**Introduction**

In a time like these with Coronavirus, helping struggling business stay open and retain good employees is extremely important. The Paycheck Protection Program (PPP) is one of the programs created that has become vital to many businesses. This program was, per the Small Business Association (SBA), “designed to provide a direct incentive for small businesses to keep their workers on the payroll.” The SBA will also forgive the loan if the funds are used for eligible expenses. This creates many questions. For example, what types of businesses are taking out these loans? What states have taken out the most loans? How many jobs was this program able to help retain? As well as, did all the business that used the money use it for it’s intended purpose or did they use it for other means?

**Analysis and Models**

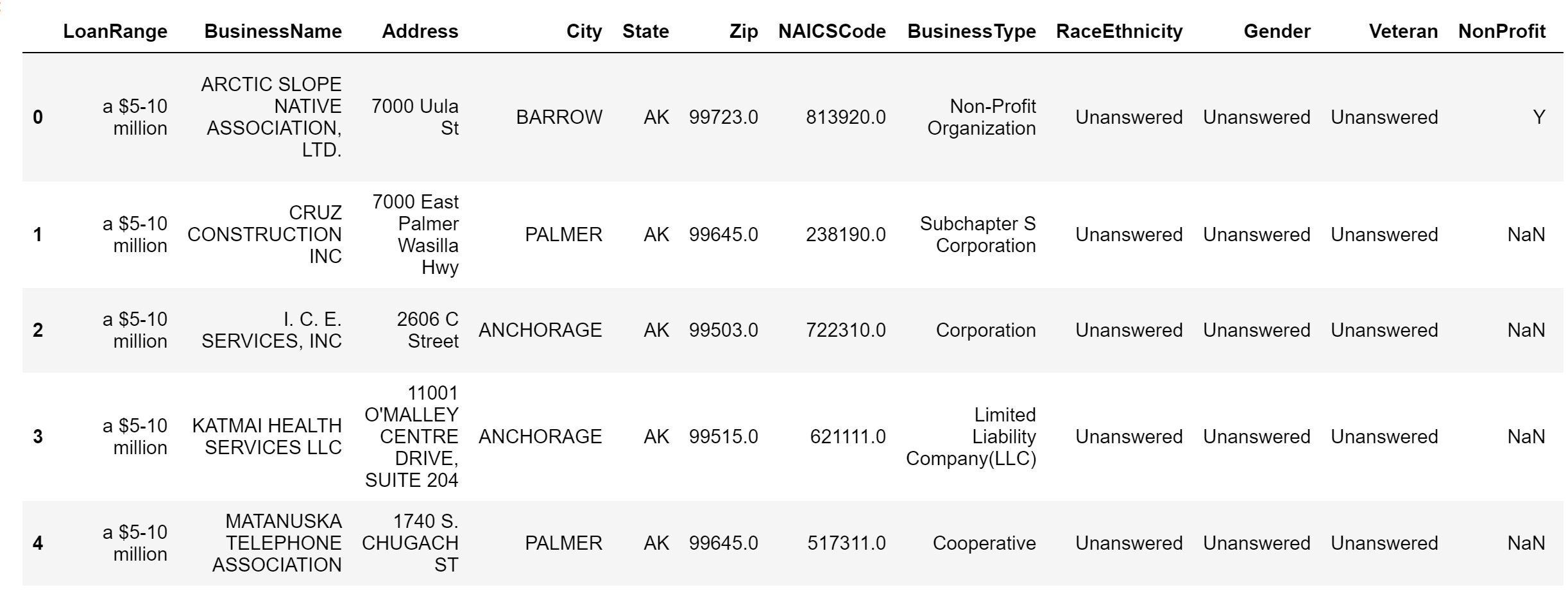
**About the Data:**

There were two datasets that were examined to seek the answers to the questions being asked. The first dataset was obtained from Kaggle.com and it consisted of a CSV file that contains 16 columns and just over 660,000 rows of data from the period of April 1, 2020 to June 30, 2020.

