Lending Club Case Study

Sivapriya Srikakolapu

Vasanth Kumar

January 17, 2024

IIIT- B ML & AI

Problem Statement

A Lending Loan company need an analytics hand in identifying their operational risks and finding the risk parameters effecting their business. As part finding risks, this case study is more concentrated on finding risk parameters effecting the Loan defaulters.

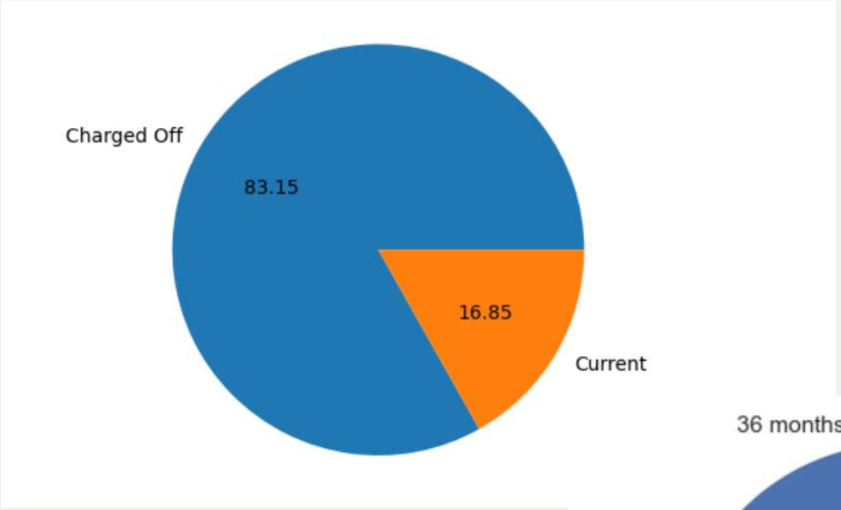
Exploratory Data Analysis (EDA) :

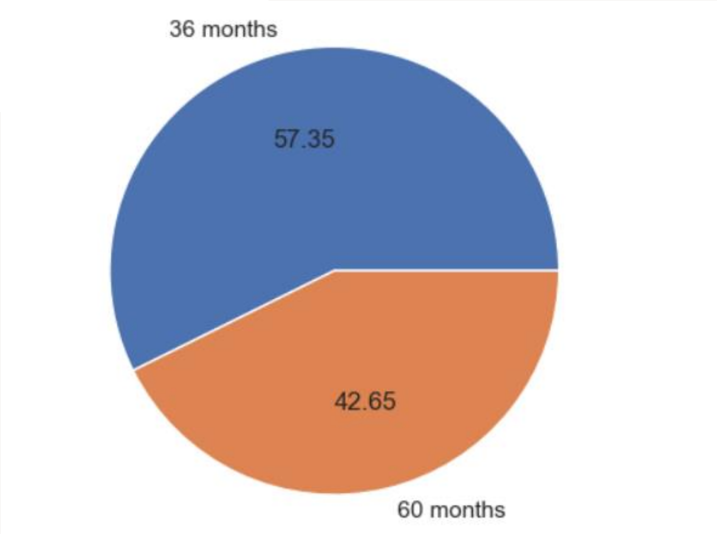
Initial Data manipulation is the necessary step in EDA, where data is explored, cleaned and formed a useful shape

This is performed in the Loan dataset, identified nulls, assigned a default value, excluding unnecessary fields from the data set, type conversions of the fields, grouping the data for graphical insights.

Going forward, various analysis and insights are expressed in detail with respect to the cleaned data set resulted from above manipulation.

