

Lending Club Case Study

Submitted by:
Ahul Vidhyadaran
Abhishek Pathak

Contents

- Problem Statement
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Conclusion

Problem Statement

- Lending Club(LC) is the largest online loan marketplace, facilitating personal loans, business loans and financing of medical procedures. It's largest source of credit loss is lending loans to 'risky' applicants.
- Lending Club wants to identify the risky loan applicants, so that such loans can be reduced thereby cutting down the amount of credit loss
- Goal of this study is to find the driving factors behind loan getting default (Charged-Off)

Analysis Approach

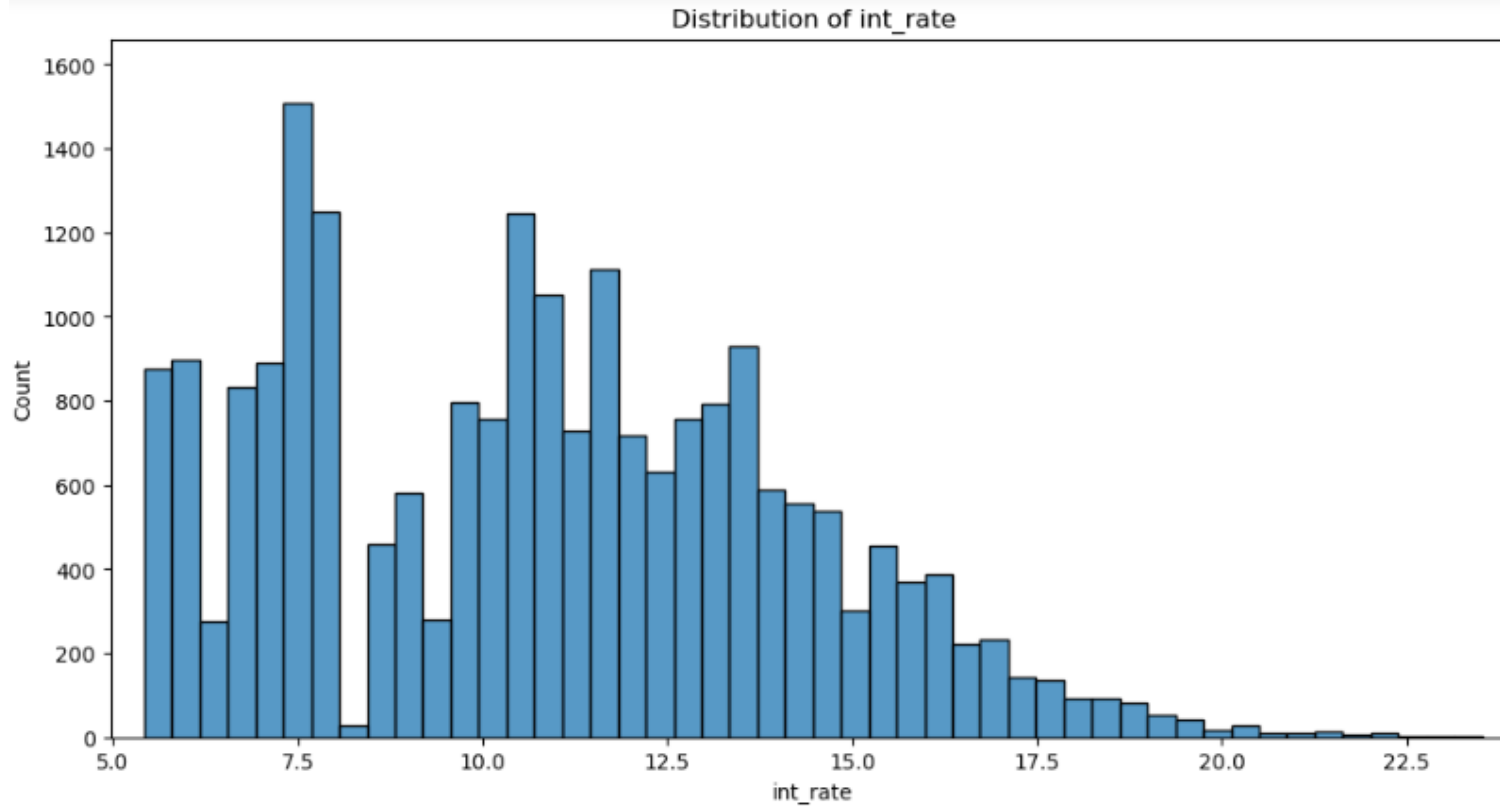
- The loan data set contain two main category of columns such as values pertaining to customer information and loan characteristics, other columns not relevant to the analysis will be excluded
- Loans with current status is not considering for analysis
- The data set will analysed in 4 different stages as follows to find out the key customer informations fields and loan characterstics fields.
 - Data Cleaning
 - Univariate analysis
 - Bivariate analysis
 - Multivariate analysis

