



LENDING CLUB CASE STUDY

Submitted by

SHAMBHAVI NIMBHORKAR
TUSHAR SAHA



Problem statement

A consumer finance company that specializes in providing various types of loans to urban customers wants to gain a deeper understanding of the driving factors behind loan defaults. Specifically, they are interested in identifying the key variables that serve as strong indicators of whether a borrower is likely to default on a loan. This analysis is crucial for the company as it can leverage this information to improve its portfolio management and risk assessment processes.

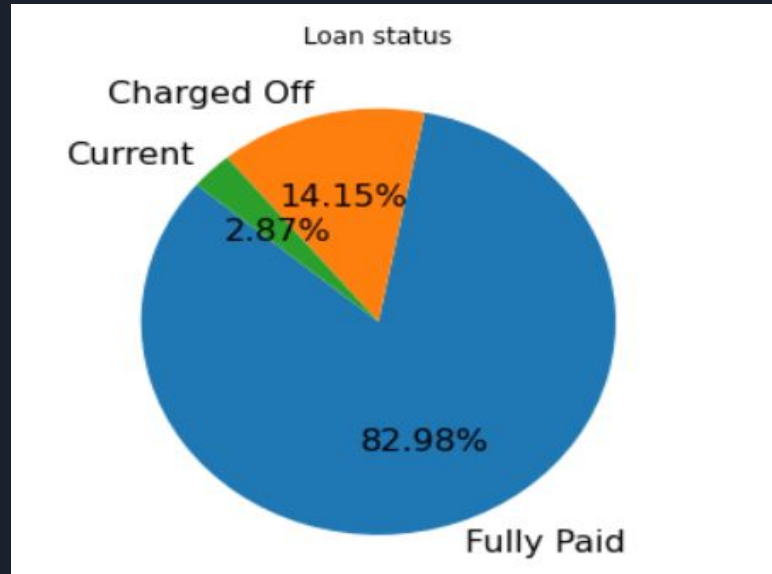
Conduct a comprehensive analysis to identify and understand these driving factors.

Data pre-processing

- The original size of the data is (39717,111).
- During our initial analysis, we realized that not all parameters were essential for this particular analysis, so we have dropped them.
- For columns with missing values, we applied the following approaches:
 - If the missing values were greater than 40%, we dropped the column.
 - If the missing values were less than 40%, we imputed the missing values with the appropriate statistical measure (median/mode).
- We corrected the data types for some variables that were not correct in original dataset.
For example-revol_util,int_rate
- There are no duplicates in this dataset.
- After data cleaning and preprocessing, we have shape of dataframe as (39667,27)

