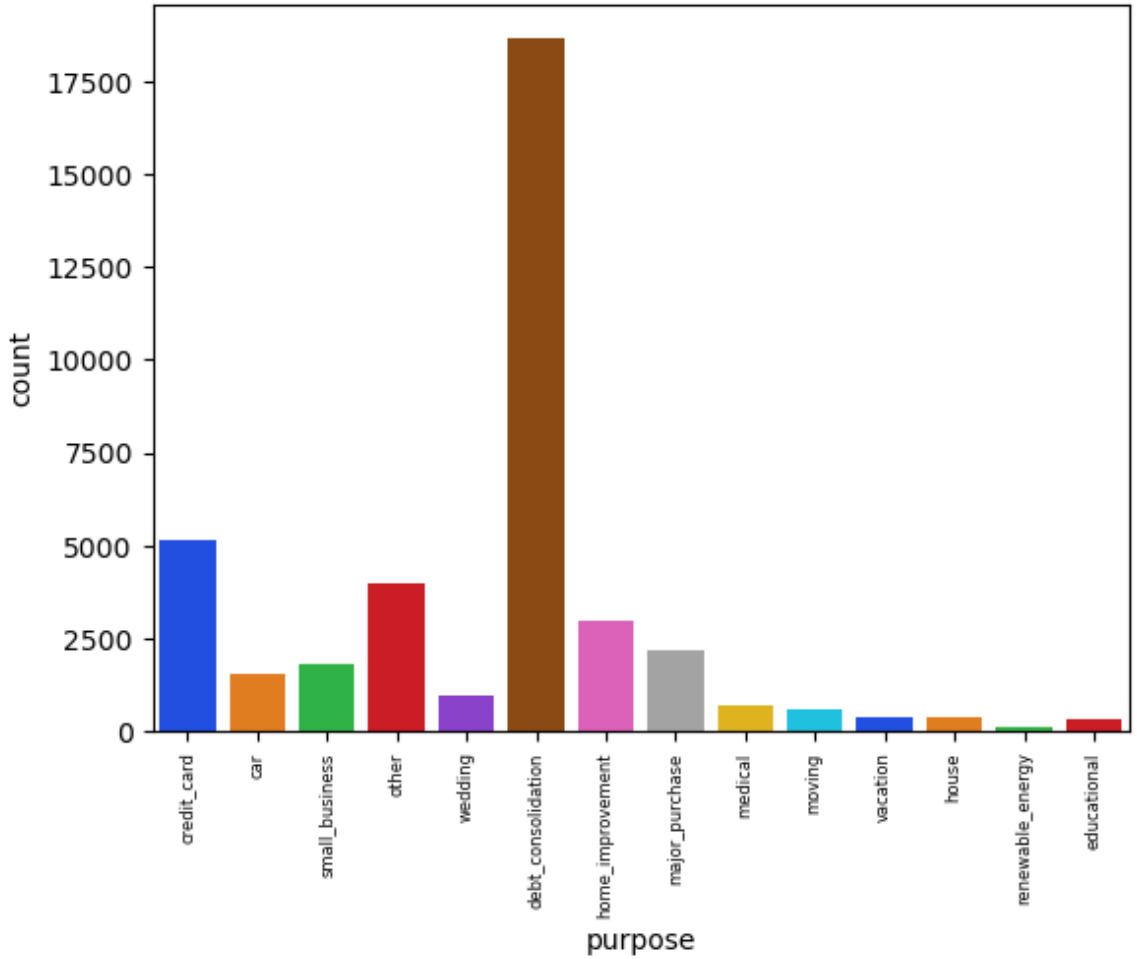




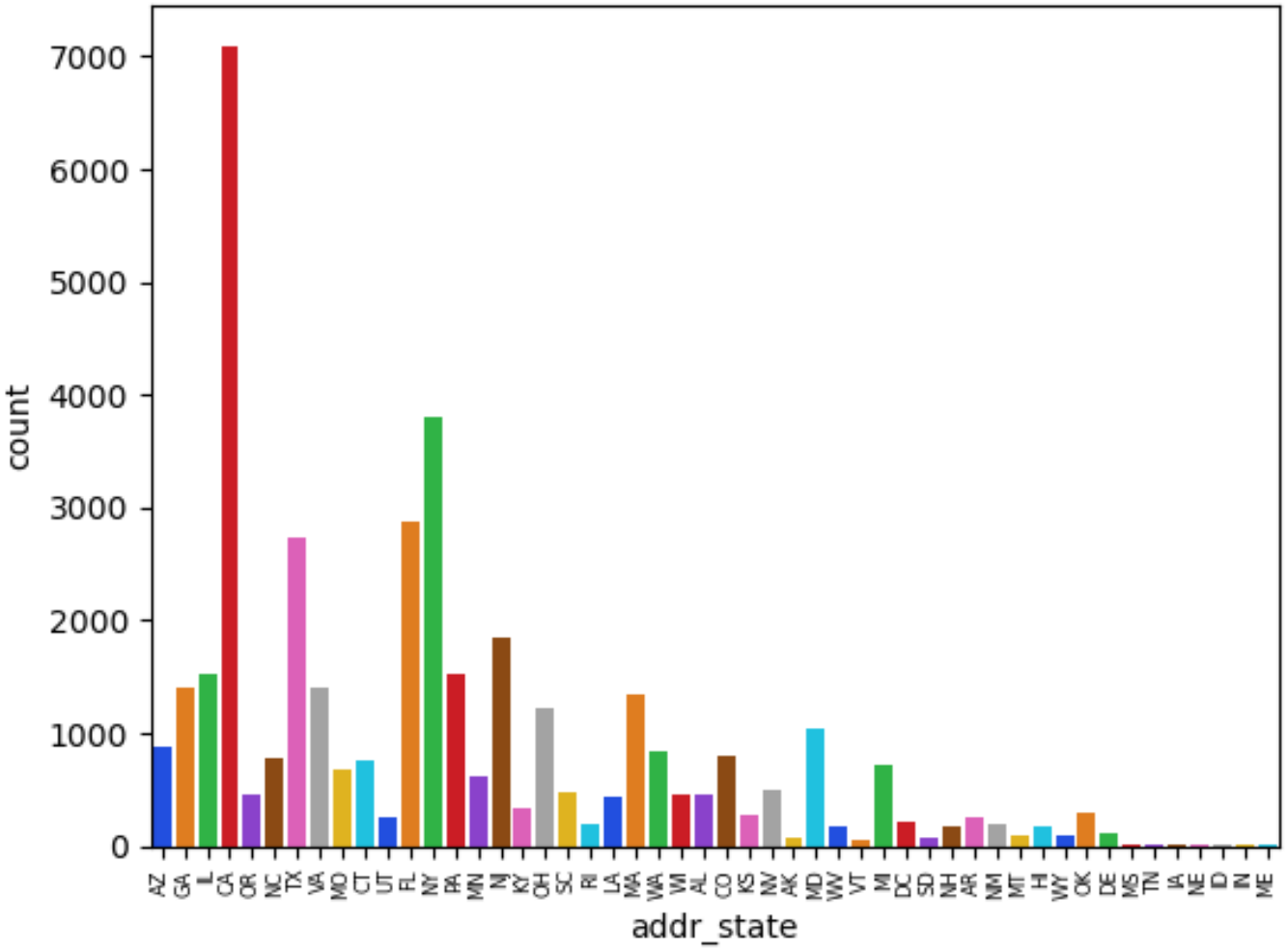


Observations:

