Lending Club Case Study

Submitted by

Hannah Abraham

Hamsaveni Ramaiah

Contents

Business Statement

Introduction: Lending loans is a risky business as it involves a great risk of ‘Defaulters”, due to increase in the defaulters the organizations lending money face huge financial loss also called “credit loss”.

Problem statement: The main objective of this case study is to understand factors that are strong indicators of defaulters.

By using EDA (Exploratory Data Analysis) concepts we can understand how each of these variables singly or jointly affect the loan repayment. This Analysis will also help us understand the co-relation between multiple factors.

The Company can utilize this knowledge for its portfolio and risk assessment, they should be able identify “risky applicants” (who are the largest source of financial loss) and decision can be made on the approval of the loan.

Note: The loans in the “Current” state, is not considered for analysis.

Data Cleaning

Before commencing the analysis the below listed data cleaning activities has been performed to ensure the data used for error due to incorrect / incomplete data does not impact the analysis

Missing values : Columns with > 50% of values missing are removed

Irrelevant Variables: variables that do not contribute to the analysis ex: URL are removed

Imputation of Missing values: applied imputation for Employee Length and Public records of Bankruptcy.

Incorrect format/ multiple format usage ex: date : all the values in the columns are aligned to single format

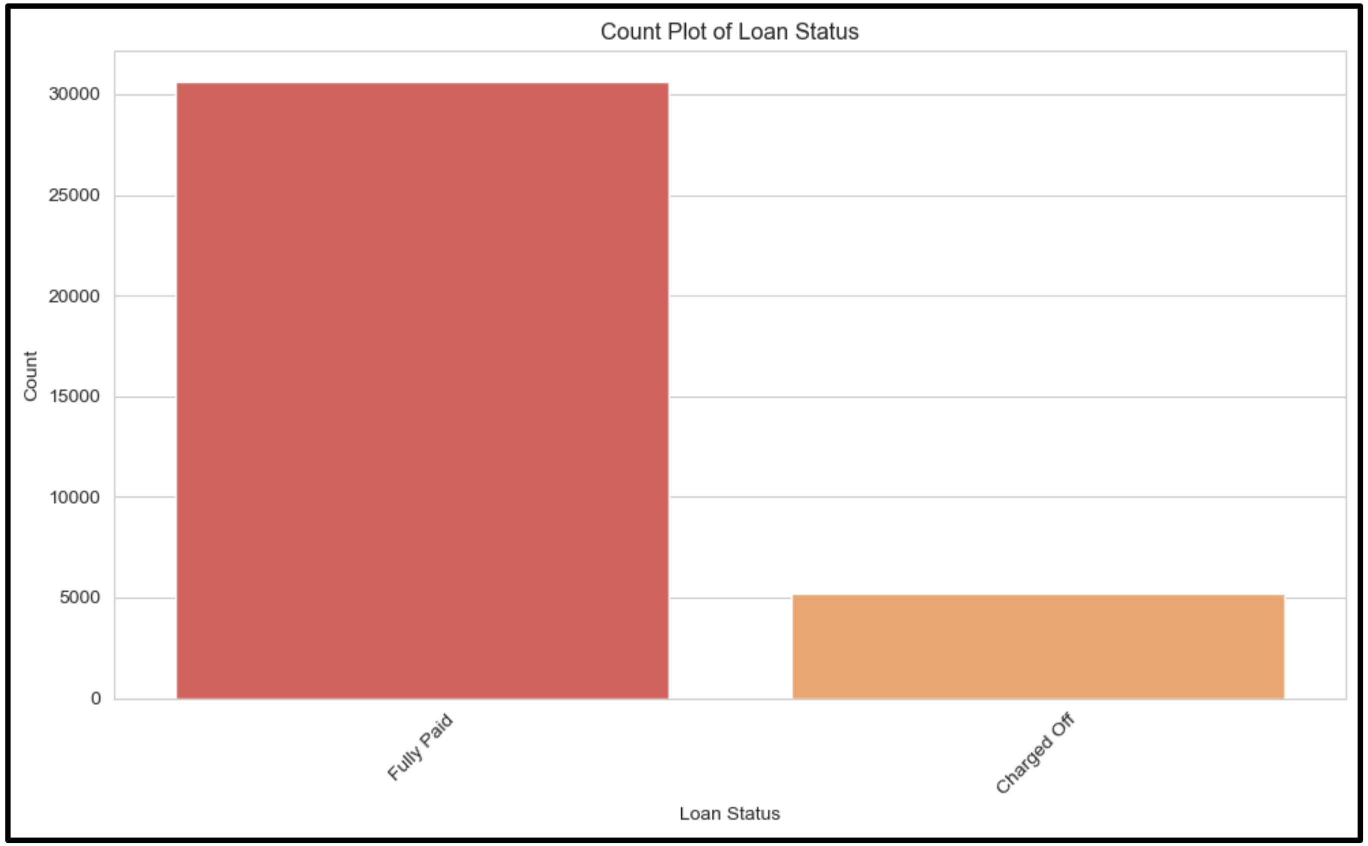
Derived Metrics: Loan issue Month and Loan issue year is derived from 'Loan Issue Date'