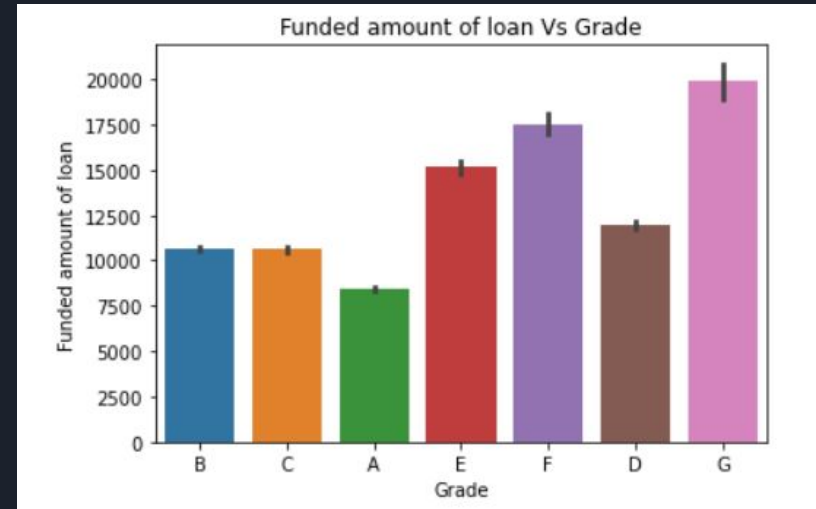
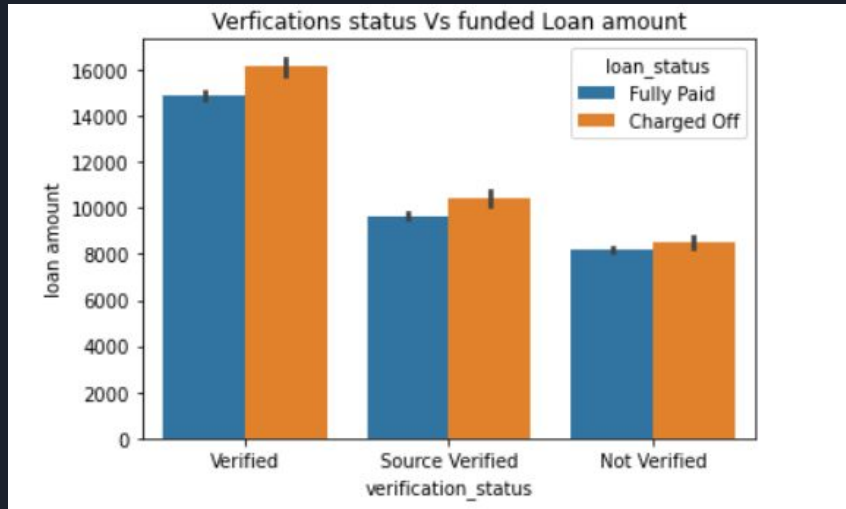
Understanding of Business Objective

- The below graph illustrates the loan status of borrowers. Our focus is on identifying patterns associated with defaults. Therefore, we will exclude borrowers with a loan status of current, and instead, concentrate on those categorized as charged off or fully paid.



Analyzing Funded amount

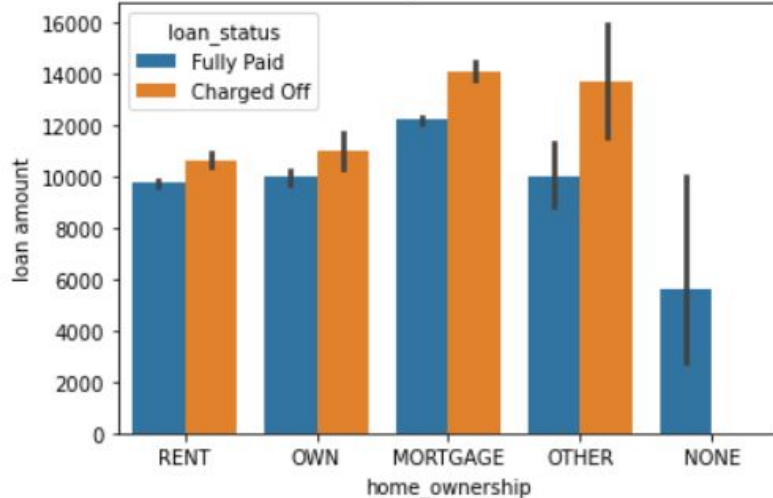
- Grade G borrowers have taken high loan as compared to other grades.
- Though verification is done , as the loan amount increases the chances of being charged off also increases.



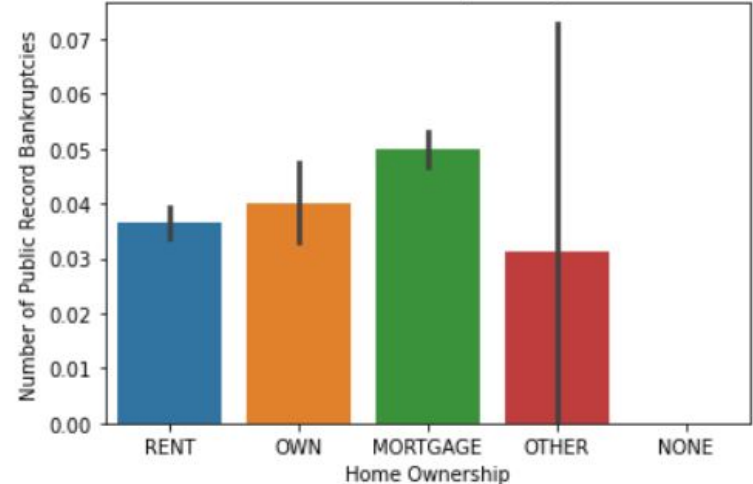
Analyzing Home ownership

- Mortgage is leading among home ownership category in terms of loan amount..
- Borrowers having mortgage have high possibility of Bankruptcies.

