Prosper Loan Data Visualization Using Tableau

Introduction:

Prosper is a platform for both lenders and borrowers in the Unites States. Borrowers apply online for fixed rate, fixed term loans between $2,000 to $40,000 and Investors like institutions and individual investors invest in these loans. Prosper on the other hand provides loan servicing on behalf of the lenders and borrowers.

Dataset:

The Prosper Loan dataset has 113937 records of various P2P loans and has 81 variables describing the details of these loans given. The variables in this dataset describe the details of borrowers (i.e. State, Loan Amount, Credit Score, Debt to Income Ratio etc) and lenders (Yield, Profits, Losses etc).

Summary:

The scope of this analysis is to see how one of the variables named *ProperRatings* is affected or described by other sets of variables. I have tried to analyze the relationship between variables related to borrower's profile (i.e. Credit Score, Income Ranges, Stated Monthly Income etc.) and Borrower Rates. Additionally, how the Borrower rates affect the Prosper Ratings. I have also tried to touch-base on the relationship between Loan Amounts and the Category Type of loans.

There are two versions created for the Prosper Loan Story. The initial version is a first draft and has received feedback which is incorporated in the second version of the story. Both the versions are listed below:

1. Version 1 : draft : You need to scroll below click on links labeled Story1, Story2. Story 3 to see all the stories.
   https://public.tableau.com/profile/radhika7971#!/vizhome/ProsperRadhika/Story1CorrelationMatrixBorrowerRatevsProsperScore?publish=yes
2. Version 2 : final
   https://public.tableau.com/profile/radhika7971#!/vizhome/ProsperRadhikaV2/ProsperLoanP2PLoanDataStory?publish=yes

Based on the story published above, we can draw the following conclusions visually:

1. The borrower rate is inversely proportional to the prosper rating (from the correlation matrix and the more detailed visual on the 2nd slide of the story). The borrower rate is the rate of interest that a borrower must pay to the money lender (in this case Prosper). The Prosper Rating represents an estimated average annualized loss rate range. Prosper ratings (in order) go from AA, A, B, C, D, E, to HR. Where AA represents a borrower with a low risk profile, and HR represents a borrower with a high-risk profile. Intuitively, this suggests that as a percentage of borrowers defaulting on the loan, borrowers in the HR bucket have a higher likelihood of not being able to pay back or defaulting on the loan compared with AA or A bucketed borrowers. Since borrowers with an AA or an A rating are more likely to repay the loan, they are offered a loan at a lower rate of interest compared to borrowers with a prosper rating of HR. This means that to get a competitive interest rate on the prosper loan, you must have a high (AA, A, etc…) prosper rating. However, if you think about it from the perspective of a borrower (an alternate

model), having a lower prosper rating (of HR) means that you are more likely to default on the loan, so intuitively an individual with a prosper rating of HR should receive a lower interest rate so that they have a fair chance of being able to pay the loan back. However, this alternate model doesn't work well for lenders, so the borrower interest rate is directly related to how risky the individual is in terms of being able to pay the loan back. i.e. higher the risk, higher the interest rate.

2. When looking at the total amount of money lent out to borrowers broken down by the type of loan (i.e. the purpose the borrower borrowed the money in the first place), we notice that majority of prosper loans were meant to consolidate debt from multiple other sources (such as credit cards). One can speculate why that may be; on average credit cards charge ~17% interest per year. However, if you are able to consolidate the debt from the cards and transfer it into a prosper loan, depending on your prosper rating, you could reduce the interest rate to ~7% per year, which is a 58% reduction!
3. Loan default rates are highest in the states of North Dakota, and Iowa. The high default rates seem to be related directly to the fact that the average monthly income is the lowest in these states compared to any of the other states in the USA. This is despite the observation that on average, each loan given out in these states has the smallest absolute amount lent.

Design:

1. After studying the data, I gathered that *Prosper Ratings* & *Borrower Rates* are the main two variables that I wanted to study and analyze. I plotted a Correlation matrix first that would help me set a path to the analysis. This matrix has led to the following insights:
    a. Borrower Rates and Lender Yields show a strong positive relation. This one felt intuitive since lenders would earn a higher when they lend at higher rates.
    b. Borrower Rates and Prosper Ratings show a negative correlation.
    c. The Lender Yield and Prosper Ratings show a negative correlation implying that lenders. earn lower yields for higher (i.e. good) prosper ratings.
    d. Inquires in last 6 months are higher for lower Prosper Ratings.
    e. Credit scores and Borrower Rates don't show a relationship as such which is surprising.
2. Income range of $25k to $49K have the largest number of loans. Adding Prosper Ratings and Loan Amounts to this chart tells us that a higher Prosper Ratings have higher loan amounts. In other words. People with a good prosper rating take bigger loans.
3. Loan Amount increased drastically in 2013. Term of the loans are 1 Yr, 3 Yrs and 5 Yrs; 3 Yr loans being popular. Debt Consolidation segment has the highest aggregate loan amount.
4. I wanted to investigate the default rate for each state based on average monthly income and average debt to income ratio using maps visualization. The average default rate is high for states having lower average monthly income. The debt to income ratio does not do a good job in explaining this effect.
5. I have also used bar charts to investigate the default rate using income ranges. Adding prosper ratings to this chart makes it clear that people with lower income ranges and lower prosper rating have a higher default rate. The prosper rating "A" and "HR" show some fluctuation as shown in the chart.
6. I tried to study the lender variables of *return* and *losses*. Plotting it on a map does not give us a true picture of their relationship. Plotting these variables based on Occupation and Employment

Status shows that the relation is linear. Also, people that are employed tend be given larger loan amounts. Additionally, we also see average number of investors are more for the debt consolidation segment. This could be because the prosper loan kitty has people taking loans more for debt consolidation compared to other segments.

Feedback:

1. Combine all stories to tell one story.
    a. Addressed in the Version 2
2. Remove Credit Score Lower and Borrower Rate sheet. It seems repetitive.
    a. Addressed in the Version 2. I made this slide to make it clear to the reader that Credit score and Borrower Rates show no correlation which I felt was strange but this is already shown in the correlation matrix so I removed it.
3. Change the title of sheets to remove the clutter
    a. Addressed in the Version 2

Resources:

1. Prosper Loan Website (https://www.prosper.com/about)
2. You Tube
    a. https://www.youtube.com/watch?v=vhrmmvAvOfQ
    b. https://www.youtube.com/watch?v=wa4lOPxNyPc
3. Udacity Data Story Telling Course
4. Prosper web page for loan consolation
5. Google to find average credit card interest rates