# Prosper Loan Data Visualization

## Summary

Prosper Marketplace is a peer-to-peer lending marketplace connecting those who need a loan to those who have extra money to lend. I investigated the performance of Prosper Loan from the Investor and Borrower point of views by looking at the height and color distribution of each bar chart for different loan types. The bar charts show how many loans of each specific types are given to borrowers with different scores and what portion of those loans are completed, defaulted, or charged off. I found two types of loans, Motorcycle & Engagement Ring, have highest ratio of successfully completed loans for all borrowers with different scores. And for borrowers with specific ProsperRating, the "Debt Consolidation" loan type with more than 11000 loans has highest chance to be invested.

## Design

I used the Martini Glass structure to provide above mentioned variables animated and then used by reader in an interactive way. This data visualization contains following sections:

**1- Indicator section:**

In order to add the "Loan Type" as a variable to the animated visualization to be used for pausing the animation for further investigation, I added it as a bar chart in indicator section with its size related to the total number of loans given to the borrowers for each type. The design characteristic of this section is as follows:

a. A bar chart built on the right side of the svg page with x axis for number of loans, "count", for each group of summarized and filtered loans that is described in details on the "ProsperLoanData_Project - Viz_filtered.Rmd" document. The y axis shows "ListingCategory" and each bar's size is based on the log value of the "count" in order to have almost similar length of bar charts due to the big differences of x values. I also removed extra title and line indications of x and y axis.

b. Added Title "LOAN TYPES" on top of the indicator bar chart and a note on how to pause and activate the animation by clicking on each bar of the bar chart for different loan types.

c. Since the loan types," ListingCategory", is a numeric variable, I manually added their related text equivalents in the indicator section.

## 2- Main Chart:

In the main chart I again chose bar chart because it the height of the chart can show the number of loans given to each borrowers with different scores and we can see how popular is a loan type among different Prosper Ratings at the same time the color coded sections of each bar shows the distribution of different loan status. This way the reader easily find out how risky a loan for a particular borrower is by comparing the completed loans (blue colored sections) compare to orange and red sections. The design characteristic of this section is as follows:

a. On the left side of the main svg I set bound for a main bar chart with x axis for the Prosper Rating assigned to borrowers between AA – HR based on several different evaluations done by Prosper organization. The y axis shows the cumulative value of "count" which is the number of each category of loans after aggregation process I did on the original Prosper Loan data. The details are described on R document "ProsperLoanData_Project - Viz_filtered.Rmd" where all cleaned up data were grouped by (ProsperRating,LoanStatus,ListingCategory,ProsperScore,Term) and then summarized for average LenderYield and BorrowerAPR, and finally filtered b specific loan status.
b. Added loan status legend on the top of the main bar chart to show filtered loan status in different colors: "Completed" in blue, "Chargedoff" in red, and "Defaulted" in orange.

## 3- Storyboard Control:

I used storyboard to animate the data visualization with following characteristic:

a. Using the dimple.storyboard for controlling the storyboard with the tick event and other methods.
b. Clicking the indicator chart of Loan Types will select a frame and pause the animation, clicking again will resume.
c. Hovering on the main bar charts for different Prosper Ratings will show additional information for each specific loans like "ProsperScore", "Term", "Lender Yield", "Borrower APR", "Loan Status", "Prosper Rating", and "count" the number that particular loan.
d. Each different colored sections of the bar chart shows the "Loan Status" of each loan. So the bar chart with bigger blue portion shows more successfully completed loans that will be more attractive for investors.

Among several different information that can be extracted from this data visualization, readers can explore the data interactively for reader driven explorations. For example:

a. The length of each Loan Type can show investors and borrowers the popular loan types through which they can see which loans can be invested quickly.

b. The distribution of the Loan Status for each loan type among borrowers with different Prosper Rating shows investors which rate has higher yield with lower risk when it shows more blue colored sections.
c. Borrowers can investigate with their particular Prosper Rating, what type of loans has higher chance to be fulfilled on time.
d. Investors can find their expected returned yield per borrowers rating and also loan types by hovering over each sections of the bar chart.
e. Borrowers can investigate the amount of APR they might end up paying per their rating and the type of loan they are requesting.