Outliers: outlier values are removed to avoid any type of bias during analysis

Duplicate Rows: Duplicate rows are removed

Univariate Analysis

Good Loan V/s Bad loan



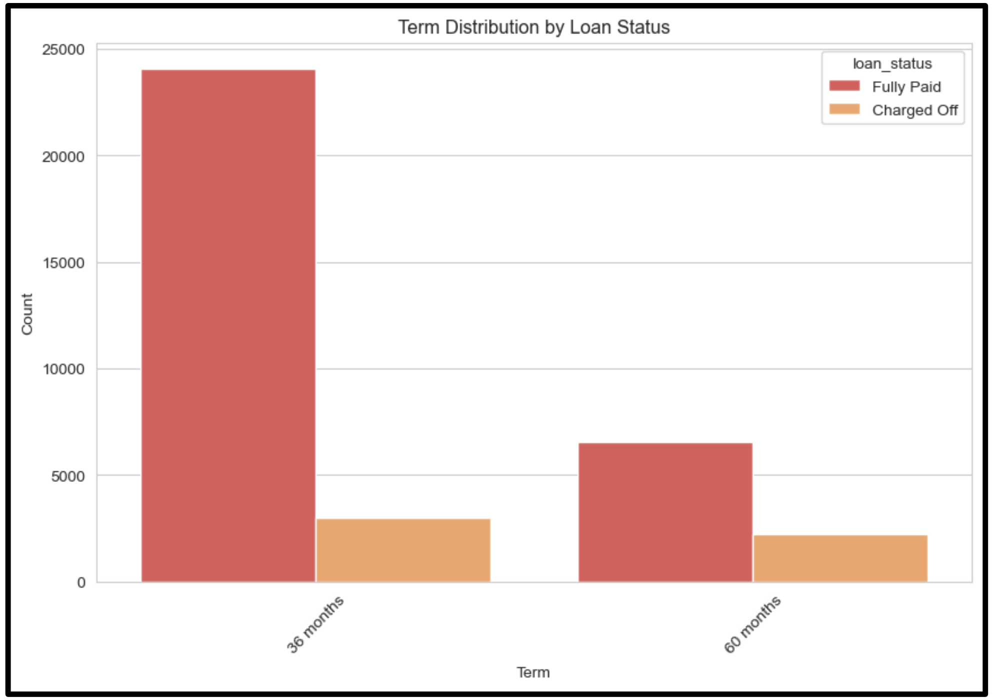
The graph depicts the count of loans that was “Fully Paid” V/s “Charged off”

