

Lending Club Case Study

Team Members : Vinit Naik, Rashmi Nivaragi

Date : 6th Feb 2024

Agenda

- 1.Problem Statement
- 2.Data Collection
- 3.Analysis Approach
- 4.Data Handling and Data Cleaning
- 5.Univariate Analysis
- 6.Bivariate Analysis
- 7.Recommendations

Problem Statement

You work for a **consumer finance company** which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

The data given for this case study contains information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Data Collection

The data has been Given by UpGrad in Landin Club Case Study Module along with the Data Dictionary in the .csv format

Analysis Approach

1. Data sourcing (Upgrad given loan.csv file)
2. After data sourcing apply EDA concept and perform the required changes/cleanup on the received data.
3. Then apply actual Data Analysis and Data Visualization concepts, which will helps to understand the trends.
4. And, share the outcome.

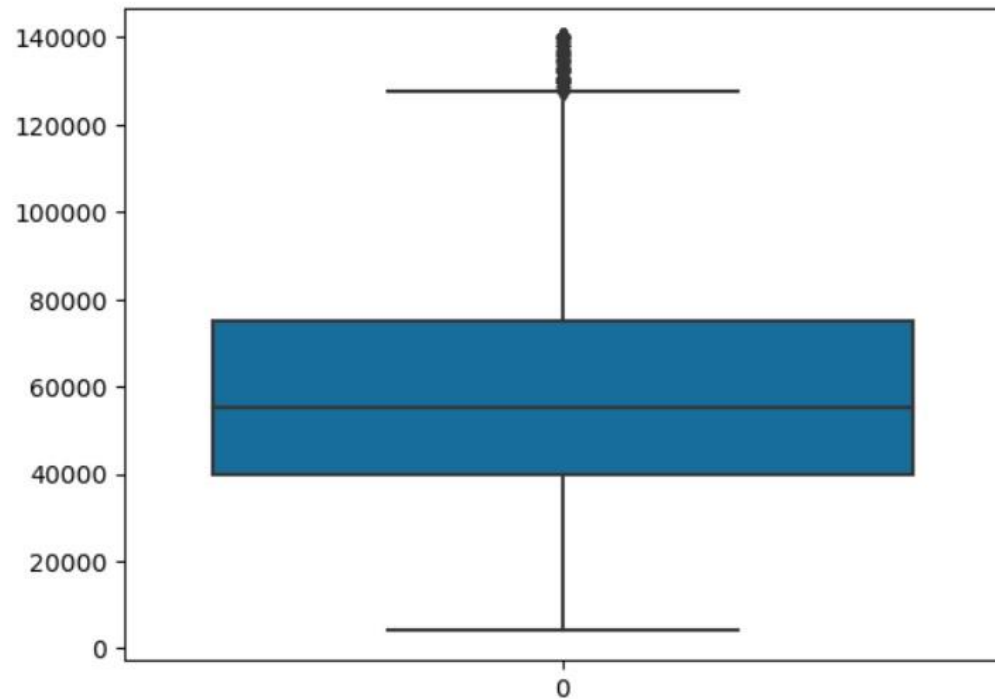
Data Handling and Data Cleaning

- First of all we gone through on dataset which we created for the loan.csv file.
- We checked what type of data it contains.
- We checked what is the shape of the dataset like total number of rows and columns.
- Then, We identified the column wise total number of null values in dataset.
- We saw multiple columns were having null values more than %60-%70 percent that's why we delete them from dataset as those where not going to help us in our analysis.
- We checked the rows where emp_length column was having NA values. we handle them using Imputation method and we fill the mode() value.
- Also, we removed the extra characters (like < , + Years) from the 'emp_length' and change data type to String.
- We drop rows wherein revol_util column value is null. Also, we remove the "%" from revol_util and change to datatype to float.

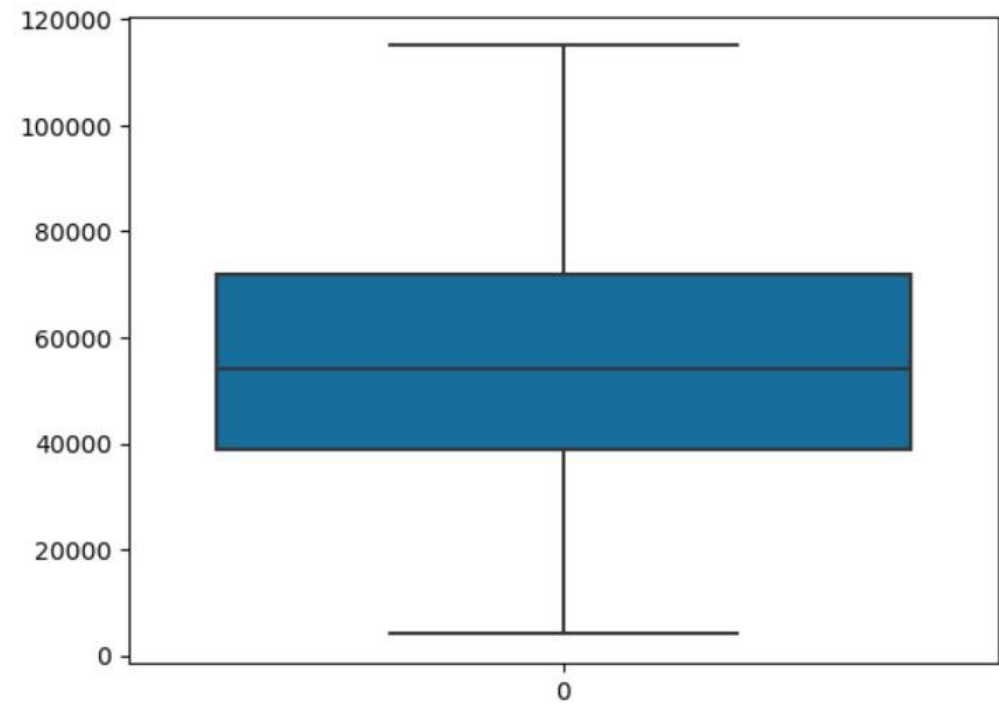
- int_rate columns was having % symbol as next to each value in column so we removed that % symbol and change the int_rate datatype to float.
- We created new column int_rate_groups and distribute the int_rate column in bins
 - ['0%-5%', '5%-10%', '10%-15%', '15%-20%', '20%-25%']
- We created new column loan_amnt_groups and distribute the loan_amnt column in bins
 - ['0%-5%', '5%-10%', '10%-15%', '15%-20%', '20%-25%']
- Also, we created new column annual_inc_groups and distribute the annual_inc column in bins
 - ['3k-31k', '31k-58k', '58k-85k', '85k-112k', '112k-140k']
- we created new column dti_groups and distribute the dti column in bins
 - ['0-6', '6-12', '12-18', '18-24', '24-30']
- We skip the people who are currently paying their loan from the analysis.

