

## Dataset

The dataset `ny-demographics.csv` contains information on the residential demographics of each census tract in New York state. The dataset contains the following variables:

<code>geoid11</code>	11-digit geographic identifier for census tract
<code>geoid11name</code>	Name of census tract, county, and state
<code>population</code>	Number of residents in census tract
<code>asian</code>	Number of tract residents who are non-Hispanic Asians
<code>black</code>	Number of tract residents who are non-Hispanic blacks
<code>hispanic</code>	Number of tract residents who are Hispanic
<code>white</code>	Number of tract residents who are non-Hispanic whites

A row in the dataset describes one census tract. For example, the row that begins

```
geoid11      geoid11name
36001000100  Census Tract 1, Albany County, New York
```

indicates that there are 2139 residents in Census Tract 1 of Albany County, New York, of whom 55 are non-Hispanic Asians.

There are 4919 census tracts in New York state and there are no missing values in the dataset. The “`geoid11`” variable has the property that the first 2 digits identify the state (36 = New York) and the first 5 digits identify the county within the state (36001 = Albany, New York). Tracts partition a county: if you add up all the residents in the 75 tracts in Albany County, this equals the population of Albany County (304,204 residents). There are nine tracts in which the population variable is the population count followed by “(rXXXXX)”, where XXXXX is a 5-digit revision number, indicating that the Census Bureau revised the population count at some point after the initial data release.

## Assignment

Your assignment is to write a short script that generates a county-level dataset describing each New York county’s demographics. In particular, please produce a CSV file containing the following variables:

<code>geoid5</code>	5-digit geographic identifier for county
<code>geoid5name</code>	Name of county and state
<code>population</code>	Number of residents in county
<code>asian_share</code>	Fraction of county residents who are non-Hispanic Asians
<code>black_share</code>	Fraction of county residents who are non-Hispanic blacks
<code>hispanic_share</code>	Fraction of county residents who are Hispanic
<code>white_share</code>	Fraction of county residents who are non-Hispanic whites
<code>tracts</code>	Number of census tracts in county with positive population
<code>asian_majority_tracts</code>	Number of tracts in county where > 50% of residents are non-Hispanic Asians
<code>black_majority_tracts</code>	Number of tracts in county where > 50% of residents are non-Hispanic blacks
<code>hispanic_majority_tracts</code>	Number of tracts in county where > 50% of residents are Hispanic
<code>white_majority_tracts</code>	Number of tracts in county where > 50% of residents are non-Hispanic whites
<code>nomajority_tracts</code>	Number of tracts in county where no demographic category has > 50% share

The dataset should include one observation for each of New York’s 62 counties and should be sorted by the 5-digit code that identifies the county. If you need to make judgment calls about how

to process the data, please write us a short note describing the decisions you made.

### **Evaluation**

We will evaluate your submitted assignment based on four criteria:

- Is the county-level dataset correct?
- Is the code clean, easy to read, and sufficiently commented?
- Is the computation performed efficiently?
- How much time was required to produce the code?

For the sake of the final criterion, please accurately report the total time you spend on the assignment, commencing when you first read through these instructions.