Univariate Analysis



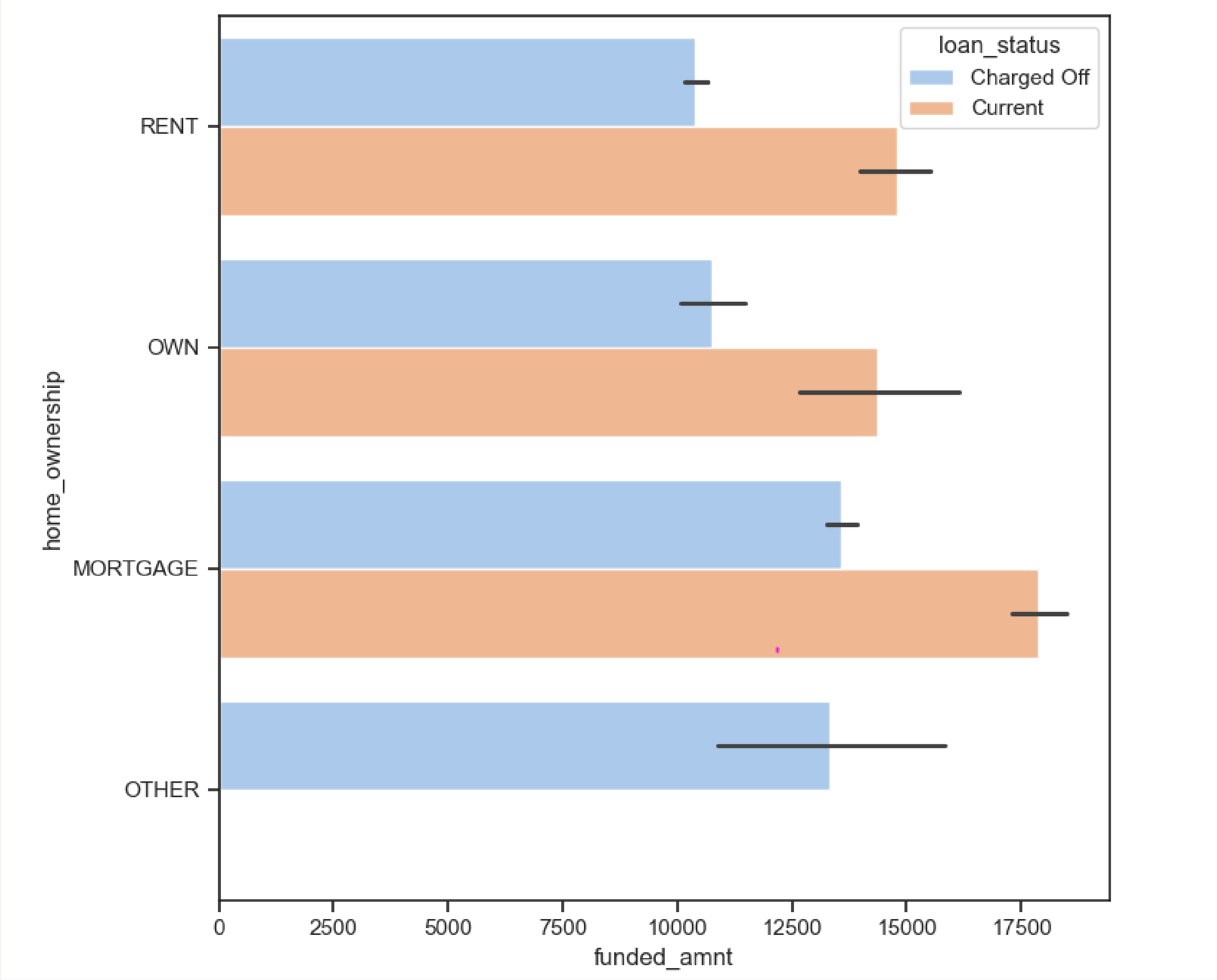


Using Univariate analysis, these insight depicts out of total loans, % of Charged off loans are higher and charged off loans have low% of loan term which implies to high risk of the Lending Club. So, funding loans for longer term will decrease the risk and charged off % .

Univariate
Segmented Analysis

Segmented Univariate analysis performed on Loan status, Funded amount and Home ownership depicts that more amount has been funded to Mortgage and rented ownership when compared to rest others.

