

Financial Management and Risk assessment of Lending Club

Introduction:

Lending club is an online platform that helps borrowers find personal loans, financing, and business loans. The club joins investors with the borrowers which allows investors to get great returns. One of the major goals of the company is to improve the loan default rates. The management's strategy to make this improve is the following:

- (1) Identify the risk of loans quantitatively (in numbers) from the credit history of the borrowers. This will help to determine whether or not to approve the loan request for a particular borrower.
- (2) Estimate how much principal will be returned from a particular borrower after completion of the loan term.
- (3) Identify the management strategies that improve the loan default rates.

In this report, we use exploratory data analysis and machine learning techniques to address these problems.

Lending Club's Performance:

The Lending club began its service in 2007. The club gained popularity quickly after it showed up in the market. In the year of 2007, the club funded only a couple of hundred thousand dollars to the borrowers. This has become in the order of several million in 2016. These numbers show rapid growth of the company's business.

In average, only 85% of the borrowers pay the loans in time. The remaining customers make late payments or are charged off. The charged off loans goes through the loan collections and some of that is recovered with a collection fee. The major goal of the company is to improve the fraction of late/charged off payments. But Over the years, this number has not improved.

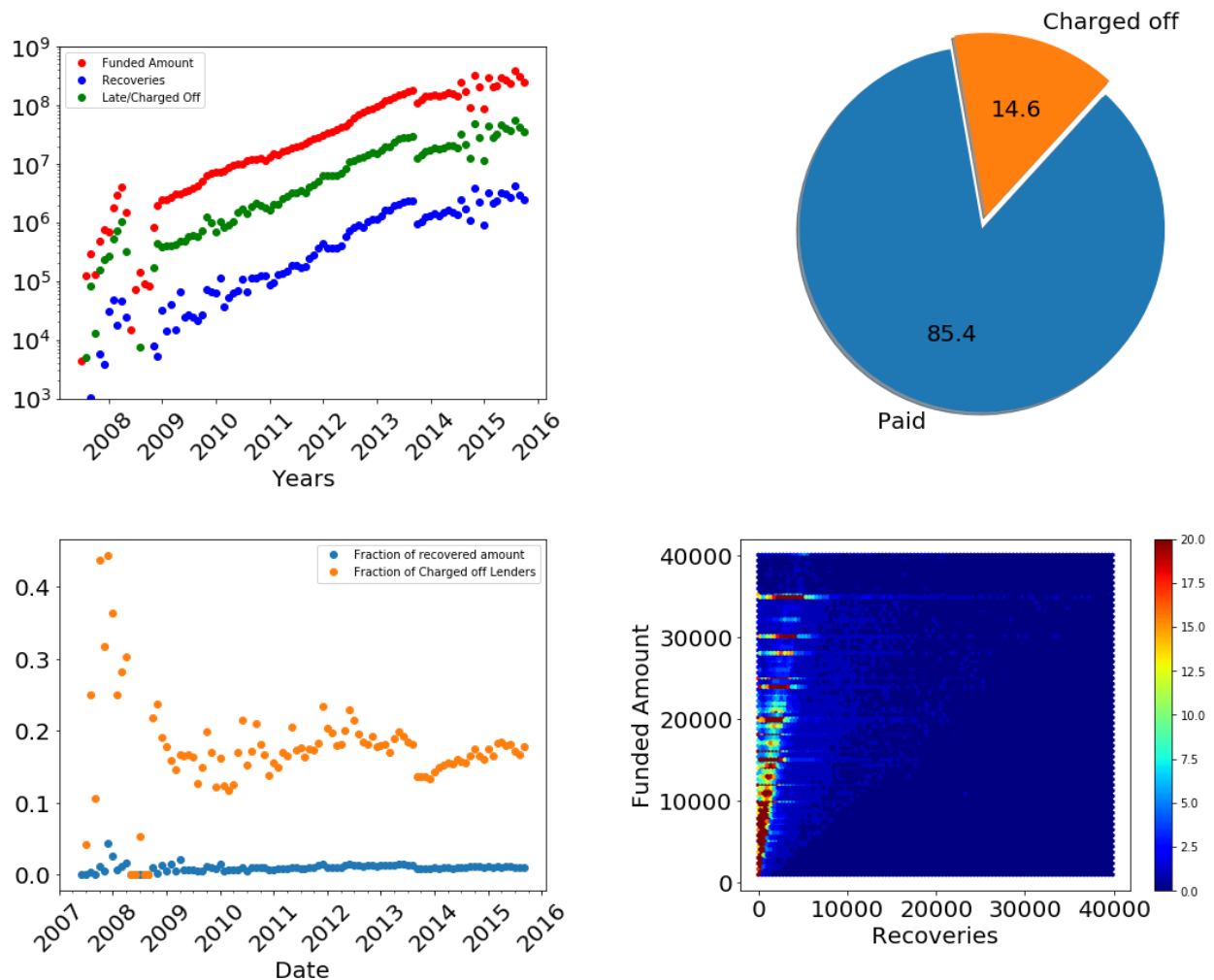


Figure 1: (upper left). Performance of the company over the past years. (upper right) The proportion of paid and late/charged off loans. (lower left) The fraction of charged-off borrowers and the amount recovered from them. (lower right) The density of funded amount and recoveries.