These are the following attributes that were in this dataset:

|  |  |  |
| --- | --- | --- |
| **Fields** | **Description** | **Example** |
| LoanRange | Range of loan approved | A $5 - $10 million |
| BusinessName | Name of the business | ARCTIC SLOPE NATIVE ASSOCIATION, LTD. |
| Address | Address of the business | 7000 Uula St |
| City | City of where the business is located | Barrow |
| State | State of where the business is located | AK |
| Zip | Zip Code | 99723 |
| NAICSCode | North American Industry Classification System | 813290 |
| BusinessType | The type of the business | Non-Profit Organization |
| RaceEthnicity | The race ethnicity | Unanswered |
| Gender | The gender | Unanswered |
| Veteran | Were they a veteran | Unanswered |
| NonProfit | Are they a nonprofit | Y |
| JobsRetained | How many jobs retained with loan approval | 295 |
| DateApproved | Date of approval | 04/142020 |
| Lender | Lending bank | National Cooperative Bank, National Association |
| CD | Business Congressional District | AK - 00 |

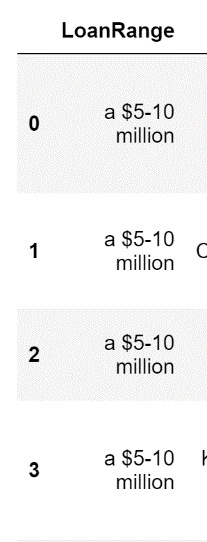
The first step before any analyzation can be done is the cleaning and preparation of the data. To be able to get a better understanding of the data, the first five rows were examined, the figure below show the first 12 attributes:



The next step was to get a better understanding of the data by performing a summary of the statistical values of all of the attributes, including the non-numeric attributes. This will help to get a better understanding of the top values to which to focus on during the analysis of the data. For example, under the Loan Range attribute, the top value is $150,000-$350,000 which can be further examined to see if in fact it is the top value. This also shows the mean or average, the minimum and maximum amounts with the quartiles, and the standard deviation for the numeric attributes, such as Jobs Retained.

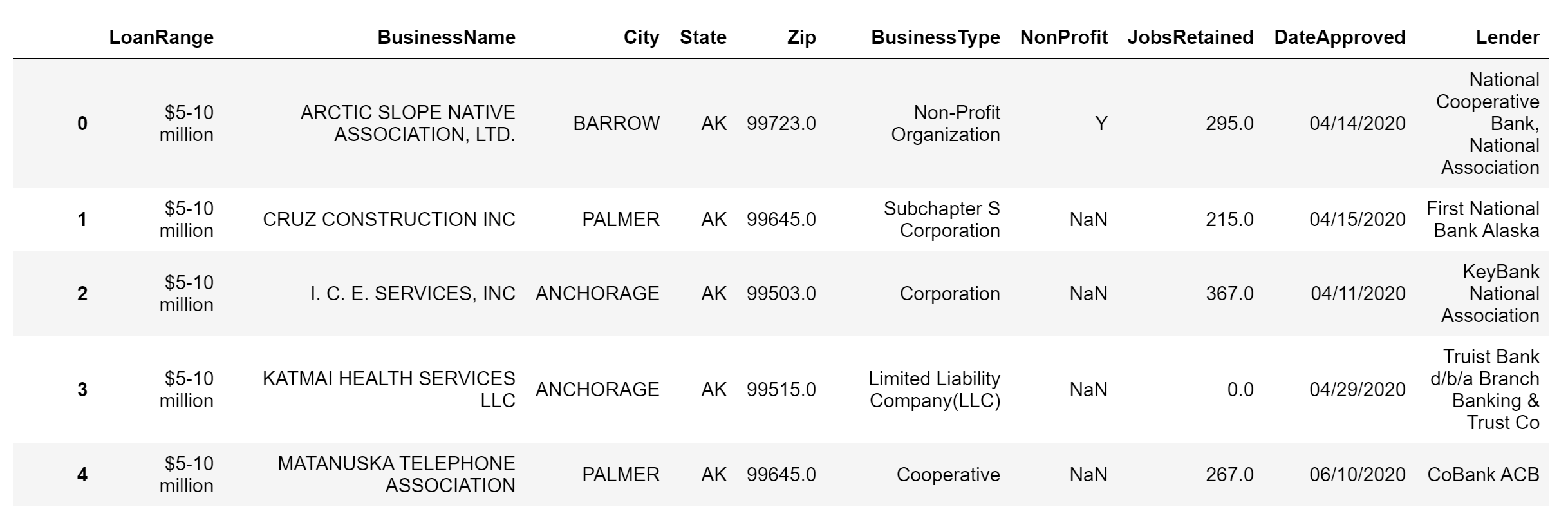


The next step was to remove the letter before the range in the Loan Range attribute. This will ensure a cleaner look when creating the graphs that will be used to analysis the data.