- We checked the outlier using boxplot and removed them like below from annul_inc and loan_amnt column using quantile function (below ex. Is for annul_inc column)

Before Removing Outlier

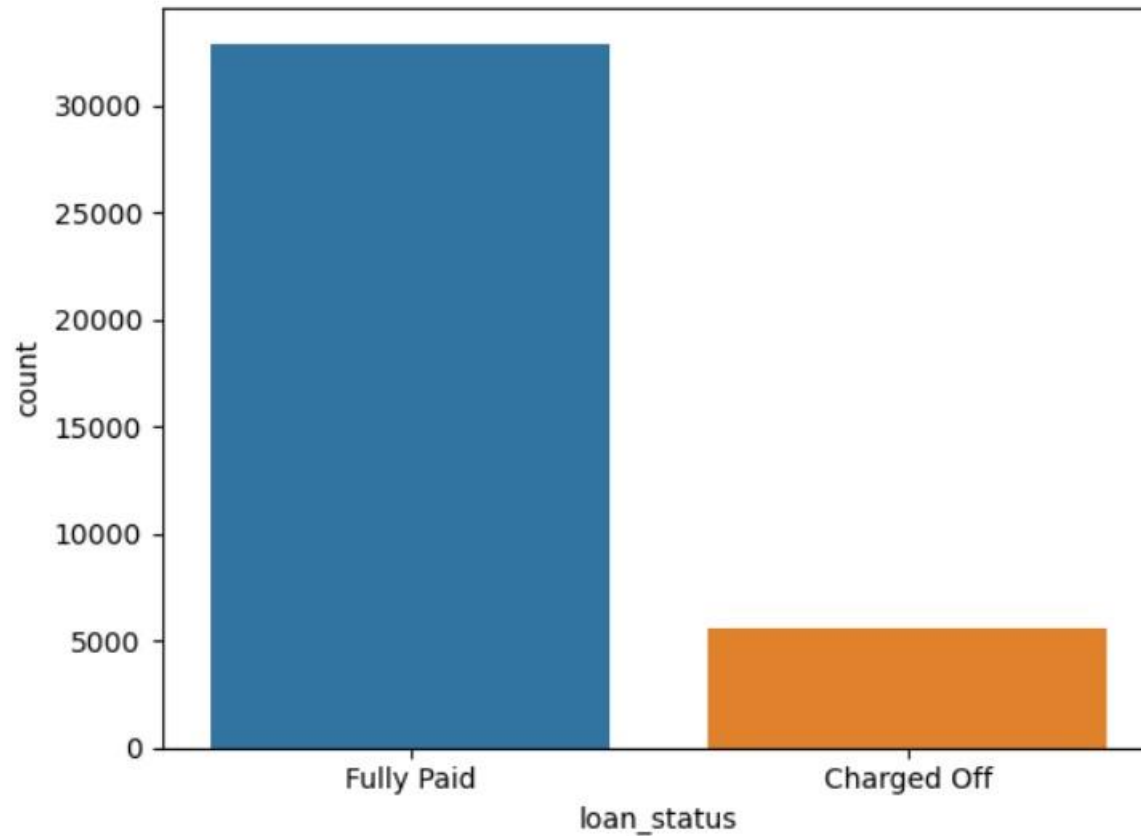


After Removing Outlier

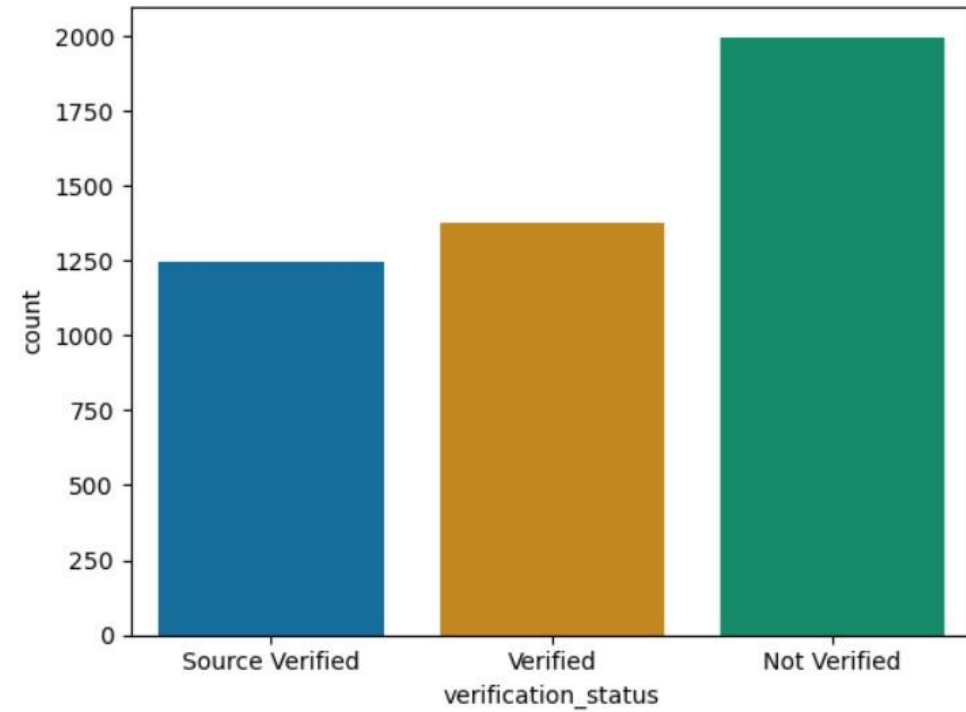


Univariate Analysis

