ETL Project

Goodman, Paul, Ryan & Kayla

1. Purpose:
   1. To collect energy usage and production data (USA) compared to population data per state.
2. Datasets
   1. US Energy Information Admin
      1. State Total Energy Rankings, 2018:

Link: <https://www.eia.gov/state/?sid=US>

* + 1. Primary Energy, Electricity, and Total Energy Price Estimates, 2018

Link: <https://www.eia.gov/state/seds/data.php?incfile=/state/seds/sep_sum/html/sum_pr_tot.html&sid=US>

* 1. Census Bureau
     1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019 (NST-EST2019-01)

Link: <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>

1. Steps
   1. Validate data
   2. Extract Data
   3. Clean in Jupyter Notebook
      1. Only looking at the year 2018
         1. Dropping all other years from dataset
      2. Combining into on data frame
   4. Load section to Postgres
   5. ERD
   6. Report