

Tableau Story about Prosper Loans

-Ratik Dugar

Links

1st draft:

<https://public.tableau.com/profile/ratik.dugar#!/vizhome/1stattemptproj/DefaultratesoftheP2PProsperloans-1stdraft>

Improved draft:

<https://public.tableau.com/profile/ratik.dugar#!/vizhome/2ndattemptproj/DefaultratesoftheP2PProsperloans-2nddraft>

Summary

The dataset is taken from the peer-to-peer(P2P) lending marketplace called Prosper. It has 81 variables and loan information from 113,937 records. Through this visualization project, I investigate the relationship between the variables in order to understand the data better. The questions I want to answer help me understand and identify borrowers who are more likely to default, and creates a foundation of things a lender should look at before investing in loans. The questions I decided on are:

Which income range has the most loan records and on average how much loan is taken out?

How does income and prosper credit rating affect loan default rates?

How do the loan records and default rates vary from state to state?

How has the loan default rate, borrower rate and number of loan records changed over the years?

Does the occupation and debt-to-income ratio of the borrowers affect the loan default rates?

How does the default rate change across the different credit ratings, income ranges and loan objectives?

Design

I used bar charts for my first two visualizations as they were comparing measures across various categories. I also showed the composition of credit ratings within each income group by color encoding it. I try and use colors that don't interfere with

color blindness and use different hues to build optimally distinct palettes. I also create a new measure that calculates the default rate.

For my second plot, I use geospatial data and divide the map into states. I encode this with marker size and color representing the number of records and default rate respectively.

For my third plot, I am trying to show how default rates, borrow rates and loan records are changing over time and this is why I use a line plot. I color encode the income ranges and then use the states as a filter. Again, I use colors that don't interfere with color blindness.

For my fourth plot, I am comparing 2 continuous variables which I suspect have some correlation, and hence, I use scatterplots. I use occupation of the borrowers to color encode the data points and the average debt-to-income ratio to encode the marker size. Here again, I use the states as filters.

For my final plot, I use a box and scatter plot to compare default rate distributions across different listing categories. The listing categories dimension originally had only numbers from 1-20 to represent the different categories. I added string name aliases to all of these such that I can use them as my X-axis categories on the boxplot. I use the prosper ratings to color encode the data points. I further use income ranges and states as filters. Here again, I use colors that don't interfere with color blindness and make sure that the boxes of the box-plots are transparent and not colored so that they can show the colored data points.

In all of the plots where a filter has been used, I display it using a single value dropdown option such as to conserve space and allow the reader to interact with the graphic. A color legend that has been manually sorted is also shown for quick reference.

The first three plots show more of the generalities while the last two add more complexities allowing us to analyze and visualize the data from a more microscopic perspective. I added some more information as tooltips in the final 2 plots after noticing some comprehension delay on the last plot. I also got a suggestion to use red and green because they are more intuitive but I am consciously not using them because they interfere with color blindness. Another feedback was adding some more details to the story texts which I did.

Feedback

I got feedback from 1 person and the answers relating to the questionnaire are listed below.

What do you notice in the visualization?

- Default rate among borrowers is not always representative of their prosper credit rating. I like the use of colors and how you can play around in the graph.

What questions do you have about the data?

- How does the company decide on the borrower rate for a particular case.

What relationships do you notice?

- There is a relationship between the borrower rate and the default rate. There is also an inverse relationship between the loan amount and the income range. The listing category also affects the outcome of a loan.

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

- The main takeaway from this visualization is that there are better ways to understand the likely default rate on a case instead of just relying on the prosper rating and income.

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

- The last boxplot took a little more time for me to interpret but I got it. May be using green and red colors for the different credit rating will be more intuitive. I also wouldn't mind reading a little more about each graphic.

Resources

Udacity Nanodegree course: Data Visualization in Tableau