**Drug overdose deaths in Connecticut from 2012 to 2020**

**Introduction**

This project was done to analyze the drug overdose death in Connecticut, USA from 2012 to 2020. The purpose of this project was to build a database that helps in analyzing the data like deathrate, population based on drug overdose death and census.The database is built on by ingesting, combining, and restructuring data from two main data sources into a conformed one PostgreSQL database. The two sources of our data are Kaggle for drug overdose data and [Census API](https://growjo.com/company_api) for scraping census data for different counties in Connecticut.

**Data Extraction**

In this project we extracted, transformed, and loaded drug overdose death and census data for Connecticut, USA from 2012 to 2020.

Our main sources :

* Drug overdose data from 2012 to 2020

Kaggle

* Census Data from Census API
* Census API

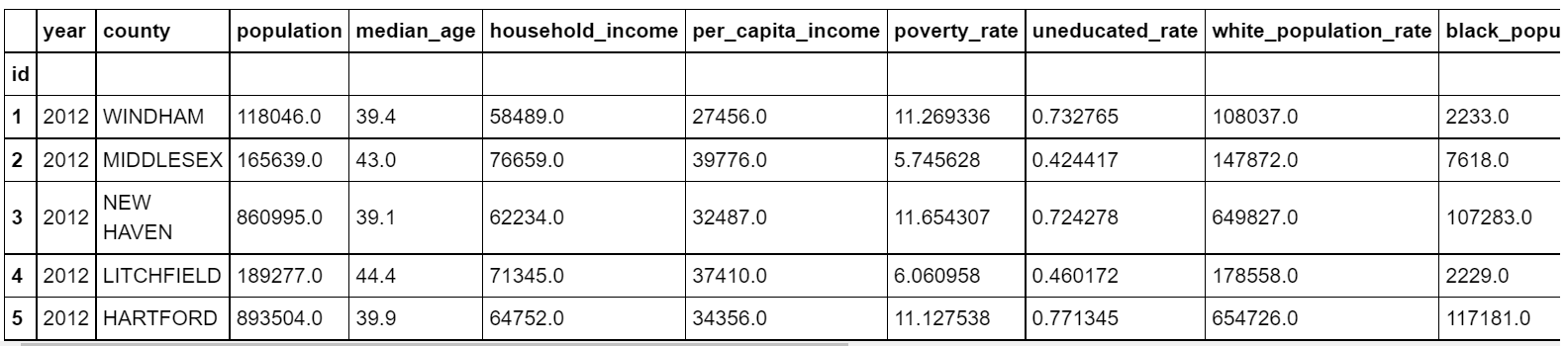
### Data Transformation

* We used a Pandas functions in Jupyter Notebook to transform all CSV files and API request responses.
* We reviewed the files and transformed into a dataframes.
* Filtered out unrelated columns.
* Divided a single column to two different columns.
* Renamed column names to match database schema.
* Removed rows with Null/Nan values.
* Multiple rows were merged into single one to avoid duplicates.
* Set one column to index.
* We conducted some aggregation to find totals for comparison in the datasets.

### Load

* For our final production, we used a relational database called Postgresql.
* We created a database called ‘drug\_connecticut\_db’.
* It has two tables census\_ct and drug\_type.
* drug\_type has 9 columns and census\_ct has 14 columns and the data loaded successfully.
* Final tables/collections are stored in the production database.

### Census table

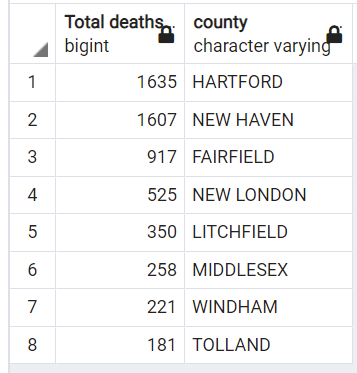


### Drug table

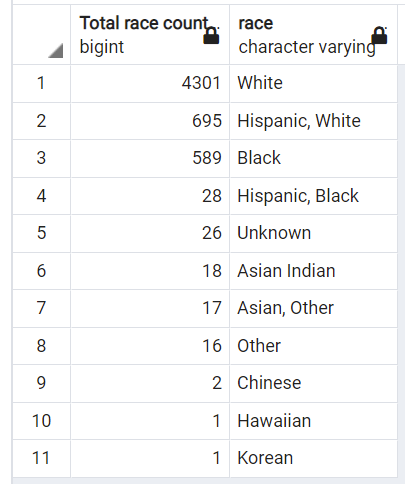
### 

### Aggregate

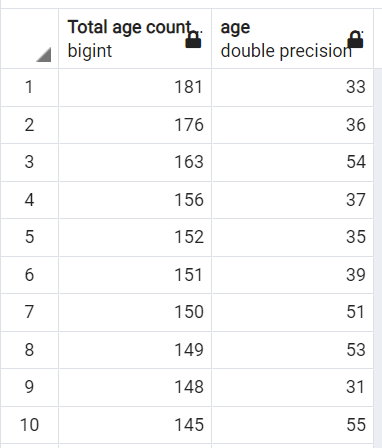
* Total number of deaths per county.



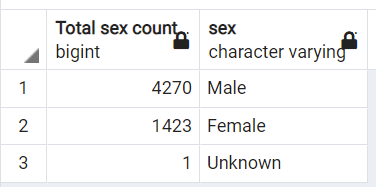
* Total number of deaths per race.



* Total number of deaths per age.



* Total number of deaths per sex.



* Deathrate per county

### Data Analysis

### We can analyze from tables that the greatest number of deaths occurred in Hartford County among whites within age group 33 and Male.

### Team Members