

Summary

The sample dataset i choose to analyse here is Loan Data From Prosper. This data set contains 113,937 loans with 81 variables on each loan, including loan amount, borrower rate (or interest rate), current loan status, borrower income, and many others.

Design

Initial Design -

I started my tableau story with time series analysis to identify the performance of Prosper loans over years. Following this i wanted to see how the loans were distributed across different state over the years. Now that we got the distribution of loans i wanted to explore and see what are the status of these loans. Next step was to explore why do people borrow loans and what purpose do they take loans. Following which i wanted to see if they are able to repay the loan on time. Over the years the number of amount of delinquent loans has reduced .

Initial Story -

<https://public.tableau.com/profile/kavitha.gopalakrishnan#!/vizhome/ProsperDataAnalysis-V1/ProsperDataAnalysis?publish=yes>

Changes Based on feedback is shown in a different font color in the document.

Feedback Received

About the figure "Loan Distribution over years"

This figure shows the number of loans originated on a yearly basis from 2005 to 2014. The data shows a low value in 2009 and the high value in 2013. While 2009 corresponds to the financial crises I wonder why the peak in 2013 ? Next, the value for 2005 is very less which makes the bar invisible unless the mouse hovers upon. Does it make sense to a different scale such a log ?

Modified the graph from bar to line and added explanation regarding the peaks.

- About the figure "Loan Distribution across each Quarter"

This figure continues from the previous figure and shows the loan originations on a quarterly basis. While this visualization is good, I am not sure why the figure collates all the quarters

together in the x-axis ? for example, why all Q1 is clustered followed by Q2. Would it make sense to show them in chronological order ?

Used the correct variable to represent the quarter and also modified the graph accordingly.

- About the distribution of loan over the years in each state

This figure shows the loan distribution on a state basis over the years. Initially it was not clear what was the effect of choosing multiple year check boxes. I was expecting some sort of animation that shows the variance like rate of growth or decline in the selected years in each state. I then figured out that it was a cumulative number. Perhaps this can be made clear in the description.

Changed the graph to represent the loan number in size.

- About the figure "Loan Origination Date Vs Count of Loan Status"

This figure shows very good visualization on the status of loans across the years. However, I have some questions. When I selected only "current", I see the data only from 2011 to 2014. Does this mean that no loans were current in the previous years ? Is this right ?

Added a few line to explain that graph only show the data available for the years based on loan status.

- About the figures "who borrows these loans", "why do people take loans?"

This visualization is good. However, there is no year information provided. Did the original dataset contain this information ? if it did, would it be useful to see a trend of changing reasons across the years ? Also is the reason the same across different states ? or is there any difference in patterns ?

Added a year filter to analyse the graph between different years.

- About the figure "which type of loan has the most delinquency"

This figure is very informative especially adding the state information. I have the same feedback as with previous figure... was there year information included in this data ? will it be useful to see the yearly pattern ?

I did not want to add the year component to this graph as i only wanted to discuss about the delinquencies in different states.

- About the figure "Has the delinquency improved over years"

The title of the figure does not match with what is being plotted. The title led me to assume that the figure is about number of delinquencies over the years... however, upon closer inspection I noticed that the y-axis is the average amount. I suggest the title to be updated. Next, I wonder if the bar chart is the right form of visualization for this data ? since no insight can be derived. I wonder if you see the delinquency across different categories, state using another form of visualization, perhaps the information can be more easily consumable.

[Modified the visualization to better represent the delinquency improvement over years.](#)

The Final updated story

<https://public.tableau.com/profile/kavitha.gopalakrishnan#!/vizhome/ProsperDataAnalysis/ProsperDataAnalysis?publish=yes>

Resources

<https://stackoverflow.com/>

<https://public.tableau.com/en-us/s/resources>

<https://github.com/jubins/Tableau-Projects>