MEMBER TASK

(SAIPAVAN TADIKONDA)

TASK-2

Enrichment Dataset (Census Demographic ACS):

I have chosen Census Demographic ACS enrichment dataset. This enrichment dataset contains all the demographic information of all states across USA by county level for the year 2021. It includes the estimates of population by age (different age groups), races (different races), sex and voting population.

Datatype-Variable dictionary:

Name	Datatype	Description
countyFIPS	Integer (int64)	Unique five-digit number for county
County Name	object	Name of the county
State	object	Name of the State
Male population	object	Estimate of total male population
Female population	object	Estimate of total female population
Under 5 years	object	Estimate of under 5 years age population
5 to 9 years	object	Estimate of population between 5 to 9 years
10 to 14years	object	Estimate of population between 10 to 14 years
15 to 19 years	object	Estimate of population between 15 to 19 years
20 to 24 years	object	Estimate of population between 20 to 24 years
25 to 34 years	object	Estimate of population between 25 to 34 years
35 to 44 years	object	Estimate of population between 35 to 44 years
45 to 54 years	object	Estimate of population between 45 to 54 years
55 to 59 years	object	Estimate of population between 55 to 59 years
60 to 64 years	object	Estimate of population between 60 to 64 years

65 to 74 years	object	Estimate of population between 65 to 74 years
75 to 84 years	object	Estimate of population between 75 to 84 years
85 years and over	object	Estimate of population of age 85years and over
Under 18 years	object	Estimate of Under 18 years age total population
18years and above	object	Estimate of total population of 18years and above age
65years and above	object	Estimate of total population of 65years and above age
Male (18years and above)	object	Estimate of total male population of age 18years and above
Female (18years and above)	object	Estimate of total female population of age 18years and above
Male (65years and above)	object	Estimate of total male population of age 65years and above
Female (65 years and above)	object	Estimate of total female population of age 65years and above
One race population	object	Estimate of total population of One race
White population	object	Estimate of total White population
Black population	object	Estimate of total Black population
Asian population	object	Estimate of total Asian population
Total Housing units	object	Estimate of total housing units
Voting population	object	Estimate of total voting population
Male voting population	object	Estimate of total male voting population
Female voting population	object	Estimate of total female voting population

How can you merge the data with the primary COVID-19 dataset. Identify the individual variable which map between the datasets?

The original raw census demographic ACS dataset contains 715 columns. Out of all columns I observed two columns GEO_ID and Name. GEO_ID column contains unique five-digit countyFIPS code in it and the Name column contained both County Name and State. These two columns are in common with the super covid-19 dataset. In order merge the census demographic dataset with super covid-19 dataset I extracted the countyFIPS code from the GEO_ID column of enrichment dataset using the slice function and changed the datatype of countyFIPS column to integer type as the datatype should match with the super covid-19 dataset countyFIPS column and I obtained the County Name and State by using the split function on the Name column in the census demographic dataset. After that I replaced the full name of the state to the two letter codes as it should match with the state column in the super covid-19 dataset. Also, out of 715 columns I only selected few columns (30) from the dataset along with the obtained countyFIPS, County Name and State columns (total 33) as these columns would help us in understanding the spread of covid-19. In future if required we may add or delete the columns.

countyFIPS, County Name and State are the variables which map between the census demographic enrichment dataset and super covid-19 dataset. Finally using the countyFIPS, County Name and State columns I merged the super covid-19 dataset with the census demographic dataset.

Describe how your enrichment data can help in the analysis of COVID-19 spread?

If we observe the census demographic enrichment dataset it contains all the demographic information based on the age, sex, and races across all states in United States by county level. By using this information, we can do different analysis (State level or County level) and get to know which race got most affected in the covid-19 spread, also which age groups are mostly affected in the covid-19 spread and which gender was more prone to this covid-19 virus.

HYPOTHESIS QUESTIONS:

1) Which age group got most affected by the covid-19 virus?
2) Which race was badly affected with the covid-19 virus?
3) Which gender was more prone to covid-19 virus?
4) Which county registered highest number of covid-19 cases?
5) Which State registered highest number of covid-19 cases?
6) Which county registered highest number of covid-19 deaths?
7) Which state registered highest number of deaths?

8) Which state has lowest number of covid-19 deaths?
9) Which county has lowest number of covid-19 deaths?
10) Does the chance of getting infected is more in higher age groups?
11) Is the death rate more in the higher age groups?