LENDING CLUB CASE STUDY

EXPLORATORY DATA ANALYSIS

Team Members

Sebananda Ghosh
Bincy James

PROBLEM STATEMENT

Introduction:

- Develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding:

Lending club is a finance company specializing in lending various types of loans to urban customers

- 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

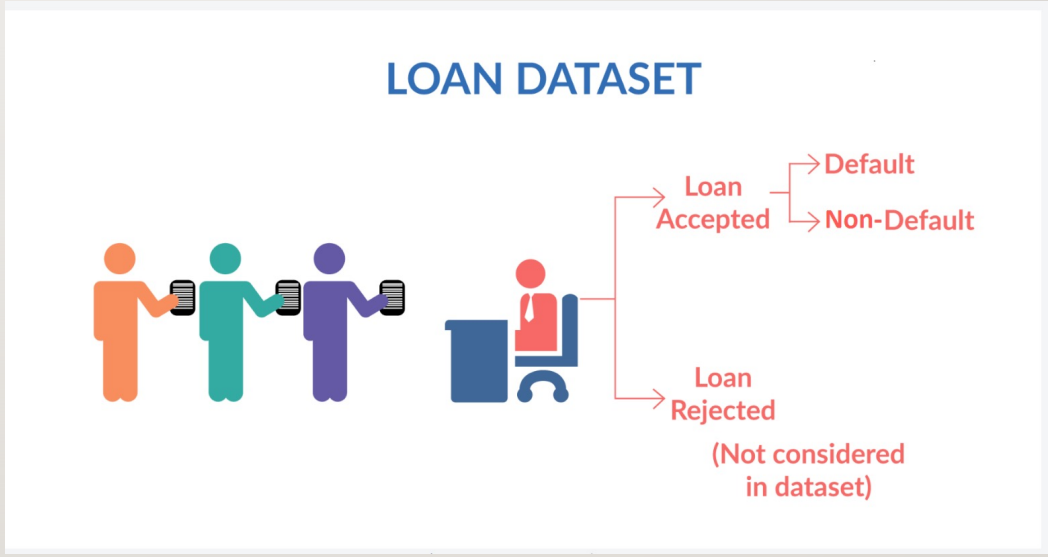
DATA SOURCE

1. loan.csv(loan data set)

2. Data_Dictionary.xlsx (explanation of variables in loan dataset)

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:
  - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
  - **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'.
  - **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)



ANALYSIS APPROACH

Data Sourcing

- Load the Loan Dataset
- Understand the data and variables

Data Cleaning

- Remove unnecessary columns based on missing value percentage threshold(50%)
- Remove unnecessary columns which have few unique values(unique/single value column drop)
- Convert loan status into numerical data i.e. Fully Paid = 0 and Charged Off = 1
- Handle the missing values properly
- Convert columns to proper datatype
- Standardize the data
- Outlier Treatment

Derived Metrics creation

- Create a derived metrics for loan issue year and month
- Create a column for defaulter type 1 yes and o no from analysing

Univariate Analysis ,Bivariate Analysis & Multivariate Analysis

- Interest rate vs loan status
- Loan Amount vs loan status
- Loan Term vs loan status
- Loan Grade vs loan status
- Home ownership Vs loan status
- Home ownership Vs loan Amount
- Term Vs loan Amount
- Heat map to understand the relation with each variable on other

