The following describes how I completed my ETL project:

\*E\*xtract: I utilized two data sources – I used Coronavirus and Ebola data from Kaggle and they both were in CSV files.

\*T\*ransform: I cleaned the data by removing columns that I did not intend to analyze. I also updated the format of the dates so that they were formatted the same in both data sets. I also filtered for the united states in both data sets so that I was only looking at cases that happened within the US.

\*L\*oad: the final database I chose was PostgreSQL, and I created a Pandemic database with two tables, a Coronavirus and Ebola table. I chose this because the data was structured so I would not use MongoDB as I would for unstructured data.