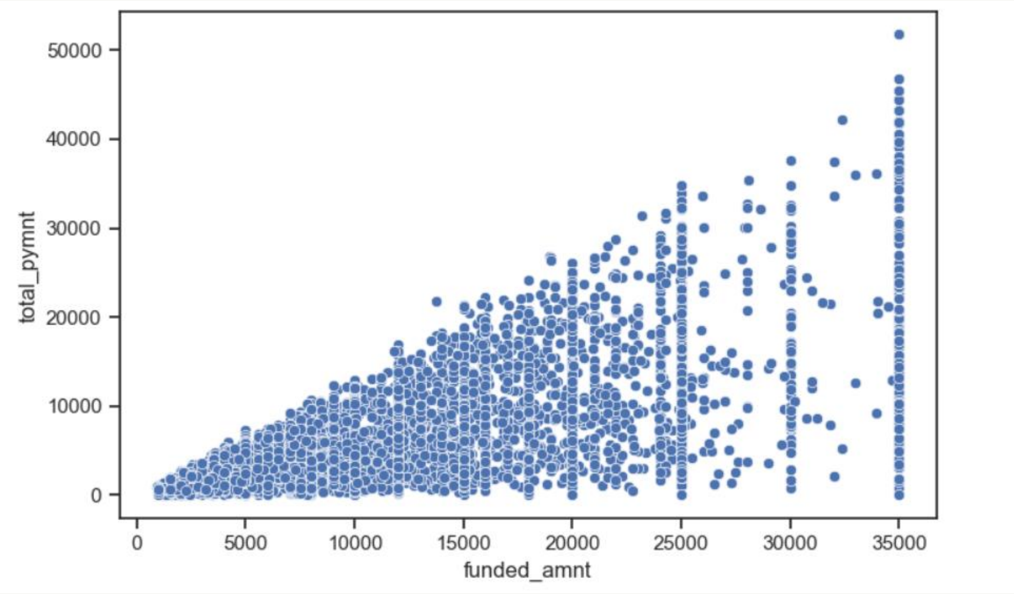
Considering Loan status, Charged off of loans is high in Mortgage and Other ownership, so distributing the funded amount to Own and Rented membership would limit the risk of Charging off further.

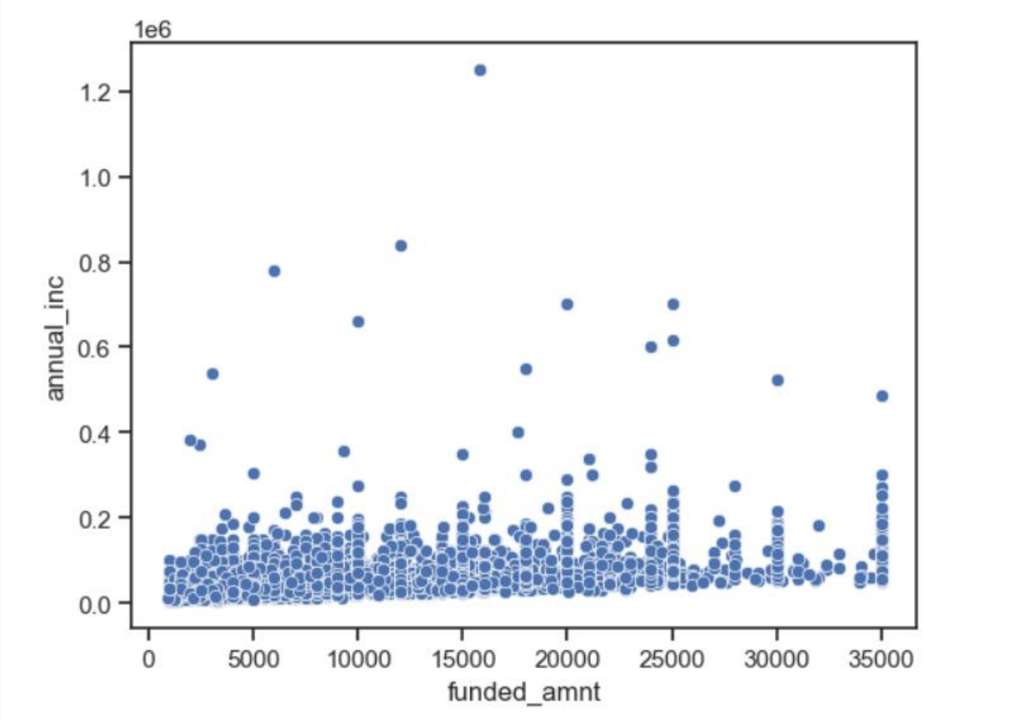


Bivariate Analysis

Performing Bivariate analysis between Total payment amount vs Funded amount and Annual income vs funded amount on Charged off loans resulted to employee with lesser annual income is also funded higher amount, which can be a reason for Loan charge off . This can be easily identified by density of the scatter plot at all the levels og funded amount.

