Dataset Summaries

The LA County Covid Cases Dataset comes from the data.gov database and details the number of cases, deaths, and other statistics of COVID from 2019-2023. This dataset keeps an updated record for Los Angeles California, one of the most highly populated cities in the United States. The columns in this dataset are as follows: County, State, fips, Date, Lat, Lon, Cases, Deaths, People Tested, State Cases, State Deaths, New Cases, New Deaths, New State Cases and New State Deaths. Many of these columns will not be applicable to my research and will be removed, as of right now these columns are people tested (all empty values), Lat/Lon, fips, county, state, and state cases/deaths. These columns will not be applicable to my research because they contain information about California as well as codes and geographical information that I do not need.

The LA Traffic Collision Data stores information about all reported motor vehicle accidents in Los Angeles California from 2010 to 2023. This dataset was also collected from the data.gov database and has over one million records of traffic collisions. The columns in this dataset are DR number, Date Reported, Date Occurred, Time Occurred, Area ID, Area Name, Reporting District, Crime Code, Crime Code Description, MO Codes, Victim Age, Victim Sex, Victim Descent, Premise Code, Premise Description, Address, Cross Street, and Location. Just as in my other dataset, many of these columns contain information that I do not need. The Columns that I do not need are DR Number, Area ID, Reporting District, MO Codes, Premise Code, Premise Description, Address, Cross Street, and Location.

The Data will be stored in the posit cloud interface and newly updated versions will be accessible from the data.gov database as it is frequently updated if more information is needed. Before joining my data, I will have to complete two major steps, one will be organizing the data by date as many rows are out of order. The other will be removing the columns from each of the tables that I will not be using in my research. After completing these steps, I will align the datasets and join them by overlapping the dates.

Data Sources:

Covid Data: <https://catalog.data.gov/dataset/la-county-covid-cases>

Traffic Data: <https://catalog.data.gov/dataset/traffic-collision-data-from-2010-to-present>