Lending Club Case Study

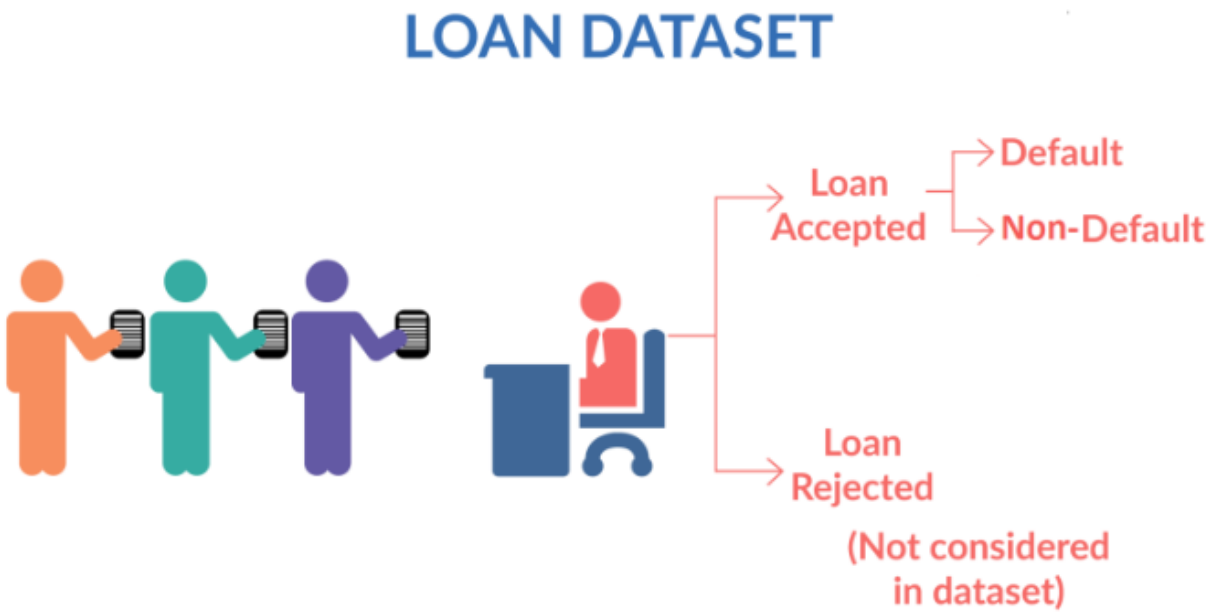
Problem Statement

Business Understanding

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 must decide 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



When a person applies for a loan, there are two types of decisions that could be taken by the company:

Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:

1. Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
2. Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
3. Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

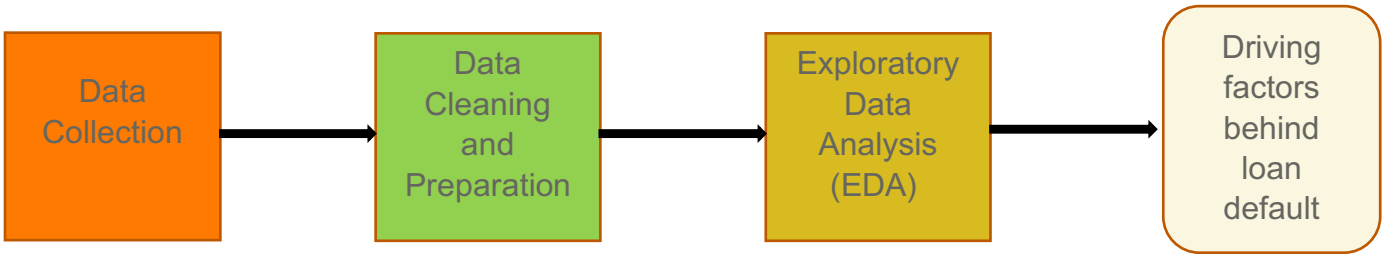
Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Business Objectives

The Company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Solution Approach:

Exploratory Data Analysis (EDA) is an approach to analyze the data using visual techniques. We will use EDA to understand the driving factors behind the loan default.



