Loan Data Analysis

MSDS DataViz Final Report

Nan Lin

May 13, 2019

Github: https://github.com/nanlin7/LoanData-DataViz-Tableau

SUMMARY

This Tableau story aims to throw light on the default rate of different borrowers in peer to peer lending investment. Through visualizing default rate data by borrowers' income, occupation, prosper score, state, and related return, the audience can choose borrowers more wisely based on their risk tolerance. To minimize the probability of default, risk-averse investors should find borrowers who have higher Prosper score, higher income, offer relatively lower interest rates are a good fit. In term of occupation, students are usually more likely to default with one exception: graduate student. Plus, a risk-averse investor should pay more attention to borrowers located in North Dakota, Maryland and Lowa State as the default rates of these are as high as 0.29, 0.37, 0.36

Design

The story is started with a line chart to indicate the trend of the number of loan records and the total amount over time. Dual axis makes it possible to display two visuals in the one plot and we can compare the patterns over time.

Next, I create a map to display the default rate of borrowers from different states. Color is used to display the magnitude of default rates clearly.

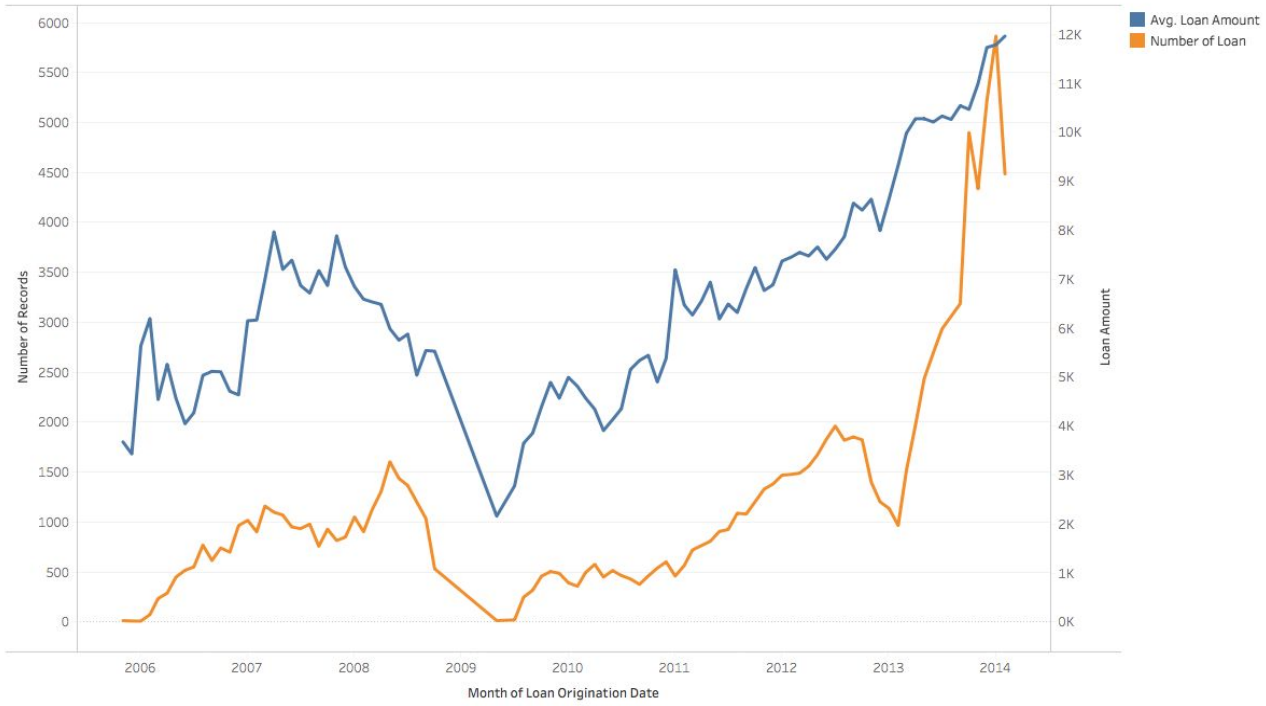
For categorical data as occupation, I choose a bar chart which allows easy comparison of the number of loan records by occupations.

Then the borrower rate and the related default rate are displayed as a collection of points. This scatter plot enables an audience to obtain a visual comparison of the two variables and observe the correlation. Further, student data is encoded in a different color to emphasize the observation I want to convey.