Term and home ownershipwise Loan amount Distribution

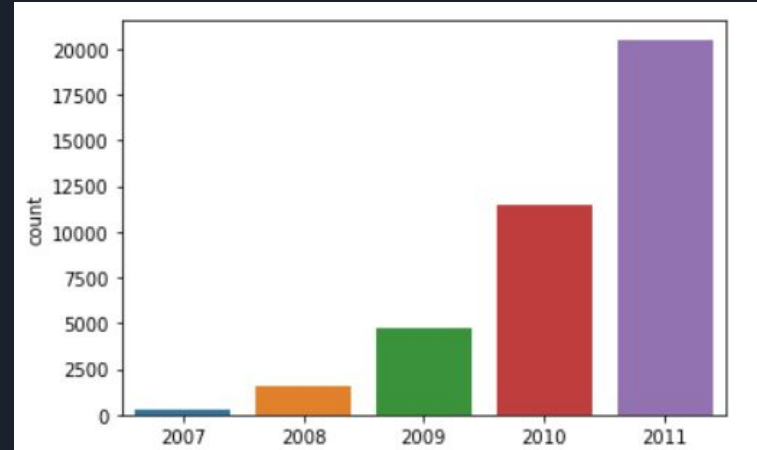
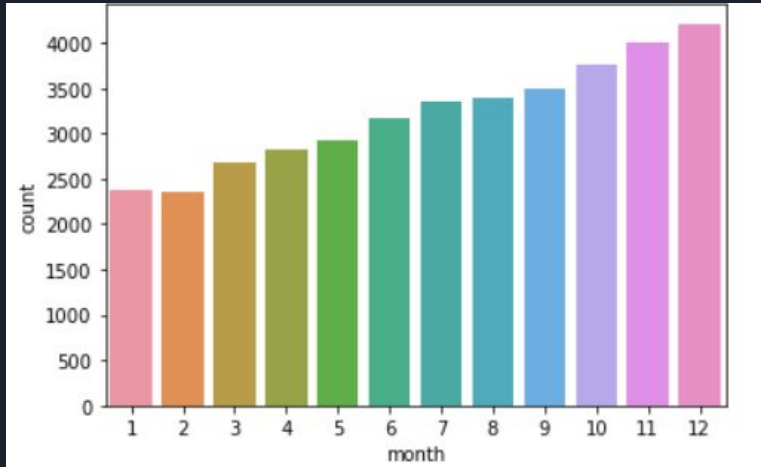


