A survey and analysis of the impact of COVID-19 on countries around the world

Xavier Zhen
xaizhen0808@gmail.com

Contents:

Table of Contents

1. Background... ***4***

1.1 Main Purpose of This Report,... **4**

2. Key Question and Question Justification.. ***5***

3. Brief data description.. ***5***

3.1 Overall trends in the global epidemic.. **6**

4.0 Analysis Process and Results Presentation on Key Question 1.................................... ***7***

5.0 Analysis Process and Results Presentation on Key Question 2.................................... ***9***

6.0 Insights... ***12***

7.0 Limitations.. ***13***

8.0 Dashboard... ***13***

References.. ***15***

Appendix A: Project code listing... ***16***

1. Background

Since the prevalence of COVID-19, it has had a significant impact on the health of people all over the world due to its robust transmission, hidden symptoms, and high mortality rate (Fu & Xiong, 2021). According to research, by the end of 2021, about 18 million people will die from the COVID-19 pandemic (COVID-19 Excess Mortality Collaborators, 2022). At the same time, COVID-19 affects the survival rate of newborn children. Compared with before the COVID-19 pandemic, the stillbirth rate and maternal mortality rate during the pandemic increased by about one-third (Chmielewska et al.,2021). Compared with high-income countries, the maternal and infant pregnancy outcomes in low-income countries are worse (Chmielewska et al.,2021). The COVID-19 pandemic also tested the entire society and all aspects of the social structure in an unprecedented way - including the test of the national medical and health system, test of social health services, test of solidarity, and the test of social contracts (Legido-Quigley et al.,2020). Therefore, in the face of sudden public health events such as COVID-19, the disease is far from a medical or epidemiological problem but involves the comprehensive social management of significant public crisis response. Because how the government formulates policies, spreads information, and how the public reacts to emotions, behaviors, and other aspects will significantly affect the development of the epidemic.

The reduction of the impact of the COVID-19 pandemic on the world and the reduction of mortality is inseparable from the government and individuals' dual efforts. For example, many governments have taken positive response measures such as isolation, school closure, closure, and requiring residents to wear masks (Van & Parolin, 2020). So it is imperative to determine what factors will affect the mortality rate of COVID-19 because it will help the government to prescribe the correct medicine for the situation. Only the prevention and control measures formulated based on scientific research results can ensure their effectiveness. At the same time, popularising the scientific principles behind the protection rules is the premise to persuade social members to accept the necessary action constraints voluntarily. According to the data released by Johns Hopkins University, COVID-19 varies significantly among different countries, with a 20-fold check (Dong et al.,2020).

1.1 Main Purpose of This Report,

The aim of this research report is to examine the current global impact of COVID-19, including an analysis of what factors affect mortality rates, why mortality and infection rates vary widely between countries, how strategies have done between countries for epidemic prevention and control have assisted in their prevention of the spread of COVID-19, and whether vaccination has been effective in preventing the spread of the virus. In addition, this report will also include an analysis of the effectiveness of existing epidemic prevention methods. The evaluation results can provide effective interventions to bring insights into countries that are not doing well in their epidemic prevention initiatives.

2. Key Question and Question Justification

This report analyzes the critical question: **Why do different countries experience vastly different death rates, and What affects the death rates of COVID-19?** The primary purpose of this question is to determine what factors will affect the mortality of COVID-19. It will not only compare the COVID-19 mortality rates of different countries but also analyze and find out the reasons for the different COVID-19 mortality rates in these countries from the perspective of policies, population structure, medical level, and other aspects. The purpose of this question is to judge how much external factors affect the mortality of COVID-19 to help the government take more scientific and practical measures to deal with COVID-19 and enhance the public's recognition of government policies and necessary action constraints.

The second question we have interest in analyzing is **Whether the current measures that different countries are taking are effective in preventing further transmission of COVID-19.** The purpose of the question is to figure out that the current epidemic prevention measures, including repetitive testing, and government reactions like lockdowns, are actually playing some roles in preventing further transmission. The condition of a country varies from one to another, including the strength of the public health system, population, and political environment. These differences make each country's response unique. Whatever a prevention policy or measure is, it does put pressure on the domestic economy. The necessity of an approach in one country may even change as the vaccination rate rises. The question aims to evaluate if the measures are effective and worth maintaining.

At the end of the article, according to the analysis results, corresponding suggestions will be put forward to reduce the mortality of COVID-19 and its negative impact on society.

3. Brief data description

The dataset used in this report is retrieved from https://github.com/owid/covid-19-data. The 'Our World in Data' (also referred to as 'owid') sourced dataset is automatically updated to provide real-time data. We used owid-covid-data.csv, which was last updated on October 18, 2022. The dataset contains 67 columns and 224,712 rows that record data from 232 countries and regions, which means that there are 67 different features in this dataset. The data date ranges from **January 1, 2020, to October 18, 2022**. Owid divided the population into four income groups, low income, lower middle income, upper middle income, and high income, to illustrate how the impact on different income groups can vary. Owid also added Africa, Asia, Europe, European Union, North America, Oceania, South America, and World to help researchers analyze regional differences. These add up to 243 different entries.

Due to different national conditions, countries with lower levels of development or higher poverty rates may face difficulties collecting data than other countries. Alternatively, some countries do not manage some of the features measured in the dataset. Hence there are undoubtedly multiple null values in the data. During exploratory data analysis, a total of 6,789,031 missing values were found in the dataset.

3.1 Overall trends in the global epidemic

Before further analyzing the two critical questions proposed, we visualized some basic information about the number of infections globally for the epidemic of novel coronaviruses. The dataset captures the number of confirmed infections from the COVID-19 outbreaks to the present. As can be observed from Figure 1, the number of confirmed cases is still on an ongoing upward trend **until October 18, 2022**.

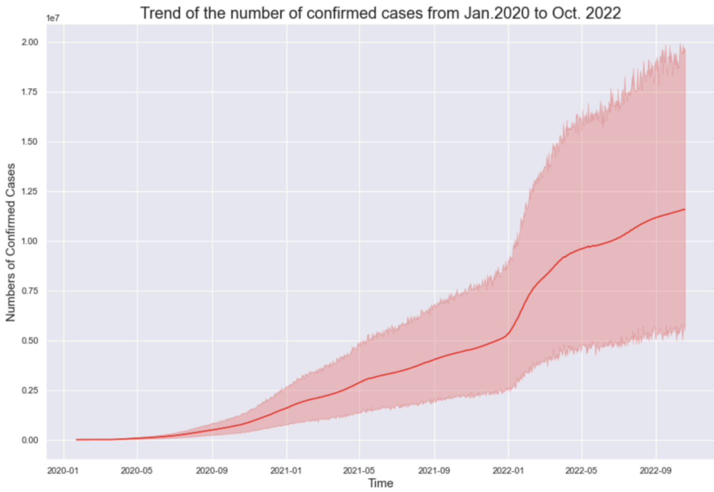

Figure 1:Trend of the number of confirmed cases from Jan.2020 to Oct. 2022

Figure 2 is a graph illustrating the trend in the number of confirmed cases for each country, with distinct color lines representing different countries, which can be viewed by clicking (in Jubtyer Notebook) on the country name in the Legend box on the right. Overall, the top five countries with the highest number of confirmed cases as of October 18, 2022, are the United States, India, France, Brazil, and Germany. This conjures up the question of whether the disparity in infection rates and death rates between countries is associated with the different epidemic prevention measures taken between countries.