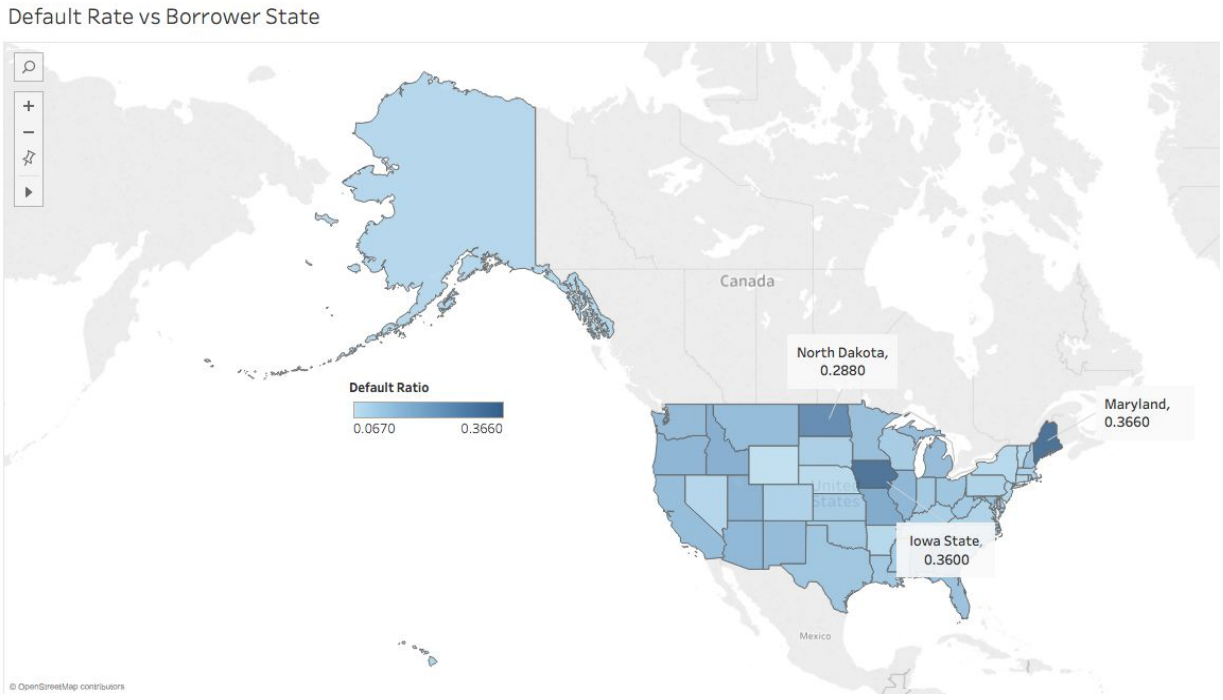
Plots

1. Line Chart
2. Chloropleth Map
3. Bubble Map
4. Connection Map
5. Barplot
6. Scatterplot

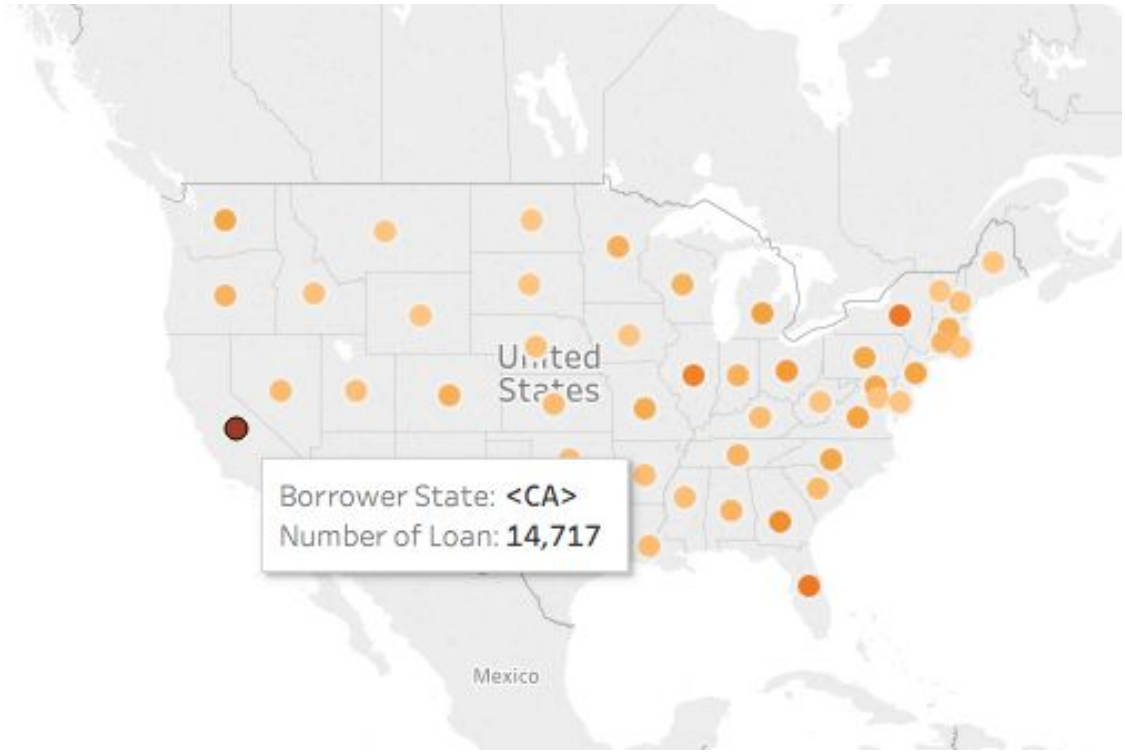
7. Heat map
8. Histogram
9. Stacked area or stream graph
10. Histogram
11. Treemapping
12. Interactive plot: check tableau story! They are all interactive