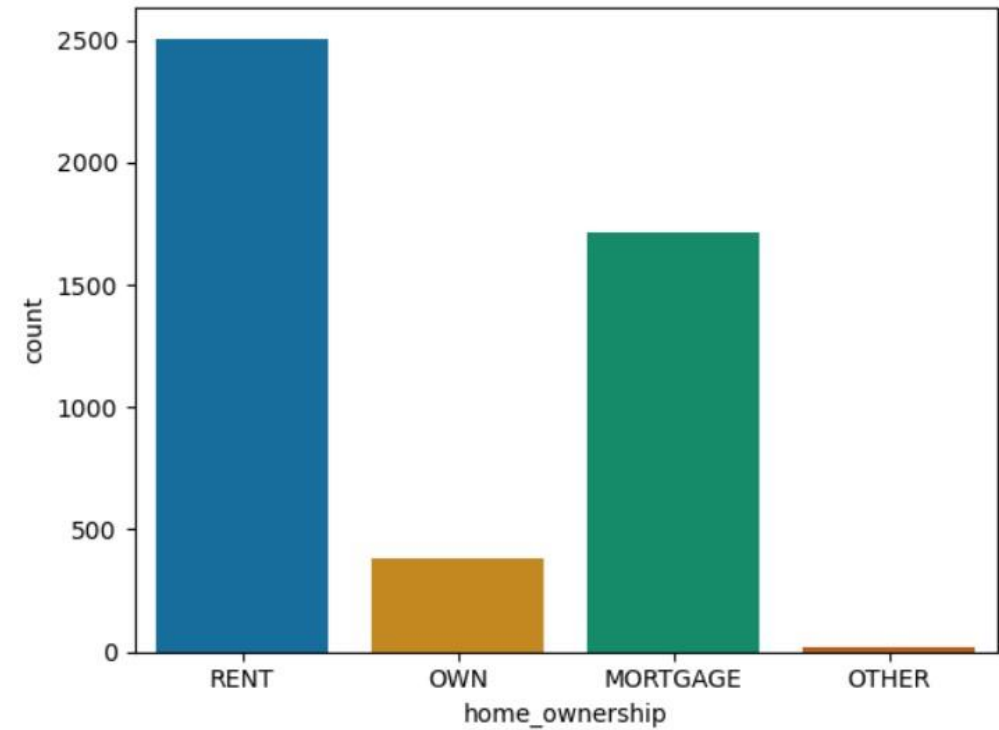
Considering loan_status column it seems that there are customer who had not paid the loan



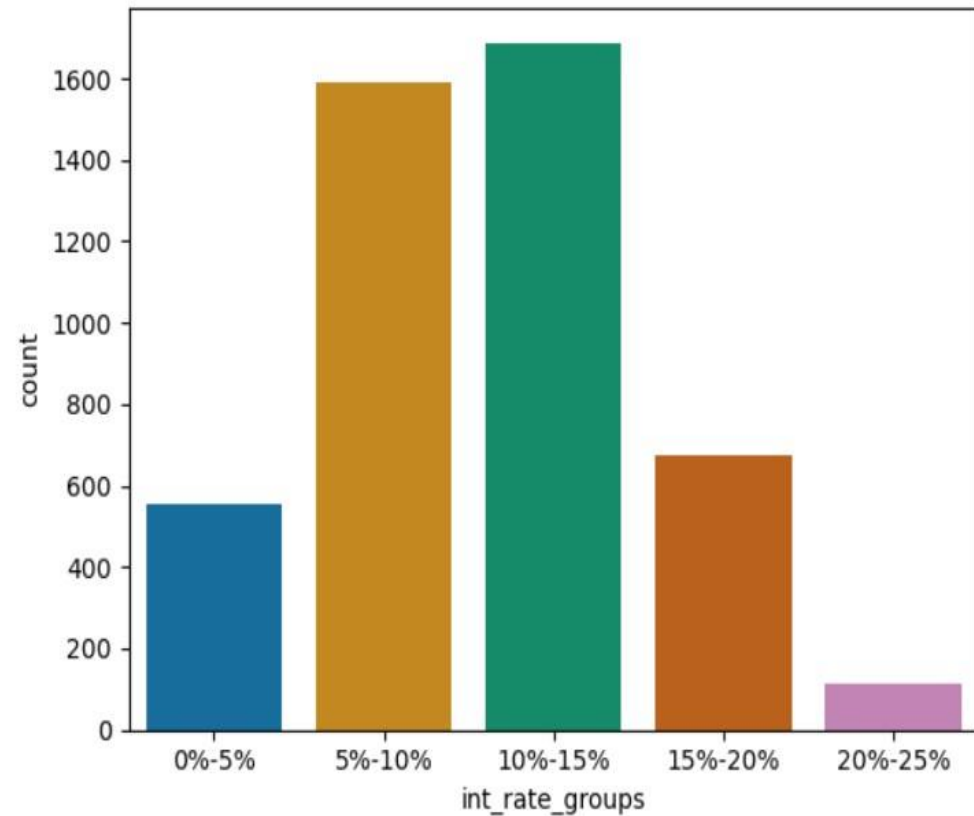
Not Verified Customer Count is Very High



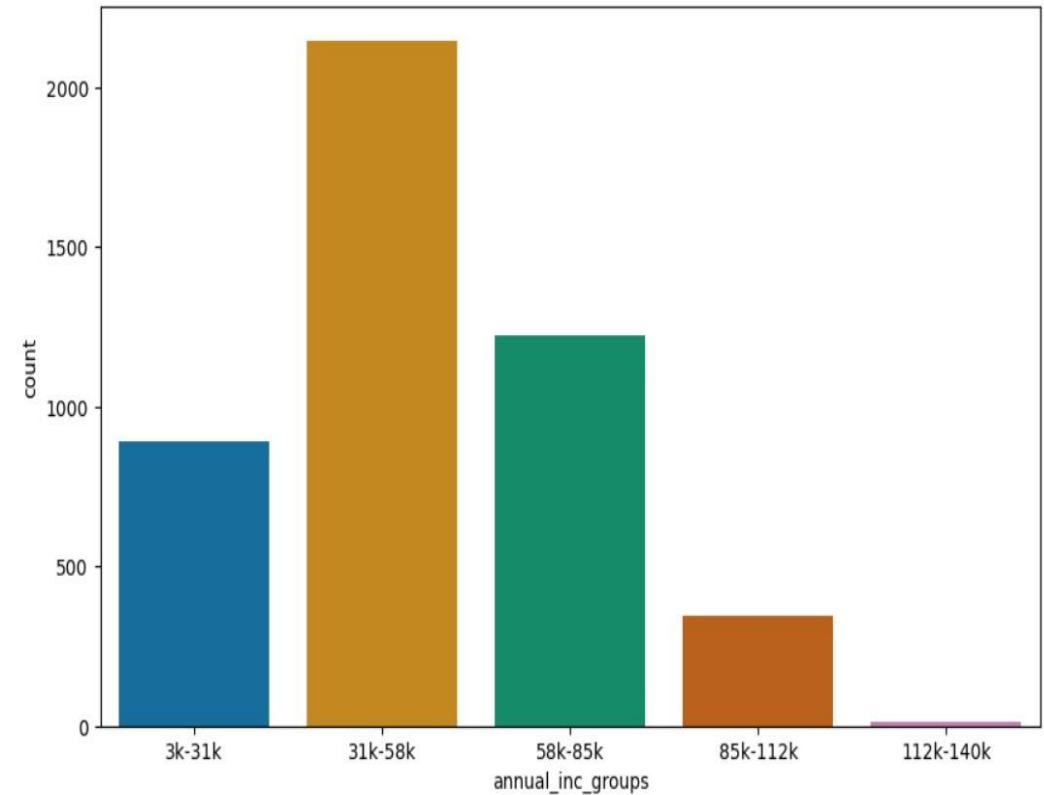
Most Charged of loan customers having home_ownership = RENT



