Create a Tableau Story using Prosper Loan Data

By Chenchen Li

May 2018

Project Links

Tableau Public Workbooks:

- First version: https://public.tableau.com/profile/chenchen.li#!/vizhome/First_version_prosper_loan_data/First_story publish=yes (https://public.tableau.com/profile/chenchen.li#!/vizhome/First_version_prosper_loan_data/First_stor publish=yes)
- Final version: https://public.tableau.com/profile/chenchen.li#!/vizhome/Revised_version_prosper_loan_data/Final_s publish=yes (https://public.tableau.com/profile/chenchen.li#!/vizhome/Revised_version_prosper_loan_data/Final_ publish=yes)

Summary

Personal online loan business has been an important way for borrowers to raise money and for investors to earn interest. However, the risk of this type of investment is relatively high. It's important to address the default risk for this business. The Prosper (https://www.prosper.com (https://www.prosper.com)) Loan Dataset contains 113,937 loans from Prosper with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others. In this project, I explored many aspects for personal loan business, including the number of loans based on different loan status or ratings over time; Loan amount and number of loans based on different states over time; the borrower rate across different states of US, sereval factors which have clear impact on borrower rate, etc. Moreover, this analysis showed that higher risk loans don't always guarantee higher return so a useful tip for investors was given.

Design

The prosper dataset is a complex dataset, so before I began to create the visualisation, I needed to figure out what is the purpose of my Tableau story using this dataset. The most important things that I cared about for this dataset are the trends of loans over time, the status and the relevant factors about the borrower rate and the default risk for this business.

In this project, I divided this idea into four groups:

- Trends of the loans over time based on loan status, Prosper rating and different states.
- The spatial distribution of the average debt to income ratio and median borrower rate.
- The relationships between the borrower rate and the relevant factors.
- Interesting patterns of risk and useful insights that were found in this process: Emphasise that higher return means higher risk, and tell the audience don't put money in C&D loans based on this dataset.

Dashboard one

- Trends of the number of loans over time based on different loan status: Line Chart;
- Trends of the number of loans over time based on different loan rating: Line Chart;
- Trends of the number of loans and loan amount over time based on different states:scatter plots, loan count was encoded as size and states were encoded as color.

Dashboard two

In order to see whether loans differ across the different US states and how does that relate to the number of the loans and the average debt-to-income ratio of lenders in these states, I chose to use spatial distribution of these two factors to stand for different state.

- Spatial distribution of the count of loans and the average debt-to-income ratio: loan count was encoded as size and debt-to-income ratio was encoded as color on top of the map of United States.

Dashboard three

Similarly, in order to see whether borrower rate differ in different states, thematic map was used to show the distribution of borrower rate for different state.

- Spatial distribution of the default rate: Thematic Maps.

Dashboard four

Here I made two visualisations to show the relatioships between two relevant factors to borrower rate, and a line chart showing the trend of borroer rate over time:

- Borrower rate & IncomeRange: Bar Chart.

- Borrower rate & ProsperRating & IncomeRange: Stacked Bar Chart, with one bar for each ProsperRating and using color to stand for each IncomeRange within that ProsperRating loans.
- Trend of borrower rate over time: Line Chart.

Dashboard five

Since ProsperRating is a good indicator for the default risk of the loans, lastly, I have looked at EstimatedReturn , EstimatedLoss , EstimatedEffectiveYield, and the Actual Net Principal Loss for loans of different rating using bar charts. Also, I used “color” to draw attention for the “C”, “D” loans in the last

Feedbacks

After completing the first version of my Tableau story I shared it with two of my friends. I got some important feedbacks mentioned below:

- Showing trends of the loans over time based on states: for this I created a scatter plot, loan count was encoded as size and states were encoded as color.
- Finding the trend of borrower rate over time : this showed an interesting finding that since 2011, the borrower rate was constantly decreased over time.
- There were also some minor feedbacks on spelling, typos, grammar and chart changes.
- All these have contributed in improving my visualization.

Resources

- Tableau Help: http://onlinehelp.tableau.com/current/pro/desktop/en-us/default.html (http://onlinehelp.tableau.com/current/pro/desktop/en-us/default.html)
- Tableau Training: https://www.tableau.com/learn/training (https://www.tableau.com/learn/training)
- Prosper Data Analysis Project on Kaggle: https://www.kaggle.com/jschnessl/prosper-analysis (https://www.kaggle.com/jschnessl/prosper-analysis)