

Lending Club Case Study (EDA)

Problem Statement

Goal

Identify the risky loan applicants from the data, so that loans can be given to only those where the amount of credit loss will be reduced.

Objective

This 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.

Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).

Borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

Target

Using EDA, understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default and derive the patterns of loan defaulters.

Approach followed

The approach for analysis is divided into four main parts:

- a. Data understanding from the csv files provided.
- b. Data cleaning (cleaning missing values, removing redundant columns etc.).
- c. Data Analysis (Univariate, Bivariate).
- d. Recommendations.

Data Understanding

Types of variables

- Customer (applicant) demographic
- Loan related information & characteristics
- Customer behavior (if the loan is granted)

Customer Demographics

Employment length

Employment Title

Annual Income

Zip Code

Description

Loan Info

Loan Amount

Funded Amount

Interest Rate

Loan Status

Loan Grade

Customer Behavior Variables

Delinquency year

Earliest Credit Line

Revolving balance

Recoveries

ApplicationType

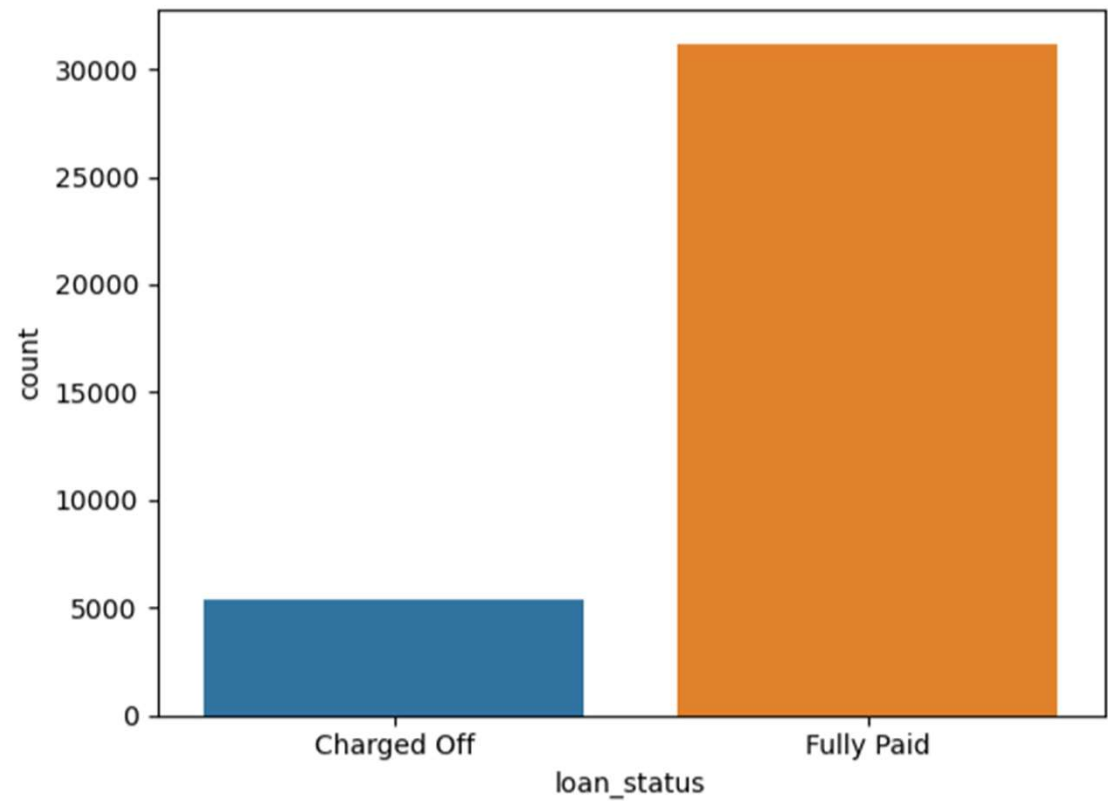
Loan Purpose

Data Cleansing

- a. There are some columns with lot of missing values. We need to drop such records.
- b. Removed all the null value columns.
- c. Found the percent of values in columns which are NA and removed the unnecessary columns which will not help in analysis.
- d. Removed text columns which are not useful for this loan disbursement analysis.
- e. Converted datatypes as needed from object to Int / float.
- f. Detecting outliers and deleting them for getting more clean data.

