

Lending Club Case Study

By

Manivannan K

Vinu Priya D


Background – Lending Club Case Study



Background

- The background of the project involves peer-to-peer lending, where borrowers can obtain loans and investors can fund those loans.

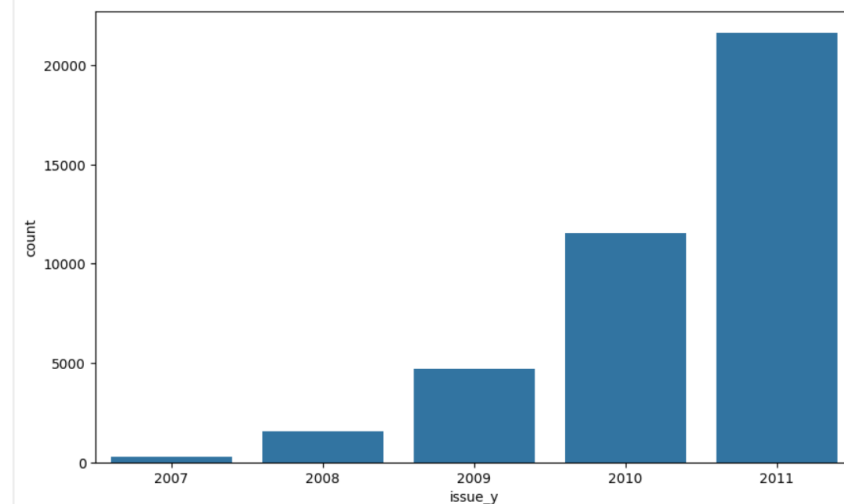
Business Objective

- The business problem being addressed is identifying factors that contribute to loan defaults and assessing the risk associated with different types of loans.
 - The dataset used includes information on loan amounts, interest rates, borrower details, loan status, etc.
- 

Univariate Analysis

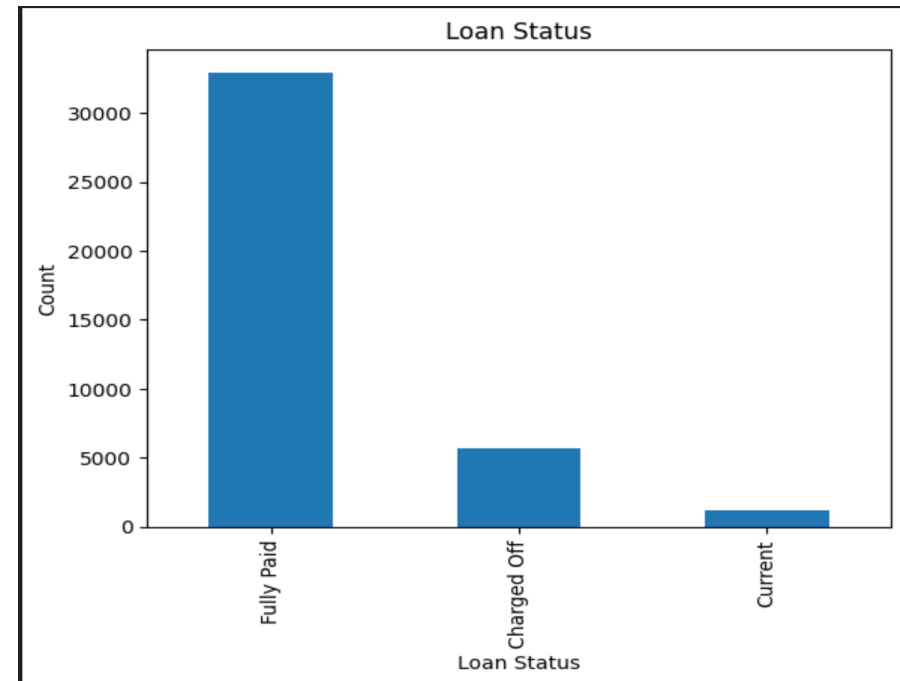
Time Frame of the Data – 2007 to 2011

- Time Frame Bar chart provides us the information on the increasing growth of lenders every year.
- We could see an exponential growth of 40% every year which is good sign.
- To know more about the potential defaulters list we will analyse loan amounts, interest rates, borrower details, loan status, etc.



