Summary

I have analyzed the Prosper loan data set. It contains 113,937 loans data with 81 variables on each loan. The dataset can be found at the below link:

https://s3.amazonaws.com/udacity-hosted-downloads/ud651/prosperLoanData.csv.

The data dictionary explaining the each variable usage can be accessed through the below link:

https://docs.google.com/spreadsheets/d/1gDyi_L4UvIrLTEC6Wri5nbaMmkGmLQBk-Yx3z0XDEtI/edit#gid=0

In this tableau story, I analyzed correlation between different variables and tried to connect them with a story.

Links to the story

Initial story link:
https://public.tableau.com/profile/sonal.garg#!/vizhome/PROJECTXX/ExplorationofProsperLoanData?publish=yes

Final story link:
https://public.tableau.com/profile/sonal.garg#!/vizhome/PROJECT-LoanData/ExplorationofProsperLoanData?publish=yes

Design

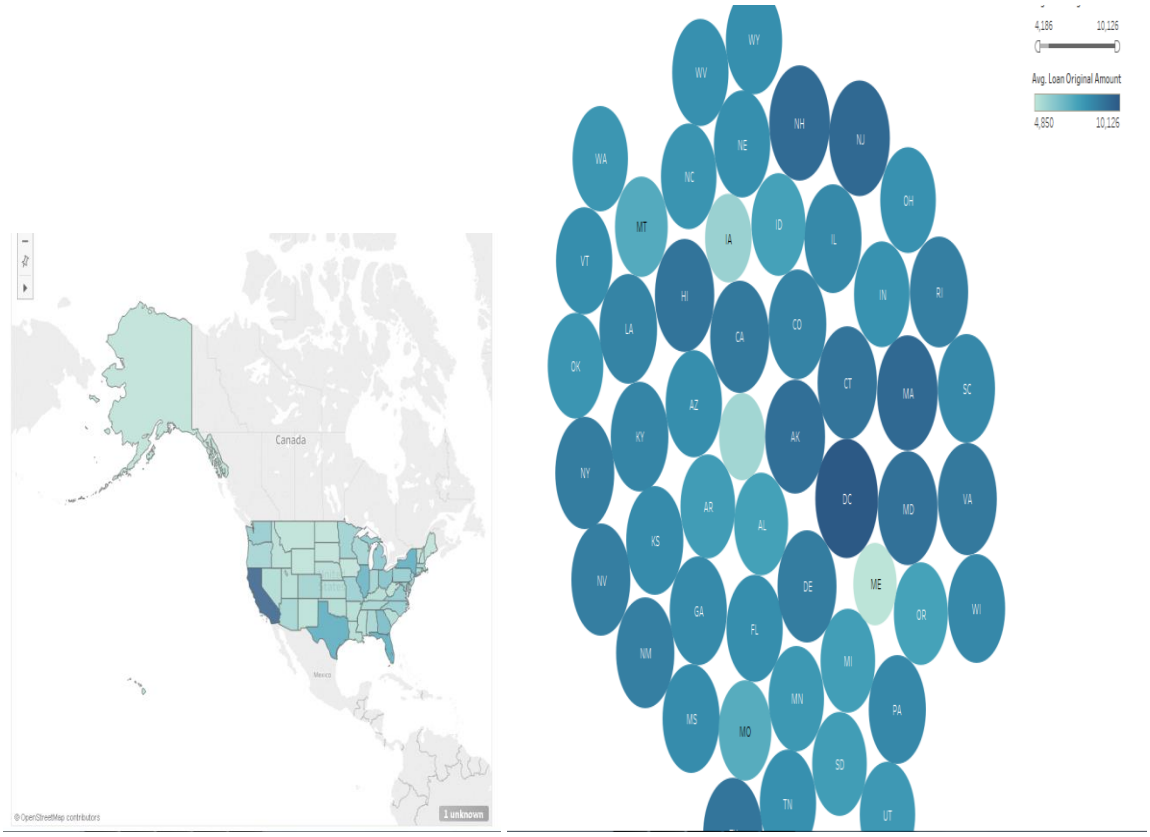
Initially, I put in some time to understand each variable and made many exploratory plots. This made me understand the uses of various variables and how they could be correlated. I also changed the data type for listing category and prosper score to string variable.

