Syracuse University

Memo

|  |  |
| --- | --- |
| To: | Dr. Landowski |
| From: | Samantha Brennen-Lisko |
| Date: | August 8, 2020 |
| Re: | Project Proposal |

**Topic: Analyze Paycheck Protection Program**

**Data Description:**

The main data set consists of a data set exported from Kaggle.com, which can be found here: <https://www.kaggle.com/susuwatari/ppp-loan-data-paycheck-protection-program> . It consists of one CSV file that contains 16 columns and just over 660,000 rows of data from the period of April 1, 202 to June 30, 2020.

|  |  |  |
| --- | --- | --- |
| **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 |

I will then look to Facebook for supplemental data. The two pages I will be looking at are the Small Business Association, which can be found here <https://www.facebook.com/SBAgov>, and a public group I found as a source for businesses looking for help and guidance, which can be found here <https://www.facebook.com/groups/paycheckprotectionprogram>. I will also be looking at news articles about the program to see how successful or unsuccessful the program is. As someone who will be auditing Coronavirus Relief Funds, I am interested in other government COVID-19 Fund Programs are working especially in the area of internal controls.

**Research Questions:**

* 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?

I will be starting with these questions but as I delve more into the data these questions may change or others added.

**Data Preparation Plan**

1. Load the data into Python
2. Concatenate the data from each file into one data set
3. Filter the data set to only attributes we are interested in analyzing
4. Explore the data for anomalies, missing data, adjusted duration, etc.
5. Resolve any issues related to the previous step, including deciding how to handle records with missing data.