**Isaac Carranza Project 2: ETL**

**Zoey Charif**

Our objective was to perform data engineering on police shootings data. We collected data from the following sources:

* https://github.com/washingtonpost/data-police-shootings
* https://www.kaggle.com/brendanhasz/fatal-police-shootings?select=CityLocations.cs

**EXTRACT:**

* imported both files to Jupyter Notebook
* converted files to separate Pandas dataframes.

**TRANSFORM:**

* To clean our data, we utilized the following functions: rsplit, join, merge, drop, dropna, rename, isnull.sum
* We saved our newly structured data into CSV and JSON files for future use.

**LOAD:**

* We loaded our updated dataframe into Postgres SQL (relational db) and MongoDB (non-relational db).
* To load to Postgres, we first created a database inside PGAdmin called “police\_shootings”. We created a connection between our Jupyter Notebook and our SQL database. We then queried the table to verify all data was properly loaded.
* To load to MongoDB, we used 2 processes. Firstly, we utilized MongoDB Compass to manually upload the new CSV file. Our second method was to use the insert\_many() function to import our dataframe directly from Pandas. Both functions adequately transferred the data to MongoDB.