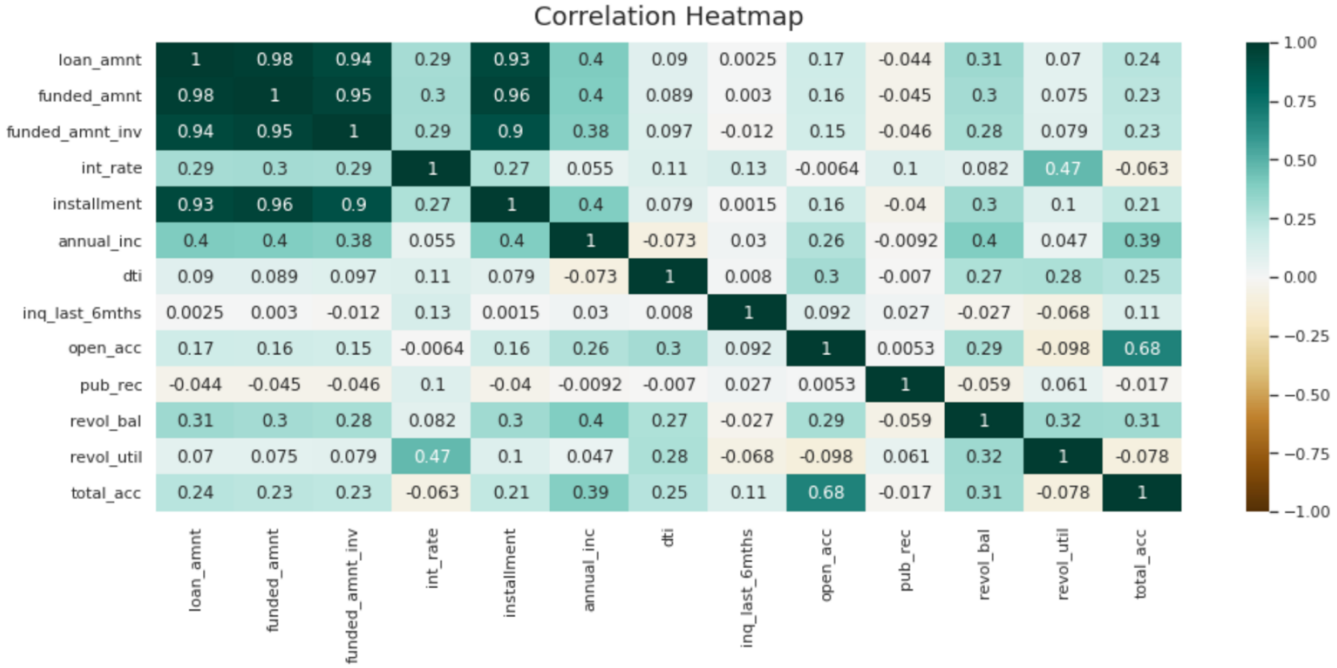
Data Cleaning:

- Check missing values columns and take appropriate action
- Check unique values columns and take appropriate action
- Analyze columns which are not useful for analysis, drop these columns
- Analyze columns which needs value formatting like remove % symbol from int_rate

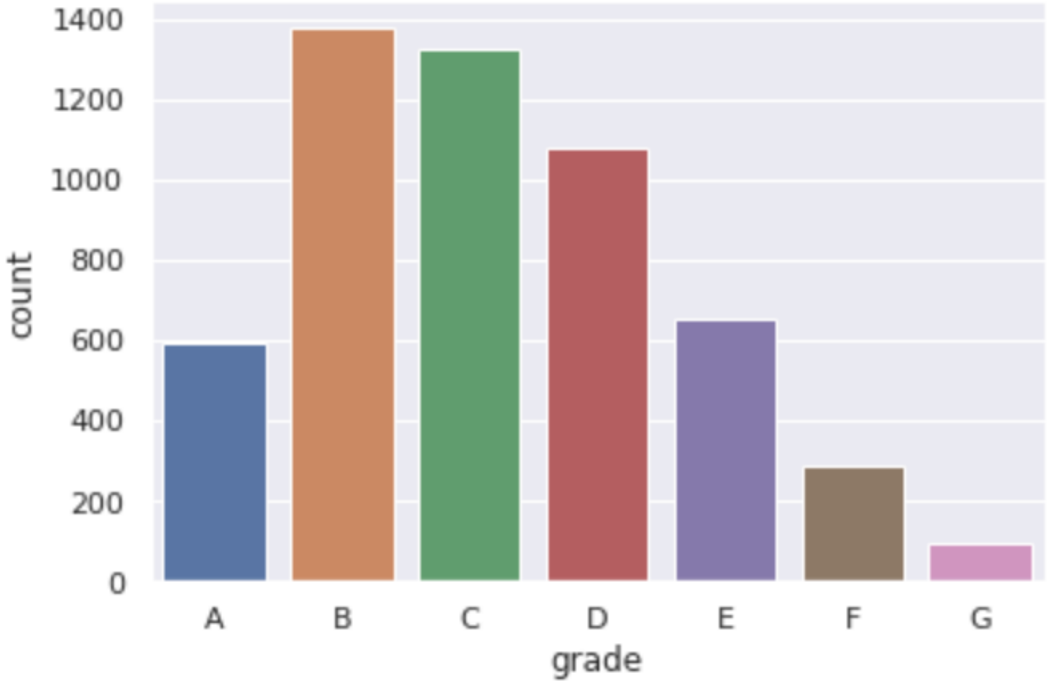
- Convert data which is needed in numeric format
- Group data with bins and create new columns if required for analysis

Exploratory Data Analysis (EDA):