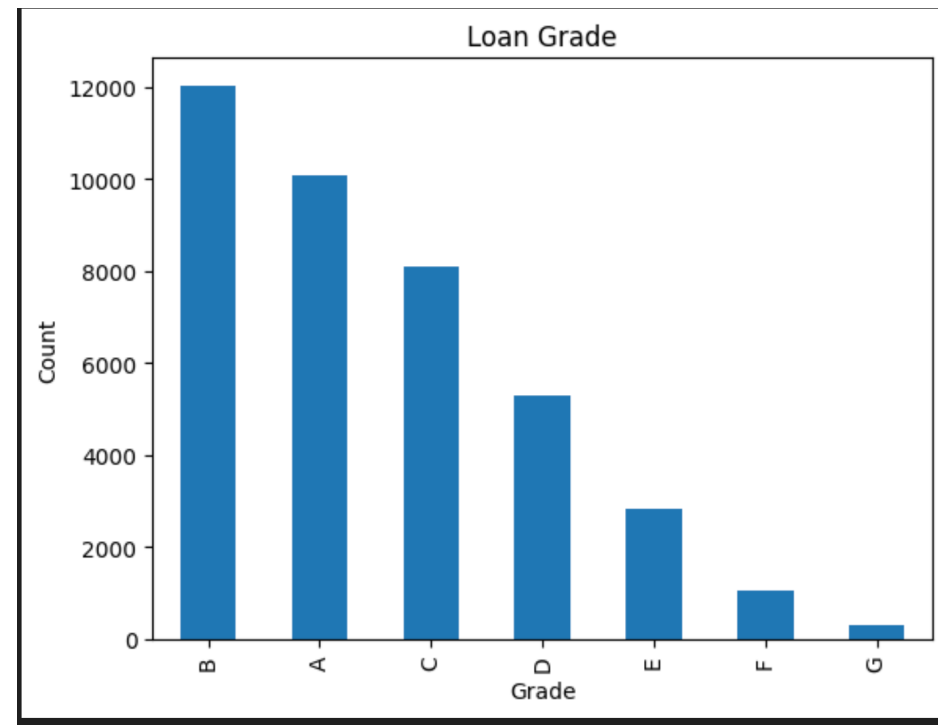
Loan Status

- Analysing the loan status we could see the majority of loans are successfully paid off, which is a positive indicator of borrower reliability.
- There is a significant number of charged-off loans, which may indicate some risk in lending practices.
- As per the Business objective to identify the risk factors in the lending practices we have considered Loan Status , charged of as the target variable



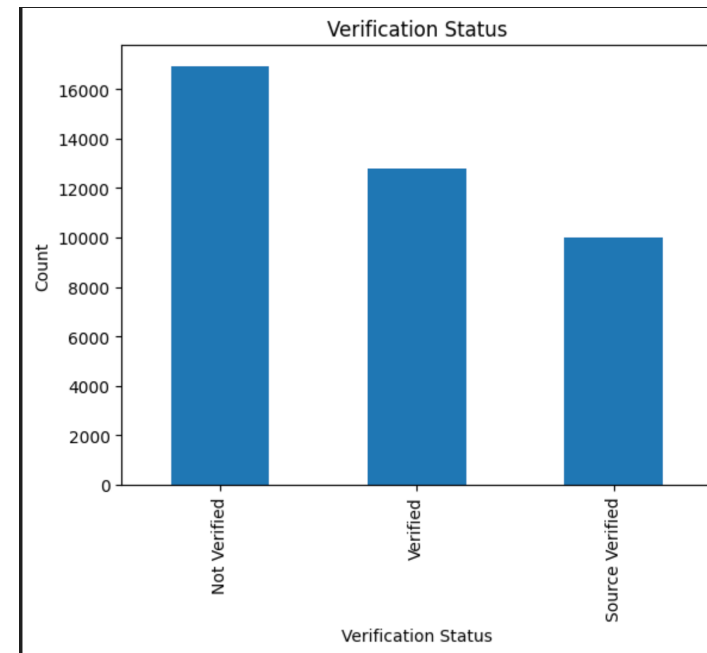
Loan Grade Assigned by LC

- Analysing the bar chart we could see Grade B has the highest count of borrowers and Grade G has got the minimum number of count.
- Though we don't have the extract data on how these grade are provided we have the data on the sub-grades .
- We have tried comparing the the loan amount with the combination of Grade and sub- grades to get more clarifications



