

Prosper Credit Loan Evaluation

Paras Patel

6/15/19

Table of Contents

- Project Overview
- Data Overview
- Exploratory Data Analysis
- Results and Discussion

Project Overview

The purpose of this project is to explore and visualize a dataset from an online loan company, Prosper Loan. Prosper Loan is a peer to peer marketplace that provides unsecured personal loans to interested borrowers and lenders looking to invest. Creditors can invest anywhere from \$2,000 - \$40,000. The goal of this analysis is to model and visualize key trends and relationships to help loaners mitigate risk based on features of the borrower.

Data Overview

The data source is provided by Udacity. The Prosper Loan dataset was last updated on March 11, 2014. Included with the dataset is a variable dictionary explaining all the features tracked by Prosper Loan. The dataset is comprised of approximately 114,000 entries and 81 variables. For this analysis, we will focus on the following features:

- Listing Creation Date
- Borrower State
- Credit Grade
- Borrower APR
- Debt To Income Ratio
- Income Range
- Prosper Rating (Alpha)
- Estimated Loss
- Estimated Effective Yield
- Occupation
- Employment Status

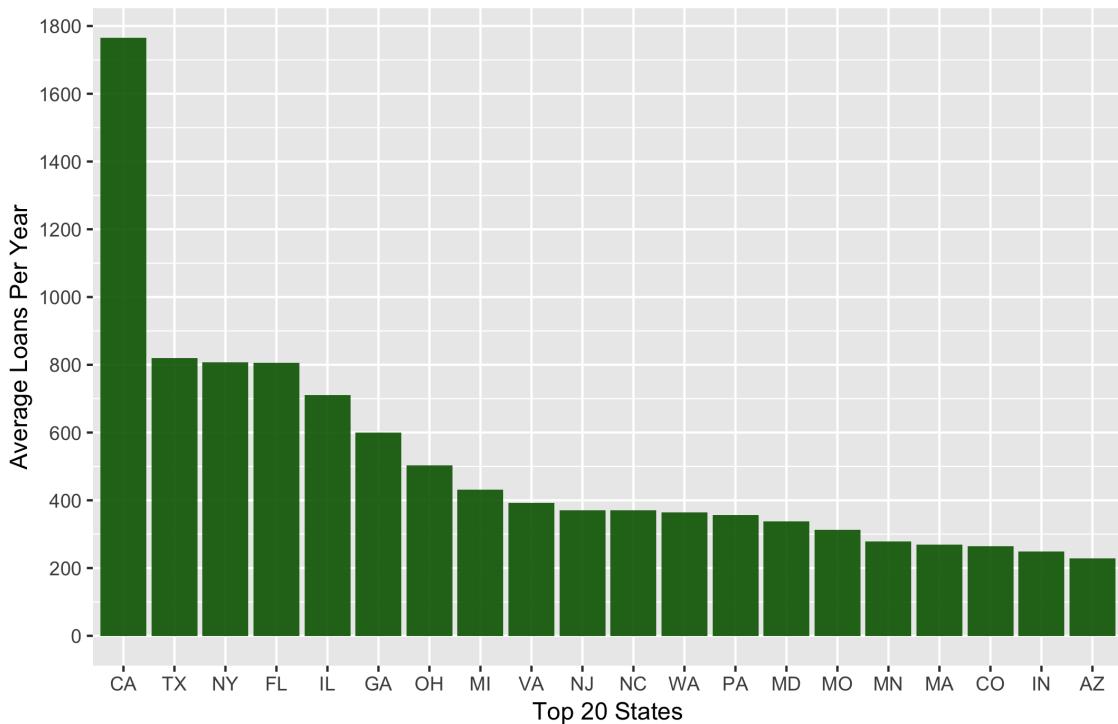
EDA (Analysis)

First, we'll examine the number of loans created by year and country. Prosper Loan experiences a huge spike in loan origination in 2013, most likely attributed to a settlement in a class action lawsuit and 2 major capital investments summing over \$45 million. The next plot shows the top 20 states by average loan origination. California is Prosper Loans leading market as of 2014, not surprising as Prosper is based out of San Francisco. Texas, New York, Florida, and Illinois are the next largest markets, respectively.