*Figure 1: Loan Range attribute with the letter before the range Figure 2: Loan Range attribute without the letter before the range*

The final step was to remove the attributes that are not necessary for this analysis. The attributes removed were Address, NAICSCode, RaceEthnicity, Gender, Veteran, CD. While the Congressional Districts (CD) were an area that was originally going to be examined, due to not all business entering that information, it was decided to break down the loan by the state. After removing the attributes, the remaining attributes were, Loan Range, Business Name, City State, Zip, Business Type, Non-profit, Jobs Retained, Date Approved, and Lender.



The other dataset that was examined were current tweets about PPP Loans. Originally the Facebook page for the Small Business Association as well as a public group about called Paycheck Protection Program, both of which are sources for businesses looking for help with their applications. Unfortunately, the setting on these two Facebook pages made it a challenge with getting information, thus a data set containing 2,000 tweets was used to see what type of information is being tweeted about PPP Loans.

The first step was to collect the tweets from Twitter. To be able to be able to do this, an account was created with Twitter allowing the user access to Twitter’s platform. This account, which contain the access user’s information and keys, allows the user to be able to pull tweets containing certain information, such PPP loan.

After the tweets were obtained, they were analyzed to see which ones were retweeted more than 100 times and which ones were retweeted less than 50. This was done to see what types of stories get retweeted faster than others and do these stories so anything about the types of businesses.