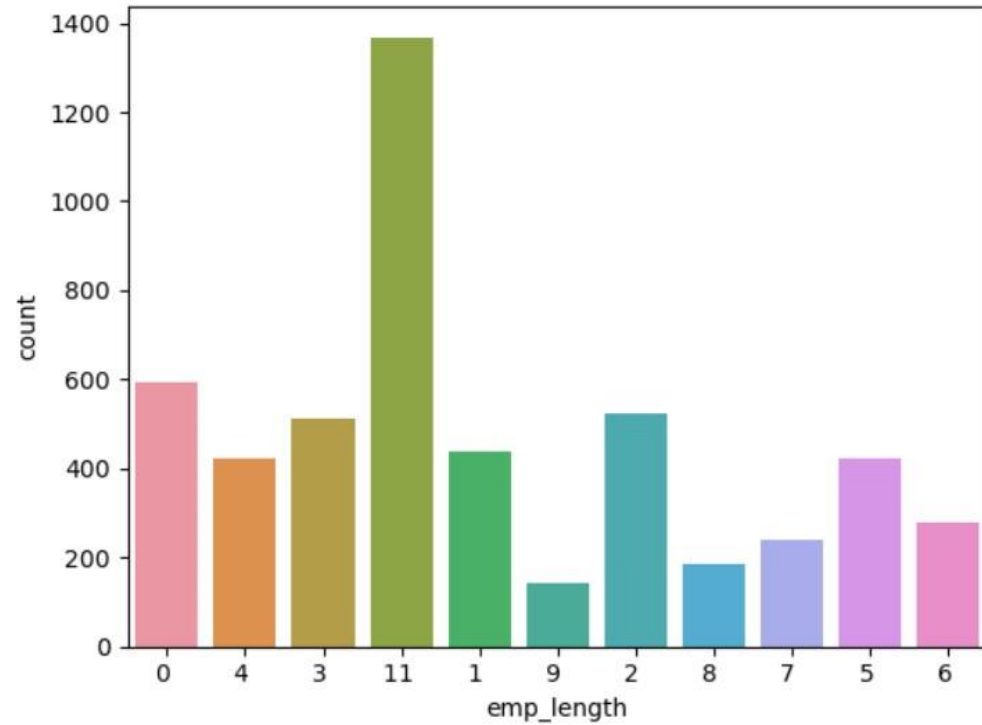
Most Charged off loans having interest rate 10-15%



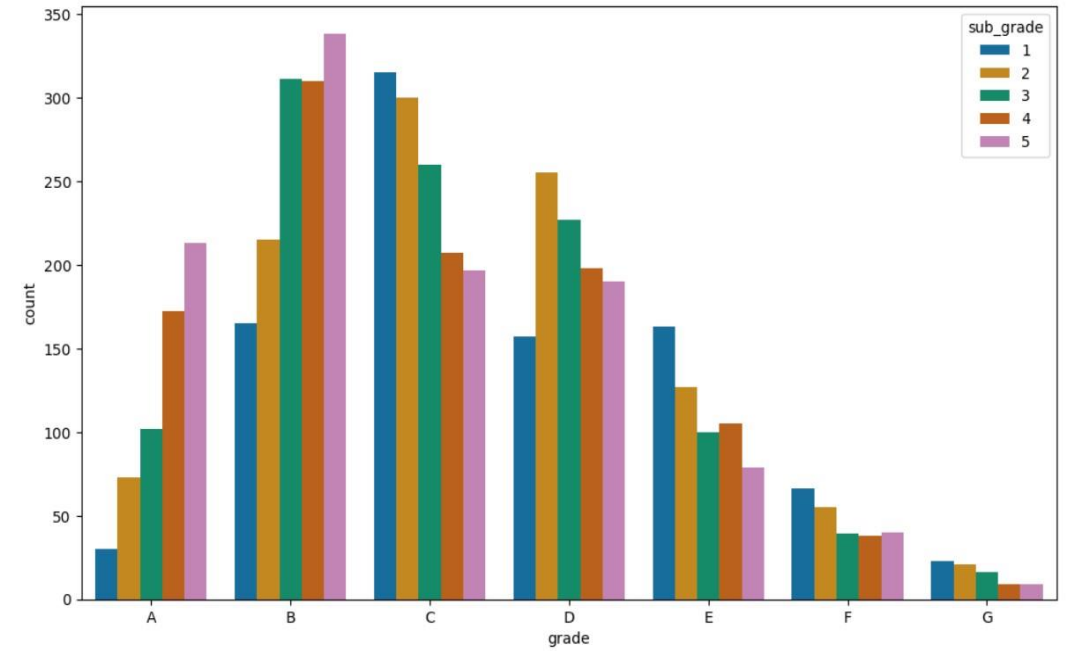
Most Charged off loans having Borrower annual income range between 31k - 58k



