

Lending Club Case Study

Exploratory Data Analysis Case Study

by

Vedant Tibrewal

Vishnu Vinod V K

Agenda

- Problem Statement
- Data Understanding
- Data Cleaning
- Exploratory Data Analysis
- Conclusion

Problem Statement

Problem:

A consumer finance company which specialises in lending various types of loans to urban customers wants to understand the driving factors (or driver variables) behind loan default. 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

Objective:

Use EDA to understand how consumer attributes and loan attributes influence the tendency to default

Target:

The target column has three variables,

Charged Off: This indicates the person has defaulted the loan

Current: This indicates the loan is currently in progress

Fully Paid: This indicates the loan has been fully paid off

Data Understanding

- Shape of raw data: 39717 rows and 111 columns
- The data is a mix of categorical, numerical, and date columns
- ~50% of the columns have 100% null values
- 9 columns are constant
- 3 columns are 100% unique
- Target variable column is “loan_status”
- We are focused on the value Charged Off and Fully Paid

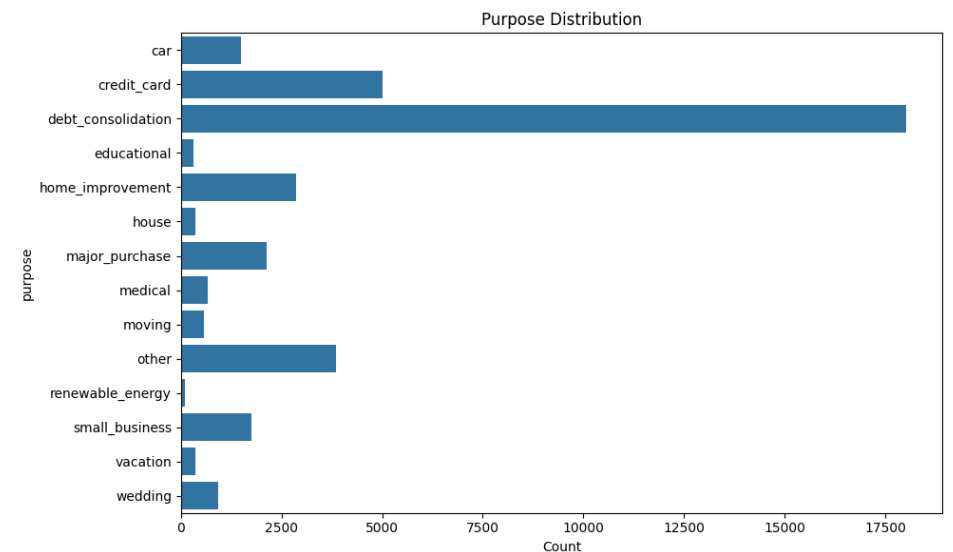
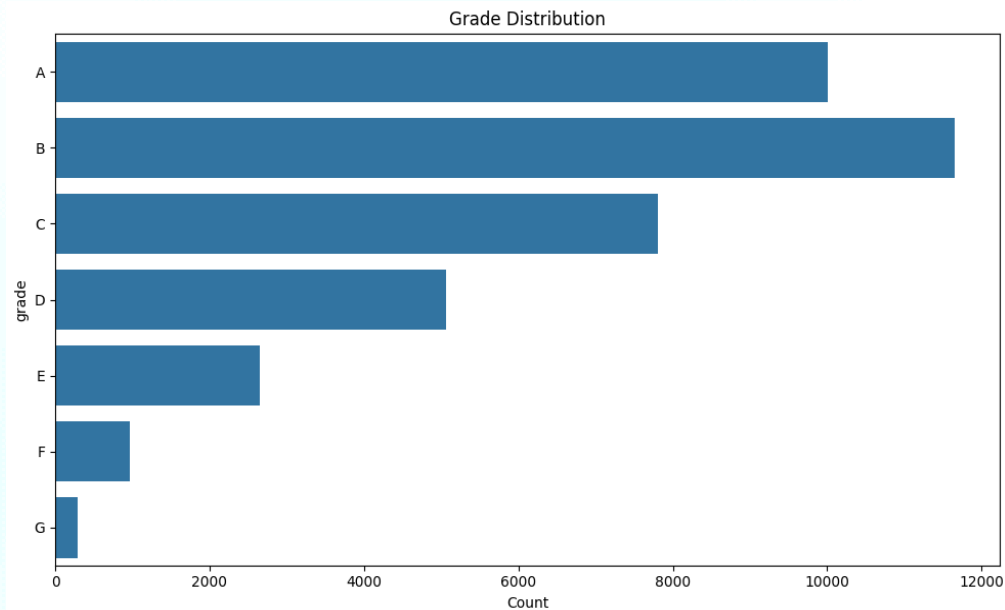
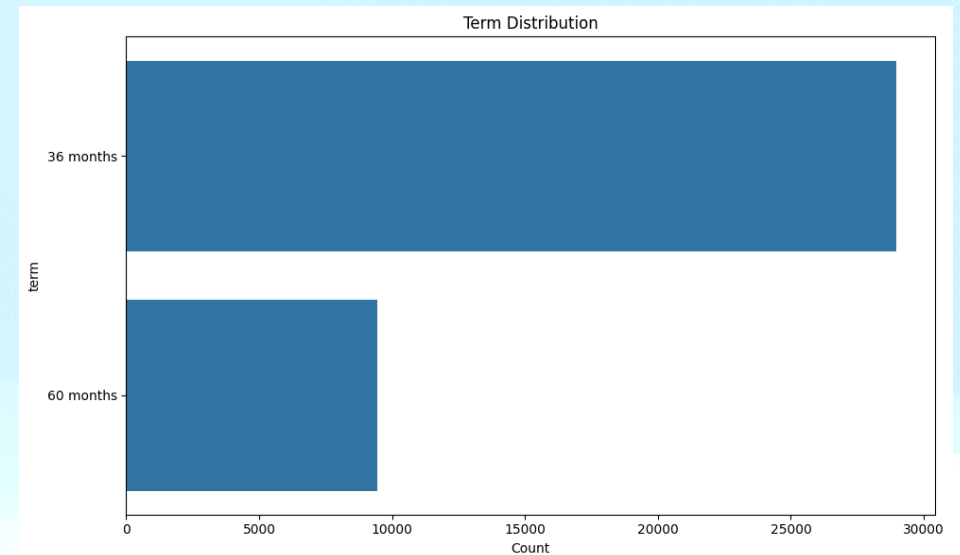
Data Cleaning