Predicting next year's business:

The amount invested to the borrowers increased very rapidly until mid-2015. After then, the growth slows down. The growth can be mathematically modeled by a logarithmic function. This model predicts that the company will require three hundred million dollars of investment in the

next year.

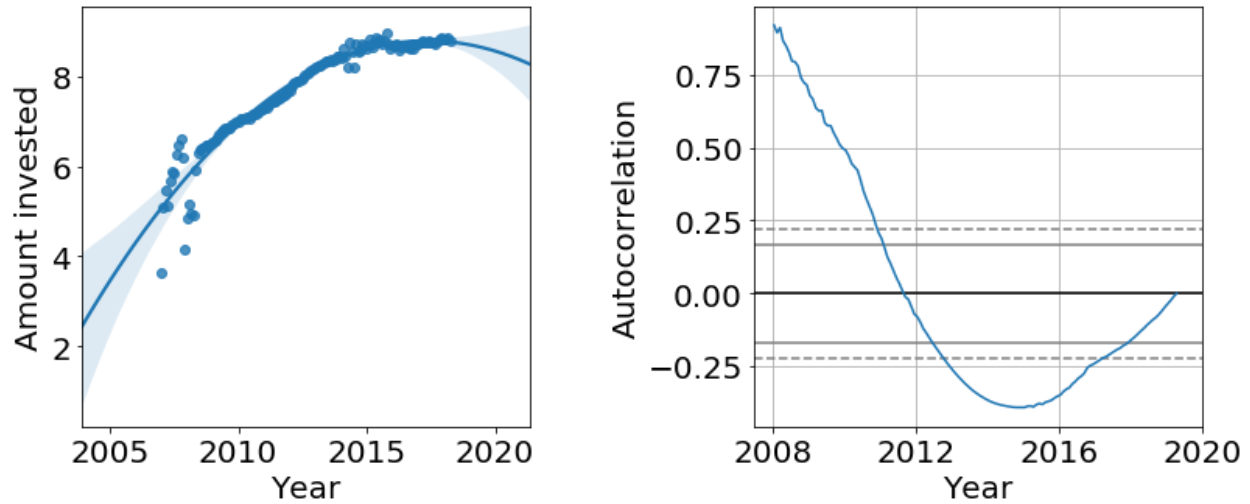


Figure 2: Logarithmic values of the amount invested over time. The solid line is a model that predicts the amount. The shaded line represents the 95% confidence intervals.

Understanding the borrowers:

Most of the borrowers have an annual income of less than a hundred thousand dollars.

Majority of borrowers have a total current credit balance below five times their current annual income. The main reason to borrow the loan is to consolidate their debt and pay their credit card loans. This indicates that most of the customers are already facing financial problems.

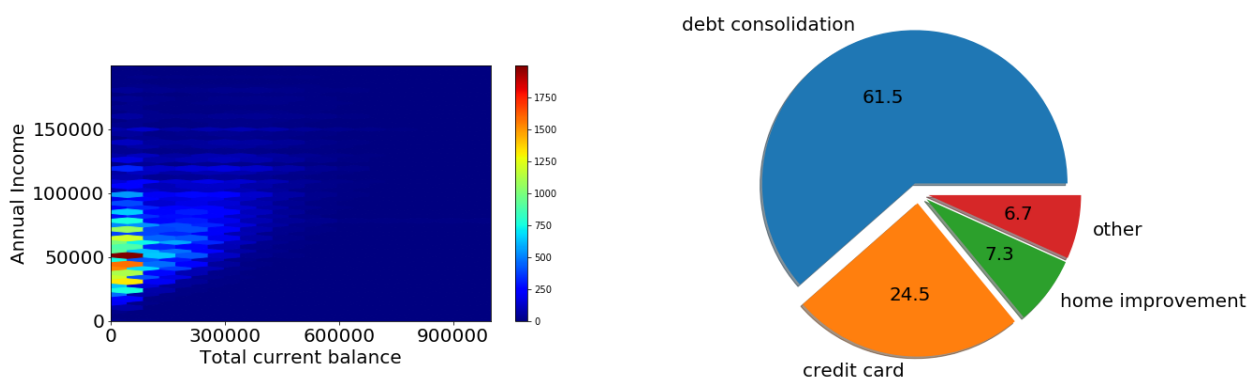


Figure 3: (left) The density of borrowers by annual income and total current balance. (right) The proportion of borrowers with the purpose of the loan.

