|  |
| --- |
| IST 652 |
| Syracuse donor data |
| Manipulating donor dataset utilizing Python pandas data frame |

|  |
| --- |
| Matthew Maloney  12-2-2020 |

Contents

[Introduction 1](#_Toc56117170)

[The data set 1](#_Toc56117171)

[Loading and formatting the data 2](#_Toc56117172)

[Simple analyses 3](#_Toc56117173)

[2021 giving by state 3](#_Toc56117174)

[Top constituents by lifetime giving 4](#_Toc56117175)

[Top cities by lifetime commitment 4](#_Toc56117176)

[Constituents ready for solicitation 6](#_Toc56117177)

[Regional gift officers 7](#_Toc56117178)

[Opportunity across regions 7](#_Toc56117179)

[Prospect lists for gift officers 8](#_Toc56117180)

# Introduction

For this project, I wanted to utilize data from my job, as the attributes and values are all very familiar to me. I work as a research analyst at Syracuse University’s Division of Advancement and External Affairs. My role is to identify potential sources of future philanthropic gifts by culling our own database and by relying on publicly available wealth indicators (real estate, stock and salary disclosures, etc.).

Within the confines of this project, I strove to find a dataset that was finite. I pulled a report of all constituents with credited non-zero values of “new business” for the fiscal years of 2019, 2020, and 2021. Our fiscal year runs from July 1 of the prior calendar year through June 30. New business denotes a new cash gift or a new pledge but does not include payments towards previous pledge commitments. The data set that I am currently using contains 3,599 observations (representing individuals) across ten attributes. Unique identifiers like name, address, and other contact information have all been suppressed.

A core part of my position is to work directly with regional gift officers, who are posted throughout the country and work with constituents within a confined geographic footprint, to ensure that they have refreshed prospect lists throughout the year. At the outset of this project, I was particularly interested in utilizing Python to create some of these prospecting lists as a use case. In addition to these deliverables, I have investigated a few descriptive statistical questions about this group of constituents (specifically related to the states and cities in which they reside).

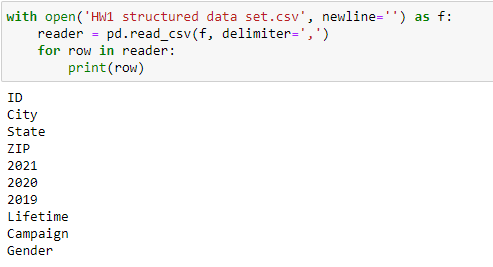
\*A note: the database utilized by our division is called CITRUS, hence the name of many of the objects in these programming snippets.

## The data set

The attributes included are:

* A system generated ID (so as to anonymize the data)
* City
* State
* Postal code
* Value of gifts made in FY 2021
* Value of gifts made in FY 2020
* Value of gifts made in FY 2019
* Value of lifetime giving
* Value of gifts during current campaign
* Gender

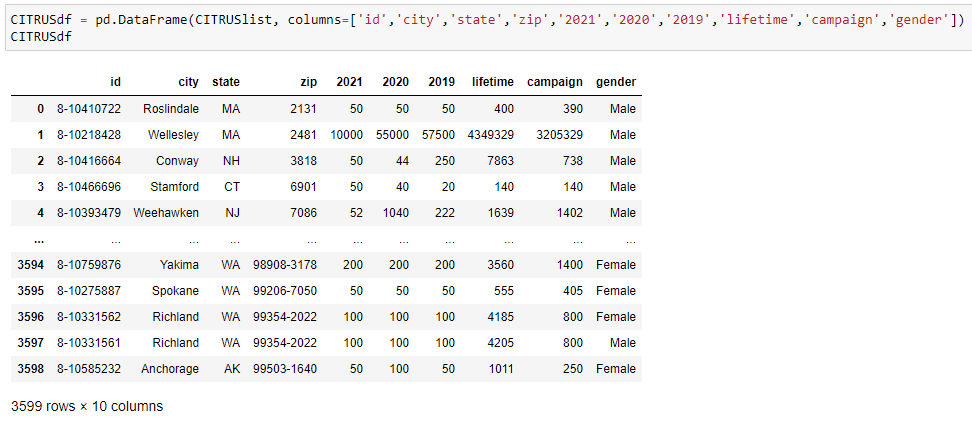
# Loading and formatting the data



My first step was to remind myself of the column names. Once I had done that, I brought the table in as a dictionary of lists. I first had to ignore the column names, so that they were not included as a row of data. I then had to provide Python with the keys for the dictionary, which were the same as the columns in the initial data set. As you can see below, the program correctly read in 3,599 records.



But a dictionary is not as easy to manipulate, so I converted CITRUSlist into a data frame utilizing the pandas library.