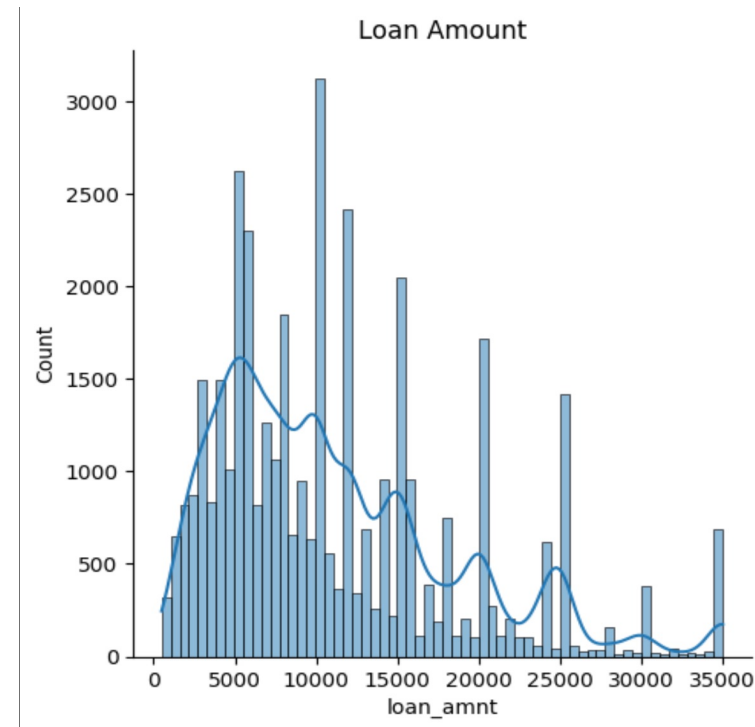
Verification Status of the Borrower

- Overall, the majority of the data remains unverified, with a significant portion verified, and a smaller portion verified using a specific source. This indicates that while some data has undergone verification, there is a considerable amount that has not yet been validated.



Loan Amount and Frequency

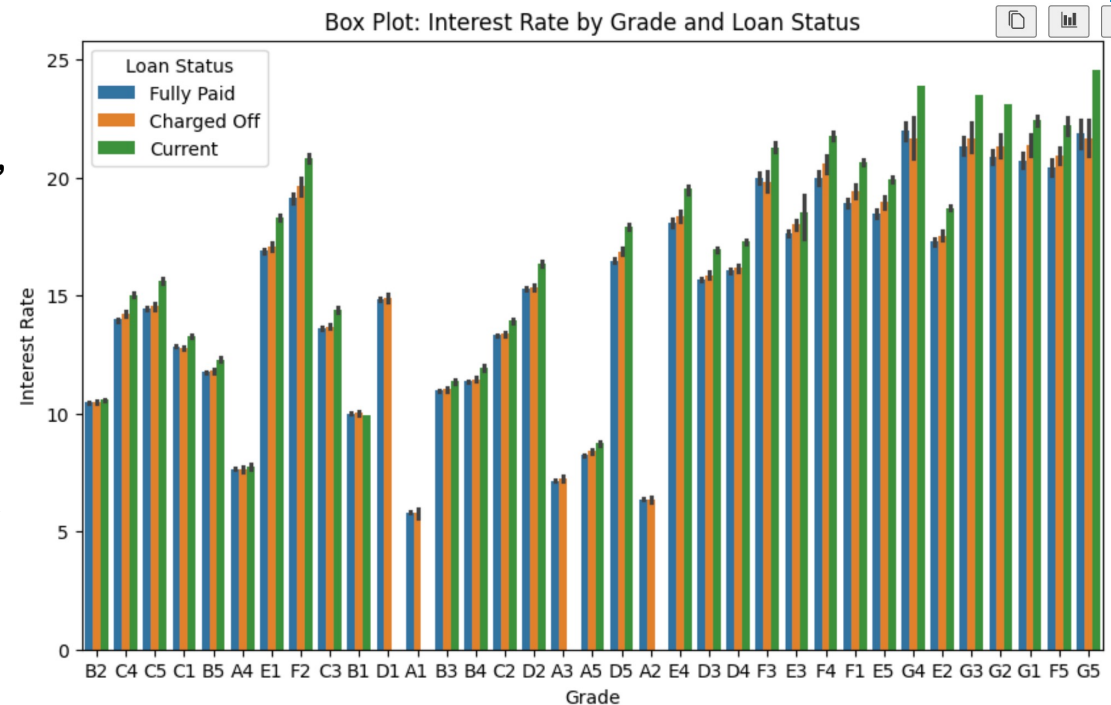
- The majority of loan amounts is between 5,000 and 15,000, with a noticeable peak around 10,000 to 12,000. This suggests that these are the most frequently requested loan amounts.
- The highest frequency of loans, with counts exceeding 2,500, is observed in the 10,000 to 12,000 range.
- Overall, the data shows that lower to mid-range loan amounts are more common, with a gradual decrease in frequency as loan amounts increase.