- Majority loans are given for grade A, B and C
- More loans are given to customers with more than 10 years of exp
- More loans are given to customers with more than 10 years of exp
- More customers with rented and mortgaged homes are opting for loans



Majority of the customers give ‘debt_consolidation’ as the purpose for loan



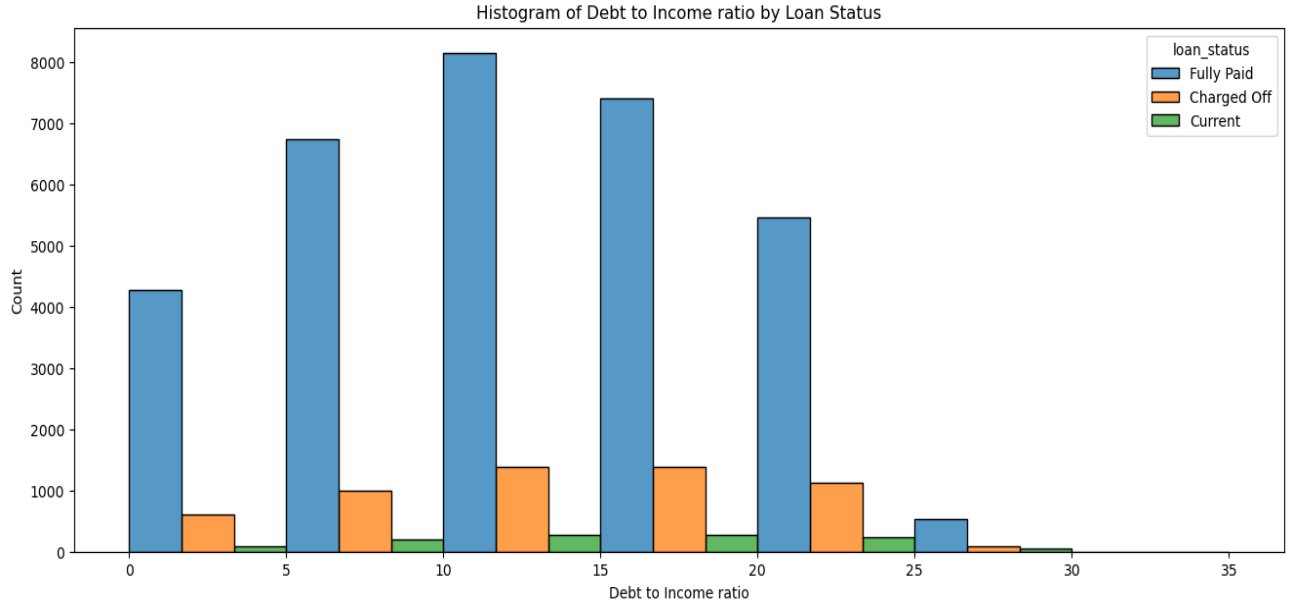
Highest number of customers are from CA and NY