- The columns having **>50%** of data null were dropped (columns - 57)
- The columns having **100%** of data unique were removed (columns - 3)
- Constant columns were removed (columns - 9)
- Dropping Columns - These are the columns which have data points once the loan is approved and there are some columns not required for our analysis (columns - 18)
- Dropping the rows with loan status as **Current** (rows - 1140)
- Checking null values and Imputing the rest of the null values
 - emp_legth, pub_rec_bankrupties, home_ownership were imputed
 - reval_util were dropped
- No duplicates were found after sanity check.
- Fixing the data type of the columns was required
 - int_rate, revol_util were % symbol was dropped
 - issue_d, earliest_cr_line the date as modified as date type
 - term where the unwanted space was removed
- The final shape of data: 38527 rows and 24 columns

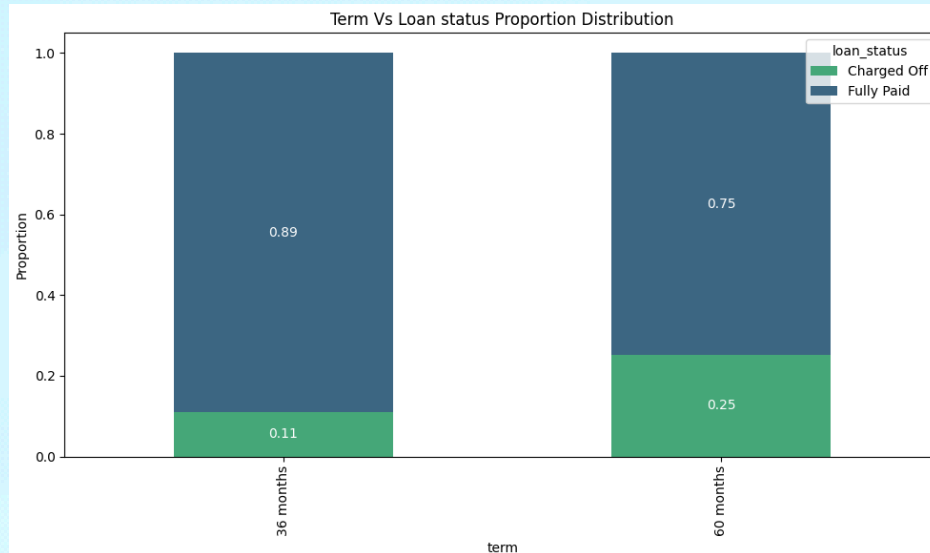
Exploratory Data Analysis

Univariate Analysis

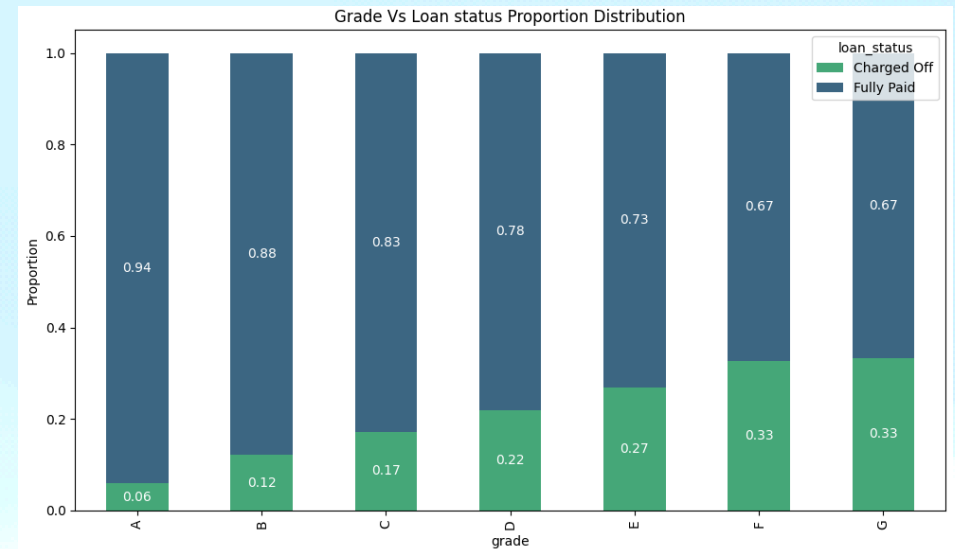
- The most common loan term is 36 months
- Highest number of loans are provided to people falling in grade B, followed by grade A and then grade C
- Mostly the loans are taken for debt consolidation (around 50% of the entire data)