Data Cleaning

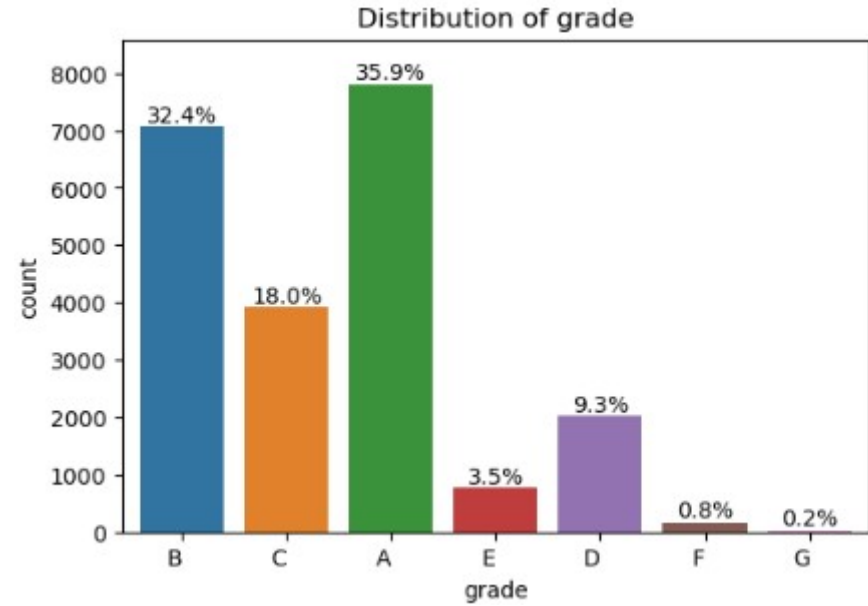
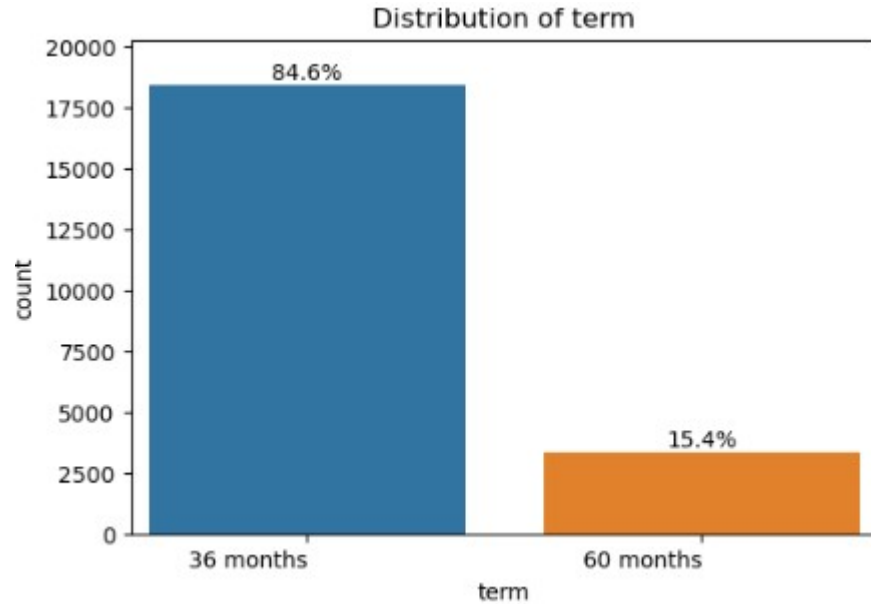
- Loan Data Set contains the complete loan data for all loans issued through the time period 2007 to 2011
- Many of the columns values are invalid or non-standardised format in the data set , in this phase of the analysis data will be cleaned/removed and standardised
- Steps to clean and standardise the values
 - Drop Columns with a High Percentage of Missing Values
 - Remove fields not relevant to the analysis such as unique ID fields
 - Standardising Values such as converting field types from string to float/int and converting and string type to datetime or better analysis