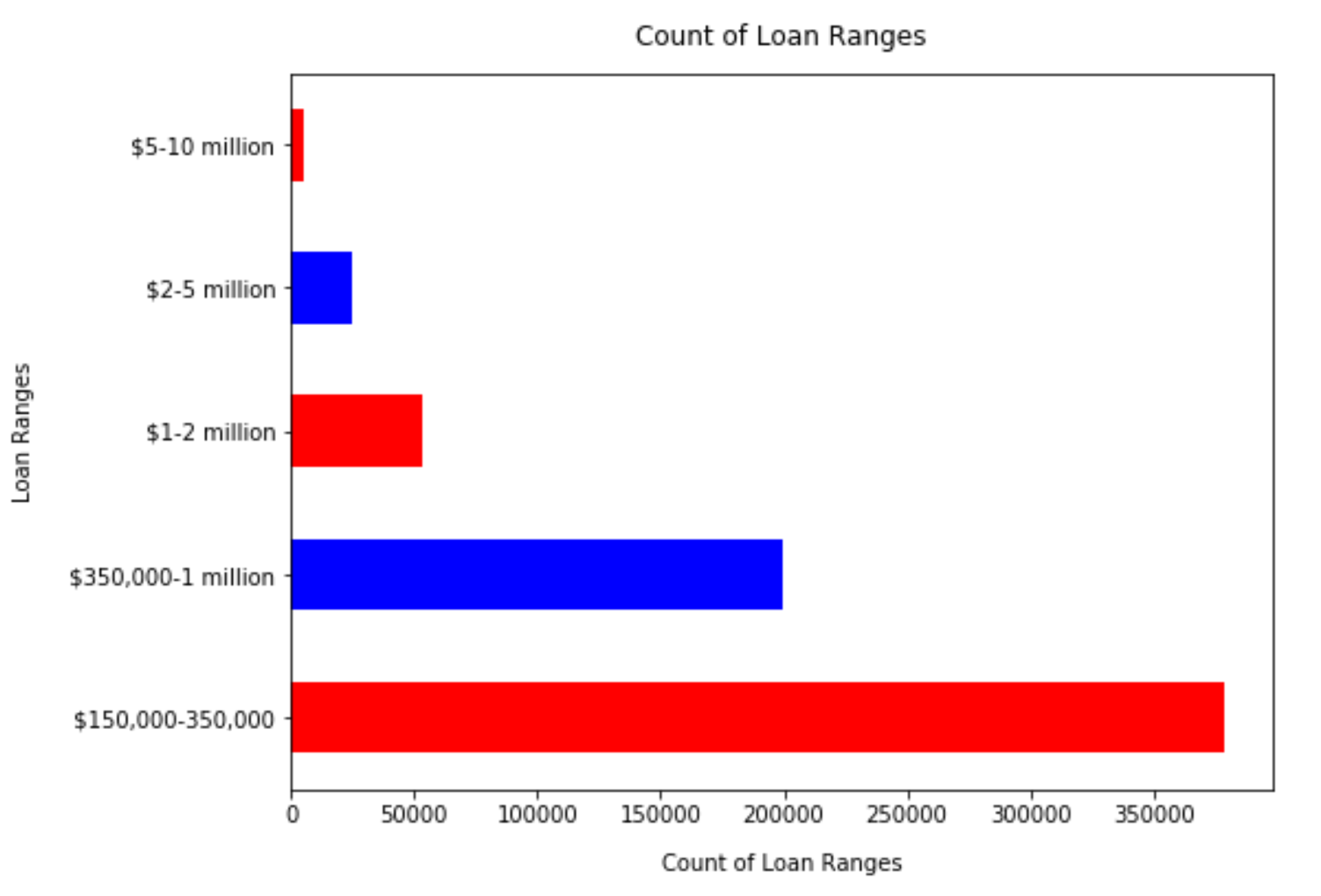
**Models and Results:**

The original questions that were going to be examined were:

* What Business Congressional Districts gave out the most loans?
* How many businesses were in each loan range?
* What is the percentage of the types of business to receive the loans?
* Which lending banks processed the most loans?

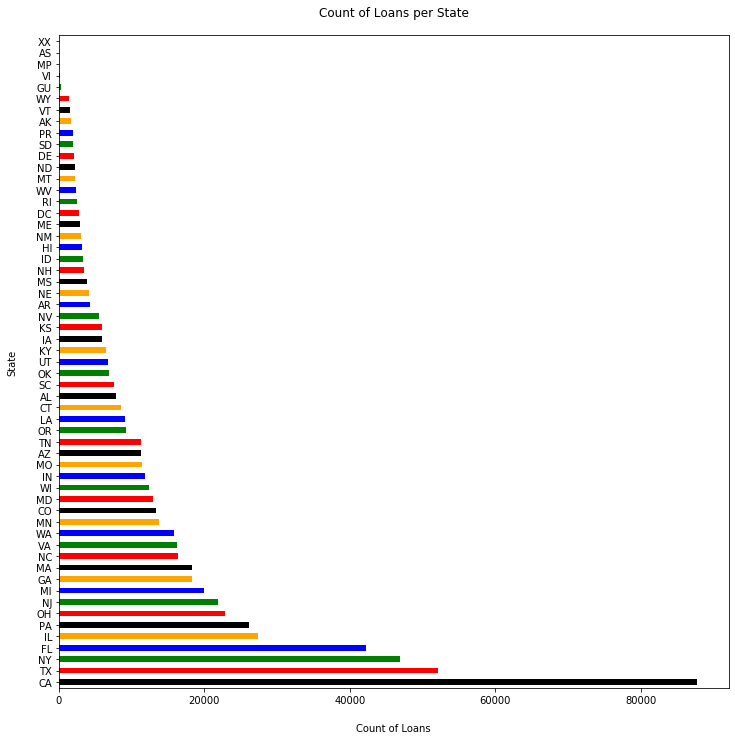
As the data was further examined some of these questions were changed and some eliminated and replaced with other questions. This will be explained further with each model.

The first thing that was done was a chart of the loan ranges to get a better understanding of what loan range had the most loans. In examining the chart below it was determined that the top two ranges were $150,000-$350,000 and $350,000-$1 million and the $150,000-$350,000 was almost double the count of $350,000-$1 million. This information could be used to show that small businesses were in fact the larger beneficiary of the PPP loan program as the program was intended but further examination should be done to truly be able to understand what types of businesses did in fact take out the loans and where are they located.