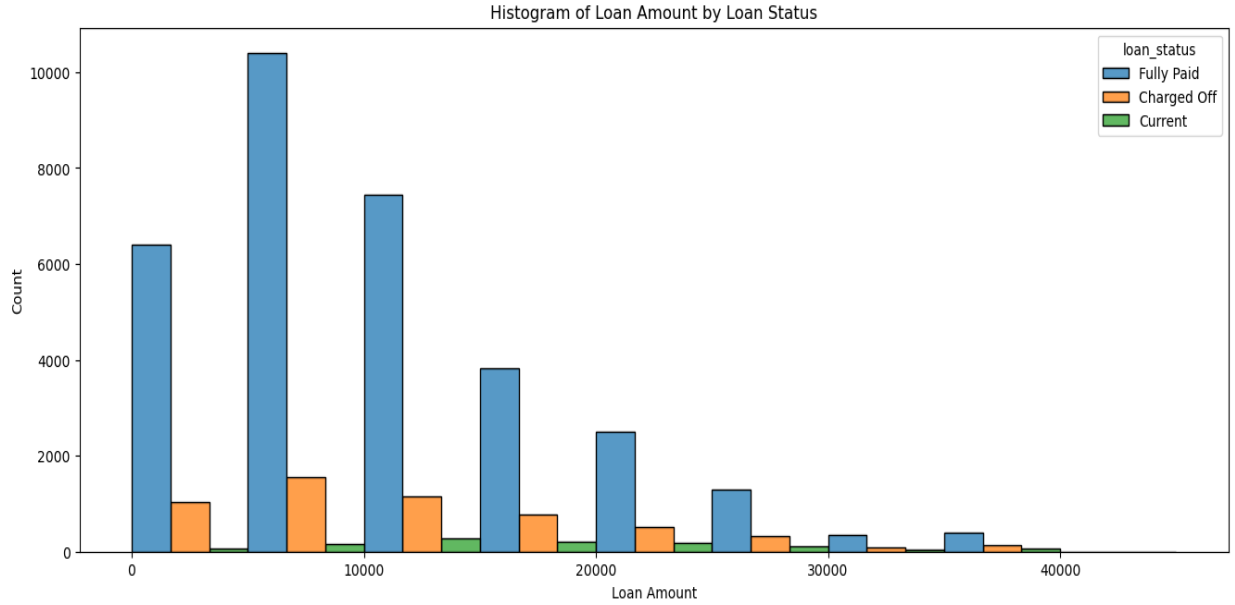
Bivariate Analysis

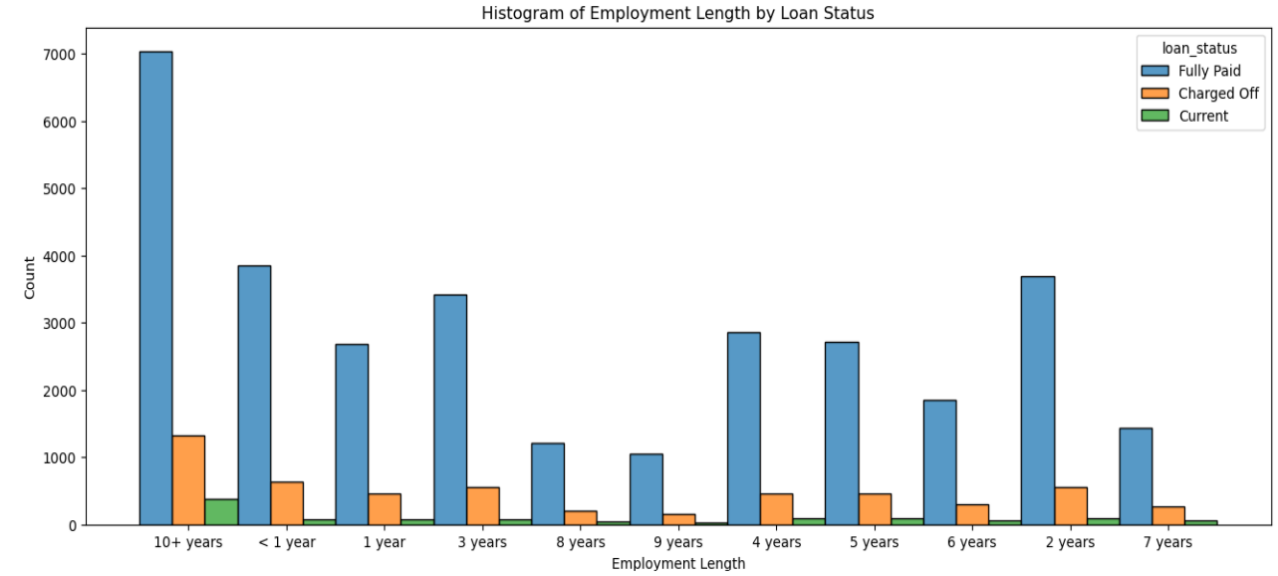
UpGrad

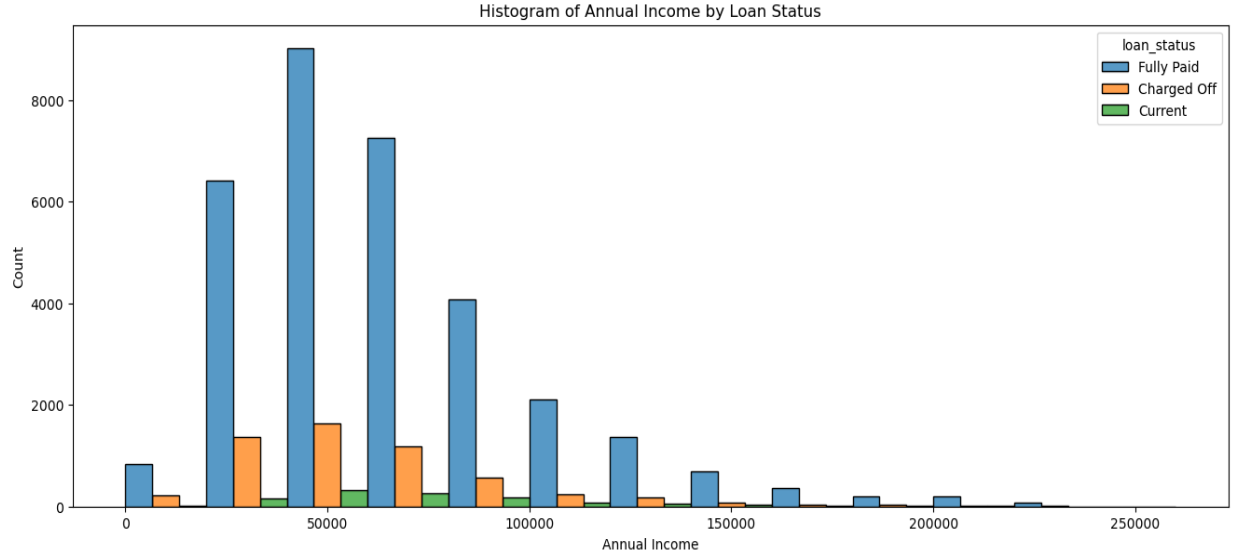
iit-b ज्ञानमुत्तमम्

The graphs below are related to various parameters divided by loan status. The pattern for both Fully Paid and Current seems to be similar. It is difficult to conclude anything from these graphs.