Number of Public Record Bankruptcies by Home Ownership



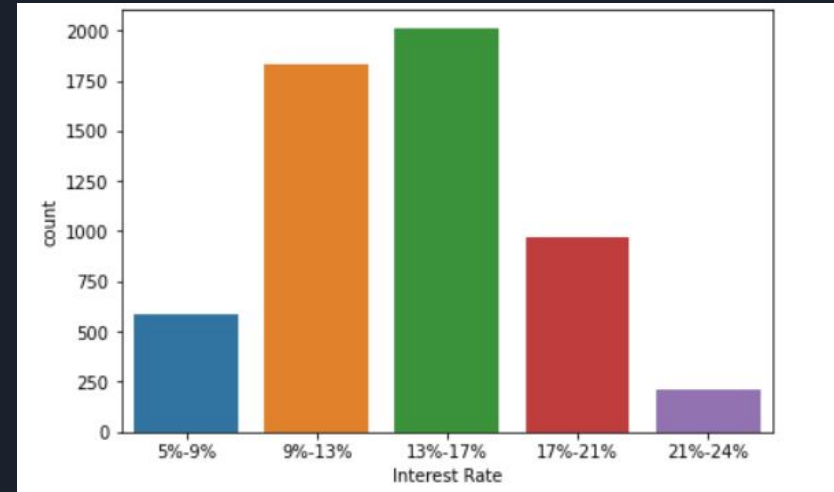
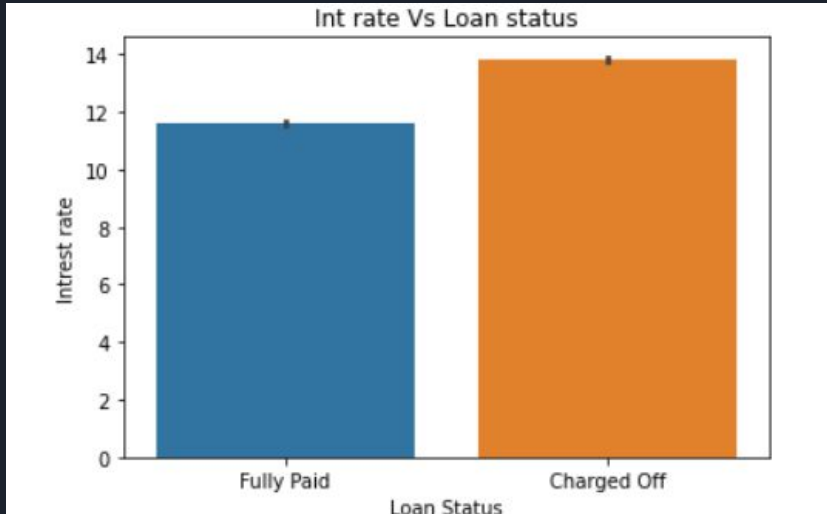
Analyzing Issue Date

- Graph shows that there is increase in number of borrowers over the year and in the month of Dec maximum loans are approved.



Analyzing Interest rate

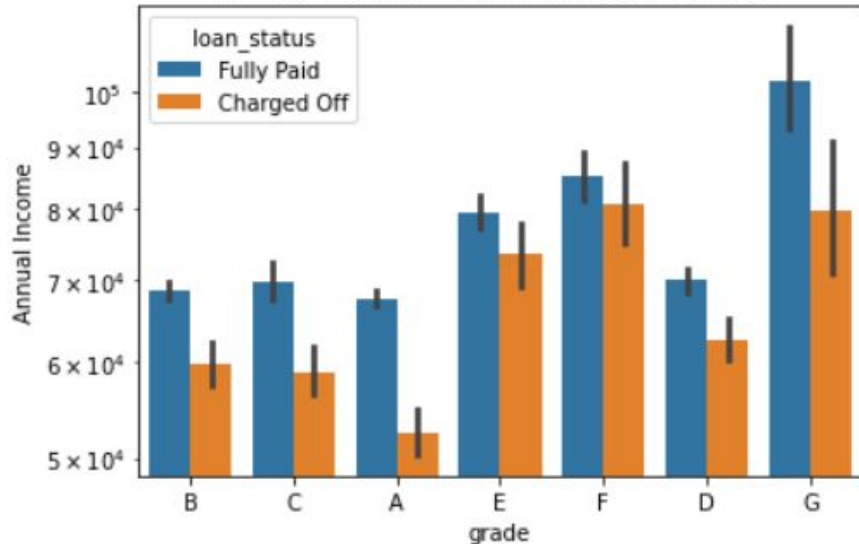
- Graph shows when interest rate is high there are more chances of being charged off.
- Borrowers who are charged off fall under interest rate of 13-17%.



