Alissa McBain
Create a Tableau Story

Summary:

I used the Prosper dataset, because I find finance incredibly interesting, and my husband and I are investors in Prosper. I poked around the data, and found several interesting things, which I made visualizations of, four in total. They tell a story about the loan data from a geographical standpoint, and also give insight into the details and complexities of borrower's financial picture and subsequent loan performance.

Design:

I made countless design choices and incorporated as much of the advice given to me from my feedback as possible. As often as possible, I tried to make my visualizations clear, interesting and informative. Detailed feedback is listed below. In general, I came away with a few good takeaways. First, Prosper loans are most common in the state of California, the best state to make a loan in is Nebraska and second place is tied between Alaska and Wyoming, while the state with the most bad loans is Maine. Secondly, judges, doctors and pharmacists received the lowest average APR's, while teacher's aides, nurse's aides and college freshman received the highest average APR's. And finally, the higher the loan amount, the more likely it was to go into NIGO (Not In Good Order) status. Loans above $11k were more likely to go into NIGO status, and specifically, for those borrowers who reported no income, loans above $10k were more likely to go into NIGO status.

For my first visual, I really wanted to see how the loans were spread out geographically, and the best way to illustrate that was using the map function. What I really wanted to do, but couldn't figure out how, was to show (and color) the proportion of good loans vs bad loans per state. Say, if California had 30% of its loans NIGO (Not In Good Order), it would be darker than say, Texas, who had only 10% of its loans NIGO (these are just numbers I'm making up to illustrate the problem). Since I couldn't figure out how to represent that in Tableau, I decided to add lists for Top 10 and Bottom 10 states instead. The map is shaded so that the darker a state is, the more loans (no matter what the status) Prosper made in that state. California is the darkest, with 13.57% of the total loans. This makes me wonder if Prosper piloted in California, or had an aggressive marketing campaign in California, because it seems disproportionately high.

For my next visualization, I wanted to illustrate how a borrower's occupation played into the details of the loan they received. I started with putting the occupation on the x axis, then switched it to the y axis for better readability. On the opposite axis I graphed the average APR for those occupations. To add interest, and look for anomalies, I shaded the graph by the average credit score for each occupation. One would assume that credit score and APR are negatively correlated (a higher credit score would yield a lower APR, and a lower credit score would yield a higher APR), so these anomalies were pretty clear after I colored the graph. For example, borrowers with the occupation "Investor" had relatively high credit score, but also received a higher than expected APR. On the flip side, some groups of students, "College Graduate", "College Senior", "College Junior" and "Technical School" students, had lower credit scores but also received lower APRs. Perhaps there is something in Prosper's risk algorithm that lowers the risk for students?

In the third visualization I wanted to explore what effect being a homeowner had on a borrower's creditworthiness (according to Prosper). I plotted the average Monthly Income vs the average Debt to Income (which I know are correlated from a previous project using the same data set), for each occupation, and then split the data points one more time on whether or not the borrow was a homeowner. This last split I have represented in green (homeowner) and red (non-homeowner). In earlier versions of the story I had the size of the dot be representative of the group's average APR, but based upon feedback, removed that aspect because it was covered in the previous graph. There were a couple interesting observations that come to light in this graph. First, I noticed that the majority of the green dots (homeowners) lie above the red dots (on the average Debt to Income axis). This makes sense to me and could be explained by the homeowner's mortgage increasing their debt. Also interesting to me were the two largest Debt to Income outliers, both of which were Homemakers. Given that their average Stated Monthly Income was above average for the whole data set, why do they have such high average Debt to Income? I'm not sure. Lastly, there is a cluster of red (non-homeowner) points between 0.5 and 1.2 Debt to Income. These were all students, and I believe their higher than average debt may be attributed to student loans.

