Exploring functions of Prosper ratings by visualizing Loan Data

Explanatory data visualization with Tableau

Links to visualizations:

- Initial version: https://public.tableau.com/views/LoanDataVisualizationswithaTableaustory2/Story1?:embed=y&:display_count=yes
- Final version: https://public.tableau.com/views/LoanDataVisualizationswithaTableaustoryafterfeedback/Story1?:embed=y&:display_count=yes&publish=yes

Summary

An explanatory data visualization was created from a Loan dataset containing 113,937 loans each with 81 features. The data visualization is focused on 3 main features:

- **Borrower rate** for interest rate of a loan
- **Prosper rating** for loan's level of risk, 7 being the lowest level of risk and 1 being the highest risk
- **Loan Status** with 3 **good status** (Current, Final Payment in Progress, and Completed) and 3 **bad status** (Past Due*, Defaulted, and Charged Off)

My two main questions were as followings:

- **"How is a loan interest rate determined?"** (answered in Slide 1 and 2)
- **"Is a higher Prosper rating loan more likely to end up in good status?"** (answered in Slide 3 and 4)

Here are the main findings:

- Prosper rating was found to be a crucial predictor for interest rates.
- Loans that received higher Prosper ratings are more likely to be in good loan status in the future.

Note: * There were originally 6 categories for past due status depending on the number of days passed from due dates, but they were combined as one category, Past Due, for simplicity.

Design

- Slide1

To show changes in interest rates for different Prosper ratings, I first deployed a simple scatter plot and added boxplots to show interquartile ranges for each Prosper rating (Top Left). Since I got a feedback saying why Prosper rating 1 has very low variability in interest rates, I created jitter to show individual

dots instead of dots overlapping too much. This change shows there were less loans with Prosper rating 1 than higher Prosper ratings, this can explain the narrower interquartile range for Prosper rating 1. Next, I broke down each Prosper rating by loan terms to show how interest rates change across different loan terms within each Prosper rating (Bottom). This was modified after getting feedbacks. The original graph was broken down by Prosper ratings for each loan term and this showed changing interest rates across different loan terms less clearly. Finally, the last graph (Top right) is a scatter plot with loan months since origination for the x-axis and interest rates for the y-axis. Different colors for different Prosper ratings help us to see two kinds of things. First, we can see how interest rates change over time for each Prosper rating. Secondly, the colors enable us to see how hierarchically interest rates are determined by Prosper ratings for different time windows; less overlapping colors for more recent loans indicate interest rates for recent loans were more hierarchically determined by Prosper ratings. The color legend for Prosper rating (bottom right) is applicable to all three graphs and an animation created also helps to check each Prosper rating in all graphs. I added a brief description of Prosper rating in the legend to follow a suggestion in the feedback.

- Slide2

United states maps were used to show Prosper ratings and interest rates averaged for each state. Color depth was used to show the averaged values. Since Prosper ratings and interest rates are negatively correlated, I used darker colors for higher Prosper ratings and lower interest rates. If darker colors were used for higher Prosper ratings and higher interest rates, it would be difficult to see the strong correlation from the maps. To show their correlation more clearly, a scatter plot was deployed with the correlation coefficient displayed in its middle. An animation was added to make each state stand out in all three graphs.

- Slide3

This slide is very similar to Slide 2. The only change made was replacing average interest rates with proportions of good status loans. Color depth of the maps also show the correlation between average Prosper ratings and proportions of good loans. Moreover, both color depth of the maps and the scatter plot on the right revealed some outliers, not very extreme though.

- Slide4

The last slide includes two bar plots and one pie chart. The bar plots were used to show the average Prosper rating for each kind of loan status (Top left) and the proportions of good status loans for each Prosper rating (Top right). The bars in the first plot were chosen to be horizontal since category names for loan status are long. The pie chart (Bottom) is somewhat redundant to the second bar plot, but added because it more clearly shows the decreasing proportion of bad status loans as Prosper rating increases. Two colors and their animation were added to separate good and bad loan status. Since the feedback suggested to add some mark labels, I added values with reduced decimal points in format to the bar graphs and removed tooltips for those.

Feedback

- What do you notice in the visualization?

The visualization is nice, clean, and easy to understand.

- What questions do you have about the data?

The interest rate for the prosper rating 1 has very low variability. Do you think there is a reason for this?

- What relationships do you notice?

Prosper rating and interest rates are negatively correlated and Prosper rating and proportions of good status loan are moderately correlated.

- What do you think is the main takeaway from this visualization?

Good prosper rating is a good indicator for a high probability of good loan status, but there are exceptions.

- Is there something you don't understand in the graphic?

No

More suggestions:

- It is difficult to see the changes in interest rates across different loans terms for each Prosper rating.
- The summary introduces Prosper rating, but it would be better if the visualization also has a brief introduction about the feature.
- Adding mark labels to some graphs (e.g. the bar graphs in the last page) would be helpful to visualize the actual values.

Resources

- How to create jitter: https://www.dataplusscience.com/TableauJitter.html
- Changing the number decimal places: http://kb.tableau.com/articles/howto/displaying-zero-decimal-places-for-mark-labels