ETL (Done through Excel/power Query)

1. ['Death Rates, 2015-2019.txt', 'Death Rates, 2015-2019.csv']
    1. Loaded Death Rates.txt document into an excel using get data
    2. Removed state codes section
    3. Changed null column that contained total into United States using find and replace
    4. Saved as CSV file

2. [2015.csv, 2016.csv, 2017.csv, 2018.csv, 2019.csv, combined_insurance.xlsx]
    1. Renamed location to state
    2. Removed Puerto Rico row using filter
    3. Combined all CSV's mentioned into one file called combined_insurance.xlsx
    4. using the append function in power query
    5. Created another excel worksheet that separates the sheets by year value
        a. Done by just loading in data using get data

3. [SAIPESNC_2015-2019.csv]
    1. Rename column State to State Code - contains state codes
    2. Rename State/County name to State
    3. Remove Columns with Under Age 18, Ages 5 to 17 and Under age 5 in header name
    4. Rename/Replace remaining column headers that have spaces with underscores
    5. Remove United States rows
    6. Filter on State/County Name - where text does not contain left parenthesis to filter down to just states
    7. Remove Column County ID
    8. Remove Blank Rows

4. [numberOfHospitalsByState.csv/Hospital Statistics by State]
    1. Rename/Replace remaining column headers that have spaces with underscores
    2. Remove Rows American Samoa, Guam, Northern Mariana Islands, Puerto Rico, and Virgin Islands
    3. Rename Washington D.C. to District of Columbia
    4. Change Gross_Patient_Revenue($) type from currency to decimal
    5. Split column State by delimiter hyphen (-)
    6. Remove column with state abbreviations
    7. Rename column state.2 to State that contain the state names

5. [api.census.gov]
    1. Call API for company summary (abscs) by state
    2. Filter for rows where NAICS2017 == '62'
        a. Returns Health care and social assistance companies

3. Replace ‘EMP’ column as integer type

6. [Death Rates and Poverty Merged.csv]
    1. Created a workbook called Death Rates and Poverty Merged
    2. Merged Death Rates, 2015-2019.csv and SAIPESNC_2015-2019.csv together
        a. Merged using the State column on both CSV’s in power query
        b. Removed Duplicated columns by using remove column
    3. Saved the current sheet as a CSV

7. [Hospital Death Rates Insurance Poverty Merged.csv]
    1. Used the SAIPESNC_2015-2019.csv and merged all values by years
        a. Merged by creating a pivot table and set row values by the State values
        b. Averaged all values so that it can be merged to other CSV’s afterwards
        c. Saved the sheet as SAIPENC_merged.csv
    2. Used the combined_insurance.xlsx and merged all values by years. Merged by creating a pivot table and set row values by the State values
        a. Averaged all values so it can be merged to other CSV’s afterwards
        b. Saved the sheet as combined_insurance.csv
    3. Merged the following 4 csvs: Death Rates, 2015-2019.csv, numberOfHospitalsByState.csv, SAIPENC_merged.csv, and combined_insurance.csv
        a. Merged through inner join on the State column
        b. Expanded the columns after merge and removed duplicated columns by just using remove column
        c. Saved the finalized dataset as a CSV file to be used for visualizations