I visualised Prosper Loan data as a part of my final project in Udacity DAND Nanodegree. Prosper is a P2P lending platform.

The following are the important URLs:

1. The data used in this analysis can be downloaded here..
2. My analysis can be found here
3. My initial analysis can be found here

Probes

- What is the growth rate of the company?
- Do loans differ across the different US states
- What is the distribution of loan amount?
- How do loans taken differ across individuals in different income groups?
- What is the loan principal amounts by prosper rating?
- How much do they spend monthly towards the loan and how the return from loans differ across these income groups?
- Which income group is the largest borrower? Also, a follow up question might be 'which income group pays the highest amount monthly'?

Summary

- The growth rate of the company is pretty impressive over the years. There are two periods when growth was negative. One was during the 2008-2009 financial crisis which was expected. The other was in 2013-2014 which caught me by surprise.
- Amount in the 4k range is given out the most. There are peaks in other amounts as well, including 10k, 15, 20k and 25k. This might be because there is a natural bias towards multiples of 5s when arriving at a figure.
- Prosper rating is highest for IL.
- Most number of loans are given out in January and the least number in April.
- Loans are mostly given to employed individuals and full time individuals.
- Individuals earning between $25k and $50k constitute the largest group of borrowers. Those with over $100k income borrow more. Naturally, this groups spends most on their loans monthly. The estimated returns is greater for lower income groups.

- Most loans originated in California. This might be due to the fact that Prosper is headquartered in CA. Furthermore, people from CA has the highest credit rating only next to average rating of IL

Design

- Line graph was chosen to visually depict the growth rate of the company both in terms of the number of customers(borrowers) as well as loan amount disbursed. Buth shows an increasing trend. However, there's a huge dip in 2008-2009, during the economic slowdown as well as in 2013-2014
- A map was chosen to visually depict the loan amount difference across different states of US. Loan amount was colored(gradient)
- My analysis mainly consists of bar charts. Bar chart is very efficient in visually summarizing a large dataset. I used size to depict the number of records in each loan (bin) amount.

Feedback

My friend Aadim Makarand who is a data scientist went through my work:

- In my initial visualisations, I had tried to incorporate many dimensions into one plot, which resulted in the whole presentation being messy. My colleague asked me to make a few corrections wrt this.
- My friend also suggested that I should add total loan amount in my growth visualisation which I incorporated.
- I also added more filters as suggested by my friend.

Resources

N/A