Description

The goal is to predict the rate of heart disease (per 100,000 individuals) across the United States at the county-level from other socioeconomic indicators. The data is compiled from a wide range of sources .

There are 33 variables in this dataset. Each row in the dataset represents a United States county, and the dataset we are working with covers two particular **years**, denoted **a**, and **b** We don't provide a unique identifier for an individual county, just a **row\_id** for each row.

The variables in the dataset have names that of the form **category\_\_variable**, where category is the high level category of the variable (e.g. econ or health). variable is what the specific column contains. Both training and test dataset are from same population.

**We're trying to predict the variable heart\_disease\_mortality\_per\_100k (a positive integer), for each row of the test data set.**

**Attributes**

**area — information about the county**

1. **area\_\_rucc** — Rural-Urban Continuum Codes "form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area.
2. **area\_\_urban\_influence** — Urban Influence Codes "form a classification scheme that distinguishes metropolitan counties by population size of their metro area, and nonmetropolitan counties by size of the largest city or town and proximity to metro and micropolitan areas."

**econ — economic indicators**

1. **econ\_\_economic\_typology** — County Typology Codes "classify all U.S. counties according to six mutually exclusive categories of economic dependence and six overlapping categories of policy-relevant themes. The economic dependence types include farming, mining, manufacturing, Federal/State government, recreation, and nonspecialized counties.
2. **econ\_\_pct\_civilian\_labor** — Civilian labour force, annual average, as percent of population
3. **econ\_\_pct\_unemployment** — Unemployment, annual average, as percent of population
4. **econ\_\_pct\_uninsured\_adults** — Percent of adults without health insurance

health — health indicators

1. **health\_\_pct\_adult\_obesity** — Percent of adults who meet clinical definition of obese
2. **health\_\_pct\_adult\_smoking** — Percent of adults who smoke
3. **health\_\_pct\_diabetes** — Percent of population with diabetes
4. **health\_\_pct\_low\_birthweight** — Percent of babies born with low birth weight
5. **health\_\_pct\_excessive\_drinking** — Percent of adult population that engages in excessive consumption of alcohol
6. **health\_\_pct\_physical\_inacticity** — Percent of adult population that is physically inactive
7. **health\_\_air\_pollution\_particulate\_matter** — Fine particulate matter in µg/m³
8. **health\_\_homicides\_per\_100k** — Deaths by homicide per 100,000 population
9. **health\_\_motor\_vehicle\_crash\_deaths\_per\_100k** — Deaths by motor vehicle crash per 100,000 population
10. **health\_\_pop\_per\_dentist** — Population per dentist (HRSA Area Resource File)
11. **health\_\_pop\_per\_primary\_care\_physician** — Population per Primary Care Physician

**demo — demographics information**

1. **demo\_\_pct\_female** — Percent of population that is female
2. **demo\_\_pct\_below\_18\_years\_of\_age** — Percent of population that is below 18 years of age
3. **demo\_\_pct\_aged\_65\_years\_and\_older** — Percent of population that is aged 65 years or older
4. **demo\_\_pct\_hispanic** — Percent of population that identifies as
5. **demo\_\_pct\_non\_hispanic\_african\_american** — Percent of population that identifies as African American
6. **demo\_\_pct\_non\_hispanic\_white** — Percent of population that identifies as Hispanic and White
7. **demo\_\_pct\_american\_indian\_or\_alaskan\_native** — Percent of population that identifies as Native American
8. **demo\_\_pct\_asian** — Percent of population that identifies as Asian
9. **demo\_\_pct\_adults\_less\_than\_a\_high\_school\_diploma** — Percent of adult population that does not have a high school diploma
10. **demo\_\_pct\_adults\_with\_high\_school\_diploma** — Percent of adult population which has a high school diploma as highest level of education achieved
11. **demo\_\_pct\_adults\_with\_some\_college** — Percent of adult population which has some college as highest level of education achieved
12. **demo\_\_pct\_adults\_bachelors\_or\_higher** — Percent of adult population which has a bachelor's degree or higher as highest level of education achieved
13. **demo\_\_birth\_rate\_per\_1k** — Births per 1,000 of population
14. **demo\_\_death\_rate\_per\_1k** — Deaths per 1,000 of population

**Expectations**

**1) PowerPoint presentation Max.10 slides (excluding, introduction slide).**

**2) Send a text document where code is written or the Project file.**

**3) If possible, use different algorithms and compare results.**

**4) Deadline for submission is 1 November 2019, 11:59 PM.**

**5) Any doubts please mail us.**