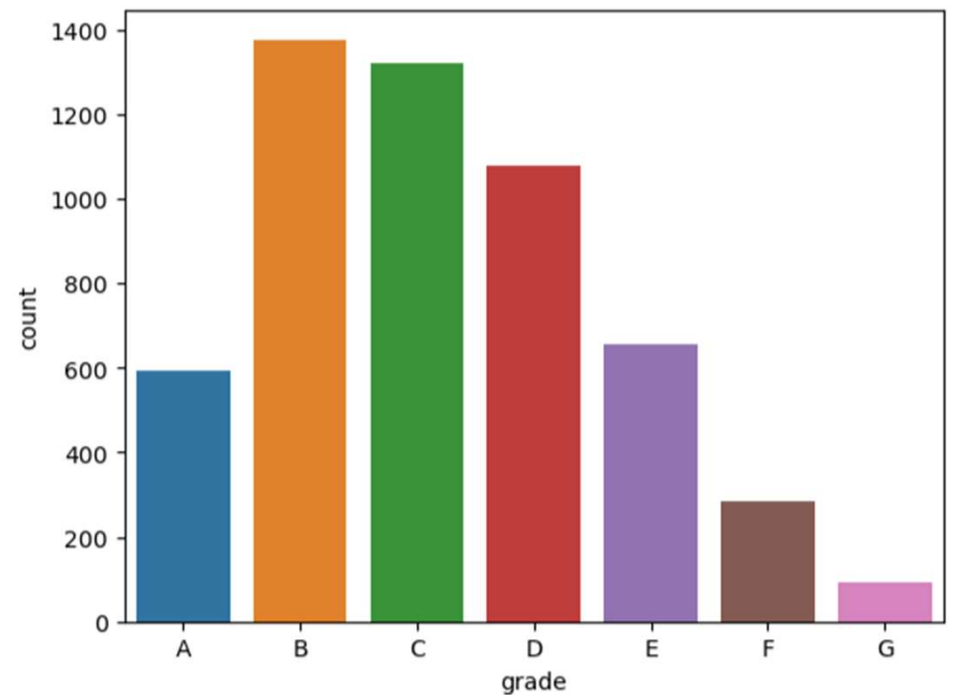
Analysis

This graph represents people who have already paid loans and defaulters. The defaulters are about 15 or 16%



