

## About Dataset

# United States Energy, Census, and GDP 2010-2014

Data set comes from Kaggle: [https://www.kaggle.com/datasets/lislejoem/us\\_energy\\_census\\_gdp\\_10-14/data](https://www.kaggle.com/datasets/lislejoem/us_energy_census_gdp_10-14/data)

The purpose of this data set is to allow exploration between various types of data that is commonly collected by the US government across the states and the USA as a whole. The data set consists of three different types of data:

- Census and Geographic Data;
- Energy Data; and
- Economic Data.

When creating the data set, I combined data from many different types of sources, all of which are cited below. I have also provided the fields included in the data set and what they represent below. I have not performed any research on the data yet, but am going to dive in soon. I am particularly interested in the relationships between various types of data (i.e. GDP or birth rate) in prediction algorithms. Given that I have compiled 5 years' worth of data, this data set was primarily constructed with predictive algorithms in mind.

*An additional note before you delve into the fields:*

- There could have been many more variables added across many different fields of metrics. I have stopped here, but it could potentially be beneficial to observe the interaction of these variables with others (i.e. the GDP of certain industries, the average age in a state, the male/female gender ratio, etc.) to attempt to find additional trends.

## Census and Geographic Data

- StateCodes: The state 2-letter abbreviations. **Note that I added "US" for the United States.**
- Region: The number corresponding to the region the state lies within, according to the 2010 census. (1 = Northeast, 2 = Midwest, 3 = South, 4 = West)
- Division: The number corresponding to the division the state lies within, according to the 2010 census. (1 = New England, 2 = Middle Atlantic, 3 = East North Central, 4 = West North Central, 5 = South Atlantic, 6 = East South Central, 7 = West South Central, 8 = Mountain, 9 = Pacific)
- Coast: Whether the state shares a border with an ocean. (1 = Yes, 0 = No)
- Great Lakes: Whether the state shares a border with a great lake. (1 = Yes, 0 = No)
- CENSUS2010POP: 4/1/2010 resident total Census 2010 population
- POPESTIMATE{year}: 7/1/{year} resident total population estimate
- RBIRTH{year}: Birth rate in period 7/1/{year - 1} to 6/30/{year}
- RDEATH{year}: Death rate in period 7/1/{year - 1} to 6/30/{year}

- **RNATURALINC{year}**: Natural increase rate in period 7/1/{year - 1} to 6/30/{year}
- **RINTERNATIONALMIG{year}**: Net international migration rate in period 7/1/{year - 1} to 6/30/{year}
- **RDOMESTICMIG{year}**: Net domestic migration rate in period 7/1/{year - 1} to 6/30/{year}
- **RNETMIG{year}**: Net migration rate in period 7/1/{year - 1} to 6/30/{year}

*As noted from the census:*

Net international migration for the United States includes the international migration of both native and foreign-born populations. Specifically, it includes: (a) the net international migration of the foreign born, (b) the net migration between the United States and Puerto Rico, (c) the net migration of natives to and from the United States, and (d) the net movement of the Armed Forces population between the United States and overseas. Net international migration for Puerto Rico includes the migration of native and foreign-born populations between the United States and Puerto Rico.

Codes for most of the data, information about the geographic terms and conditions, and more information about the methodology behind the population estimates can be found on the US Census website.

## Energy Data

- **TotalC{year}**: Total energy consumption in billion BTU in given year.
- **TotalP{year}**: Total energy production in billion BTU in given year.
- **TotalE{year}**: Total Energy expenditures in million USD in given year.
- **TotalPrice{year}**: Total energy average price in USD/million BTU in given year.
- **TotalC{first year}—{second year}**: The first year's total energy consumption divided by the second year's total energy consumption, times 100. (The percent change between years in total energy consumption.)
- **TotalP{first year}—{second year}**: The first year's total energy production divided by the second year's total energy production, times 100. (The percent change between years in total energy production.)
- **TotalE{first year}—{second year}**: The first year's total energy expenditure divided by the second year's total energy expenditure, times 100. (The percent change between years in total energy expenditure.)
- **TotalPrice{first year}—{second year}**: The first year's total energy average price divided by the second year's total energy average price, times 100. (The percent change between years in total energy average price.)
- **BiomassC{year}**: Biomass total consumption in billion BTU in given year.
- **CoalC{year}**: Coal total consumption in billion BTU in given year.
- **CoalP{year}**: Coal total production in billion BTU in given year.
- **CoalE{year}**: Coal total expenditures in million USD in given year.
- **CoalPrice{year}**: Coal average price in USD per million BTU in given year.
- **ElecC{year}**: Electricity total consumption in billion BTU in given year.
- **ElecE{year}**: Electricity total expenditures in million USD in given year.
- **ElecPrice{year}**: Electricity average price in USD per million BTU in given year.
- **FossFuelC{year}**: Fossil fuels total consumption in billion BTU in given year.

- GeoC{year}: Geothermal energy total consumption in billion BTU in given year.
- GeoP{year}: Geothermal energy net generation in the electric power sector in million kilowatt hours in given year.
- HydroC{year}: Hydropower total consumption in billion BTU in given year.
- HydroP{year}: Hydropower total net generation in million kilowatt hours in given year.
- NatGasC{year}: Natural gas total consumption (including supplemental gaseous fuels) in billion BTU in given year.
- NatGasE{year}: Natural gas total expenditures in million USD in given year.
- NatGasPrice{year}: Natural gas average price in USD per million BTU in given year.
- LPGC{year}: LPG total consumption in billion BTU in given year.
- LPGE{year}: LPG total expenditures in million USD in given year.
- LPGPrice{year}: LPG average price in USD per million BTU in given year.

*Notes:*

- BTU stands for British Thermal Unit and is a unit of measurement for energy. One BTU is equal to the amount of energy used to raise the temperature of one pound of water on degree Fahrenheit.
- Many other types of energy and their associated consumption, production, expenditure, and price totals can be found from the EIA; this is where I received the data I used in compiling this dataset.

## **Economic Data**

- GDP{year}{quarter}: The GDP in the provided quarter of the given year (in million USD).
- GDP{year}: The average GDP throughout the given year (in million USD).

*Notes:*

- The GDP is reported by the Bureau of Economic Analysis from the U.S. Department of Commerce and measures the value of the goods and services produced by the economy in a given period.
- The quarterly GDP data can be downloaded from the BEA.
- The yearly GDP data can be downloaded from the BEA.