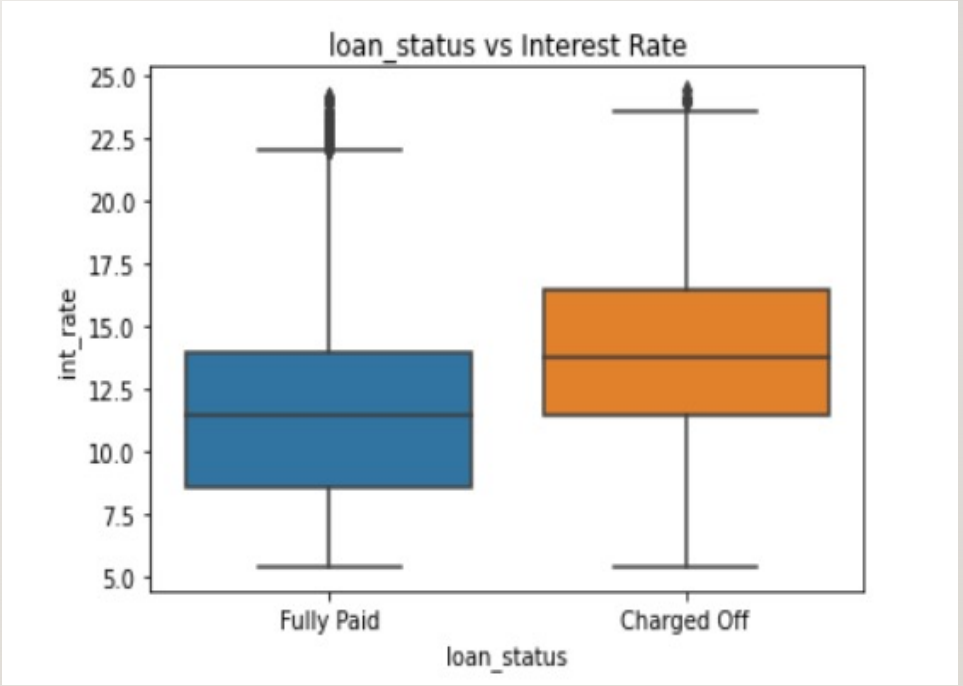
Conclusion

- Concluded with an analysis recommendation to Lending club

ANALYSIS

- INTEREST RATE VS LOAN STATUS

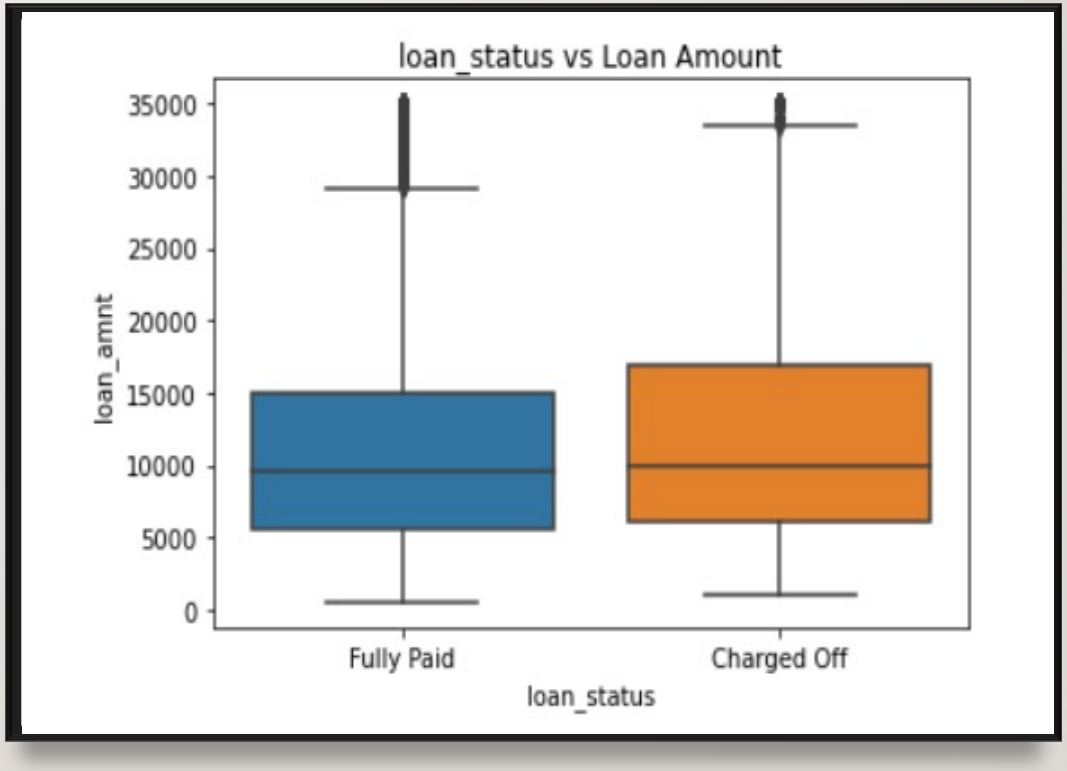
  This clearly tells us, greater the interest rate more the chance of defaulting the loan.