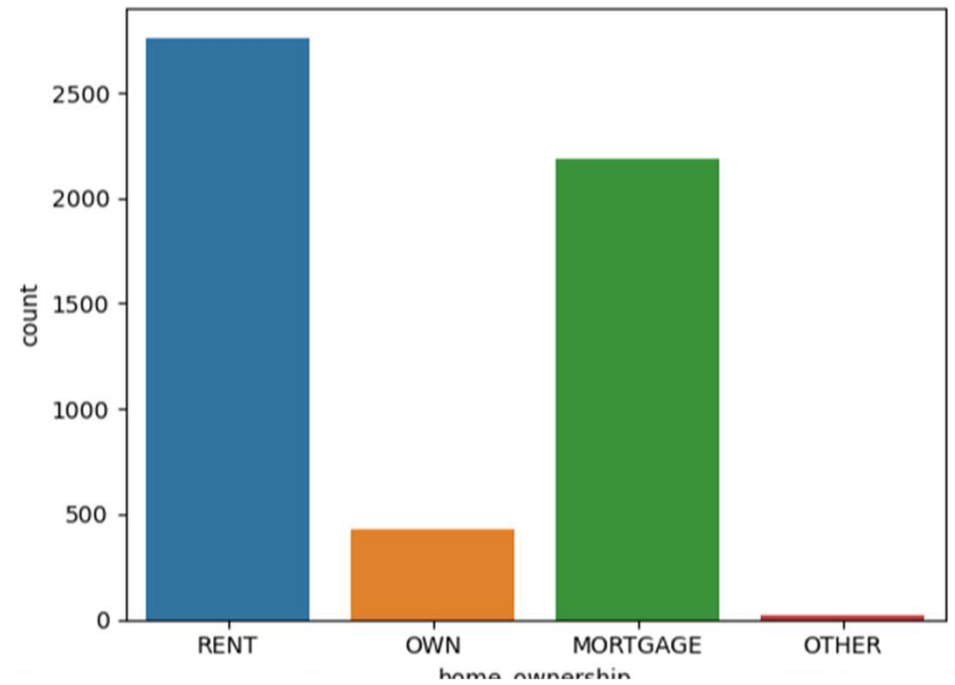
Analysis

This graph represents that majority of defaulters are in grade B



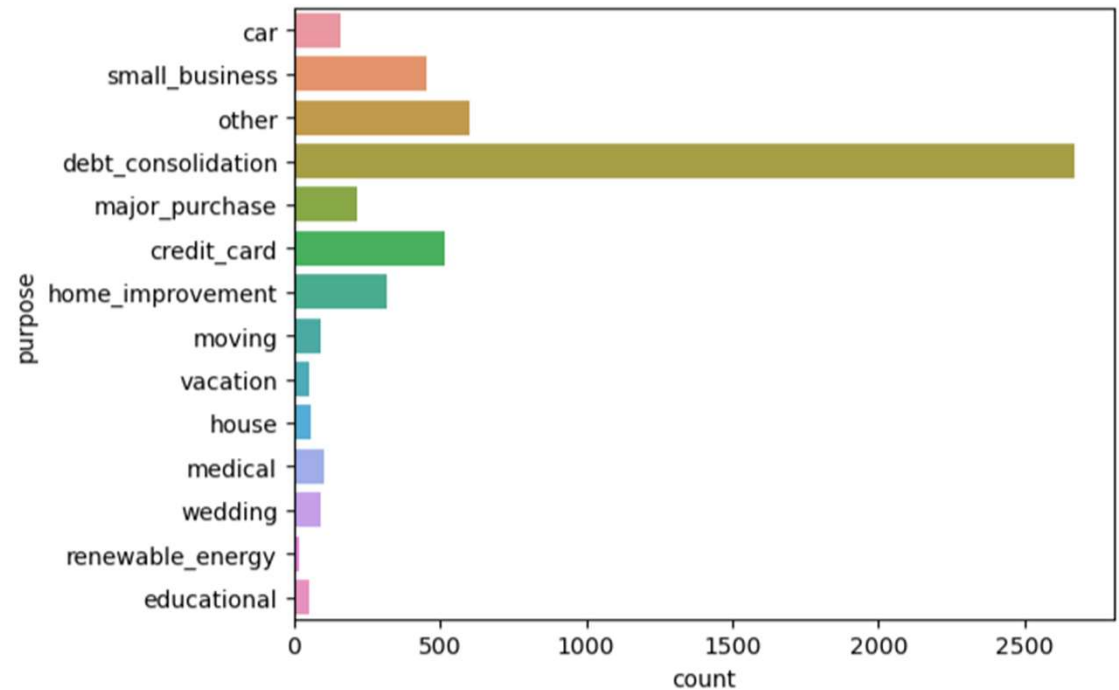
Analysis

With respect to home ownership, the members who own houses are less likely to default on the loans.