Trends in the total number of confirmed cases by country From Jan. 2020 to Oct. 2022

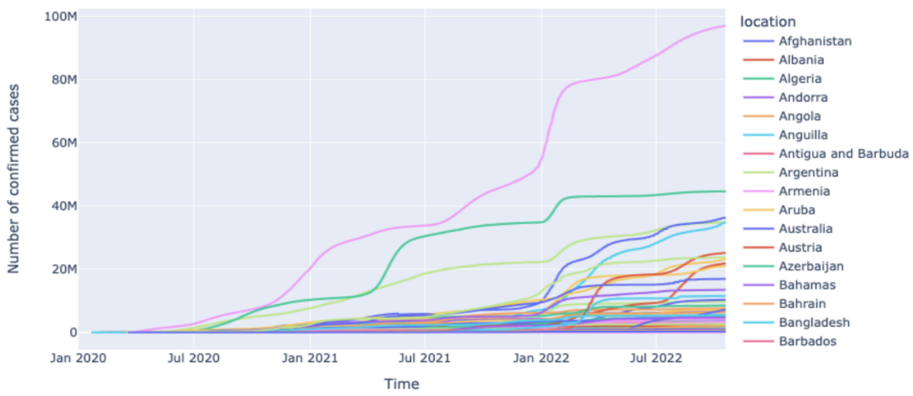

Figure 2: Trends in the total number of confirmed cases by country From Jan. 2020 to Oct. 2022

4.0 Analysis Process and Results Presentation on Key Question 1

The first question presented in Section 2.0 is about the analysis of mortality rates, the difference in mortality rates between countries, and what factors contribute to this discrepancy. The analysis will be taken done regarding the level of health care, the country's GDP, and the human development index. The study of this question will focus on external factors in society and provide insights from the study results in Section 6.0 to provide strategies for improving preventive measures.

Total numbers of death in 5 countries from Jan.2020 to Oct. 2022

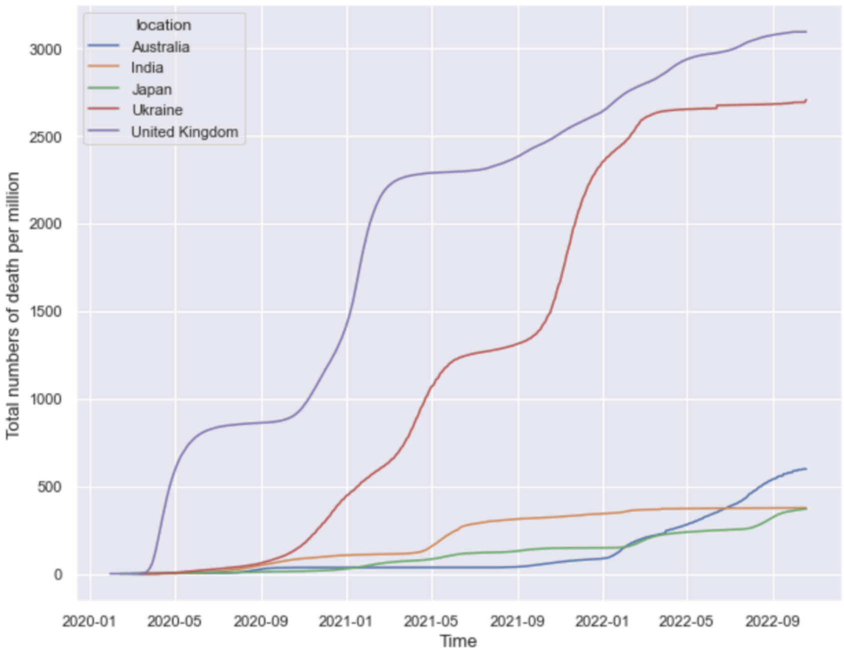

Figure 3: Total numbers of death in 5 countries from Jan. 2020 to Oct. 2022

The first step was to extract data from the original dataset for Japan, the UK, India, Australia, and Ukraine for this comparative analysis. Figure 3 contains the analysis of the number of deaths in the five selected countries and it can be observed that during the two years of the pandemic, the number of deaths increased in all five countries, but the total number of deaths in Australia, Japan, and India is lower compared to the UK and Ukraine. And has not shown extraordinary increases so far, which indicates that the effectiveness of their prevention measures for the epidemic was relatively satisfactory in comparison.

Figure 4 is a combined graph that compares GDP per capita, Human Development Index, hospital beds per thousand people, and population density for five countries. The confusing point that emerges from the graph is that India performs very low in all aspects and has the highest population density. Yet, it shows a lower number of deaths than Australia in Figure 3. Therefore, after browsing the data provided by the official website of WHO (World Health Organization, n.d.), it was noticed that the statistics of the dataset in Github for India data are not well matched with the WTO data, which will be highlighted in Section 7.0 (Limitation).