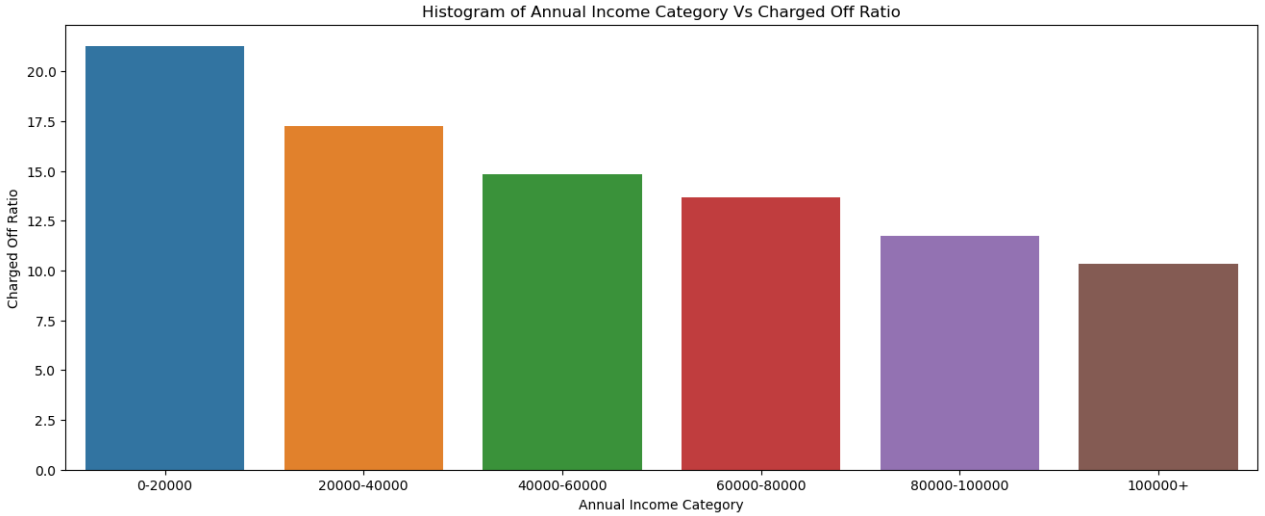




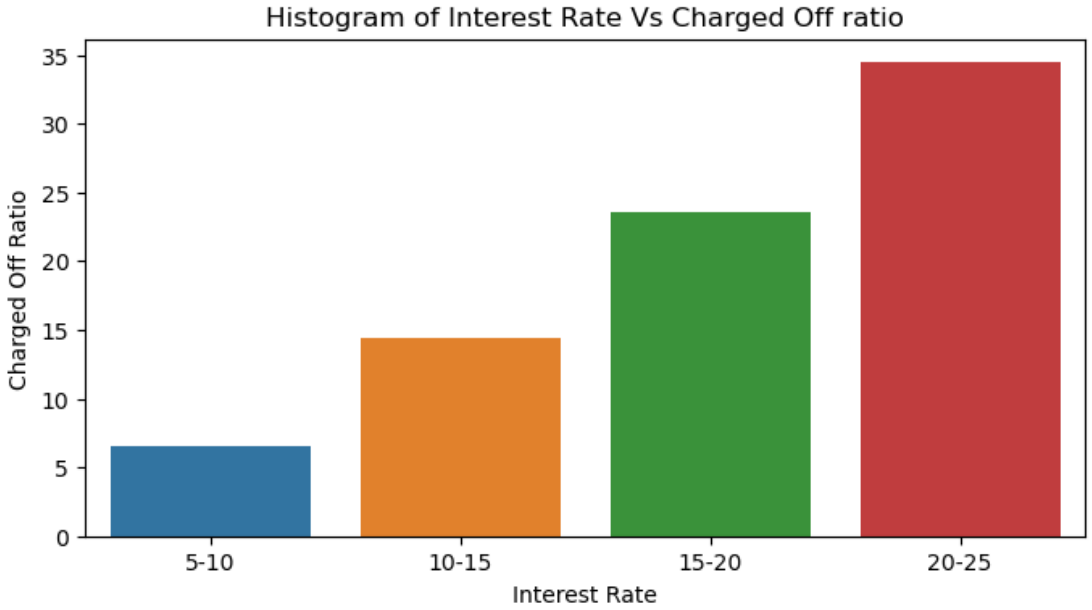




Observations

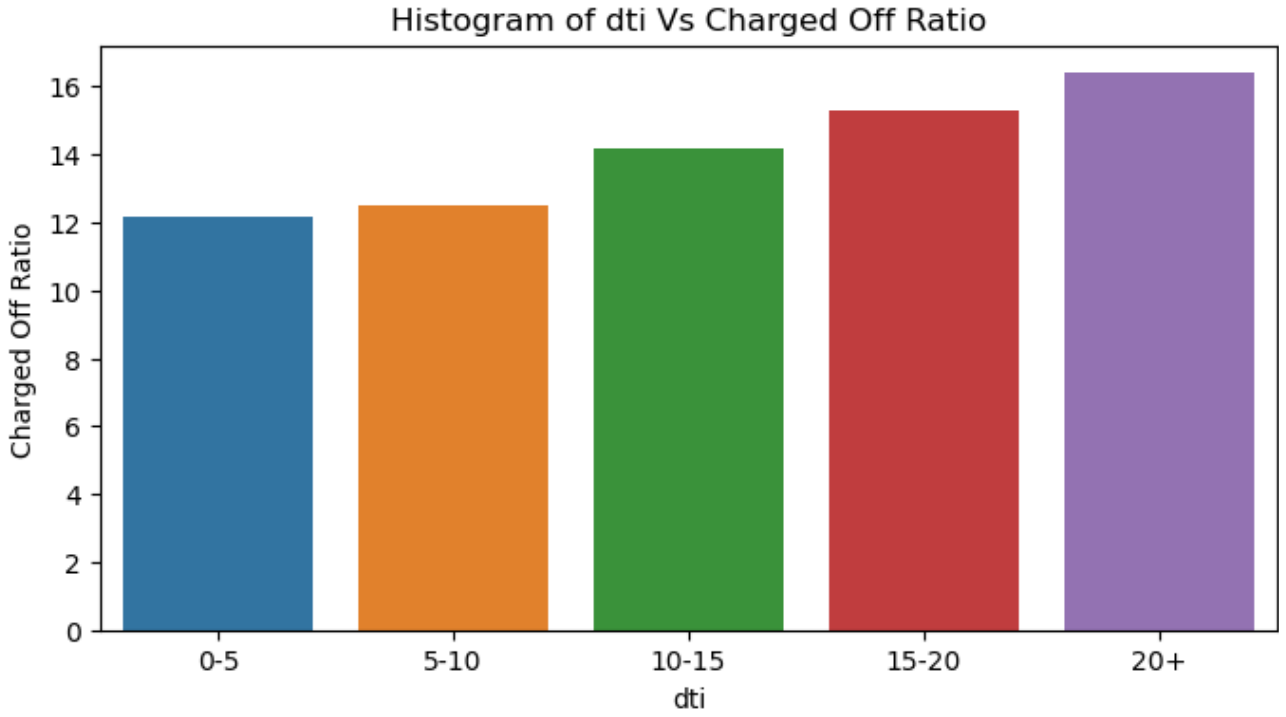


With the increase in salary, the loan charge off has reduced

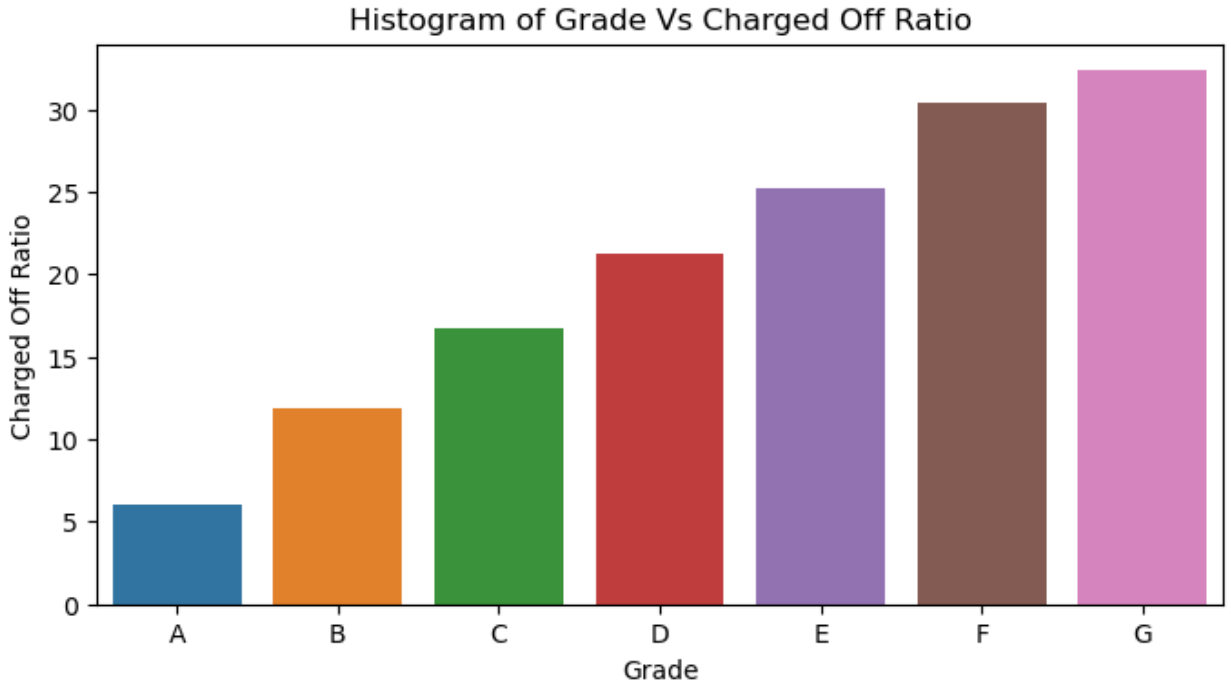