ANALYSIS

- LOAN AMOUNT VS LOAN STATUS

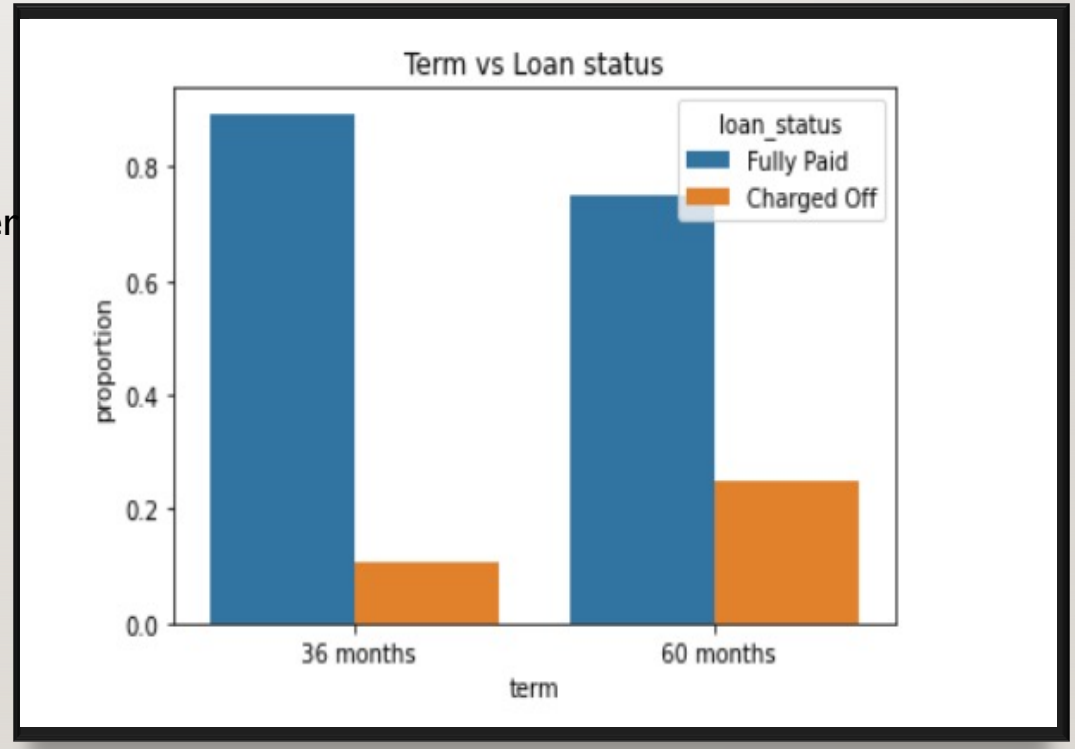
  Observations: Charged Off loans have higher amounts than Fully Paid ones.