# Feedback

I received following feedbacks from three of my colleagues and the forth one from Udacity Reviewer. The final version of the data visualization is based on all above mentioned feedbacks. I have put the initial versions of the README.rmd and index.html files with a sample dataset in public Gist and GitHub but haven't received any feedback yet.

Below are the feedbacks and their follow ups that have received so far:

### Feedback 1 – M. Makarian

Feedback point 1:

- *The first sketch visualization was Writer Oriented and didn't show enough information on the graph.*

Follow-up 1:

- I added more specifications by adding legends to the graph.

### Feedback 2 – J. S PANNU

Feedback point 1:

- *The second sketch was not animated and still was Writer oriented with no interactive functionality for the reader.*

Follow-up 1:

- I used the Storyboard Control with the tick event for the reader to visualize the data properly.

## Feedback 3 – M. BURKE

Feedback point 1:

- *The y axis originally was showing the commutative values for "Lender Yield" which was not informative enough for readers.*

Follow-up 1:

- Based on Mark's feedback, I changed it to "Investors" variable which gave better sense to the investors and borrowers to compare and learn about the loans. The reader can interactively find out about a loan type as if it is popular enough for investing or if it is requested by a borrower with specific rating, it will be fulfilled quickly.

## Feedback 4 – Udacity Reviewer-1

Feedback point 1:

*Since you will be using the data this way, there is no need to use the smaller "prosperLoanData_sample5.csv" file. To conclude, the process is as follows:*

- *Decide on what finding(s) you wish to present to readers (more on this later when we talk about explanatory data visualization).*
- *Process the entire dataset with Python code, extracting the aggregated data to visualize.*
- *Write visualization code with this much smaller aggregate data.*

Feedback point 2:

- *This specification requires that the visualization made focuses on a specific, clear finding in the data, which means it should be explanatory, following the Martini Glass principle. To do this properly, the first thing you need to do is to find an interesting insight that you'd like to present.*
    - *•Which loan types are more profitable for lenders (i.e. most of them completed)?*
    - *•What do people commonly taking loans for? How much?*

- *When you have found a question to answer with this visualization, make sure to present it in a way that leaves no ambiguity to readers about your finding.*
    - *•You may update the title of this visualization to include the finding directly.*
    - *•Or expand the paragraph below it.*
    - *•In addition to either of the above, you may also update the visualization so it tells a narration that leads to a finding. For this, you will need to add some interactivity to the visualization. You'll get to learn much useful data visualization techniques by doing this.*

Feedback point 3:

- *Word document is not a universal format for readme, I strongly suggest to use .md or .pdf format instead.*
- *Include the main finding in Summary section*

Feedback point 4:

- *Note that for this section, we need to see the reasoning behind design decisions made in the visualization, rather than just a statement of what they did.*

Feedback point 5:

- *To pass this specification, we need to see that feedback has been collected from at least three people. Currently, all the three points could have come from the same person. If these feedback pieces were indeed coming from three different people, please properly layout them to show this, for example, something like (of course, this isn't the only format to use.*

**Follow-up: Explains the correction made for feedback 4**

## Follow-up point 1:

I used R to wrangle the dataset by selecting few number of parameters and running exploratory analysis on them to finalize the selected variables for visualization. The data wrangling process details can be found in “ProsperLoanData_Project - Viz_filtered.rmd” file.

- I decided to present the risk of investment to the investors and chance of quick approval of loans for borrowers.
- I aggregated the Prosper Loan dataset significantly from 113398 records to 1749. I removed records with “na” or blank data and also outliers first, then I grouped them by the selected variables and summarized them based on the average values of “LenderYield” and “BorrowerAPR” for each group.
- Finally I filters all the records to only contain the Loan Status that we are interested in for our analysis like “completed”, “Chargedoff”, and "Defaulted" that can tell us about the risks and benefits of the loans.
- The new final visualization code, index_final.html code is running very quickly for the final selected variables. The previously submitted code is now called index4.html in the zip file.

