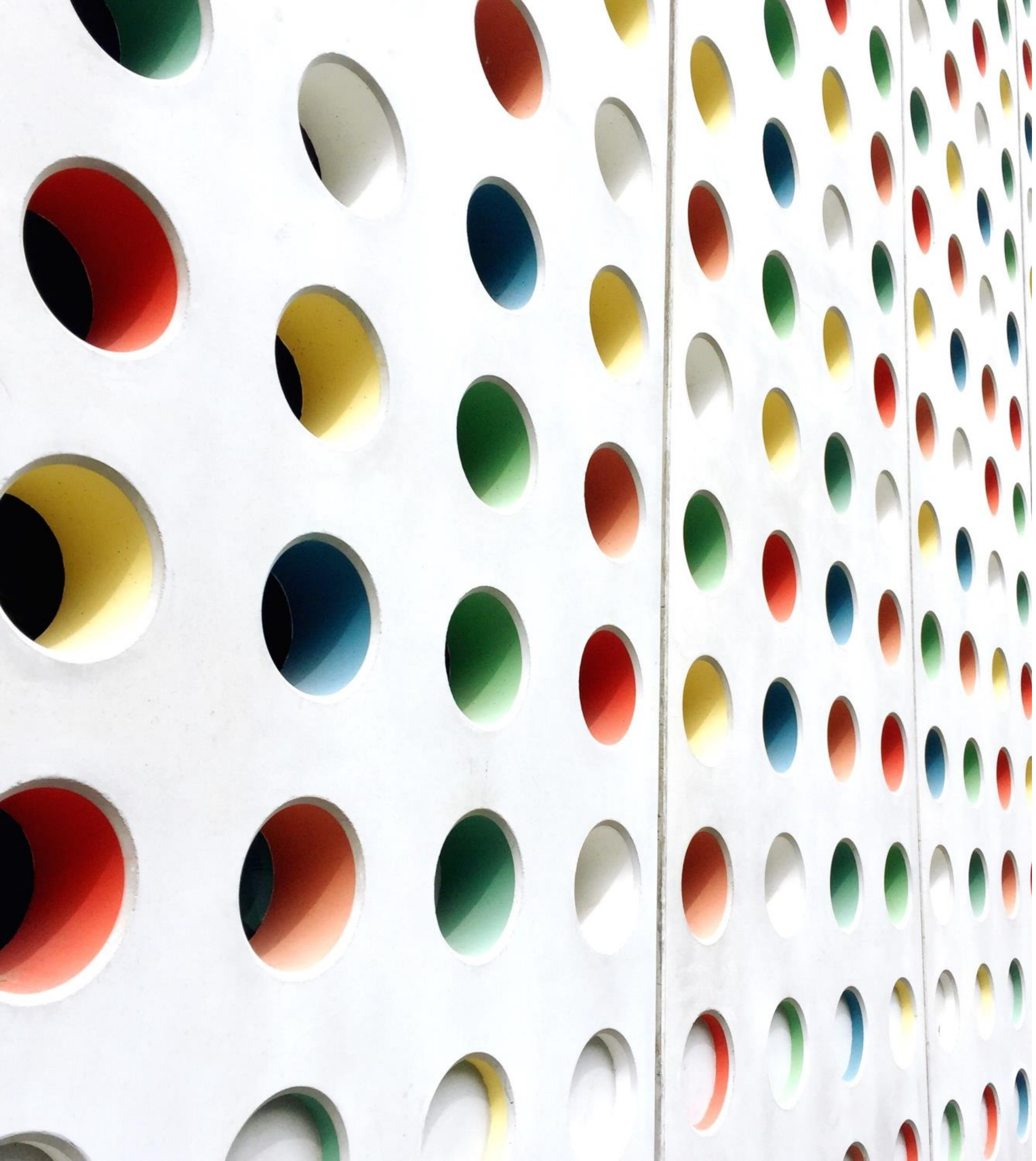


Lending Club Case Study

CHRIS FLOOD FLOODYC@GMAIL.COM

JOSEPH HANNA ECONNECTFOREVER@GMAIL.COM

UpGrad & IIITB
Global Program in Machine Learning and AI

2023-02-08

Key Takeaways

- The overall risk of lending was found to be 14%.
- Loan Grade, Home Ownership, Loan Purpose, Interest Rate, Employment Length and Loan Amount were the greatest drivers (variables) in determining a potential default.
- For certain loan types interest rates should be increased.
- For certain borrower types the amount funded should be reduced by 30%
- For borrowers with higher grades, the interest rate can be reduced.
- For borrowers with unverified income the interest rate should be set highest.
- For borrowers with fully paid homes decrease the interest rate. Raise it for renters.

Problem Description