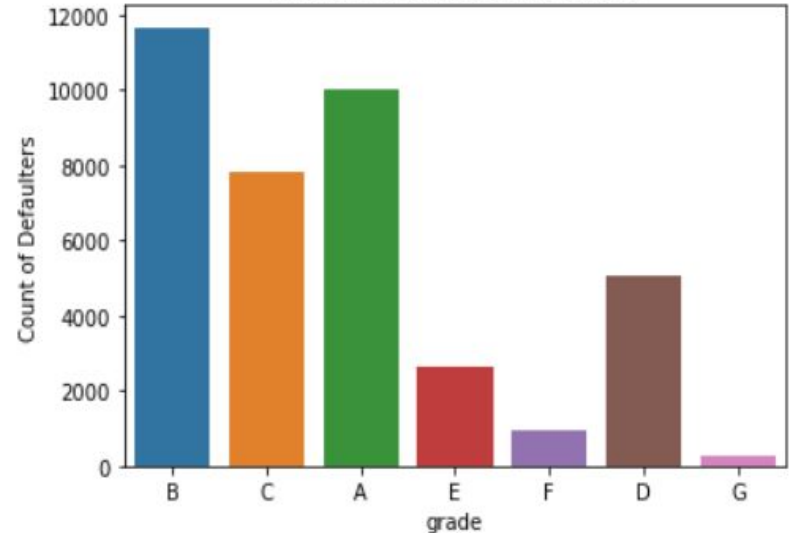
Analyzing Grade and Income wise loan status

- Borrowers belonging to G grade have the maximum annual income and majority of them have paid loan
- Grade B has maximum number of Defaulters.

Income-Grade wise loan status Distribution



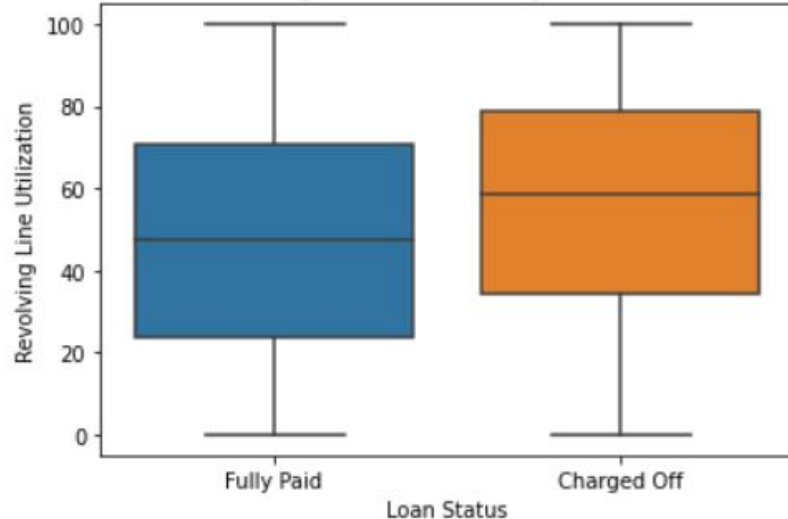
Grade wise count of Defaulters



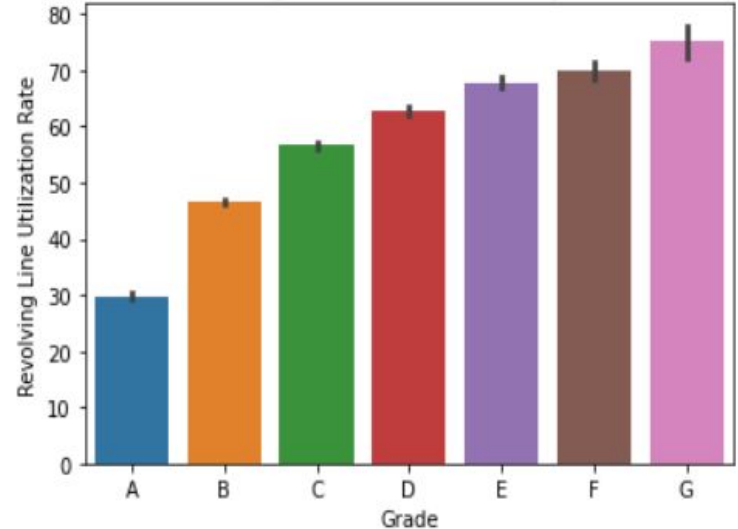
Analyzing of Revolving Line Utilization rate

- Borrowers having high Revolving line utilization have more possibility of getting charged off.
- The revolving line utilization rate having value of 30 shows that borrowers exhibiting responsible behaviour as they are using only 30% of their available revolving credit.