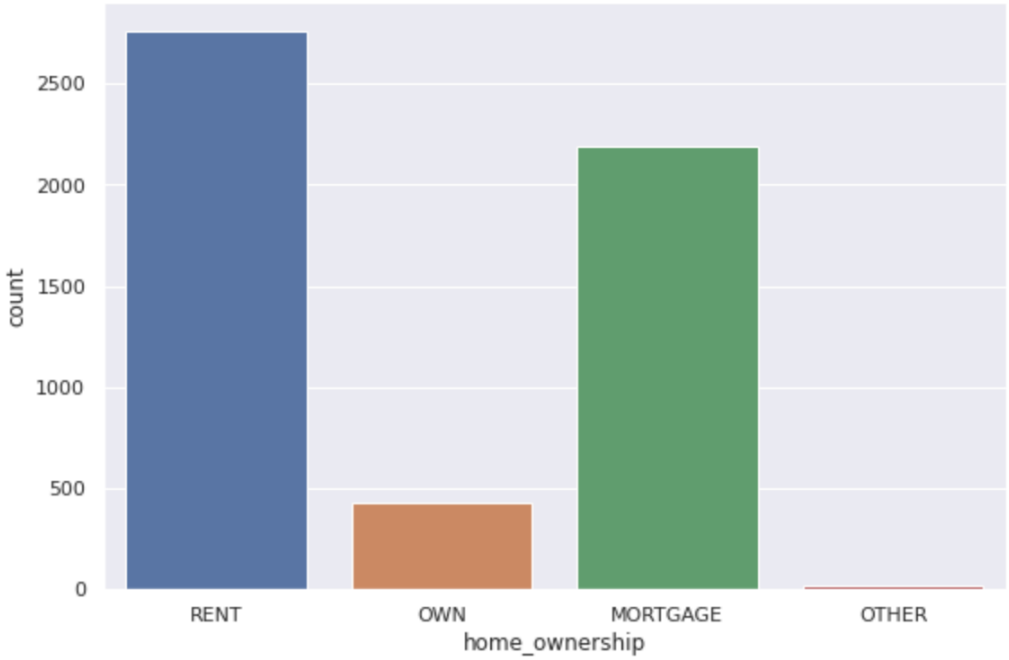
Check Correlation



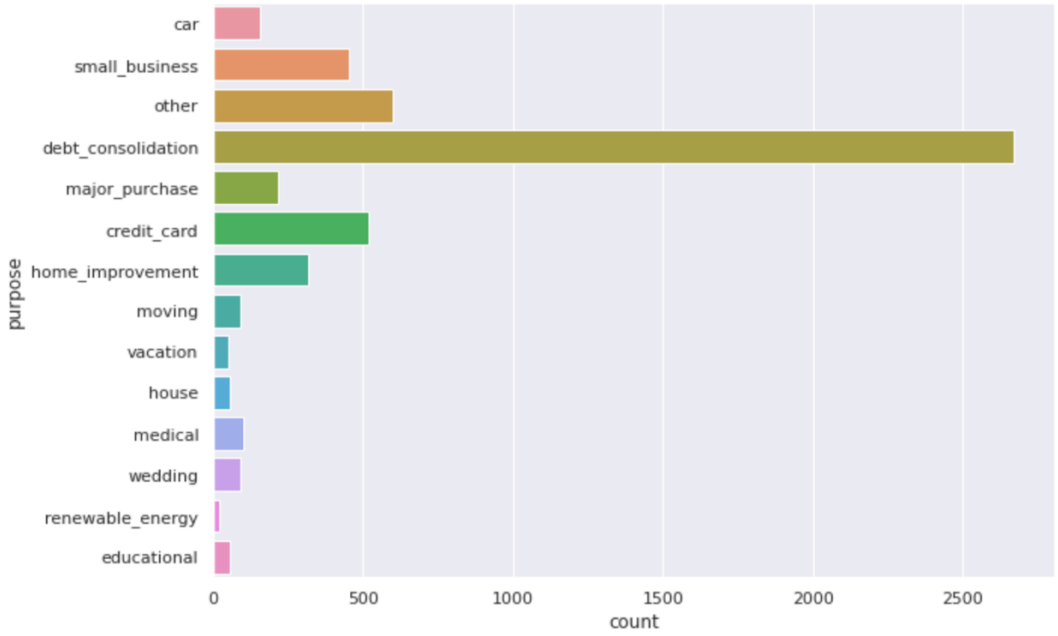
 # Grade vs Charged Off



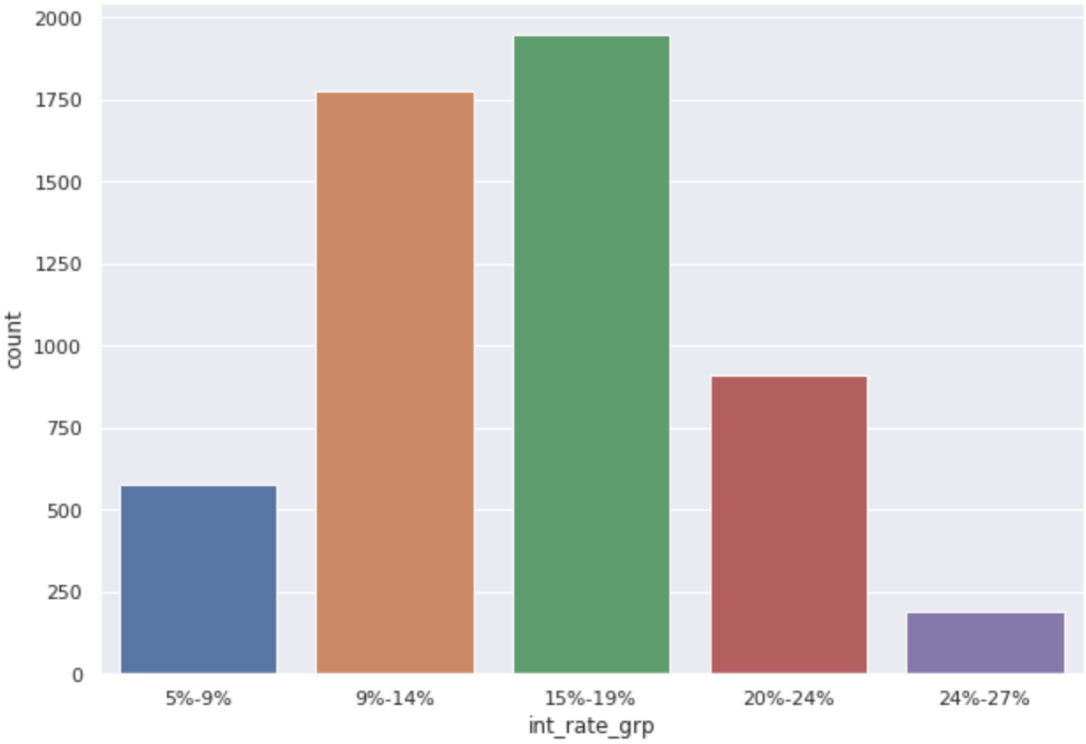