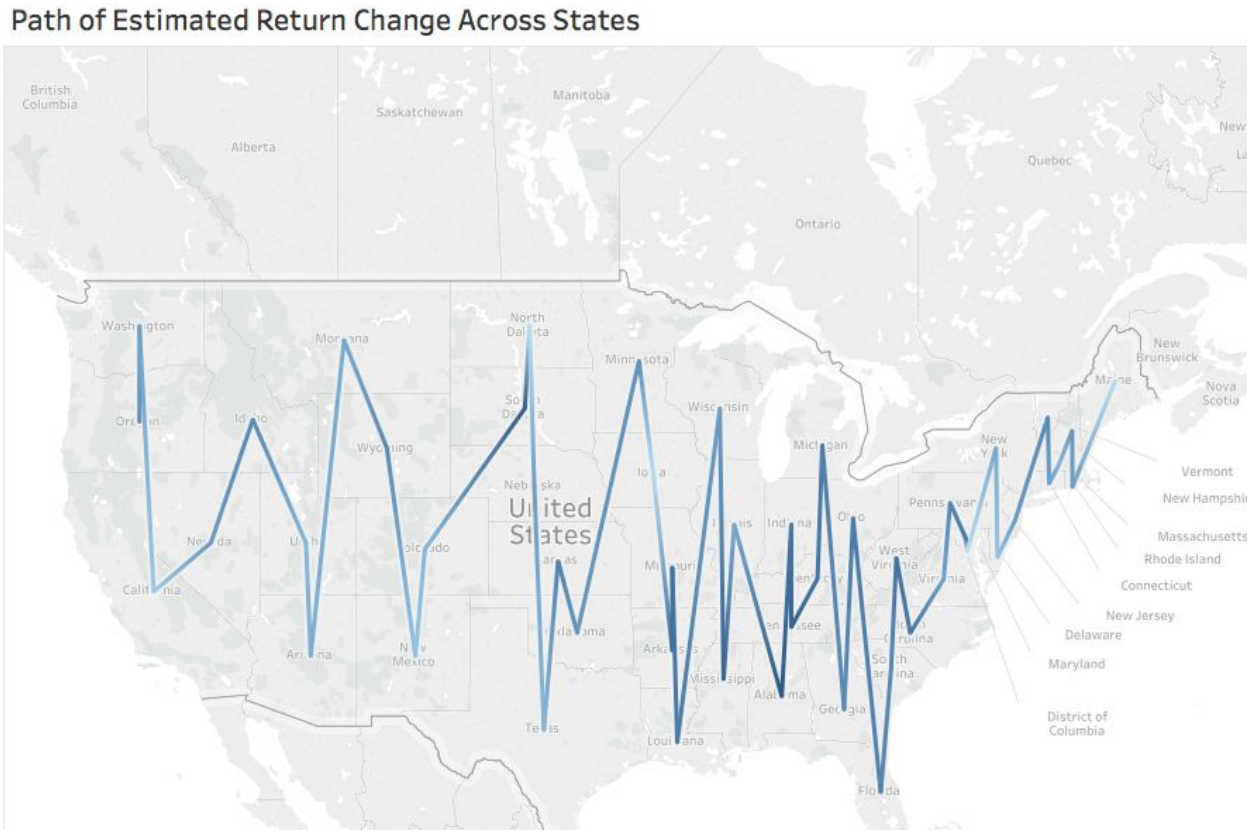
Since **2010**, the average loan amount and the number of loan records have increased steadily. The number of records climbed up quickly since **2013** and dropped down at the first quarter of **2014**. On the other hand, the average loan amount showed a clear upward trend from **2010**.



