ETL Project Write-Up

**Subject:** Mortality Rates

**Question:** How do mortality rates compare between insured individuals and the general population in the United States?

**Data Sources:** <https://mort.soa.org/> (For insured individuals)

<https://api.census.gov/data/1990/pep/int_natcivpop> (For general population)

**Steps**

**Extract (Randall: CSV & JSON)**

* Put CSV files in a directory.
* Sign up for Census API.
* Make API call in Jupyter Notebook.
* Use Pandas.read\_csv to pull in CSV data.

**Transform (Ben: CSV, Sasmita: JSON)**

* Ben: Create Pandas dataframes of CSV files.
* Ben: Clean Pandas CSV dataframes.
* Sasmita: Create Pandas dataframe of JSON dictionaries.
* Sasmita: Clean Pandas JSON dataframe.
* Sasmita: Join the Pandas dataframes.
* Ben: Create equivalent tables in PgAdmin and assign Age as Primary Key.

**Load (Sasmita)**

* Use Pandas.to\_sql to load the joined table to the PostgreSQL database.