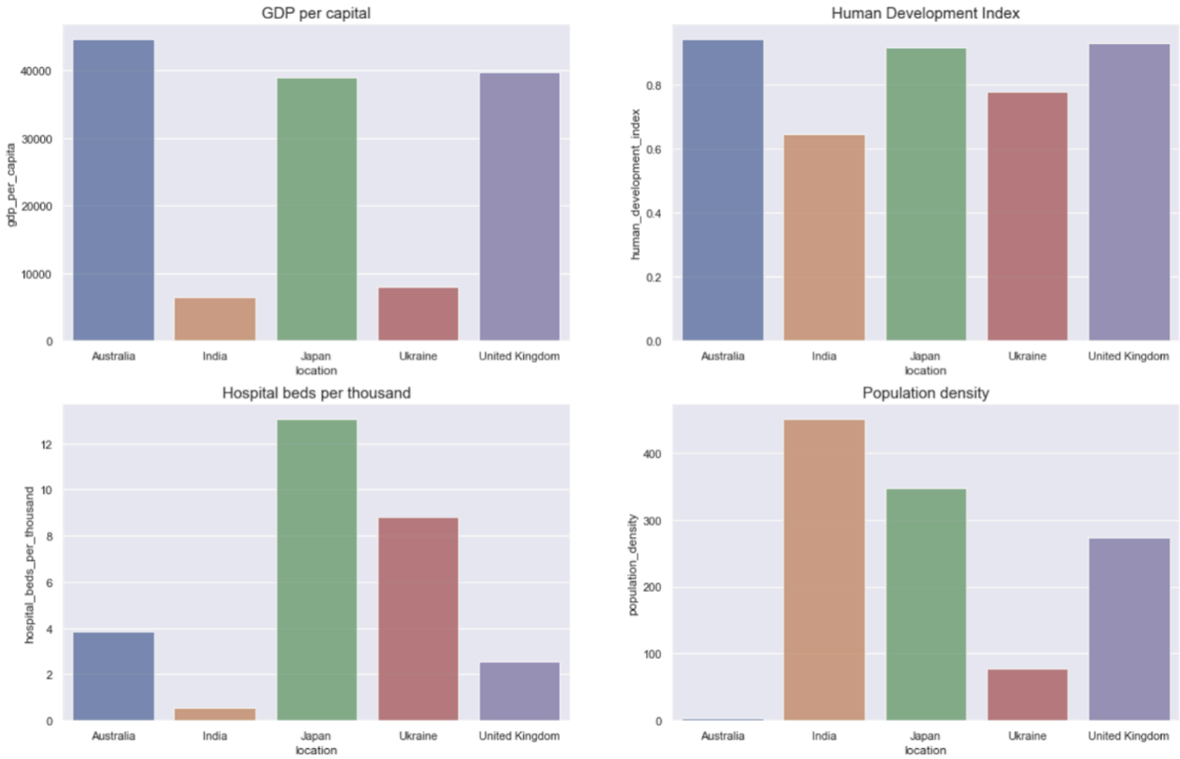

Figure 4: Difference between GDP per capital, Human Development Index, Hospital beds per thousand and Population Density in these five countries.

Overall, Japan and Australia have done a remarkable job of controlling patient mortality caused by COVID-19, attributed to their level of medical care (in this case, measured by hospital beds per thousand) in proportion to their population density. They also have the strength to build urban infrastructure and develop epidemic prevention strategies, as shown by the GDP per capita, and the ability to ensure that hospital beds are available to treat the patient in times of crisis. In contrast, the UK has a strong economy but is not doing well in the face of COVID-19.

With the fact that all countries have different population sizes, it is not impartial and rigorous to compare and analyze mortality numbers exclusively. Therefore, the data needs to be

processed to calculate the mortality rate. The World Health Organization defines the death rate as the Case Fatality Ratio (CFR), which measures the proportion of individuals who die from that diagnosed disease (Estimating mortality from covid-19, 2020). CFR was calculated by dividing the number of death from the disease by the number of confirmed disease cases. First, the mortality rates of the five previously selected countries were compared (as shown in Figure 5), and it was found that the mortality rates of the United Kingdom were significantly higher than those of the other countries. Then, four additional European countries were added to the analysis to investigate whether most European countries' mortality rate was higher. The results are shown in Figure 6, which shows that European countries have relatively high mortality rates compared to other countries.

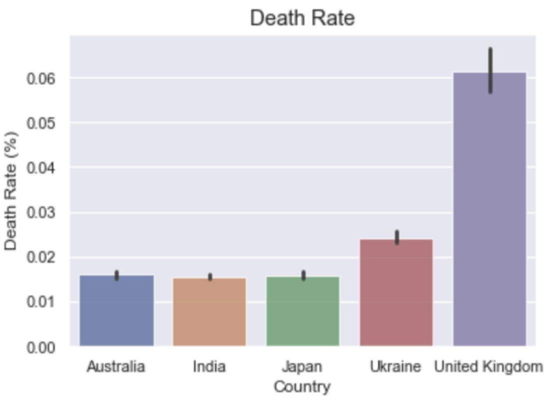

Figure 5: Comparison of mortality rates among five countries

Figure 7 presents a review of the number of COVID-19 patients in intensive care units (ICUs) and the total number of deaths, with France having the highest number of ICUs and the highest mortality rate. Several countries, including the UK, are ranked higher in this relationship showcase. Also, there is a positive linear relationship exists between the two variables.

Further research revealed that Ellyatt (2021) proposed that the UK government's response to

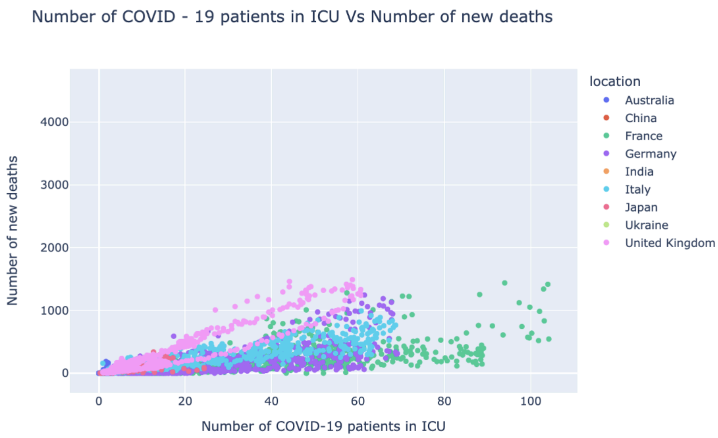

Figure 7: Number of COVID-19 patients in ICU Vs Number of new deaths

the COVID-19 outbreak was the absolute worst because they allowed the COVID virus to spread through the population to achieve "herd immunity". This response not only resulted in the rapid spread of the virus across the country but also caused many infected people to travel to other countries because they hesitated to blockade the country.

Thus, from the analysis and studies in this section, it can be concluded that the differences in mortality in different countries are determined by how they implement their epidemic prevention policies and measures for COVID-19. For countries with strong economic power, scientific and technological power, as well as medical level, if they match their capabilities and appropriate epidemic prevention initiatives together, they will implement substantial strength in their epidemic prevention capabilities.

5.0 Analysis Process and Results Presentation on Key Question 2

As question 2 concerns the comparison of epidemic prevention practices between countries, thus data from the various countries will be examined in this section, including a comparison of the number of confirmed cases and deaths between countries, a comparison of vaccination registration rates, and a comparison of the measures taken by the government.

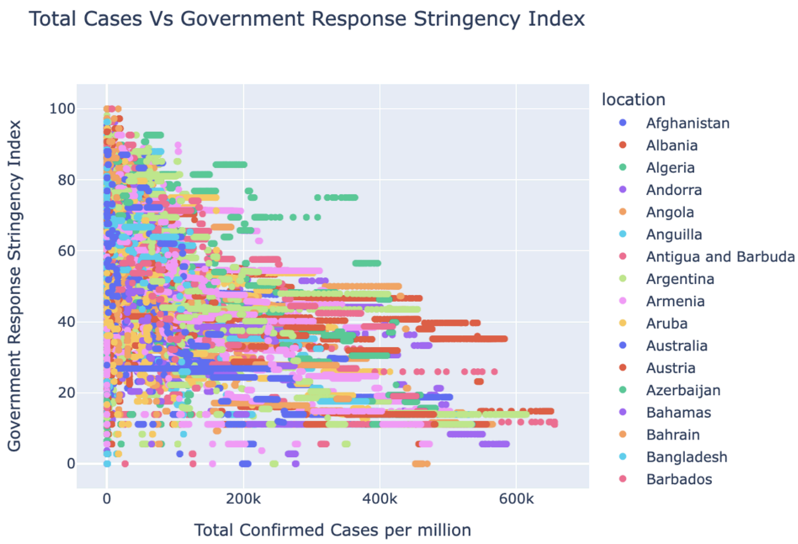

Figure 8: Total Cases Vs Government Response Stringency Index

Firstly, an analysis was performed for all countries in the dataset on the correlation between the total number of confirmed cases in the country and the government response stringency index. The dataset provider Githubs explains the stringency index measure, where they are scoring the national government's response index (on a scale from 1-100) using nine response indicators (including school closures, workplace closures, travel bans, etc.). Conversely, for some countries where the government does not undertake initiatives, their confirmed cases are much more numerous. However, it can be noted that some countries with indexes below 20 also have few confirmed cases (e.g., Algeria, Gabon, etc.) because the country's original population is smaller than others.