Revolving Line Utilization by Loan Status

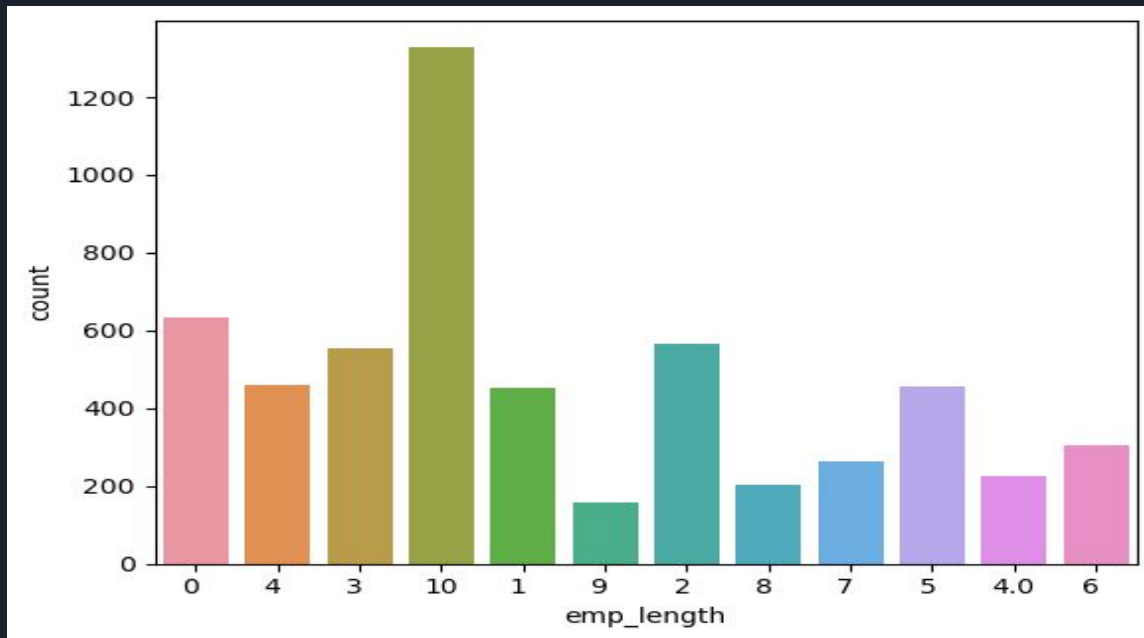


Revolving Line Utilization Rate by Grade



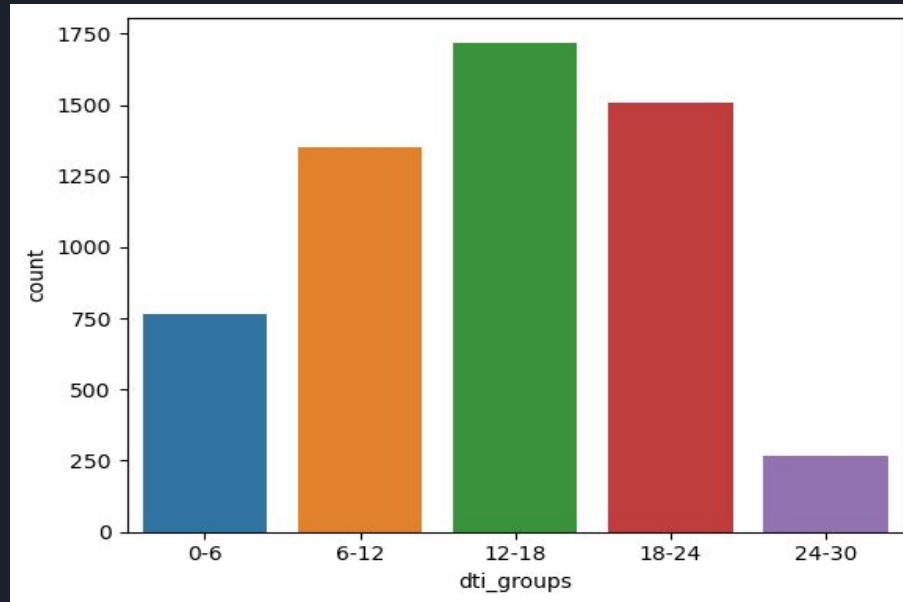
Analyzing employment length

