An extensive study of the COVID-19 data

Authors:

Sri Naga Sai Sushma Kondabolu

Summary:

COVID-19 is a Coronavirus disease caused by severe acute respiratory coronavirus 2 and it's a contagious disease that can be transmitted through many media. It weakens the human immune system and creates respiratory/other complexities that have the potential to end life. This disease was first detected in Wuhan, China in the year 2019. Transmission among humans was first recognized and confirmed on January 20, 2020. The World Health Organization declared announced the Public Health Emergency of International Concern on 30 January 2020 citing the COVID-19 outbreak and a pandemic was declared on 11 March 2020. The WHO has been collecting daily statistics of the positive cases of COVID-19 and deaths due to it from 237 countries/regions. Additionally, the WHO is recording the vaccination status of the countries/regions. This analysis is based on the data collected by WHO and aims at understanding the progress of the pandemic and its status in the world, the trend in positive cases, and deaths.

The following data from WHO is utilized:

- **Daily cases and deaths by date reported to WHO**

  This data consists of the date at which the reporting was made, the ISO Alpha 2 country code, name of the country, WHO region, number of new cases, cumulative cases up to that day, number of new deaths, and cumulative deaths due to COVID 19 up to that day.

- **Latest reported counts of cases and deaths, and transmission classification**

  This data consists of the Name of the country/territory/area, WHO Region, cumulative cases, cumulative cases per 100000 population, cases newly reported in the last seven days, cases newly reported in the last seven days per 100000, cases newly reported in the last 24 hours, total deaths, total deaths per 100000 population, total deaths newly reported in the last seven days, total deaths newly reported in the last seven days per 100000 and deaths newly reported in the last 24 hours.

- **Vaccination data**

  This data contains the Name of the country/territory/area, ISO 3 code of the country, WHO region, source of data, date of update of the status, total vaccinations, people vaccinated one plus dose, total vaccinations per 100, persons fully vaccinated, persons fully vaccinated per 100, vaccines used, first vaccine use date and a number of vaccines type used.

➔ **Vaccine metadata**

Consists of information on first authorizations and usage of different vaccines globally.

The goal here is to analyze the data and answer questions like finding regions having the highest death percentage and COVID-19 cases, the trend in the number of cases and deaths, progress in vaccination around the world, and vaccination impact on the rise or fall of cases/deaths. To facilitate the study of the data, the Python programming language is used with help from libraries such as NumPy, Pandas, Seaborn, etc. There'd also be a visual representation of the statistics of cases, deaths, and vaccination on the world map.

The current data indicates that the number of deaths due to COVID is on the decline, but the cases are fluctuating.

Analysis of the COVID-19 data is performed by many institutions but primarily the work of John Hopkins University is notable as their Coronavirus Resource Center is based on an independent data collection mechanism rather than the data provided by WHO.

Methods:

The data is directly read from the links to files provided on the WHO COVID-19 website. The data is uniform and there were no abnormalities in reading the data.

For the daily cases data, the country of Namibia has no country code. The unique value of country codes was inspected and a code of 'NM' was manually assigned to Namibia using fillna() method.

The latest cases data has few null values, they don't require processing as they are for 'Other' regions and Global data which would not cause a problem for the further steps aimed to take here

The vaccination data too has certain null values, they are filled with 'Not Available'. Since filling with Not Available changes the datatype of numeric features, another copy of the data is read and stored for statistical computations.

The columns that hold date are transformed using the datetime module in Pandas.

To compute and visualize the number of cases and deaths per day, all the records of countries/region for the day is summed up and plotted using the Seaborn line plot.

To have monthly data and yearly data, new columns have been added to the daily cases data as part of feature engineering, the Month_year and Year columns that have the year and month of the date reported. The cases and deaths are grouped by month and plotted with year as hue to understand the difference between the data for 2020 and 2021.

To see the trend for any country a function is designed such that it takes the valid country name as input and displays the cases details and visualizations for the country. This function is reused

multiple times to check the status of countries that have decent vaccination and the ones that are highly populated.

Choropleth maps are plotted to visualize the total number of cases, deaths, and vaccination on the world map.

Results:

The analysis on the data has provided many insights on the progress of the pandemic:

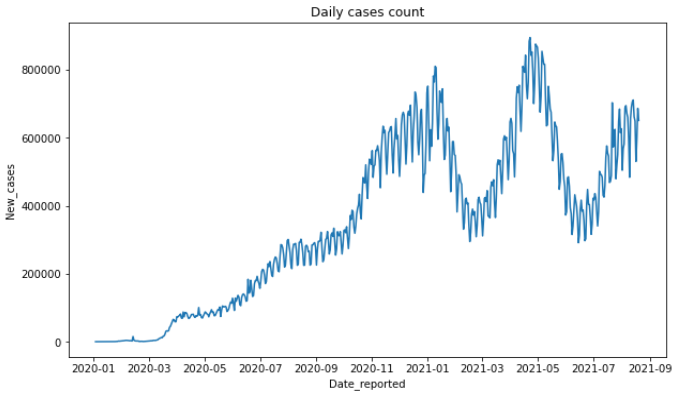


Fig1(Left) The daily trend in cases

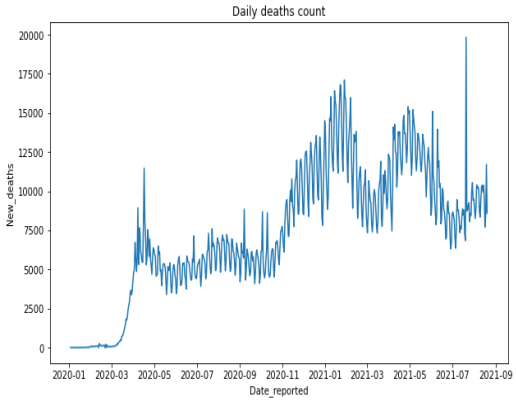


Fig 2(Right) The daily trend in deaths

Fig 1 depicts that the world has witnessed two massive peaks in the number of cases. Currently, the world is marching towards another peak going by the visualization of the daily cases count. The number of deaths too is on a rise now compared, but the highest peak of the number of deaths was between Jan 2021 and March 2021. The highest number of reported positive cases (894883) was on 23 April 2021 and the highest number of deaths in a single day is 19834 which was reported on 21 July 2021. The country that reported the maximum number of cases in a single day is 414188 by India on 7 May 2021 and the maximum number of deaths in a single day is 8786 in Ecuador on 21 July 2021.

