COVID-19 cases in the US(Hypothesis Formation)

COVID-19 pandemic has affected people globally. The purpose of this project will be to analyze the time series and forecast outcomes of the pandemic.

1 Context

COVID-19 pandemic has impacted people globally both economically and physically. The virus originated in Wuhan, China from person-to-animal contact and continues to spread at a fast rate. The time series data will be used to forecast the trend in the near future to safeguard solutions to prevent further spread of the virus.

2 Criteria for success

Identify a time series trend and forecast cases to safeguard better solutions to lower the rate of spread.

3 Scope of solution space

Create a forecast of the trend via the time series data

4 Constraints within solution space

The datasets being used will have to be formatted properly in order to work with. Since it is a time series data, the number of cases given any day is the cumulative number. The number must be subtracted from the data of previous day to get the accurate number of increase in cases for each day.

5 Stakeholders to provide key insight

Meant to provide key time series insights to the healthcare industry and public health officials to mandate better safeguards and approaches to lower the rate of spread.

6 Key data sources

- countries_by_population_2019.csv
- country_codes_2020.csv
- covid_19_data.csv
- covid_us_county.csv
- COVID19_line_list_data.csv
- COVID19_open_line_list.csv
- time_series_covid_19_confirmed.csv
- time_series_covid_19_confirmed_US.csv
- time_series_covid_19_deaths.csv
- time_series_covid_19_deaths_US.csv
- time_series_covid_19_recovered.csv