 # home_ownership vs charged off



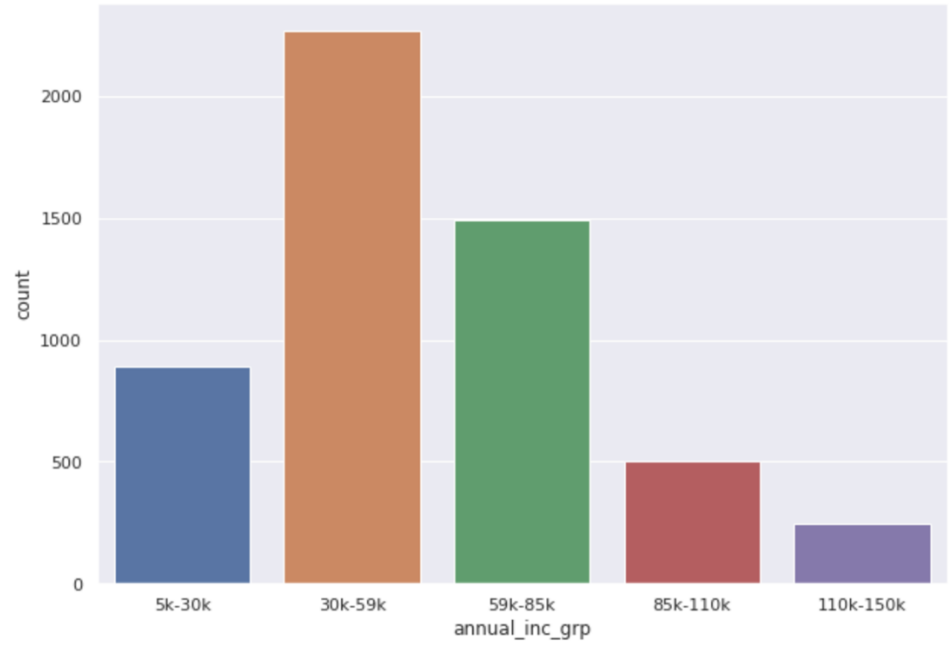
purpose vs charged off



interest rate vs charged off

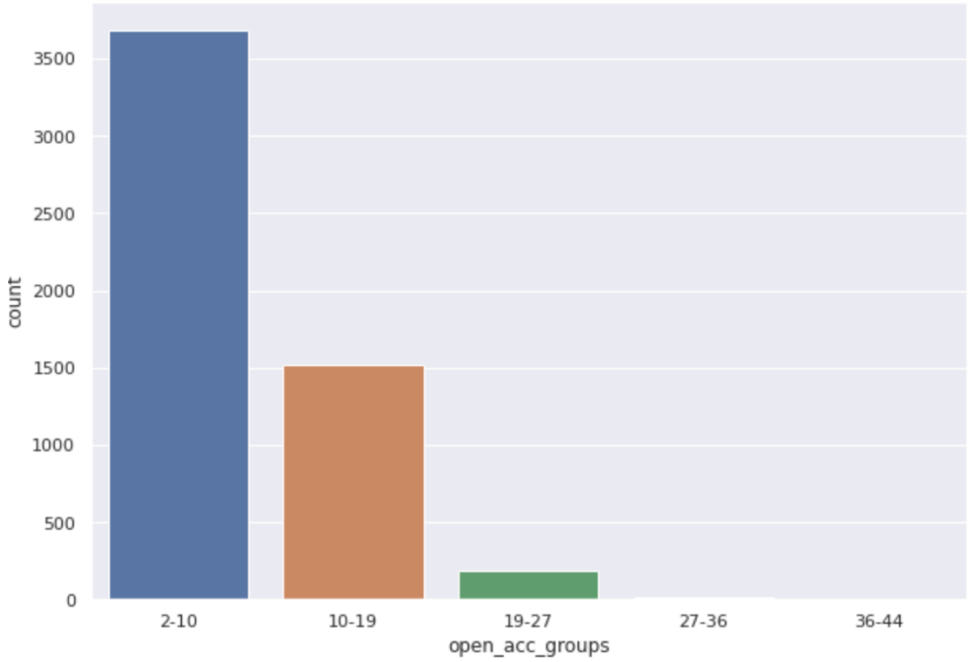


Annual Income vs Charged Off