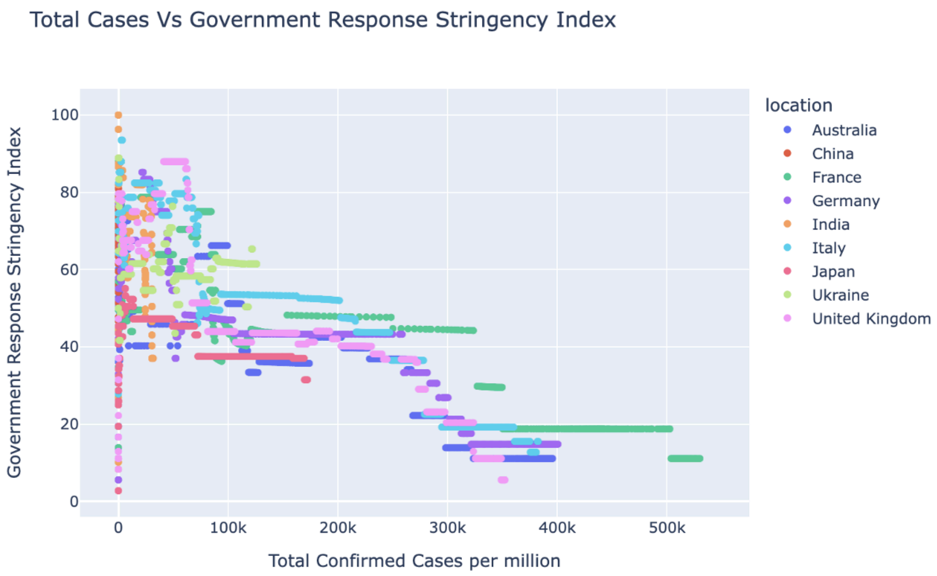

Figure 9: Total Cases Vs Government Response Stringency Index in selected 9 countries

To identify the direct effect of government interventions and confirmed cases, an additional graph was created using the nine countries investigated in section 4. As can be seen in Figure 9, when the stringency index is 0, it is possible that the epidemic has just started. Countries are unaware of the severity, or it has not spread to this country yet, so without considering this case, it is nevertheless clear that the response of the national government to COVID-19 transmission is related to the number of confirmed cases.

Following this, data from the United States and China were selected for analysis and comparison. the plots on the left in Figures 10 and 11 show a comparison of all confirmed cases and deaths, with the green portion representing the excess of confirmed cases over deaths. Still, since the number of deaths is so rare compared to the number of confirmed cases, only a red line can be seen near the x-axis. The graph on the right shows the number of new cases and new deaths per day over time. it can be found that the COVID-19 epidemic has been sweeping the United States since January 2020. Up until today, the U.S. has not had a perfect response to the coronavirus outbreak. The response of China at the beginning of the outbreak has been outstanding, such as city closures and tightened regulations on entry and exit. As can be observed in Figure 11, there were no significant new cases until May 2022. Although the May 2022 outbreak was in Shanghai and was caused by an Omniclonal mutation, the Chinese government responded rapidly to the epidemic wizth large-scale COVID-19 testing and strict lockdown measures (Bruno, 2022).

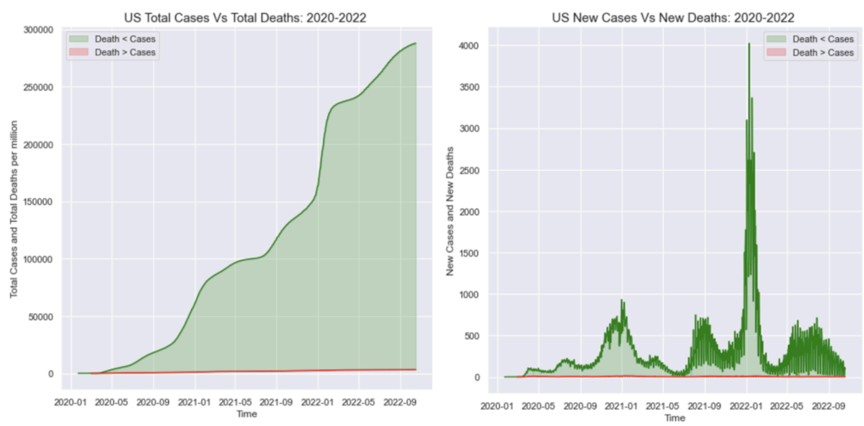

Figure 10: (Left) United States total confirmed cases and deaths from 2020 -2022

(Right) United Stated daily new confirmed cases and deaths from 2020-2022

In addition to how governments respond to emergencies among countries can impact the effectiveness of managing an epidemic, it is also essential to encourage people to get vaccinated voluntarily. From Figure 12, the three most apparent scatter plots are the United States, India, and China, three countries with large populations. It can be seen that countries with more significant numbers of vaccinated people have lower infection rates. From Andrews et al. (2022) research on effectiveness in preventing virus infection, they expressed that existing vaccines are only marginally protective and weak in being effective against Omicron variant virus strains. Vaccines can only partially enhance protection by preventing patients infected with coronavirus from turning critically ill or dying.

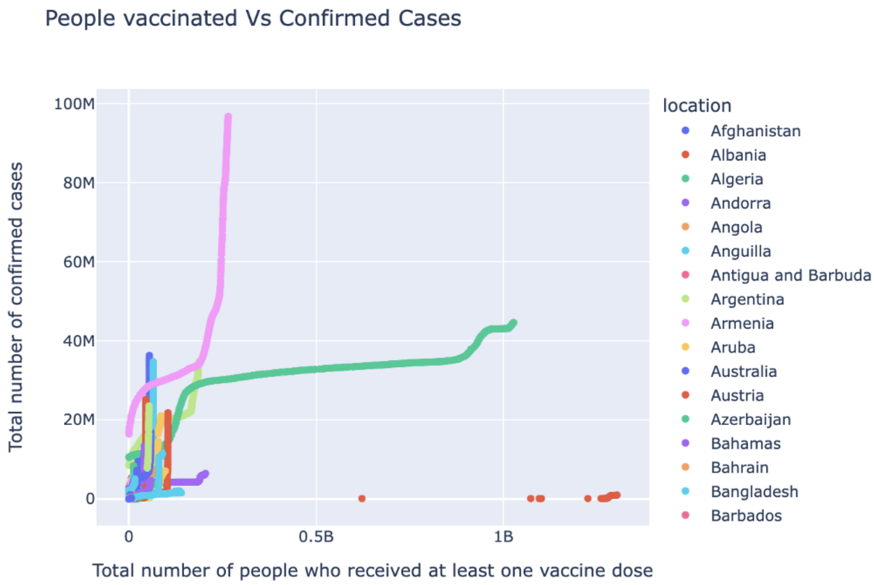

Figure 12: People vaccinated Vs Confirmed Cases

To conclude the findings in this section, the effectiveness of the measures taken by different countries depends on how the government implements its policies. Encouraging people to vaccinate is also a crucial step to protecting citizens. It has been two years since the beginning of the epidemic, and will soon enter the 'post-covid era,' how countries take the following steps will depend on government intervention and the cooperation of the people.