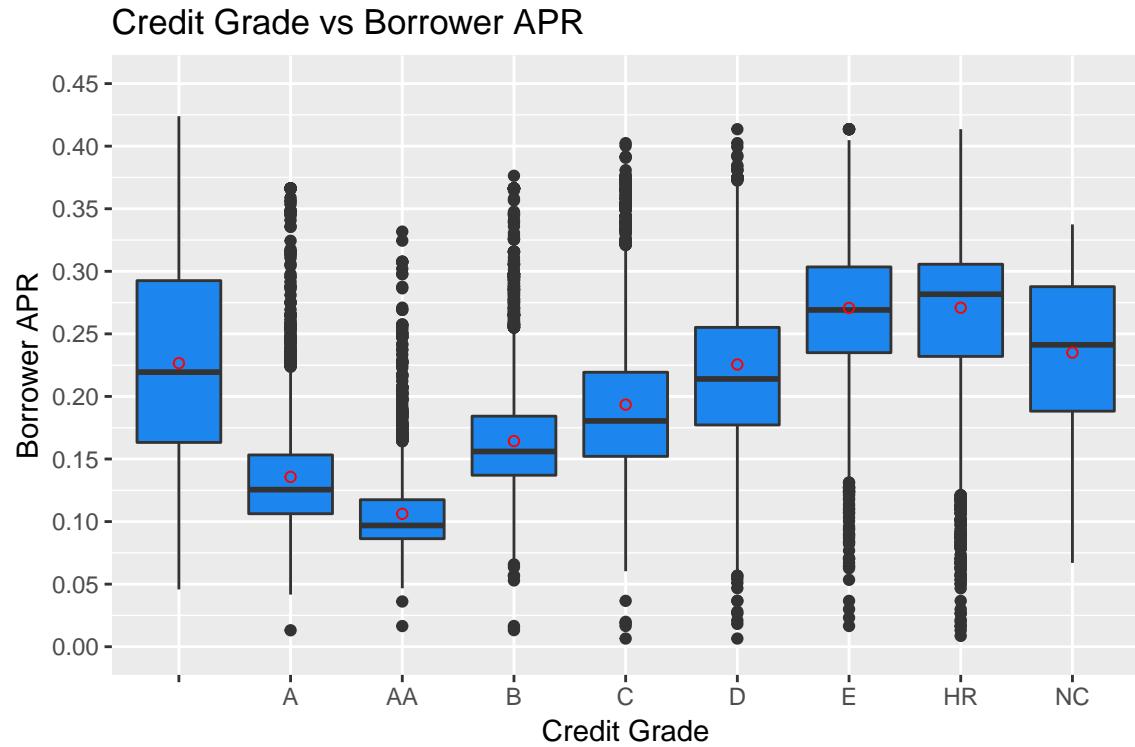
Listing Creation Distribution



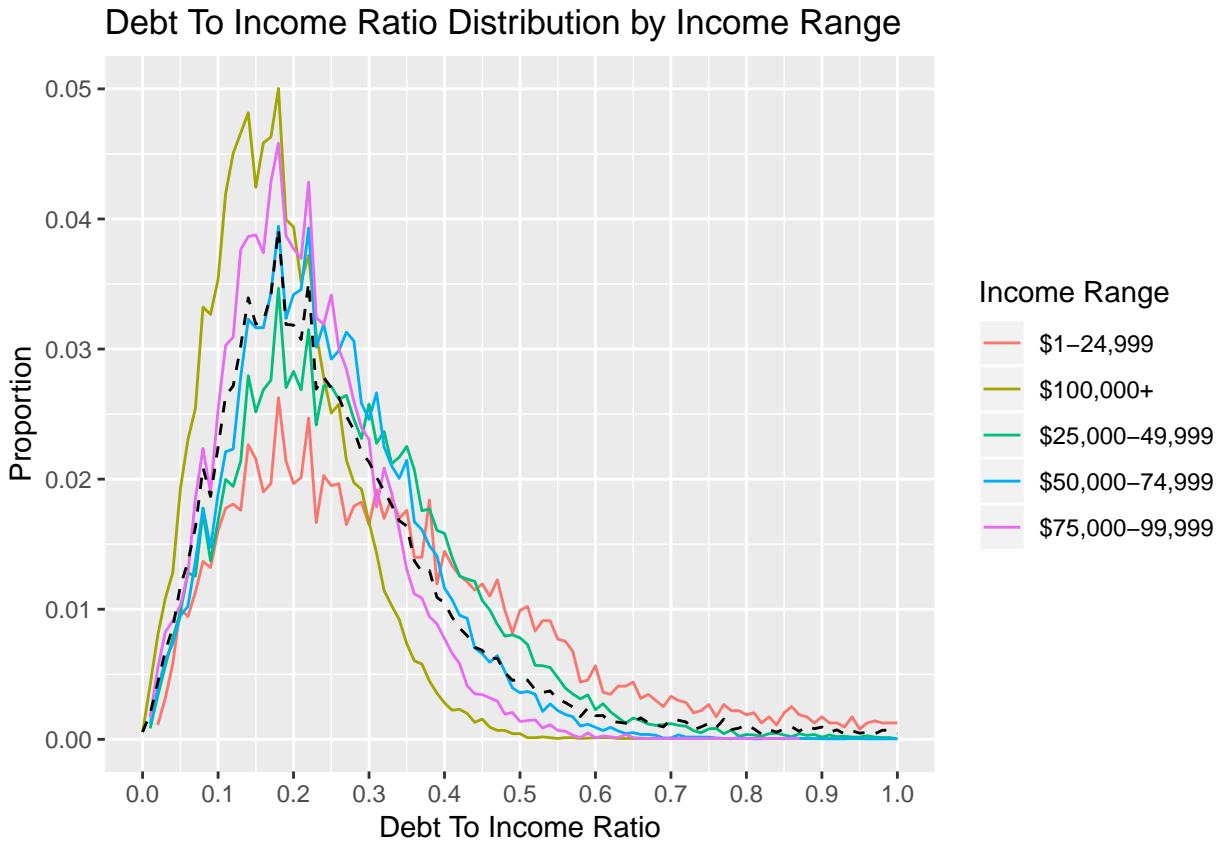
States With Most Loans Borrowed



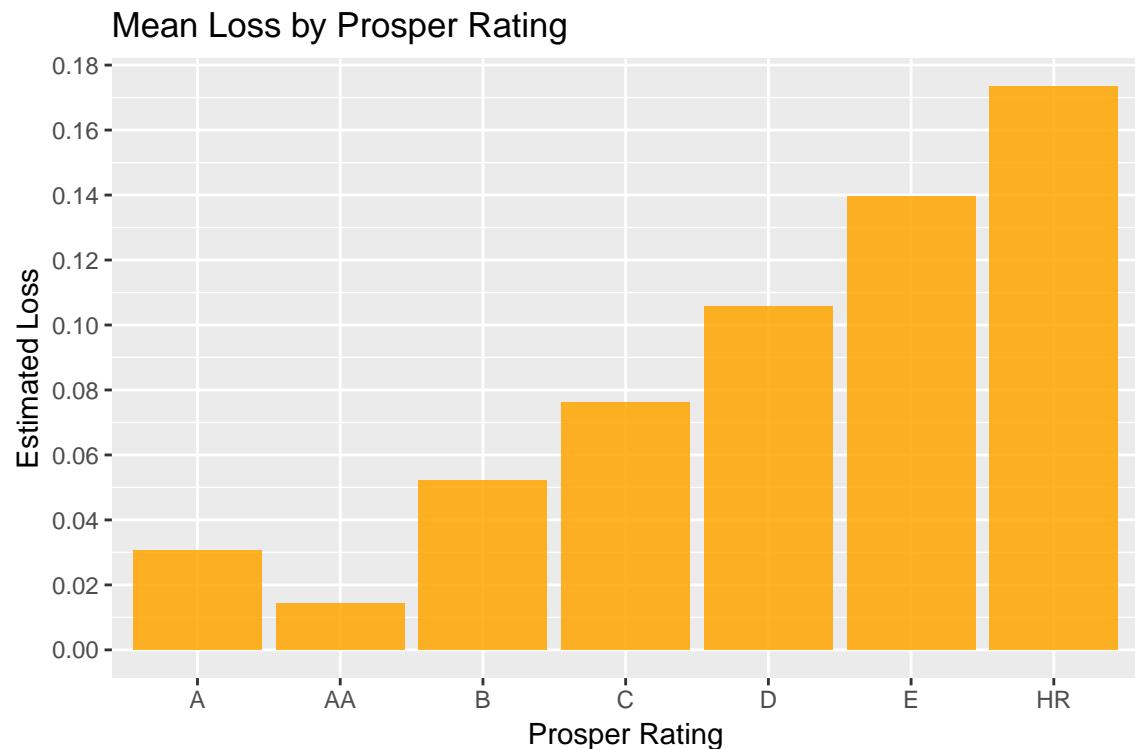
Below, a bivariate relationship between Credit Grade and Borrower APR is illustrated. As expected, mean Borrower APR (Annual Percentage Rate) decreases with a better Credit Grade. The median APR for borrowers with an AA Credit Grade is 10%. On the other end, High Risk (HR) borrowers are hit with a median APR of 27% on their loans.