I tried to ask different questions in the dataset. Some of them were like:

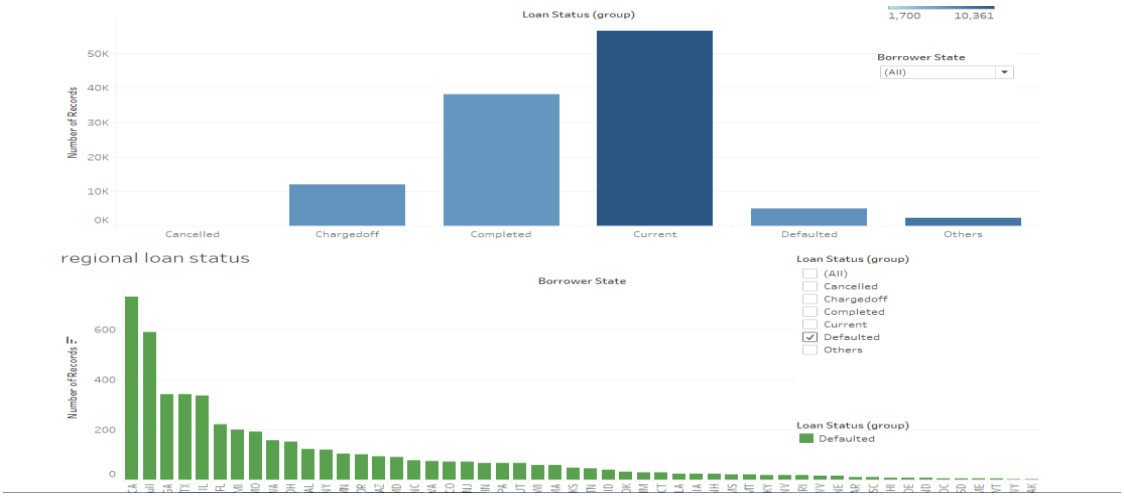
- How trends have changed over time for loan taken and amount of loan taken
- Loan status over time
- Loans compared to their return and losses
- What is the region of most number of loans

In the process of finding answers to such similar questions, I came up with the first draft of the story. Below is the story points explained.

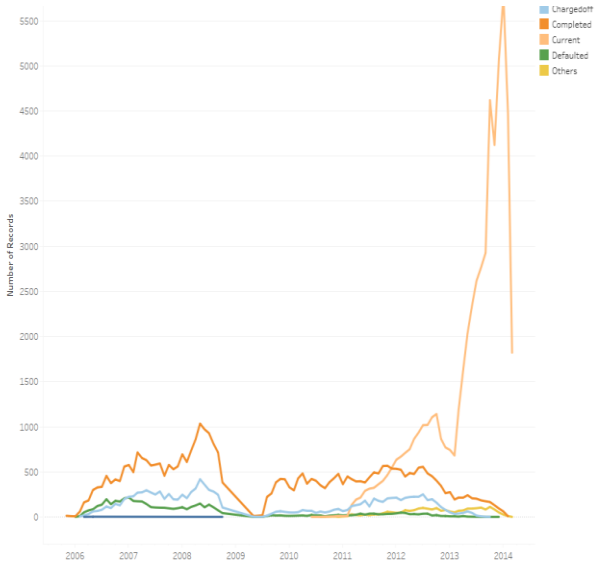
A map exploring the loan taking patterns was used. Then a packed bubble chart was used to depict average amount of loans taken in different regions. This helped define regional characteristics clearly.