## Follow-up point 2:

- The current animated visualization is following the Martini Glass principle by showing the number of loans per Borrowers rating, ProsperRating, for each type of loans, ListingCategory, and color coded for the status of Loans. My focus was on finding which loan characteristics are safe and beneficial for investors. The data visualization clearly shows that "Motorcycle” & “Engagement Ring” loan types have lowest ratio of failed loans. The chart also shows that "Debt Consolidation" loans are the most popular loans among investors and borrowers with about 12000 loans that majority of them were successfully completed. So there is a very high chance for borrowers with different scores to request that loan and have it invested quickly instead of waiting for long time to receive their loans.
    - By clicking on each button of the Listing Category indicator chart, the animation will be paused and the reader can visually recognize how risky is that type of loan among borrowers with different ratings by comparing the ratio of the color distribution to color blue for the completed loans. At the other hand borrowers with specific rating like “B” can see which type of the loans has high chance of getting the loan quickly by looking at the number of loans in that category.

## Follow-up point 3:

- The new Readme document is in pdf format. I have also added the rmd format of this file on the Github: https://github.com/Manonuro/Data_Visualization and Gist: https://gist.github.com/Manonuro
- The Summary section of the Readme document has detailed information of the findings.

### Follow-up point 4:

- The design section explained all the reasoning behind design decisions.

### Follow-up point 5:

- The suggested format for feedback and Follow up is used in the README document.

## Feedback 5 – Udacity Reviewer-3

### Feedback point 1:

*The visualization needs to be explanatory*

- *One way to highlight a finding is by including some introductory text before the graphic that explains the data and what you are trying to show (similar to the README but more concise).*
- *The second thing that you can do is provide a descriptive title that gives some idea of the finding ('Prosper loans increasing and defaults down' for example).*
- *Last, you can take advantage of preattentive visual cues and use encodings in a way that grab the viewers attention most effectively (http://flowingdata.com/2010/03/20/graphical-perception-learn-the-fundamentals-first/).*

### Follow-up point 1:

- I added following descriptive title to the chart:
  "COMPLETED LOAN RATIO"

"Motorcycle & Engagement Ring loan types have lowest risk of investment","(The Bigger Blue portion of bars indicates the lower risk of investment in each category of loans)",

Feedback point 2:

- *Because a clear explanatory finding couldn't yet be identified, this specification 'requires changes' for the time being.*
- *Two important things that I would add to the design (as mentioned above) are a title and text. The other thing that I would add to the design is a label for loan status near the color legend. A person without any familiarity to prosper loan might have a hard time understanding what the color encoding represents.*

Follow-up point 2:

- Added a label for loan status near the color legend.

Feedback point 3:

- *There is a lot of exploratory language used in the summary and design sections.*

Follow-up point 3:

- Modified the exploratory language used in the summary, design and Follow-up sections by focusing more on the findings.

Feedback point 4:

- *The design section does well to mention the various aspects of design but it doesn't talk about why those design decisions were made. Design decisions should revolve around making the visualization as explanatory as possible. So what about those design decisions would help in easily communicate the finding to a viewer? This sort of 'requires changes' for the time being because an explanatory visualization has not quite yet been created. When one is created, be sure to talk about how those design decisions make the finding 'stand out'.*

Follow-up point 4:

- The design decisions included in the design section as suggested through following the suggestion on feedback-1 and also explanatory part of the design section.

