

First Version of My Tableau Story:

<https://public.tableau.com/profile/xiaoxuan.jiang#!/vizhome/UdacityNanodegreeProsperLoansAnalysis/Story1?publish=yes>

Final Version of My Tableau Story:

<https://public.tableau.com/profile/xiaoxuan.jiang#!/vizhome/ProsperLoansAnalysisaguideforlenders/Story1>

Introduction

This is for people who want to lend their money and gives information on a lending market. Loan data information from Prosper, a lending platform, 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. Important information for lenders is visualized.

Summary of Findings

1. Loan Market is growing fast without sacrificing the loan quality. In fact, higher rating loans grows the fastest.
2. Lend out your money cautiously. There are two rules, the first is the higher the return, the higher the risk. The second is larger loans tend to have higher default risk.
3. Geographically wise, the East and West Bay areas plus Illinois have more peer-to-peer loan activities and thus more investment opportunities. Those are the states that have active and energetic economics activities as well. Furthermore, North East Bay areas where most financial services are located provide longer term loans. So those lenders who would like to learn more about longer term loans are advised to dive deeper and investigate into loan dataset of North East Bay areas, United States.

Design

1. Visualization for Overview of Prosper Loan Market

1.1 graph 1 Loan Growth Over Time

type of visualization: lines

x-axis variable: year

y-axis variable: count of loans

marks/color embedding: loan status

filters: loan status

design consideration: lines are best for showing changes over years, marks/color embedding is used to distinguish different trends for different loan status, filter is used to extract key information. Here, we do not need to show all the loan status. Some key loan status to represent high quality and low quality loan data are enough.

1.2 graph 2 Loan Status Breakdown by Year

type of visualization: stacked bars

x-axis variable: year

y-axis variable: count of loans

marks: loan status

design consideration: Stacked bars are best to show the quality contrast between different categories of data. It has showed clearly that despite high growth rate, the quality of the loans is not compromised.

1.3 graph 3 Term VS Year

type of visualization: lines

x-axis variable: year

y-axis variable: average term of loans

design consideration: As completed loans didn't grow as much as current loans, a hypothesis is that the longer term of loans resulted in the unique trend. So I decide to show the yearly changes of term of loans. A line graph is the best graph to represent changes over years.

2. Visualization for Yield and Loss Analysis

2.1 graph 1 ROI

type of visualization: side-by-side bars

x-axis variable: year

y-axis variable: ratings

marks/color embedding: yield/loss

filters: loan status, loan originated date, ratings

design consideration: side-by-side bars are best for showing the contrast between quantities of two categories of data. With this bar chart, we can see clearly the higher the yield, the higher the risk. The rule holds across years.

2.2 graph 6 Income VS Yield/Loss

type of visualization: side-by-side bars

x-axis variable: income

y-axis variable: loan yield/loss

marks: yield/loss

design consideration: side-by-side bars are best for showing the contrast between quantities of two categories of data. With this bar chart, we can see clearly the higher the yield, the higher the risk and the lower the income, the higher yield/loss.

2.3 graph 7 Analysis of Factors that Result in Default Loan

type of visualization: side-by-side bars

x-axis variable: income

y-axis variable: debt-to-income ratio, loan originated amount, monthly loan payment

design consideration: side-by-side bars are best for showing the contrast between quantities of two categories of data. With this bar chart, we see interestingly that no matter how rich the group of people is, the more money they borrow, the higher the risk of going default. The default loans tends to have a higher amount than the completed loans across different income groups.

3. Visualization for Geographic Factors

3.1 graph 8 Geographical Breakdown of Market

type of visualization: symbol maps

size of the symbol: sum of loan originated amount

location: borrower state

palette: average term of loan

design consideration: symbol maps are best to show both the location and intensity of events. The location information and the intensity of peer-to-peer activities are both key informations we would like to know in this case. So symbol maps are the best choice for overview of geographical breakdown of market. From the symbol map, we can see that though peer-to-peer loan activities happen across US, the types of the loans and the frequency and amount of the loans vary a lot. The Northeast Bay areas provide the longest term of loans.

3.2 graph 9 Top Market

type of visualization: maps

limit(set under filter function): top 5 borrow state by loan originated amount

location: borrower state

palette: borrower state

design consideration: maps are best to show the locations of events The location information is the single most important information we would like to know in this case. So we choose maps for this visualization. From the map, we can see clearly that bay areas plus Illinois have most active peer-to-peer loans.

3.3 graph 10 Not So Active Market

type of visualization: maps

limit(set under filter function): bottom 5 borrow state by loan originated amount

location: borrower state

palette: borrower state

design consideration: maps are best to show the locations of events The location information is the single most important information we would like to know in this case. So we choose maps for this visualization. From the map, we can see clearly that central US has the lowest peer-to-peer loan activities.

4. Visualization for Geographic Factors

4.1 graph 8 Overview of Current Loan Status

type of visualization: packed bubbles

size of the symbol: sum of loan originated amount

location: borrower state

palette: average term of loan

design consideration: I choose packed bubbles graph because it is a very intuitive way to show the percentage of quantities of different categories to total quantity. The graph showed that the credibility of the borrowers and quality of loans is good now with most of the loans at completed and current status and small amount of loans going default.

Feedbacks and Improvements:

1. When clicking the publish-to-public button, one need to select out the first page of the story or else the published story will be showing to audience whatever page you are reading at the time you publish the story.
2. The line graph(1st graph) for prosper loans overview over the years has too many lines and the data in the 2014 should be taken off as it appeared to be a drop in the trend but actually it's just because the data collection in 2014 was incomplete.
3. In the 2nd graph, one of my friends has noted that the completed loans had decreased, though the default loans kept low. This might be explained by the longer terms of the loans because people have more confidence in borrower's credibility. An improvement will be inserting a graph(3rd graph) to illustrate the relationship of term VS year.
4. The scales of the 5th graph for average estimated lender yield and average estimated loss are different. So I later forced a scale of 0-0.34 on both scales to ensure consistency.
5. The AA and A rating rows should switch their positions so that the order of layout is consistently from highest rating to lowest rating.

6. Didn't illustrate the average term information encoded in the colour information in the overview of geographical loan data(8th graph), The east bay of United States has been providing longer term of loans. So I added the explanation in the story text section of that graph.
7. We can see from the above comments that indexing the graphs is important for future references, so I added index for each graph such as "1st graph", "2nd graph", etc.
8. Udacity Review: In the 6th slide, the bar chart, Y-axis is labeled as 'value'. You need a descriptive axis label to tell readers what variable this axis represents.
9. Udacity Review: In the 7th slide, the line chart is not appropriate for this data type since it is best to be used with the time series data like what you have created in the first slide. you may change it to a bar chart, have a look at the image below.
10. Udacity Review: In the design section of the document, please include details about charts used with an explanation for its usage.

Resources

Github Repo:

<https://github.com/igorstojanovic91/udacity-create-a-tableau-story>

<https://github.com/latinacode/Create-a-Tableau-Story>

Tableau Published Page:

https://public.tableau.com/profile/steve7an#!/vizhome/prosper_loan_data_0/Story1