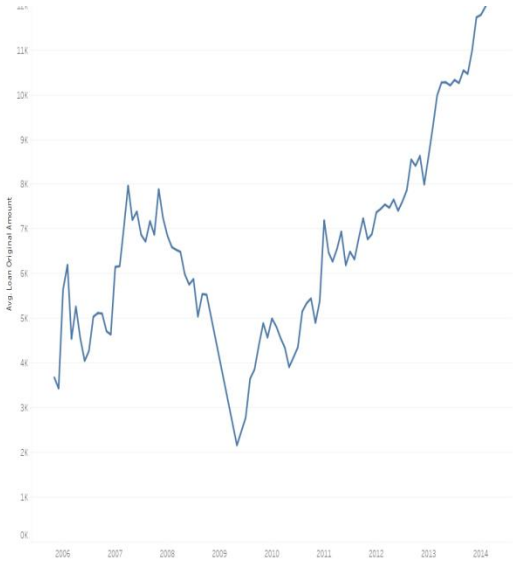


After this I tried to correlate the loan status in various regions along with the loan status in general over the years. Regional wise, I focused on which state was the most defaulted region. For this purpose, bar charts were used. They present the clearest depiction for categorical variables.

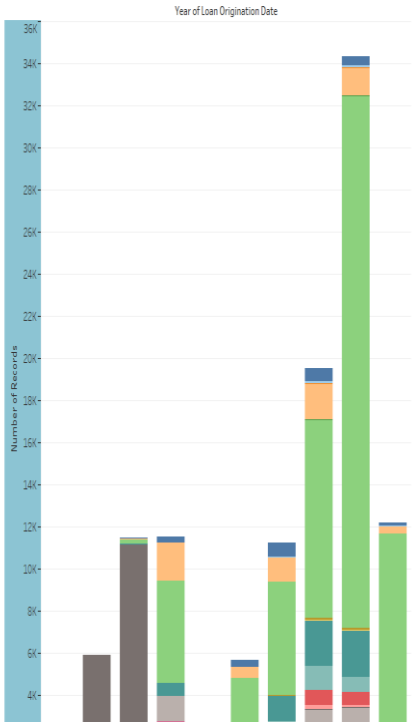


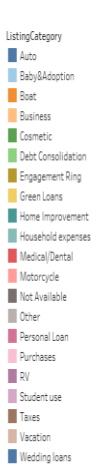
Then, I formed two charts, one depicting number of loans and average amount of loans over the years. For these trend observing plots, line charts are the most useful.

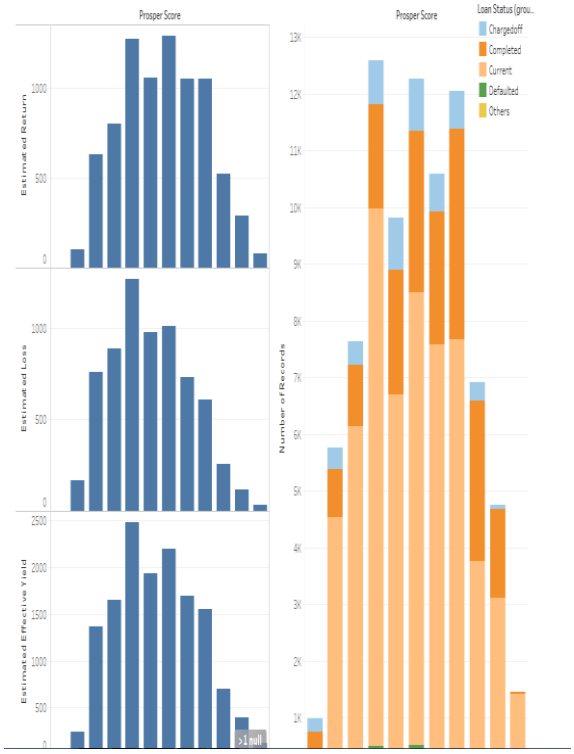




A bar chart depicting listing category and their popularity over the years was formed. Finally, a dashboard was created. This had 4 plots, each correlating prosper score with returns and losses as well as loan status.