Bi-variate Analysis

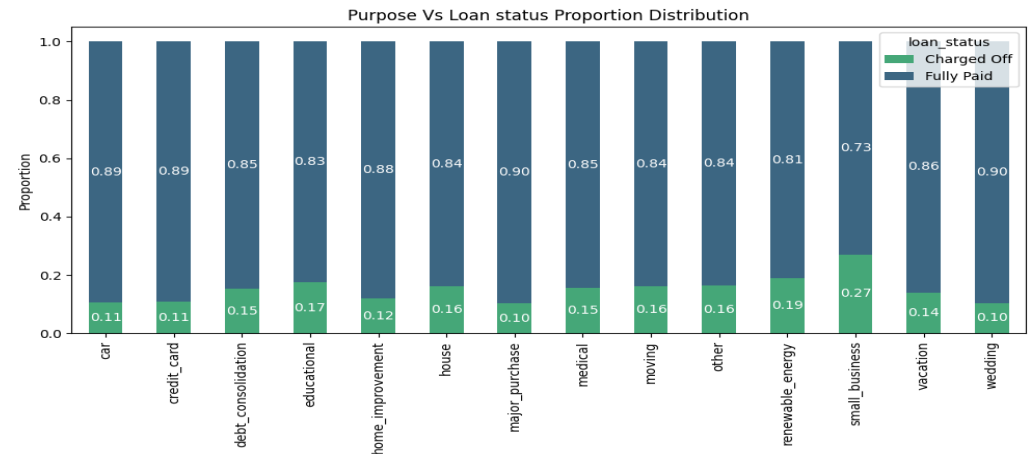


- For loan term the proportion of charged off for term 