

Create a Tableau Story

Larry Henry Jr.

July 6, 2021

Table of Contents

- [Links](#)
- [Summary](#)
- [Design](#)
- [Feedback](#)
- [Resources](#)

Links

- [Initial Version](#)
- [Final Version](#)

Summary

This project is part of the Udacity's Data Analyst Nanodegree Program. I conducted an Exploratory Data Analysis (EDA) using Tableau on one of the curated data sets, provided by Udacity, from Prosper, which is America's first marketplace lending platform, with over \$12 billion in funded loans as of June 2021.

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, borrower employment status, borrower credit history, and the latest payment information. The dataset contains loans created between 2005-Q4 and 2014-Q1 with a last updated date of 03/11/2014. Here's a link to the variable definitions for this dataset.

The main purpose of this project is to create a tableau story where we explore the dataset and identify relationships among multiple variables using Exploratory Data Analysis (EDA) and data visualizations.

In this visualization, I start by exploring the dataset geographically to visualize the distribution of loans originated by state using a Packed Bubble graph. As a result, we see that California had the highest number of loans. Then,

subsequently, I decided to take a deeper look into some key attributes in the dataset by state (i.e., Prosper Rating, Employment/Loan Status, Income Range, Borrower/Interest Rate, etc.) using a map and bar charts to see how variables vary depending on the state. As an example, for loans originated in California, the average loan amount is \$8,974, Average Debt-to-Income is 26.19%, and Average Borrower Rate is 19.23% where most borrowers were employed/full-time that a majority had an income range that fell between the ranges of \$25,000 to \$74,999. On the other hand, in Texas, the average loan amount and Debt-to-Income was slightly higher, but with an average Borrower Rate that was 0.01% lower at 19.22%.

After exploring a few key attributes in the dataset, I decided to see if I can identify any relationships with Borrower Rate and Debt-to-Income ratio. I started by creating scatter plots. In the first one, I use Borrower rate and Loan Original Amount based on income range. As a result, the correlation between the two variables is dependent upon Income Range. As an example, borrowers with an income range of \$100k + have a strong negative correlation. Whereas borrowers with an income range between \$1 - 25k have a weak negative correlation. In addition, the maximum value of Loan Original Amount for borrowers with an income range of \$100k is \$35,000. Whereas the maximum for borrowers who reported an income less than \$100k is \$25,000. In my last scatter plot, I wanted to take a deeper look into the relationship between borrower rate and debt-to-income ratio based on income range. As a result, there was a correlation between borrower/interest rate and Debt-to-Income ratio in which it is generally positive. However, this is dependent upon correlation that varies based on factors such as the income range, employment status, and credit score. In addition, the average Debt-to-Income ratio for borrowers with an income range of \$1 - 25,000 is significantly higher than borrowers in other income ranges. By looking further into this income range by examining the employment status and using a text table to show very specific values, we find that the median Debt-to-Income ratio for self-employed borrowers is 2.13. In the end, this leads us to think that self-employed borrowers with an income range between \$1 - 24,999 tend to acquire loans way higher than their income level. Therefore, they can be considered high risk to investors.

Design:

- Audience:
 - Udacity Reviewer's and Students.
 - Potential investors and borrowers.
 - Individuals interested in Peer-to-Peer lending.
- The visualization will address the following:
 - Key attributes in the dataset I will be focusing on are interest rate and risk. This is to help potential investors/borrowers make informative decisions.
 - Relationships between the key attributes.

- Interesting/Unusual data points.
- Interactive views showing how the key attributes change based on the different variables/values.
- Chart Types:
 - Packed Bubbles: this chart will be used to visualize the number of loans originated per state based on the loan origination year using a cluster of circles.
 - Map: this chart will be used to visualize the data geographically (per state) and help the readers understand and explore how the key attributes vary depending on the state.
 - Bar chart: bar chart will be utilized significantly as the dataset include many categorical variables (such as: employment and loan status, income range, risk rating, etc.). Using bar charts will help compare specific measures across the different categories.
 - 100-stacked bar charts: this type will be used to show a breakdown of loan status (Current vs. Delinquent) across specific categorical variables.
 - Scatterplot: this visual will be used to plot two measures (such as borrower/interest rate and debt-to-income) to identify any possible relationships.
 - Text table: this visual will be used conservatively to show very specific values to help the reader spot the exact values explained in the description.
- Encoding:
 - Color Marks: I will use colors mainly to differentiate between loans that are current and those that are delinquent.
 - Filters: I will use interactive filters and shelf filters.
 - Field/value calculations: calculate specific values for the reader or fields to be used to organize (sort) the 100-stacked bar chart.

Feedback

I asked for feedback on all stages of design from 4 different people. Sometimes the questions and answers were verbal and others were documented. Please see the following feedback.

Review	Reviewer	Feedback	Action
1	Peer	The Summary by State was cut off when viewing this page on my desktop and the zoom was at 100%. I can see the wording if I zoom out more. This also applies to the other pages too. On the far right, some of the wording and graphics were cut off because of the zoom.	Rearranged graphs
2	Peer	The income range was cut off at the bottom, and I was unable to scroll.	Rearranged graphs
3	Peer	A short description that stands out a little bit on the landing page might help the viewer understand what it is all about.	Updated the introduction section to include and highlight key information the reader needs to focus on.

Review	Reviewer	Feedback	Action
4	Peer	It may look better to number your pages and move the information to the page's subtitle	Updated story points to numbers and moved text to subtitle
5	Udacity	Note that for the ranking, since AA is higher than A but since Tableau can't sort properly, please try this way as shown below: https://help.tableau.com/current/reader/desktop/en-us/reader_sort.htm	Moved 'A' rating to the right of 'AA' rating.

Resources

- <https://www.prosper.com/>
- <https://www.youtube.com/watch?v=EFEK0m4xK1o>
- <https://kb.tableau.com/articles/howto/showing-hawaii-and-alaska-near-continental-us>
- https://help.tableau.com/current/pro/desktop/en-us/maps_howto_simple.htm
- https://help.tableau.com/current/pro/desktop/en-gb/dashboards_organize_floatingandtiled.htm
- http://gamapserver.who.int/gho/interactive_charts/mbd/life_expectancy/atlas.html