ANALYSIS

- LOAN TERM VS LOAN STATUS

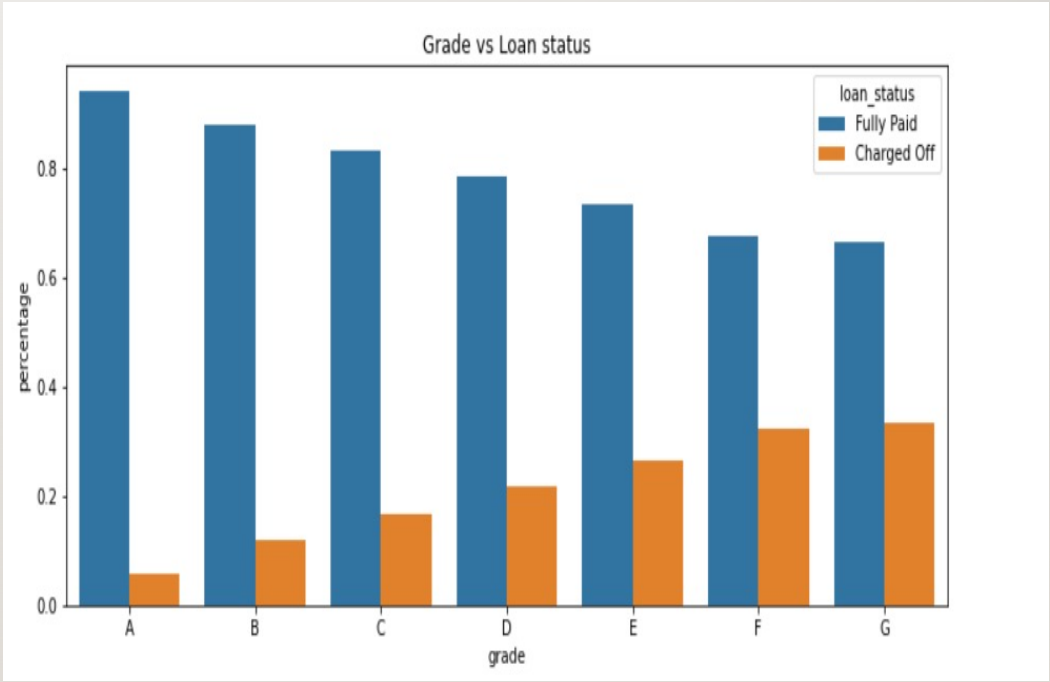
  There are more proportion of borrowers defaulted loan in 60 months term, then 36 months. Also, the Fully Paid rate is higher in 36 months tenure.



ANALYSIS

- LOAN GRADE VS LOAN STATUS

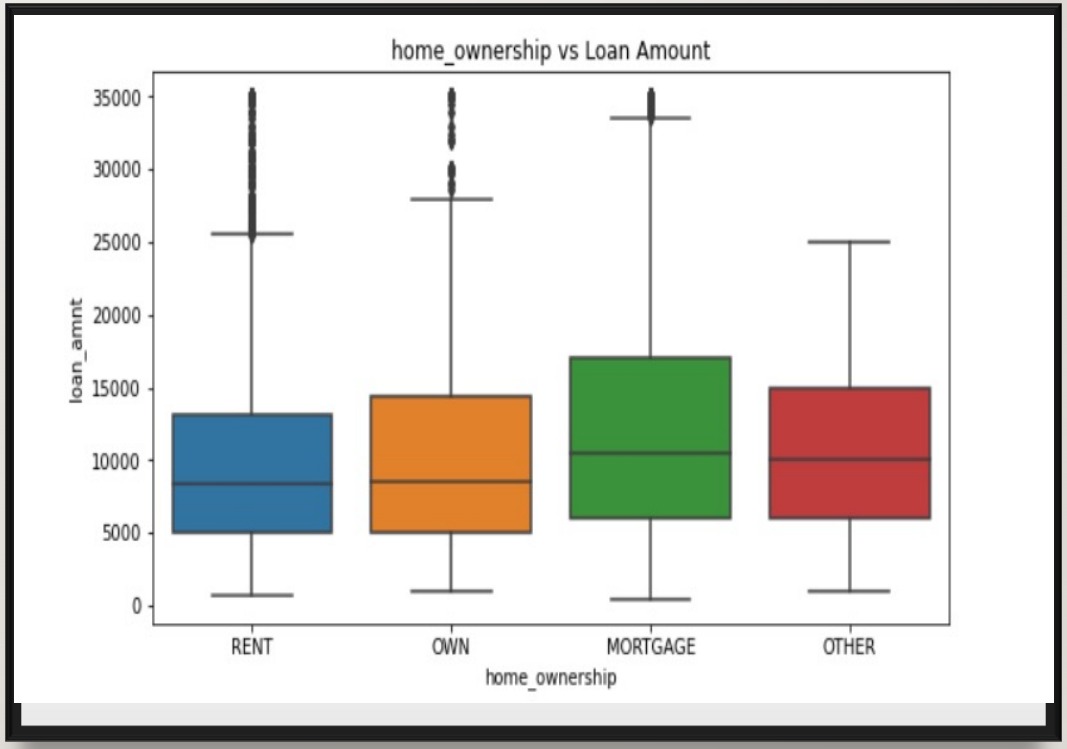
The graph says that the Charged off increases as grades decreases.