60 months is double of 36 months

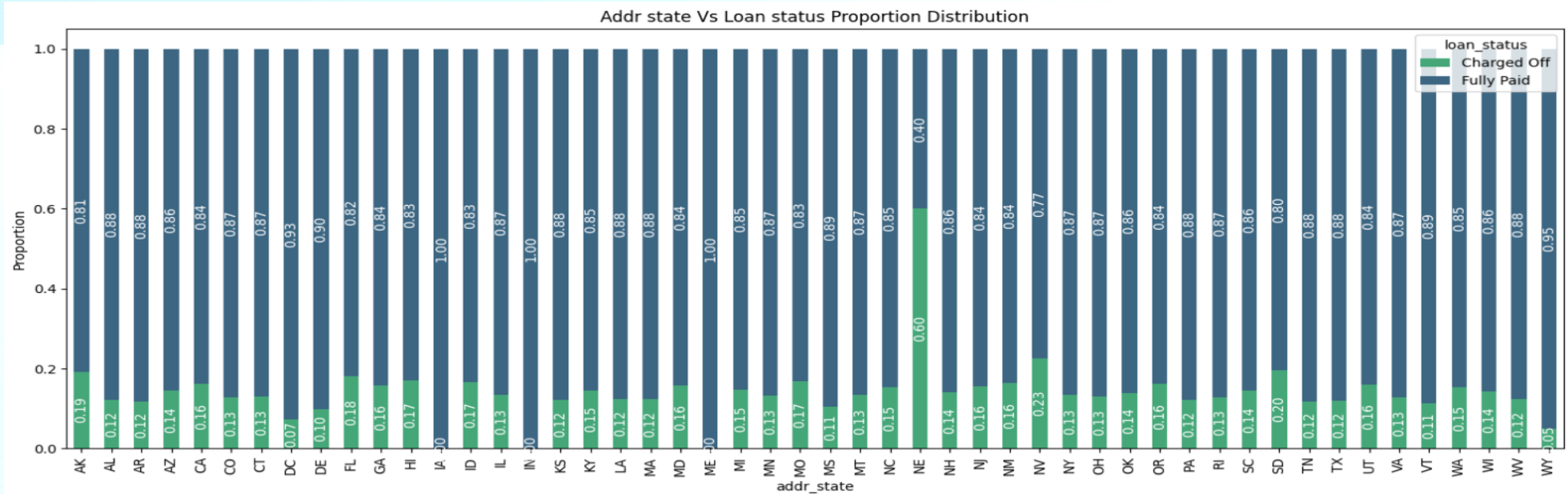
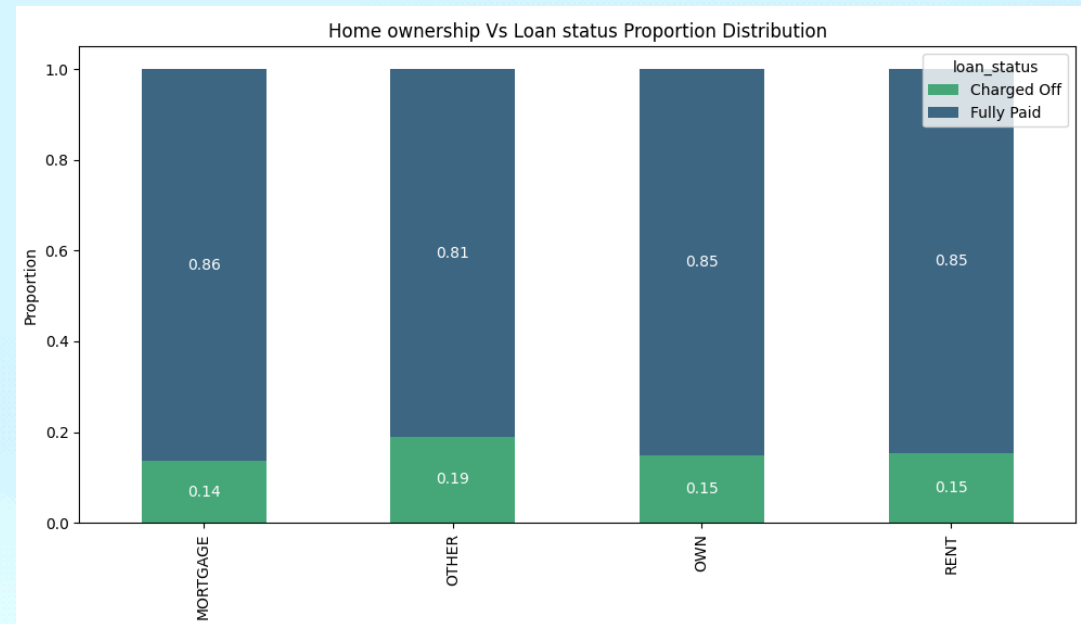