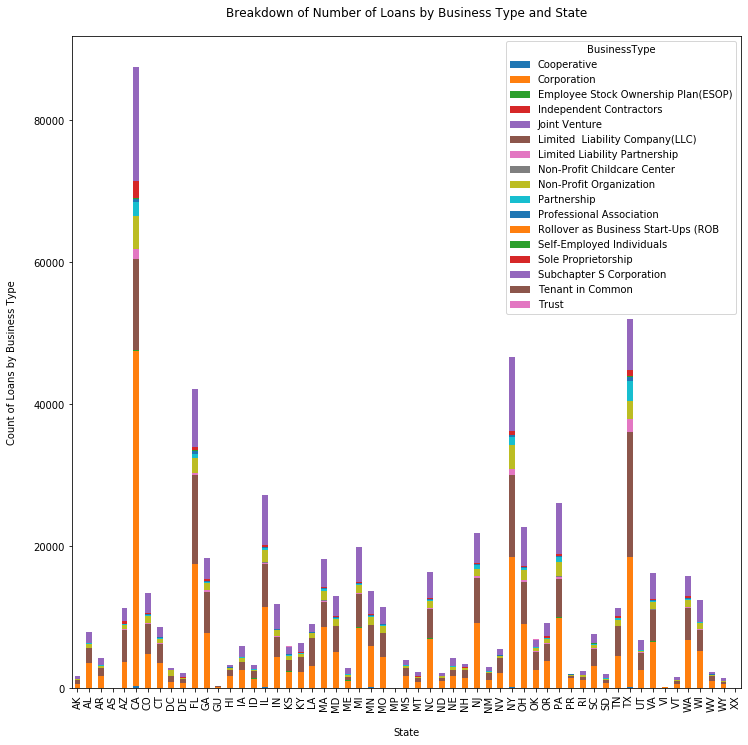
This leads to the question: What Business Congressional Districts gave out the most loans?

Since not all business listed their districts, it was decided to use the question: What States gave out the most loans? The bar chart was created to examine the amount of loans for each state by counting each loan in the Loan Range and then grouping them by their State location. This showed that California had the highest amount of loans followed by Texas, New York, and Florida. When looking at this from a population aspect these four states being in the top make sense. What should be noted is that California has almost 40,000 more loans than Texas yet population wise they are only about 10 million apart per <https://worldpopulationreview.com/states>.

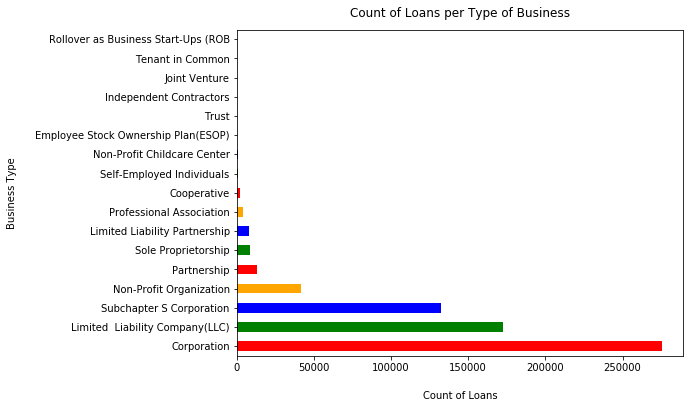


This led to the next question: How many businesses were in each loan range?

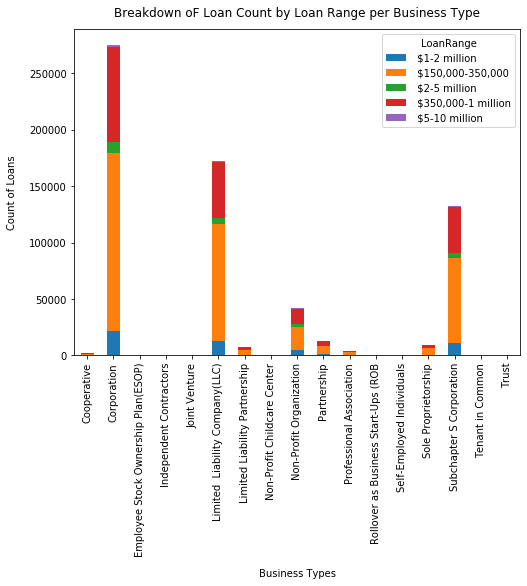
Since this data type breaks down the businesses by types, it was decided to examine the the amount of loans each Business Type received and then break down that information by State. This was accomplished by examining the count of each business type that has a loan an group them by the state. This again confirmed that top four states but another observation from this stacked bar chart is that C corporations (C corp) are one of the biggest recipients of the PPP loans.

The observation about the corporations led to the next question: What is the percentage of the types of business to receive the loans?

Instead of examining percent, the count of loans per each business type was examined instead so that it coincided with the stacked bar chart. By counting the loans from each business type, it confirmed what was observed with the corporations being the largest recipient of the PPP Loans. This bar shows that Corporations, Limited Liability Company (LLC), and Subchapter S Corporations (S corp) are the top three. This leads to more questions such as, are all the businesses listed under corporations C corps or could some be S corps? Also what are the size of the business under the corporation? It should be noted that with C corps tending to be a larger business, this is an area that should be further examined through other means to ensure that the loans they received were not intended from smaller businesses that may have needed them.

