A Story about the Online Peer-to-Peer Loan Business

Author: Jubin Soni ***Tableau Project: Data Analysis Nanodegree***

Visualization: Story Version 1, Story Final Version, GitHub

About

Prosper is a peer-to-peer lending platform that aims to connect people who need money with those people who have the money to invest. In this data analysis project, I have explored the Prosper dataset and used Tableau to create my visualizations.

Dataset

The Prosper loan 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, and many others. The dataset and data-dictionary can be found on below links.

- Prosper Loan Dataset
- Prosper Loan Data – Variable Definitions

Summary

In peer-to-peer lending, there are three main stakeholders: borrowers, lenders and the company itself. In my Tableau story I have done exploration on the relationship between these people, what affects borrowers Prosper Score and who defaults the most. First, I have done a time series analysis ranging from 2009 – 2014 about the number of loans taken by borrowers, the amount of their loans and how their ProsperScore got affected in this duration. I noticed that since 2009, the loan business increased and climbed up quickly since 2013 and then dropped down at beginning of 2014 while the borrower credit scores constantly dropping over this time and some states having default rates more than 30%.

Then I have explored the defaults, I found out that the people with $0 income have highest default rates and most defaulters invest in the loan type 'D'. Breaking down to occupation wise, an interesting pattern was found that the college student group which are enrolled in higher grade studies have more loans, higher borrower and default rate. While this made sense with $0 income, the sophomore students were the top defaulters and having lower number of loans. Lastly, I have looked at the income and losses on different loan ratings – the 'HR' loan rating had the highest loss even though this type of loan is given to most credit-worthy borrowers. However, looking at the net principal returns over the time I noticed loans C&D had highest losses than other loans and are most risky.

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 borrowers purpose for loan and some minor spelling mistakes. All these have contributed in improving my visualization.

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:

- Showing state-wise default rate – for this I created a new numerical variable called Default Rate from Loan Status and mapped with Borrower State, this showed an interesting finding that in states like CA, TX, NY, IL the default rates were quite high with CA having highest defaulters (>700).
- Finding why people borrow money – this was tricky but from data dictionary I found a variable called Listing Category that had reasons for loans. I found out that since 2007 borrowers took most Personal loans which suddenly dropped in 2009. Then since 2009 the most loans borrowers took were for Home Improvement or Business which steadily kept increasing with steady-drop in borrowers' credit rating.
- There were also some minor feedbacks on spelling, typos, grammar and chart changes.
- I really appreciate all the feedback I got which helped me in improving my plots.

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
- Learned a lot and had to do to lot of googling to figure out how to do stuff