Visual Analysis of Prosper Loans

– *Tarun Lingam*

Tableau Public link to the Visualization

https://public.tableau.com/profile/tlingam#!/vizhome/ProsperLoanDataVisualization_0/Story1

Summary

The Prosper Loan dataset contains 113,937 loans with 81 variables for each loan, dated from Q4 2005 to Q1 2014. With my visualization story, I want to show the impact of Prosper Rating in forecasting Loan Status.

- My visualization story begins by giving an introduction to Prosper and showing the loan status distribution, loan amount distribution and number of loans by state.
- In the next page, I show the effect of the recession in 2008, which led to a steep fall in the number of loans and average loan amount. It also shows that borrowers had very high Debt to Income ratio with poor Credit Scores before the recession.
- Prosper was forced to shut down operations due to the recession and register with the SEC. It restarted operations in Q2 2009 with stricter regulations. It started rating the loans according to their risk from HR (highest risk) to AA (least risk). From Q2 2009, we see a steady increase in business for Prosper.
- In the next three pages, I show the factors impacting Prosper Rating. I chose Monthly Income, Current Delinquencies, Public Records in last 10 years, Bankcard Utilization, Debt to Income Ratio, Credit Rating and Income range to show the major variables impacting Prosper Rating. Also, we see that borrowers with a higher Prosper rating have lower APR on their loans.
- From the previous pages, I established that Credit Rating and Income range have a significant impact on Prosper Rating, so in the next page, I show the impact of Credit Rating and Income Range on Loan Status. The percentage of bad loans (Delinquent, Defaulted, Chargedoff) progressively decreases with increasing Credit Rating and Income Range.
- One notable finding is that more than 50% of loans are Defaulted or Chargedoff for Poor Credit Rating loans. Also, there are no Current Loans with Poor Credit Rating as Prosper stopped offering loans to borrowers with Poor Credit (0-600) after the recession.
- In the final page of my story, I show the impact of Prosper Score and Prosper Rating in forecasting Loan Status. The percentage of Bad Loans (Delinquent, Defaulted & Chargedoff) reduces progressively with Higher Prosper Rating.

Design

After an initial look at the data and information about Prosper, I decided to focus my visual story on the impact of Prosper Rating in forecasting the outcomes of loans (Loan Status).
First, I wanted to show the factors that impact Prosper Rating, so after analyzing the data, I focused on Credit Rating, Income Range, Stated Monthly Income, Monthly Payments, Bankcard Utilization, Debt to Income Ratio, Current Delinquencies and Public Records in the last 10 years.

- To introduce my story, I used a map to show the distribution of loans by state, a histogram to show distribution of loan amounts and a bar chart to show the distribution of loan status.
- In the second page, I used a time series plot to show the changes over time while highlighting the effect of the recession on the business of Prosper.
- I used a line graph with a consistent frame of grouping by Prosper Rating to show the changes in Credit Score, Debt to Income Ratio and Borrower APR across Prosper Rating.

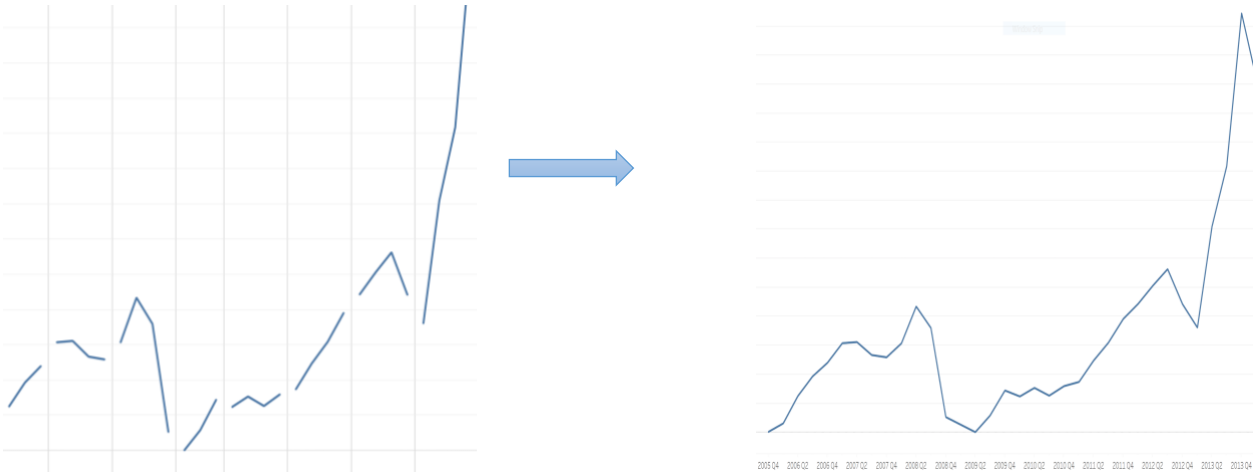
- In the rest of the pages I used bar charts with a consistent frame of grouping by Prosper Rating and Credit Rating to show the relationship of my chosen variables to Prosper Rating.
- In the final pages, I used bar charts, colored by loan status to show the impact of Credit Rating, Income Range, Prosper Score & Prosper Rating.

