

Prosper Loan Visualization

This project is on a data set from Prosper, which is America's first marketplace lending platform, with over \$7 billion in funded loans. This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, borrower employment status, borrower credit history, and the latest payment information, etc. The main purpose of this project is to summarize the characteristics of variables that can affect the loan status and to get some ideas about the relationships among multiple variables using summary statistics and data visualizations.

Structure of your dataset?

There are 113937 rows in the dataset with 81 features. Most variables are numeric in nature(61), but the variable ProsperRating is ordered factor variable with the following levels (best) —> (worst) AA-A-B-C-D-E-HR

Main feature(s) of interest in the dataset?

I'm most interested in figuring out what:

- What affects the borrower's APR or interest rate?
- Are there differences between loans depending on how large the original loan amount was?

My assumptions based on the dataset of interest:

Granting a borrower's loan and assigning an annual percentage rate(APR), it has to be based on type of loan, borrower credit score and risk profile. The better the score, the lower the APR - and the less the borrower pay over time. If borrower score is lower, there's possibility of borrower's defaulting on payment, so therefore higher interest rate since considered more of a risk.

To confirm this assumption, exploring of data has to be done to confirm whether my assumption are right or wrong.

Feature(s) of interest?

- **BorrowerAPR:** The Borrower's Annual Percentage Rate (APR) for the loan
- **BorrowerRate:** The Borrower's interest rate for this loan.
- **ProsperRating:** The Prosper Rating assigned at the time the listing was created between AA - HR. It is ordered factor variable with the following levels (best) —> (worst) AA-A-B-C-D-E-HR
- **Prosper score:** A custom risk score built using historical Prosper data. The score ranges from 1-10, with 10 being the best, or lowest risk score.
- **Listing Category:** The category of the listing that the borrower selected when posting their listing: 0 - Not Available, 1 - Debt Consolidation, 2 - Home Improvement, 3 - Business, 4 - Personal Loan, 5 - Student Use, 6 - Auto, 7- Other, 8 - Baby&Adoption, 9 - Boat, 10 - Cosmetic Procedure, 11 - Engagement Ring, 12 - Green Loans, 13 - Household Expenses, 14 - Large Purchases, 15 - Medical/Dental, 16 - Motorcycle, 17 - RV, 18 - Taxes, 19 - Vacation, 20 - Wedding Loans
- **Employment status:** The employment status of the borrower at the time they posted the listing.
- **Loan Original Amount:** The origination amount of the loan.
- **credit score range lower:** The lower value representing the range of the borrower's credit score as provided by a consumer credit rating agency.

- **credit score range upper:** The upper value representing the range of the borrower's credit score as provided by a consumer credit rating agency.
- **debt-to-income-ratio:** The debt to income ratio of the borrower at the time the credit profile was pulled. This value is Null if the debt to income ratio is not available. This value is capped at 10.01 (any debt to income ratio larger than 1000% will be returned as 1001%)
- **Loan status:** The current status of the loan: Cancelled, Chargedoff, Completed, Current, Defaulted, FinalPaymentInProgress, PastDue. The PastDue status will be accompanied by a delinquency bucket.

Insights

- **Loan status:** The plot shows that 113,937 individual has a loan with Prosper, almost 54,000 borrowere are still owing while significant number has completed the loan. The number of past due and cancelled are quite low
- Borrower Annual Percentage Rate(APR) distribution shows multimodal distribution. There's a small peak at 0.1, another peak at 0.21, with another peak at 0.3. In addition, there is one exceptionally high peak at o.35. It's distributed around 0.05 to 0.41
- **Prosper Rating** is expressed as a letter grade that indicates the expected level of risk associated with a loan listing. The rating ranges from (Highest level of risk)HR----AA(least level of risk). The above plot shows that most loan takers has ratings of B and C.
- **Prosper Score** Most borrowers has score ranging from 4 to 8 Score, with 4, 6 and 8 having the highest frequency
- **Debt-To-Income-Ratio(DTI)** ratio exhibits slight right skewed, with peak at 0.26 and mean of 0.22. There are more borrowers with $DTI < 0.5$. which indidcate that majority of the borrower qualify for loan. To qualify for Prosper loan, borrowers needs $DTI \leq 50\%$
- There's is a strongly negative correlation between prosperScore and [BorrowerAPR & BorrowerRate] This indicate that increase in one will lead to decrease in the other. i.e if Borrower's Prosperscore is high, APR[Annual % rate and Interest rate] will be low
- Negative correlation between the creditScore and [BorrowerAPR & BorrowerRate]. This means the better the creditscore, the lower the APR and the lesser the borrower will pay overtime. This can be use to determine how reliable a borrower will be in paying the loan
- Loan Original Amount has a weak correlation with BorrowerAPR and BorrowerRate. The amount of loan borrowed also depend on BorrwowerAPR, as it can be seen that the higher the amount, the lower the APR
- Borrower's with full time job has slightly lower APR while borrowers that are not employed rarely qualify for best interest rate and APR
- Employment status and the purpose of loan clearly influence the amount of loan been given.e.g Borrowers with part time job, but with very crucial loan purpose like medical/dental tends to be granted higher loan amount
- One interesting thing to note is that there are a very large number of Debt Consolidation in the data, almost twice as many as the second-most borrower loan purpose, Not Available(they didn't provide the purpose of the loan.