

Data from the United States Census Bureau discussed below focuses on biographical information, including but not limited to, overall population count, different age groups, sex, and race of different subsets of the population dataset, per state.

Name	Definition	Data type	Possible values	Required?
State_Name	Name of the state	String	Alabama, California, North Carolina	Yes
Total_Population	Population per State	Integer	2,445,896, 14,000,631	Yes
Age_Groups	Subsets of Population that falls under a specific age group	Integer	2,445,896, 14,000,631	Yes
Male	Subset of Population that falls under a specific sex	Integer	2,445,896, 14,000,631	Yes
Female	Subset of Population that falls under a specific sex	Integer	2,445,896, 14,000,631	Yes
Race	Subset of Population that falls under a specific race	String	White, Black or African American, Asian	Yes

This dataset can be merged with the primary COVID-19 dataset using the State_Name or by a simple data wrangling process, where a derived variable can be generated from the State_Name and converted to the unique state identifier.

An approach that can be addressed while diving into the comprehensive dataset would be using the age groups in different states and analyse the actual numbers in those different age groups, moreover, a great conclusion would be comparing death rates

among the different age groups, with the hypothesis of senior citizens are the most vulnerable.

Another aspect worth exploring is if there are any correlations between race, or sex and infection rates, and death rates.