- Default rate is highest where employment length is 10+ years



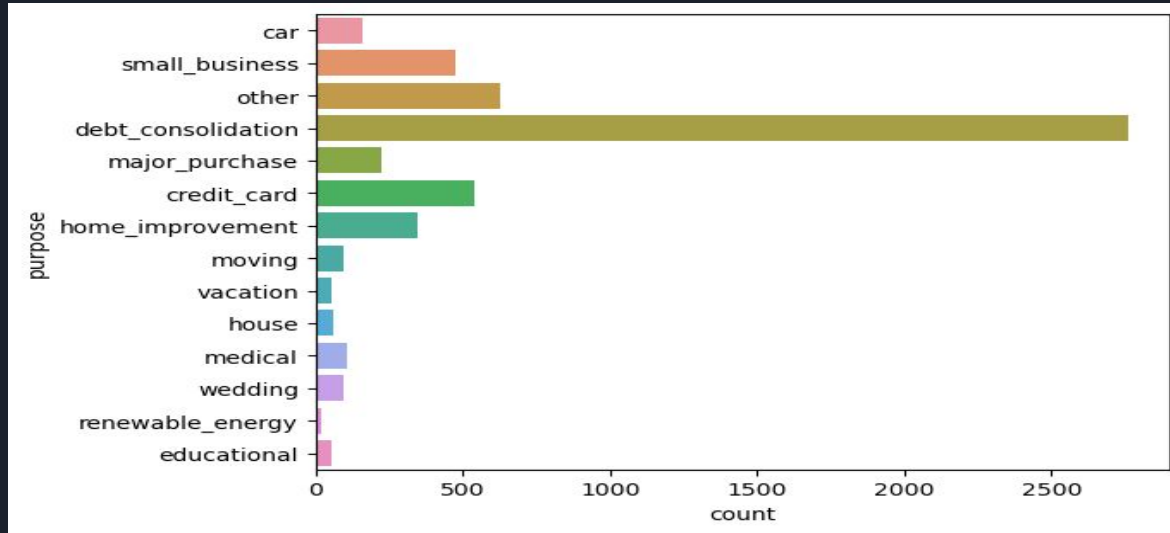
Analyzing DTI

- DTI is a ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
- Default rate is maximum for dti group 12-18