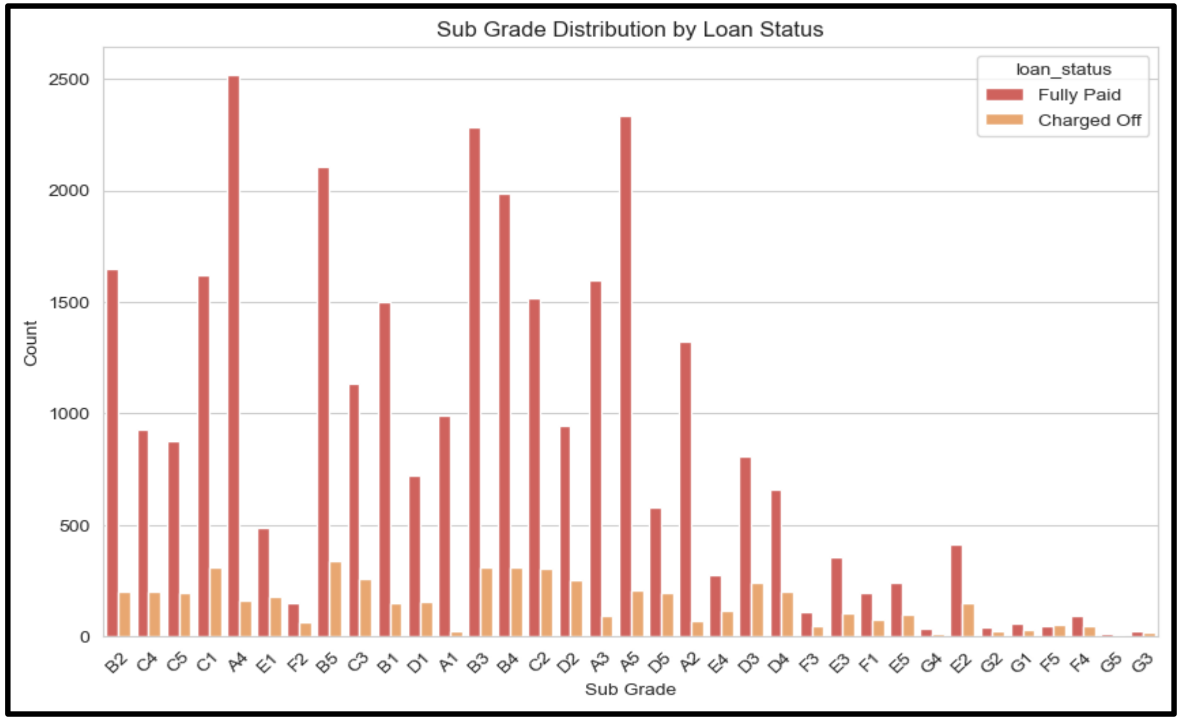
~75% of the loan was fully paid

Inferencing that ~ 10% (considering the ‘current loan’) of the loan was charged off. This will amount to huge volume of loan amount, requiring thorough analysis of the variables that can help identifying the key factors that pose the “Risk” of “Credit losses”

Univariate Analysis (Segmented)

Impact of “Term of loan” & “Sub-Grade”