Distribution of Borrowers:

Loan above \$25,000 is mostly given to the income verified borrowers. There is seemingly no big difference between the loan distribution between paid and charged-off borrowers.

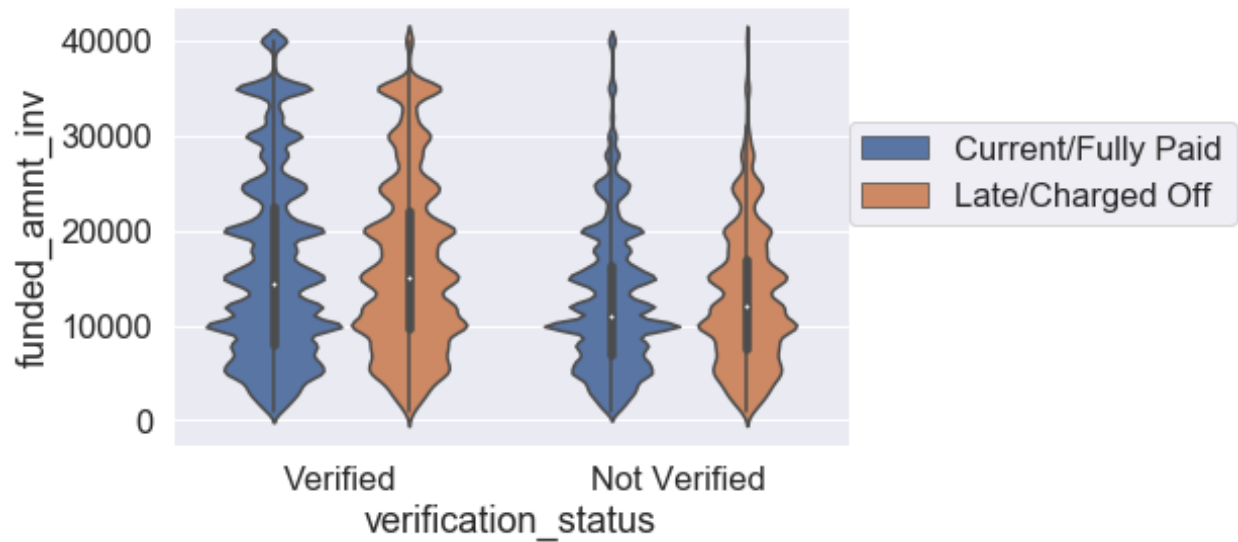


Figure 4: Kernel density plot of loans for verified and unverified borrowers.

Important features:

Understanding how the credit history of borrowers is related to the loan status (paid or charged off) is one of the important questions for the management. We did a quick survey on the relation between loan status and remaining features. It shows that “*account opened in past two years*” is the most important feature among others. Important features affecting the loan status are below.

