

Objective:

To extract data from at least two separate sources and prepare it for later analysis via processes like filtration and aggregation. The data must then be transformed and merged before loading it into an open-source relational database (PostgreSQL).

Project Overview:

According to the Centers for Disease Control, drug overdoses in the United States increased nearly three-fold during the first fifteen years of the twenty first century. By 2014, over half of all drug overdose deaths involved an opioid. There have been many studies linking this rise to doctor prescribed pain medications such as Hydrocodone (Vicodin®), Oxycodone (OxyContin®, Percocet®), Oxymorphone, and Morphine among others. Studies have deemed communities with high poverty rates as more susceptible to these types of overdoses. In this project, data has been prepared for potential analysis to determine if a relationship exists between state unemployment rates and the annual number of opioid-related deaths for the years 1999-2014.

Project Structure:

In an effort to regionalize this database synthesis, two CSV files were pulled from kaggle.com, the first of which focuses on annual opioid-related deaths for 1999-2014. It is broken down by state and includes state populations and crude rates or, the number of occurrences in a specified population per 100,00 people. This dataset had to be cleaned as there are several states incomplete death-toll numbers or unreliable crude rates due to population limitations.

The second dataset concentrates unemployment rates, expressed as a monthly percent per county. In order to join this data, which spans 1990-2016, the county and month totals had to be aggregated and averaged into state and year totals respectively. Furthermore, since this database spans a much longer time-period, the rows then had to be filtered to drop incongruous data in preparation for the merge. Once complete, the newly created relational database was loaded into PostgreSQL using SQLAlchemy.

Methodology:

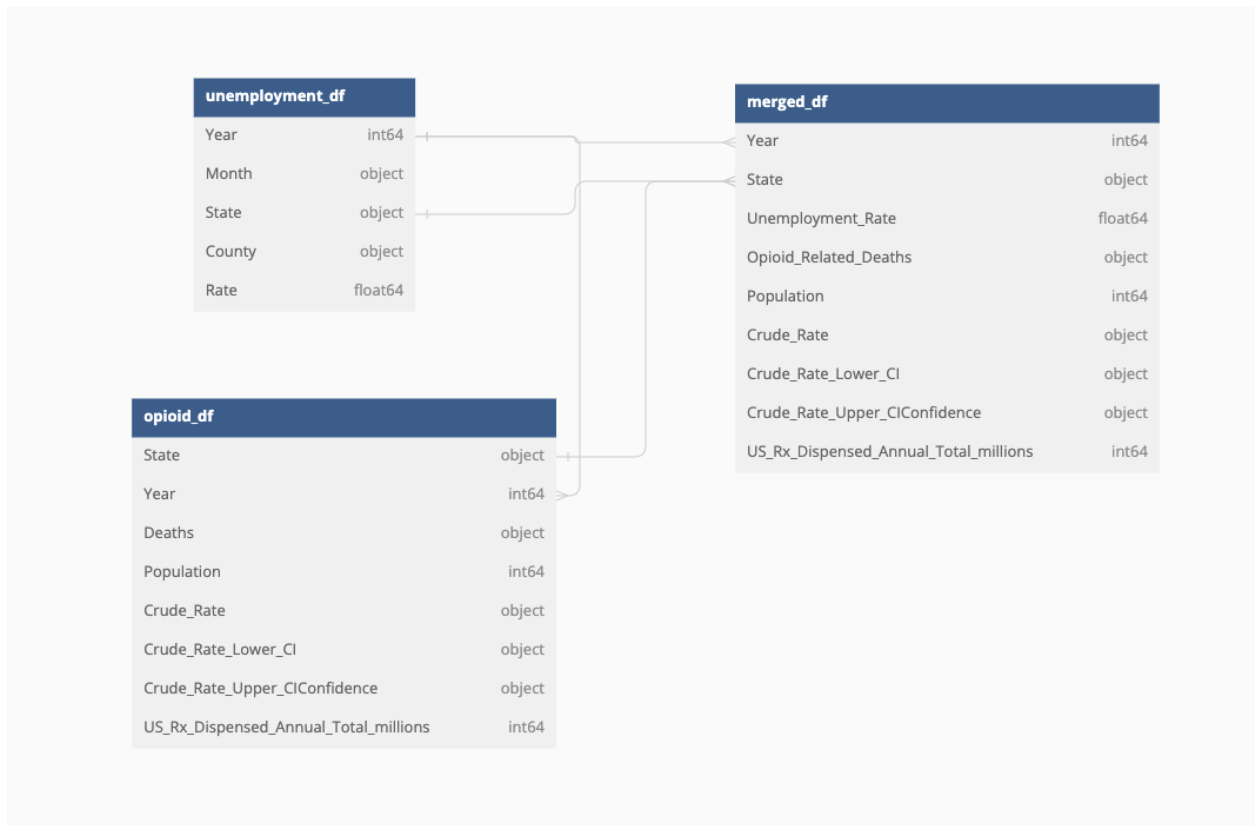
Using Python and Pandas, the downloaded CSV files were read into an interactive notebook where the data was cleaned. SQLAlchemy was then used to connect to a local database and upload the newly created database into PostgreSQL.

Sources:

- <https://www.kaggle.com/datasets/jazzang/opioid-overdose-deaths>

- Which references: <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

Schemata:



Opioid Database Preview:

	A	B	C	D	E	F	G	H
1	State	Year	Deaths	Population	Crude Rate	Crude Rate Lower 95%	Crude Rate Upper 95%	Prescriptions Dispensed by US Retailers in that year (millions)
2	Alabama	1999	39	4430141	0.9	0.6	1.2	116
3	Alabama	2000	46	4447100	1	0.8	1.4	126
4	Alabama	2001	67	4467634	1.5	1.2	1.9	138
5	Alabama	2002	75	4480089	1.7	1.3	2.1	142

Credit: <https://www.kaggle.com/datasets/jazzang/opioid-overdose-deaths>

Unemployment Database Preview:

	A	B	C	D	E
1	Year	Month	State	County	Rate
2	2015	February	Mississippi	Newton County	6.1
3	2015	February	Mississippi	Panola County	9.4

<https://www.kaggle.com/datasets/jayrav13/unemployment-by-county-us>