ETL Project

Analysis of the contaminants of Striped Bass and Common Carp that reside in California

Fatima Donato

10/06/20

**Extraction**

The datasets are extracted from California Environmental Data Exchange Network (CEDEN), the records compiled are both from the government and private sectors. These records date back from 1977 until 2019.

1. <https://ceden.waterboards.ca.gov/AdvancedQueryTool>

**Transformation**

1. The extracted files are in extensive xlsx formats, several records are dropped to accommodate data loss from the conversion of xlsx to csv.
2. Records with null values are dropped for data uniformity, to avoid complexity and issues with constraints such as primary key and foreign key: these keys may not contain null values. This process reduced the data from almost 15,000 records to 4,300.
3. Columns are converted to lower case letters, postgres does not allow capitalization on column matching.

**Load**

1. The analyzed data in jupyter notebook is connected to postgres path and engine specification. Pandas is utilized to load the converted data frame into postgres, it is also used to confirm the connection by the use of query.
2. Two tables area created in postgres to accommodate both the Striped Bass and Common Carp data frames.