Here, we evaluate the Debt to Income Ratio's by Income Range class. Approximately 4% of all borrowers have a Debt To Income Ratio of 0.18 with a gradual decline at 0.22. The relationship is expected. Borrowers with higher incomes tend to have lower Debt to Income Ratios than borrowers in the lower income class. Little variation from the mean line and the \$50,000-\$75,000 income range group is observed, indicating sample size bias towards this group.



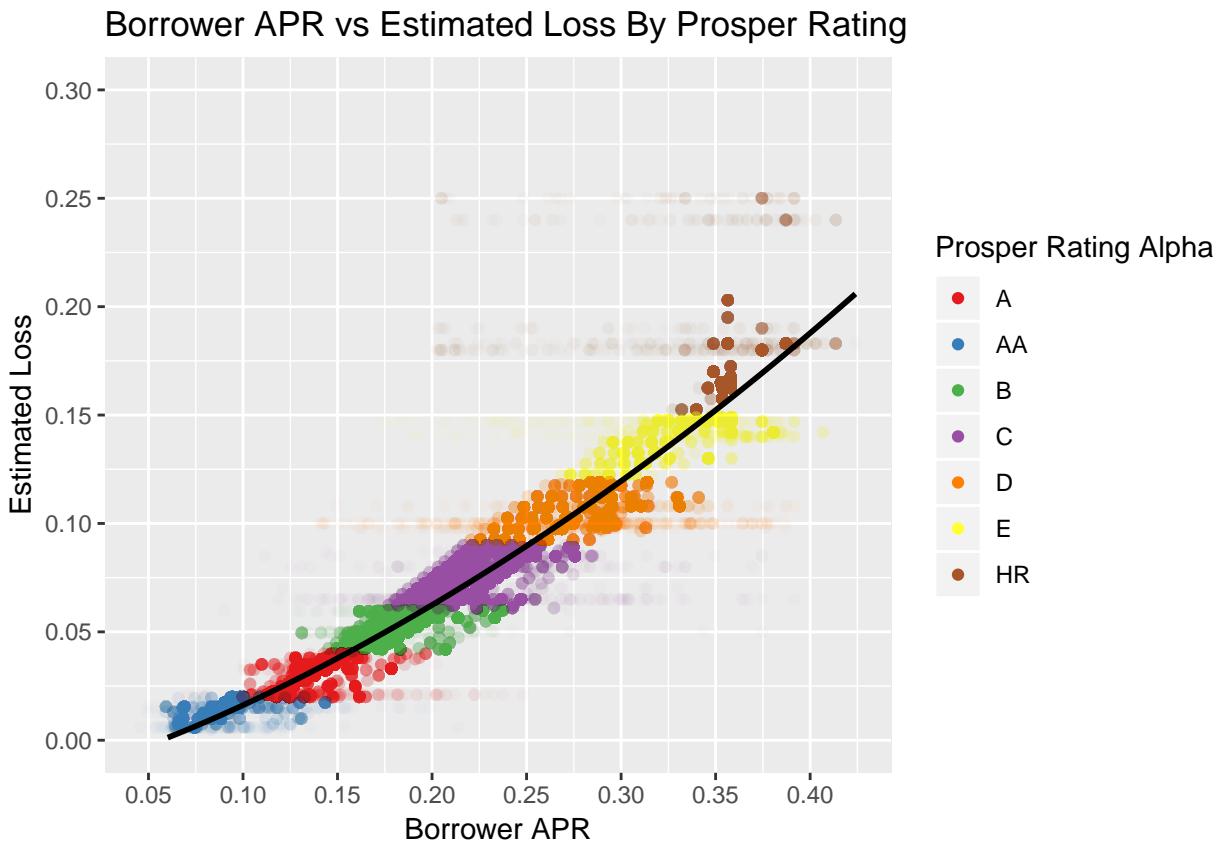
It's important for creditors to understand the potential risk's for a loan. The Estimated Loss is defined as the estimated principle loss on charge off's (an amount of debt unlikely to be collected due to delinquency). Prosper Loan provides a rating for loans with unspecified Credit Grades. Average loss' are skewed heavily towards High Risk borrowers (Mean Estimated Loss of >17%).