Bivariant Analysis

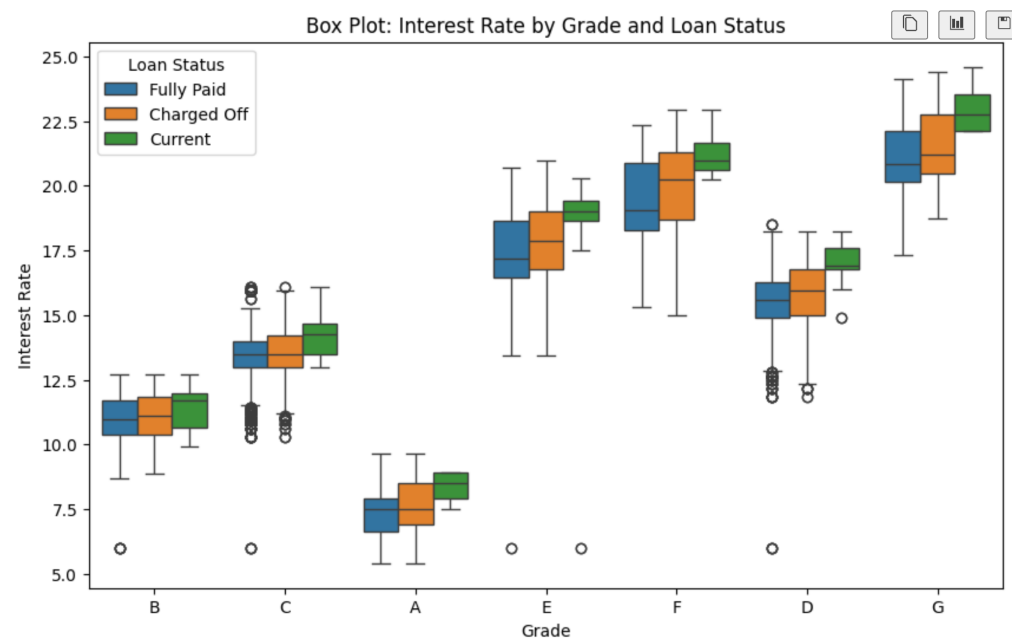
Loan Amount Vs Grades

- From the distribution of interest rates, we could see it's apparent that the median interest rates vary across different loan grades and statuses.
- The box plot indicates the spread of interest rates within each grade and loan status category, providing insight into the variability and central tendency of interest rates.



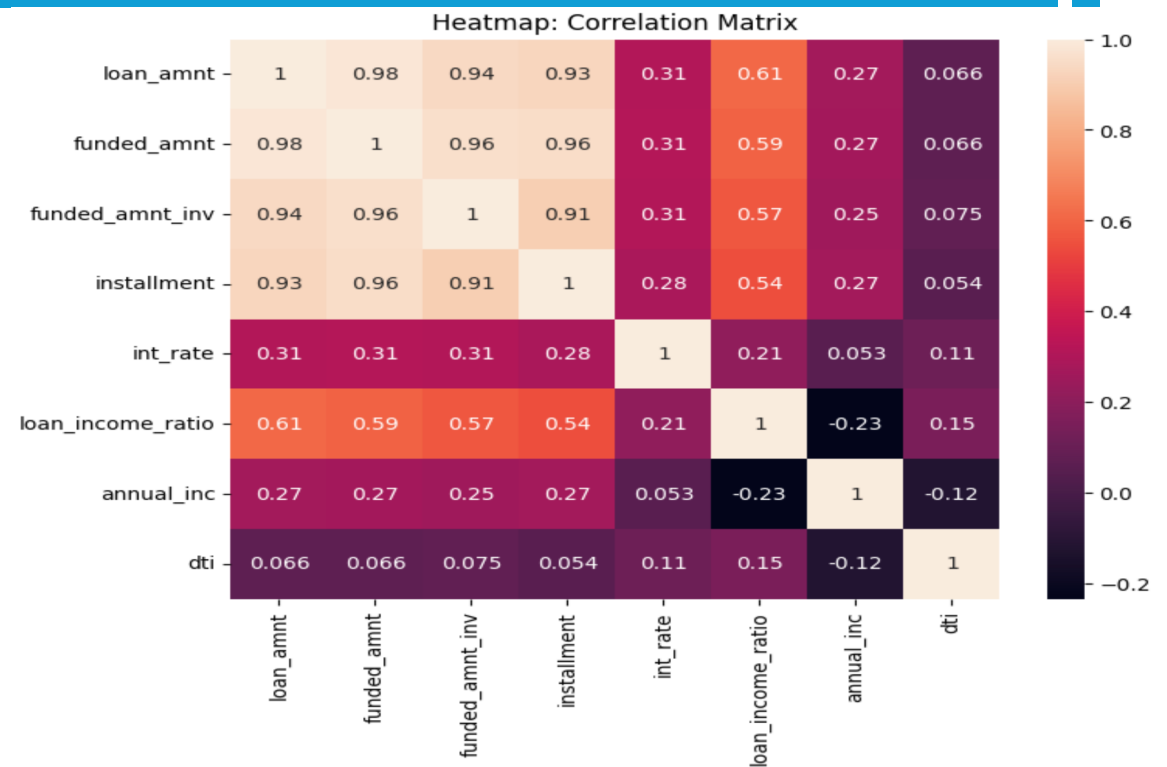
Box Plot: Distribution of interest rates across different grades