Analyzing Total payment correlation with funded amount for charged off loans depicted that more loans have been charged off for the lesser funded amount on the exploring scatter plot density.

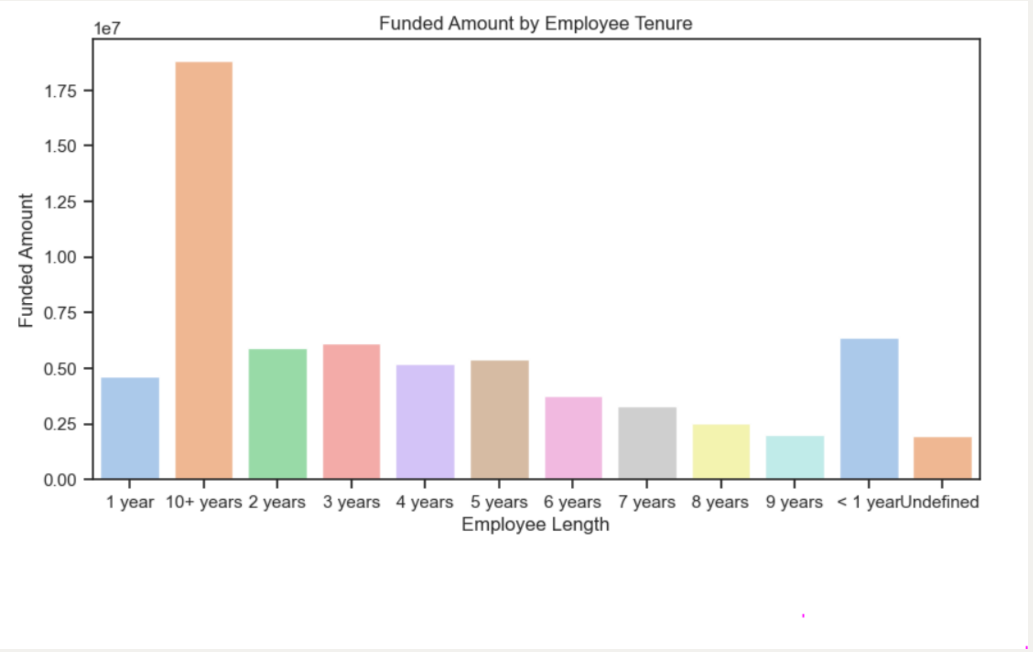


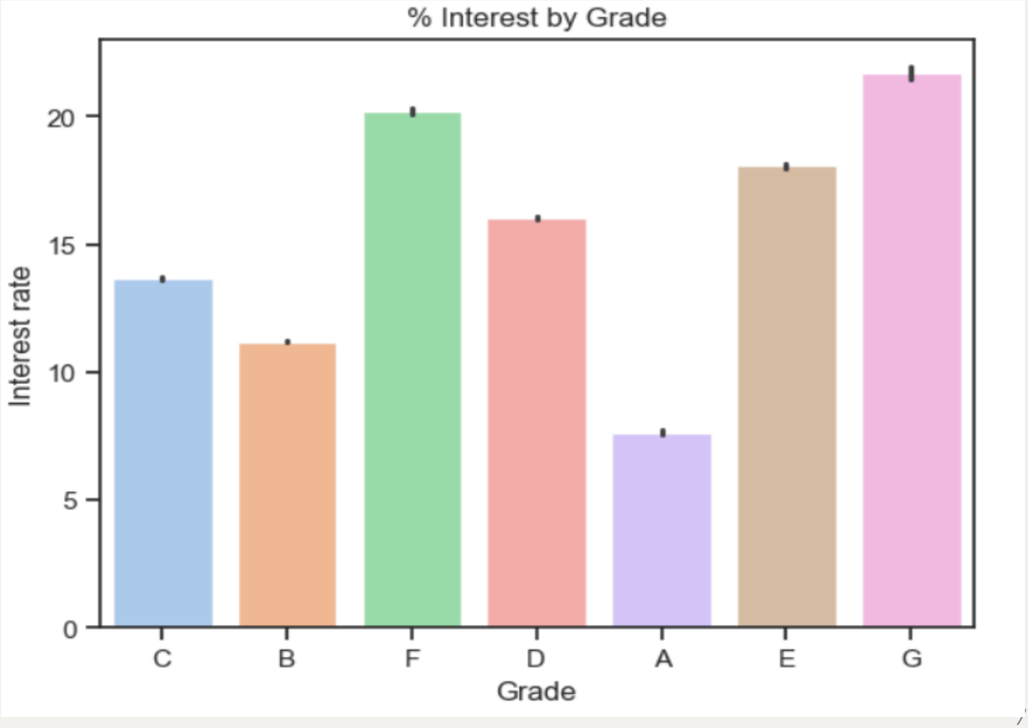


Bivariate Analysis

On correlating Employee experience (employment length) to Funded amount of charged off loans, we see most of the loan amount is funded for higher experience employees , which can be distributed along other buckets of experiences for low risk.

Comparison of interest rate across different grades of borrowers (employees) , Interest rate is ‘F ’ and ‘G’ can be normalized for decreasing the charge offs in the loans.

