Capstone Project - Biotech Space (Part 1)

Description of issue and background:

A well-known biotechnology company is considering opening a new location in the Phoenix Metropolitan area. Their main objective is to research the coronavirus, implement clinical trials, and ultimately find a cure to save lives. They prefer to have a location where there have been many confirmed cases, therefore geographic data would be useful. Foursquare can elaborate on the geographic data with nearby locations.

Description of data and usage:

State level data was extracted from the CDC database from the National Center for Health Statistics (NCHS) and the Division of Health Informatics and Surveillance (DHIS). These data sets provide updated statistics on specific cause mortality, confirmed coronavirus cases, demographics on patients, and geographic locations.

1. The first data set sample labeled Provisional Death Counts for Coronavirus Disease (COVID-19): Weekly State-Specific Data Updates and the source can be found here: <https://data.cdc.gov/resource/pj7m-y5uh.json>
2. The second data set sample labeled NNDSS – Table – 1FF. Severe acute respiratory syndrome-associated coronavirus disease to Shigellosis and the source can be found here: <https://data.cdc.gov/resource/x8jf-txib.json>
3. The last data set sample labeled Provisional Death Counts for Coronavirus Disease (COVID-19) includes state level data for deaths related to pneumonia and influenza and the source can be found here: <https://data.cdc.gov/resource/hc4f-j6nb.json>

Reasoning: Each data set was collected with the intention of cleaning and wrangling the data to find the appropriate jurisdictions that are going to be examined. Three data sets were examined to enhance the overall knowledge of the area. Several other factors related to the coronavirus were given in the data set to include all variables, and to provide a more comprehensive analysis.

Limitations: Each data set has delays in reporting, some data is incomplete and it does not provide an entire picture at the state level. Therefore, some data may be underestimated. Another consideration is the level of comorbidities that are associated with the disease. Although this is preliminary data, these factors were not included in the data sets.