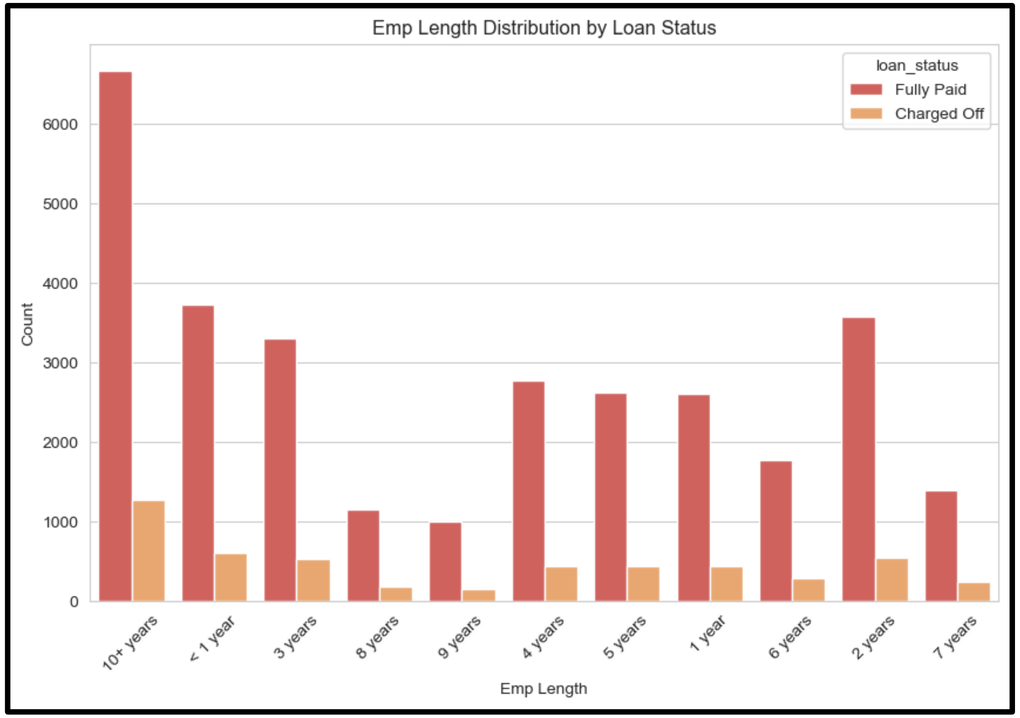
Inference:-

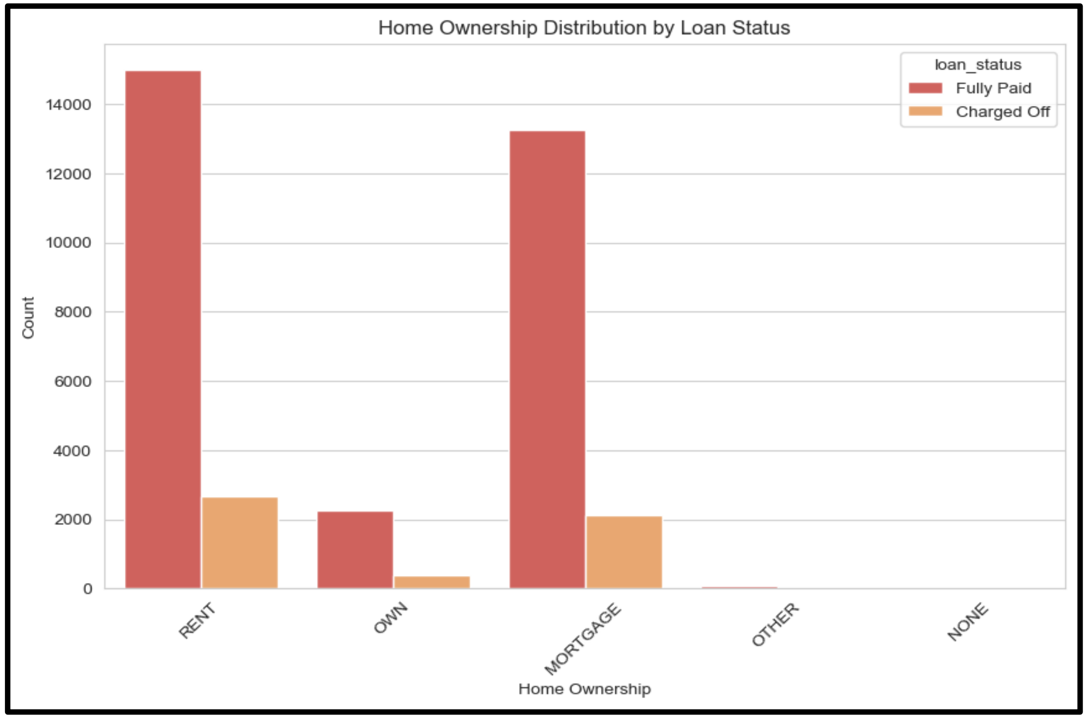
Term: The percentage of defaulter is relatively higher when the repayment term is 60 months

Sub Grade : The charged off loans significantly increase with subgrades of D, E, F , G and loans availed by applicants with A, B, C grades are better with Paying off the loans. The grades of the applicant can be used as an important factor to determine their repayment capabilities.