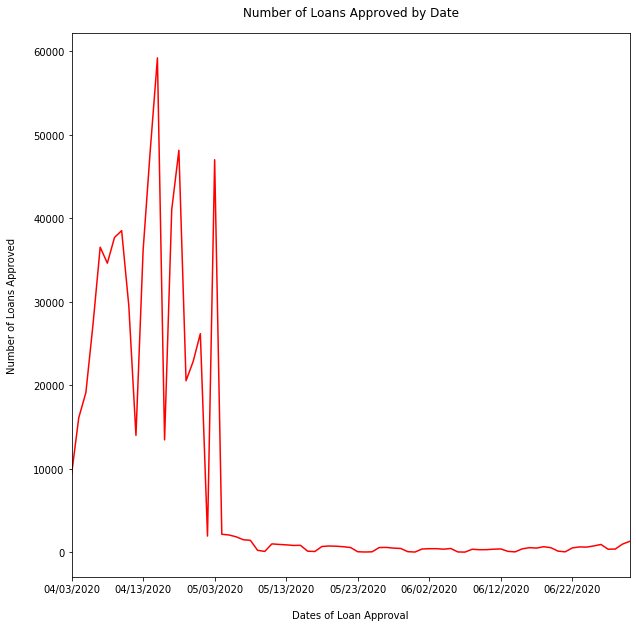


A stacked bar chart was created to get a better understanding of the breakdown of the counts of loan range by the type of business. When examining the stacked bar for the corporation, what pops out is that the majority of the loans are in the $150,000-$350,000 and $350,000-$1 million range. Which confirms what we have examined in other charts. It should be noted that while this confirms that the corporations are following the same trends as the LLCs and S corps, this data set does not show if the corporation took out loans for different locations. As reported by NBC news (<https://www.nbcnews.com/business/business-news/which-companies-are-returning-their-ppp-loan-here-s-list-n1194566>), many larger corporations took out loans that were intended for small business. This situation caused the SBA and the Treasury Department to issue new guidance on the loans.



Now knowing about the situation with big corporations taking out loans and how some are now being returned by May 7, 2020, it creates the question: when are these loans being approved?

While this data set is dated from April 1, 2020 to June 30, 2020, the majority of the loans were approved before May 3, 2020 in this dataset. The largest number of loans were approved around April 13, 2020. This leads to the question of how many of the loans in this dataset had to be returned due to the new guidance.

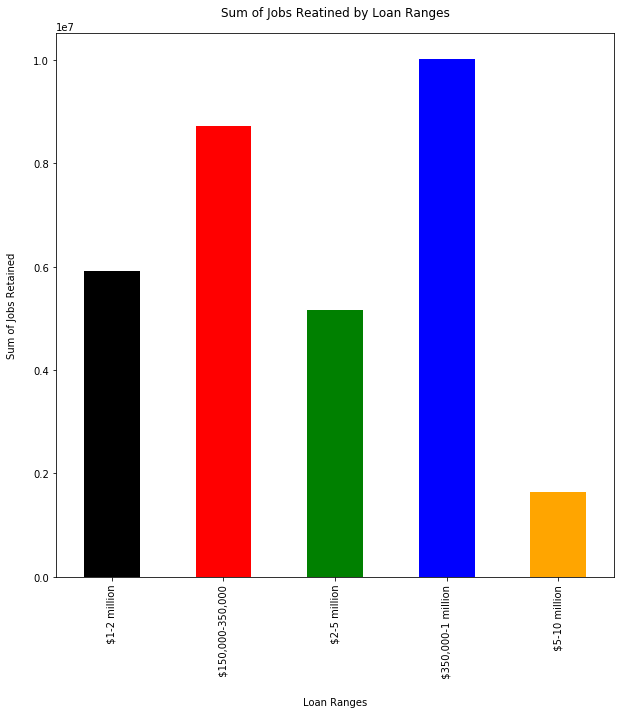


The final beginning question was: Which lending banks processed the most loans?