Lending Club data analysis involves exploring and analyzing data from the peer-to-peer lending platform, Lending Club. The data can be used to understand trends and perform risk analysis. Analysts can use various statistical and machine learning techniques to gain insights and make informed investment decisions.

A default occurs when no payment has been received for over 120 days. This requires 'charge-off'. This is a loss to the company. The goal is to identify the risk associated with each borrower so as to minimize risk and hence loss. This is performed by understanding what causes lead to default. The answer is in the data provided.

Approach

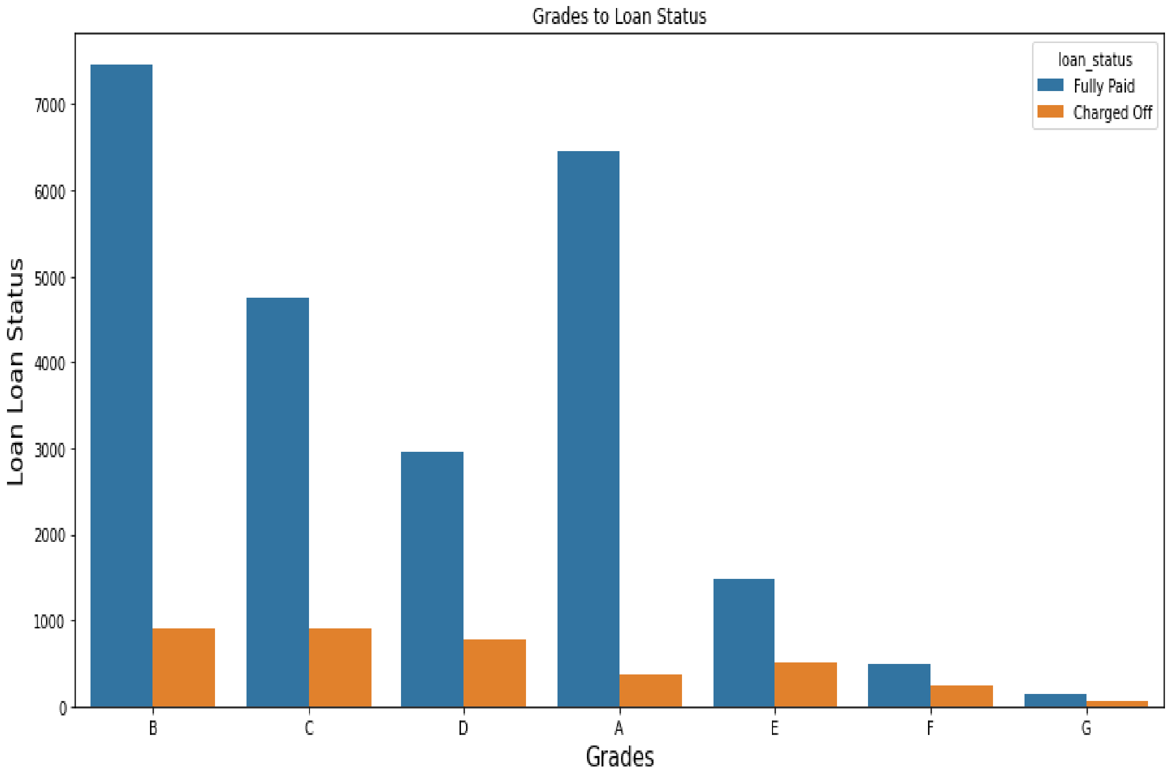
A dataset was provided. The data was imported to a Python Notebook, cleaned and analyzed. The analysis was developed after the process of cleaning and sorting the data was completed. Any data not useful to the analysis was discarded and the analysis then focused on data around those borrowers who paid their loans and those who defaulted.