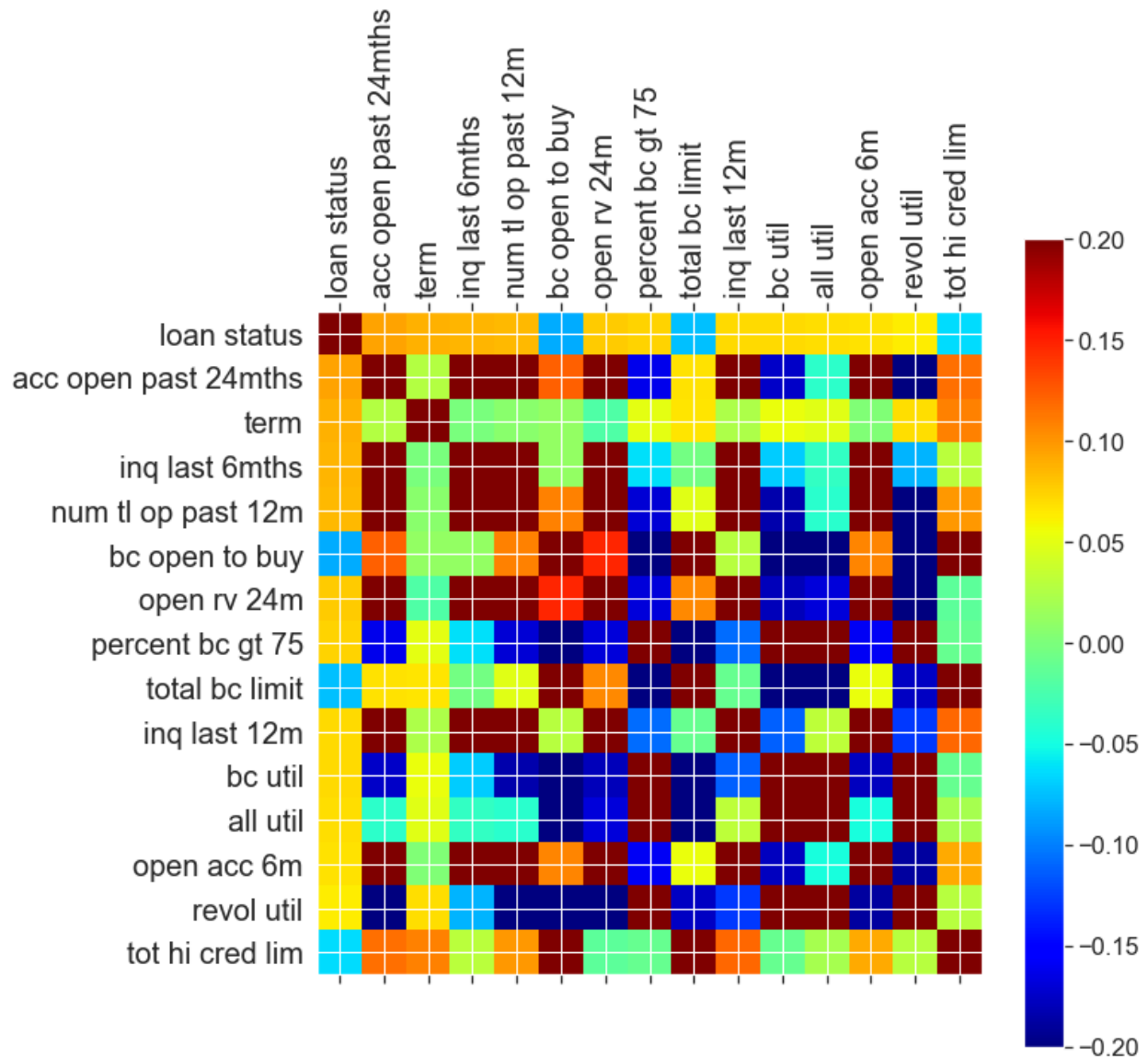


Figure 5: Important categories that impact loan status.

Role of recent history:

The credit history of borrowers tells a lot about the loan status. A borrower with better credit conditions is more likely to pay the loan. For example, borrowers with high open to buy ratio (The ratio of maximum credit limit to the current balance on the account) have a significantly smaller risk. Borrowers with more accounts open in past years or inquiries are less likely to pay the loans in time. The error bars represent 99.99% confidence interval.