- From the distribution of interest rates, and grade shown as E, F and G are paying more interest rates



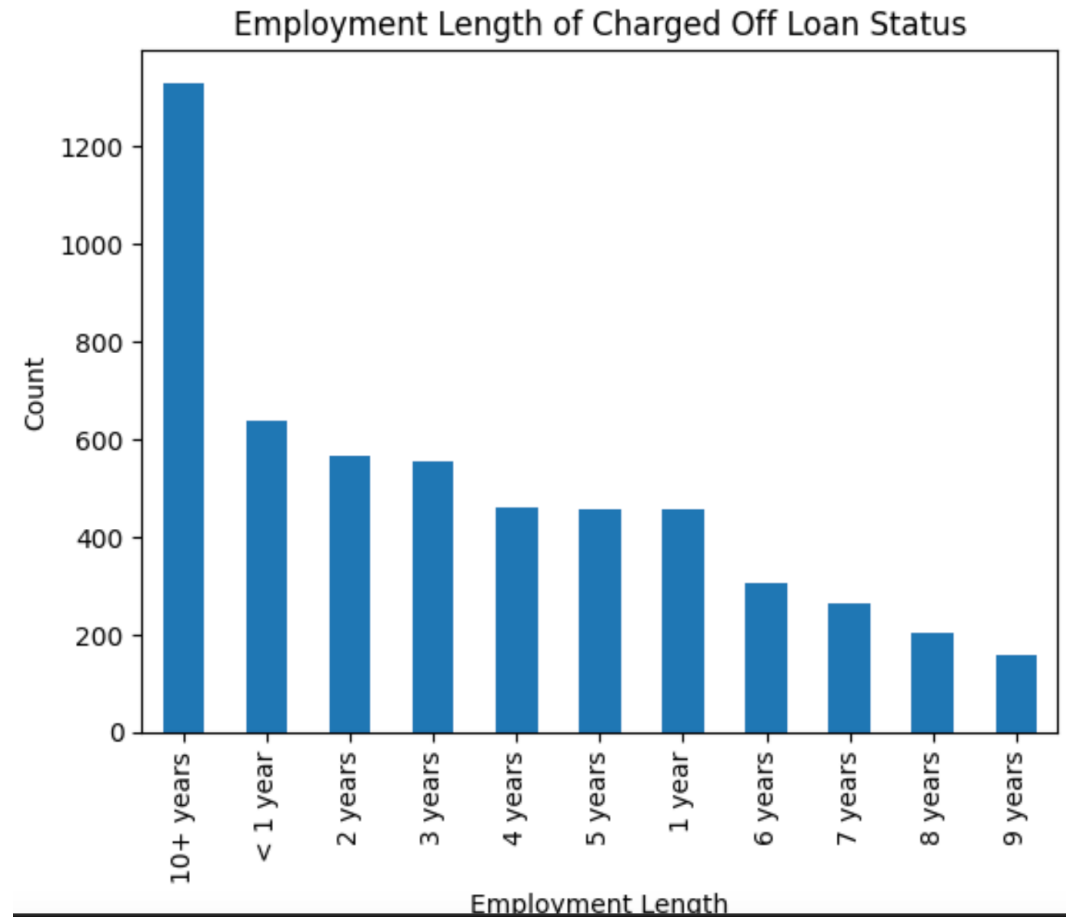
HeatMap: All continuos variables Insights

It is clear from the Heatmap that how 'loan_amnt','funded_amnt' & 'funded_amnt_inv' are closely interrelated. So we can take any one column out of them for our analysis.