- Remove columns with only one unique value as one unique value column not impacting the analysis result
- Remove outlier data , outliers are calculated using IQR method
- Imputing data values with media/mode
 - Median is used for imputing values for numeric fields and Mode is used for categorical values
- Removed duplicate columns
- Analyse the categorical data columns and standardise value

Univariate Analysis

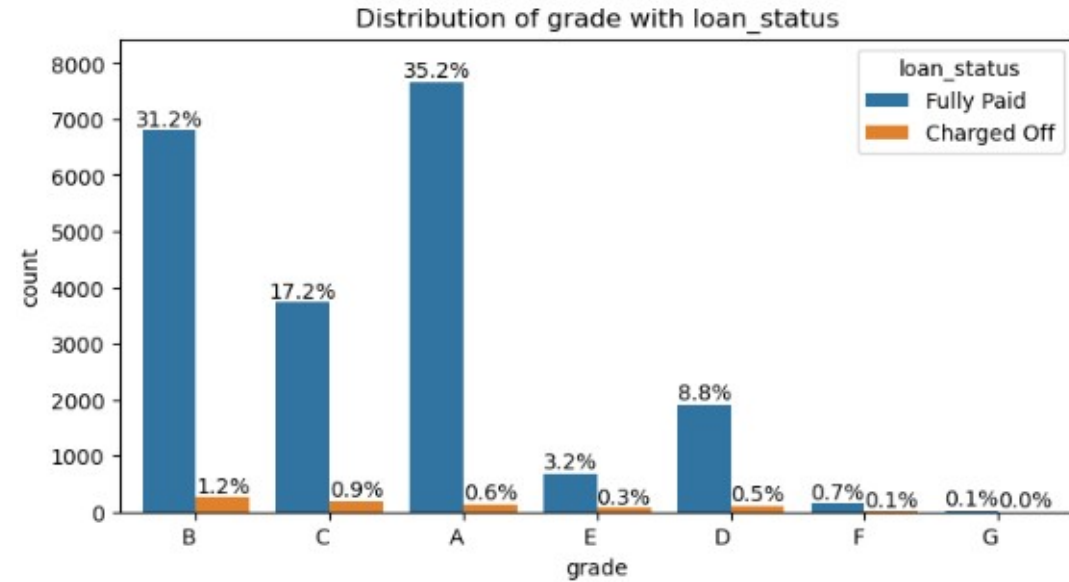
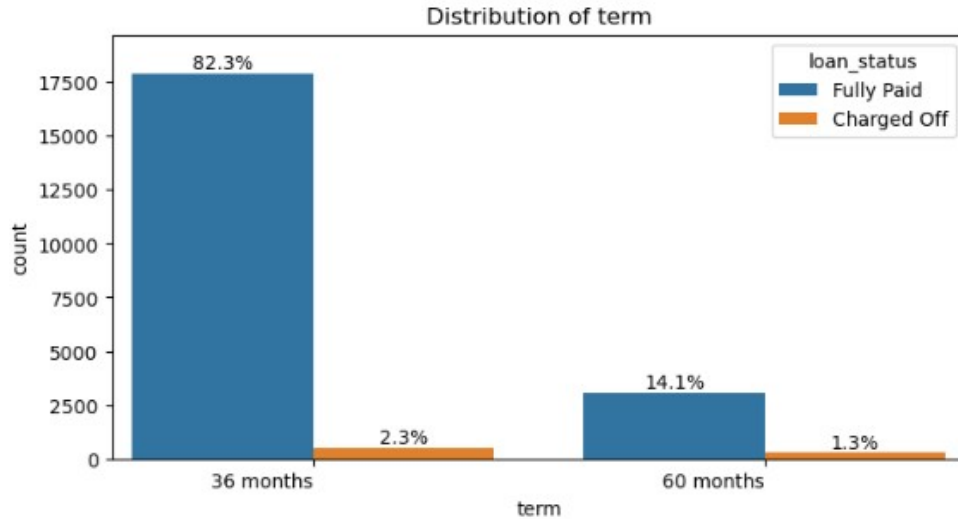