With the increase in interest rate, the loan charge off increases

Observations

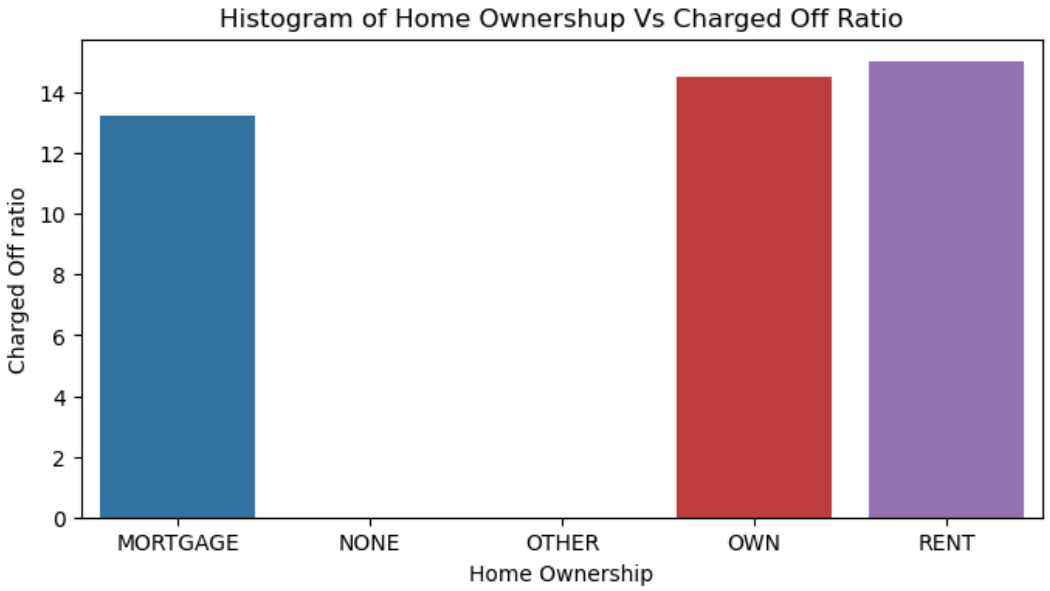


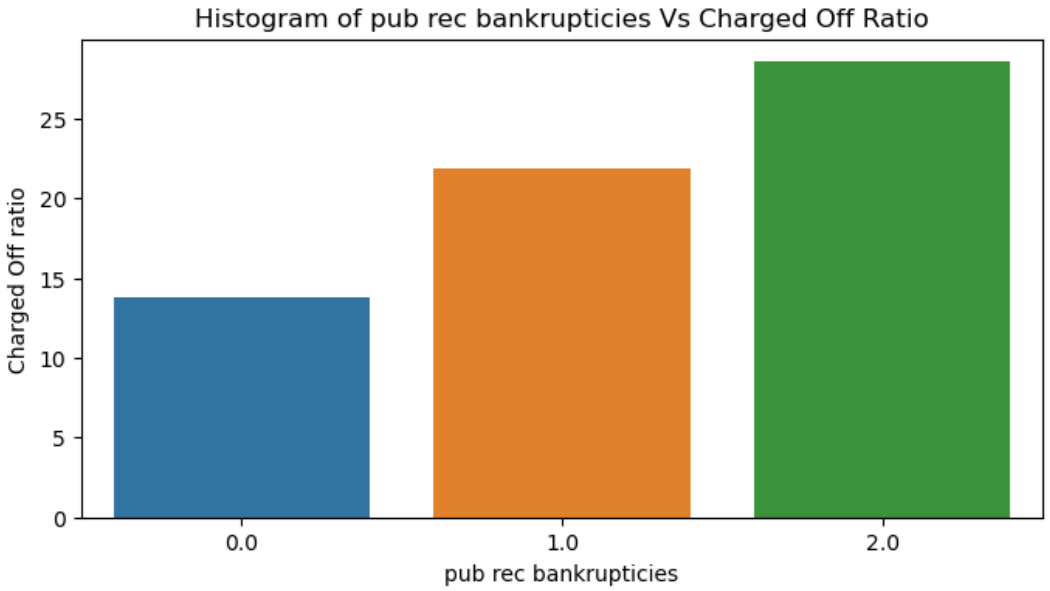
With increase in dti, the charge off increases