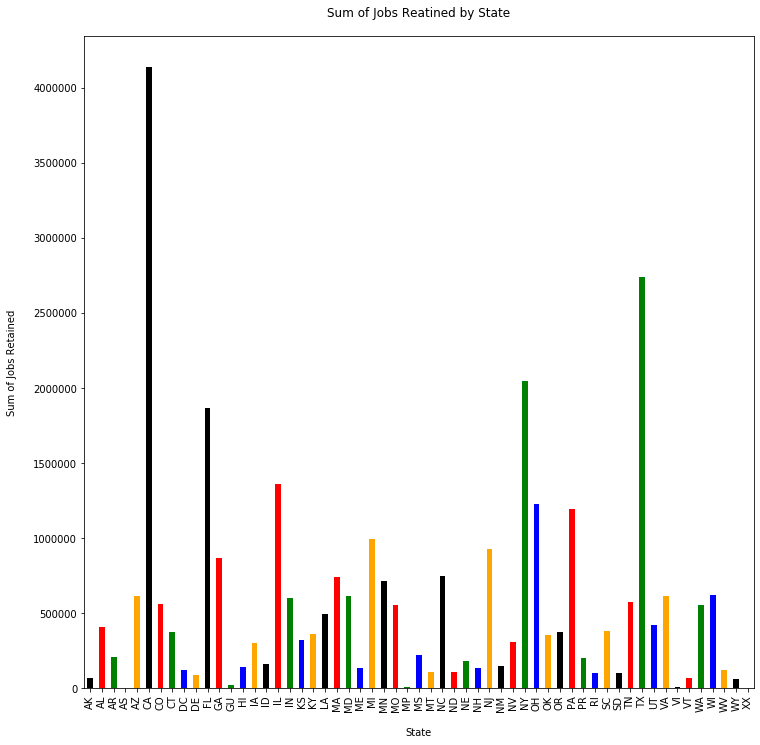
Upon examining the data to be able to create this chart, Juypter Notebook, the software used to examine the data, crashed. After a few more tries, a bar graph finally came through but it really did not answer the questions, thus this question will not be examined.

One question that was not asked in the beginning but is probably the most important is: How many Jobs did this program retain?

First it was examined how many jobs were retained by each Loan Range. While there were more loans taken out in the range of $150,000-$350,000, almost twice as many as the $350,000-$1 million range, there were actually more jobs retained in the range of $350,000-$1 million. It should be noted that between the two ranges close to two million jobs were retained but there may need to be more research into this area to find out why the range that has more loans didn’t retain as many jobs.



When the jobs retained are broken down by State, California, Texas, New York, and Florida are again in the tops spots which coincides with all the other analysis that has been done.



It should be noted that when examining the Tweets collected that there are quite a few people who are tweeting and retweeting opinions and stories about positive outcomes with the program as well as possible misuse of the funds being provided with this program.

Retweets: 108

Tweet text: RT @SenMcSallyAZ: Over 1 million Arizona jobs were saved by the Paycheck Protection Program! Last Saturday, I visited with @PrescottBrewing owners John & Roxane who shared that they were able to pay their hourly workers using their PPP loan.

Retweets: 5381

Tweet Text: RT @TheRickyDavila: I still can’t get over the fact that Steven Mnuchin funneled $500B in PPP loan covid relief to entities of his choice including himself, Devin Nunes, Jared Kushner, Kanye West, Moscow Mitch’s wife Elaine Chao, but $600 for struggling Americans is too much. Evil is as evil does. (One of the retweets can be found here: <https://twitter.com/MarshaDB54>.)

Retweets: 584

Tweet text: RT @KlasfeldReports: Gas station secured small business bailout money, then paid for Trump billboards, via CNNPolitics https://t.co/y7ZqB9G… (The CNN story, which can be found here: <https://www.cnn.com/2020/08/28/politics/trump-billboards-ppp-loan-invs/index.html>)

It should be noted that the majority of the Tweets collected discuss more possible misuse within the program than possible success with the program.

**Conclusion:**

During this pandemic, we as a country need to do as much as possible to be able to help out fellow man. Having programs such as the Paycheck Protection Program, helps to ensure that our fellow citizens will be able to afford to afford their household needs, purchase food, and have access to health insurance. This program will also ensure to help keep people off of Unemployment Insurance as this program is already extremely strained by employees of businesses that had to furlough them. While there are some areas of the program that should be further investigated, for the most part that program has been successful in helping businesses retain their employees.