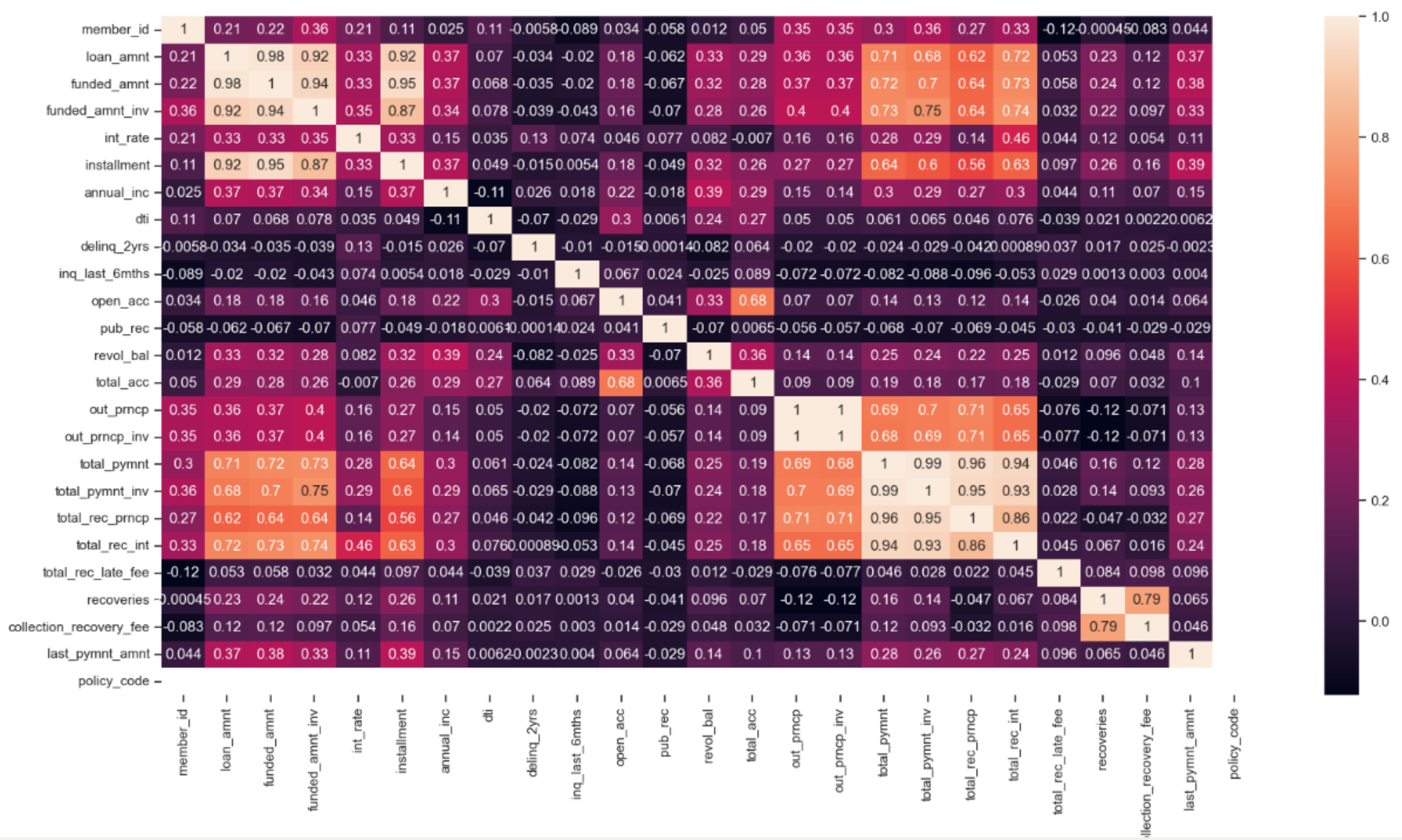


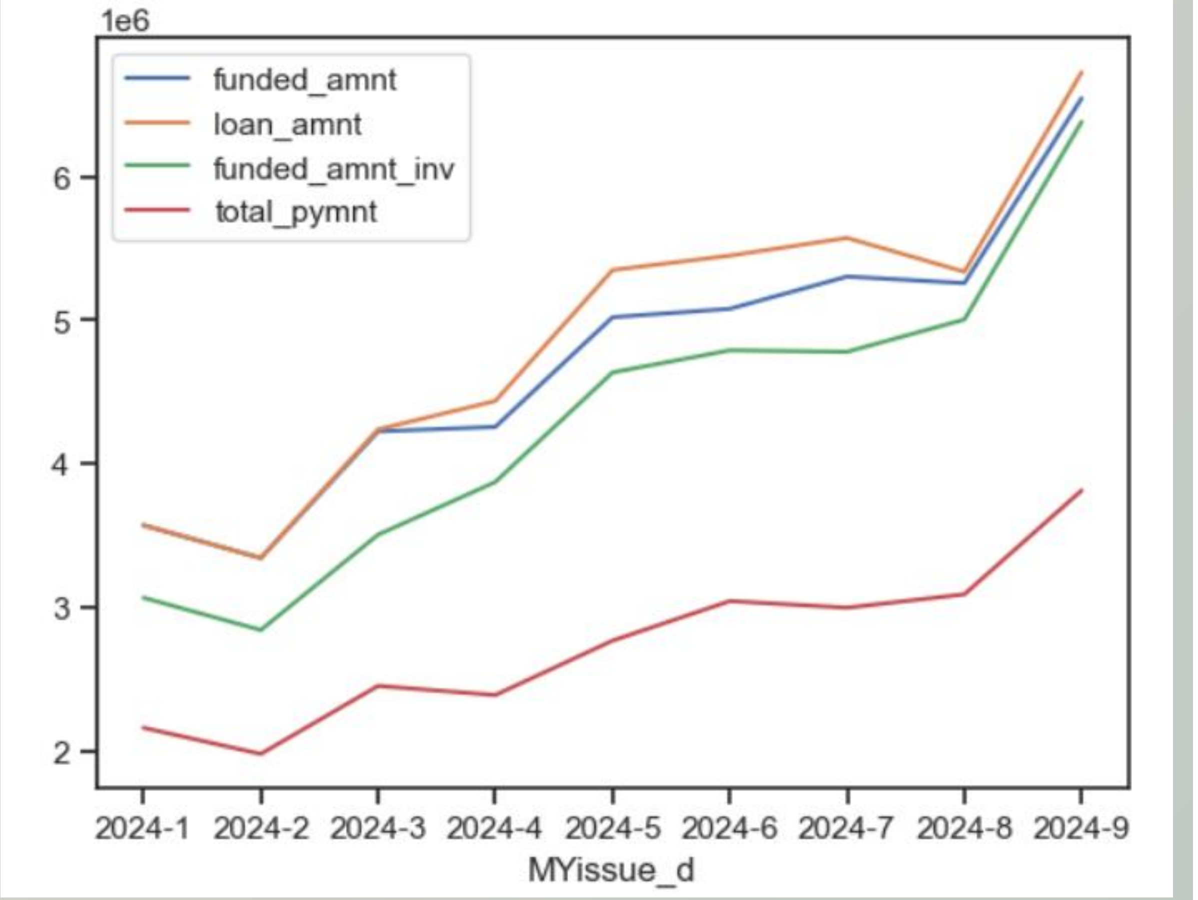


Multivariate Analysis: Correlation Matrix

In Multivariate analysis, correlation matrix, easily displays the correlation of complete data on a single frame for easy and better understanding.