ANALYSIS

- HOME OWNERSHIP VS LOAN AMOUNT

There is bit high percentage of defaults are recorded in other home ownership category. Also People with Mortgage tent to take more loan Amount than others.



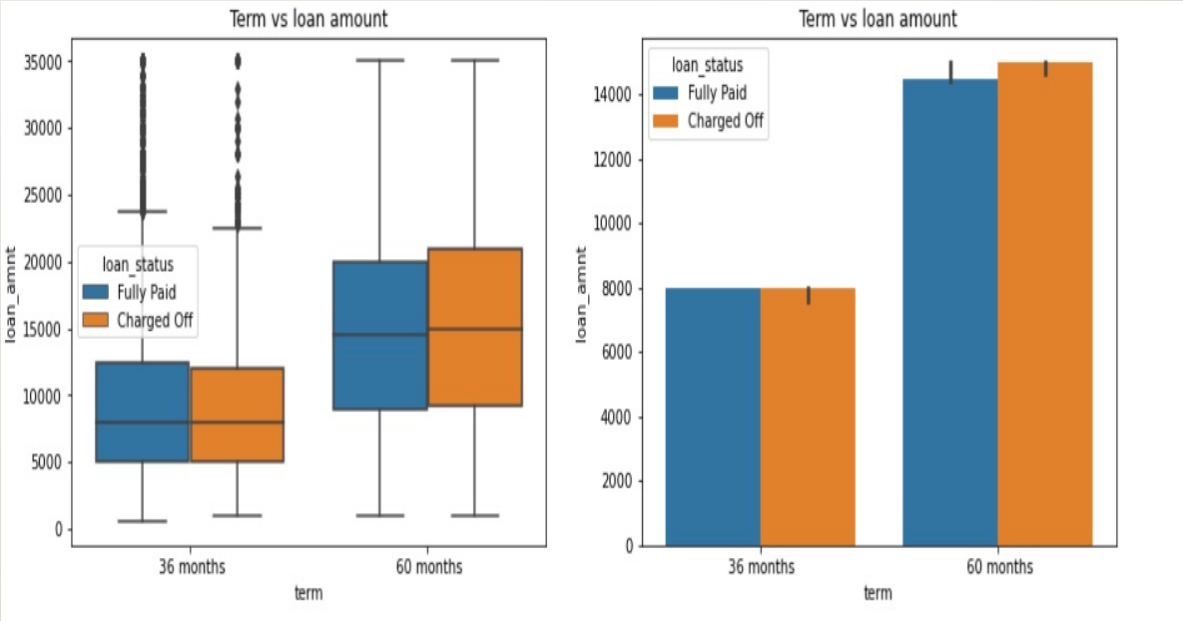
ANALYSIS

- TERM VS LOAN AMOUNT

People are taking more loan Amount for 60 months than 36 months.

There are more proportion of borrowers defaulted loan in 60 months term, than 36 months.

