Data Visualization

Lanmixue Mao 2018/4/16

initial version link:
https://public.tableau.com/profile/michelle4362#!/vizhome/ProsperLoan_4/Story1

final version link:
https://public.tableau.com/profile/michelle4362#!/vizhome/Loan_finalversion/Story1

[Summary]

The prosper loan dataset is comprised of 81 variables and contains 113937 entries.

The prosper loan data can allow me to explore:

1. Customer Quality: What is the loan amount in each state, and what is the borrower's income range in terms of different prosper rating?
2. Loan Provider Profitability: How different is the borrower interest rate from time to time, and how much interests and fees the loan provider can earn in each year?
3. Credit Risk: By examining the relationship between delinquencies and prosper rating, we can identify which prosper rating borrower is more likely to generate credit risk.

[Design]

{Customer Quality}

Chart 1: Bar Plot: Columns: Prosper Rating Group, Rows: SUM[Number of Records] (Color: AVG[Debt To Income Ratio], Filter: ProsperRating, Year, Loan Original Quarter) Find Prosper Rating, group exclude NULL, then set a new group named "Prosper Rating Group".

Edit: Make a new group named "Prosper Rating Group" to get the prosper rating data without missing values. Sort the prosper rating to show the value in the sequence "AA, A, B, C, D, E, HR". Enlarge the grey text box to display more text.

Chart 2: Bar Plot: Rows: Income Range (group), Columns: SUM[Number of Records] (Color: Employment Status, Filter: Year, Loan Original Quarter) Find Income Range, group $0 and not employed as "$0", then set a new group named "Income Range (group)".

Edit: Make a new group named "Income Range (group)" to combine $0 and not employed to $0.

Chart 3: Map: Borrower State (Color: SUM[Loan Original Amount], Filter: Year, Loan Original Quarter, Detail: Borrower State)

Edit: Change the color scheme to brighter color. Show the label of the sum of number of records.

Chart 4: Map: Borrower State (Color: Number of Records, Shape: Prosper Rating Group, Filter: Year, Loan Original Quarter, Detail: Borrower State) Find Prosper Rating, calculated field: “IF NOT ISNULL([Prosper Rating]) THEN [Prosper Rating] END” as “Prosper Rating Group”

Edit: make a new group named “Prosper Rating Group” to select non-missing values. Change the shape to show Prosper Rating Group.

{Loan Provider Profitability}

Chart 5: Comparison (Line Chart & Dual Axis & Trend Line): AVR[Borrower APR], AVG[Borrower Rate], AVG[Lender Yield]

Edit: Change the loan date from year to quarter which shows the volatility of lines. Predict the trend of lines. Show the start and end values.

Chart 6: Pie Chart: Angle: SUM[LP InterestandFees], Color: ProsperRating, Label: SUM[LP InterestandFees], ProsperRating

Edit: Show the percentage of each pie.

Chart 7: Columns: AVG[Loan Amount] , Rows: AVG[Estimated Return] (Percent of Total), Color: ProsperRating, Shape: ProsperRating

Edit: Remove average Borrower APR from shape. Calculate “Estimated Return Amount” by “Estimated Return * Loan Amount”. Show the average estimated return amount label.

{Credit Risk}

Chart 8: Dual Combination: Columns: Year(Loan Date), Rows: AVG[Current Delinquencies], Color: ProsperRating

Edit: Remove average debt to income ratio. Not show the null prosper rating.

Chart 9: Amount Delinquent (calculated field: [Amount Delinquent] > 0 AND NOT ISNULL([Amount Delinquent])) as ‘Delinquencies’ group. Find this group: Pie Chart: Columns: Delinquencies, Color: Prosper Rating, label: SUM[Number of Records](Percent of Total) and Prosper Rating, angle: SUM[Number of Records]

Edit: Change the bar chart to pie chart. Display “Not Delinquent” and “Delinquent” in two pie charts. Remove average debt to income ratio and average public record last 10 years. Show each prosper rating percentage in the pie chart.

Chart 10: Revolving Credit Balance (calculated field: ([Revolving Credit Balance] / [Available Bankcard Credit] >= 0.7) AND NOT ISNULL([Revolving Credit Balance] / [Available Bankcard Credit] >= 0.7)) as ‘Liabilities’ group. Find this group: Pie Chart: Columns: Liabilities, Color: Prosper Rating, label: SUM[Number of Records](Percent of Total) and Prosper Rating, angle:

SUM[Number of Records]

Edit: Change the bar chart to pie chart. Display “High Liability” and “Low Liability” in two pie charts. Remove average debt to income ratio and average bankcard utilization.
Show each prosper rating percentage in the pie chart.

[Questions]

1) What do you notice in the visualization?
2) What questions do you have about the data?
3) What relationships do you notice?
4) What do you think is the main takeaway from this visualization?
5) Is there something you don’t understand in the graphic?

[Feedback 1]

Chart 1: Bi modal distribution Null has the highest frequency. What is on the horizontal axis?
Chart 2: Looks like normal distribution 25 000 to 74000 have the highest count. Why is zero different to non-employed? What is not displayed and why?
Chart 3: Just shows the US. Why is California so different? Heading should be more obvious.
Chart 4: What does the rating mean. Sees to be concentrated on the west coast.
Chart 5: The variables need more explanation. Should include reserve ratio from the federal reserve.
Chart 6: Seems 3 colours occupy everything. Needs labels.
Chart 7: Don’t understand this.
Chart 8: Not sure what the lines are. 2007 average debt to income was the highest, why did it fall too much in 2008? Lines are too crowded.
Chart 9: Don’t understand this.
Chart 10: Two graphs look different. Why is C so high? Don’t understand this properly.

[Feedback 2]

- The text in the grey boxes on top of your charts (the text bar) is not fully readable. The text is clipped. I see there is a scroll bar, but it is a bit taxing.
- The sort order for the credit grade is wrong: AA is the best credit grade and should therefore be at the top.
- You could maybe use more distinct colors for your variables (it is the same color scheme during all the story).
- Some graphic is too small and deserve a whole page (Chart 2 and Chart 6).
- I do not think the missing data in Chart 8 is important to show.
- I do not understand the definition of high liability in the Chart 10, it needs explaination.

[Resources]

Subprime mortgage crisis:
https://en.wikipedia.org/wiki/Subprime_mortgage_crisis