Tableau Story Write-UP of Prosper Loan

January 30, 2019

1 Tableau Story Write-UP of Prosper Loan Business

1.0.1 Introduction

The story was developed from a dataset about Prosper loan, containing over 100k observations with 81 variables spanning across 10 years. In this storty, I have done Tableau visualizations to explore the overall Prosper loan business, why people borrow money from Prosper, what are key features indicating higher default risk. Hopefully this exploration would help investors to make descisions when they invest their money.

1.0.2 Summary

This story is an extension of the EDA project I did before. In this Tableau story, I looked at the key features that could indicate if a borrower has higher risk of default, such as credit score, interest rate, employment status, monthly income, geographic region, etc. Discoveries include: People borrow money from Prosper for the main reason of debt consolidation; the loan types with highers default risk are student loan, personal loan, and business loan; borrowers with higher risk of default include retired people, students, monthly income <5k, and credit score <600; people also tend to have higher default risk at higher interest rate.

1.0.3 Design

the Purpose of this story is to dig out the reasons behind loan default, and give some suggestions investors about if someone is reliable to lend money. So the structure of the story come up with the following parts.

Questions to ask

- What are the borrowers' loan status?
- Why people borrow money from Prosper?
- What are the key features could indicate if a borrower will default or not?
- How those features related to the loan status?
- If you are an Prosper investor, what would you do?

Process

1. Data Wrangling of the original data set using R, including cleaning missing values, removing irrelevant columns, creating new columns for analysis. A final dataset was created for Tableau story telling.

2. page1-An overview of the Prosper loan business.
3. Page2-What are the borrowers' loan status?
4. Page3-Why people borrowe money from Prosper?
5. Page4-Why do they default 1 --Are they employed? How much money do they make each month?
6. Page5-Why do they default 2 --What's their occupation?
7. Page6-Why do they default 3 --Can the credit scores tell if a borrower will default or not?
8. Page7 Why do they default 4 --Do the living regions have an impact on default?
9. Page8-Why do they default 5 --does interest rate have an impact on default?

Overall design decisions

- story captions were used to orient the readers navigation
- dashboard captions were used to uncover the findings and conclusions of the charts
- the color and sizing of different types headings wre kept consistant
- all the titles, axises of charts were labled properly for readers to easily understand the chart information
- observations were ordered by descending so the readers could clearly see the patterns and trends
- filters or legends were located in proper loacations for visulizations
- the overall design is to keep concise so readers could directly understand the purpose of each chart and how the story is connected

Page 1

- a time series plot could help readers easily understand the business growth of the Prosper loan
- an histogram of loan distribution having the highest loan amount marked could directly tell the readers the Prosper loans are small loans, so you could assume borrowers borrow money for personal use. this help us to dig further for the later story telling

Page 2

- loan status is the target variable, a bar chart could give readers direct sneak peek about how many people might default

Page 3

- by parallel comparison between the counts of loan purpose and loan default risk, readers might easily understand why did people borrow money from Prosper
- by sorting the default risk, the trend could be clearly seen

Page 4 & 5 & 6

- the purpose of these three dashboards is to look at borrowers' different status, including employment status, monthly income, occupation, and credit score, try to find how those features are related to default.

- there are 81 variables for this data set. The correlated features with default are not limit to the above chosen features. They just give an general idea about the borrowers status ##### Page 7
- this dashboard is to tell one of the causes that could lead to borrowers to default from the investor side, higher default rate is corresponding to higher default risk. But it is hard to conclude higher interest rate is a trigger to default because investor could charge higher interest rate when they see the borrower has low credit score. Further study is needed to make a solid conclusion

1.0.4 Feedback

Initial Feedback

User1 "Please communicate your findings to the reader by placing it as captions. You can place 2 plots per dashboard if they are connected in some way else you use one plot per dashboard."

User2

1. in the very first chart, I would rename the y-axis - it wasn't clear to me what it measures.
2. in the second slide, by default the charts show only the defaulted category. however, the first chart means to show what is the most popular use for all loans, so i would not even give the option for interactivity, since that would change the meaning of the chart (the title of the chart and the chart itself will not correspond to each other anymore)
3. same for the second chart, where if you preserve the interactive character of the chart you allow the user to uncheck defaulted loans and then check some other category. again, the connection with the title of the chart would be lost
4. i think if you want to preserve the interactive character of the charts you need to give them a more general title ("How loan use varies with loan status" for example - for the first graph). you could even have first that general interactive graph, and then other non-interactive graphs that show your own conclusions.
5. third slide: the interactive tab is aligned with the first chart, but it modifies the second one. not sure why....
6. fifth slide: I would change causation to correlation - in this case in particular the causation might be going the other way: it might be that because people have a lower credit score, the bank is offering them a higher interest rate since they are at greater risk to default. if you just posit a correlation claim, you won't have to answer this objection
7. btw, in the last comment i was referring to the title of the first slide

Changes made

1. Changed the captions for eache sheet with my findings to make it communicate with readers
2. Remade most of the plots to make the story more related my goal--finding the reason why people would default, but not limited to the features I listed in the Tableau story
3. set the default ratio as target variables, by grouping the loan status 'Defaulted', 'Chargedoff', and all loans past due over 60 days, then calculate the default rate. The reason why I define these loan status as default loans could refer to the Delinquent loans and the default risk prediction project I did previously.
4. rearranged the sheets to make the story more clear and connective

1.0.5 My Tableau Story of Prosper Loan

- Here is the initial version of my Tableau story Proser loan (Version 1)
- Here is the final version of my Tableau story Proser loan (version after feedback)

1.0.6 Resources

- https://community.tableau.com/ideas/7517
- https://onlinehelp.tableau.com/current/pro/desktop/en-us/buildexamples_histogram.htm
- https://community.tableau.com/thread/120387