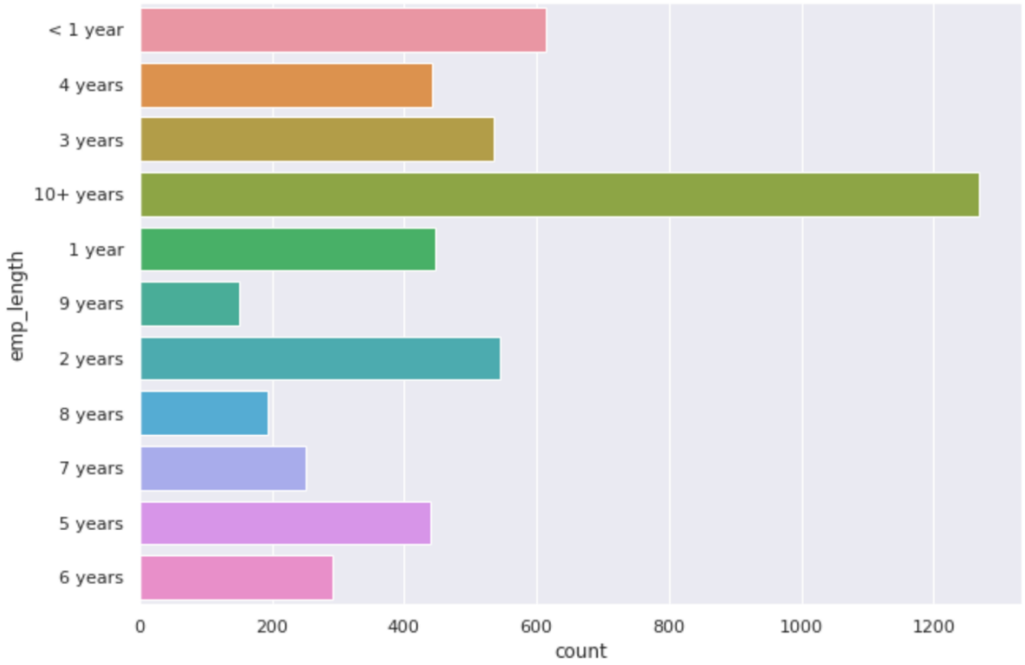


Open Accounts vs Charged Off

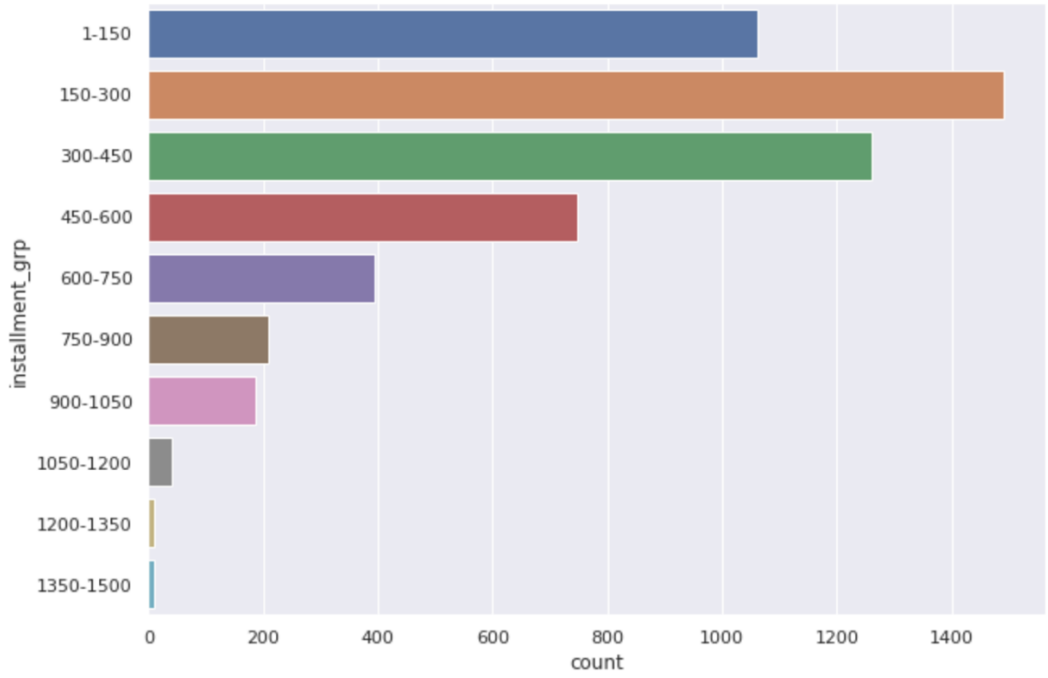
Open Accounts (The number of open credit lines in the borrower's credit file)



