Two data sets were sourced from cdc.gov and NYTimes (via Kaggle) to illustrate the COVID-19 death rates across states and compare to COVID-19 case numbers. This data can also be used to compare COVID death rates against flu and pneumonia death rates by state or country as a whole.

The CDC dataset included COVID-19, Pneumonia, and Influenza deaths for each state for each year from 2020-2022. It also included age groups and sex. We created three dataframes, one for each disease. For each dataset, we filtered it to only include data for 2021 and we used all Age Groups and Sexes. We dropped unused columns leaving us with 'Year', 'State', and 'Number of Deaths' for each dataframe. The NYTimes Kaggle dataset included COVID-19 Deaths and Cases for each state for each year from 2020-2022. We filtered to only include data for 2021 and then removed the Deaths column since that data was previously pulled from the CDC dataset.

Once exported to pgAdmin and formatted into tables with state serving as both a primary and foreign key, we joined the ‘covid\_deaths’ table and ‘covid\_cases’ table. We completed this join in order to computer the death rate per cases for COVID-19. Through dividing COVID-19 deaths by COVID-19 cases and multiplying by 100, we were able to compute this value, which was then stored in a new column called ‘death\_perc.’ We computed this figure because it quantifies the lethality of COVID-19.

Technologies used:

1. Pandas
2. SQL Alchemy
3. Postgres

We decided to create four tables in postgres, in order to load our data in. The first table looked at covid deaths in 2021 across all states in the US. The primary key of our table was the state column, and the following two columns were for year and deaths. Our next table looked at flu deaths in 2021 across all states in the US. Similarly, the primary key from this table was also the state. The following columns were for the year and flu deaths. Our third table looked at total covid cases in 2021 across all states in the US. The primary key again was the state column, and the corresponding columns were for year and total cases. Lastly, our fourth table looked at pneumonia deaths in 2021 across all states in the US. The primary key for this table was, indeed again, the state column.conclusions, indications for use. In conclusion, our project looked at total covid deaths, total covid cases, total pneumonia deaths, total flu deaths for all states within the US in 2021. We were able to clean, load and transform our data successfully into pgadmin. Once we looked our data into pgadmin, we were able to calculate the percentage of deaths by state for COVID. For a further analysis, it would be interesting to investigate subsequent years such as 2022 to have a comparison of covid deaths and covid cases to 2021. Additionally, it would be useful to explore possibly comparing US total covid death and covid cases for 2021 to other countries. All in all, this project helped combine our knowledge of pandas and postgres successfully.