Univariate Analysis (Segmented)

Impact of “Length of Employment” & “Home Ownership”



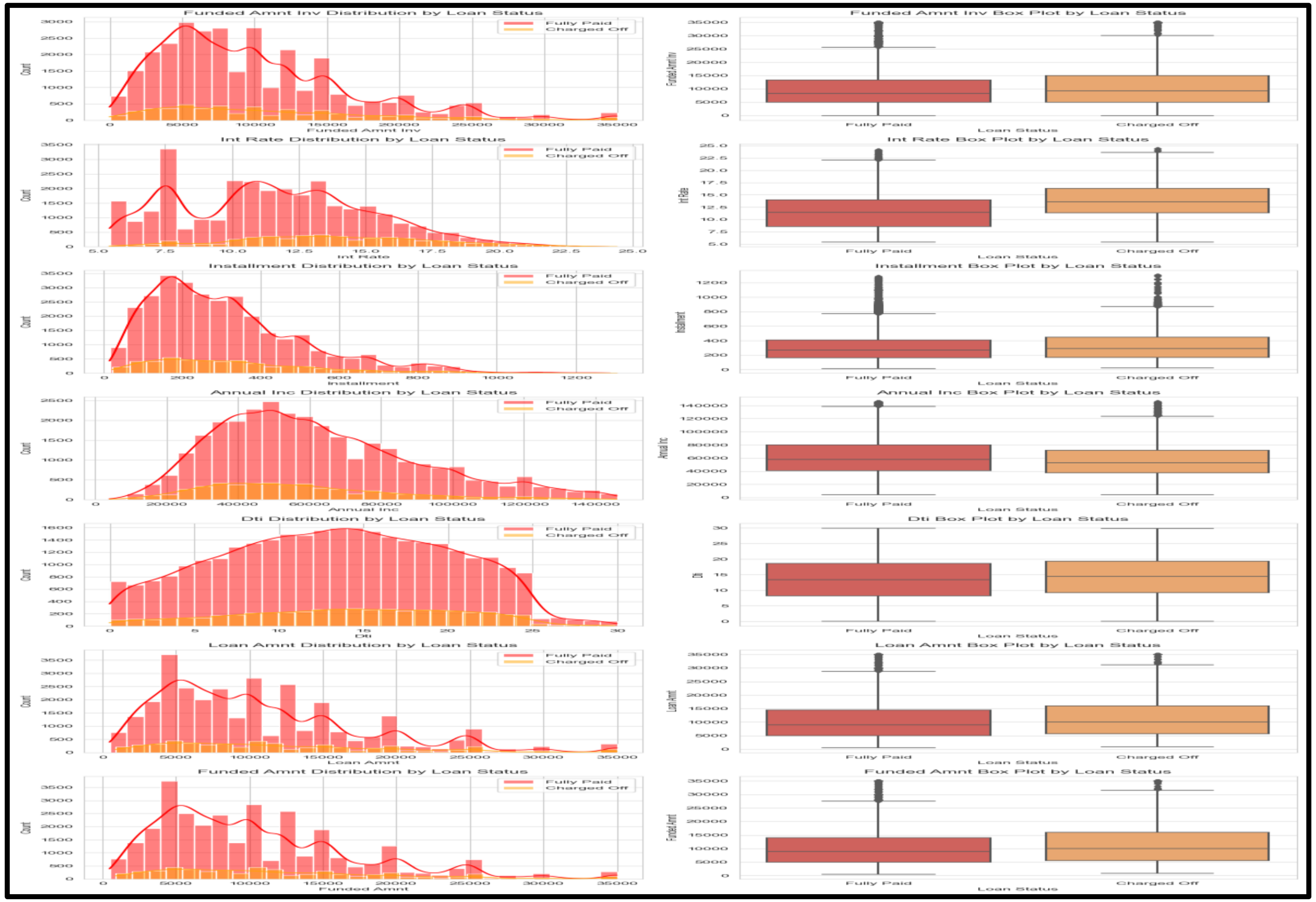


Inference:

Length of Employment: Majority of Borrowers have 10+ years of work experience, it is also observed that applicants having longer years of experience are awarded bigger loan amounts and the percentage of defaulting also reduces.

Home Ownership: The majority of borrowers have rented or mortgaged houses

Segmented Univariate Analysis (Numerical)



The graphs shows the impact of Numerical factors on the "loan status"

Funded Amount By Investor: Higher the funded amount by investor higher are the chances to default.

Interest Rate: Higher the interest rate higher are the chances to default.

Installment Higher the interest rate higher are the chances to default.

Annual Income: Lesser the annual income higher are the chances to default.