The charge off keeps increasing as we move from Grade A to G

Observation

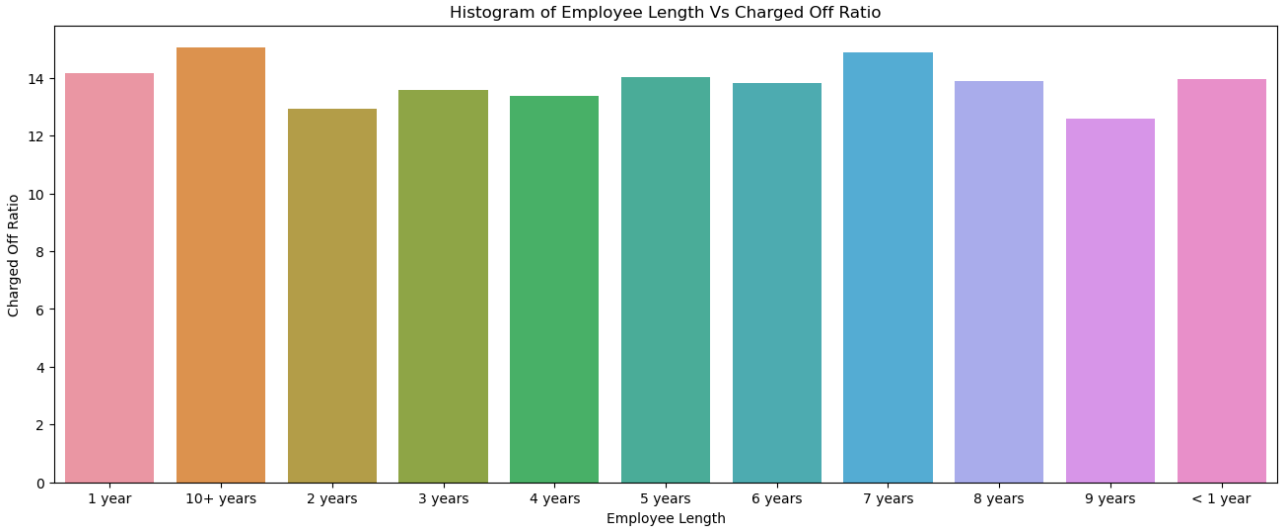




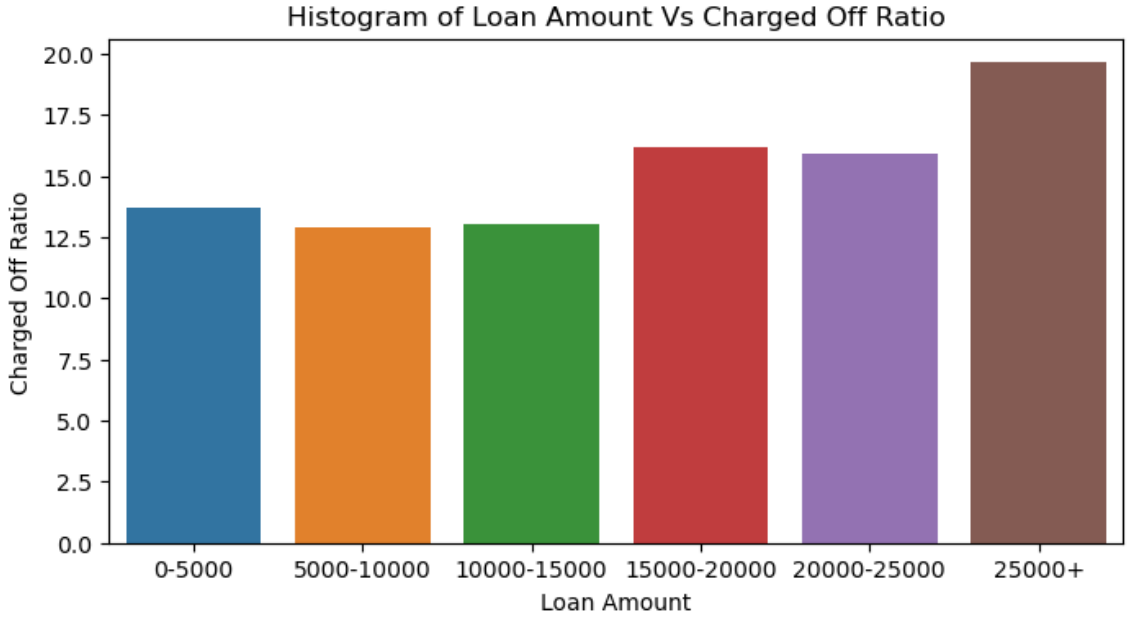
Customers with Mortgaged homes Charge off less. Own and Rented house customers are charging off equally.

Charge off increases with higher public record bankruptcies. Data for 2 bankruptcies is very less but we have enough data for 1 bankruptcies where charge off is higher than 0 bankruptcies

Observations

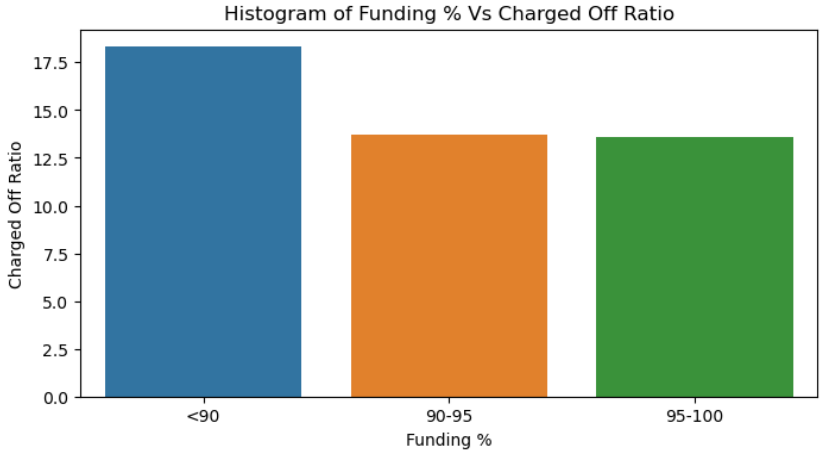


There is no clear visible pattern in the emp length and the charged off ration therefore no conclusion can be drawn from emp length

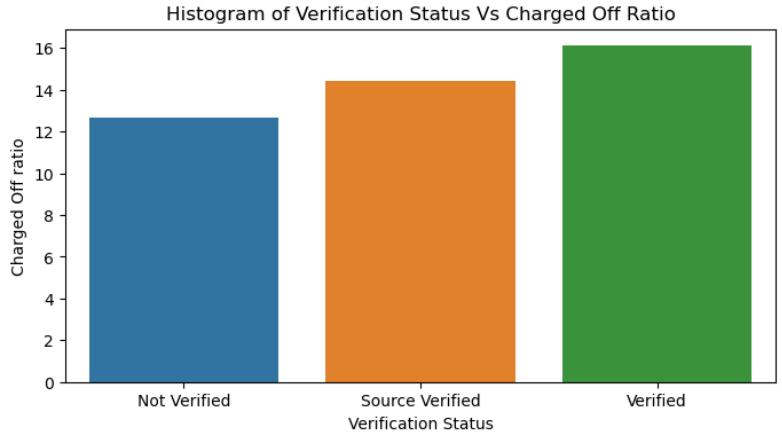


Higher charge off with higher loan amount. Highest charge off ration is for the high loan amount 25,000+