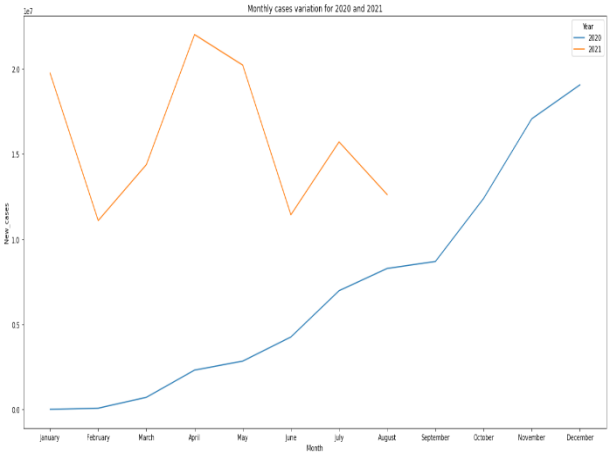


Fig 3(Left) Month cases for the years 2020 and 2021

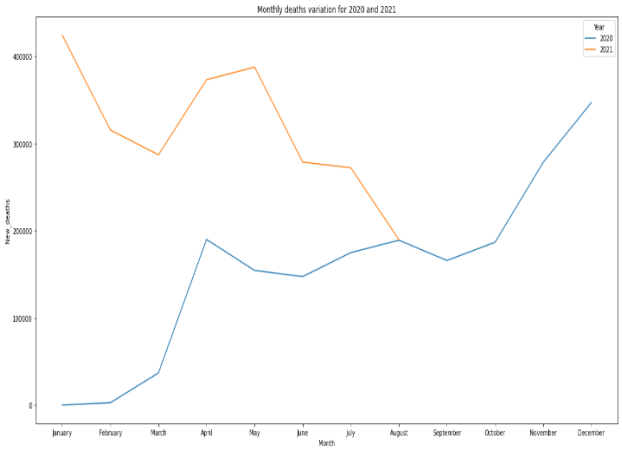
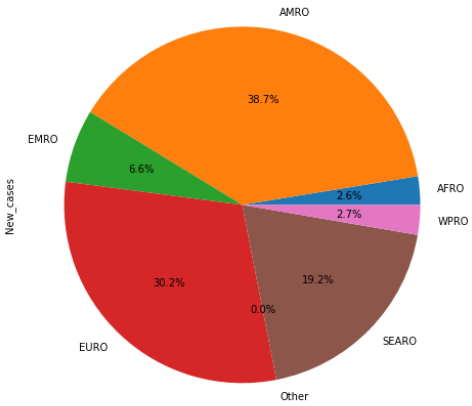


Fig 4(Left) Month cases for the years 2020 and 2021

The visuals suggest that the number of confirmed COVID-19-positive cases in 2021 is more than the 2020 cases for any corresponding month so far. Also, there was a steady increase in the number of cases per month in 2020, whereas 2021 saw a decline in the beginning and went on a rise and decline alternatively twice so far. Coming to the number of deaths, the count of deaths was on the increase and became steady in 2020 but had an increasing trend at the end of the year Whereas 2021, had higher deaths at the beginning of the year but currently is indicating a decline in the number of deaths though there were rises and falls in between. Overall, the highest number of positive cases were reported in the month of April 2021 (220012706 cases) and the greatest number of deaths were reported in the month of January 2021 (424185 cases).



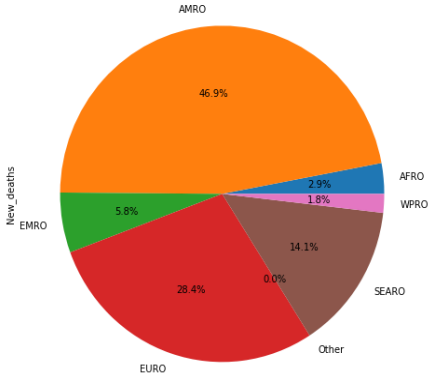


Fig 7(Left) Cases distribution by WHO regions

Fig 8(Left) Deaths distribution by WHO regions

From the pie chart, the AMRO region of the WHO has the most number of total cases and a total number of deaths when compared to other WHO regions.

All over the world, 1029756684 people are fully vaccinated and a total of 4562256778 doses were administered.