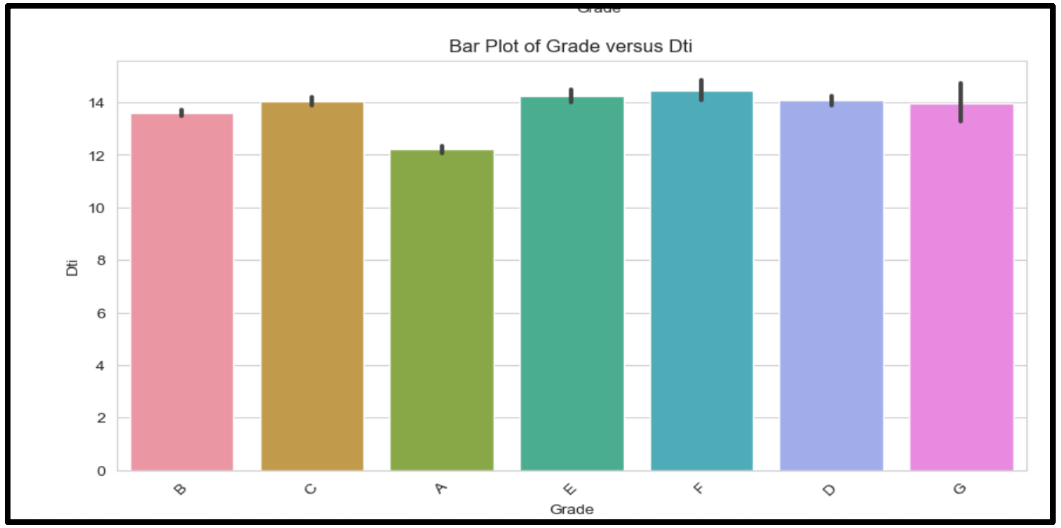
Debt-to-Income Ratio: Higher the DTI rates higher are the chances to default.

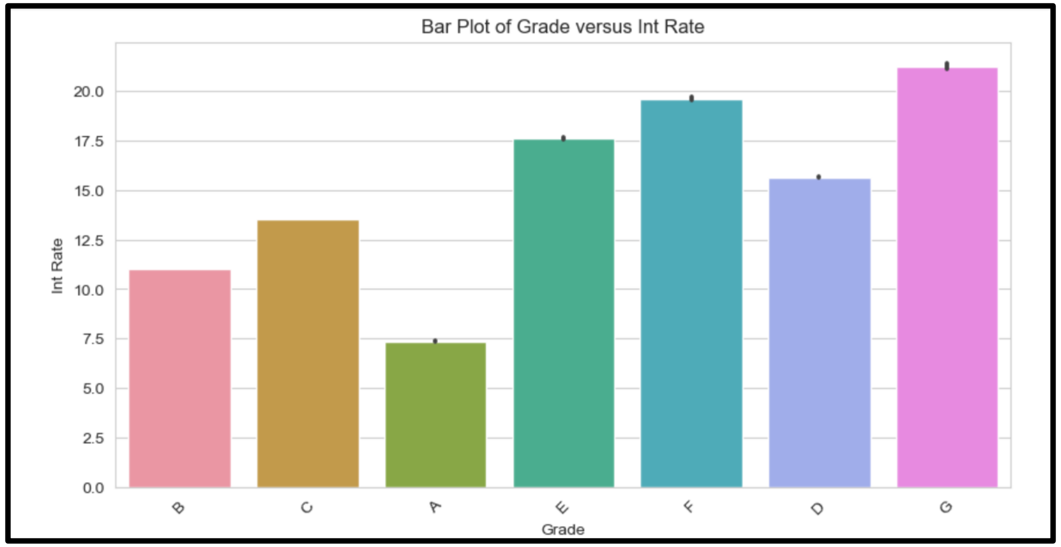
Funded Amount: Higher the funded amount higher are the chances to default.

Loan Amount: Higher the loan amount by investor higher are the chances to default.