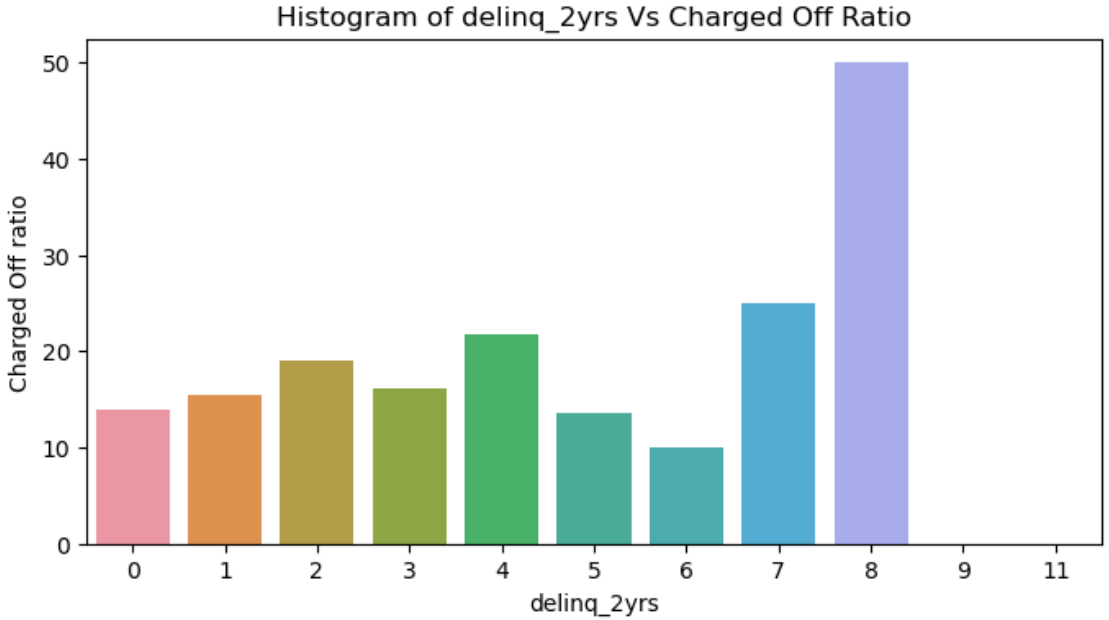
Observations



Loans with lesser funding are more prone to charge off

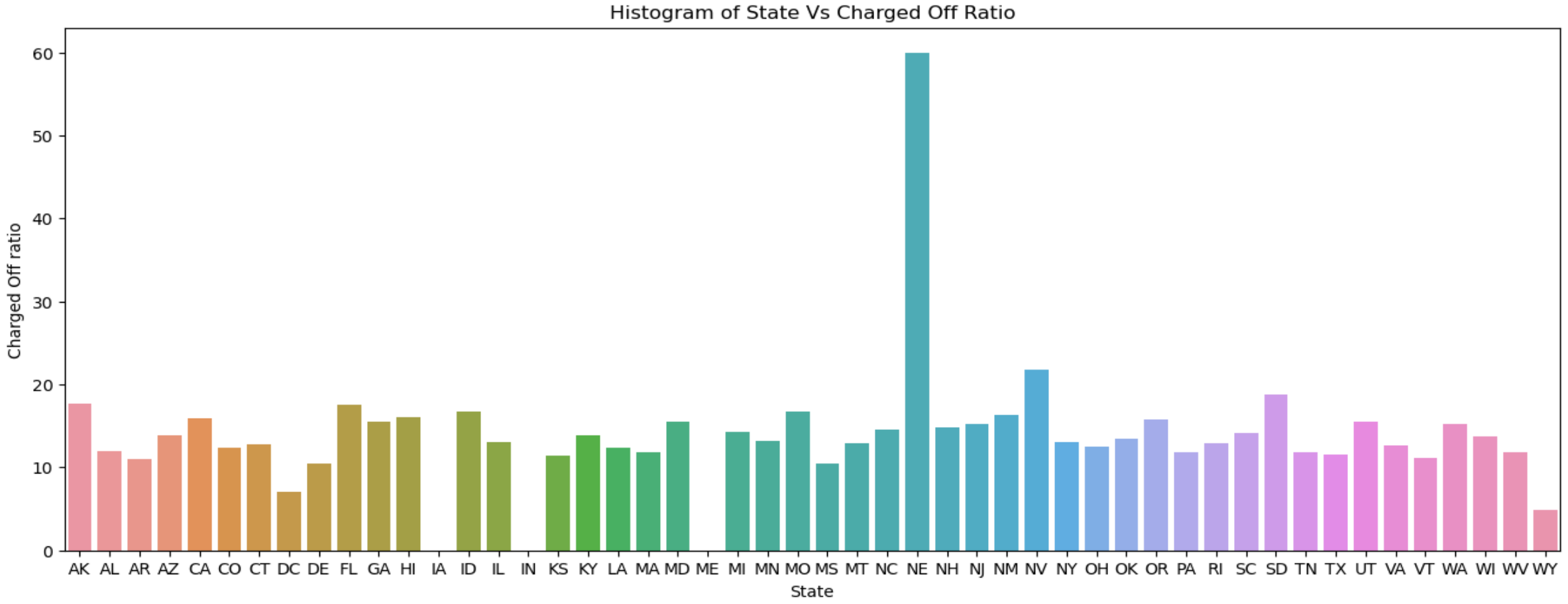


Difficult to conclude anything related to verification status



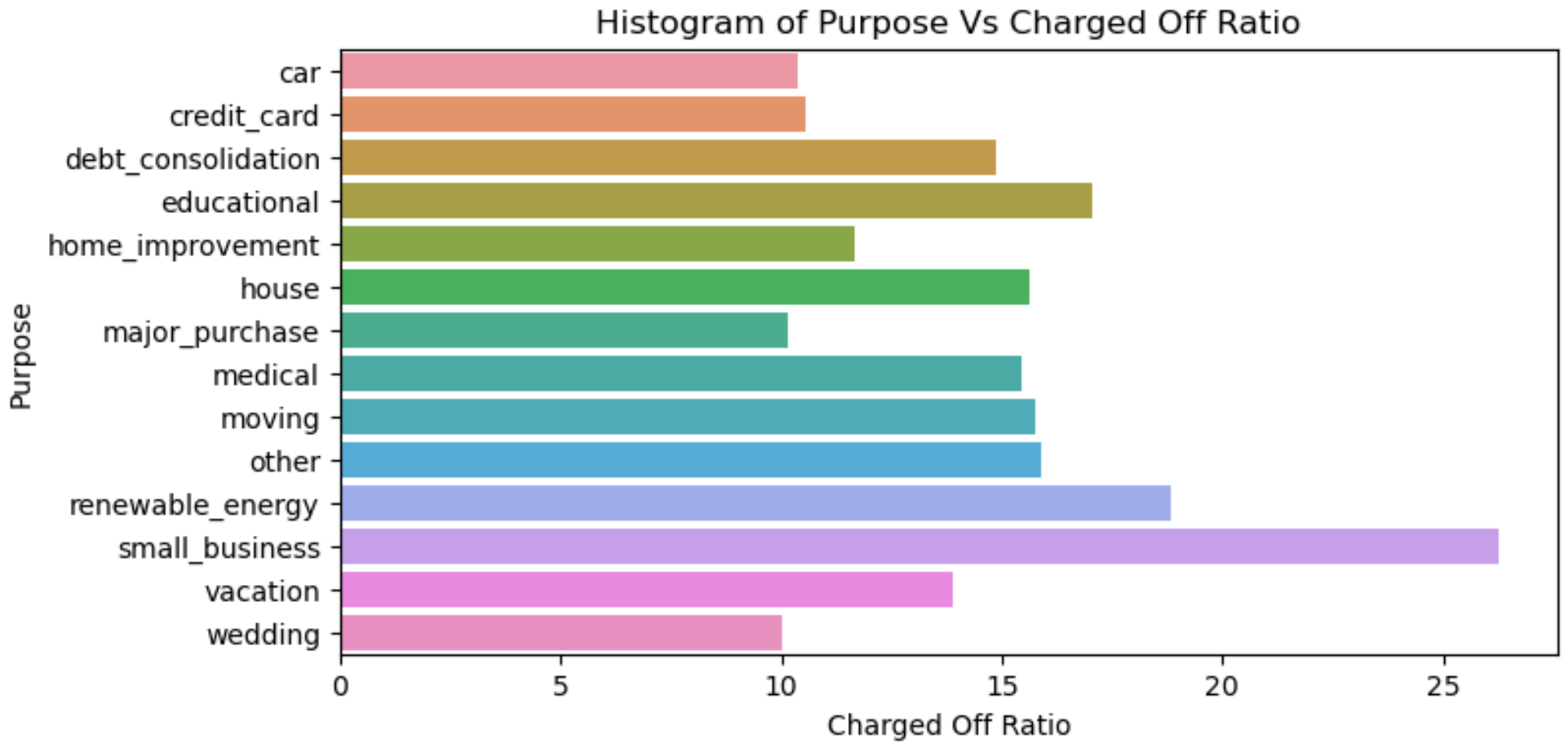
Nothing can be concluded from this variable as almost 90% of data is 0 for delinq_2yrs and its charge off is lesser than others. For higher number of delinq_2yrs the data is too low to conclude anything.