Choosing map for reason that it's a direct and interactive way to show and filter out other visualizations. Choosing histograms for reason that it's a direct way to get frequency of each variable, especially for continuous data. Choosing time series for reason it's a great method to show data points indexed in time order. Choosing bar charts for reason that it's convenient to compare changes over fixed time period for more than one group. Choosing line charts for reason that it can easily show how one variable is affected by another. Choosing the layout of the pages containing bar charts and line charts for reason that it's convenient to compare and filter different metrics of the variable. And the design of legends and filters are due to the fact that each categorical variable has different levels. Legends will change according to the filters and be compared by color. Choosing hover effects in the dashboards to highlight the changes between the groups.

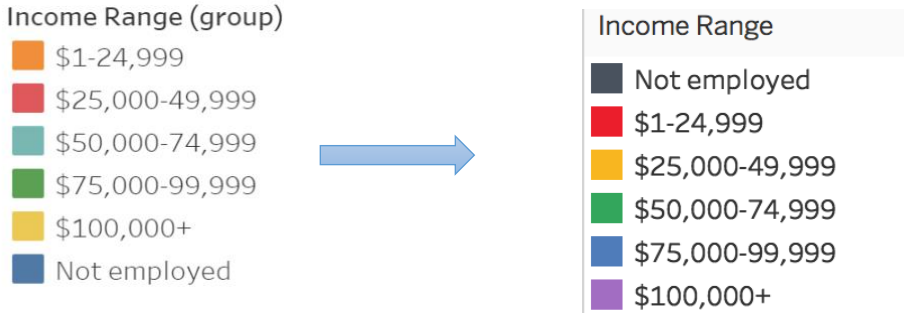
Feedback

I got feedback from my Udacity mentor while designing my Tableau charts, dashboards and story.

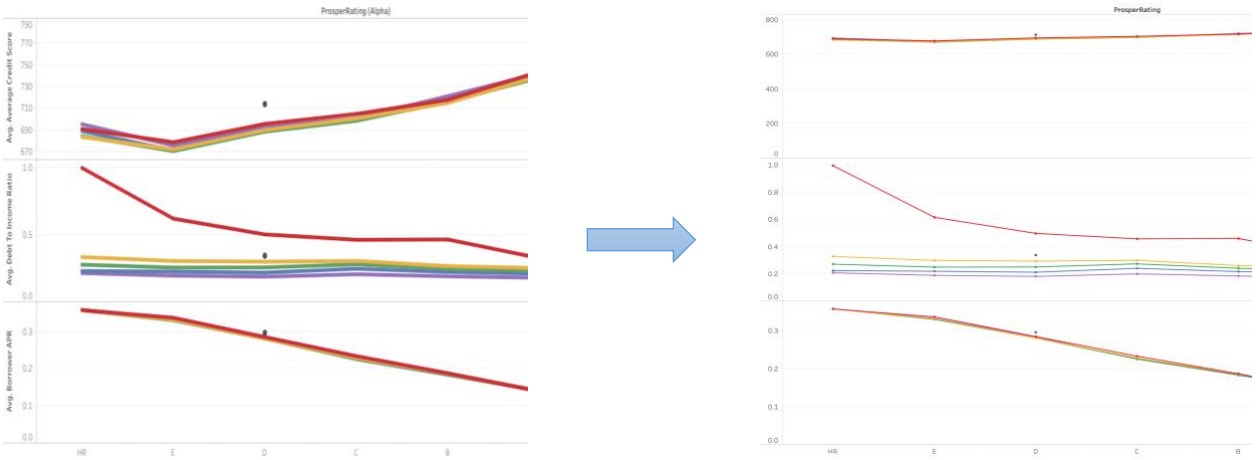
- I was suggested to use a continuous line chart instead of a broken line for the time series plot.



- In all my legends, I was suggested to use a color progression from red to blue to signify better Loan Status and higher Income Range. I was suggested to use a qualitative accent palette.



- I was suggested to make the Credit Rating group more descriptive, by displaying the credit score range in each Credit Rating group.
- I was suggested to make the lines in the line chart thinner to have better visibility when lines overlap.



- I was suggested to avoid starting the Y-axis on values greater than 0 as it is misleading. For example (Please refer to the previous images), in the initial line chart, the viewer gets the impression that the average credit score of AA Prosper Rating is 4 times higher of HR Prosper Rating.
- Finally, I was suggested to rename the axis titles, so as to cater to all kinds of viewers (data-literate and not data-literate). For example, "number of loans" is more descriptive than "number or records".

Resources

1. Information on Prosper
    a. https://www.prosper.com/
    b. https://en.wikipedia.org/wiki/Prosper_Marketplace
2. Udacity course on Tableau
3. Tableau training & learning

Data Source

The original Prosper loans dataset is available here.