Bivariate Analysis (Numerical V/s Categorical)

Grade V/s Int Rate
Grade V/s DTI





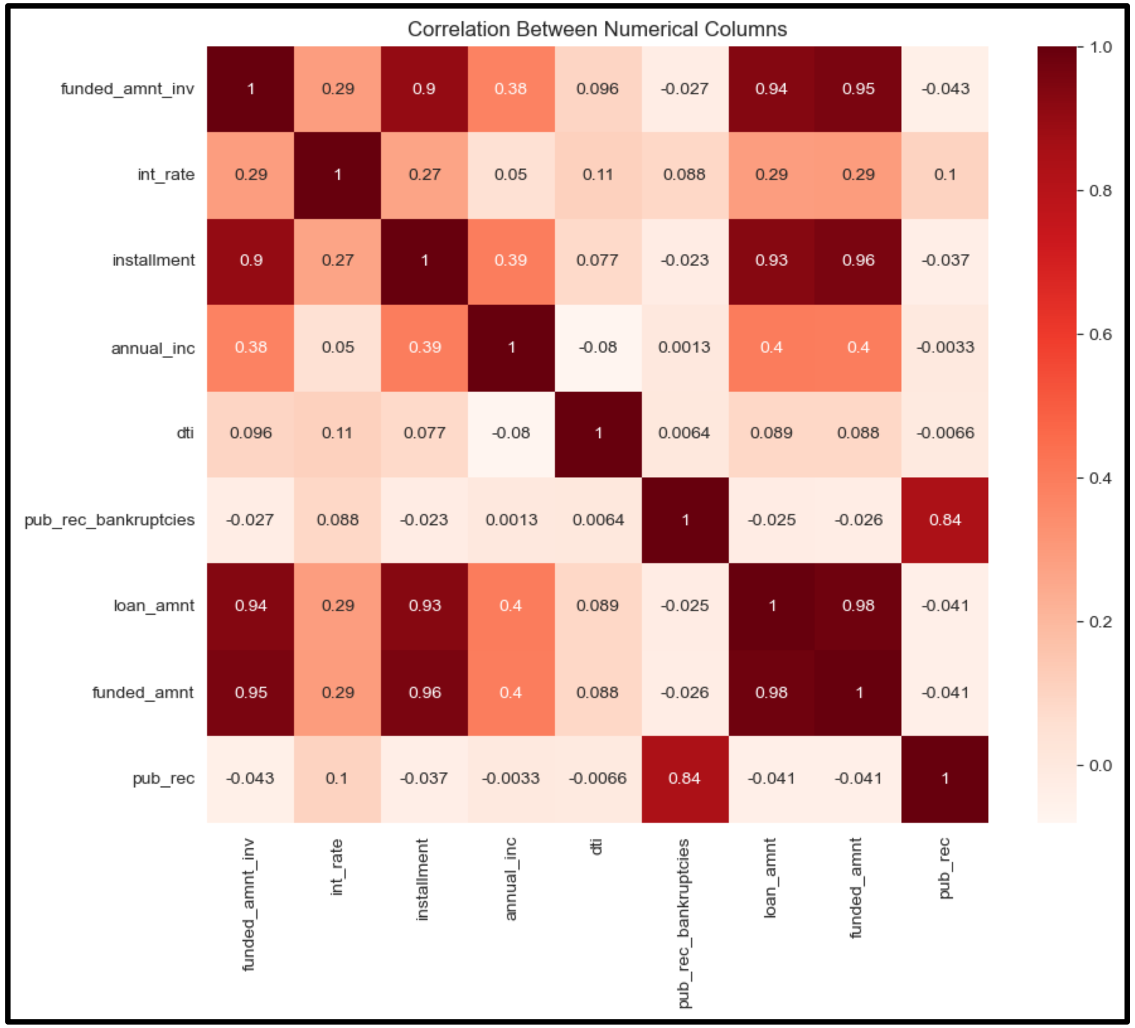
Inference:

Grade V/s DTI : Grade A has the least DTI, which depicts that the applicants with Grade A , have lesser debt and a higher income.

Which implies that they will get better interest rates as depicted in graph 2.

Grade V/s Interest Rate: As the DTI percentage increases inferencing that applicants Debt increases with same level of income (or probably slightly higher income) the grades are moved to E/F/G,   this is a measure to show their repayment capability hence applicants with Grade E/ F/ G have higher interest rates

Multivariate Analysis



Highly Correlated Pairs:

- loan_amnt and funded_amnt have a very high correlation of 0.98.
- funded_amnt and installment also have a high correlation of 0.96.