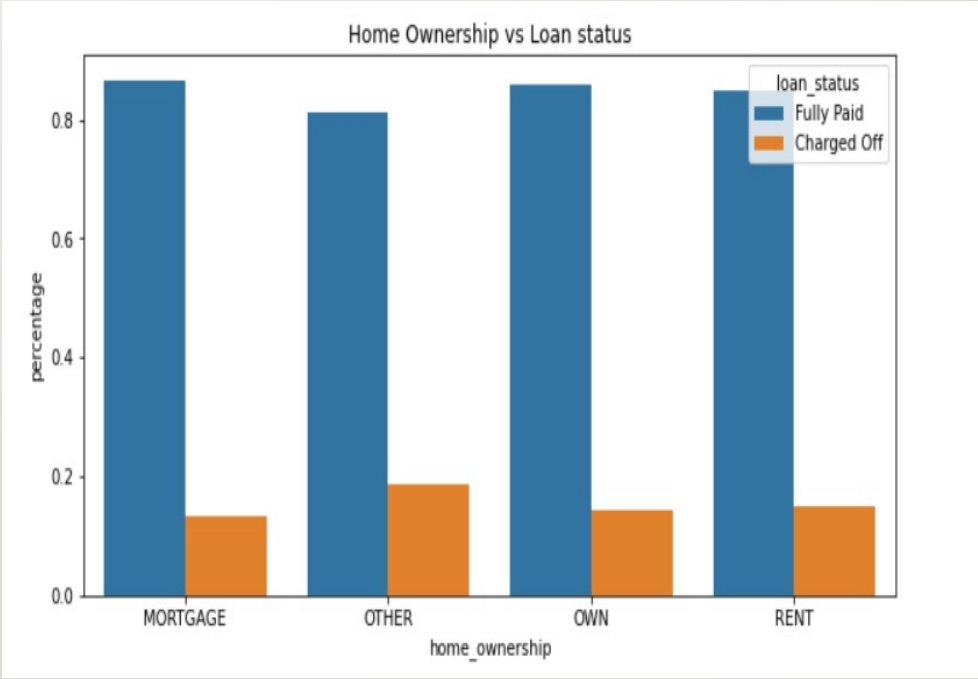
Also, the Fully Paid rate is higher in 36 months tenure.



ANALYSIS

- HOME OWNERSHIP VS LOAN STATUS

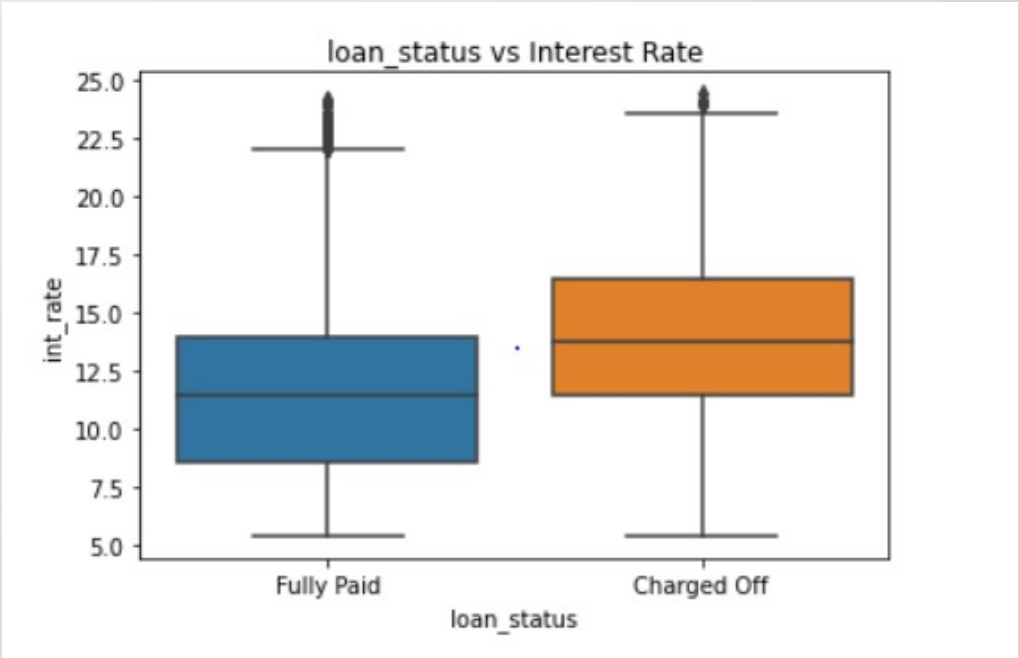
There is bit high percentage of defaults are recorded in other home ownership category. Also People with Mortgage tent to take more loan Amount than others and high default followed by Rent and other.