Technologies used included Python as well as Pandas, Numpy, Matplotlib and Seaborn libraries. The information was presented in the form of tables, charts and heatmaps. The observations from each stage of the analysis and from the visual presentation of the data was gathered into a set of conclusions and recommendations to address the business objectives as stated.

Categorical Univariate, Categorical Bivariate, examining data in detail for outliers, and finally producing a correlation heatmap allowed the analysis to make definite conclusions. These conclusions allow the lender to reduce the risk that a loan will be provided to a borrower with a higher likelihood of default..

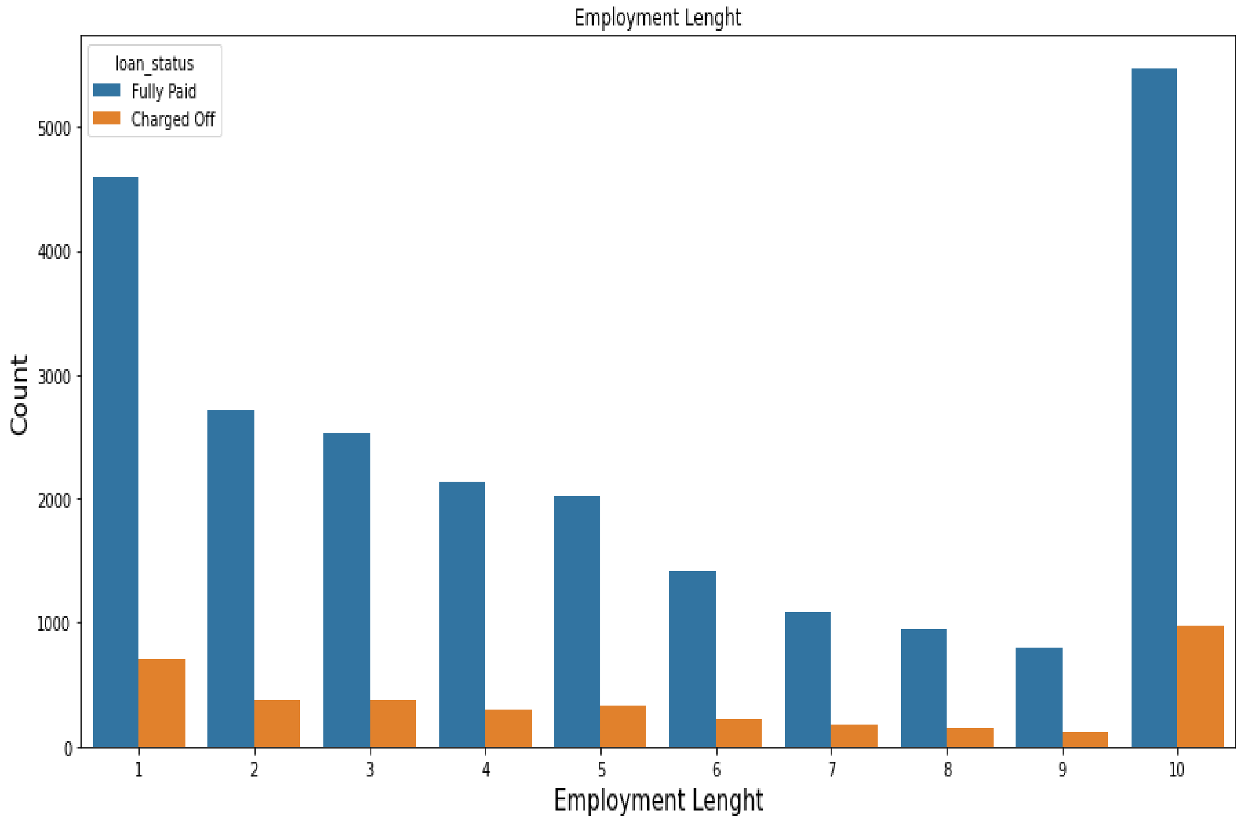
Observations – Grades to Loan Status

- Grades F and G showed the highest default rate of 32% and 29% respectively.
- Loans of these grades represented the highest risk of default amongst the grades.
- The largest volume of loans belonged to grades A and B.