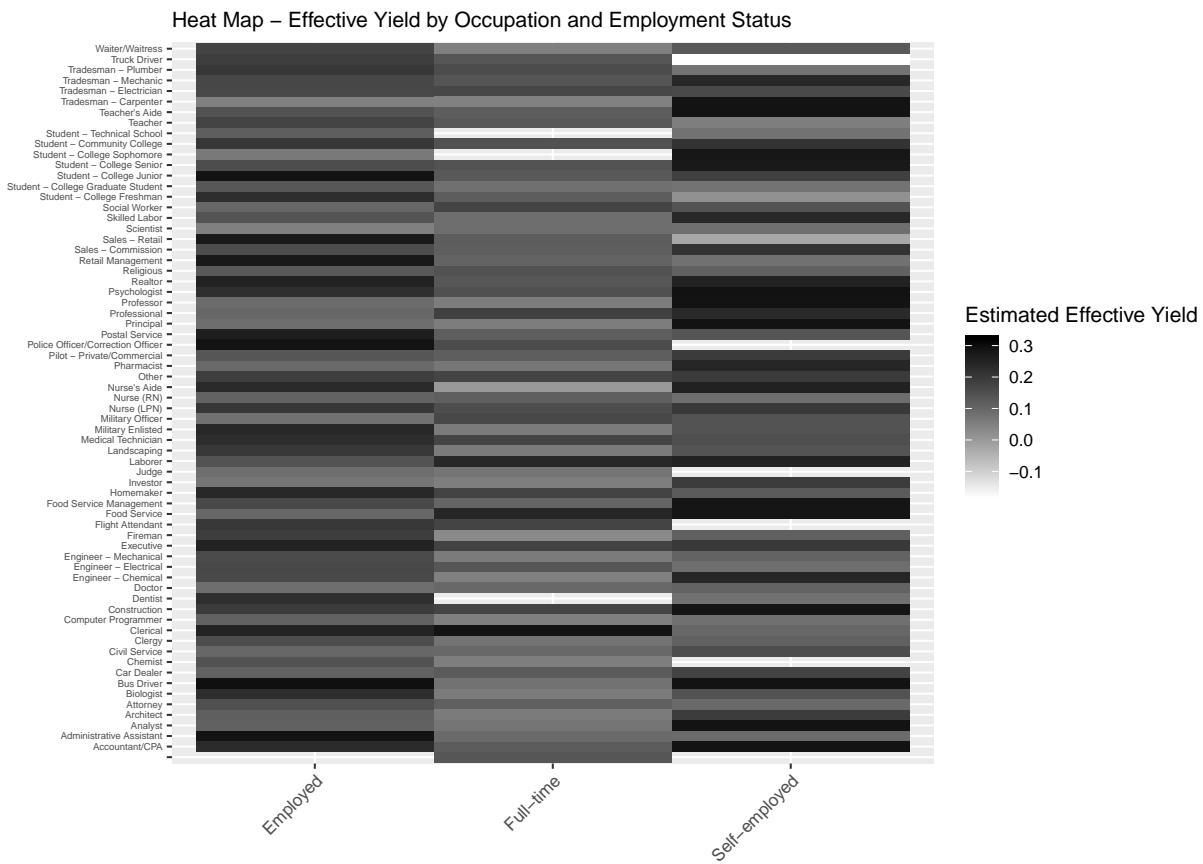
Creating a linear model with Prosper Rating and Borrower APR as responses and Estimated Loss as the predictor, an R-squared of 0.95 and p-value of ~0 are computed, with a residual error of 0.01 (Note the model relationship). R-squared and p-value suggest a strong correlation, however, it's important to note the residual standard error is significant (~10% of the mean/median). The model is still able to provide a fair range for Estimated Loss given these two predictors.

```
##
## Call:
## lm(formula = y$EstimatedLoss ~ exp((y$BorrowerAPR)^1.25) + y$ProsperRating..Alpha.)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.046750 -0.005385 -0.000224  0.004295  0.213759
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           -0.1693392  0.0017225 -98.31  <2e-16 ***
## exp((y$BorrowerAPR)^1.25) 0.1837457  0.0015802 116.28  <2e-16 ***
## y$ProsperRating..Alpha.AA -0.0094136  0.0001735 -54.25  <2e-16 ***
## y$ProsperRating..Alpha.B   0.0142868  0.0001333 107.20  <2e-16 ***
## y$ProsperRating..Alpha.C   0.0306383  0.0001703 179.89  <2e-16 ***
## y$ProsperRating..Alpha.D   0.0496040  0.0002495 198.81  <2e-16 ***
## y$ProsperRating..Alpha.E   0.0726776  0.0003382 214.89  <2e-16 ***
## y$ProsperRating..Alpha.HR  0.1006817  0.0003901 258.07  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0102 on 84845 degrees of freedom
## Multiple R-squared:  0.9524, Adjusted R-squared:  0.9524 
## F-statistic: 2.425e+05 on 7 and 84845 DF,  p-value: < 2.2e-16
```