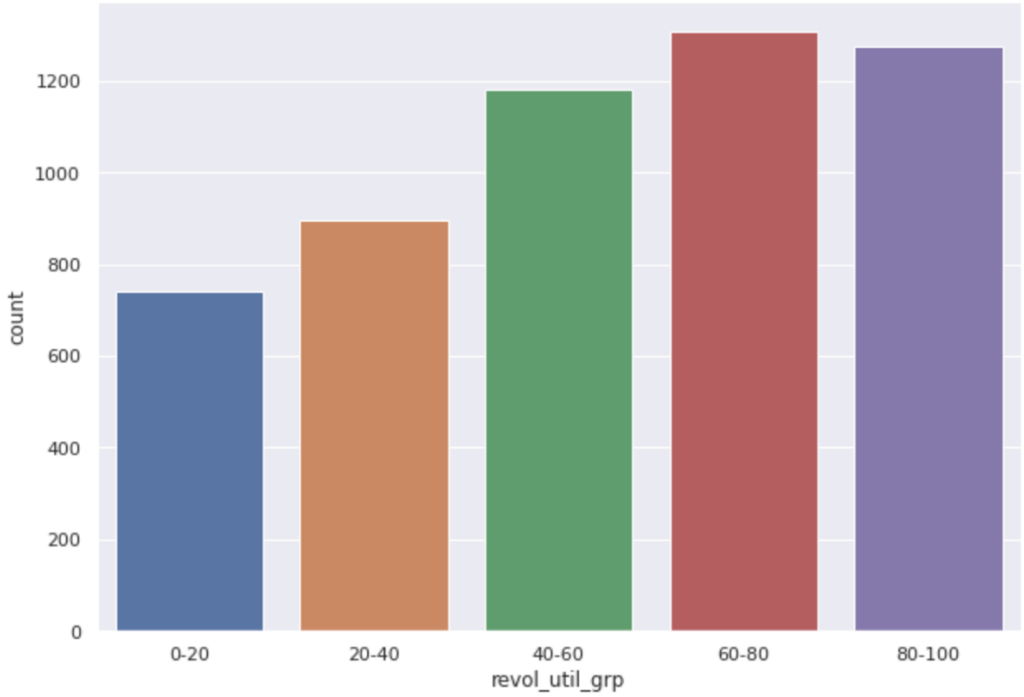
Employment Length vs Charged Off



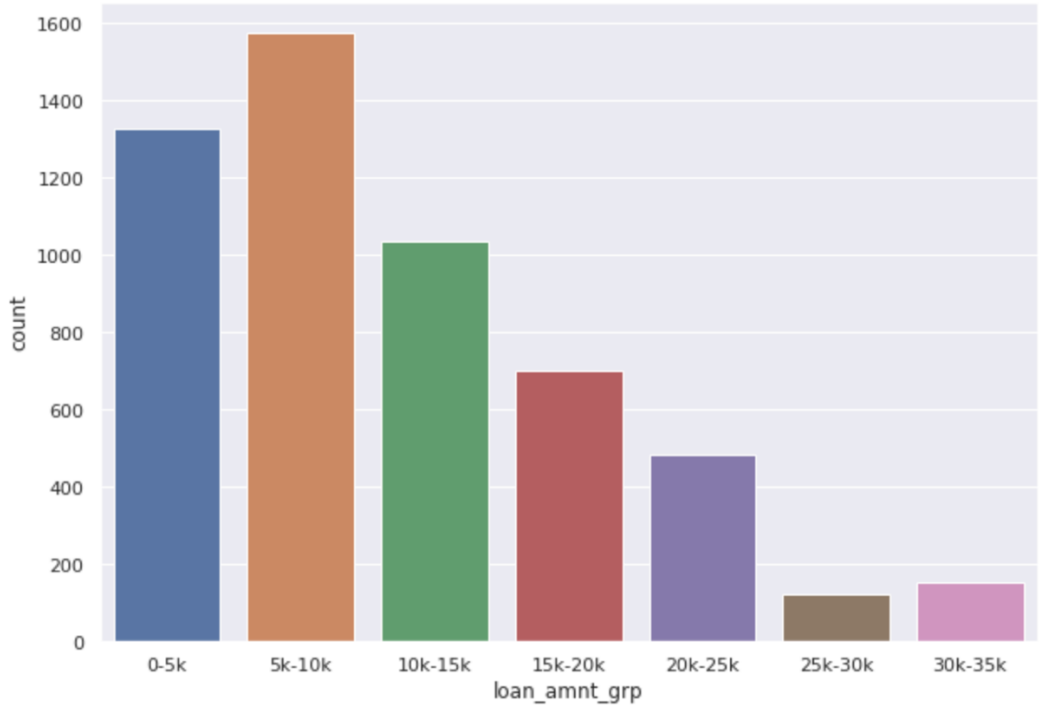
Installments vs Charged Off



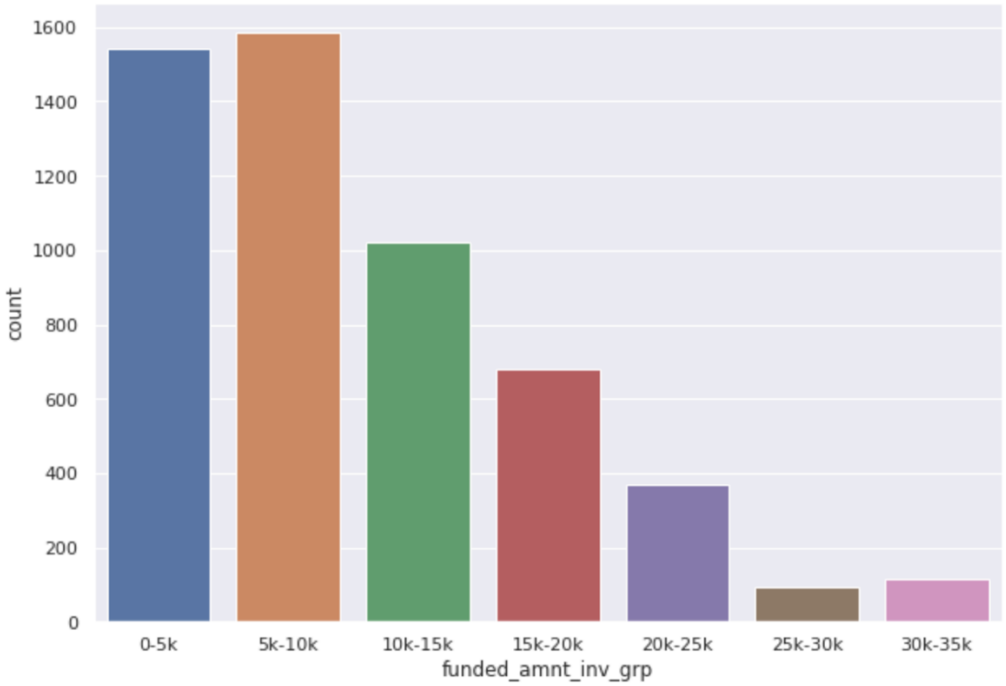
Revolving Line Utilization Rate vs Charged Off