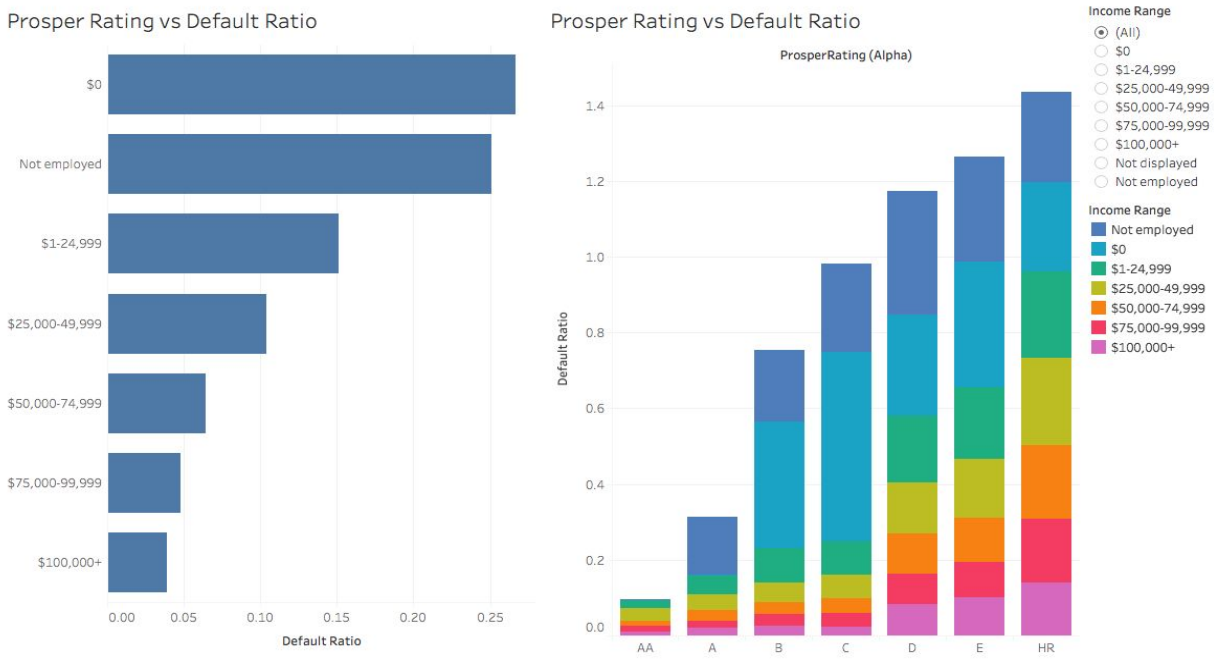
Default Rates for different states of the US have obvious variance. Borrowers from North Dakota, Maryland, Iowa State tend to have higher default rates.



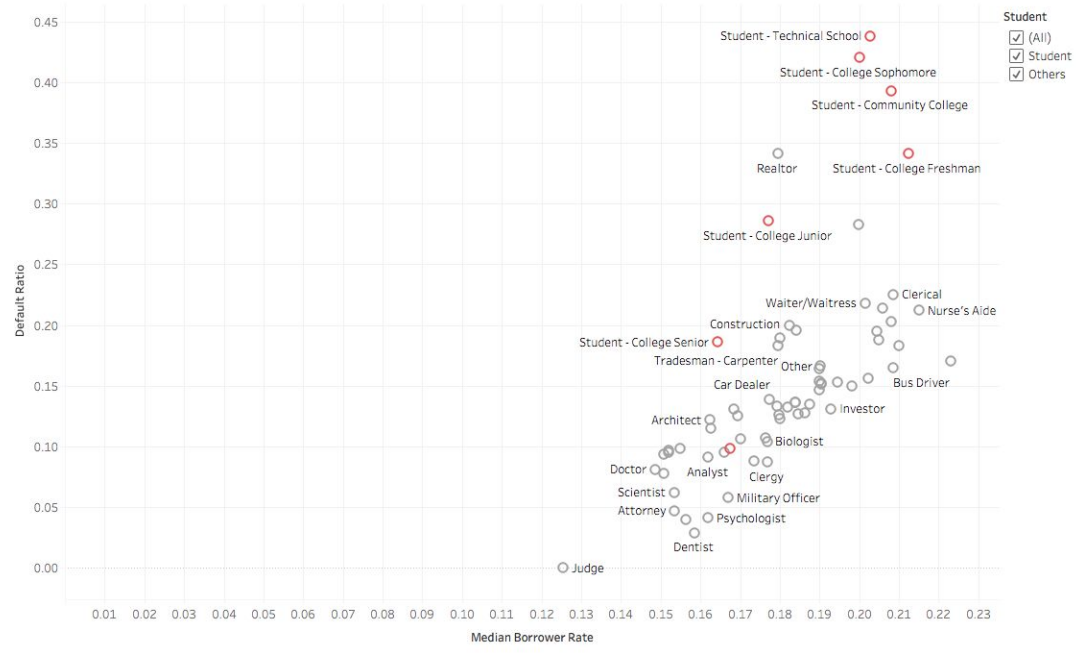
If we look at the number of load records by borrowers' states, California stands out as the state with the largest amount of records but not the highest default ratio as we show above.