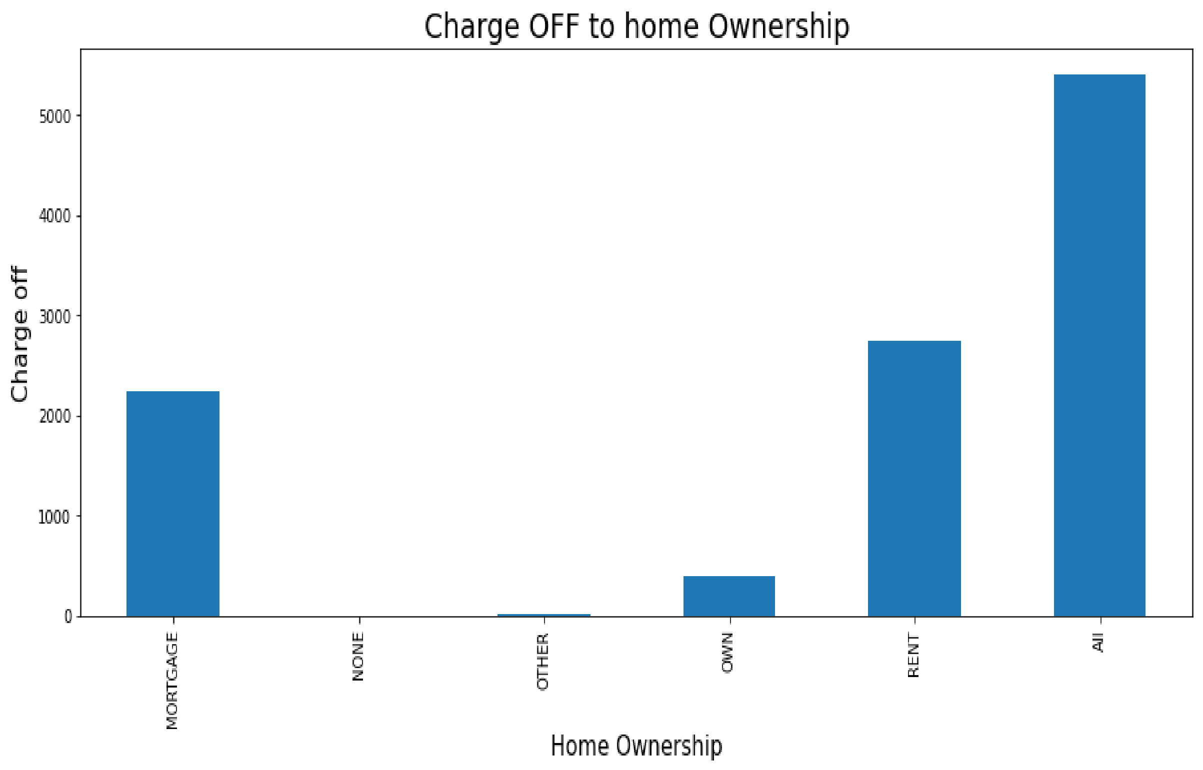
Observations – Employment Length

- For borrowers whose employment was 10 years or more a charge-off rate of 15% was observed.
- The highest volume of loans were taken by those with greater than 10 years employment.
- Borrowers with 1 year of employment represented the second most risk in terms of default.



