TASK 2

ENRICHMENT DATASET: CENSUS DEMOGRAPHIC ACS

This dataset contains demographic information of each county in the United States and can be combined with Covid-19 dataset. The data has been taken from the American Community Survey in the United States Census bureau website for the year 2019.

This information shows population estimates of categories such as sex, age, race, citizens and voting age of people in the country.

DATA DESCRIPTION

Sex: Total population estimate of males, females, and sex ratio of males per 100 females across theUnited States.

Age: The total population estimate by different age groups.

Race: The total population estimate of the different races like White, Black or African American, Indian etc.) in the country.

Citizen, 18 and Voting population: Total population estimate of citizens that are in voting age.

DATATYPE AND VARIABLE DESCRIPTION

Column Variables	Data types	Variable description
GEO_ID	Object	Geography id and the last fivedigits contains the FIPS code
Name	Object	Name of county and state
DP05_0001E	Int64	Total population estimate
DP05_0002E	Int64	Total Male Population
DP05_0003E	Int64	Total Female population
DP05_0004E	Int64	Total population estimate formale per 100 females
DP05_0005E	Float64	Total population estimate ofages under 5years
DP05_0006E	Int64	Total population estimate ofages from 5 to 9 years
DP05_0007E	Int64	Total population estimate ofages under 10 to 14 years
DP05_0008E	Int64	Total population estimate ofages under 15 to 19 years
DP05_0009E	Int64	Total population estimate ofages under 20 to 24 years
DP05_0010E	Int64	Total population estimate ofages under 25 to 34 years
DP05_0011E	Int64	Total population estimate ofages under 35 to 44 years

DP05_0012E	Int64	Total population estimate ofages under 45 to 54 years
DP05_0013E	Int64	Total population estimate ofages under 55 to 59 years
DP05_0014E	Int64	Total population estimate ofages under 60 to 64 years
DP05_0015E	Int64	Total population estimate ofages under 65 to 74 years
DP05_0016E	Int64	Total population estimate ofages under 75 to 84 years
DP05_0017E	Int64	Total population estimate of ages under 85 years and above
DP05_0018E	Float64	Total population estimate ofthose in median years
DP05_0037E	Int64	Total population estimate ofwhite race
DP05_0038E	Int64	Total population estimate ofBlack or African American
DP05_0039E	Int64	Total population estimate of American Indian and Alaskannative
DP05_0044E	Int64	Total population estimate ofthe Indian race
DP05_0052E	Int64	Total population estimate ofNative Hawaiian and other Pacific Islander race

IMPORTANCE OF DEMOGRAPHIC ENRICHMENT DATA IN COVID ANALYSIS

Covid has significant impact on everyone in the United States in a variety of ways. An analysis of the most impacted races, gender and age group would be extremely helpful in determining how government agencies, non-governmental organizations, and individual organizations might better share resources in a balanced manner.

MERGING COVID AND DEMOGRAPHIC DATA

To analyze covid effects on demographic data, we need to merge both the datasets. We can do that by using a common column named ‘County FIPS’ in Covid dataset and ‘GEO_ID’ in demographics dataset. However, the County FIPS needs to be extracted from GEO_ID column by eliminating first 9 characters.

For instance: GEO_ID: 0500000US01001. The County FIPS for this will be 01001.

SAMPLE HYPOTHESIS QUESTIONS

1. Do seniors have a greater death rate than those in lower age groups?
2. Are there significant number of cases and deaths among children under any specific age group?
3. What is the overall death rate?
4. Is any specific gender highly prone to virus?
5. Is there a particular race that was hit the hardest by the virus when the various races were compared?