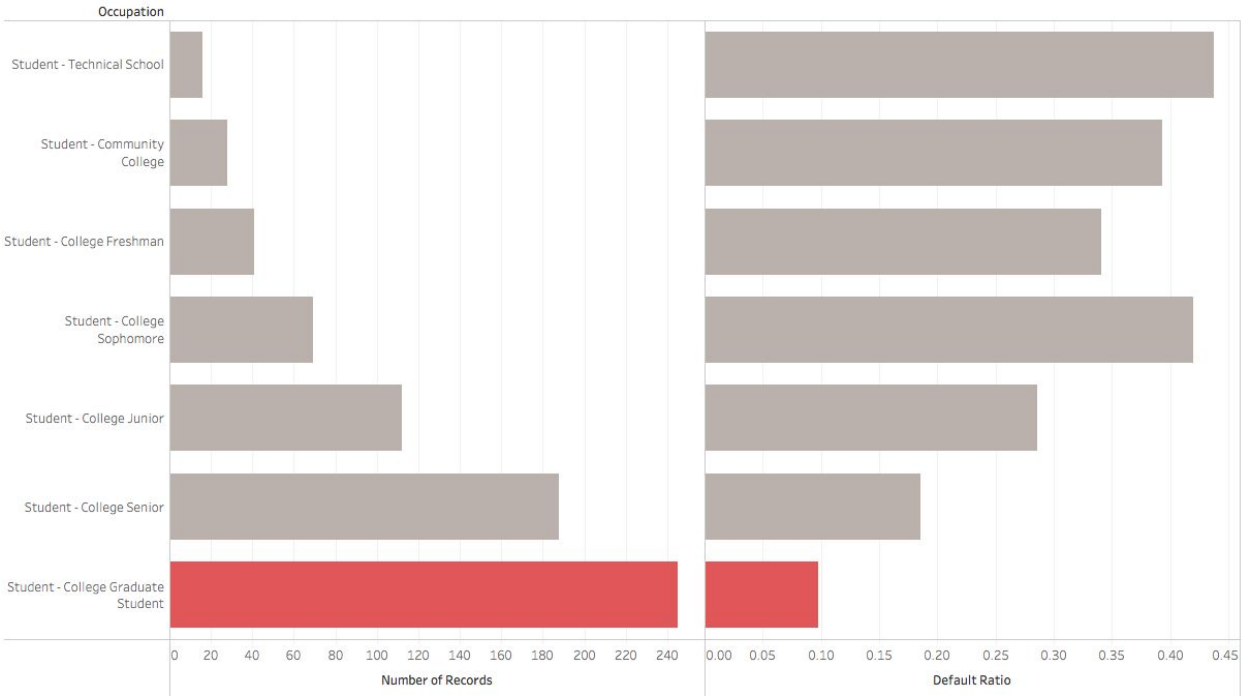
Through a connection map, we can see how the estimated return is changing from state to state.



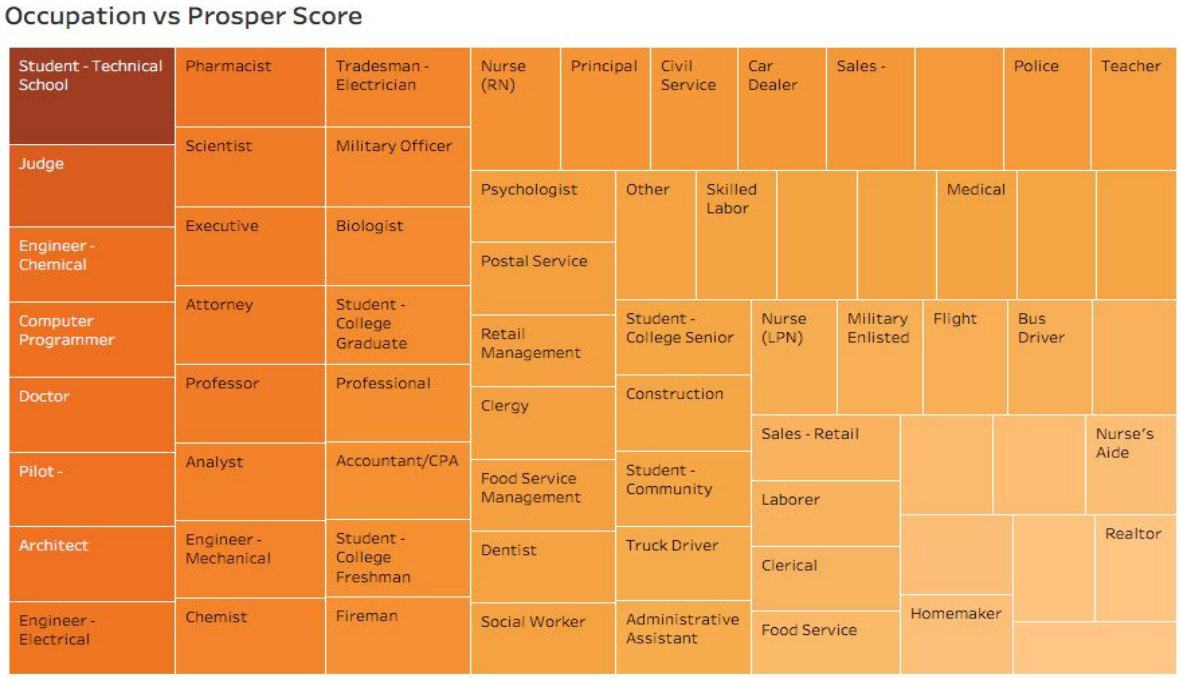
Income and credit rating are two factors that influence default rates. Borrowers who have higher income are less likely to default. Better credit rating means lower default rate.