6.0 Insights

Covid-19 is a crucial challenge posed before all countries in the world, and the price for failing is precious lives of citizens. Countries which are economically strong tend to be more resistant to the pandemic because they can financially afford the cost of epidemic prevention initiatives, invest in Anti-Covid medicines and supplies development, and provide medical infrastructures to treat and quarantine confirmed or suspected cases. However, if a country takes passive epidemic prevention initiatives like UK's "herd immunity" strategy, it will pay the price of rapid increases in infected citizens and death rate. Whereas, countries that are economically less developed can protect their citizens well by implementing active interventions.

The future development of Covid-19 remains an unsolved mystery to the world. The vaccines currently available are not as effective for the Omicron variant as they are for Covid-19. In addition, there are no mature technologies to predict future variants of the virus, which increases difficulties in preventing virus from further spreading.

In addition to the fact that the economic burden that pandemic has put on all countries remains, citizens' dissatisfaction with the inconvenience brought by strict and long-lasting epidemic prevention initiatives grows. Countries may loosen some epidemic prevention policies in exchange of financial gains in tourist industry, foreign trading and investment, and other industries so that they can mitigate the economic burden. All the countries in the world are closely connected in the 21st century and no country can detach itself from our interconnected world. Just like the wooden barrel effect, countries like the UK which decided to do nothing will hinder other countries' commitment in preventing the epidemic, resulting in increased cost and resources input.

However, it is yet too soon to surrender to the pandemic. Vaccinations and active government interventions have been proved to be effective in protecting citizens from having severe symptoms or dying. It will be an endurance contest between the virus and countries which are close to their limit of economic capacity.

7.0 Limitations

In Section 4, while analyzing the COVID-19-related features and basic demographic information about the target five countries, it is noticeable that although India has poor performance in economic development and medical infrastructures, it still has a lower death rate compared with Australia. Besides, this dataset mismatches with the WHO records, therefore, the authenticity and accuracy of the data need further examination. However, most insights based on the dataset are still trustworthy except for the less developed countries, data is not collected correctly due to unsatisfied communication links or accessibility issues. More updated datasets with higher data precision will be recommended for further investigation into research questions.

Furthermore, the efficiency of the Anti-Covid actions taken by countries depends on policy executions and correlates with their economic strengths and different cultural ideologies of understanding Free Will. For impoverished countries, providing free COVID tests and vaccination for the whole population is money and time-consuming, which makes it difficult to actualize, and weak performance in intervention does not necessarily mean the governments are not efficient; financial burdens versus taxation income should be considered. Also, the level of difficulty in implementing Anti-Covid interferences could be culturally based. It is easier for Eastern nations to implement lockdown policies since Asian culture is group-oriented and strongly emphasizes society and community connection. They tend to obey professional guides rather than break the rules and get judged. At the same time, the Western world values freedom and individualism, making it harder to implement policies against Free Will, for instance, border closure, lockdown policies, daily COVID-19 tests, Booster Doses for vaccines, etc.

8.0 Dashboard

The Interactive Dashboard can be accessed from the link “http://127.0.0.1:8832/ ”. To ensure that the site can be opened properly, it is necessary to run all the code (please refers to the submitted zip files) in the Jupyter note book and then open it in the link shown in the last output column. This page should look like Figure 13, which previews the dashboard. Users can navigate the dashboard to see the outbreak's status by country on the map at different dates, including all confirmed cases, new confirmed cases, all deaths, and new deaths. In addition, users can also see three other graphs in the dashboard, including a graph indicates the number of people get vaccinated by country and region over time, a graph shows the number of confirmed cases and deaths worldwide, and a graph indicates the proportion of inpatient admissions to the ICU.

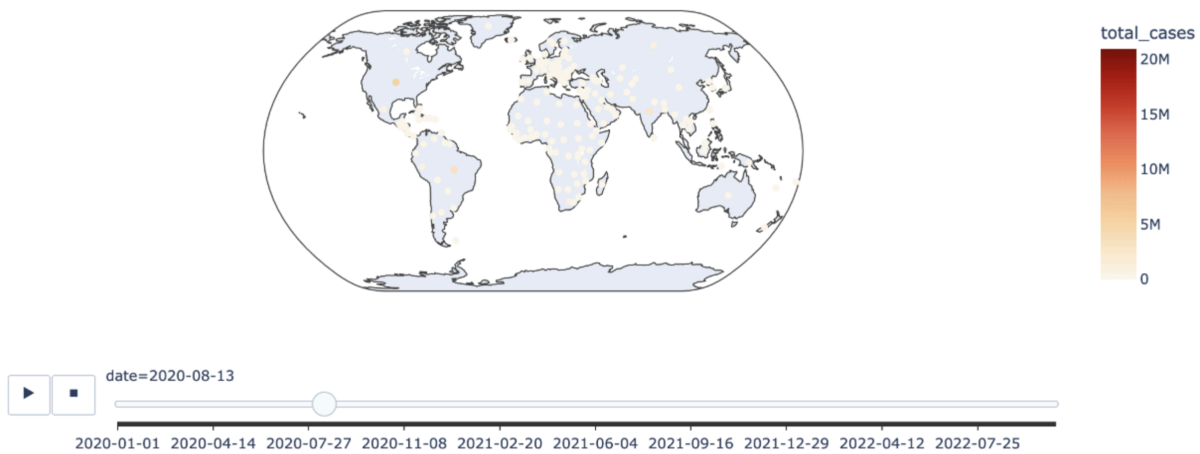

Figure 13: Preview of the Dashboard

References

Andrews, N., Stowe, J., Kirsebom, F., Toffa, S., Rickeard, T., Gallagher, E., Gower, C., Kall, M., Groves, N., O'Connell, A.-M., Simons, D., Blomquist, P. B., Zaidi, A., Nash, S., Iwani Binti Abdul Aziz, N., Thelwall, S., Dabrera, G., Myers, R., Amirthalingam, G., … Lopez Bernal, J. (2022). Covid-19 vaccine effectiveness against the Omicron (b.1.1.529) variant. *New England Journal of Medicine*, *386*(16), 1532–1546. https://doi.org/10.1056/nejmoa2119451

Bruno, M. (2022, July 6). Shanghai's new round of covid-19 testing reignites lockdown fears. Port Technology International. Retrieved October 29, 2022, from https://www.porttechnology.org/news/shanghais-new-round-of-covid-19-testing-reignites-lockdown-fears/

Chmielewska, B., Barratt, I., Townsend, R., Kalafat, E., van der Meulen, J., Gurol-Urganci, I., O'Brien, P., Morris, E., Draycott, T., Thangaratinam, S., Le Doare, K., Ladhani, S., von Dadelszen, P., Magee, L., & Khalil, A. (2021). Effects of the COVID-19 pandemic on maternal and perinatal outcomes: a systematic review and meta-analysis. The Lancet. Global health, 9(6), e759–e772. https://doi.org/10.1016/S2214-109X(21)00079-6.

