

Visualization-analysis of Loan dataset

This analysis is a part of Advanced data-analysis nanodegree provided by Udacity and sponsored by FWD.

Dataset

This project is on a data set from Prosper. 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. The dataset can be found in the repository [here]

<https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv>.

The main purpose of this project is to summarize the characteristics of variables that can affect the loan status and borrower APR and to get some ideas about the relationships among multiple variables using summary statistics and data visualizations.

Summary of Findings

- The length of the loan seems not to have an impact on the outcome of the loan: no matter what status a loan has, the most common length is 36 month and least common is 60 months
- The Prosper Score seem to affect the outcome of the loan: so, have the most borrower an Prosper Score of 8 when a loan is completed, while most borrower with defaulted and charged-off loans have and Prosper Score of 6. Furthermore, the most common Prosper Score for borrowers with loans that have past due payments is 4
- The figure says that most listings fall under 50000 dollars, followed by 50000-74999 dollars and so on. This also concludes that the more the income rate the less the listings count.
- Most of the borrower have a low Prosper Score of 4 (custom risk score) while most of the customer have a score between 5-8.
- it's clearly appeared that the number of borrowers of non-homeowners higher than the number of borrowers of homeowners
- This graph shows that the loans with higher Borrower APR and Lender yield had a lower prosper score.

Keys Insights for presentations

For the presentations, after introducing dataset, I focused on answering question which factors effect on Loan Status in different ways by using univariate and bivariate analysis. I've made sure to use different color to make it clearly, with making legend and summary after all figures.