- More loans are taken with interest rate ranging from 6-17

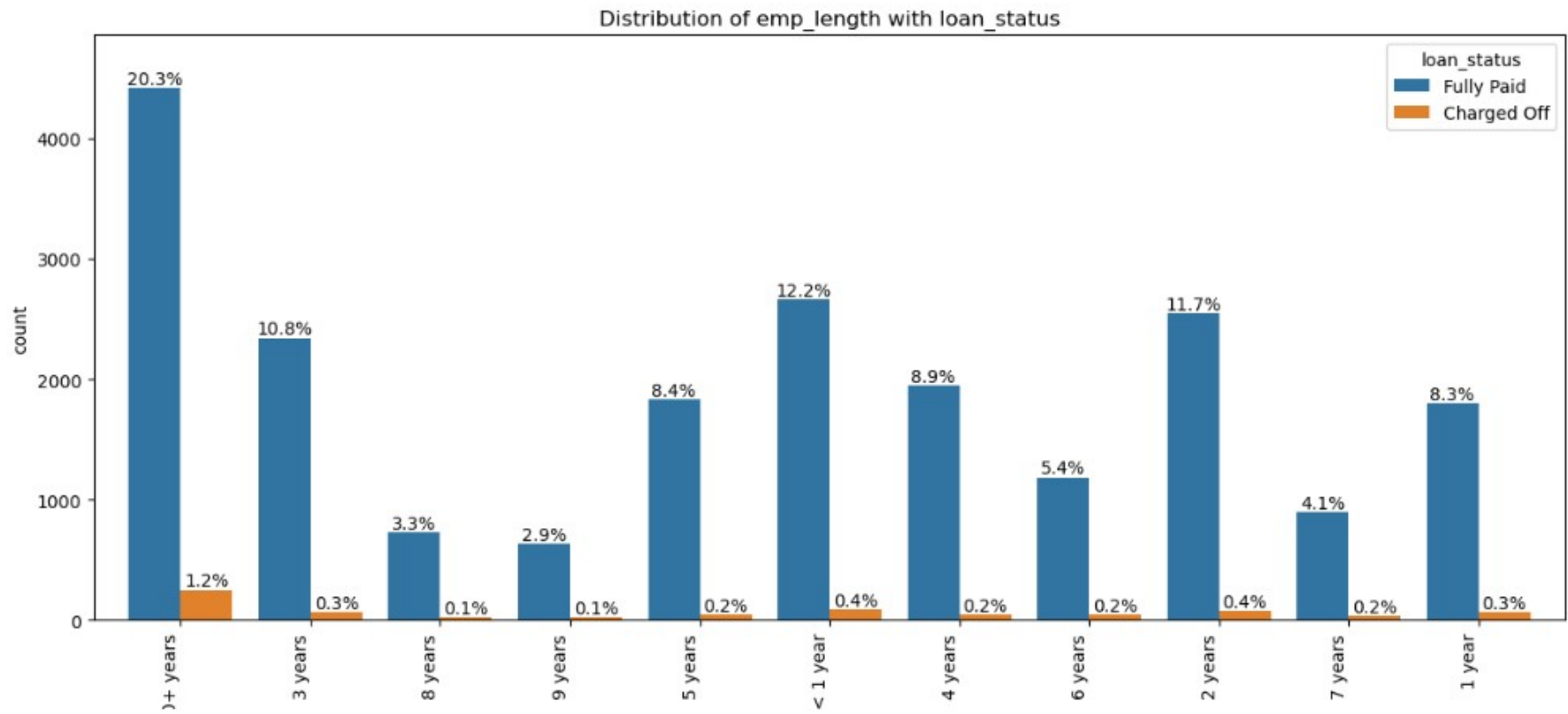


- 85% of the loan taken was 36months term and only 15% is 60 month
- More than 70% of the loans belongs to A,B,C grade, very few percentage belong to the lower grade loans