COVID-19 Excess Mortality Collaborators (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. Lancet (London, England), 399(10334), 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3.

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020; 20: 533–34.

Ellyatt, H. (2021, October 12). Lawmakers Slam UK's Covid response, say 'herd immunity' strategy a public health failure. CNBC. Retrieved October 28, 2022, from https://www.cnbc.com/2021/10/12/uks-herd-immunity-covid-strategy-a-public-health-failure-inquiry.html

Fu, Y., & Xiong, S. (2021). Tagged extracellular vesicles with the RBD of the viral spike protein for delivery of antiviral agents against SARS-COV-2 infection. Journal of controlled release : official journal of the Controlled Release Society, 335, 584–595. https://doi.org/10.1016/j.jconrel.2021.05.049.

Legido-Quigley, H., Mateos-García, J. T., Campos, V. R., Gea-Sánchez, M., Muntaner, C., & McKee, M. (2020). The resilience of the Spanish health system against the COVID-19 pandemic. The Lancet. Public health, 5(5), e251–e252. https://doi.org/10.1016/S2468-2667(20)30060-8

Van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: a social crisis in the making. The Lancet. Public health, 5(5), e243–e244. https://doi.org/10.1016/S2468-2667(20)30084-0

World Health Organization. (2020, August 4). Estimating mortality from covid-19. World Health Organization. Retrieved October 28, 2022, from https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19

World Health Organization. (n.d.). Who coronavirus (COVID-19) dashboard. World Health Organization. Retrieved October 29, 2022, from https://covid19.who.int/?mapFilter=cases

Appendix A: Project code listing

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:


##import all the packages that needed
import matplotlib.pyplot as plt
import pandas_bokeh
from bokeh.plotting import figure, output_file,show

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from functools import wraps
import plotly.express as px


# In[2]:


#Read in the Data
data = pd.read_csv('owid-covid-data.csv')
pd.set_option('display.max_columns', None)
data


# In[3]:


# Check the total number of missing values
data.isnull().sum().sum()


# In[4]:


#preview of the dataset
#check the column name to have a breif understand of the
features
data.info()


# In[5]:
```

```
#Convert the format of 'date' (from object to datetime)
#https://pandas.pydata.org/docs/reference/api/pandas.to_dateti
me.html
data['date'] = pd.to_datetime(data['date'])


# In[6]:


#Check the total number of countries in the data and the
number of times they are counted in the data
data['location'].value_counts()


# In[7]:


data.describe()


# # Data Cleaning Process

# While doing data inspection, we found out that in "location"
coloumn, there are some countries named "Asia", "Africa",
etc., these are not country names but continent names, and
others named as "high income", "low income", etc. With country
names like these, they don't have corresponding continnent
names in the "continent" column. **Since the research question
is country based, not continent based or income based, we
should drop rows that don't have appropriate location names.**

# In[8]:


des = data.describe()
des.loc['skew', : ] = data.skew()
des.loc['kurt',: ] = data.kurt()
des

#If skewness value is greater than 1 or less than -1 indicates
a highly skewed distribution.
#A value between 0.5 and 1 or -0.5 and -1 is moderately
skewed.
#A value between -0.5 and 0.5 indicates that the distribution
is fairly symmetrical.
#for the extreme skewness, log transformation is needed.


#If kurtosis value is greater than +1, the distribution is too
peaked.
```

```
#Likewise, a kurtosis of less than -1 indicates a distribution that is too flat.
#data need to be transformed into its power of 1/2 or 1/4.


# In[9]:


data1_countryname = data[data["continent"].notna()]


# # Trend of the number of confirmed cases from 2020 to October 2022

# In[10]:


data_trend = pd.DataFrame()

data_trend['Time'] =  data['date']
data_trend['Confirmedcase'] = data['total_cases']
data_trend['Country'] = data['location']
data_trend['Continent'] = data['continent']

data_trend


# In[11]:


#Make a figure that shows the overall trend of the number of confirmed cases
plt.figure(figsize = (15, 10))
sns.set(style='darkgrid')
sns.lineplot( x = 'Time', y = 'Confirmedcase', data = data_trend, color='red')
plt.xlabel('Time', fontsize = 15)
plt.ylabel('Numbers of Confirmed Cases', fontsize = 15)
plt.title('Trend of the number of confirmed cases from Jan.2020 to Oct. 2022', fontsize = 20)
plt.show()


# In[12]:


#to see the differences of confirmed case between country
data1_countryname = data[data["continent"].notna()]
fig = px.line(data_frame = data1_countryname, x = 'date', y = 'total_cases', color = 'location',
```

```
        title = 'Trends in the total number of confirmed cases
by country From Jan. 2020 to Oct. 2022')
fig.update_layout( xaxis_title = "Time",yaxis_title = "Number
of confirmed cases")

fig.show()


# # Question 1

# In[13]:


data_q1 = data.query('location == ["Japan", "United
Kingdom","India","Australia","Ukraine"]')


# In[14]:


#for comparison use
plt.figure(figsize = (10,8))
sns.lineplot(data = data_q1, x = 'date', y
='total_deaths_per_million', hue = 'location')
plt.ylabel('Total numbers of death per million', fontsize =
13)
plt.xlabel('Time', fontsize = 13)
plt.title('Total numbers of death in 5 countries from Jan.2020
to Oct. 2022', fontsize = 15)
plt.show()


# In[15]:


#for comparison use
fig,axs = plt.subplots (nrows = 2, ncols = 2, figsize = (8,5))
plt.subplots_adjust(top = 2 , right = 2)

sns.barplot(data = data_q1, x = 'location', y
='gdp_per_capita', alpha = 0.8,
            ax = axs [0][0])
sns.barplot(data = data_q1, x = 'location', y
='human_development_index',
            alpha = 0.8, ax=axs[0][1])
sns.barplot(data = data_q1, x = 'location', y
='hospital_beds_per_thousand',
            alpha = 0.8, ax=axs[1][0])
sns.barplot(data = data_q1, x = 'location', y
='population_density',
            alpha = 0.8, ax=axs[1][1])
```

```
axs[0][0].set_title('GDP per capital', size = 15)
axs[0][1].set_title('Human Development Index', size = 15)
axs[1][0].set_title('Hospital beds per thousand', size = 15)
axs[1][1].set_title('Population density', size = 15)