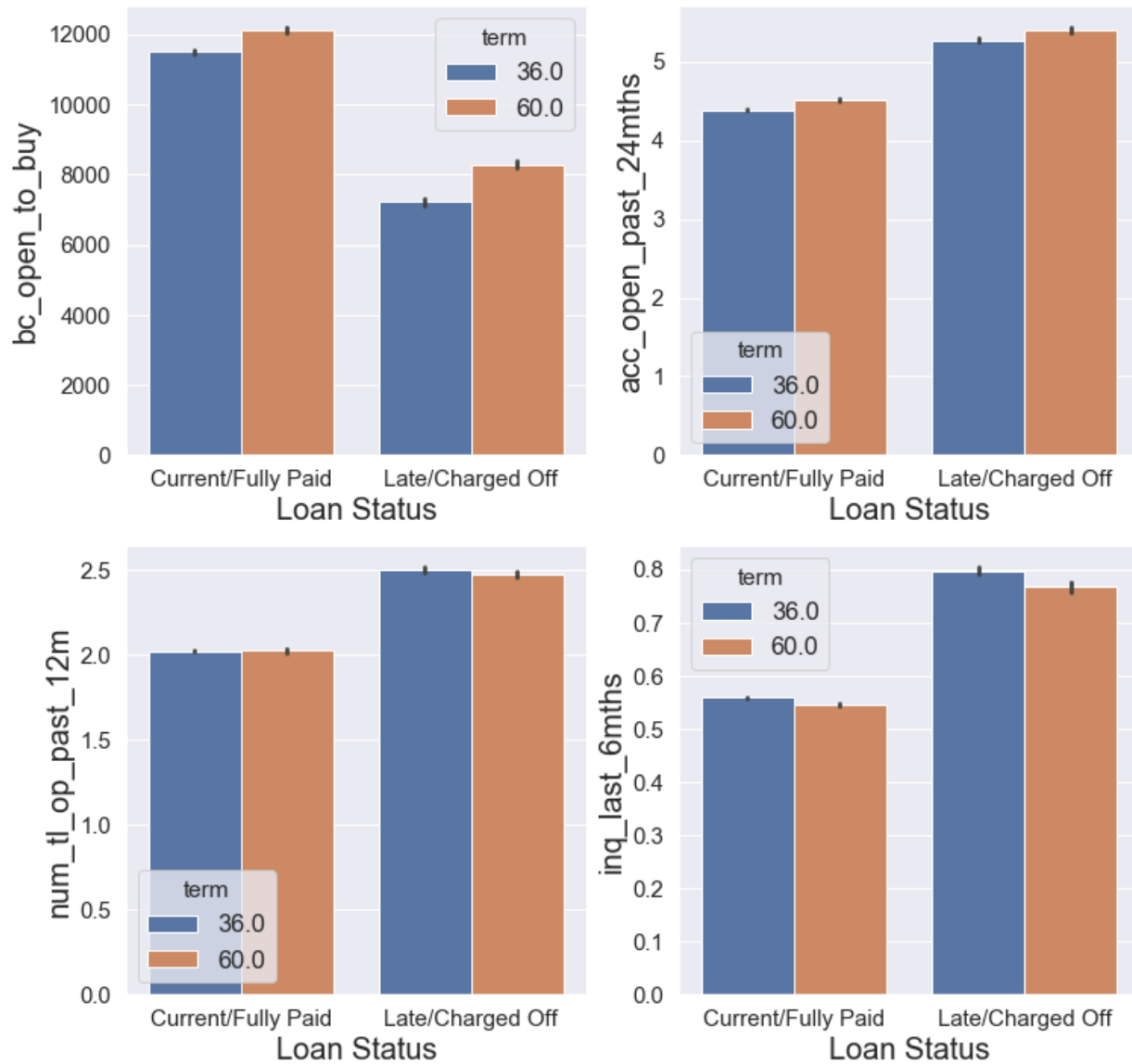


Figure 6: Risk of loans for different categories.

Role of long-term credit history:

The long-term credit history impacts the risk of loans but not as much as the recent credit history. Below are the histograms or Kernel density Estimation plots of some of the features. Borrowers having shorter credit history have only slightly more risk compared to those who have a long credit history. The annual income does not make much difference in the risk factor. This is because the real income as income of many borrowers is not verified. But the DTI (A

ratio of the borrowers total monthly debt to the total debt obligations), which reflects the current financial condition of borrowers makes more contribution to the risk factor. Borrowers with lower DTI are more likely to pay the loans in time.