Feedback point 5:

- *Great job documenting design changes in response to feedback. For this specification, it's also required that 1 or more **previous versions** of the visualization be submitted. For the next submission you can use this version as evidence of a previous version, or if you'd like, you can submit a version that was created before this submission. Sketches or image files can also serve as evidence of a previous version.*

Follow-up point 5:

- This is an unfortunate that a project is not being reviewed by the same reviewer. Since there is no standard and specific answer to each project, the reviews can easily be subjective based on reviewer's background or area of interest.
  Since my previous submission with several csv and html files was confusing for the reviewer and he checked the wrong files even after I was mentioning the file specifications on my note to the reviewer and also through a separate email. I decided to just send the final versions of those files to avoid further confusion. I have also mentioned this in my note to the latest submission if check it again.
  I will put all these files again in one zip file and I hope you would be the next reviewer in order to be in the same page through this review.

**Feedback 6 – Udacity Reviewer-4**

Feedback point 1:

- *Next time would you mind knitting any RMD files into html as it makes it much easier to look at and review?*

Follow-up point 1:

- I knit the rmd file to ProsperLoanData_Project_-_Viz_filtered.html file and added to submission.

Feedback point 2:

- *the OnClick event - after a while I worked out that it is functioning as it should but the graphic still crashed on me several times. Why? A combination of the following:*

    - *The dimple chart takes a while to reload - how about the earlier reviewer suggestion to go back and aggregate the data further in Python (ie. Prosper Rating, Loan Status, Listing Category, Category Loan Count etc...) You'd have a considerably smaller file and therefore much quicker loading time. It is also worth asking yourself "Does this feature actually contribute to my data story?" You might not need so many features either?*
    - *It is not clear whether or not the animation has been paused*
    - *The onClick event is triggered when you click the number, not the text.*

Follow-up point 2:

- I aggregated the dataset further and reduced the number of records to 533 by removing "ProsperScore" from grouped by data.
- I changed the order of typing the text value and numerical value of the "ListingCategory" which made it possible to pause onClick event by clicking on the text instead of the number.

Feedback point 3:

- *Please could you add a brief title and description? It doesn't have to be more than a sentence or two. It would be nice to mention both the borrower and investor perspective and perhaps give a specific example comparing one loan type/prosper rating to another. This would be a good thing to add right at the start as it would give the user something to focus on while the dimple chart loads. If you decide to add a progress bar/loader during the onClick event you could incorporate it here too.*

Follow-up point 3:

- I added following note on the indicator section to show the Borrowers take from the chart:
  (The length of the bar charts in the indicator section corresponds to the likelihood of loan type approval.)

And added following note on the title of the main chart for Investors: Motorcycle & Engagement Ring loan types have lowest risk of investment (The Bigger Blue portion of bars indicates the lower risk of investment in each category of loans)

## Feedback point 4:

- *Bar Chart labels - "Loan Count" and "Prosper Rating Score" would be clearer.*
- *Legend - ideally this would be closer to the bar chart with a title and possibly the same font size as the loan type so it's more visible*
- *Could you make the SVG a little smaller so it fits on the page?*
- *Finally, it would make all the difference if you could sort the data by Loan Status after filtering it by Prosper Rating - this would be solved if you aggregated the data! In terms of pre-processing it enable the user to get an instant grasp of what % are Completed, Charged Off or Defaulted in each category.*

## Follow-up point 4:

- Bar Chart labels are changed to- "Loan Count" and "Prosper Rating Score".
- Legend - A title with the same font size as the loan type is added.
- The SVG is smaller now it fits on the page?
- I sorted the data by Loan Status after filtering it by Prosper Rating – But still the bar charts are not completely sorted by color.

**Gist**:

https://gist.github.com/Manonuro/4543fcfc35981d42f5db70382760bfe2

Github:

https://github.com/Manonuro/Data_Visualization

**Resources**

Dimple:
http://dimplejs.org/

Margin Convention:
http://bl.ocks.org/mbostock/3019563

Storyboard Control:
http://dimplejs.org/advanced_examples_viewer.html?id=advanced_storyboard_control

dimple.chart:
https://github.com/PMSI-AlignAlytics/dimple/wiki/dimple.chart#methods-1

Create Data Visualizations in JavaScript using Dimple and D3:
https://www.sitepoint.com/create-data-visualizations-javascript-dimple-d3/