- The Charged off proportion increases as grade moves from A to G

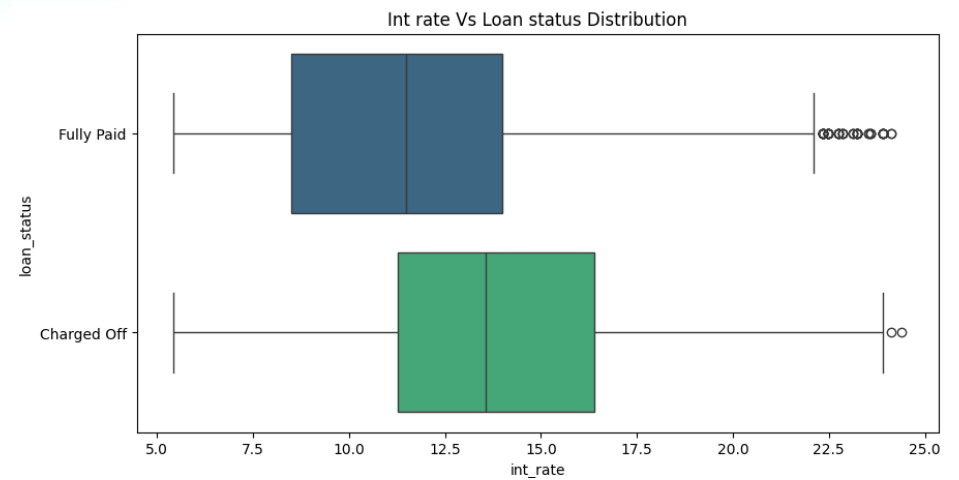
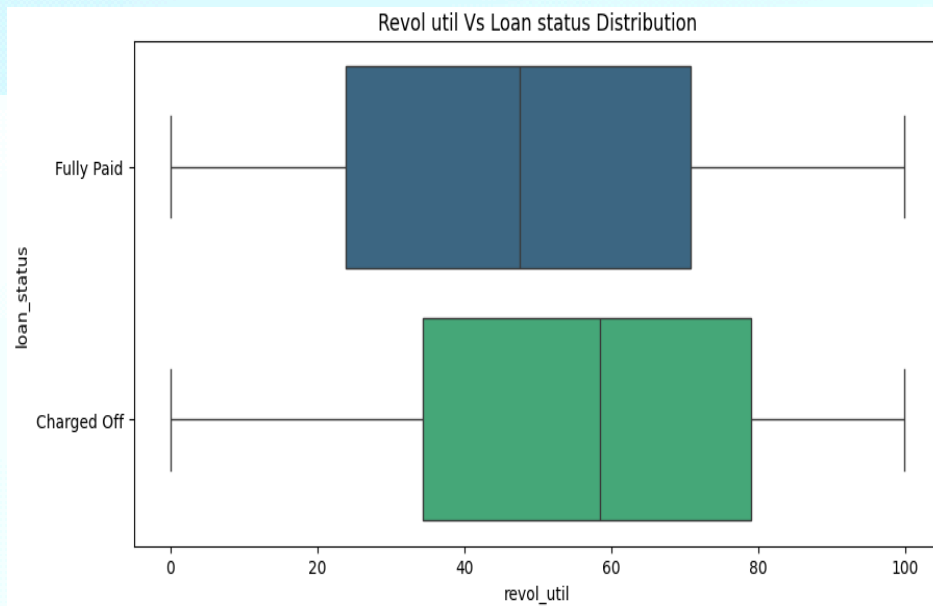
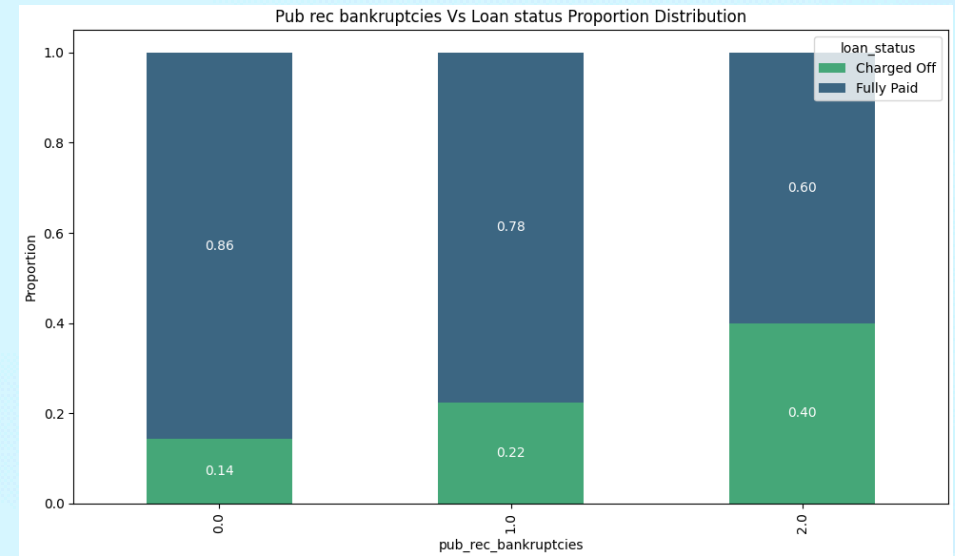
- For Purpose, small businesses tends to have more charged off rates