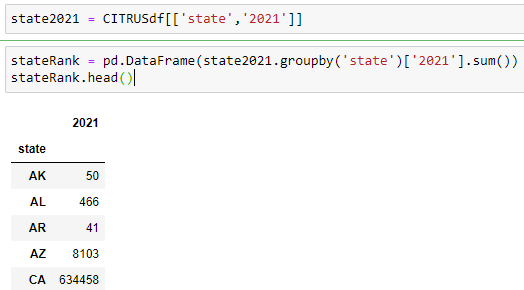


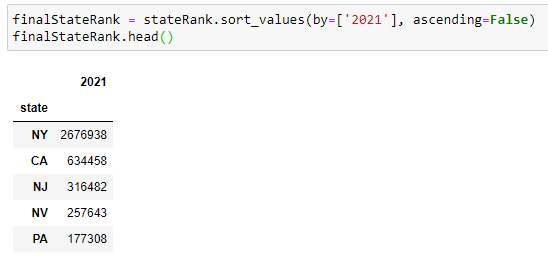
# Simple analyses

With the data now easily accessible in a pandas data frame, I could now begin to perform some initial analysis.

## 2021 giving by state

First I decided to extract only the columns for state and fiscal year 2021 giving. I then grouped all constituents by state and summed their 2021 giving.

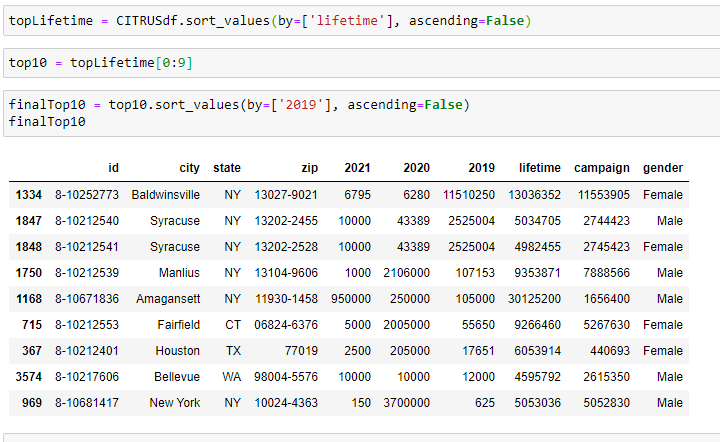




This final list is ranked in descending order, and demonstrates that some of the most populous states, and the states with the greatest number of Syracuse alumni, are best represented in this list. Speaking with lots of anecdotal knowledge of the constituency, Nevada is a bit anomalous.

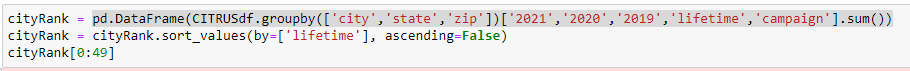
## Top constituents by lifetime giving

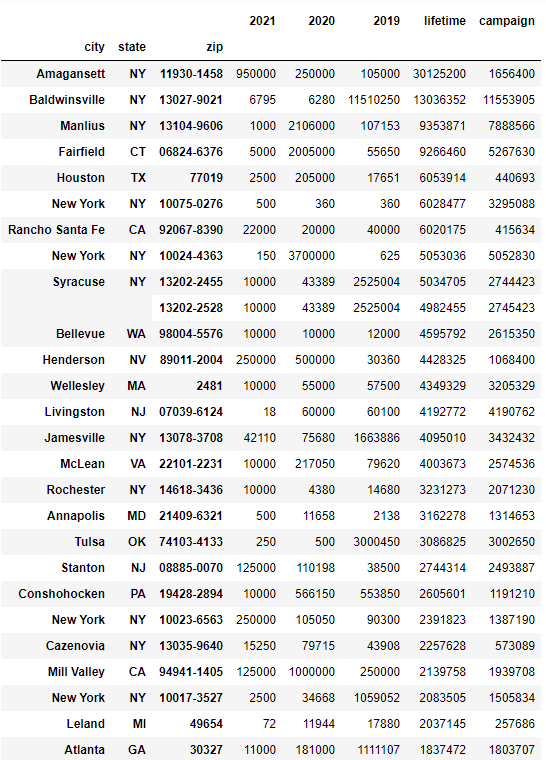
Next, I chose to sort the data frame by lifetime giving (in descending order). Here we can see the recent contributions for those who consistently give and who have committed the most over the course of their relationship with the University.



