Loan Listing Data Exploration

Dataset

The dataset consists of information regarding 113,937 loan listings including loan status, borrower APR, stated monthly income, loan original amount, prosper rating, prosper score, and other loan listing attributes. However, after data cleaning, approximately 85,000 loan listings were left and used for the analysis.

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

The features dictionary can be found here: https://docs.google.com/spreadsheets/d/1gDyi_L4UvIrLTEC6Wri5nbaMmkGmLQBk-Yx3z0XDEtI/edit#gid=0

Summary of findings

In my exploration, I found that there is an inversely proportional relationship between both prosper rating and prosper scores with borrower's APR. As prosper rating and scores increase, the borrower's APR decreases, this shows that borrower's APR is affected by prosper variables. Increased Borrower's APR was also associated with failed loan outcome status as loan statuses with late payment or no payment at all had high borrower's APR. Loan outcome status is therefore affected by borrower APR. Failed loan outcome status is seen to be associated with high loan original amount and low stated monthly income while loan success is associated with low original amount and high monthly income. Loan Status is therefore affected by both the loan original amount and stated monthly income.

Outside of the main variables, I verified that a large portion of the applicants had a verified income and that employed applicants were given a higher loan original amount.

Key Insights for presentation

For the presentation, I focus on just the effect of the loan original amount, stated monthly income, prosper rating and prosper score on loan status and borrower APR of the loan listings. I started by introducing the Stated monthly income, followed by the pattern in loan original amount distribution and then the distribution of borrower APR. I went on to introduce each of the categorical variables in focus. I started with the borrower APR across loan status using boxplot. Here, I am trying to figure out maybe the higher ranges of borrower Apr is associated with a negative loan outcome. The prosper variable (prosper rating and prosper score) were then plotted against borrower APR using a box plot too. Both of them have a similar interaction with borrower APR. The better the rating or score, the lower the borrower APR.
The other two numeric variables, stated monthly income and loan original amount were covered afterwards across loan status. To expand my investigation on loan status and these two numeric variables, I went further to make use of point plots and included prosper rating, generating a multivariate plot.