

README

March 16, 2023

1 Loan Data Exploration

1.1 by Albert Atsu Dellor

This project explores a dataset containing 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others

1.2 Dataset

The data can be found here: `'https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv'`

The main interest of this project is to study the borrowers of prosper bank, especially their loan status along with other features to understand what factors affect a loan status of a borrower. Due to the large set of features, this analysis won't analyse each but will focus on these:

- ListingCategory: This is a categorical variable that represents the loan's category or purpose, e.g. debt consolidation, home improvement, business, etc.
- LoanStatus: the ratio of the borrower's debt to their income
- BorrowerAPR: the annual percentage rate charged to the borrower
- LoanOriginalAmount: the original amount of the loan
- BorrowerRate: The interest rate that the borrower is paying on the loan.
- CreditScoreRangeLower: The lower limit of the range in which the borrower's credit score falls.
- CreditScoreRangeUpper: The upper limit of the range in which the borrower's credit score falls.
- StatedMonthlyIncome: the amount of monthly income stated by the borrower
- DebtToIncomeRatio: the ratio of the borrower's debt to their income
- EmploymentDuration: the length of time the borrower has been employed in their current job.
- Employment: indicates if a loan is current, completed, or in default.
- Occupation: the borrower's occupation
- IsBorrowerHomeowner: indicates if the borrower owns a home
- InquiriesLast6Months: the number of credit inquiries made on the borrower in the last 6 months.
- PublicRecordsLast12Months: the number of public records in the borrower's credit history in the last 12 months.
- DelinquenciesLast7Years: Refers to the number of times the borrower has been delinquent on a payment in the last 7 years

1.2.1 Data wrangling

- Remove columns not directly needed for analysis.
- Rename list category from numeric to corresponding string value and rename column name for context.
- summarise the various statges of Past Due (1-120days) into a single Past Due value for simplicity.

1.3 Summary of Findings

1. Additionally, the credit score distribution of borrowers shows that the majority of borrowers have credit scores ranging from 700 to 850, with most having scores of 700 to 749. A smaller percentage of borrowers have scores ranging from 650 to 699 and an even smaller percentage have scores below 650. This indicates that the majority of borrowers have good credit and are less likely to default on their loans compared to those with lower credit scores. This is supported by the previous observation that a majority of loans are in good standing, with only a small percentage being in bad standing with the bank.
2. Further investigation reveal that, for borrowers in good standing with the bank the top 5 reasons for taking loans (in descending order of frequency) are for: Debt consolidation > undefined > home investment > business > household expensis While that of the bad standing borrowers are for : Not available > debt consolation > business > other > personal loan. They both have **debt consolation** and **business** in common. However the reason of taking loan for **personal loan** is exclusive to those in borrowers in bad standing. Therefore, borrowers who take personal loans are more likely to be in bad standing with the bank by defaulting, charged off or have their loans cancelled. It's also worth noting that the data for reasons for taking loans for borrowers in bad standing is not comprehensive, as a significant portion of their reasons are marked as "not available". This may indicate a higher likelihood of these borrowers not accurately disclosing their reasons for taking loans or a lower level of record-keeping by the bank.
3. It is also possible that borrowers with lower credit scores, higher debt to income ratio, or other risk factors may have to pay higher interest rates, which could explain the frequency peak at 35%. This highlights the importance of borrowers having good credit scores and low debt to income ratio for accessing loans at lower interest rates.
4. However, it is worth noting that the data does show a weak positive correlation between being a homeowner and having a higher credit score, which could suggest that homeowners are more financially stable. Additionally, the data shows a weak positive correlation between being a homeowner and borrowing larger loan amounts, which could suggest that homeowners have more financial resources to tap into when taking out loans.
5. In conclusion, from the analysis of the data, it can be seen that the majority of borrowers from Prosper Bank are in good standing, employed, and make an average salary ranging from \$2500 to \$5500. They also have a relatively low debt to income ratio and the majority of loans given out were for debt consolidation and business purposes. Additionally, most borrowers have only one delinquency in the last 7 years and have had their credit reports checked once within the past 6 months. The bank's lending practices seem to be fair, as there is no discrimination between homeowners and non-homeowners, and the majority of loans given out were for amounts ranging at \$4K, \$10K , \$15K , \$20K , \$25K.
6. This suggests that home ownership may play a role in a borrower's ability to repay their loan, as those who are homeowners tend to have a better loan repayment status compared to those who are not homeowners. However, it is important to note that this may not always be the case, and other factors such as income and credit score may also play a significant role in loan repayment.
7. In conclusion, the analysis of the Prosper loan dataset shows that the majority of borrowers are in good standing with the bank and have taken loans for debt consolidation or business

purposes. The data also shows that home ownership is not a major factor for lending out loans but there are more borrowers who are current or have fully paid off their loans and are home owners than those who are not. Borrowers who borrowed the most amount, on average, are those who are current and fully paid off their loans. Borrowers who had their loans cancelled borrowed the least amount, suggesting that their loans were cancelled and forgiven because the amount borrowed was small and it might have cost more to retrieve those loans.

8. Borrowers with lower BorrowerAPR% tend to be in good standing, while those with higher BorrowerAPR% tend to have defaulted or are past due on their loans. This suggests that borrowers with higher risk of defaulting may be charged higher interest rates. .

Interestingly, the borrowers who have had their loan cancelled have the lowest BorrowerAPR% and yet they couldn't pay their loans. This is not surprising because as seen from previous plot, they have the lowest reported income and were lent the lowest amount amongst the categories.