Moderately Correlated Pairs:

- loan_amnt and annual_inc have a correlation of 0.4.
- funded_amnt and annual_inc have a correlation of 0.4.

Low Correlations:

- DTI with other columns.
- pub_rec with other columns.

Negative Correlations:

- funded_amnt_inv and pub_rec_bankruptcies
- installment and pub_rec_bankruptcies

Summary

Interest Rates & Loan Duration

- Average interest rates have declined since 2012 this could have led to increase in the volume of loans.
- Loans that have a high interest rate(above 13.23%) are more likely to become a bad loan .
- Loans that have a longer maturity date (60 months) are more likely to be a bad loan.
- Interest Rates: In all reasons for application except (medical, small business and credit card), the low income category has a higher interest rate. Something that could possibly explain this is the amount of capital that is needed from other income categories that might explain why the low income categories interest rate for these purposes are lower.

Income Group

- Borrowers that made part of the high income category took higher loan amounts than people from low and medium income categories.
- Loans that were borrowed by the Low income category had a slightly higher chance of becoming a bad loan. (First row to the right of the subplots)
- Borrowers with High and Medium annual incomes had a longer employment length than people with lower incomes.(Second row to the left of the subplots)
- CA, NY, FL regions seems to be the most attractive in term of funding loans to borrowers.
- Borrowers with a lower income had on average higher interest rates while people with a higher annual income had lower interest rates on their loans.

Summary … contd

Purpose of Loan

- Mortgage was the variable from the home ownership column that used the highest amount borrowed within loans that were considered to be bad.
- There is a slight increase on people who have mortgages that are applying for a loan.
- Bad Loans Count: People that apply for educational and small business purposed tend to have a higher risk of being a bad loan. (% wise)
- Most frequent Purpose: The reason that clients applied the most for a loan was to consolidate debt.
- Less frequent purpose: Clients applied less for educational purposes for all three income categories.
- Bad/Good Ratio: Except for educational purposes (we see a spike in high income this is due to the reasons that only two loans were issued and one was a bad loan which caused this ratio to spike to 50%.), but we can see that in all other purposed the bad good ratio is lower the higher your income category.

Key Driving Factors

Minor Impact

- ✓ Higher loan amount (above 16K)
- ✓ Higher instalment amount (above 327)
- ✓ Lower annual income (below 37K)
- ✓ Higher debt to income ratio (above 15%)
- ✓ Applicant's address state (NV, SD, AK, FL, etc.)

Heavy impact

- ✓ Higher interest rate (above 13%)
- ✓ Repayment term (5 years)
- ✓ Loan purpose (small business, renewable energy, educational)
- ✓ Derogatory public records (1 or 2)
- ✓ Public bankruptcy records (1 or 2)

Combined impact

- ✓ High loan amount & interest rate for lower income group
- ✓ High instalment and longer repayment term
- ✓ Home ownership (other) and loan purpose (car, moving or small business)
- ✓ Residential state and loan purpose
- ✓ Income group and loan purpose

Business Recommendations

Risk Assessment Guidelines

- Higher interest rate , typically greater than 13% has greater risk of defaulting
- When the repayment term gets higher (around 60 months) rate of defaulting increases
- Loans availed for small business, renewable energy, educational, have a tendency to default
- Recommendation to consider applicants with public bankruptcy records (1 or 2) to be of high risk
- Higher loan amount (above 16K) have chances of defaulting.
- Applicants with Lower annual income (below 37K), can be considered as “Risky”
- Higher debt to income ratio (above 15%) pose risk of defaulting.
- Applicant residing in states like NV, SD, AK, FL, etc. can pose risk of being defaulters.

Risk Mitigation Strategy

When the Lending club receives loan application with any of the above factors, they must exercise extra caution to ensure the loan is not defaulted. Some of the recommendations

- ✓ Perform additional checks of the applicant
- ✓ Suggest joint application with lower risk applicant
- ✓ Reduce the loan amount
- ✓ Reduce the term
- ✓ Decline the loan

Thank you