# In[16]:


#Add a new feature - death rate
data1_countryname['death_rate'] =
((data1_countryname['total_deaths']/data1_countryname['total_c
ases'])*100)
data1_countryname['death_rate']


# In[17]:


#Make a barplot that can indicates the diffrences of death
rate betwen country
data_q1 = data1_countryname.query('location == ["Japan",
"United Kingdom","India","Australia","Ukraine"]')

fig = sns.barplot(data = data_q1, x = 'location', y
='death_rate', alpha = 0.8)

plt.xlabel ('Country')
plt.ylabel('Death Rate (%)')
plt.title('Death Rate', fontsize = 15)


# In[18]:


#Check the correlation between these two variables
data_q1 = data1_countryname.query('location == ["Japan",
"United
Kingdom","India","Australia","Ukraine","Germany","Italy","Fran
ce","China"]')

df = px.data.tips()
fig = px.scatter(data_q1, x='icu_patients_per_million',
y="new_deaths", color ='location',
                title = "Number of COVID - 19 patients in ICU
Vs Number of new deaths")
fig.update_layout( xaxis_title = "Number of COVID-19 patients
in ICU",
                  yaxis_title = "Number of new deaths")
fig.show()
```

```
# In[19]:


#for comparison use
data_q1 = data1_countryname.query('location == ["Japan",
"United
Kingdom","India","Australia","Ukraine","Germany","Italy","Fran
ce","China"]')
fig = sns.barplot(data = data_q1, x = 'location', y
='death_rate', alpha = 0.8)

plt.xlabel ('Country')
plt.xticks(rotation = 40)
plt.ylabel('Death Rate (%)')
plt.title('Death Rate', fontsize = 15)


# # Question 2

# ### US

# In[20]:


#Trend of total cases and total deaths, in US
data_US = data.query('location == ["United States"]')

sns.set(style = 'darkgrid')

Time = data_US['date']
Cases = data_US['total_cases_per_million']
Death = data_US['total_deaths_per_million']

#Initialize figure and axis
fig, ax = plt.subplots(figsize = (8,8))

#Plot lines
ax.plot(Time, Cases, color = 'green')
ax.plot(Time, Death, color = 'red')

#Fill area when confirmed case > death with green
ax.fill_between(Time, Cases, Death, where = (Cases > Death),
                interpolate = True, color = 'green', alpha =
0.2,
                label = 'Death < Cases')

#Fill area when death > confirmed case with red
ax.fill_between(Time, Cases, Death, where = (Cases <= Death),
                interpolate = True, color = 'red',alpha = 0.2,
                label = 'Death > Cases')
```

```
#Add labels, title
ax.set_xlabel ('Time')
ax.set_ylabel('Total Cases and Total Deaths per million')
ax.set_title('US Total Cases Vs Total Deaths: 2020-2022 ',
fontsize = 15)
ax.legend()


# In[21]:


data_US['new_cases']


# In[22]:


data_US['new_deaths']


# In[23]:




#for comparison use
data_US = data.query('location == ["United States"]')

sns.set(style = 'darkgrid')

Time = data_US['date']
newCases = data_US['new_cases_per_million']
newDeath = data_US['new_deaths_per_million']

#Initialize figure and axis
fig, ax = plt.subplots(figsize = (8,8))

#Plot lines
ax.plot(Time, newCases, color = 'green')
ax.plot(Time, newDeath, color = 'red')

#Fill area when confirmed case > death with green
ax.fill_between(Time, newCases, newDeath, where = (Cases >
Death),
               interpolate = True, color = 'green', alpha =
0.2,
               label = 'Death < Cases')

#Fill area when death > confirmed case with red
ax.fill_between(Time, newCases, newDeath, where = (Cases <=
Death),
```

```
                interpolate = True, color = 'red',alpha = 0.2,
                label = 'Death > Cases')

#Add labels, title
ax.set_xlabel ('Time')
ax.set_ylabel('New Cases and New Deaths')
ax.set_title('US New Cases Vs New Deaths: 2020-2022 ',
fontsize = 15)
ax.legend()


# ### China

# In[24]:



#for comparison use
#Trend of total cases and total deaths, in China
data_CN = data.query('location == ["China"]')

sns.set(style = 'darkgrid')

Time = data_CN['date']
CasesCN = data_CN['total_cases_per_million']
DeathCN = data_CN['total_deaths_per_million']

#Initialize figure and axis
fig, ax = plt.subplots(figsize = (8,8))

#Plot lines
ax.plot(Time, CasesCN, color = 'green')
ax.plot(Time, DeathCN, color = 'red')

#Fill area when confirmed case > death with green
ax.fill_between(Time, CasesCN, DeathCN, where = (Cases >
Death),
                interpolate = True, color = 'green', alpha =
0.2,
                label = 'Death < Cases')

#Fill area when death > confirmed case with red
ax.fill_between(Time, CasesCN, DeathCN, where = (Cases <=
Death),
                interpolate = True, color = 'red',alpha = 0.2,
                label = 'Death > Cases')

#Add labels, title
ax.set_xlabel ('Time')
ax.set_ylabel('Total Cases and Total Deaths per million')
```

```
ax.set_title('China Total Cases Vs Total Deaths: 2020-2022 ',
fontsize = 15)
ax.legend()


# In[25]:


#for comparison use
data_CN = data.query('location == ["China"]')

sns.set(style = 'darkgrid')

Time = data_CN['date']
newCasesCN = data_CN['new_cases_per_million']
newDeathCN = data_CN['new_deaths_per_million']

#Initialize figure and axis
fig, ax = plt.subplots(figsize = (8,8))

#Plot lines
ax.plot(Time, newCasesCN, color = 'green')
ax.plot(Time, newDeathCN, color = 'red')

#Fill area when confirmed case > death with green
ax.fill_between(Time, newCasesCN, newDeathCN, where = (Cases >
Death),
               interpolate = True, color = 'green', alpha =
0.2,
               label = 'Death < Cases')

#Fill area when death > confirmed case with red
ax.fill_between(Time, newCasesCN, newDeathCN, where = (Cases
<= Death),
               interpolate = True, color = 'red',alpha = 0.2,
               label = 'Death > Cases')

#Add labels, title
ax.set_xlabel ('Time')
ax.set_ylabel('New Cases and New Deaths')
ax.set_title('China New Cases Vs New Deaths: 2020-2022 ',
fontsize = 15)
ax.legend()


# ### Correlation between variables

# In[26]:


#Make a new DataFrame that ready to be use in this section
```

```
data_corr =
data[['total_cases','new_cases','total_deaths','new_deaths',

'total_cases_per_million','new_cases_per_million','total_death
s_per_million',

'reproduction_rate','icu_patients','icu_patients_per_million',
'hosp_patients',

'hosp_patients_per_million','weekly_icu_admissions','weekly_ic
u_admissions_per_million',