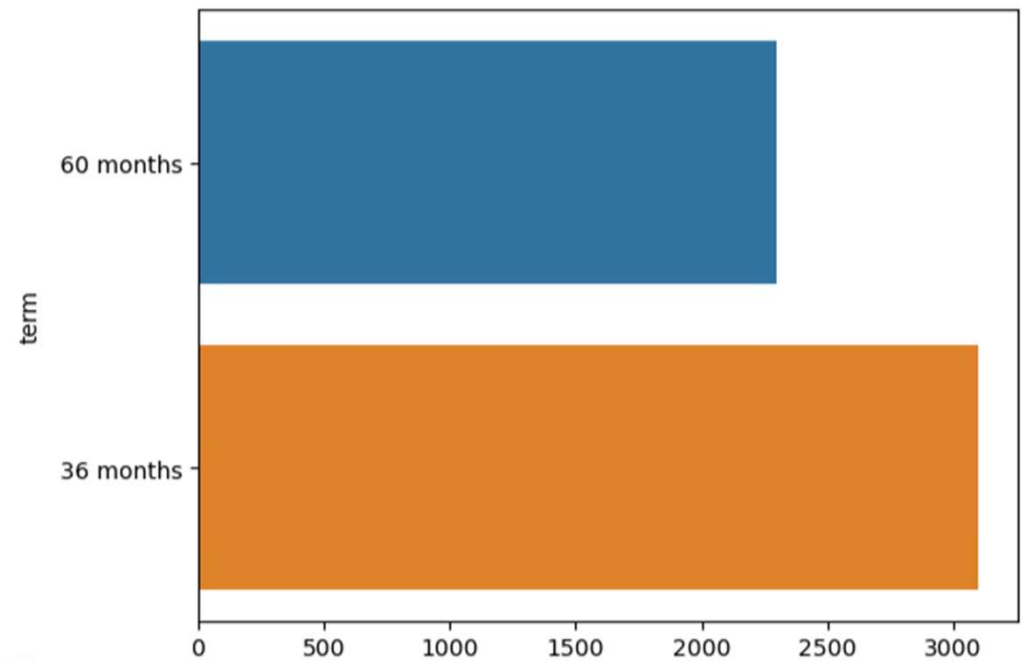
Analysis

Most of the members have a purpose of debt_consolidation which turn out to be defaulters. So when new loan is being issues, if the purpose entered is debt_consolidation then it may result in a defaulted loan.



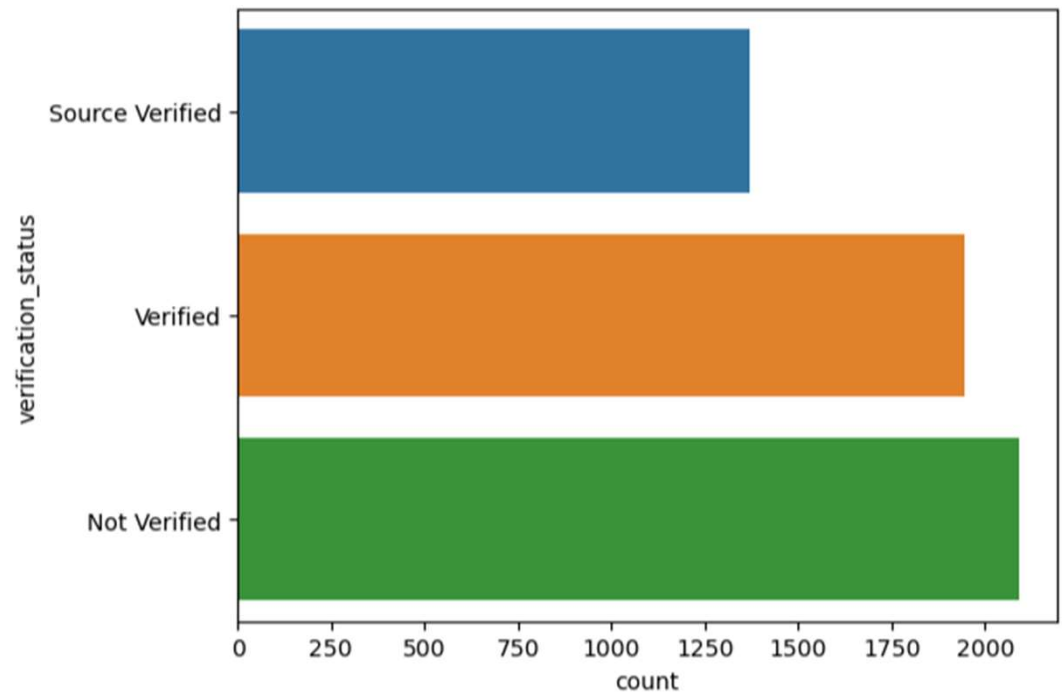
Analysis

Members who have lesser time to pay off debts are more likely to default. When given time, most of them have paid. When time is less, defaulters count is more.



