Story of Prosper Loan Data

by

Shruti Tiwari

First version (https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data/Story4 (https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data/Story4))
Second version (https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data_final/Story4?publish=yes (https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data_final/Story4?publish=yes))
Final version https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data_final_reviewed/Story4?publish=yes (https://public.tableau.com/profile/shruti8810#!/vizhome/loan_prosper_data_final_reviewed/Story4?publish=yes)

Dataset

This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, borrower employment status, borrower credit history, and the latest payment information.

Summary

In peer-to-peer lending, there are three main stakeholders: borrowers, lenders and the company itself. In this Tableau story, I have tried to shed lights on the who are the borrowers and what affects the default rate of the loans. First, I show which states have more borrowers then extended this to see which state has more defaulted loans. Turns out California has both most number of borrowers and defaulters and North Dakota least number of borrowers and zero defaulters. Next, I explore what is the distribution of loan status. I found that nearly 50% of the loans have "current" status and 33.42% are "completed". 4.42% of all loans resulted in "defaulted". next I wanted to see how borrowing and average loan amounts have changed with time. I found from 2006 to 2004 the trend of loaning an average loan amount has increased except a minimum around 2008-09. I concluded that dip in the trend may be a result of recession at that time. My next question was what are loans being used for? Most of the loans are being used for "debt consolidation".

My next question was how the borrowers' behavior varies with their income range. I found better the income range less the default rate. Similarly better the Prosper score and rating lower the chances of defaults. Interestingly, same can not be said for employment status. I found that most defaulters are full time employed. My next inquiry was the role of the loan term. I found most of the loans are taken for 3 years and most defaulted loans are also for 3 years term. That made me think what if borrowers get more time maybe there will be fewer defaults but term 1 also has a low default rate. Then I visualized how loan terms are related to average loan amounts and I saw that if the large loans are borrowed for 3 years they might result in defaults. My last question was which students are borrowing money and the answer was more senior students borrowing more number of loans.

Design Decisions

I first sifted through the dataset and roughly thought about which variables I was interested in exploring and which were outside the domain of my exploration. I first planned on making another file which would contain the subset of variables I want to explore but instead opted for using the entire dataset in case I got new ideas midway through my exploration.
I wanted my Tableau story to have a purpose, to show to those who are interested in the facts about the loan data. My focus was loan borrowers, which occupation defaults, their income, and which loans borrowers took. I got a few amazing feedbacks like mapping state-wise default rate, including the loan status distribution, label editing, and some minor spelling mistakes. All these have contributed to improving my visualization.
• I used geomaps to show the number of loans by states for users to recognize the states with higher number of borrowers with darker shades of blue.

• I chose stacked bars to show the users defaulted and other loans by states. Stacked bars help the users to see the pattern in total number of loans and defaulted loans as well as ratio of defaulted to total number of loans.

• I used highlight tables and packed bubbles to show the counts of loan status. these visualizations clearly communicate with the users the counts of different loan status.

• Line plots have been used to show the time evolution of the number of loans and loan amounts as line plots give the best estimate for time series.

• Side by side bars may give the best idea to usera of the effect of income range on defaulter rate.

• Vertical bar chart so that users can see default rates for Prosper score and Prosper rating.

Feedbacks

After completing the first sketch of my Tableau story I shared it on Udacity Slack group and emailed the link to two of my friends. I received few important feedbacks mentioned below:

• Changing the number of defaulters by states. My previous visualization was not clear on sorting which state has the most number of loans and defaulted loans.

• Including a table showing the distribution of loan status.

• Removing Listing category "Not available".

• Fixing labels and formatting e.g. changing "count of loan key" to "Number of loans".

References

- Prosper About
- Prosper Data Analysis Project on Kaggle
- Repo on the same project
- MIT page using stats on the same project
- Udacity Tableau course
- Tableau tutorials