Taking borrowers' occupation into consideration, students are usually more likely to default with one exception.

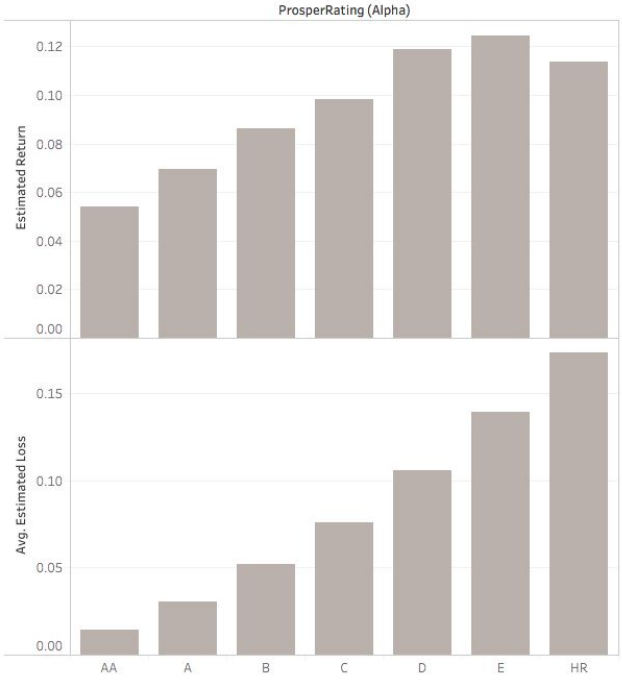


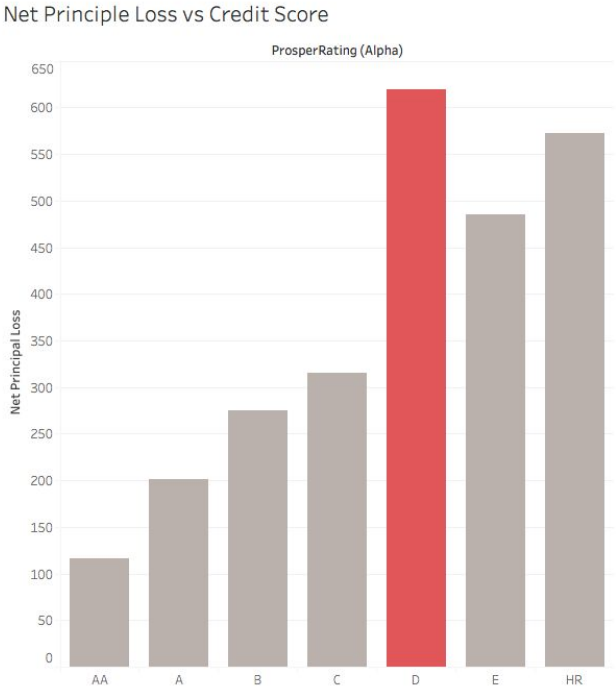
Students do not seem to be a good business with one exception for graduate students. The graduate student group has more loans but the lowest default rate compared to college students, technical students, and community college students.