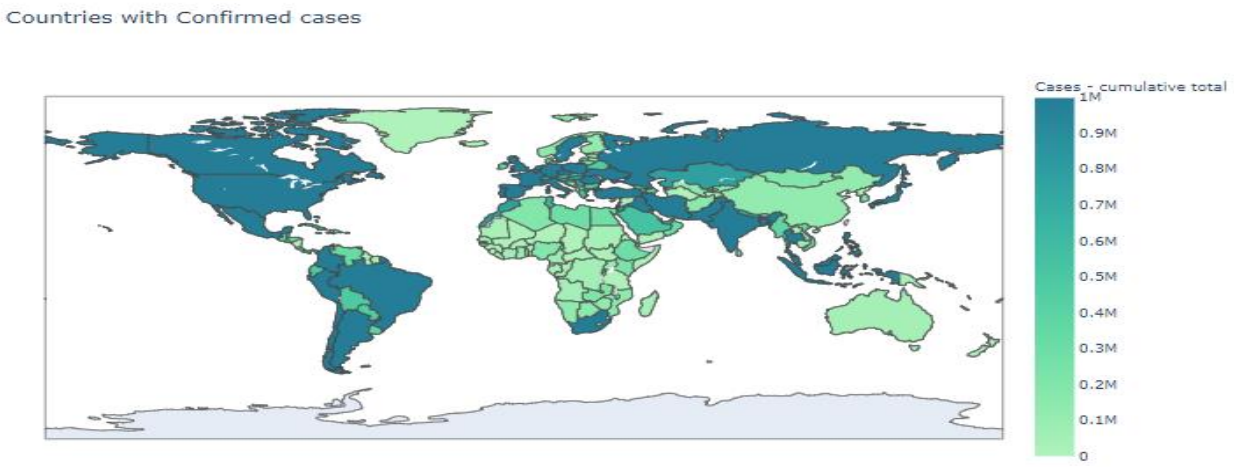


Fig 9 Globally confirmed cases on map

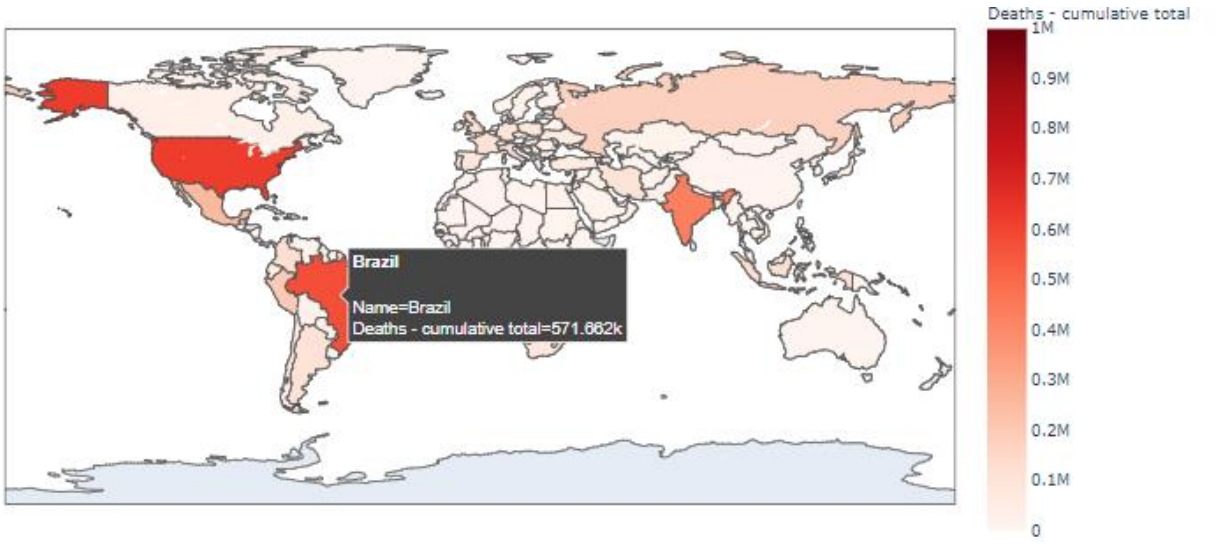


Fig 10 Globally confirmed deaths on map

Global vaccination[Total]

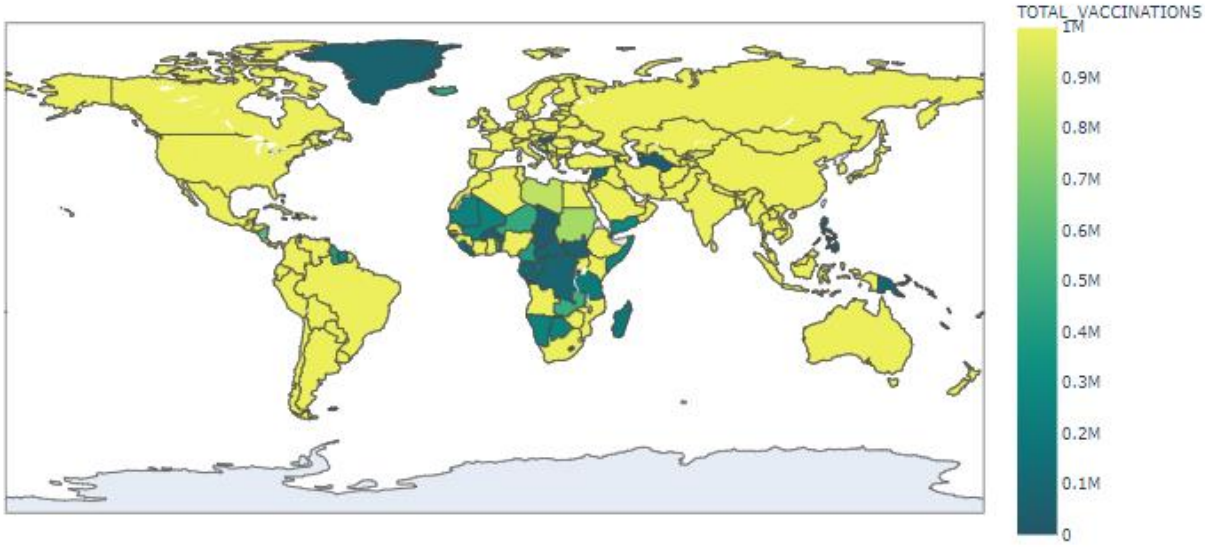
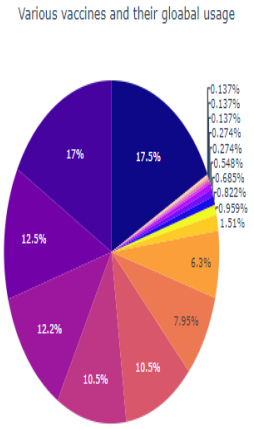
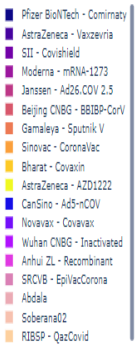


Fig 11 Global visualization of Vaccine administration

From the above 2 maps, it can be observed that the landmasses of Africa, Australia, Greenland, and China are the countries that reported less than half a million COVID cases as of August 19th,2021. The trends are in sync with the deaths reported in Figure 10 where we can see that the US, Brazil, and India are the worst-hit countries with over 0.5 million casualties in each country. From the vaccination chart, it is evident that many countries have more than one million vaccinations administered.

A thorough study of vaccines shows that Comirnaty [US], Vaxzevria [British], Covishield [India], mRNA-1273[US] together account for 60 percent of the global vaccination share.





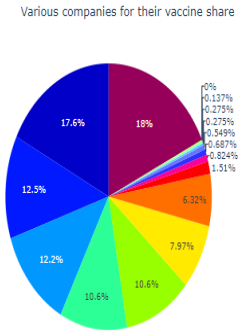


Fig: 12 Vaccines and their global usage

Fig 13 Vaccines manufacturers with their share

In Fig13, we can see the companies' share globally for their vaccines. The interesting point is that AstraZeneca which produces Vaxzevria has a higher production share than usage which hints at wastage. Below is the per country fully vaccinated people trend, India[purple] and the USA [moss green] are depicted as the highest fully vaccinated countries.