Segmented Univariate Analysis



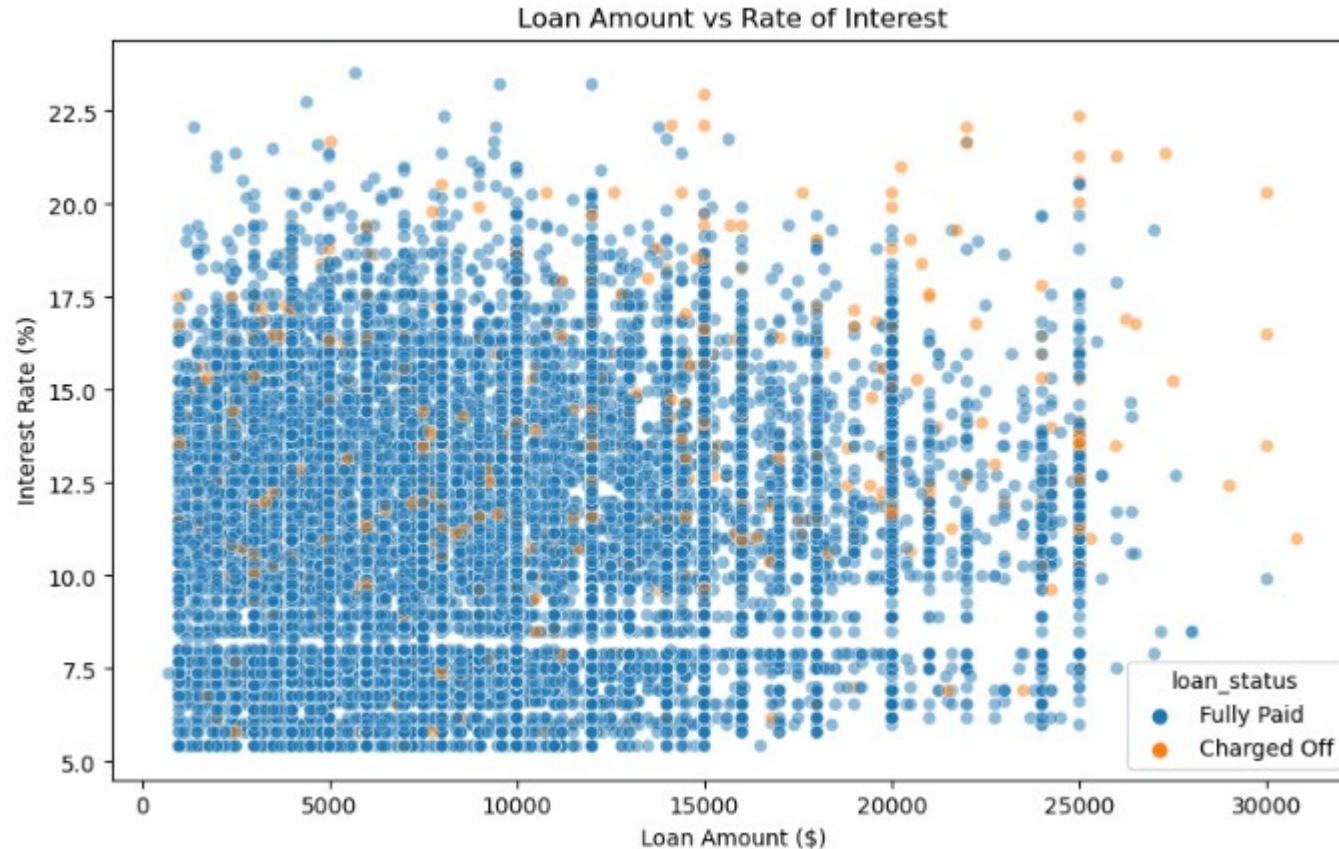
- Distribution of term with loan_status shows more loans are defaulted when compared with fully paid loan count for 60 months term
- Distribution of grade with loan_status shows more loans are defaulted for lower grades such as D, E, F when compared with no. of fully paid loans for that grade



- More loans are availed by category 10+ years and < 1 year but more Charged-off loans belongs to categories <1 Year, 1Year, 2Year, 3 Year ie junior employee segment