Additionally, with the exception of Past Due borrowers, borrowers in good standing have a higher credit score than those not in good standing. Therefore credit score and BorrowerAPR% have a negative correlation. Thus, a borrower with high BorrowerAPR% has a weak credit score and vice versa. And increasing credit score positively increases loan original amount given to borrower.

9. Yes, that is correct. Borrowers in good standing with the bank tend to have higher monthly income and borrow more than those who are not in good standing. The highest average monthly income is among those currently paying their loan, those who have completed their loan payment, and those making their final payment. On the other hand, borrowers who had their loans cancelled have the lowest average monthly income. The trend holds across multiple employment fields, except for those who are not employed.

10. Borrowers who are not employed and self-employed tend to have a higher debt to income ratio compared to other employment categories, leading to a higher proportion of defaulted loans among these groups. This highlights the importance of a stable source of income in being able to pay back loans on time.
11. This means that as the original loan amount increases, the borrower APR% tends to decrease. This also suggests that as the original loan amount increases, the credit score range upper increases, and there is a positive relationship between the two. However, this relationship is weak.

The relationship between home ownership and original loan amount is also weak, meaning that it does not play a significant role in determining the original loan amount. Similarly, the relationship between stated monthly income and original loan amount is weak, indicating that monthly income is not a strong factor in determining the original loan amount.

Finally, the relationship between home ownership and credit score, as well as home ownership and employment status duration, is also weak, meaning that these factors do not have a strong impact on each other.

1.4 Key Insights for Presentation

I study loan status of borrowers along with their other features to understand what factors affect a loan status of borrowers. Those features are:

LoanStatus, BorrowerAPR, LoanOriginalAmount, ListingCategory, BorrowerRate, CreditScoreRangeLower, Cred

I start by investigating the univariate distribution of borrower's loanstatus using barplots.

I then proceed to investigate the other features one by one with barplots and histograms in the following order: BorrowerAPR%,IsBorrowerHomeowner,EmploymentStatus,LoanOriginalAmount,DebtToIncomeRatio,Employment

Afterwards, I study these feature together with loan status using bi-variate and multi-variate plots. These are the key findings:

- 1. It turns out that majority of the borrowers have some form of employment. Specifically, 60.3 % of them having employment, with 21.6% being full time employment and 5.5 % being self-employed, With the most frequent reported employment status duration being 0 to 50. Additionally most borrowers make a salary ranging from \$2500 to \$5500 and also having a debt to loan ratio of only 0.2. No wonder most of the borrowers are in good standing because they have some means of paying back the loan. It is also no surpriseing that the frequency distribution of original loan amount given to borrowers show large peaks in frequency at \$4K, \$10K, \$15K, \$20K, \$25K being given to borrowers. However, only few loans above 26K were offered. Interestingly, the majority of borrowers have only one delinquencies in the last 7 years and have had their credit reports check only once within the past 6 month, meaning that they are not hopping from bank to bank borrowing.
 - 2. Borrowers who borrowed the most of average of \$10360 and \$8346 respectively are those consistently rapaying (**curent**) and those making their last payment(**finalpaymentprogress**). Also borrowers who have fully paid off their loans (**completed**) borrowed an average of \$6189. Borrowers who had their loans cancelled borrowed an average of \$1700 which the lowest of all amount borrowed. It can be suggested that their loans were cancelled and forgiven because the amount borrowed was small and it might cost more to retrieve those loans from borrowers. So in a nutshell, banks biggest borrowers are also its most compliant and are in good standing.
- 3. It's clear that borrowers in good standing with the bank tend to have lower borrower APR and receive higher loan amounts compared to those who are not in good standing. The inverse relationship between credit score range lower and borrower APR also highlights the importance of having a good credit score in determining loan terms. Borrowers with low credit scores are more likely to receive a higher borrower APR and receive lower loan amounts.
- 4. It is evident that borrowers who have a good repayment history with the bank have a higher income compared to those who have defaulted or had their loans cancelled across various employment categories. However, the exception to this is borrowers who are not employed, who tend to have the highest debt-to-income ratio and a higher likelihood of defaulting on their loans. Self-employed borrowers follow closely behind, with a similarly high debt-to-income ratio and a higher proportion of loan defaults compared to those who are currently making timely payments.
- 5. This statement accurately describes the relationship between debt to income ratio and loan default rates across different employment fields. It is important to note that while non-employed, self-employed and retired borrowers tend to have a higher debt to income ratio and higher default rates, there could be other factors that contribute to loan repayment behavior, such as credit history and income stability.
- 6. It is observed that there is a weak inverse relationship between borrowerAPR% and original loan amount, with borrowers having a lower borrowerAPR% borrowing higher amounts. Also, there is

a weak positive correlation between credit score range upper and original loan amount, with those having higher credit scores borrowing larger amounts. There is a weak positive correlation between being a homeowner and the loan original amount, with homeowners borrowing more than non-homeowners. Additionally, there is a weak positive correlation between stated monthly income and loan original amount, with those having a higher monthly income borrowing larger amounts. Finally, there is a weak positive correlation between being a homeowner and credit score, as well as a weak positive correlation between being a homeowner and employment status duration.

1.4.1 Conclusion

Based on the insights and findings from the loan data analysis, it is clear that the employment status, credit score, income and homeowner status play a significant role in determining the loan amount, borrower APR, loan standing and likelihood of default. Given this information, it would be wise for lenders to consider these factors when evaluating loan applications. Additionally, borrowers can take steps to improve their credit score, increase their income and consider homeownership as a way to increase their chances of getting approved for a loan with better terms and conditions. By being proactive in improving these factors, both borrowers and lenders can benefit from better loan outcomes.