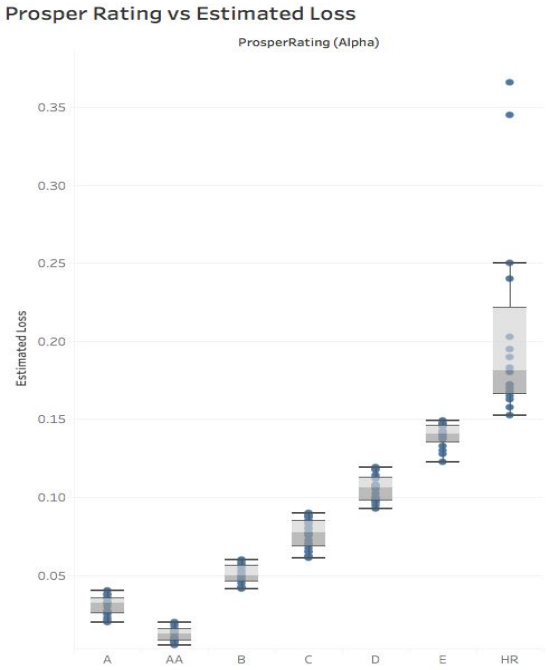
We are also curious about the prosper score, which occupation tend to have the highest prosper score? Students are also the group of borrowers who have highest prosper

scores. From this angle, this also confirms with the previous finding that students have the lowest default ratio compared to people of other occupation.