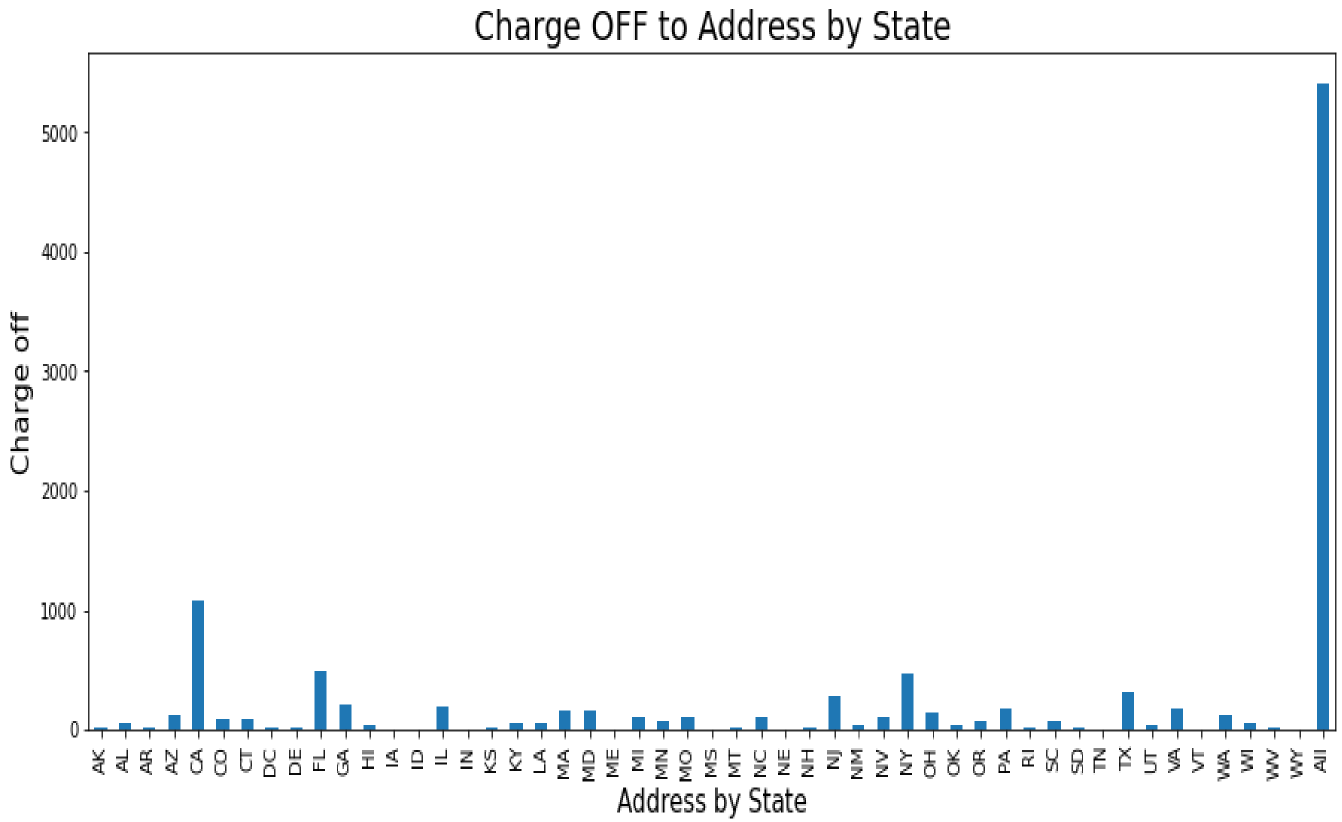
Observations – Home Ownership

- From the perspective of home ownership, the greatest risk of charge-off was associated with those who were renting their accommodation.
- Those with a mortgage represented the next greatest risk.
- Those who owned their homes represented the least risk of default.