'weekly_hosp_admissions','weekly_hosp_admissions_per_million',
'total_tests',

'new_tests','total_tests_per_thousand','new_tests_per_thousand
','positive_rate',

'tests_per_case','tests_units','total_vaccinations','people_va
ccinated',

'people_fully_vaccinated','total_boosters','new_vaccinations',

'stringency_index','population','population_density','median_a
ge',

'aged_65_older','aged_70_older','gdp_per_capita','extreme_pove
rty',

'cardiovasc_death_rate','diabetes_prevalence','female_smokers'
,

'male_smokers','handwashing_facilities','hospital_beds_per_tho
usand',

'life_expectancy','human_development_index','excess_mortality'
]]


# In[27]:


#Heatmap use to check the correlations between variables
corr = data_corr.corr()
n_var = len(corr)

plt.figure(figsize = (25,20))
plt.imshow(corr, cmap = 'winter')

plt.xticks(range(n_var), corr.columns, rotation = 90)
plt.yticks(range(n_var), corr.columns)
```

```
for i in range (n_var):
    for j in range(n_var):
        plt.text(i, j, '{:.2f}'.format(corr.iloc[i,j]), ha =
"center", va="center",fontsize = 7)


plt.colorbar()
plt.title("Correlations", fontsize = 20)
plt.show()


# In[28]:


#Check the correlation between total cases and other
Case_n_other =
data_corr.corr()['total_cases'].sort_values(ascending =
False).round(3)
Case_n_other


# In[29]:


#Add a new feature - People fully vaccinated rate
data1_countryname['fully_vaccinated_rate'] =
((data1_countryname['people_fully_vaccinated']/data1_countryna
me['population'])*100)
data1_countryname['infection_rate'] =
((data1_countryname['total_cases']/data1_countryname['populati
on'])* 100)


# In[30]:


#Plot that shows relationship between Government response
stringency index and total cases
# All countries
fig = px.scatter (data1_countryname, x =
'total_cases_per_million', y ='stringency_index',
                 color = 'location', title = "Total Cases Vs
Government Response Stringency Index")
fig.update_layout( xaxis_title = "Total Confirmed Cases per
million",
                  yaxis_title = "Government Response
Stringency Index ")
fig.show()
```

```
# In[31]:


#Plot that shows relationship between Government response
stringency index and total cases
#Selected Countries
fig = px.scatter (data_q1, x = 'total_cases_per_million', y
='stringency_index',
                 color = 'location', title = "Total Cases Vs
Government Response Stringency Index")
fig.update_layout( xaxis_title = "Total Confirmed Cases per
million",
                  yaxis_title = "Government Response
Stringency Index ")
fig.show()


# In[32]:


#check the correlation between people get vaccinated and total
cases of that country
fig = px.scatter(data1_countryname, x = 'people_vaccinated', y
='total_cases',
                 color = 'location',title = "People vaccinated
Vs Confirmed Cases")
fig.update_layout( xaxis_title = "Total number of people who
received at least one vaccine dose",
                  yaxis_title = "Total number of confirmed
cases ")
fig.show()


# # Dashboard

# In[33]:


##import all the packages that needed
import plotly.express as px
import pandas as pd
import numpy as np
import math
import pandas as pd

import plotly.express as px
import plotly.io as pio
from plotly import graph_objects as go
pio.renderers.default = 'browser'

df= pd.read_csv('owid-covid-data.csv')
```

```
# In[34]:


#Preview of the data
df.head()


# In[35]:


#Get the name of each column
df.columns


# In[36]:


#Dealing with missing values
df['total_cases'] = df['total_cases'].fillna(0)

df = df.sort_values('date')


# In[37]:


#Get readt for draw the first figure
df_fig1 = df[['date','iso_code','continent','total_cases',
        'location','total_deaths','new_cases','new_deaths']]


# In[38]:


#get a pandas pivot tables to check the data
data=
data1_countryname.pivot_table(index='location',columns='date',
values='people_vaccinated', fill_value=0, aggfunc='mean')
data


# In[39]:



head_20_index = data.loc[:,'2022-10-18'].sort_values(ascending
= False).head(20).index
head_20 = data.loc[head_20_index]
head_20
```

```
# In[40]:


#The second figure
fig2 = go.Figure()

x= head_20.columns
for idx in head_20.index:

    fig2.add_trace(go.Scatter(x =x, y =
head_20.loc[idx].values.tolist(),
                              text =
head_20.loc[idx].values.tolist(),
                              mode = 'lines+markers',name=
idx,
                              hovertext =
'people_fully_vaccinated',
                              hoverinfo = 'all',opacity=0.5))


# In[41]:


##import all the packages that needed
import dash
import flask
import dash_core_components as dcc
from dash import html
server = flask.Flask(__name__)


# In[42]:


World_df =
df[df['location']=="World"][['date','total_deaths','total_case
s']].tail(5)
World_df


# In[43]:


#Third figure
fig3 = px.bar(World_df,  x='date',
y=['total_deaths','total_cases'],
              title="The number of cases and deaths around the
world ",
              barmode='group' )
```

```
# In[44]:


#make a dataframe that contains all the data needed to make
figure 3
icu_df =
df[['date','location','weekly_icu_admissions','weekly_hosp_adm
issions_per_million']].dropna()
icu_df


# In[45]:


icu_df['year'] = icu_df['date'].map(lambda x:x[:4])
icu_df


# In[46]:


icu2022 =
icu_df.groupby(['location','year']).sum().reset_index().query(
'year=="2022"').drop('year',axis=1)
icu2022


# In[47]:


fig4 = px.bar(icu2022,  x='location',

y=['weekly_icu_admissions','weekly_hosp_admissions_per_million
'],
              title="icu number comparison different country
", barmode='stack', )


# In[48]:


#!pip install dash_bootstrap_components


# In[49]:


from dash.dependencies import Input, Output
import dash
```

```
from dash.dependencies import Input, Output
import dash_core_components as dcc
from dash import html
import dash_bootstrap_components as dbc


# In[ ]:


fig = px.scatter_geo(df_fig1, locations="iso_code",
color="total_cases",

color_continuous_scale=px.colors.sequential.OrRd,
                     hover_name="location",

hover_data={"total_cases":True,"new_cases":
True,"total_deaths":True,
                                 "new_deaths":True,

'iso_code':False},projection="natural earth",

animation_frame='date',animation_group='location')
def graph1():
    return dcc.Graph(id='graph1',figure=fig)

external_stylesheets = [dbc.themes.BOOTSTRAP]

app = dash.Dash(__name__,
external_stylesheets=external_stylesheets)

server = app.server
app.layout = html.Div(children=[html.Div(children='''Global
impact of COVID-19 '''),
    dcc.Graph(
        id='graph1',
        figure=fig), html.Div(children=''''''),dcc.Graph(
        id='graph2',
        figure=fig2),
    dcc.Graph(
        id='graph3',
        figure=fig3),
    dcc.Graph(
        id='graph4',
        figure=fig4)])
if __name__ == '__main__':
    app.run_server(debug=False,port=8832, use_reloader=False)
```