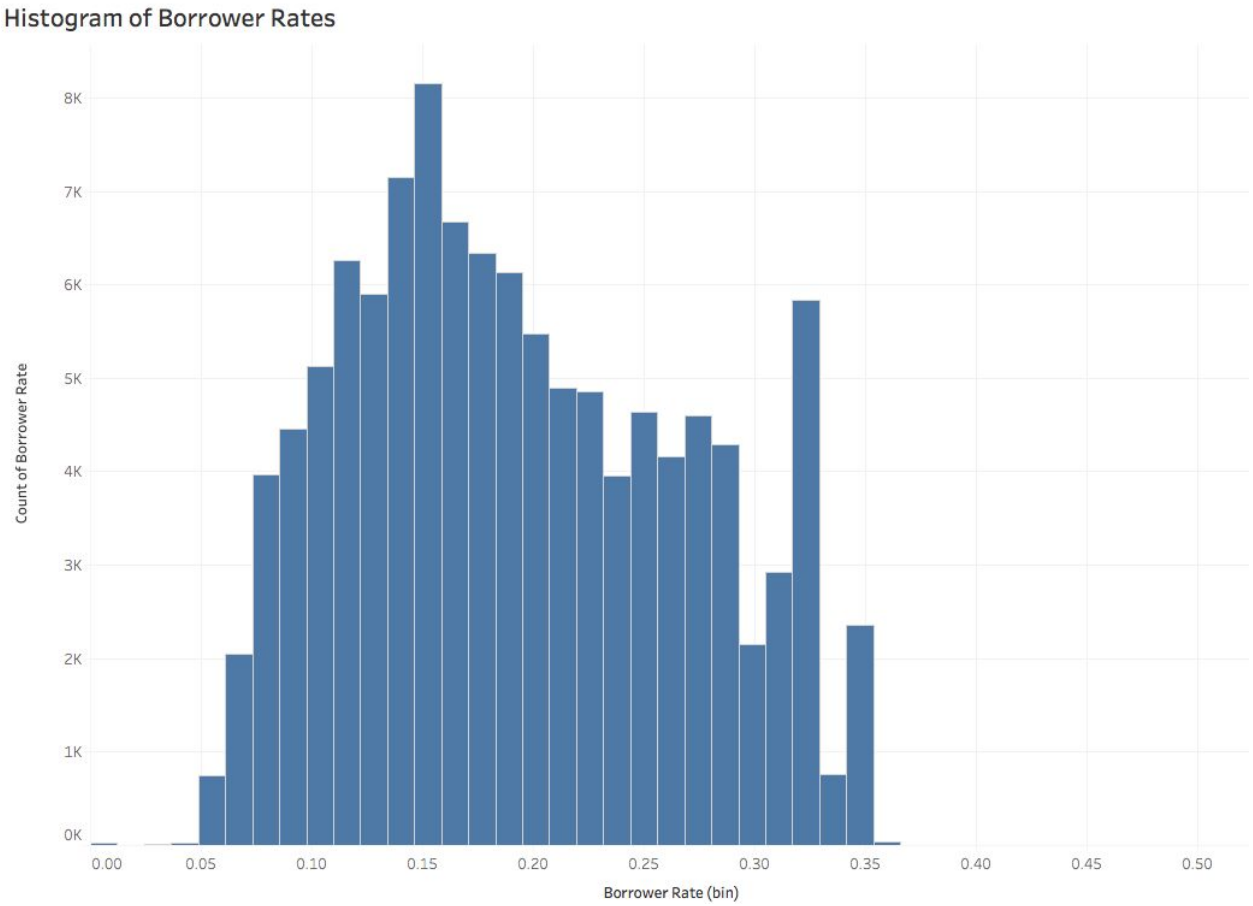




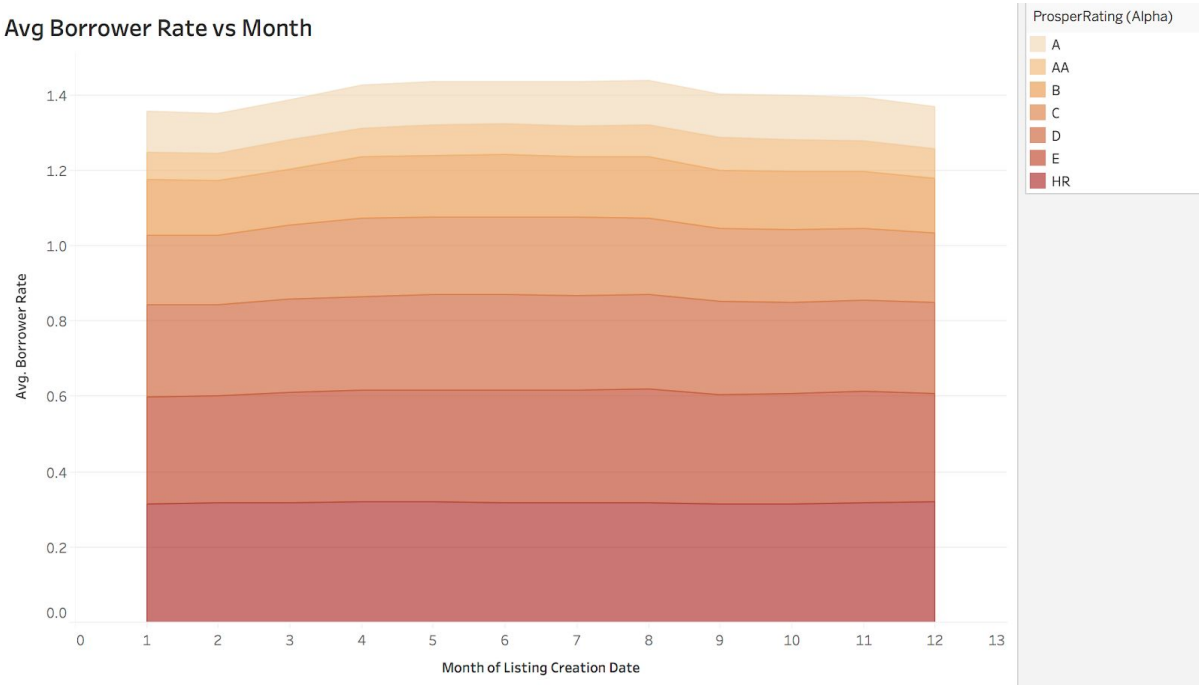
Higher return means higher risk. There are always trade-offs. But loans rated D have the highest net principle loss.



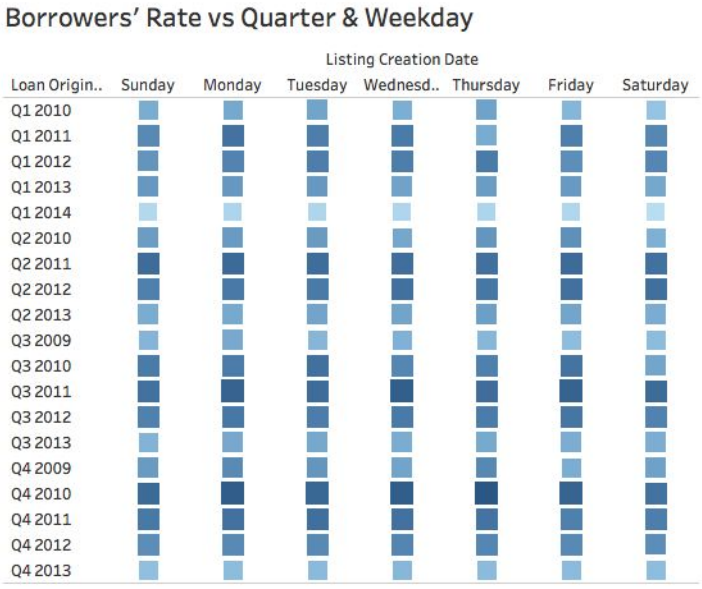
If we look into the relationship between Prosper rating and the related loss. We can see as the rating increase(AA is the highest rating), the estimated loss decreases. If we combine the estimated loss and default ratio, they both decrease as the rate increase.



How is the distribution of borrowers' rate look like in our dataset? From the above histogram, we get a sense of the distribution. Most borrowers have rates between 0.10 and 0.20.



When we look into the borrower rate by Prosper Rating and the month of listing creation, we can see the overall trend is similar for different rates. Summertime tends to have a higher borrower rate. This might be due to the large demand in the loan industry.



To study the seasonality of borrowers' rate. I utilized heatmap with color gradients. The deeper the color, the higher the rate.