**Purpose:** Take our knowledge of the ETL process and implement it into a real world scenario.

**Sources of the data:**

- Police Deaths (https://www.kaggle.com/mysarahmadbhat/police-deaths?select=clean_data.csv)
- 2015 Annual Survey of State Government Finances Tables (https://www.census.gov/data/tables/2015/econ/state/historical-tables.html)
- US Police Shootings (https://www.kaggle.com/ahsen1330/us-police-shootings)
- Census python module for the US Census API (https://www.census.gov/data/tables/2015/econ/state/historical-tables.html)

**Extraction & Transformation of the data:**

- **Police Deaths:**
    - This data was downloaded from Kaggle as a CSV file.
    - Once imported I used the loc function to extract data for the year of 2015 and gunfire as the cause of death.
    - Grouped data by state to make data linkable.
    - Changed column headers to make data outcome more understandable.
    - Dropped Puerto Rico (PR), Tribal Data (TR), and United States (US) data to make data comparable to the other datasets.
- **State_info table:**
    - 2015 ASFIN State Totals:
        - The data was downloaded from the US Census website as an excel file.
        - Reformatted the excel to transpose the rows & columns so that the state could be set as the index with the corresponding information in columns and then reformatted it as a csv to be imported into a pandas DF.
        - Once imported, pulled out only columns relevant to state general expenditures where it would be within the discretion of lawmakers with potential relation to crime/policing. Re-named those columns as well for clarity and formatting consistency.
        - The data imported as string objects so it was converted to numeric using pandas
        - Iterated through columns & rows to run transformations on the raw expenditure values and added new columns with the per capita expenditures and expenditures as percentage of total expenditures for better comparison across states
    - 2015 US Census State demographic info:

        - Used the US Census python module to request relevant demographic data regarding population, race, and poverty by state
        - for year 2015 from the US Census API and then imported that data into a pandas DF.
        - The columns were renamed for clarity.
        - Normalized the poverty count into a poverty % rate and population counts into percentages of total population for each state for better comparison across the states and added as a new column.
    - Merged the two tables together in pandas so population can be used to normalize the state budgetary info on a per capita basis. This will allow better comparison across states.
- **US Police Shootings:**
    - This data was downloaded from Kaggle (US Police Shootings) as a csv file.
    - Once this was imported into the jupyter notebook, I used "pd.options.display.max_columns" to get a clear look at all the headers from the csv in my pandas display.
    - Used the shape function to calculate how many columns were in the dataframe.
    - After acquiring the numbers of columns present, drop any unnecessary columns using the column numbers. (Checked results.)
    - Polished the data frame by renaming the remaining columns so that it would be more comprehensible.

**Type of final production database data is loaded into:**

- We used a relational database (PostgreSQL) to store our datasets in one place.

**Final tables/collections that we used in the production database:**

- Compare the number of police deaths caused by gunfire to the number of police shootings that end in fatality for each state.
- Compare the Police protection expenditure to the number of police shootings that end in fatality for each state.
    - Do states that spend a larger proportion of their expenditures have higher # of fatalities?
- Compare the percentages of fatalities by race for each state with the overall state population.
    - Are the relative race breakouts equivalent between the killings and the overall population?
- Compare the public welfare expenditures by state with the police fatalities.
    - Do states that spend a larger percentage of their expenditures on welfare have a lower rate of police killings?

- Compare the poverty levels by state with police fatalities
    - Do poorer states have higher rates of police killings?