**Final Report of ETL Project**

1. Data sources
   1. Bureau of Economic Analysis (BEA) of U.S. Department of Commerce (API, JSON)
   2. The Safe Drinking Water Information System (SDWIS) of United States Environmental Protection Agency (csv)
   3. FIPs code information that scraped from website
2. I first got the personal income (GDP), energy and economics data from the BEA’s website through the query of API.
3. Then got the water contamination/violation data from the SDWIS system. All the data sources were converted into csv file with python and loaded to MySQL.
4. The data sources including personal income for all counties, states energy profiles and drinking water violation cases through all united states.
5. I chose the relational database because the different tables correlate together with certain columns such as states, counties or FIPs code. Combining the different tables with groupby or filter function in SQL could give us some insights on understanding the relationship between economic development, states energy profile and drinking water pollution status.
6. Drinking water quality is often dependent on the wealth and racial makeup of communities. Small, poor communities and neglected urban areas are sometimes left to fend for themselves with little help from state and federal governments. While, areas with heavy industries always correlate with heavy water pollution problems. I want to get concepts of how economic development related to contamination and drinking water clean level through United States.