Task 2 part 1

The entire team looks at the COVID-19 Dataset and understands the type of variables present in each of the data. (10 pts)

Deliverable:

Section in the report describing the COVID-19 dataset and datatype - variable dictionary

Preliminary intutions from the data

Article primarily for use with creating the variable dictionary of the COVID-19 dataset and datatype.

- https://analystanswers.com/what-is-a-data-dictionary-a-simple-thorough-overview/

County Wise Population Data Dictionary

Name	Definition	Data type	Possible values	Required?
County_FIPS	Unique id to distinguish different counties	Integer	1001, 1234,1111	Yes
County_Name	Name of a county	String	Bibb, Guilford, Chatham	Yes
County_State	State where the county is located	String	NC, NY, VA	Yes
County_Population	Population of the county	Integer	100000, 122324,13243	Yes

County Wise Covid Confirmed Cases Data Dictionary

Name	Definition	Data type	Possible values	Required?
County_FIPS	Unique id to distinguish different	Integer	1001, 1234,1111	Yes

	counties			
County_Name	Name of a county	Char	Bibb, Guilford, Chatham	Yes
County_State	State where the county is located	Char	NC, NY, VA	Yes
State_FIPS	Unique id to distinguish different states	Tinyint	1,2,3	Yes
Cases_By_Date	Total Covid cases for a given date	Date	2020-01-22, 2021-01-22, 2022-01-22	Yes

County wise Death Cases Data Dictionary

Name	Definition	Data type	Possible values	Required?
County_FIPS	Unique id to distinguish different counties	Integer	1001, 1234,1111	Yes
County_Name	Name of a county	Char	Bibb, Guilford, Chatham	Yes
County_State	State where the county is located	Char	NC, NY, VA	Yes
State_FIPS	Unique id to distinguish different states	Tinyint	1,2,3	Yes
Death_By_Date	Total Covid related death for a given date	Date	2020-01-22, 2021-01-22, 2022-01-22	Yes

Simply based on the data as its given in the COVID-19 datasets for Population, Cases and Deaths, we can start to make some preliminary intuitions about the data.

One obvious intuition from the dataset is that counties with high rates of confirmed cases also have a higher rate of death compared to other counties. One factor about COVID-19 that has been relatively well explored is that people with heavier rates of exposure to the virus, regardless of whether or not they have it, are at a higher risk of more dangerous prognosis and death. It would seem to follow that counties with more confirmed cases spread the virus quickly, and therefore increase exposure of individuals to the virus.

Another intuition we can gleam from the datasets is that counties with a higher population have higher risk of confirmed cases and death on average versus counties with lower populations. The COVID-19 virus is spread through symptomatic contact with infected individuals. If there is a county with a high population, the risk of exposure is much greater than in lower populated counties. Thus, it is fairly simple to see that a greater population per square mile, in places like New York, we see a higher rate of infection and death. Not only that, but inside certain counties we notice higher rates of infection and death over other counties in the same state. Take for example Boston, Massachusetts, it has a higher population density than a lot of other places in Massachusetts and has a higher number of confirmed cases and death than anywhere else in the state.