**Data Cleanup & Analysis**

Once you have identified your datasets, perform ETL on the data. Make sure to plan and document the following:

* The sources of data that you will extract from.
  + <https://data.ca.gov/dataset/directory-of-service-providers1> (csv)
  + Data.World
  + Hud
* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).
  + Reduce columns to focus on County population, homeless population by county, number of available beds.
  + Reduce rows
    - Filtering using 2018 data only where applicable
  + Resolve many-to-one between COC number and County Name
* The type of final production database to load the data into (relational or non-relational).
  + SQL/relational
* The final tables or collections that will be used in the production database.
  + CA population by county

You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process.

**Project Report**

At the end of the week, your team will submit a Final Report that describes the following:

* **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).
  + <https://data.ca.gov/dataset/directory-of-service-providers1> (csv)
* **T**ransform: what data cleaning or transformation was required
  + Remove columns to narrow scope
  + Remove rows; focus on CA counties
* **L**oad: the final database, tables/collections, and why this was chosen.