Project Title:

ETL - Regional Opioid Death and Unemployment Rates for the United States

Team Name/Member:

Han Solo/Tanner Horton

Project GitHub Link:

https://github.com/tannerhorton/regional_opioid_project

Project Description/Outline:

Data will be extracted from two relational, CSV databases into an interactive python Jupyter Notebook. The first dataset spans 1998-2014 and lists annual opioid-related deaths by state and includes the state's recorded population as well as crude rates. In this instance, the number of deaths occurring in a specified population per 100,00 people. This dataset will first need to be cleaned as there are some years and states with incomplete information. The second dataset holds unemployment data, expressed in percent, from 1990-2016 but is broken down by county and month. Since this database spans a much longer time period, the rows will need to be filtered by year to exclude unneeded data. Although state columns do already exist in this dataframe, the county unemployment percentages will need to be aggregated and averaged before the tables can be joined. *Once the two dataframes have been merged, they will be loaded into a relational PostgreSQL database using SQLAlchemy.*

Datasets to Be Used (need at least two sources):

- *https://www.kaggle.com/datasets/jazzang/opioid-overdose-deaths*
- *(https://data.world/health/opioid-overdose-deaths)*
- *https://www.kaggle.com/datasets/jayrav13/unemployment-by-county-us*
- *https://data.bls.gov/lausmap/showMap.jsp;jsessionid=69D7D1AF451AC910D864E1FFEC668631._t3_08v*

Findings:

Although the unemployment database started out with more than 800,000 unique rows due to its larger time frame (26 years in total) and breakdown by county, the scope of the final merged database was limited by the opioid data, which covered only fifteen years. If given more time to complete this project, per county death tolls would ease the process of establishing whether or not a connection existed between regional unemployment rates and the proliferation of the opioid epidemic for the given years.