

A story on prosper loan business

Bo Fan

1st story: https://public.tableau.com/profile/bo.fan#!/vizhome/prosper_bf_v1/Story1

2nd story: https://public.tableau.com/profile/bo.fan#!/vizhome/prosper_bf_v2/Story1

Summary

In the prosper loan data set, there are 81 variables. To reduce the number of features, I mainly consider about the variables from the borrowers, lenders, and the prosper company.

In the prosper business story, I explore the trends of the total (average) amount of loan in terms of months, states, loan status, and occupations. I am also interested in the relationships among borrower APR, prosper score, and credit score limit. The average yield also helps the lenders to select loan types to invest.

In figure 1, I find that the top 5 occupations with the largest average loan original amount are Judge, Pharmacist, Doctor, Principal and executive, and the last 5 occupations are from students. It may be because student loans are usually used to pay tuition and rent, which cost less than the other listing categories. The loan status of all the borrowers are mostly 'current', and 'completed' is the 2nd popular one. In the group with low salaries from 1 to 24,999, the number of records of the borrowers who have completed the loan is higher than in the 'current' status, because they may want to keep good records.

In figure 2, I draw the original total loan amount in different states. The highest loan amount is from CA, due to the high overall cost of living. From 2009 to 2014, the average prosper score generally goes down. The original loan amount increases since 2009, while, in 2013 Q1, it drops down a little bit and bounces up again in Q2.

In figure 3, the median borrower APR is highly correlated to the borrower rate. When credit score is less than 400, the median borrower APR goes higher. When the credit score is over 659, the median borrower rate starts to drop down. The prosper score, however, always goes up as the credit score increases.

In figure 4, I notice that the homeowners are likely to receive a higher average loan, and CA also has the largest number of homeowners.

In figure 5, I show the estimated return, estimated loss, estimated effective yield and the lender yield vs different loan types. It seems that higher return means higher loss, and type E and HR have the highest estimated yields.

Design

According to the idea of exploration, analysis, and feedbacks, I design the worksheets as follows:

1 I sort the bar plots of average loan original amount vs occupation. Users can easily find the loan amount from different occupations. I also add a new plot about the number of records vs loan status to uncover the loan status from the borrowers.

2 In figure 2, I add filters on average prosper score and loan original amount to make it more interactive.

3 In figure 3, I exploit the relationship between borrower APR and borrower rate, and analyze the relationship between average prosper score and credit score.

4 In figure 4, I use line plots instead of bar plots to remove ambiguity, and apply Geo map to show the number of homeowners in each state.

5 In figure 5, I add the average lender yield plot to compare with the estimated effective yield.

Feedback

After completing my first Tableau story, I shared it with two graduate students (Greg and Lee), and listed the main feedbacks as follows:

1 In figure 1, the plots of average loan original amount (or the term) vs occupation are not sorted.

2 The lender yield can also be added to figure 4 to make comparisons.

3 Figure 5 is confusing, it does not visually show homeowners receive higher loan amount.

4 What's the relationship between median borrower APR and borrower rate?

5 It is suggested to use filter bars, dashboards and captions.

Resources

Prosper <https://www.prosper.com/>

Jubin's project <https://github.com/jubins/Tableau-Projects>

Prosper EDA project <https://jasoncarter.github.io/Prosper-EDA-R/>

Udacity tableau course