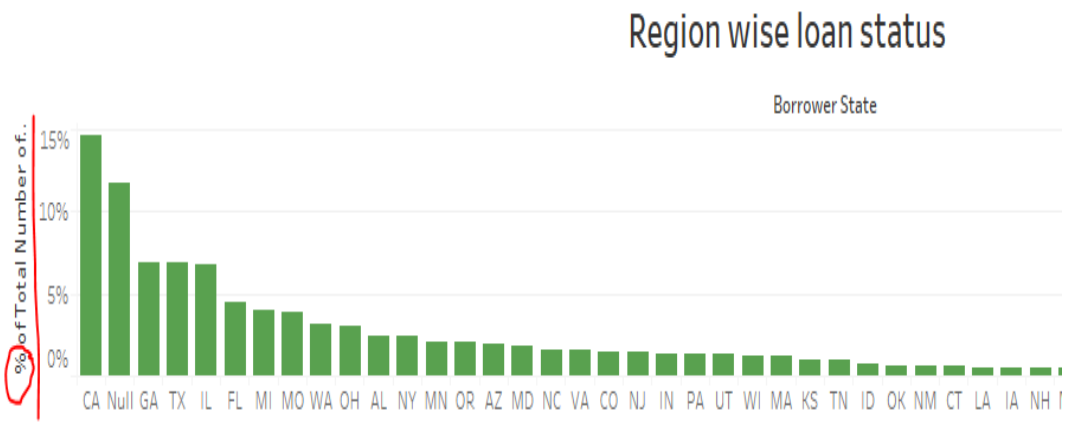


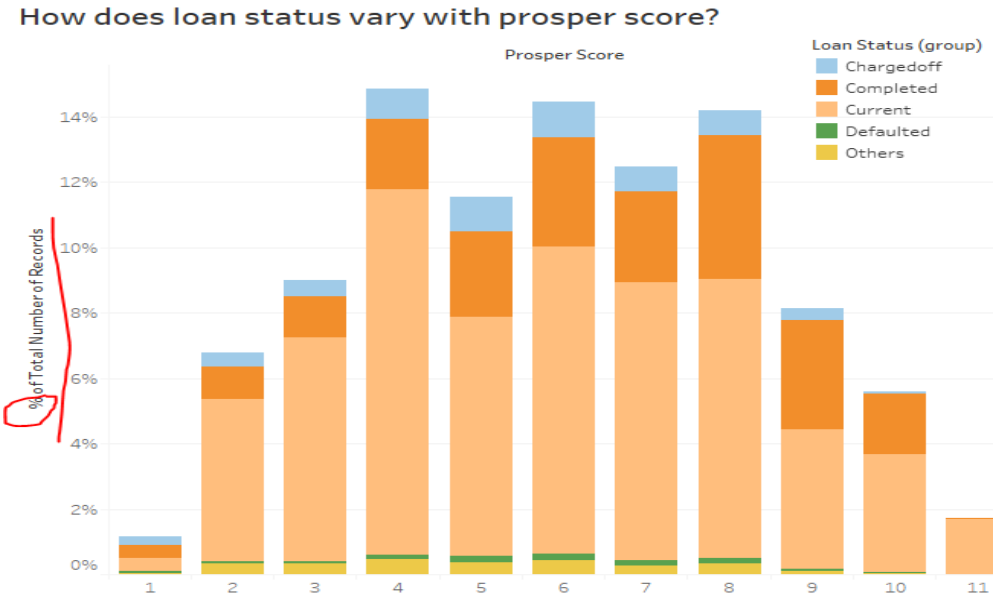


Changes made after feedback-1

After the feedback I made certain changes. They are explained below:

1. I changed the y axis of two plots showing % of total rather than just than just the number of records. This helped in comparing the data for individual region as compared to other regions and loan statuses for an individual prosper score relative to loan status for other prosper scores.