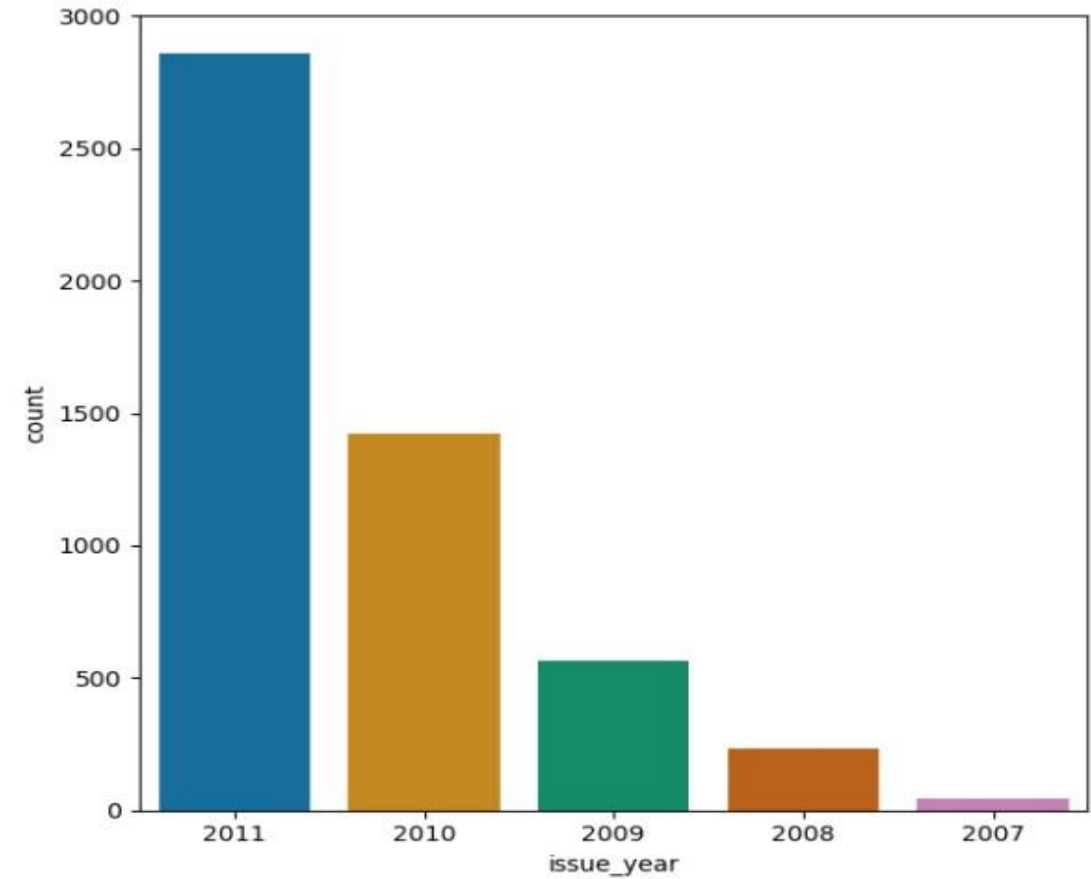
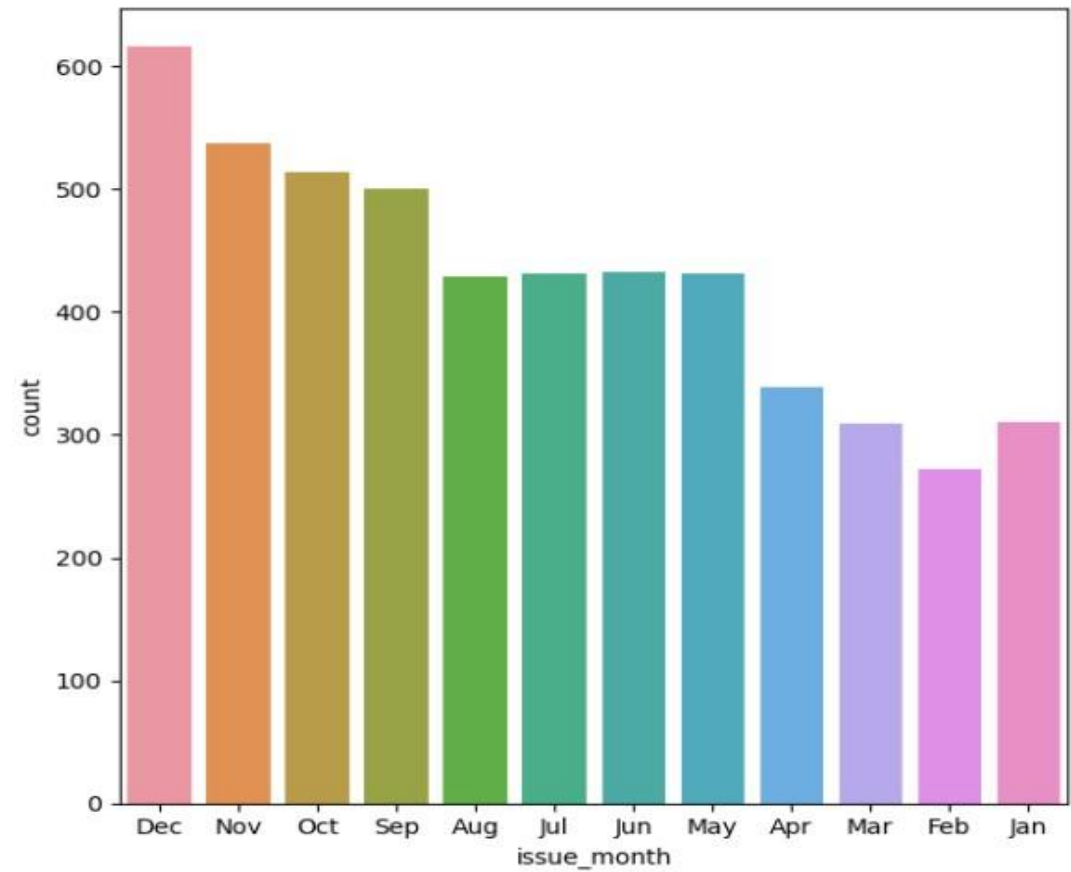
Most Charged off loans borrower having employment length > 10 years