Loan Amount vs Charged Off

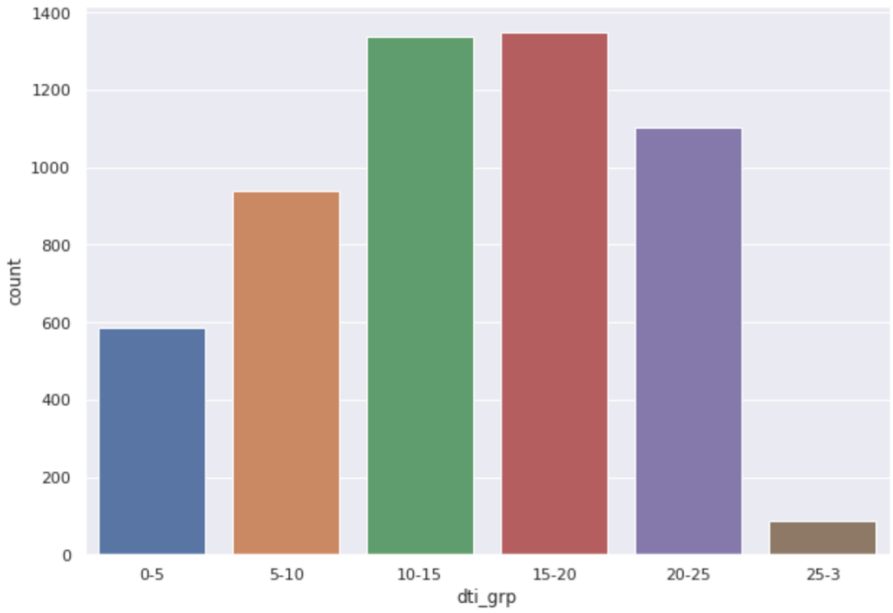


Total amount committed by investors vs Charged Off

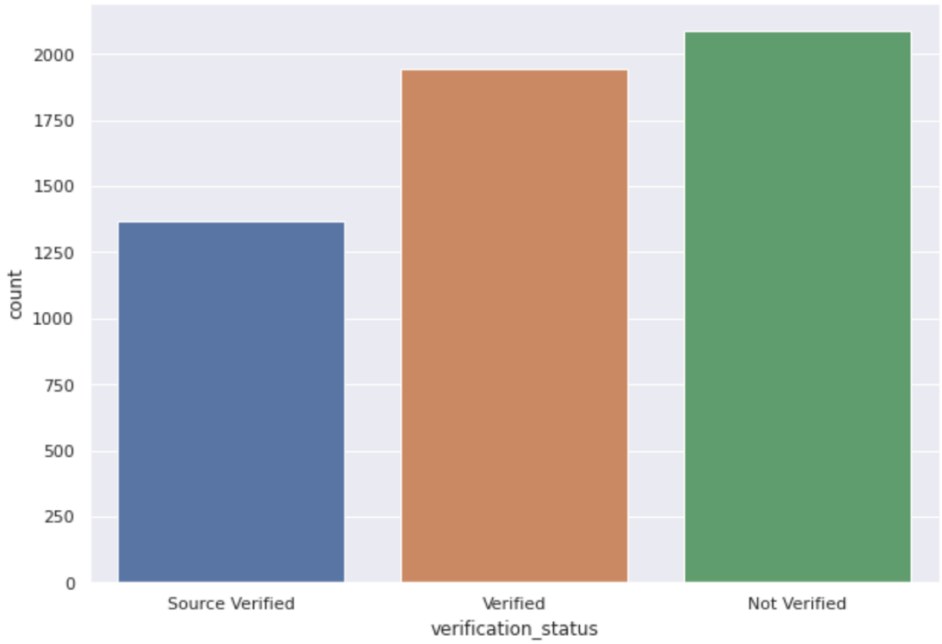


Dti vs Charged Off

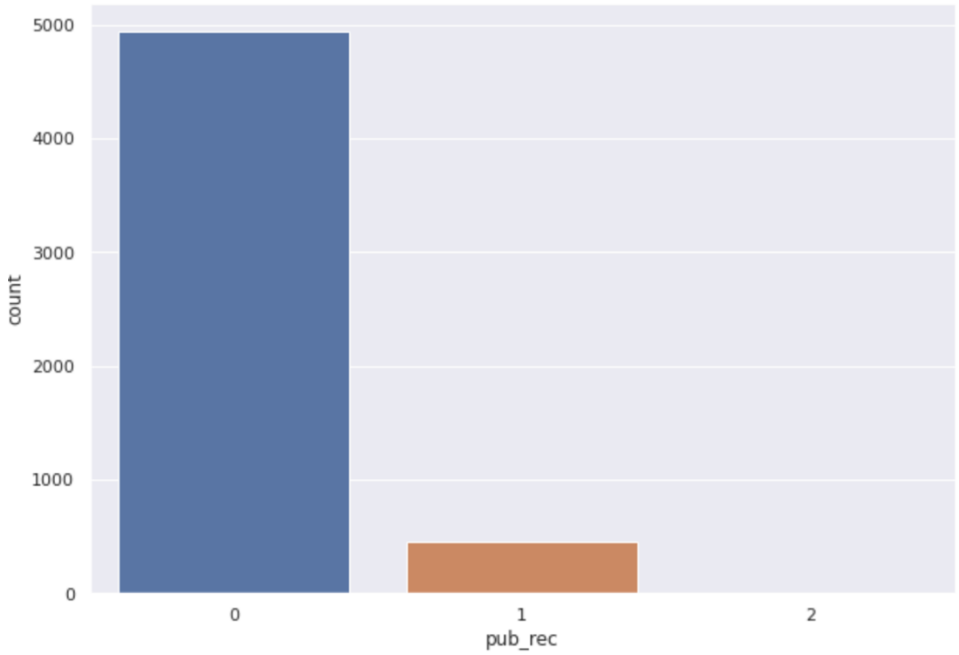
Dti- A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and requested LC loan, divided by the borrower's self-reported monthly income.