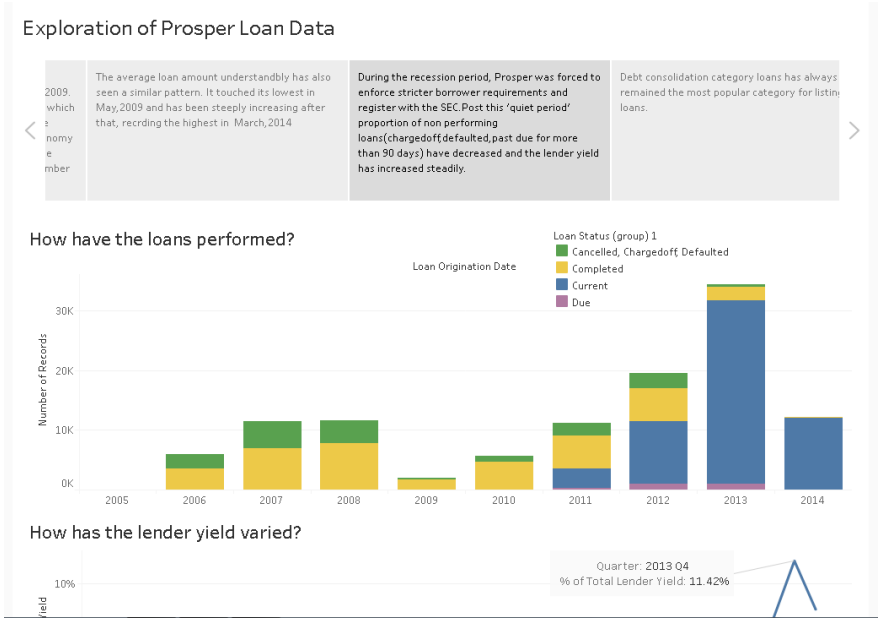




2. Appropriate titles added to plots in the dashboard.
3. In the 2nd plot, I highlighted the top 3 regions with highest average loan amount. Since, the plot is a little clustered and the color difference between each circle is not that evident. Making this change made more sense.
4. I added a plot each depicting income range and occupation of borrowers.

Changes made after feedback-2(Review)

1. I did some research which helped me explain the findings beyond what the dataset tells us. Accordingly, I have edited the text in some of the story points and added one more new story point.



2. Added some annotations for easier understanding of the visual.

3. Resized the captions so that no scrolling is required to read the text.

4. Added titles to all the charts in the story.

Feedback-1

- What do you notice in the visualization?

Usage of variety of charts and visualization techniques in representing data

- What questions do you have about the data?

Do we have insight on what the borrowers do or how much they earn vs. there loan amount

- What relationships do you notice?

Geographical distribution of loan among various US states, Distribution of loan status, Correlation between risk and estimated margin.

- What do you think is the main take away from this visualization?

Loan trends over the years

- Is there something you don't understand in the graphic?

- Yes, there is no data diagram depicting risk score vs. loan amount.

- There were no plot titles.

- Any other relevant point to be noted down for further improvement of the story?

-% calculation should be done.

-In the second plot, highlighting the relevant areas would help.

Feedback-2(Review comments)

- *See if you can go back and explain in addition to describe some of the time series trends you uncovered. Why was there the noticeable dip in loans at the start of 2009? Why has there been such a noticeable spike in lending in the most recent years in the dataset? Answering these questions may require that you do a bit of research beyond the dataset. Also, Tableau annotations might be useful in communicating what you find. They combine very well with line charts.*
- *Resize the caption boxes to eliminate the need for the end user to have to scroll through the text. Click and drag on the bottom or a side margin to resize the captions.*
- *At least one of your charts is still without a title. You may have noticed that Tableau deactivates the titles of charts that are added to a story point as a single worksheet. To get around this, try first placing the single chart onto a dashboard, then into the story point.*

Resources

- https://www.tableau.com/learn/training
- https://en.wikipedia.org/wiki/Prosper_Marketplace
- https://onlinehelp.tableau.com/current/pro/desktop/enus/annotations_annotations_add.htm
- https://www.lendacademy.com/a-look-back-at-the-lending-club-and-prosper-quiet-periods/
- https://www.lendacademy.com/prosper-review/