Number of Grade and subgrade for Charged Off loan



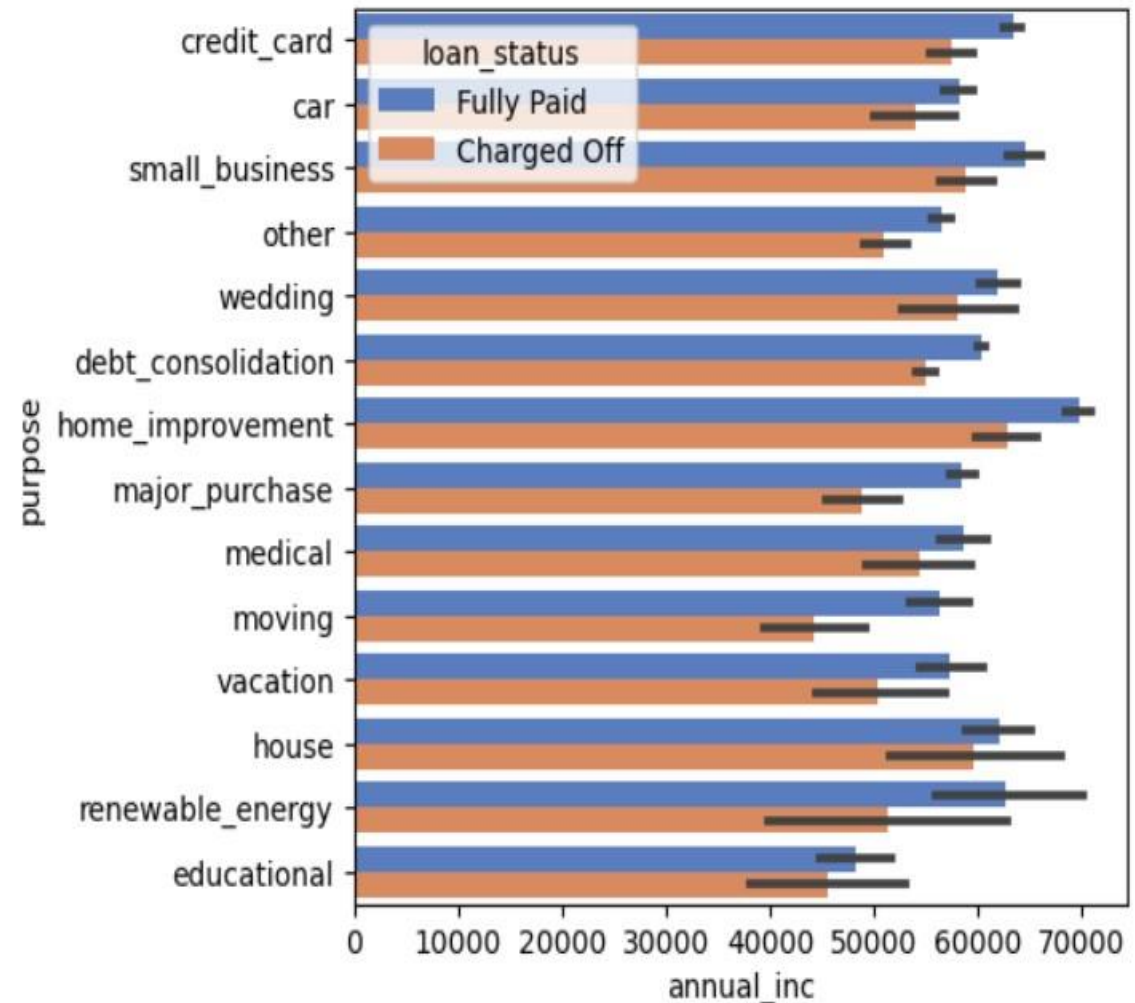
Most defaults are in December and in 2011 (due to the financial crisis in USA)



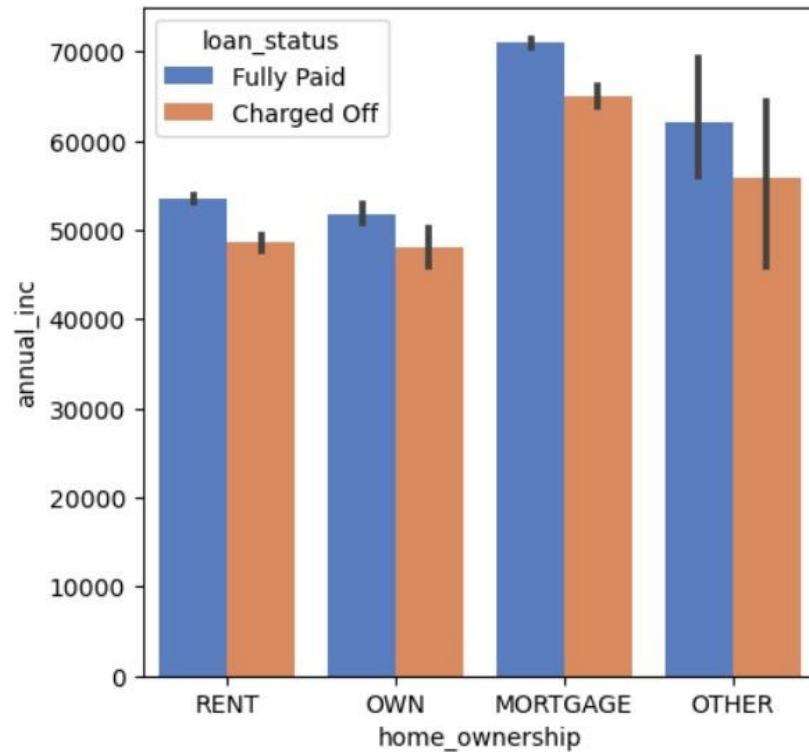
Bi-Variate Analysis

In this section we have compare the 2 variables and shown the results

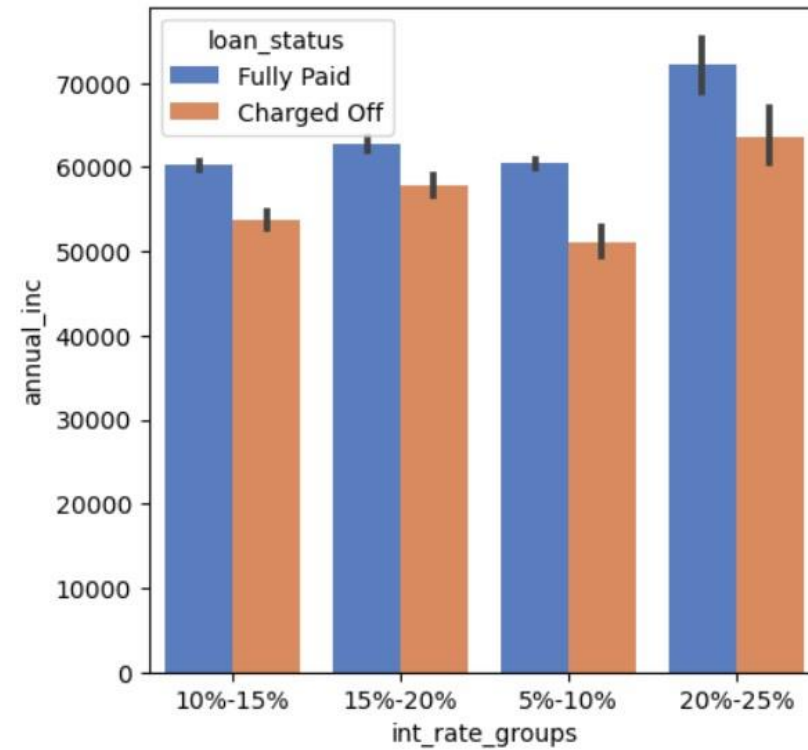
Like in the given Graph Most applications with purpose 'home_improvement' or 'house' and having income between 50k - 70k are prone to default



