`

CSC 605 Project
Stage 1

This document was written and edited by:

Venkat Narayan Reddy Cheedu
Apoorva Gnana Saraswati Tangirala
ALURU PRANEETH
Nikhil Bolisetty
Akash Suresh

University of North Carolina at Greensboro

Covid-19 Datasets

Number of Cases:

This dataset contains the number of covid instances reported on each day in every county across the United States from 22ndJanuary 2020-17th February 2022.

Datatype-variable dictionary: Covid_confirmed_usafacts dataset

Name	Datatype	Description
countyFIPS	Series (unique id) int	Unique Id of County.
County Name	object	Name of county in each state
State	object	State names
stateFIPS	Int	Unique id of state
Different dates	Int	Columns ranging from 22 Jan 2020 to 17 Feb 2022 and cases reported on that day

Number of Deaths:

This dataset contains the number of death instances reported on each day in every county across the United States from 22ndJanuary 2020-17th February 2022.

Datatype-variable dictionary: Covid_death_usafacts dataset

Name	Datatype	Description
countyFIPS	Series (unique id)int	Unique Id of County.
County Name	object	Name of county in each state
State	object	State names
stateFIPS	Int	Unique Id of State.
Different dates	Int	Columns ranging from 22 Jan 2020 to 17 Feb 2022 and deaths reported on that day

County Population:

This dataset contains the population in each county in United States.

Datatype-variable dictionary: Covid_county_population_usafacts dataset:

Name	Datatype	Description
countyFIPS	Series (unique id) int	Unique Id of County.
County Name	object	Name of County in each state
State	object	State names
Population	int	Population in each county

Preliminary Intuitions:

- We can see that states with a large population have a higher number of cases and deaths from covid-19. For Example, California State has higher number of cases and deaths

- From the analysis we can figure out that the greatest number of cases are reported in month of Jan 2022.

- Through brief observation of the data, even though cases Jan 2022 are higher, but the death ratio is quite less comparatively.

- From the Data we can observe that at the cases and deaths curves is not continuous there is spike at few intervals.