

HCSC Code Screen

Problem statement

As part of HCSC's COVID19 response, the Data Science team needs to prepare daily/weekly updates of nationwide infection counts, organized by county. We use numeric FIPS code https://en.wikipedia.org/wiki/FIPS_county_code rather than stand and county name to serve our results

For every FIPS code and date, your end user will want to know: population, daily cases, daily deaths, cumulative cases to date, and cumulative death counts to date.

Create a repeatable pipeline that merges these two datasets into a single Dataframe and output file.

Your pipeline script will be scheduled to run daily and must provide the latest COVID-19 data available at time of execution.

New York Times COVID-19 Data

Source	https://github.com/nytimes/covid-19-data/blob/master/README.md
Download Link	https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv

2019 Population Estimate Data

Source	https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html
File Layout	https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010-2019/co-est2019-alldata.pdf
Download Link	https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/totals/co-est2019-alldata.csv

Rules of Engagement

- Submission will be critiqued on design, readability, and cleanliness of your code.
- Please use Python. You can use any open source python packages. Using pandas is recommended.
- Your code must work as you document it in a README.md file; we will not fiddle with it to make it work.
- Consider adding to the README.md
 - Any learnings or thoughts about the data sources
 - How you sanity checked your output
 - Things you would add given more time
- Have fun! This challenge is meant as a way for you to show your skills and start a conversation with the team about how to approach solving problems.