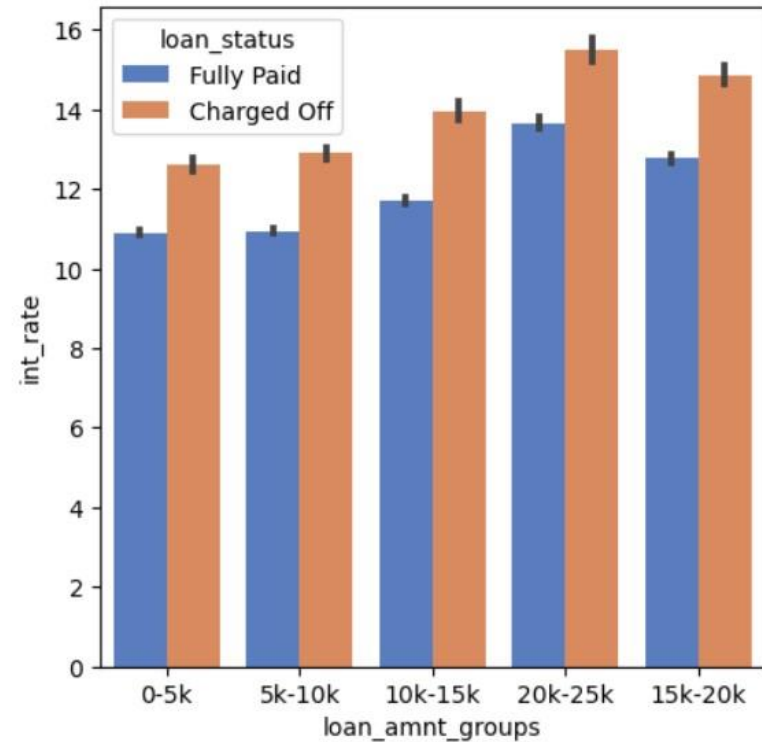
Applications whose home ownership is 'MORTGAGE' and have income of 60-70k are mostly defaulters



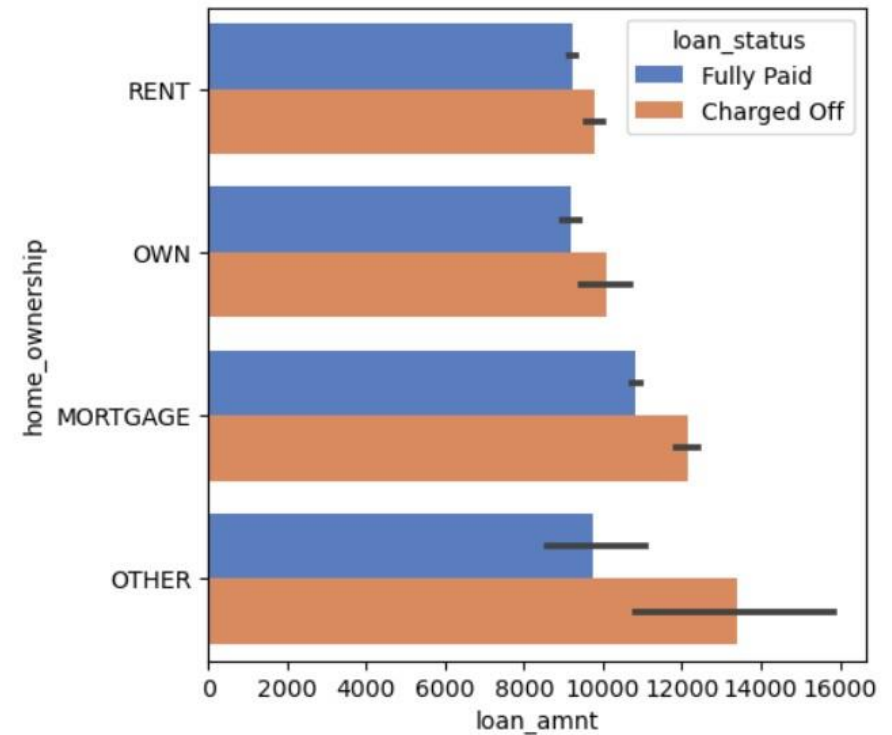
Applicants who receive interest rate between 20-25% and have an income of 50k-60k are defaulters



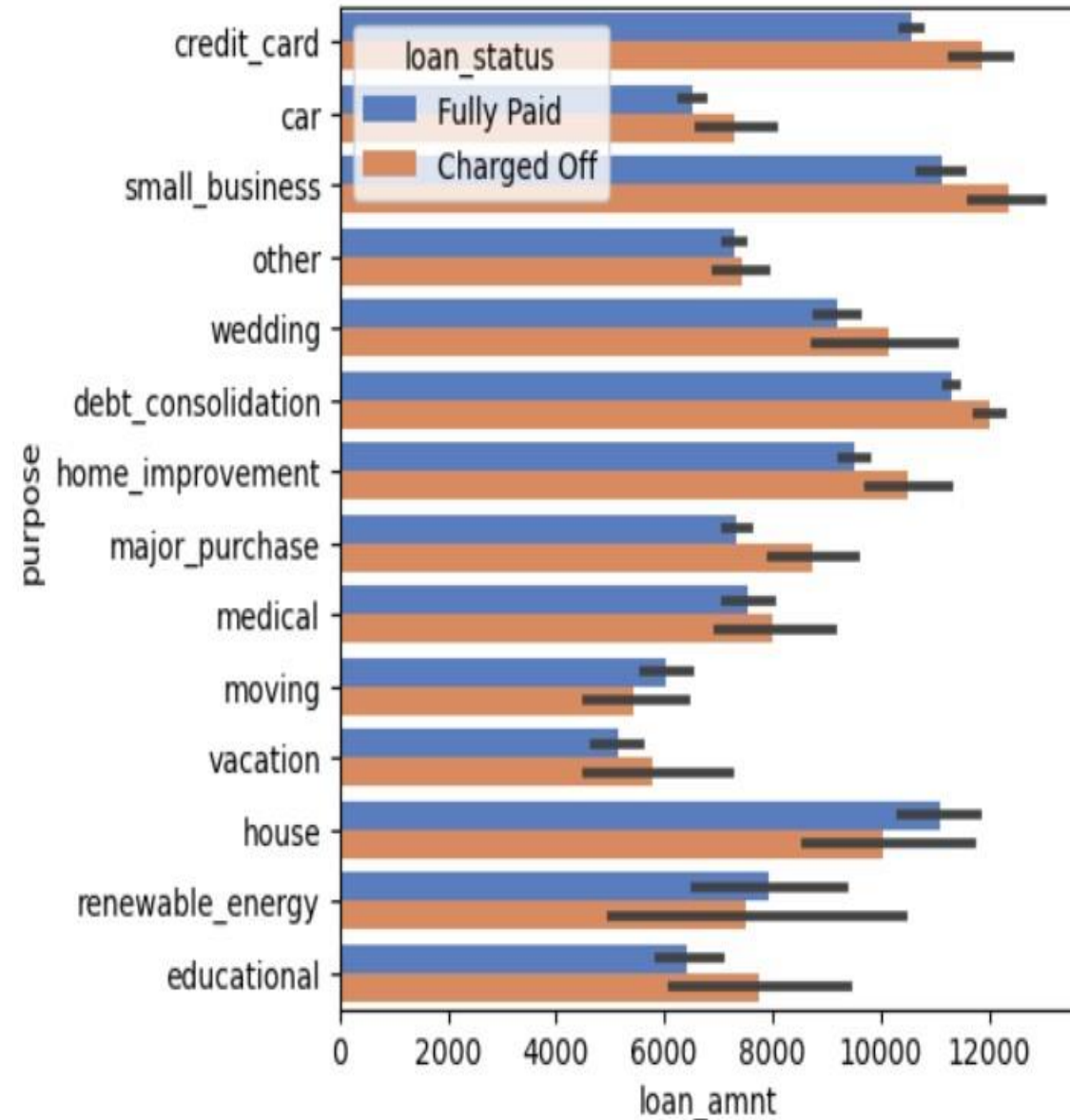
Applicants who have taken a loan amount between 20k - 25k and are charged interest rate of 14-16% are defaulters