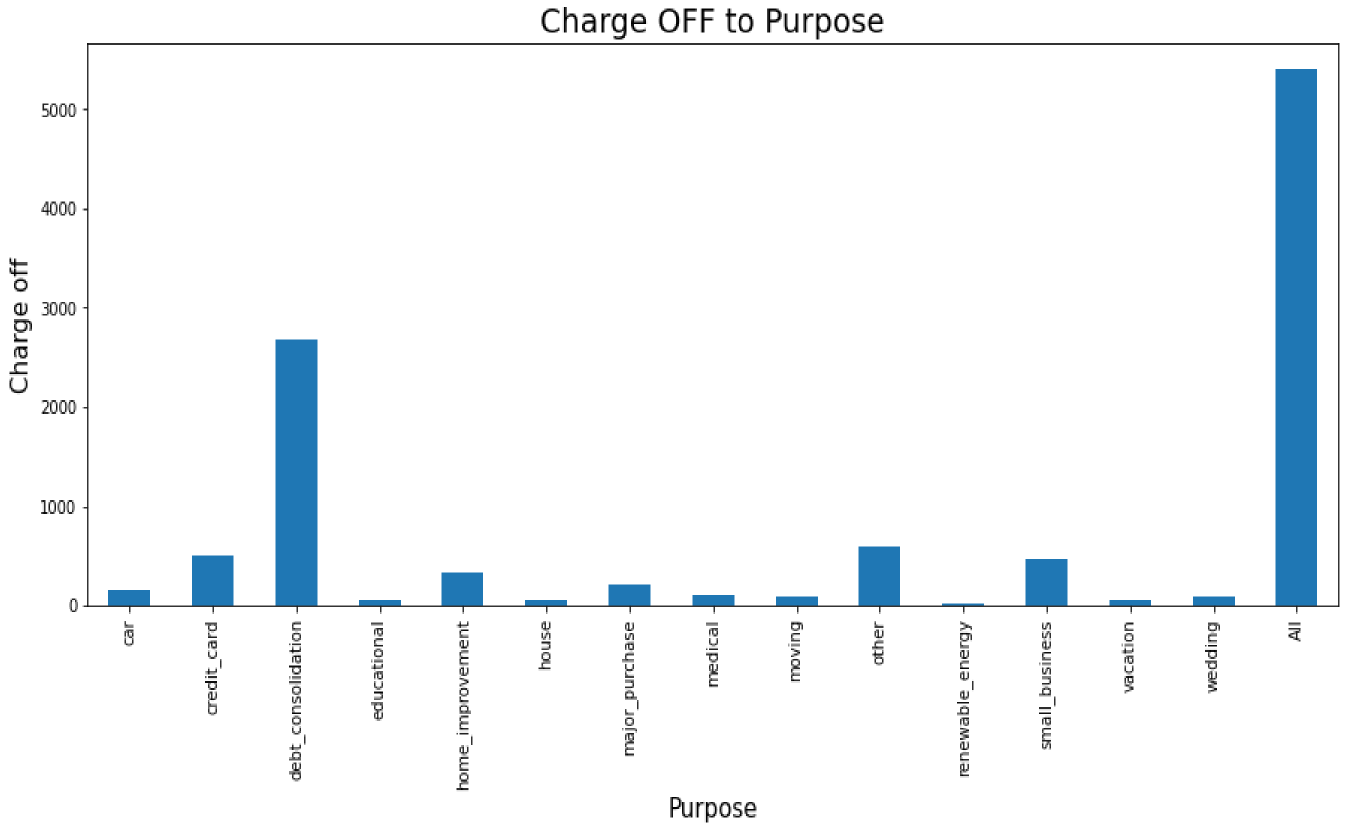
Observations – By State

- Borrowers residing in California represented the greatest risk of default.
- Followed by Florida and New York.