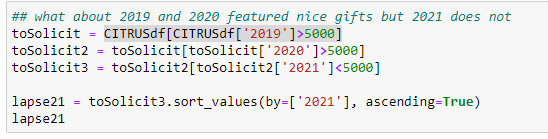
## Top cities by lifetime commitment

Here, I looked at those cities where the most dollars have been raised. Perhaps unsurprisingly, lots of cities in New York are featured. Because the grouping includes zip codes as well as the municipality, this list is quite specific. Due to the small size of the sample, it is likely that many of these zip codes only contain one constituent. While the idea of a “giving by zip code” is interesting and may have some utility, it might be a better application for a larger data set.

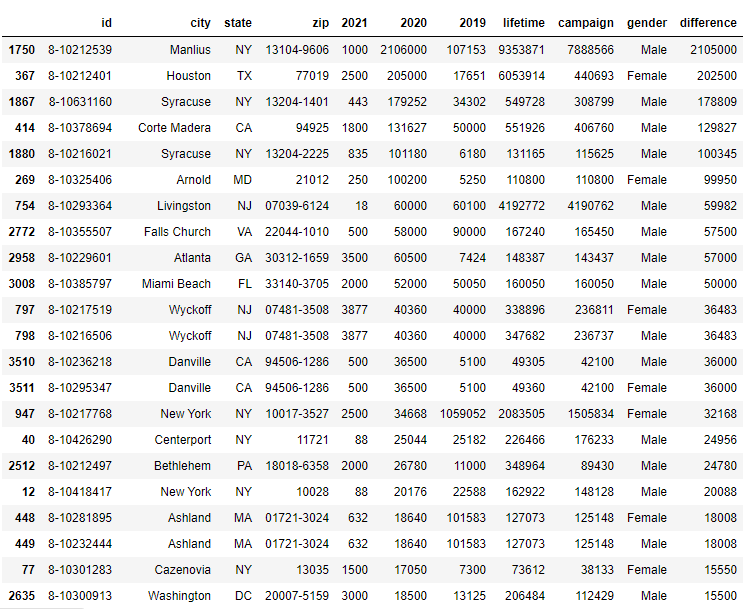
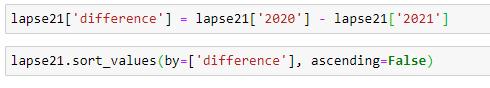




# Constituents ready for solicitation



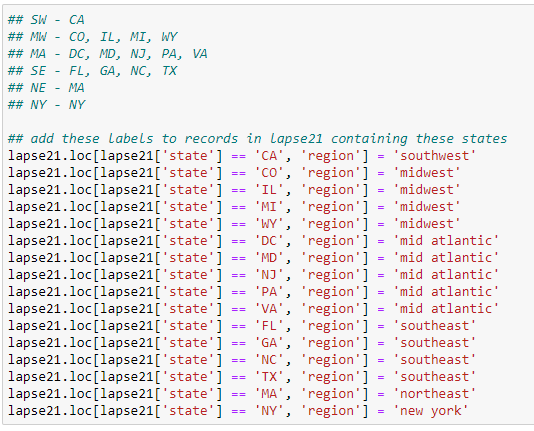
My next undertaking was to create a subset of the data that included individuals who had made significant contributions in fiscal years 2019 and 2020 (greater than $5,000) but had contributed less than $5,000 so far in fiscal year 2021. A great strategy for some of our gift officers might be to solicit these “warm leads” ahead of their annual contribution, either to increase the gift, or to simply ensure that it arrives. Once I had created the subset, I also added a column that subtracted the constituent’s 2021 giving from their 2020 total, calling it “difference.” I then sorted the constituents in descending order by that difference.



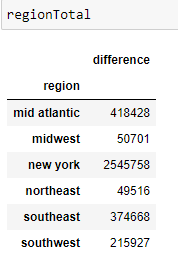
# Regional gift officers

## Opportunity across regions

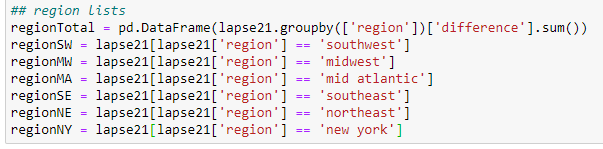
The figure generated in the previous section, the difference between the last two fiscal years, is a valuable piece of information. But to establish a plan moving forward, it would be important to decide which gift officer will solicit whom. In the following section of code, I established a new column (“region”) that buckets groups of states in the geographic footprints of our regional gift officers. I then grouped all constituents by this newly created field in order to view the amount that gift officers might stand to raise (the difference column) in their regions.







## Prospect lists for gift officers



Before hitting the road and raising the funds, the gift officers will need their marching orders. To provide them each with a list of the exact constituents that they ought to solicit, I created new data frames that provide each region separately.