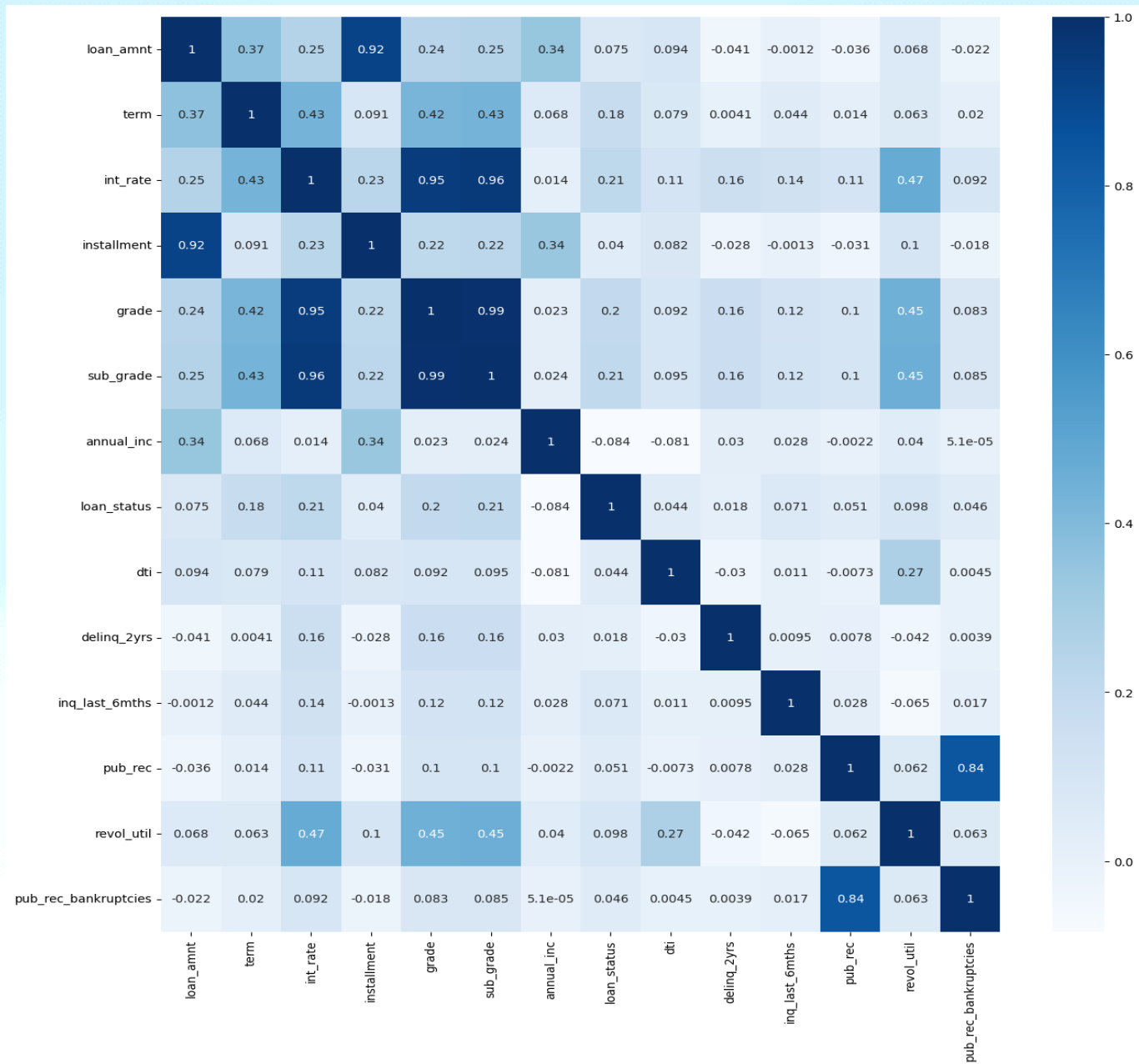


- In-case of home ownership the charged off rate is highest for category "OTHER", compared to the rest of the categories
- State Addr "NE", has the highest charged off rate



- For pub_rec_bankruptcies, charged of rate increase with count
- High revol_util rate have more chances of Charged Off
- High interest rate have higher chance of Charge Off





Conclusion

- Interest rate more than 13% have higher chance of Charge Off
- Those who are not ownig a home have high chances of Charged Off
- Loan for small business have high changes of Charged Off
- High revol_util rate have more chances of Charged Off
- Higher the number of bankrupties, higher the chance of Charged Off
- Loan Grades F and G have high chances of Charged Off
- Lower the Grade, higher the chance of Charged Off
- 5 year loans have more chances of Charged Off
- State NE have the high chance of Charged Off