ANALYSIS

- LOAN STATUS VS INTEREST RATE

  This clearly tells us, greater the interest rate more the chance of Defaulting the loan.

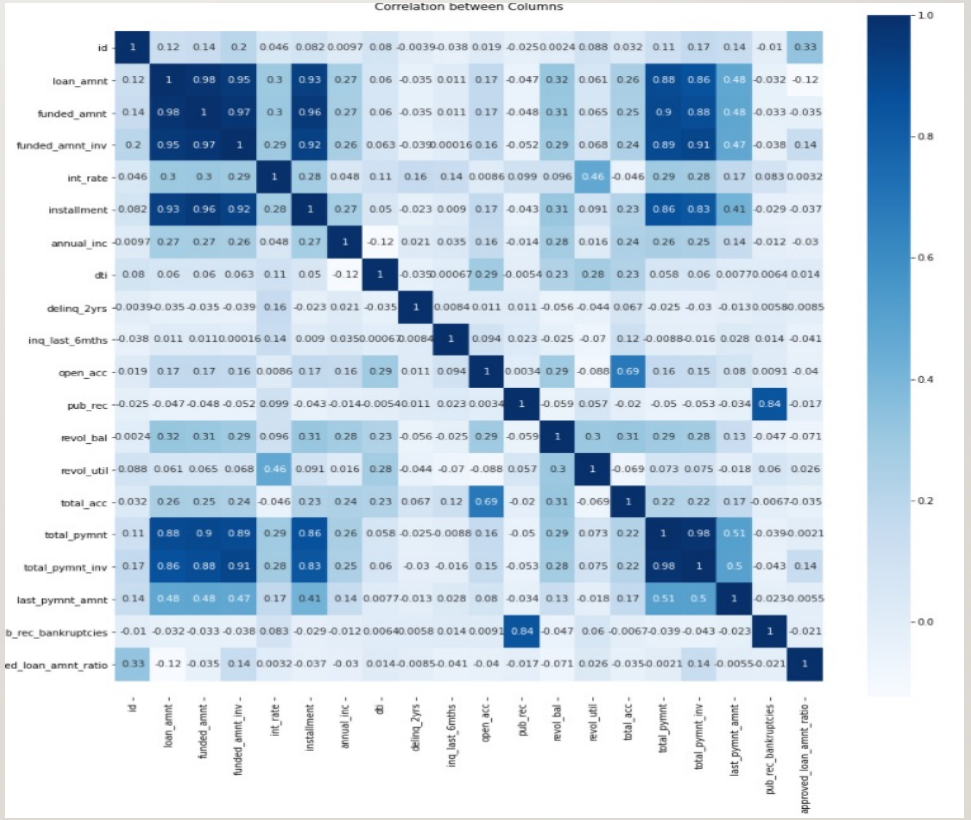


ANALYSIS

- NUMERICAL VALUE ANALYSIS

  The public derogatory records columnn is highly (+ve) correlated with public bankruptcies records.

  Interest rates are high for people with high revol utilisation.



RECOMMENDATIONS BASED ON ANALYSIS

- Greater the interest rate more the chance of Defaulting the loan
- People are taking more loan Amount for 60 months than 36 months.
- For loan amount of 60 months the risk is more as there are more defaulters
- When loan amount increases more chance of defaulting the loan
- E,F and G Grade loans are at risk as they have higher loan default
- Applicant who took loan for debt_consolidation has high risk.

CONCLUSION

- Completed the case study and lending club can use this to analyze. and mitigate the risks and there by cutting down the overall credit loss.