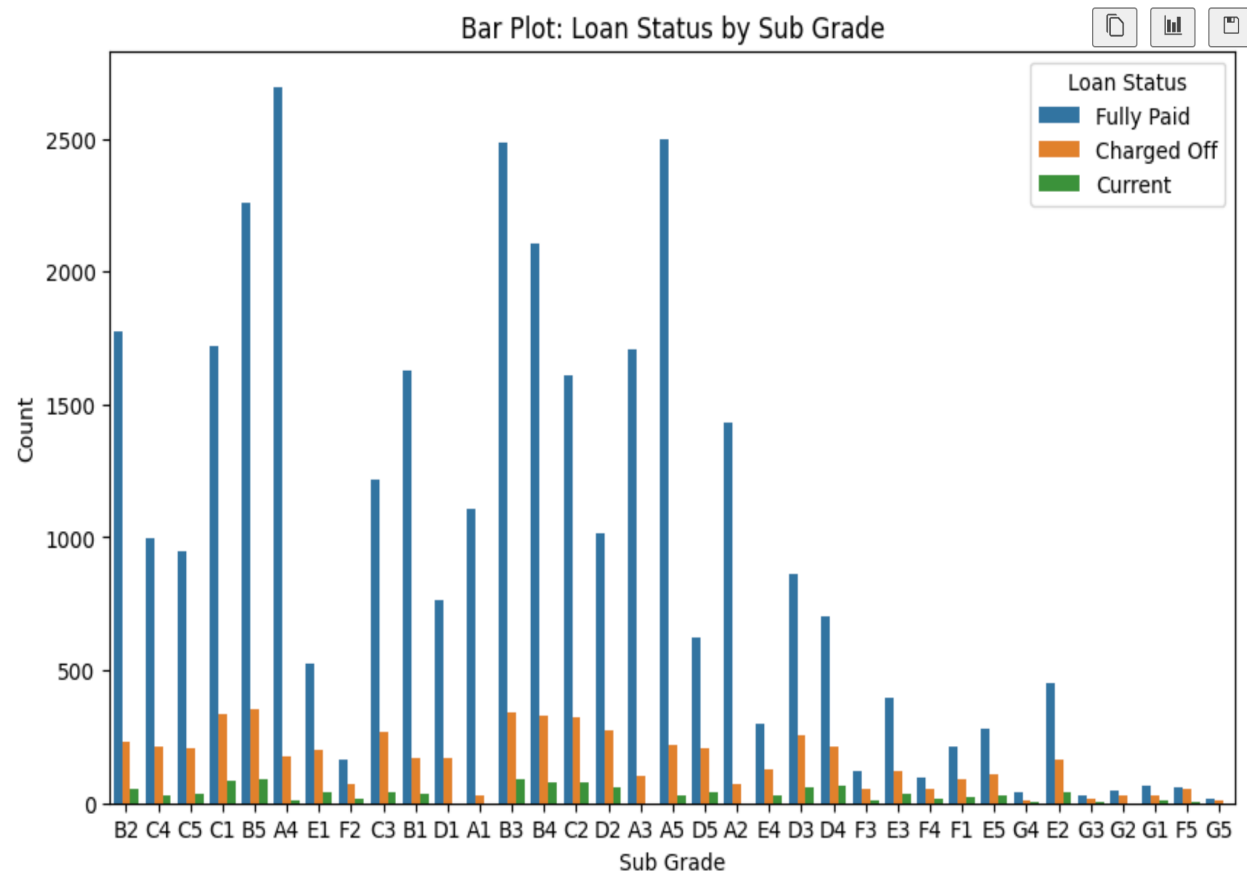
Employment Length of Charged Off Loan Status

Applicants less than 1 year of experience or more than 10 years experience probable of charged off



Bar Plot: Count of loan statuses by grade

The Bar Plot shows for Sub Grade by Loan status that the B5, A5 and A4 are good at paying loan



Outcome of EDA

- Borrower credit scores are a significant factor in loan performance.
- Higher interest rates are associated with higher default rates.
- Income levels of borrowers impact their ability to repay loans.
- Certain loan purposes have higher default rates compared to others.



Thank You