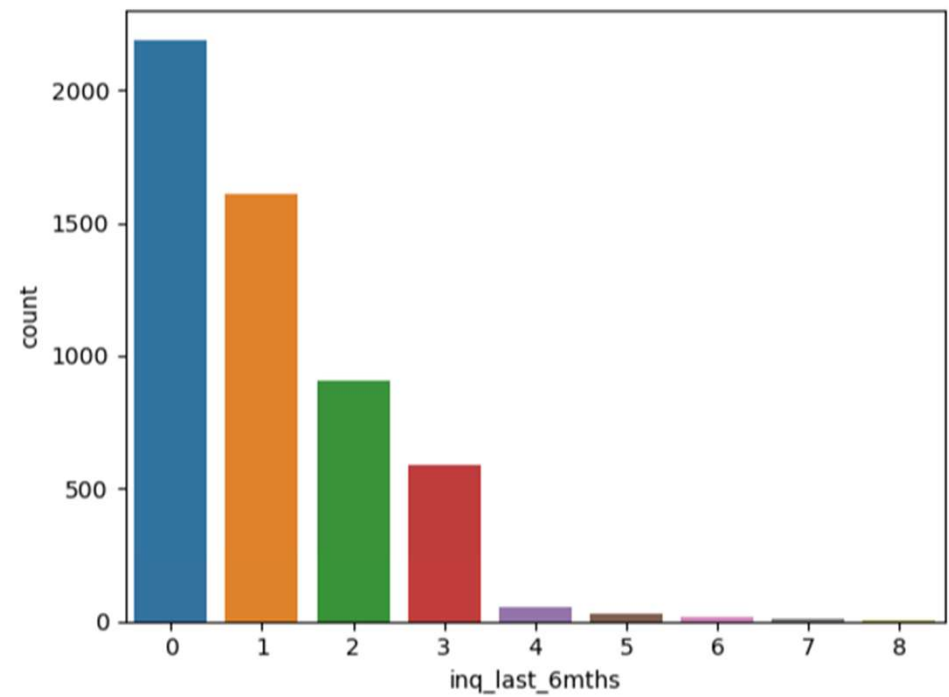
Analysis

Members whose verification status is not verified are more likely to default. This is a good indicator to make sure all the sources are verified when issuing a loan.



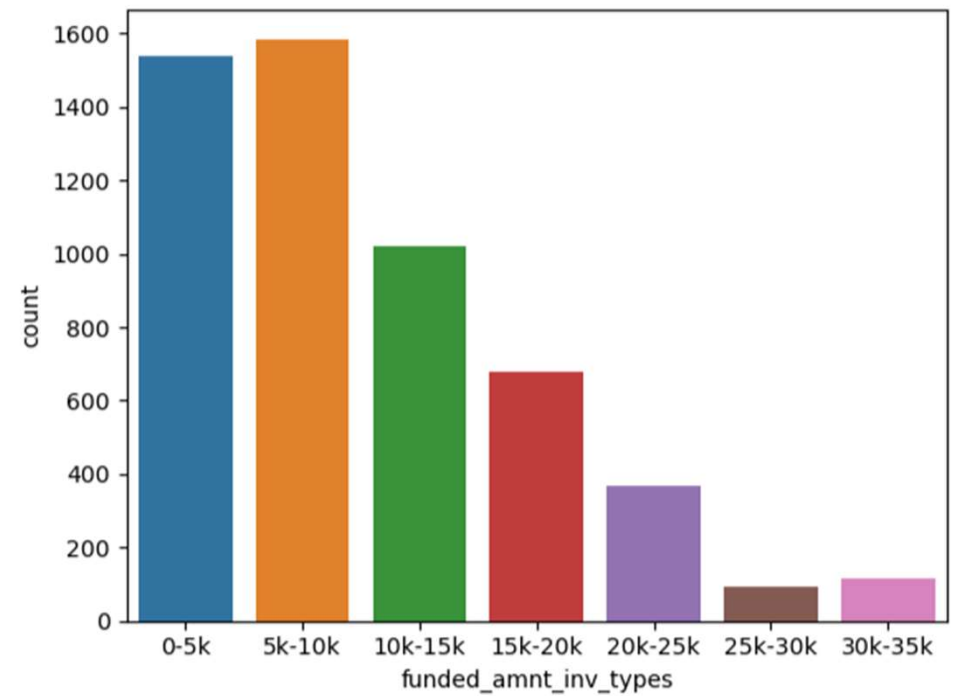
Analysis

One of the strong indicator is how many inquiries are being made over last 6 months. Here its listed that more defaulters tend to make more inquiries.



