

Group 2 Contributors:

Ying Sun
Indranil Roy
Jenna Jorstad
Nina Padgett

Proposal:

In order to have a fair comparison between the US States for COVID-19 infections and deaths, we joined two datasets related to COVID-19 and census data to create a relational database. We took a closer look at the US states, their populations, and their case/death counts during the pandemic. The resulting dataset has two new columns about the percentages of cases and deaths for every state to compare death counts and case counts between each state.

Findings:

Adding in the two columns of Percent Cases and Percent Deaths gave us better insight of how different states were affected during the pandemic. We believe the percentage comparison gives a clearer picture of the number of deaths in comparison to the case counts in more populated /larger states versus less populated states/smaller states.

Methods Used:

Cleaning, joining data, adding columns

Technologies:

Pandas, PostgreSQL

Data Sources:

Data Source #1: COVID-19Activity.csv
<https://query.data.world/s/vk5mf4hyssjgd5iqtolztihap6xjpi>

Data Source #2: Uscensuspop2020.csv

Installation:

Code tested using Python 3.8. The environment also needs pandas and sqlalchemy.
The environment was setup as follows:

```
Import pandas as pd  
From sqlalchemy import create_engine
```