Through this matrix, we can decide correlation between Loan, funded, total payment amount are higher due to charging of the loans as they are indirectly proportional to one another depicting the Charged off loans principle.



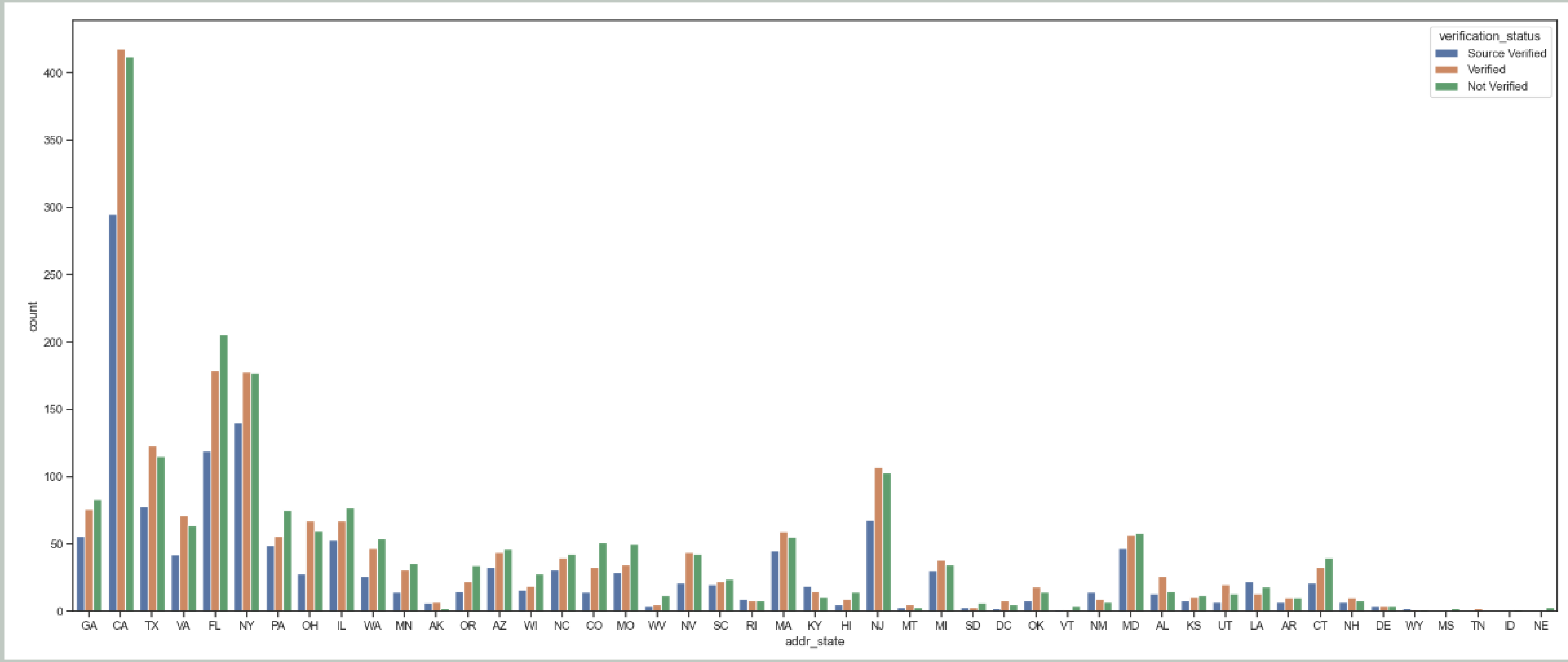


Multi variate Analysis:
Trend Analysis on time period

Multi variate analysis of amounts of charged off loans across different time period shows us the trend of payments with respect to the Month and year of issue date.

As we observe Funded amount is varying from loan amount from March24 and the trend of it with respective to other loan amounts is also same. On the trend, we can conclude that February, April and October has dip in the amounts, might be a reason for less risk for Lending club operations .