Applicants whose home ownership is 'OTHER' and have loan amount of 12k-14k are defaulters

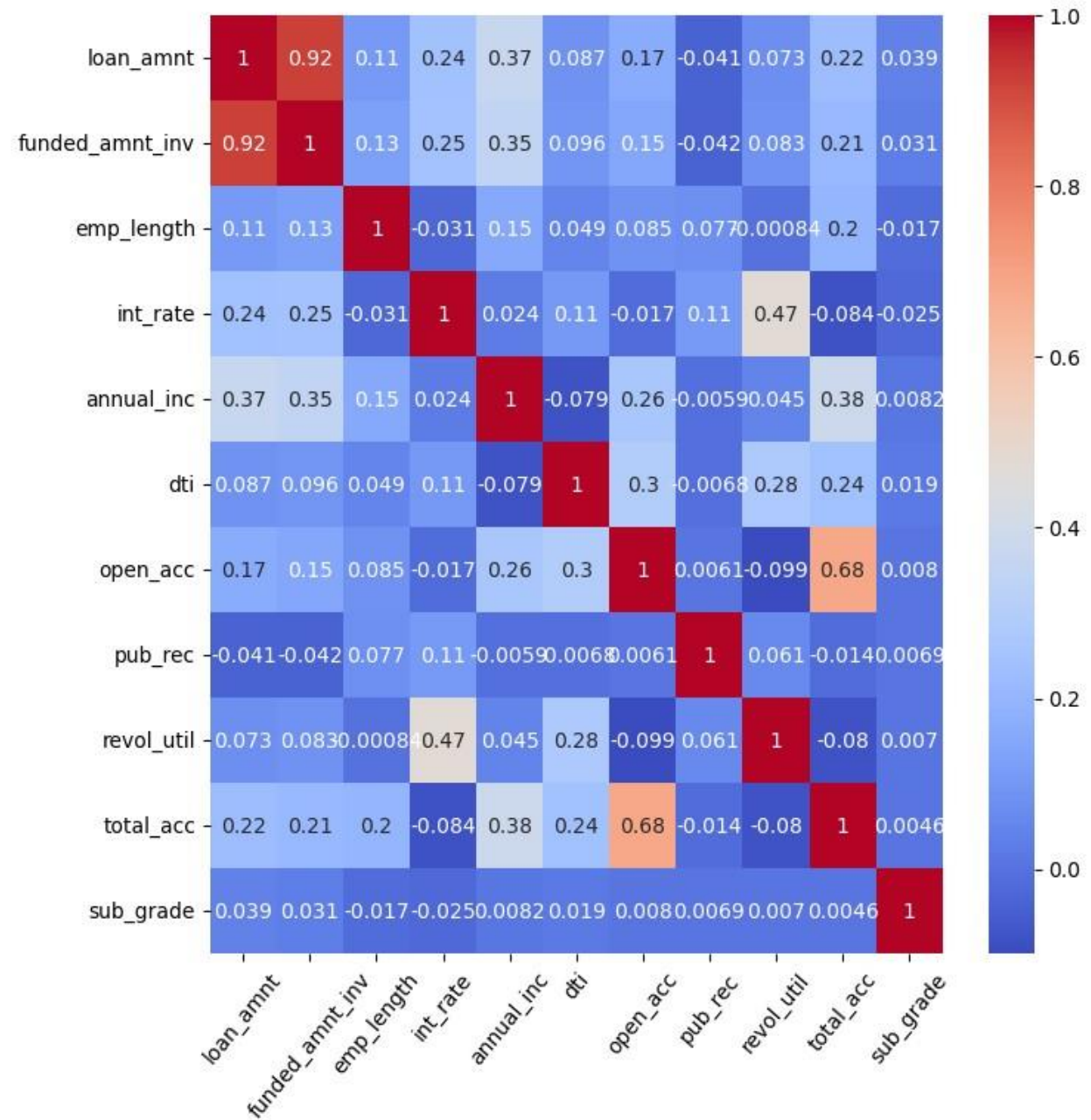


Applicants who have taken a loan for debt consolidation, small business and credit card and the loan amount is greater than 10k are defaulters



Positively Co-related attributes:

1. Loan Amount and Funded amounts
2. Annual Income and loan amount
3. Interest rate and annual income



Recommendations

1. If a Customer comes/agreed to take the loan on high interest rate then please do more background checks on that customer.
2. Try to Verify customer fully while giving loan.
3. If a Customer income is very low and he is coming for loan then check if that person will be able to pay the loan or not. Because due to the Low Income the chances are that customer will become defaulter so either do more background checks or reduce the loan amount.
4. Max number of people are taking loan for debt consolidation so it's good to check the more things for these loan Univariate Analysis
5. Customer whose home ownership is 'MORTGAGE' is having high Charged Off amount so it's good to have more background checks for these type of Customers.
6. Also, It seems Customer whose purpose for taking loan is for debt consolidation, small business and credit card are having high Charged Off amount so it's good to have more background checks for these type of Loans.