Observations – By Purpose

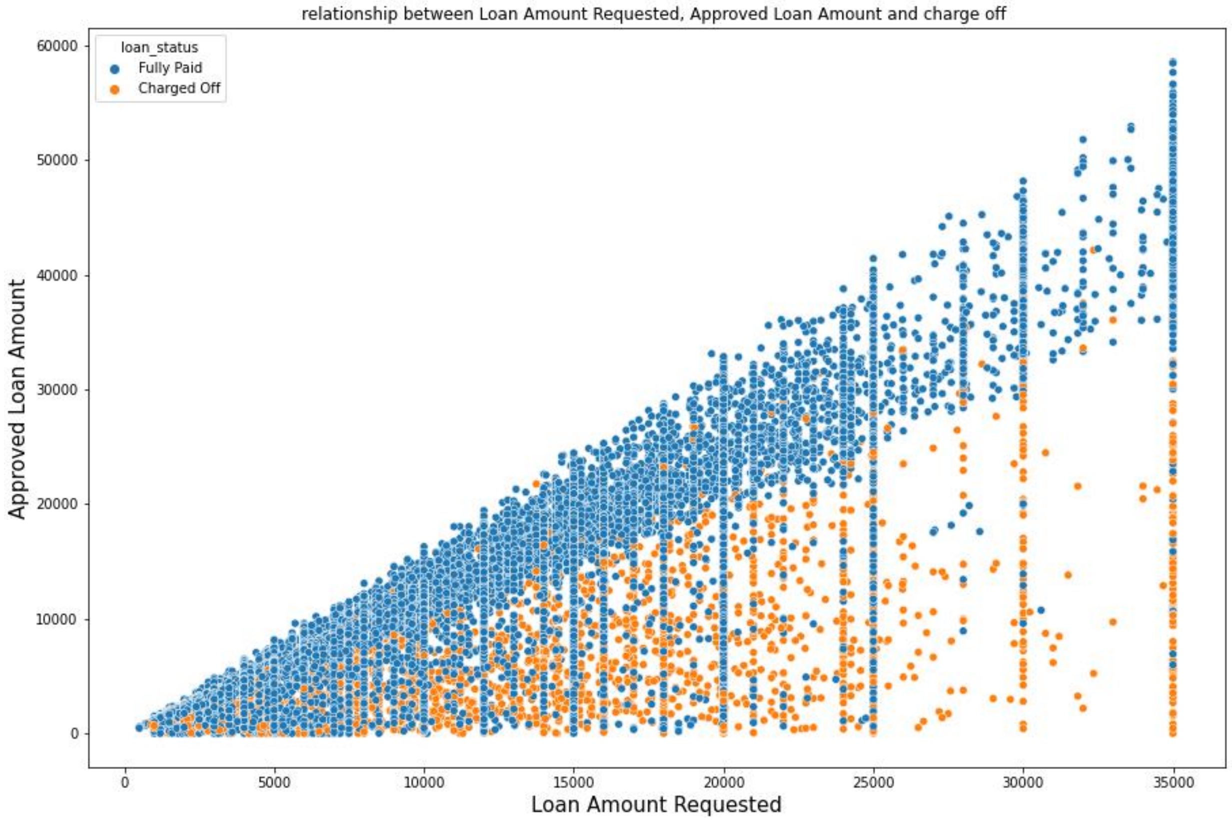
- Borrowers who required a loan for the purposes of debt consolidation represented the greatest risk of default from the perspective of purpose-of-loan.
- Credit-card was the 2nd category showing greatest risk.



Observations – Amount Requested, Amount Approved & Charge-Off

- This plot shows the relationship between how much was requested to how much was approved and which amongst those were charged-off.