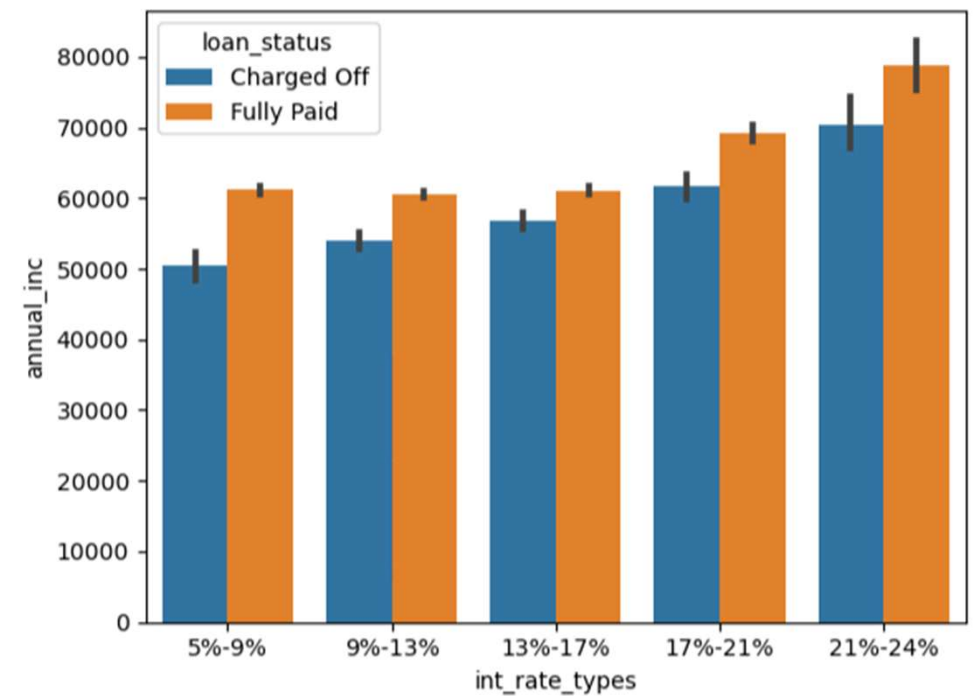
Analysis

More people are defaulters when salary is lesser than 10K. About 50% of defaulters fall into this category.



