**Executive Summary**

On December 31, 2019, The World Health Organization China Country Office is informed of a number cases of pneumonia of unknown etiology (unknown cause) detected in Wuhan, Hubei Province. All cases connected to the Huanan Seafood Wholesale Market in Wuhan. ( Source: <https://www.cdc.gov/museum/timeline/covid19.html>) Within a month, the virus was seen in many parts of the world including the United States of America. By March 11, 2020 The World Health Organization declared COVID-19 a pandemic. Today United States has the highest cumulative number of confirmed Covid-19 cases in the world which is approximately 47.42 million followed by India 34.44 million cases (Source: https://ourworldindata.org/covid-cases).

This report captures the extent and spread of COVID-19 in the US and the 19, and the corresponding government response. It studies how the key intervention measures have affected the data trend. The project is based on our analysis of data from an open-source dataset Our World in Data.

First, we created a map to visualize how the global spread of cases through out the years using geo pandas.

The total Covid-19 cases in present day is shown below.

Chart

Description automatically generated

Then, we narrowed our focus to data on United States.

Our findings were as follows:

* **Actions such as social distancing, mask mandate and stay at home orders were immediately taken to flatten the curve of the virus and it had a positive effect. US initially saw a very slow growth in cases. But US failed to use that phase to figure out more targeted and state-level interventions.**
* By spring of 2020, Covid-19 cases were spiking, and hospital systems were in the risk of being overwhelmed by the patients with life-threatening symptoms in ICU. US saw increasing death tolls.
* **After the vaccines were available starting December 12, 2020, we could see a steep fall in daily new cases as well as hospitalizations and death.**
* **US saw first case of delta variant of covid in march, 2021 and started seeing a growth in covid cases again.**
* **Thus, the interventions such as masks mandate, social distancing and lockdowns were very effective in containing the virus but planning on next steps after that is equally important.**

**Data Collection**

Ways to access data, authors, date, git repository, etc.

**Data cleaning and Feature engineering**

The dataset has a lot of features such as 'iso\_code', 'continent', 'location', 'date', 'total\_cases', 'new\_cases', 'new\_cases\_smoothed', 'total\_deaths', 'new\_deaths', 'new\_deaths\_smoothed', 'total\_cases\_per\_million', 'new\_cases\_per\_million', 'new\_cases\_smoothed\_per\_million', 'total\_deaths\_per\_million', 'new\_deaths\_per\_million', 'new\_deaths\_smoothed\_per\_million', 'reproduction\_rate', 'icu\_patients', 'icu\_patients\_per\_million', 'hosp\_patients', 'hosp\_patients\_per\_million', 'weekly\_icu\_admissions', 'weekly\_icu\_admissions\_per\_million', 'weekly\_hosp\_admissions', 'weekly\_hosp\_admissions\_per\_million', 'new\_tests', 'total\_tests', 'total\_tests\_per\_thousand', 'new\_tests\_per\_thousand', 'new\_tests\_smoothed', 'new\_tests\_smoothed\_per\_thousand', 'positive\_rate', 'tests\_per\_case', 'tests\_units', 'total\_vaccinations', 'people\_vaccinated', 'people\_fully\_vaccinated', 'total\_boosters', 'new\_vaccinations', 'new\_vaccinations\_smoothed', 'total\_vaccinations\_per\_hundred', 'people\_vaccinated\_per\_hundred', 'people\_fully\_vaccinated\_per\_hundred', 'total\_boosters\_per\_hundred', 'new\_vaccinations\_smoothed\_per\_million', 'new\_people\_vaccinated\_smoothed', 'new\_people\_vaccinated\_smoothed\_per\_hundred', 'stringency\_index', 'population', 'population\_density', 'median\_age', 'aged\_65\_older', 'aged\_70\_older', 'gdp\_per\_capita', 'extreme\_poverty', 'cardiovasc\_death\_rate', 'diabetes\_prevalence', 'female\_smokers', 'male\_smokers', 'handwashing\_facilities', 'hospital\_beds\_per\_thousand', 'life\_expectancy', 'human\_development\_index', 'excess\_mortality\_cumulative\_absolute', 'excess\_mortality\_cumulative', 'excess mortality', 'excess\_mortality\_cumulative\_per\_million'.

Describes steps and why we decided to omit some data. Did we check for any outliers?

**Data Analysis and Visualization**

By spring of 2020, Covid-19 cases were spiking, and hospital systems were in the risk of being overwhelmed by the patients with life-threatening symptoms.

**Summary**

Overall

**Appendix**