

# ETL Project Report

Correlating between Population Density per State and GDP per Capita

Alan McCabe || Siraj Mohammed || Chad Davis

## Extract:

- Original data sources came from the following Websites:
  - [https://en.wikipedia.org/wiki/List\\_of\\_states\\_and\\_territories\\_of\\_the\\_United\\_States\\_by\\_population\\_density](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density)
  - [https://en.wikipedia.org/wiki/List\\_of\\_U.S.\\_states\\_by\\_GDP\\_per\\_capita](https://en.wikipedia.org/wiki/List_of_U.S._states_by_GDP_per_capita)
- The data from both sites was in a Table Format and We used 2 Different Methods to extract the data.
  - We used Web Scraping on one Site.
  - Downloaded the Data to a CSV File from the other site.
  - Pulled population density of states of the United States from Wikipedia using read\_html
  - Pulled GDP of states of the United States from Wikipedia and captured as a CSV file
  - Read\_html provided multiple tables. Table needed was at index 0.
  - The table returned had a tuple for table headers. Renamed columns for easier management of data frame.
  -

## Transform:

- Determined which columns were needed
- One row had a footnote tag which caused erroneous data e.g. 155,959[5] which needed to be cleaned
- Targeted year 2015 to match data from the population density data and renamed to 'gdp'

## Load:

- Connect to postgres DB
- Confirm tables
- Load data into states\_db.states\_pop and states\_db.states\_gdp
- Query DB with an inner join to bring data together
- We chose only the information below to show Population Density per State and GDP per Capita

|   | state                | gdp    | den_per_sqm | pop_numbers | land_sqm |
|---|----------------------|--------|-------------|-------------|----------|
| 0 | District of Columbia | 159497 | 11011       | 672228      | 61       |
| 1 | New Jersey           | 55750  | 1218        | 8958013     | 7354     |
| 2 | Rhode Island         | 46356  | 1021        | 1056298     | 1034     |