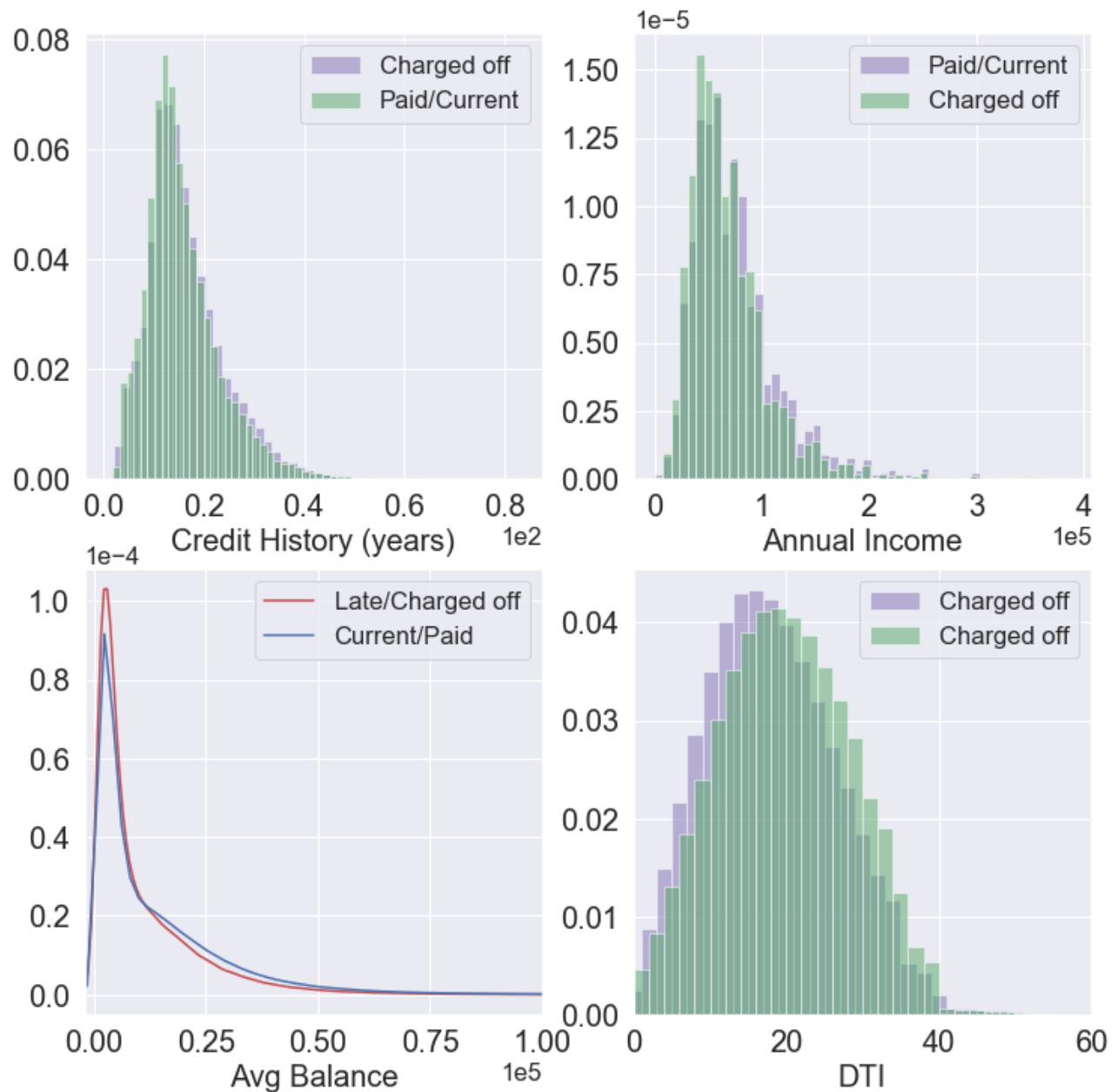


Figure 7: Risk of loans for different categories.

Role of loan terms:

Independent of whether the loan term is three or five years, the risk of default loan increases sharply until a year and falls down over time.

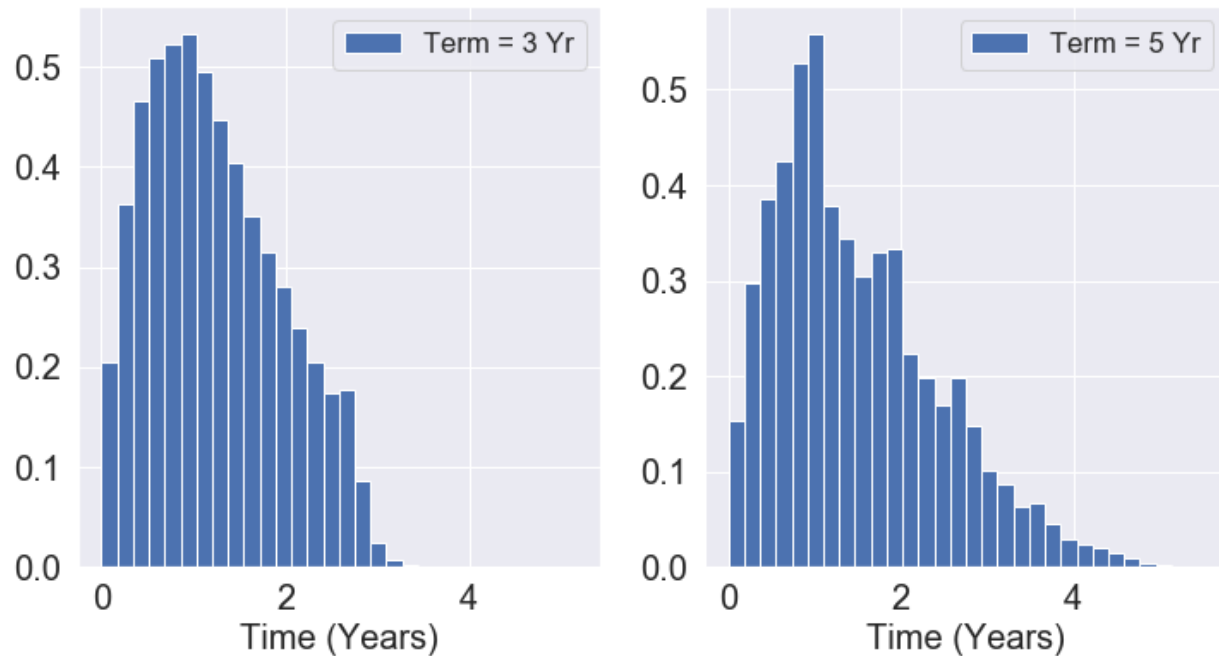


Figure 8: Histogram showing the risk of late/charged off loans.

Quantifying the risk:

The five-year loan term has double the risk compared to the three-year loans. However, the risk does not change much with the loan amount.

The risk also depends on the purpose of the loans. Loans for small business and education have a much higher risk compared to other categories.

