UCLA School of Law COVID-19 Behind Bars Data Project

Dataset Description for MIT Policy Hackathon

The enclosed csv files include data describing COVID conditions in the correctional institutions of New Mexico, Washington, Arkansas, and Pennsylvania. This data has been "web scraped" from the COVID dashboard updates from these state's Departments of Corrections.

The following variables are present in each of the datasets:

Facility: The type of facility or entity reported. These designations are internally produced by the project and have not been checked so should not be relied upon.

Name: The name of the facility or entity for which the reporting numbers.

State: The state reporting the associated COVID data.

Date: The date the data was collected.

Residents.Confirmed: The number of confirmed COVID cases amongst inmates for each facility.

Staff.Confirmed: The number of confirmed COVID cases amongst staff for each facility.

Resident.Deaths: The number of confirmed COVID deaths amongst inmates for each facility.

Staff.Deaths: The number of confirmed COVID deaths amongst staff for each facility.

Residents.Recovered: The number of confirmed COVID recoveries amonst inmates for each facility.

Staff.Recovered: The number of confirmed COVID recoveries amongst staff for each facility.

Residents.Tested: The number of confirmed COVID tests amongst residents for each facility. States report various metrics in these fields including number of persons tested or total tests conducted. Hackathon participants should look at the state dashboard to determine what metric each state is using, if determinable.

Staff.Tested: The number of confirmed COVID tests amongst staff for each facility. States report various metrics in these fields including number of persons tested or total tests conducted. Hackathon participants should look at the state dashboard to determine what metric each state is using, if determinable.

Residents.Pending: The number of pending tests amongst residents for each facility. What metric a state uses should be determined by hackathon competitors like with the testing variables.

Staff.Pending: The number of pending tests amongst staff for each facility. What metric a state uses should be determined by hackathon competitors like with the testing variables.

Residents.Negative: The number of negative tests amongst residents for each facility. What metric a state uses should be determined by hackathon competitors like with the testing variables.

Staff.Negative: The number of negative tests amongst staff for each facility. What metric a state uses should be determined by hackathon competitors like with the testing variables.

Residents.Quarantine: The number of inmates quarantined. What metric a state uses should be determined by hackathon competitors like with the testing variables.

Staff.Quarantine: The number of staff quarantined. What metric a state uses should be determined by hackathon competitors like with the testing variables.

hifld_id: The corresponding id for each facility with the HIFLD Prison Boundaries dataset, a census of prisons produced by the Department of Homeland Security.

Count.ID: The ID for each facility in our dataset.

Population_HIFLD: The population reported by the Department of Homeland Security for each facility.

For Pennsylvania there is an additional variable called **Resident.Population**. This is the inmate population per facility reported by the state.

Wherever possible and necessary we have calculated cumulative numbers for each date based on the numbers reported by a state.

There may be some spikes or drops in particular variables for each state. Unless competitors can independently confirm these spikes or drops are due to actual numbers reported by the state agency, competitors should assume these outliers are the result of errors in the projects scraper program.