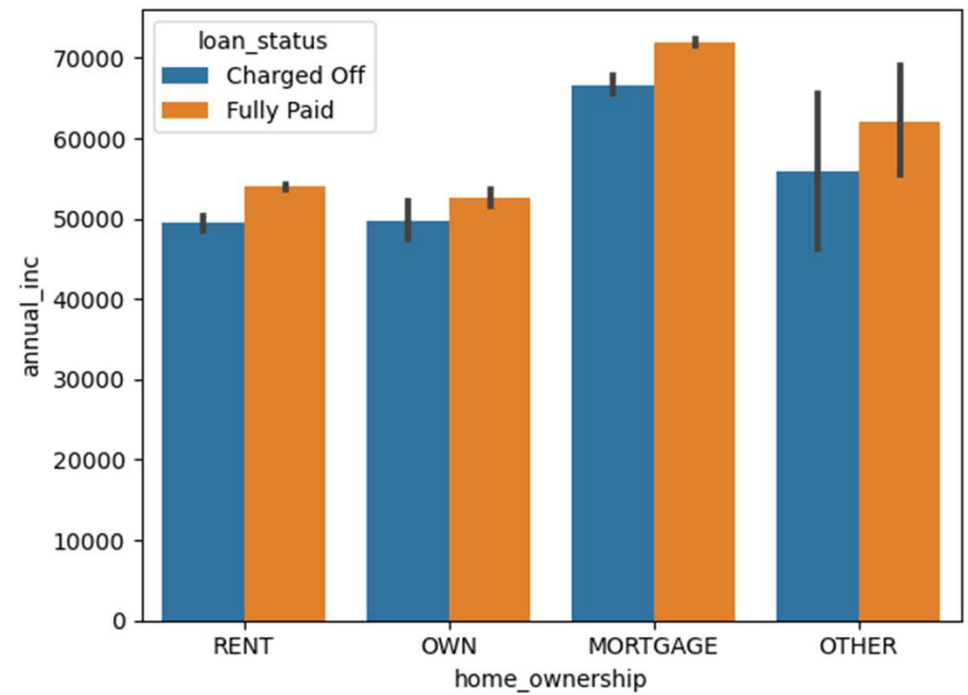
Analysis

When interest rates are higher, there is a very high chance of defaulting on a loan. When the interest is around 21 to 24% has the highest defaulters.



Analysis

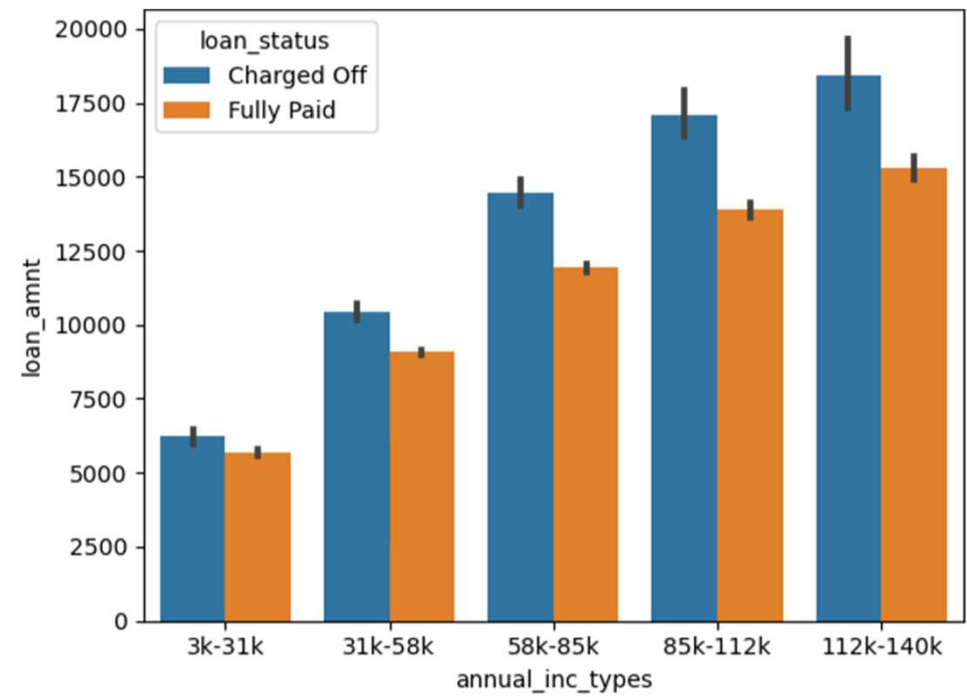
Owning a home reduces the defaulter probability. If there is already a mortgage, then defaulter probability is higher.



Analysis

Even though annual income is way higher, due to very high loan amount, the defaulters percentage is higher.

Loans can be issued to members who have high income and ask for less loan amounts.



Conclusion

From the data provided, there are 15% default rate in the data

The Probability of default cases are

- a. When applicant has house ownership as "rent".
- b. When applicant is from grade B at the 5 level.
- c. When applicant has purpose of debt clearance.
- d. When applicant term is 36 months.
- e. Most of the default cases when the applicant details are not verified.
- f. When the fund amount by investor is between 5k to 10k.
- g. When the loan amount between 5k to 10k.
- h. When the public records are 0.
- i. When the annual income types between 31k to 50k
- j. When the dti is between 12 to 18
- k. When monthly installment between 145 to 274
- l. When the interest rate is 13% to 17%
- m. When the application of the length 10