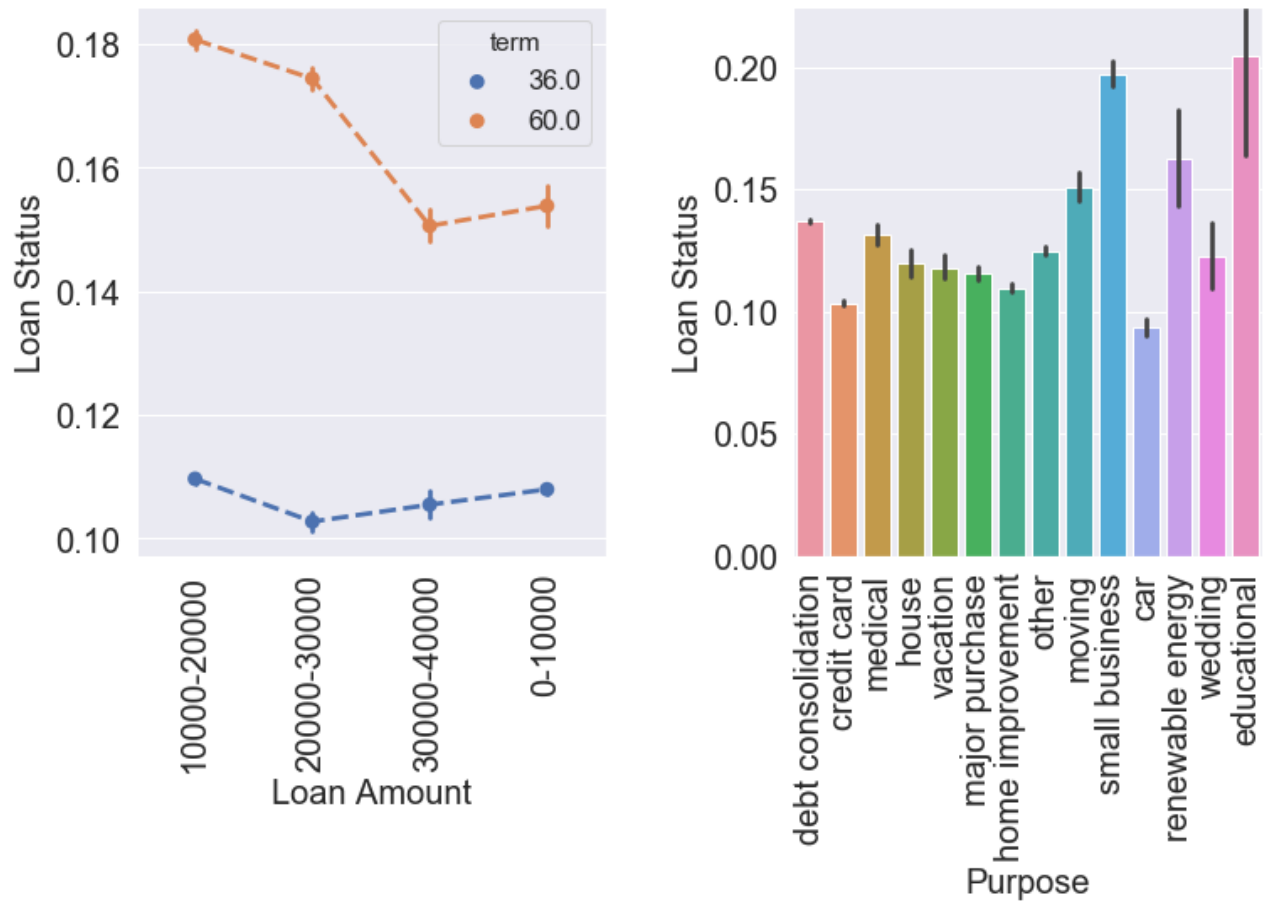


Figure 9: (left) Loan status for five- and three-year terms. (right) Risk of loans for various purposes.

Predicting returns:

One of the company's goal is to estimate the fraction of the loans returned to the company or the investors. Using linear regression, I estimated the amount returned to the company/investors. Figure 10 shows the predicted values of the 'total received principal' versus the real test values.

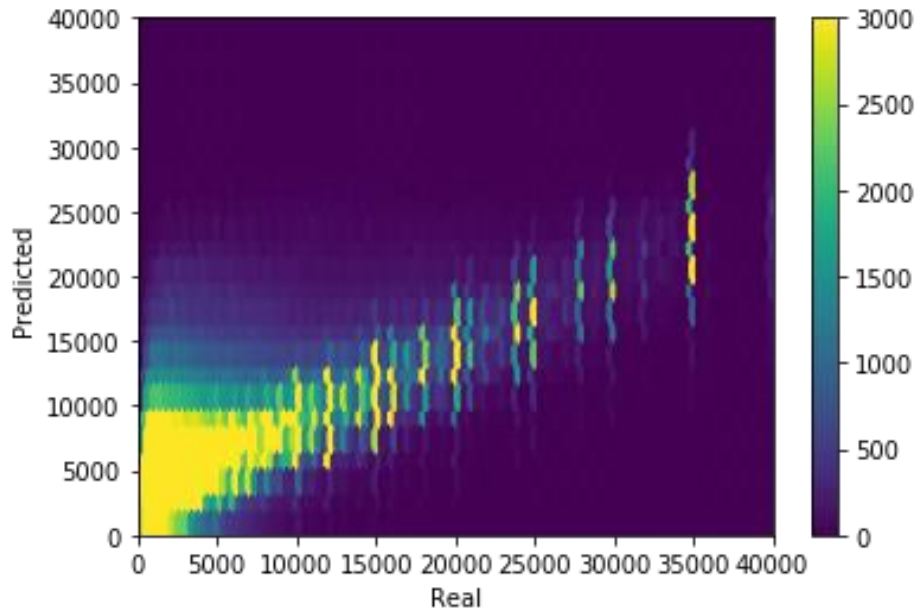


Figure 10: Predicted versus actual values of “total received principal”.