Observations



NE state seems to have the highest charge off ratio but the number of loan applications are too low to generalize. NJ, GA, MD, CA and FL have good number of loan applicants and charge off is also higher.

Observation



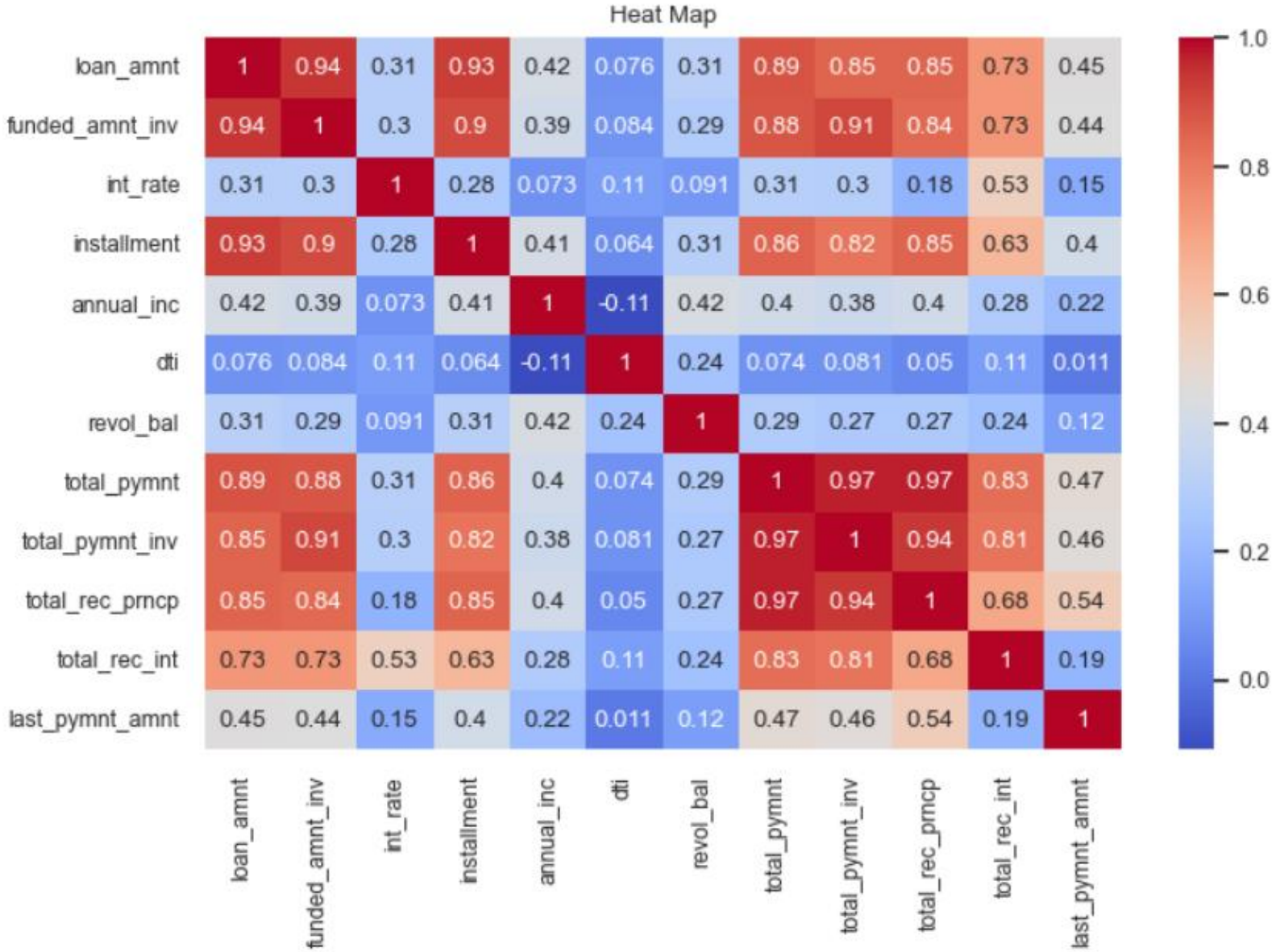
* Small businesses seems to have the highest charge off ratio

Multivariate Analysis

UpGrad

iit-b ज्ञानमुत्तमम्

HEAT MAP



Observations:

- loan_amt is having strong positive correlation with funded_amnt_inv, total_pymnt, installment
- Loan_amt is having intermediate positive correlation with annual_inc , int_rate, revol_bal.
- Total_pymnt is having the strong positive correlation with total_pymt_inv.
- Annual income is having the negative correlation with dti.

UpGrad

ज्ञानमुत्तमम्

Recommendations

UpGrad

iit-b
ज्ञानमुत्तमम्

Recommendations:

- Customers with lower salary shall not be given loan of higher amount. Also, the interest charged from them shall be lower.
- We see the owned and rented home customers are risky applicant and we should limit the loan amount based on the dti ratio.
- We see small business loan customers are risky applicants and they charged off more, so we can reduce the interest rate to avoid delay in payments.
- Increase the interest for the customers which are in less loan amount for the in moving, vacation , wedding etc.
- Loan amount for customers with grade E, F and G shall be limited, as they charge off more.
- Customers with less or no funding tend to charge off more so those customer shall be verified on parameters like dti, home ownership etc. before giving the loan.

UpGrad

ज्ञानमुत्तमम्

Thank You !!!