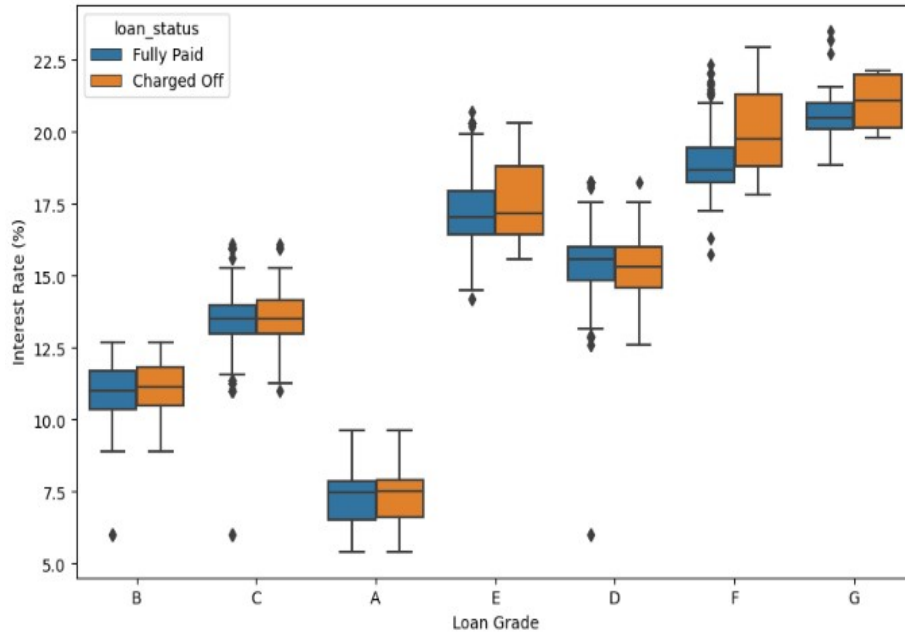
Bivariate Analysis

Analysis is done with combination of customer information data and loan characteristics data values



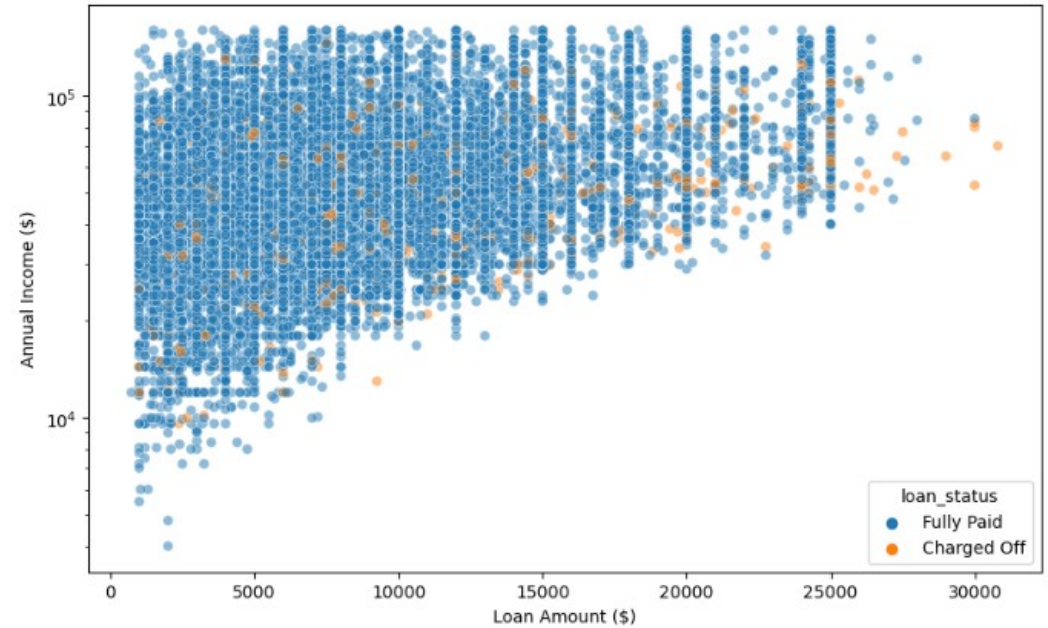
- When loan amount and interest rates are higher the chance of loan getting charged-off is also higher
- more loans are in charged-off state when interest rate is more than 17.5 %