The linear model between these 3 variables are plotted out below. The relationship between Borrower APR and Estimated Loss can be better expressed as exponential than linear.



Estimated Effective Yield is equal to the borrower interest rate minus (the servicing fee rate and estimated uncollected interest on charge-offs) plus estimated collected late fees. A heat map of Employment Status, Occupation, and Estimated Effective Yield indicates no clear correlation. Therefore, a creditor can use this heatmap to generalize potential profitability from borrowers based on occupation and employment status.



Results and Discussion

In this analysis, we analyzed a few critical metrics from Prosper Loan's trade data for the purpose of helping lenders understand borrower demographic and features to mitigate risk and increase profitability.

Prosper Loan's largest market is in its home state of California. Borrowers with better credit grades typically receive lower APR on their loans, however, there are cases where they receive APR's as high as 30%, which is what we would expect for High Risk borrowers.

We also analyzed the bivariate relationship between Debt To Income Ratio and Income Range. Typical Debt To Income Ratio for all income range classes fall at roughly 0.18 with borrowers with \$100k+ income having the smallest range.

Another valuable relationship for lenders to be aware of is the Estimated Loss by Prosper Rating. Prosper Loan uses a proprietary risk rating system. Lenders can take advantage of the data here to determine estimated loss's (principle loss on charge-off) for their loans based on the borrowers rating.

Finally, a heat map was created correlating Estimated Effective Yield for the lenders based on the Borrowers Occupation and Employment Status. Even though this is a generalization, lenders can use this as a potential guideline to evaluate certain investing opportunities before deeper investigation.