And for my last graphic I wanted to explore how the loans for various borrower demographics performed, and if there were any significant outliers to help predict whether or not a loan would go past due. My aim was to break up the data enough to show patterns, but not too much to be overwhelming. I first tried scatter plotting every data point, but it was so overplotted that any patterns were lost. So, I segmented the data by Credit Grade, Income Range (by color) and NIGO/IGO (Not In Good Order/In Good Order, defined above), represented by X's and O's. I then plotted the points on their average Credit Score (x-axis) vs their average Loan Amount (y-axis). The graph shows seven distinct vertical trends, corresponding with each Credit Grade (AA, A, B, C, D, E and HR). In each of these vertical trends, the lower incomes are at the bottom (lower Loan Amount) and the higher incomes are closer to the top, shown by the colors, red through green. In nearly each of these vertical trends, for each color, the X is above the O, meaning that the average Loan Amount, for a given Credit Range and Income, is higher for those loans that are NIGO, and the average Loan Amount is lower for those loans IGO. In short, if Credit Grade and Income Range are held constant, a higher Loan Amount is more likely to go into NIGO status, than a lower Loan Amount. Another interesting finding that came to light is that for nearly all Credit Grades and Incomes, if the average Loan Amount is greater than $11,000, it was in NIGO status. The only exception to this was the group of loans where the Credit Grade was A and the Income Range was greater than $100,000. These could be good starting points to begin to develop a credit risk model.

Feedback: see below.

Resources: none.

Links:
Version 1:
https://public.tableau.com/profile/alissa.mcbain#!/vizhome/ProsperDatav1/ProsperDatastory

Version 2:
https://public.tableau.com/profile/alissa.mcbain#!/vizhome/ProsperDatav2/ProsperDatastory

Version 3 (final):
https://public.tableau.com/profile/alissa.mcbain#s!/vizhome/ProsperDatav3final/ProsperDatastory

Feedback on version 1 from Graham McBain (husband and software developer):

- Viz 1: Loan Status by State
    - Information would be more useful if it was % of loans by state, instead of raw count, California leads the count, but the visualization doesn't show how many of those loans in each category on an even playing field as the other states
    - Would be helpful to see a list of the top 5 and bottom 5 states to make a loan in, ie, 5 states with the most current and completed loans, and 5 states with the most NIGO loans in
- Viz 2: APR and Credit Score by Occupation
    - Change colors, too hard to tell the differences with the range of blues
    - Remove Null occupation
    - Is there a better graph to represent this data, other than a bar chart?
    - Limit y axis to better represent the spread of data
    - Interesting observation: women dominated fields on the low end, male dominated fields on the high end
- Viz 3: Debt to Income vs Income by Occupation
    - Nothing to add
    - Notice the students cluster, and the homemaker anomalies
    - Green bubbles tend to have higher debt to income ration than red bubbles, due to weight of mortgage on debt to income ratio
- Viz 4: Loan status, credit score and loan amount
    - Least interesting graphic; can we add more interest?
    - Is there a loan amount that performs the best? Say, if someone asks for $2500, Prosper can be sure that it will be paid back

Feedback on version 2 from Megan Slape (sister):

- Viz 1: Loan Status by State
  - Define what makes “IGO” and “NIGO”
  - Note: CA with the most loans, is neither best nor worst
  - “Worst Loans by State” is on two lines, making both tables uneven with each other
- Viz 2: APR and Credit Score by Occupation
  - Occupations at an angle to make them easier to read
  - Switch rows and columns
  - Axis on top too
- Viz 3: Debt to Income vs Income by Occupation
  - In legend, make T/F into Homeowner/nonhomeowner
  - Size of bubble (APR) was covered in the previous graph, so doesn’t add anything new
  - Dots not visually appealing, prefer to look at lines
- Viz 4: Loan status, credit score and loan amount
  - Take grey dots out of the graph
  - Make “not employed” blue, and move to the top of the legend
  - Most helpful, informative graph
  - Low income people take out smaller loans, and any loans about $10k for low income people, defaulted
  - The higher loan amount, the more likely to default, no matter income or credit score
  - Almost all loans over $12k defaulted
  - For all incomes, as their credit score went up, the amount they borrowed went up