**ETL Project**

**Data to be Extracted:** Population for each Ohio County

**Data Source:** Wikipedia info page for Ohio Counties:

<https://en.wikipedia.org/wiki/List_of_counties_in_Ohio>

**Extraction:** Since the data was provided in a table on the Wikipedia website, Pandas was used to extract the table. After cleaning the table (described below), the table was saved as a CSV file for upload into the relational database.

**Transformation (Cleaning):** The table extracted had several additional columns that were not needed so they were deleted (county seat, establishment date, origin, etc.) Only the following rows were kept: County Name, County Code, and Population. County Code was originally named FIPS Code but it was changed to match other data sources. The County Name field had ‘County’ after each county name and that was also removed to match other data sources.

**Load:** The table was loaded into PgAdmin along with the other data collected as part of the relational database.