Distribution of Charged Off Loans:



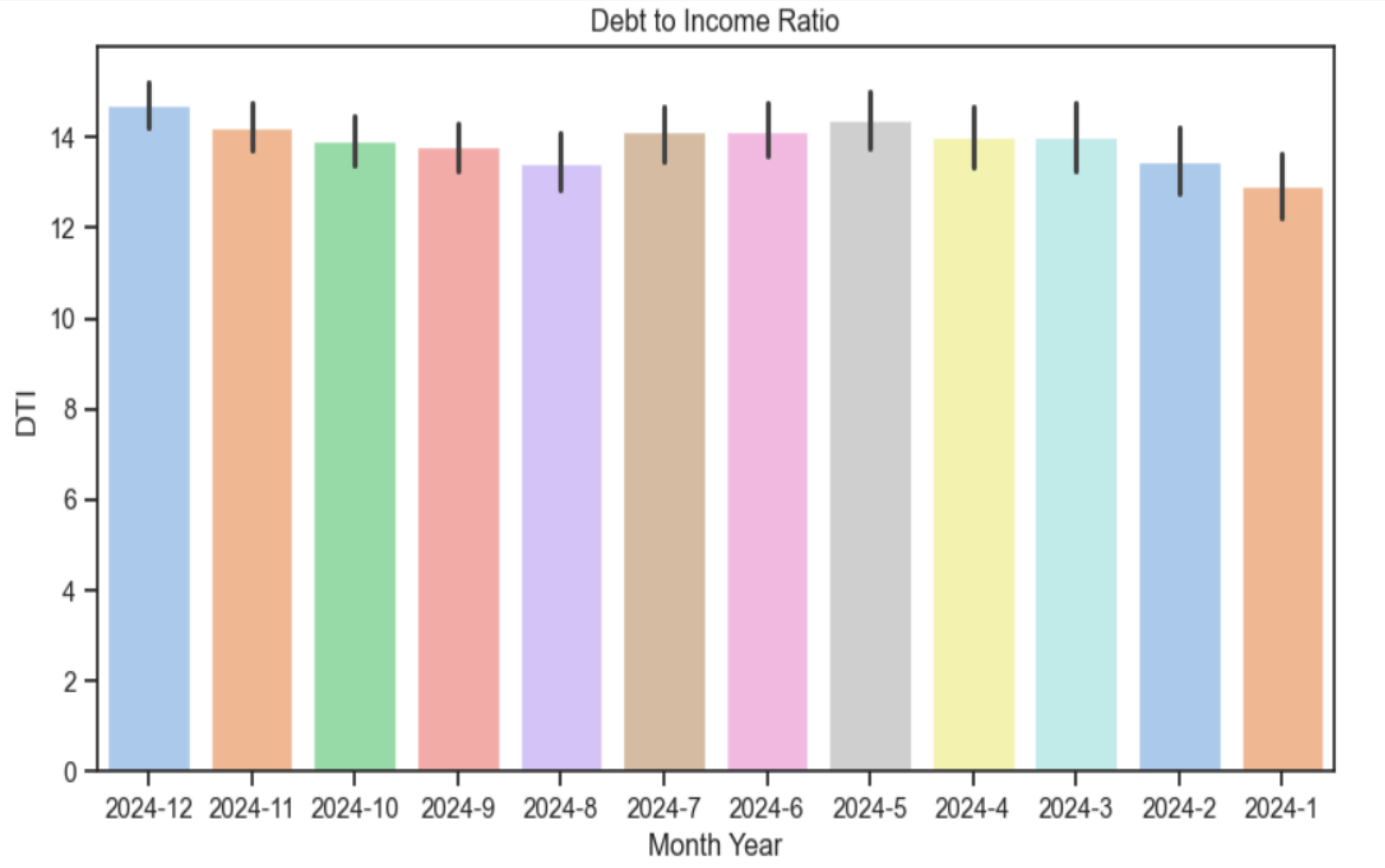
On summarizing charged off loans across different states, Source verified loans are significantly lower compared to rest which is of high risk and distribution of loans in specific location can also be at high risk even if California has high income employees

Debt to Income Ratio (DTI):

DTI determines the ratio of debts of lending club to investors vs income receiving from borrowers through lending money.

Lower the ration higher the profit earned by the Lending club.

While plotting DTI against trend of periods, it is evident that DTI is bit constant but starting at lower DTI, ending to higher by the year end which causes high risk.



Conclusion

This concludes our presentation. Thank you for taking the time to view it. We hope that you found the story that we unlocked from the Lending Club loan data to be informative and useful.