Loan Grade vs Interest Rate



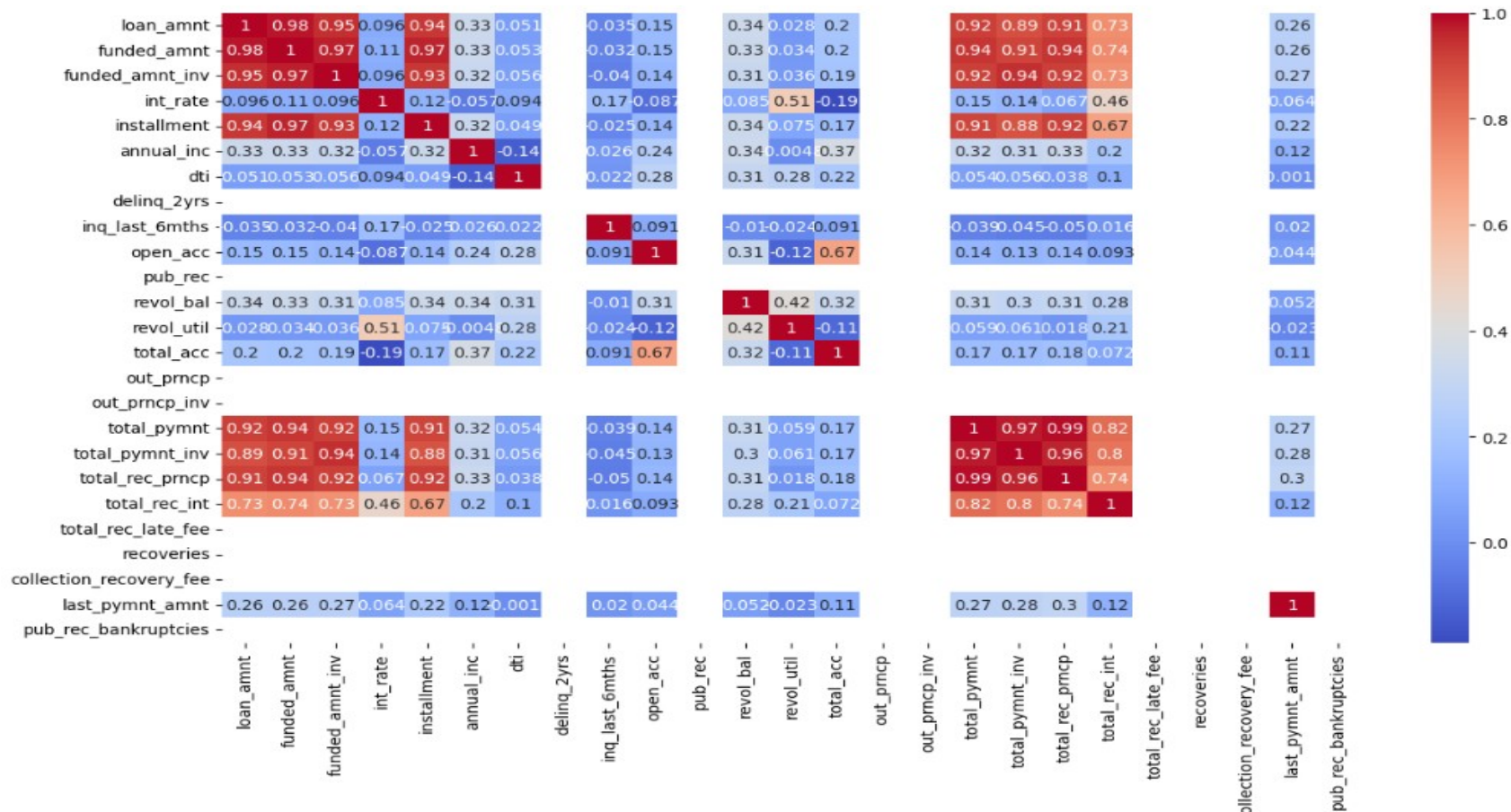
- Fig above shows loans with grade (A,B,C) comes under interest rate below 17.5 most of the cases and lower grade loans do have high interest rate and more change of default

Loan Amount vs Annual Income



- Loan Amount vs Annual Income showing inclination trend , such that when annual income increases(not all cases) the loan amount is also showing upward trend

Multivariate Analysis



- The previous plot shows the heat map of correlation between different numeric data type values
- In the plot annual_inc and dti is negatively correlated, there is strong correlation between loan_amnt, funded_amnt, and installment

Conclusion

Following factors should be considered when providing a loan to the customers

- Creditworthy Customers

- Loan amount in the range 5000 – 20000 USD
- Customer Work experience is more than 3 years
- Customers taking loan for 36 Month term
- Loan grade belongs to either A,B or C category
- Customer owning a home
- Customer dept to income ratio is below 20%

▫ Loan Risk factors

- Loan interest rate is more than 18%
- Applied loan amount more than 25000 USD
- Loan grade belongs to either D or above category
- Customer staying in rented home or mortgage home
- Customer dept to income ratio more than 25%
- Customer work experience is less than 3 years