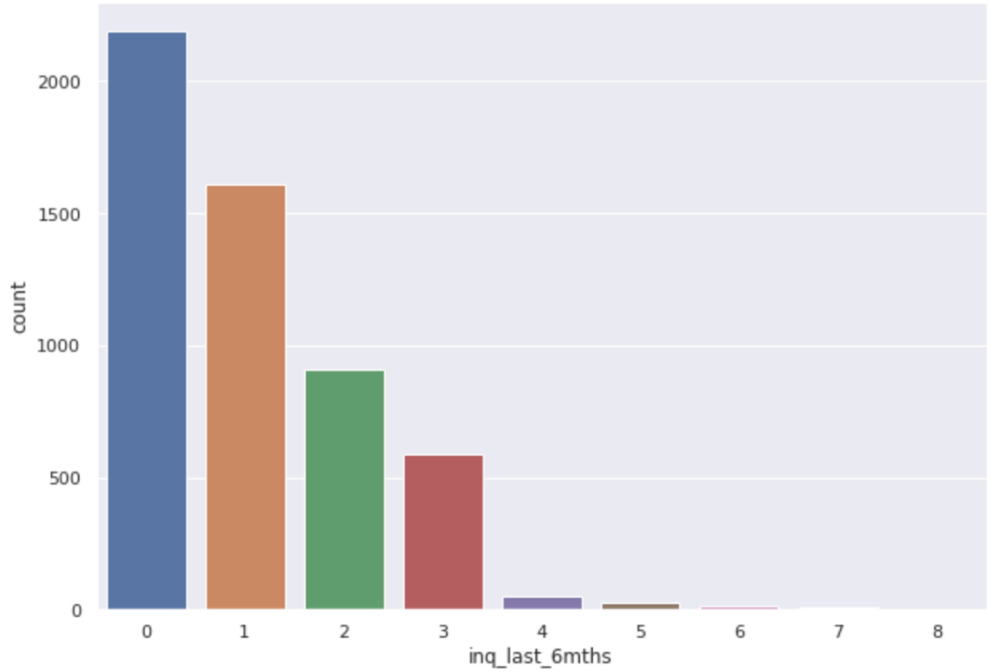
Verification Status vs Charged Off



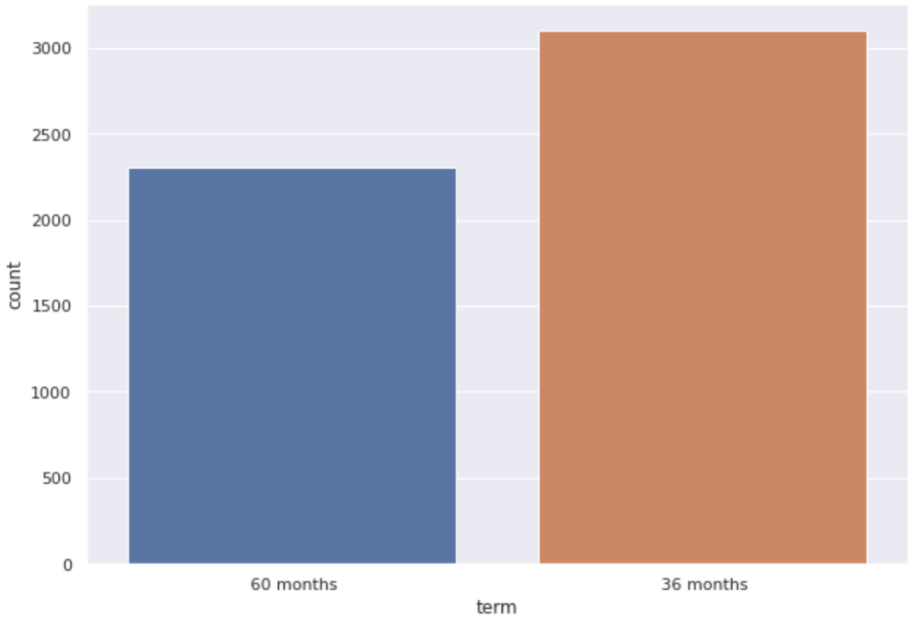
Number of derogatory public records vs Charged Off



Number of inquiries in past 6 months vs Charged Off



Term vs Charged Off



Driving factors (or driver variables) behind loan default:

Below are the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

- If grade (LC assigned loan grade) is B
- If people are renting house
- If loan purpose is debt consolidation
- if interest rate is between 15%-19%
- if annual income is between 30k-59k
- if number of open credit lines in the borrower's credit file is between 2-10
- if employment length >= 10 years
- if installment is between 150-300
- if Revolving line utilization rate is between 60-80
- if loan amount is between 5k-10k
- if total amount committed by investors for that loan at that point in time. is between 5k-10k
- if dti is between 15-20 (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.)
- if verifications tatus is Not Verified
- if number of derogatory public records is 0
- if number of inquiries in past 6 months (excluding auto and mortgage inquiries)) is 0
- if loan term is of 36 months