Customers Classification:

To make a better prediction on whether the loans will be returned within the terms, I built up a statistical model to classify borrowers based on their credit history and amount of loans. This model classifies the borrowers with 67 percent of accuracy.

	precision	recall	f1-score	support
paid	0.66	0.68	0.67	134824
not-paid	0.67	0.65	0.66	134843
Avg	0.67	0.67	0.67	269667
AUC: 0.73				

Risk Assessment:

The classification scheme described in the previous paragraph also quantitatively calculates the risk (the probability that the loan will not be returned in time). I compared the risk analysis of our model with the Lending club's original risk calculation. Based on the club's website, it determines the interest of loans based on the credit history of borrowers, i.e. the customers with smaller risk will have lower interest rates and vice versa. Using this information, I estimated the initial risk determined by the Lending club.

Fig. 11 provides a comparison between new and old risk assessment schemes. As expected, customers with higher calculated are less likely to pay the loans and their distribution is represented by the red color histograms as shown in Fig. 11. To estimate the quality of these risk assessment scheme, I defined a new term quality as,

$$Quality = \left(1 - \frac{Area\ of\ overlap}{Total\ area}\right) \times 100$$

Ideally, model with 100 % accuracy would give quality = 100 %. Zero quality on other side mean that the model has no predictive power.

Our new analysis resolves the two distribution significantly (25 %) better than that initially determined by the lending club.

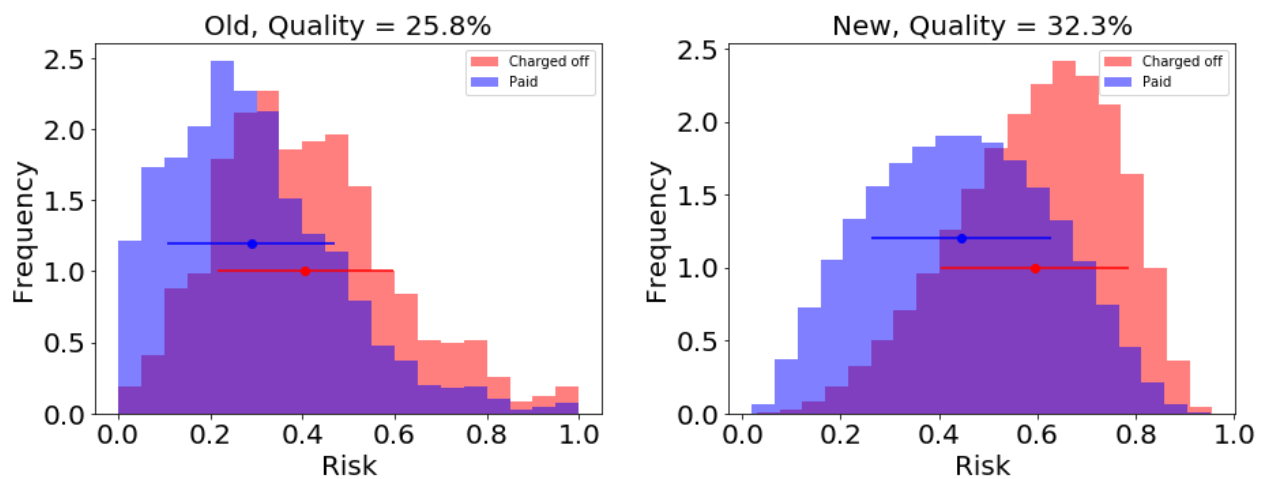


Figure 11: Histogram distribution showing the density of loans that are paid (blue) or charged-off (red).

Recommendations:

This work makes the following suggestions to improve the loan default rates:

- (1) For the lending club, it is beneficial to use our risk assessment scheme for better estimation of the risk of loans. Our model provides a significantly (25 %) better estimation of risk compared to the previously used approach.
- (2) To improve the loan default rate, the company should promote loans with shorter-term compared with five years of loans. As the loan term increases, it is more likely that the borrowers run into financial problems.
- (3) Loans on 'Education' and 'small business' have a significantly higher risk compared to other categories. These loans do not contribute much (less than 5 %) of the company's business. The company should seek more stringent criteria when lending loans on these categories.
- (4) In general, the credit history of borrowers depends on the long-term credit history of customers. However, short term credit history of borrowers is more important than the long-term history. In the future, the lending club should focus on collecting short term credit history (within 2 years) of borrowers.