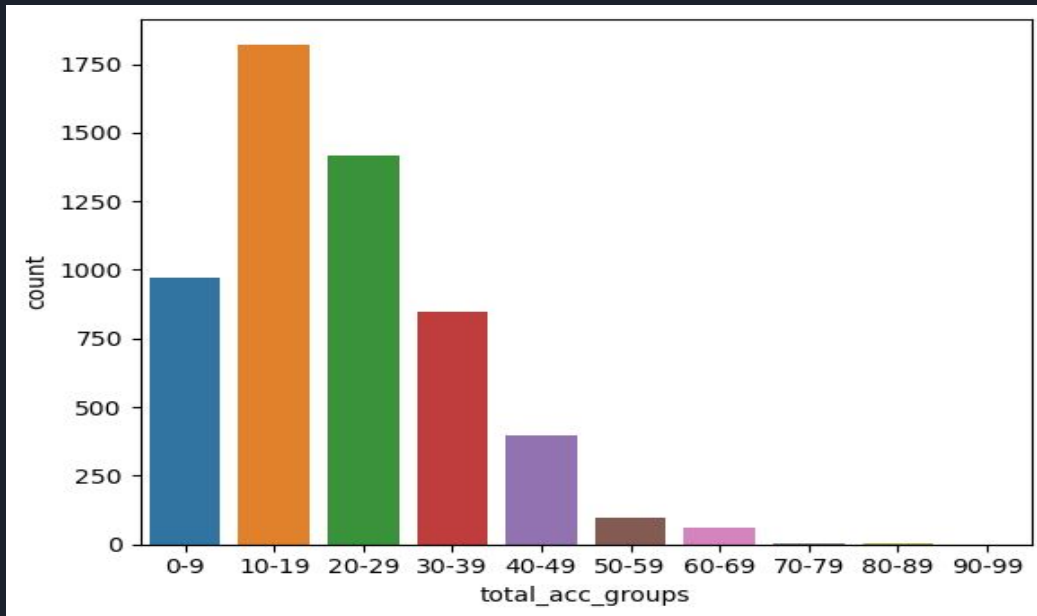
Analyzing purpose

As we can see in the graph below that most default happens when loan is taken to repay another loan



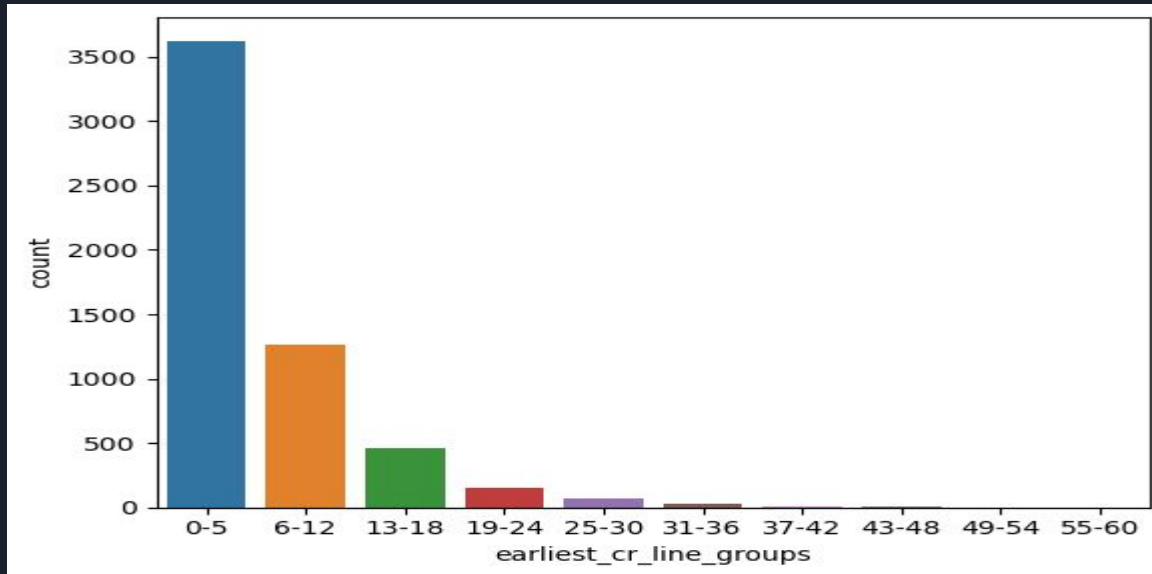
Analyzing total accounts

Most default happens when the customers have around 10-19 loan accounts



Analysing credit line age

Older the credit line, more are the chances of loan repayment , that's why we can see maximum default in initial years.





Observation

- Higher loan amounts are associated with a higher chance of default.
- The interest rate range of 13-17% has the highest default rates.
- Borrowers with high incomes have fully paid their loans.
- Borrowers belonging to Grade G have the highest annual income, and the majority of them have paid their loans.
- Borrowers belonging to Grade B are prominent among defaulters.
- The default rate is highest when loans are taken for small businesses.
- The default rate is lowest when loan taken for vacation purpose.
- Borrowers with rented or mortgaged homes have a high chance of defaulting.
- Applicants with a recent credit history and having 10-19 loan accounts also have a high chance of defaulting.