

PPP Loan Distribution and Job Retention

Undergraduate Group 5
Clare DuVal, Brayden Fidler, Jeffrey Wang
CPCS 4030/6030, Fall 2020

Project Report

Introduction

The Paycheck Protection Program (PPP) was passed as part of a series of government bills designed to help Americans during the outbreak of the Coronavirus pandemic. PPP was widely criticized for its shortcomings, inefficiency, and potential for abuse. There were even concerns that PPP was being used to curry political favors by the Trump administration. Others speculated whether or not this program would be an effective means of economic stimulation during the pandemic. We attempt to get to the bottom of the story through careful examination of the data available and unbiased visual analysis. In doing so we seek to answer three simple questions:

- Who is receiving the loans?
- Where are the loans going?
- How well are the loans doing?

Dataset

This PPP data comes from the United States Small Business Association in response to keeping transparency between American taxpayers while protecting the confidentiality of businesses included in the dataset. In total there were 661,218 entries. The table of data included 16 columns of data including: LoanRange (Categorical), BusinessType (Categorical), Address(Categorical), City (Categorical), State (Categorical), Zip (Categorical), NAICSCode (Categorical), RaceEthnicity (Categorical), Gender (Categorical), Veteran (Categorical), NonProfit (Categorical), JobsRetained (Numeric), DateApproved (Ordinal), Lender (Categorical), and CD (Congressional District) (Categorical). Several self-reporting columns like Race, Gender, and Veteran status were largely null. We chose not to focus on those for our data visualizations given it wouldn't serve as an accurate representation to have a majority of empty values.

Dataset: <https://www.kaggle.com/susuwatari/ppp-loan-data-paycheck-protection-program>

Design solution

1. Loan Numbers by Business Classification (Clare)

This visualization is a hierarchical bar chart showing the Amount of Loans (Quantitative) distributed first by Business Type (Categorical), then drilled down to Loan Ranges (Categorical). When the visualization first loads on the website, each Business Type has a corresponding horizontal bar with the length representing the amount of loans. The bars are sorted ascending down the page. When the user wishes to drill down to see how the chosen Business Type amount of loans is split into the different loan ranges, they click a bar/label and it will break down to the categories. The axis changes according to the new amount of loans. Again, each loan category is sorted ascending. When the user wants to return to the breakdown by Business Type, they can click on the background.

This visualization conveys how the amount of loans by Loan Range have more significance when shown first the distribution over Business Type. When a small business does research to predict what kind of loan they would receive, if they are limited to only the amount by loan range, they are stuck because Corporation and LLCs own a large majority of the total amount of loans. With this hierarchical visualization, they are able to drill down to the Business Type they identify as, and see their desired distribution.

2. Jobs Retained by Location (Brayden)

The second visualization is a geometric graph of the United States. Therefore, the position of each state lies in accordance with the standard map of the United States and identifies each unit of data - a state. Furthermore, the primary attribute displayed is the average number of jobs retained per loan (quantitative) in each state. The saturation and hue of the state is what distinguishes the values of one state from another. A dark blue state (like Utah) has a relatively high average number of jobs retained while a light blue state (like Virginia) has a lower number of average jobs retained. Hovering over each state provides some additional attributes via text labels for each state, such as the state's name

(categorical), the total number of jobs retained (quantitative), the total number of loans awarded (quantitative), and the actual average value mentioned previously (quantitative).

The primary purpose of this visualization is to compare and summarize the role that location of loan awards played in determining the effectiveness of the PPP loan. We wanted to see if outcomes were fairly homogeneous or if there were clear differences in each corner of the country. To that end, our visualization was quite successful because it is clear that on average, loans went further in some states than others. The causes and reasons that explain why some states were more successful than others, however, gives way to further debate. Perhaps it is political or perhaps work cultures, costs of living, local industries, or other factors play a significant role.

3. Business Types Job Retention(Trend) (Jeffrey)

I decided to do a scatter plot with color channels for the different categories and also linear regression lines to help the user make sense of trends. The color choices were to help the user distinguish business types and also to associate the points with their respective lines. The points are averages of the categories, and I came upon that solution after plotting a scatterplot of all the data points. It seemed messy and it didn't tell any narrative or give any hint of a trend. So I decided to plot the averages and also calculate linear regression lines. The linear regression lines are somewhat misleading because they are based on categorical data, but it helps add more depth to the graph. A scatter plot with only 5 categories was not a good visualisation.

I think that the visualization effectively conveyed the differences in how the business categories used their PPP loans to save jobs. It was fairly clear that nonprofits did a lot better than the other categories from the trend and also the average points. Most of the other points and lines were close together. The points were more sporadic, so I think the use of a linear regression helps bring out the overall trend.