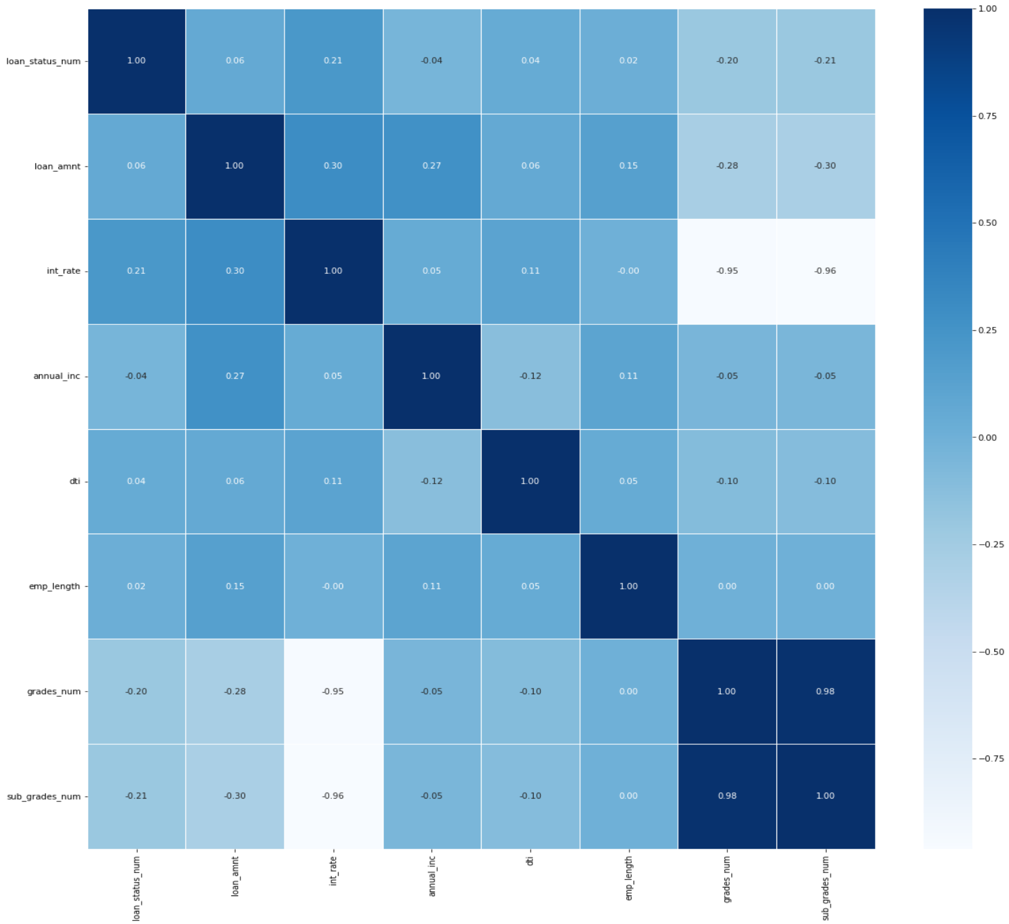
- It appears that when the amount approved aligned with the amount requested, the number of loans that went into default was lower. This relationship was very linear.



Observations – Correlation Heatmap

- The heat map has shown that there exists strong correlations between interest rate & loan amount, annual income & loan amount, interest rate & loan status.

- There is low correlation between grades & employment length also employment length & debt-to-income ratio.



Conclusions

- The study analyzed the pre-loan information of the dataset, ignoring the post-loan information as it is not relevant to the lender's decision-making process. The overall risk was found to be 14%. The goal was to identify the major factors affecting the loan outcome.

- Loan Repayment: The majority of fully paid loans were more than 80%, while loans with more than 10% charge off were less frequent.

- Loan Grades: The largest volume of loans belonged to grades B, A, and C. However, grades F and G had the highest default rate of 32% and 29%, respectively.

- Home Ownership: Borrowers who rent showed the highest charge-off rate, while those who own their homes had the lowest. Borrowers still paying for their homes through mortgage had higher charge-off rates and higher debt-to-income ratios.

- Loan Purpose: The highest probability of loan default was observed for debt consolidation and credit card payoff loans.

- Charge-off Rate: The majority of charge-off rates were between 0.11 and 0.16. Borrowers with lower interest rates tended to fully repay their loans.

- Employment Length: A charge-off rate of 15% was observed for borrowers with employment length over 10 years.

- Loan Amount: A strong relationship was observed between the loan amount requested and the approved loan amount. Borrowers who received a lower amount than what they requested showed a spike in the charge-off rate.

Recommendations

Based on the findings of the study, the following recommendations are proposed:

- Raise the interest rate for loans with purposes of debt consolidation and credit card payoff.
- Decrease the amount of loan funded by 30% for borrowers with grades F and G and their subclasses.
- Lower the interest rate for borrowers with grades A, B, and C.
- Set the highest interest rate for borrowers with unverified income.
- Decrease the interest rate for home owners with fully paid homes and raise it for home renters.