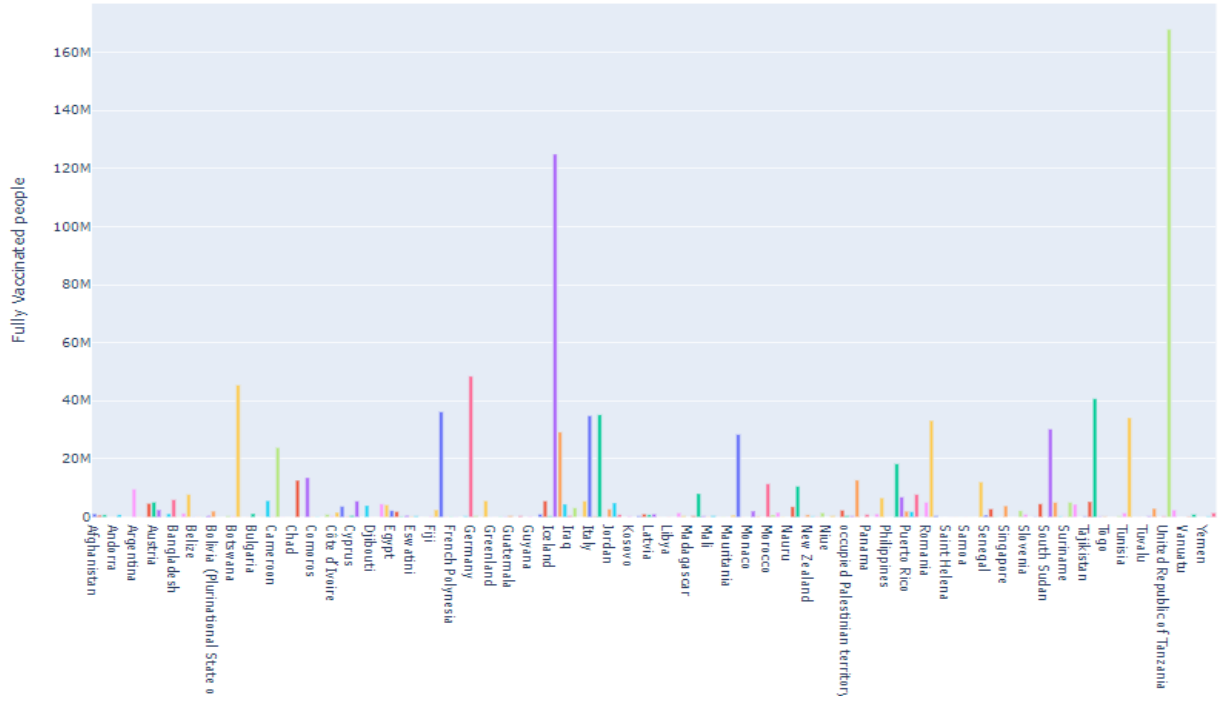


Fig 14 Major countries and their full vaccination population count

From the latest cases data, French Polynesia, Guadeloupe, Martinique, and Dominica are seeing a high rise in the cases per 100000 population. The USA, India, and Iran are seeing a greater number of cases in the last seven days.

Island countries like Palau, and Tuvalu are seeing no cases and they are unaffected since the beginning of the pandemic.

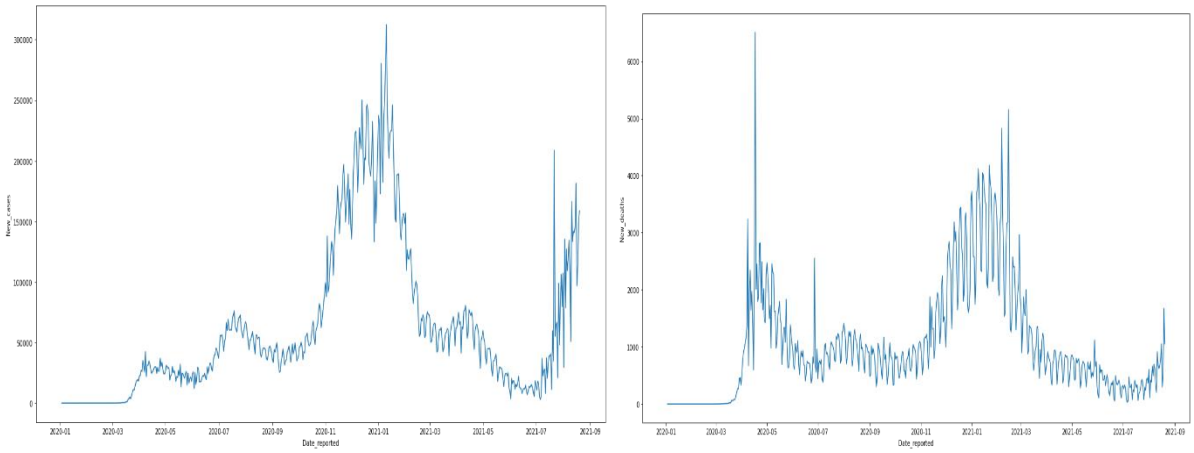


Fig 15 Number of cases (left) and deaths (right) over time for USA

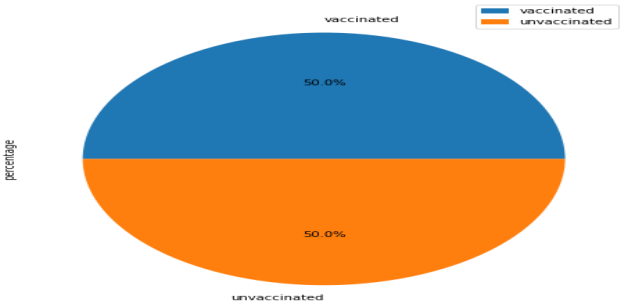


Fig 16(left) Vaccination percentage of USA

Except for China all top ten populated countries are experiencing a high volume of cases though they are not at peaks. From observing the trend in the USA, though there is noticeable vaccination, the cases are on the rise now, but the mortality has lowered.

Discussion:

The analysis results in a prediction that the World might be close to a third peak, but the death rate is on the decline due to COVID-19. The number of cases in 2021 is more than in 2020. The current decline in the death rate could be attributed to immunization as per the trend the in the USA. USA and India are the worst hit countries by the pandemic judging by the number of cases and deaths. Only a little over 10% of the world's population is completely vaccinated, which means that the demand and production are not coherent. There should be an increase in the vaccine percentage in densely populated countries to control the spread of the disease.

This analysis could be helpful for health officials to set targets for their work in the field of tackling COVID-19, they can check the intensity and statistics of their country and other countries and impose health measures they feel suitable. Governments can plan immunization procurement to speed up the vaccination drive by assessing their vaccination level. Businesses and individuals can decide upon travel and commerce activities by monitoring the status of cases worldwide using visuals presented here.

Currently, there are no active cases data available with WHO, if this data is included it would present a wholesome picture of the spread and incubation of the disease. The collection of

vaccination data is not reported daily, data collected real-time would be beneficial in calculating the daily rate of vaccination. A real-time visualizer/update section can be built for the web using this analysis in the future to make it open and accessible immediately on search. Also, a correlation study of public health safety measures with the number of cases, deaths, and vaccination can be done.

Statement of Contribution:

Sri Naga Sai Sushma Kondabolu: Data Cleaning, EDA, Theoretical Analysis, Documentation

References:

1. Data and references are taken from: https://covid19.who.int/info/
2. Motivation for work: https://coronavirus.jhu.edu/about
3. WHO Regions: https://en.wikipedia.org/wiki/List_of_WHO_regions

Appendix:

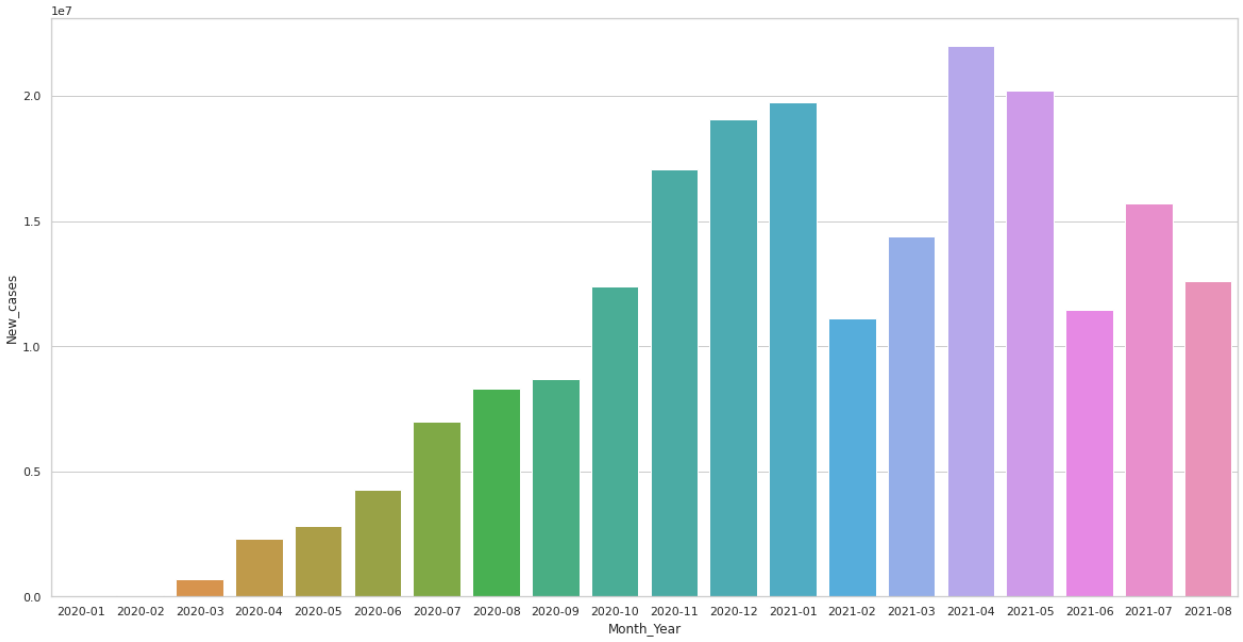


Fig: Monthly count of data

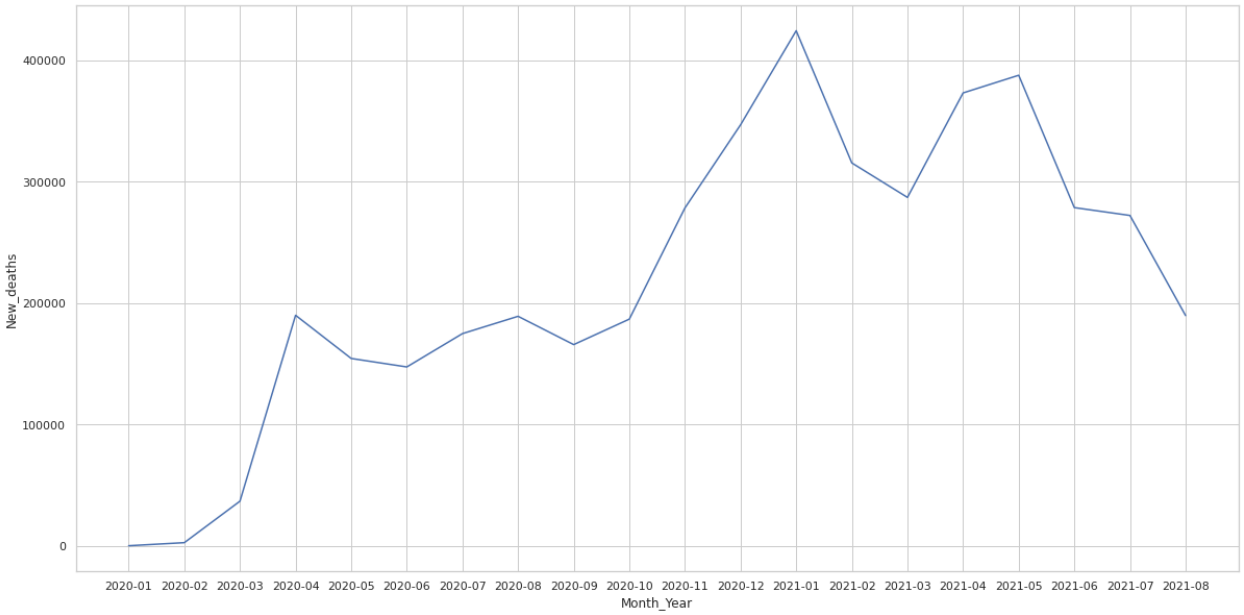


Fig: World Monthly death count

Countries trend data:

Populous (Only two countries data is shown here due to document size, others are available in code):

India:

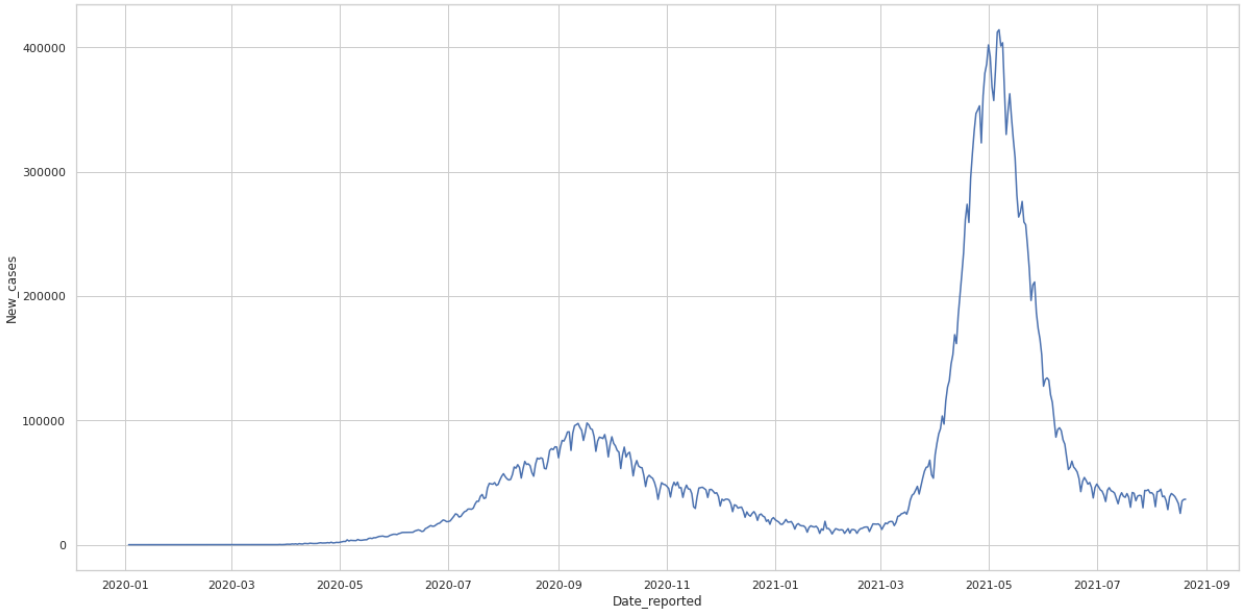


Fig: Cases by date count

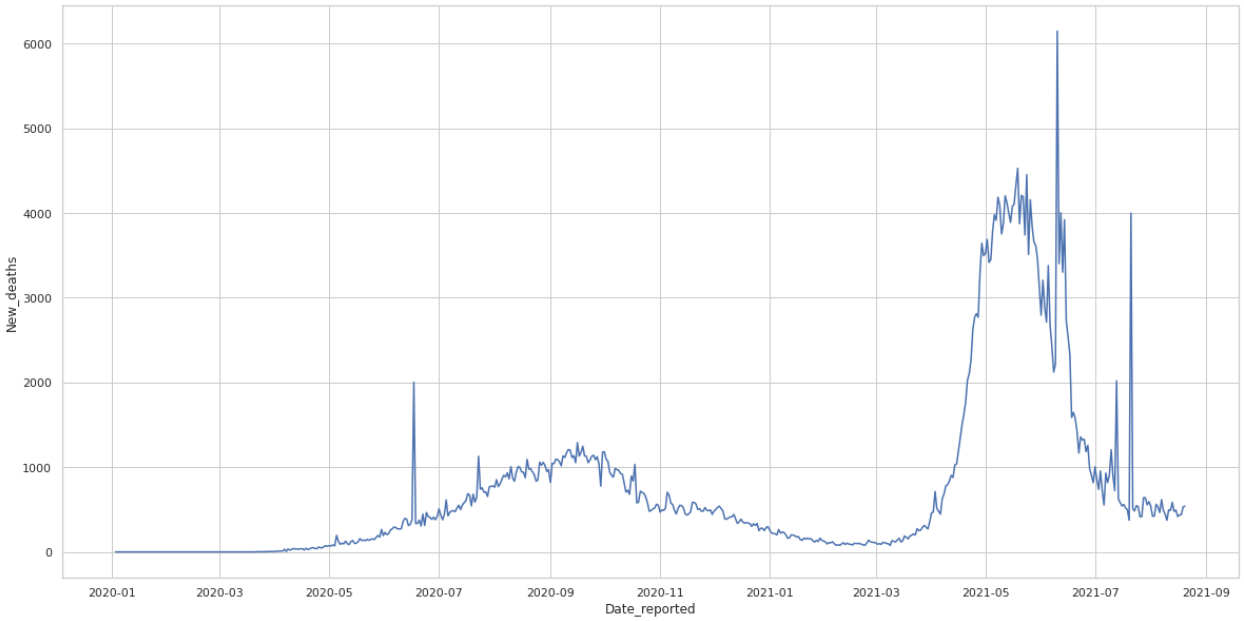


Fig Deaths by date count

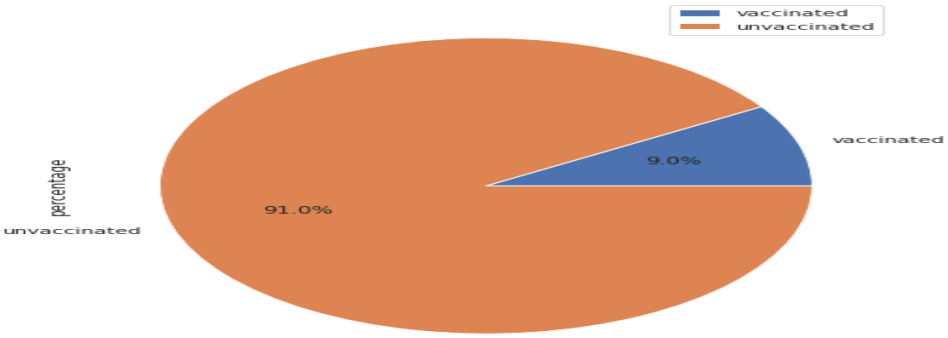


Fig Vaccination percentage of India

China:

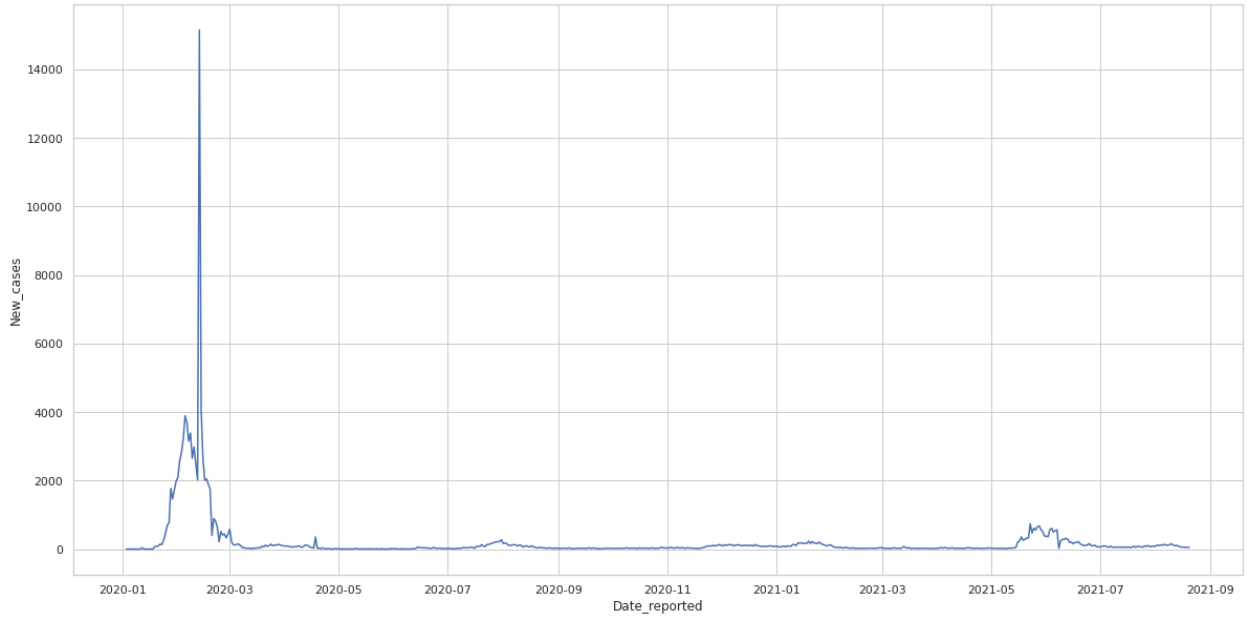


Fig Daily cases count

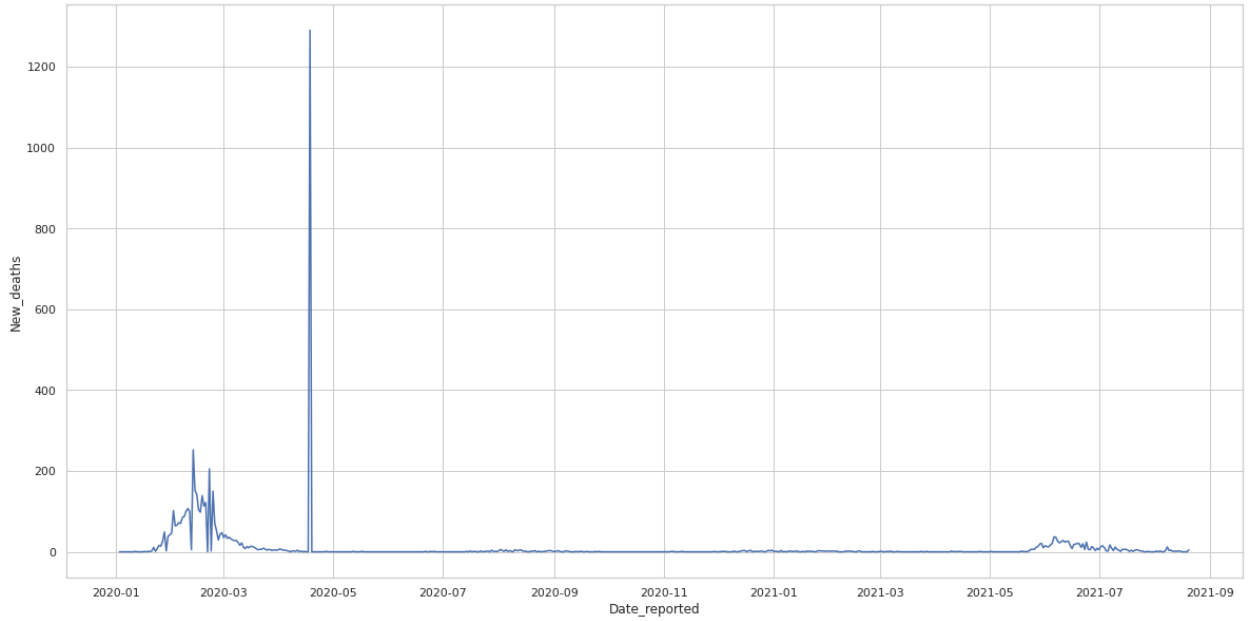


Fig Daily deaths count

China has no reported count of the population that is vaccinated

Data visuals of Countries with the highest vaccination percentage:
Malta:

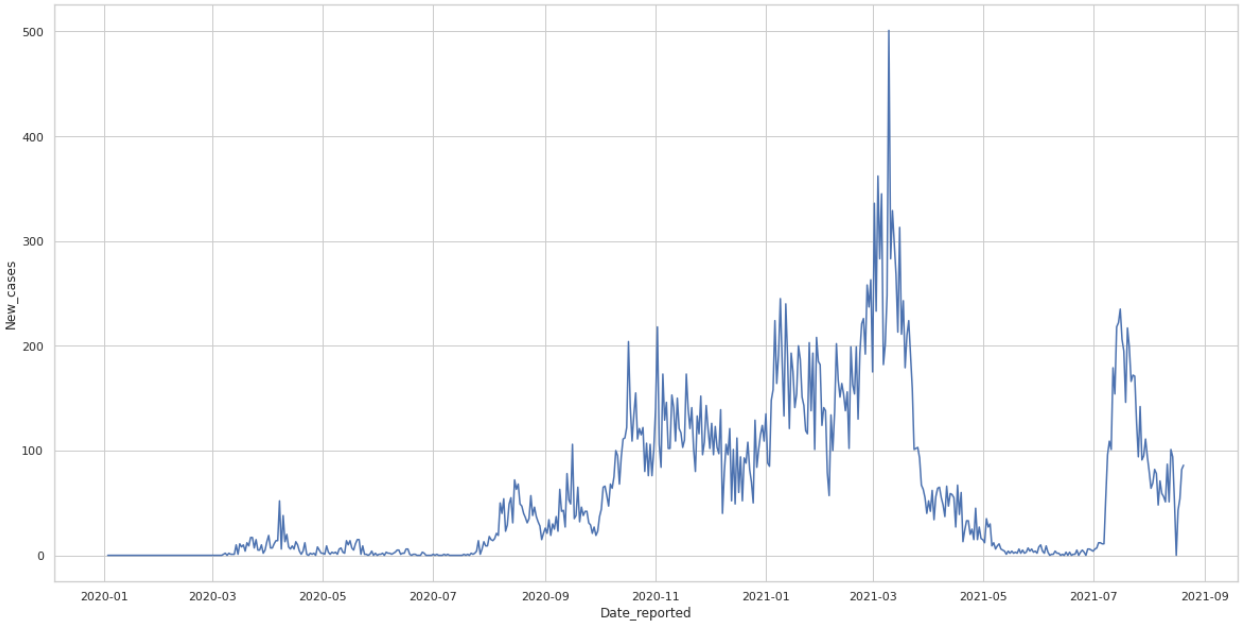


Fig Daily cases count

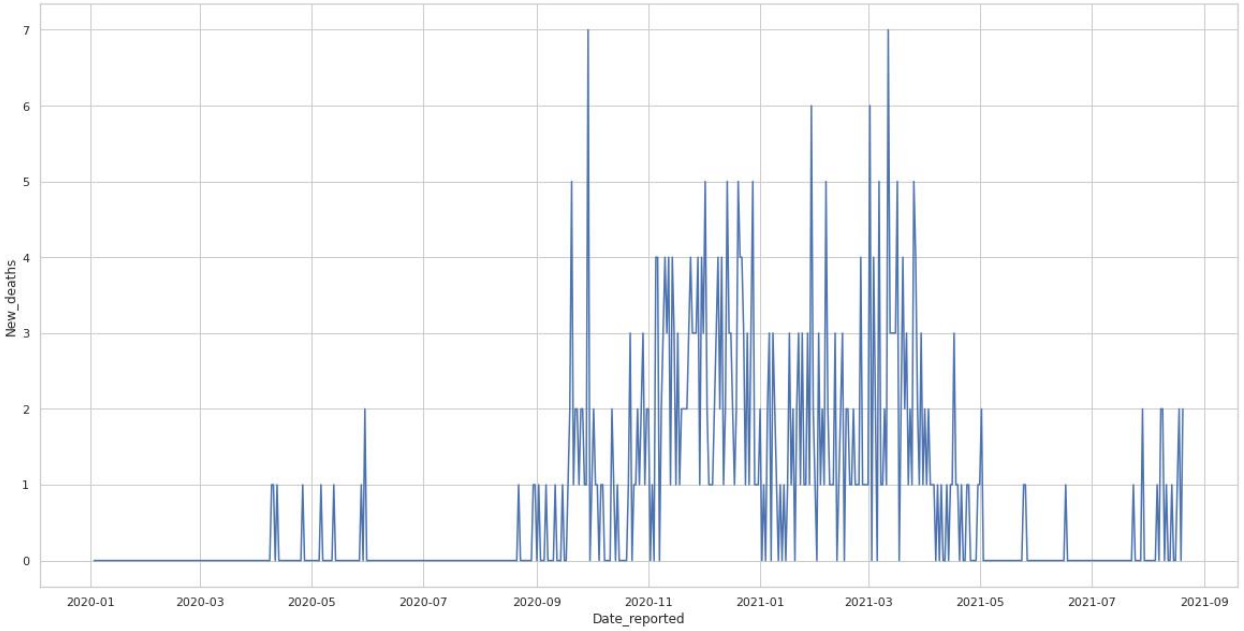


Fig Daily deaths count

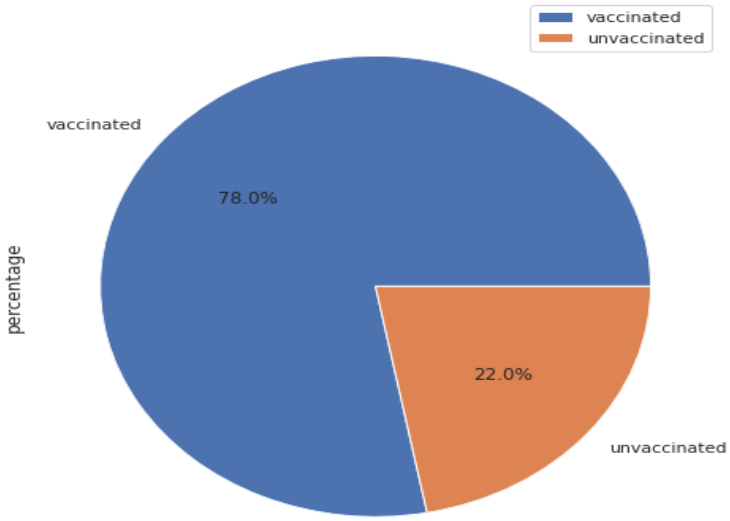


Fig Vaccination percentage

Gibraltar:

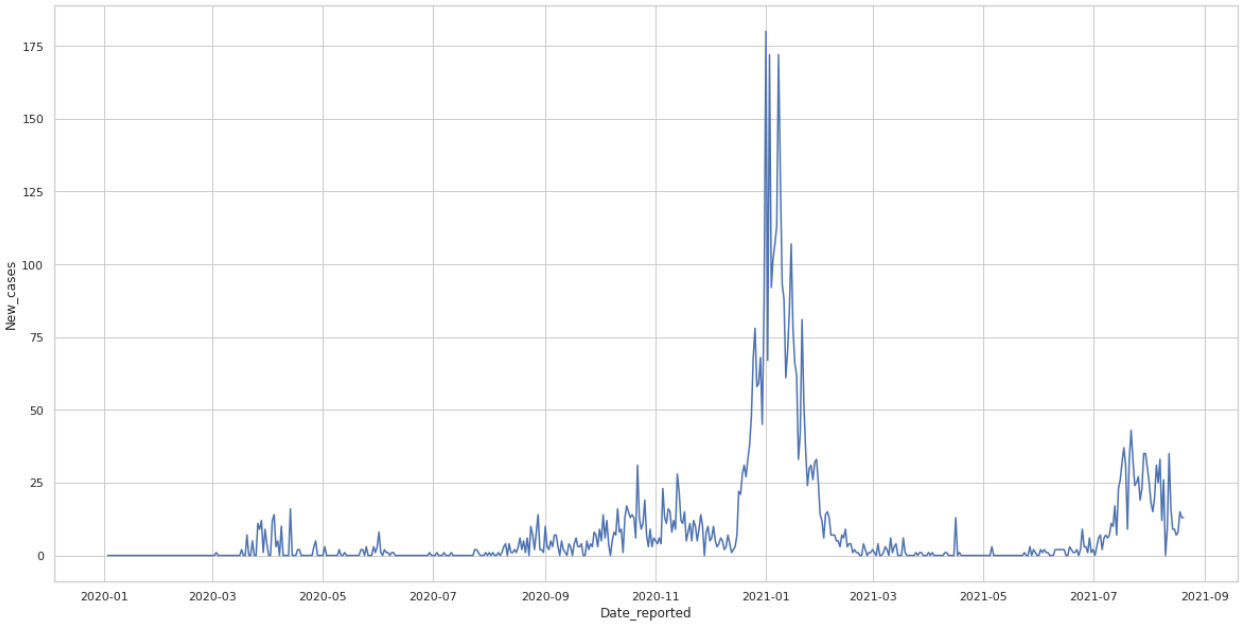


Fig Daily cases count

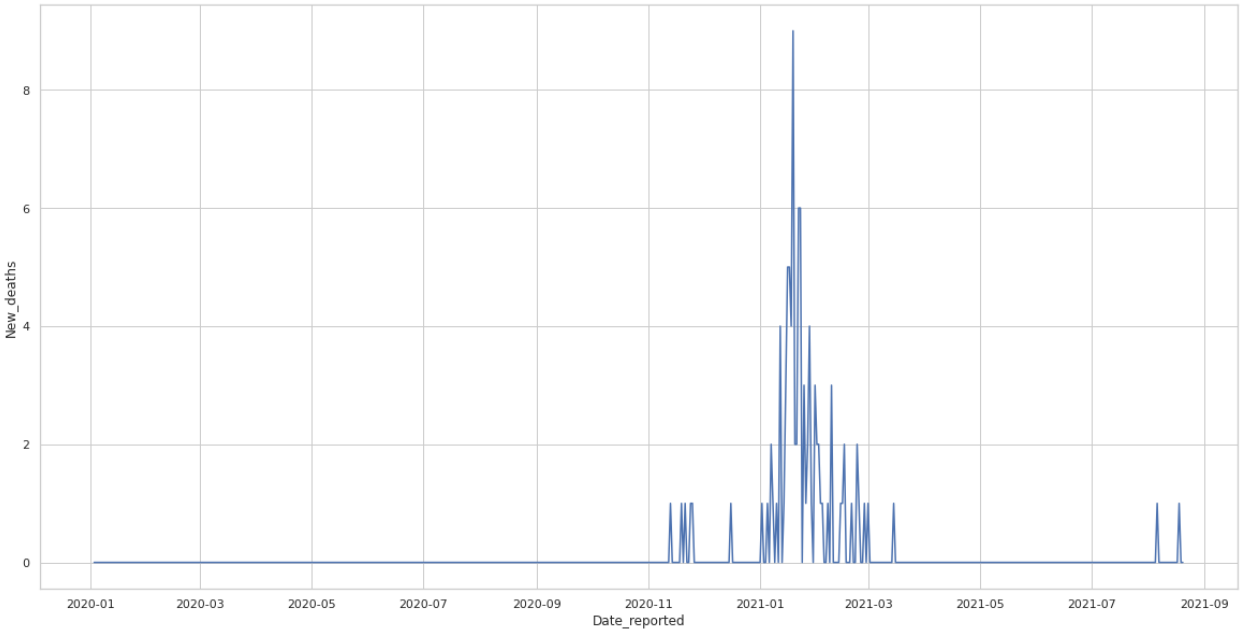


Fig Daily deaths count

Gibraltar has more than 100 percent vaccination, so omitted for piechart

Palau:

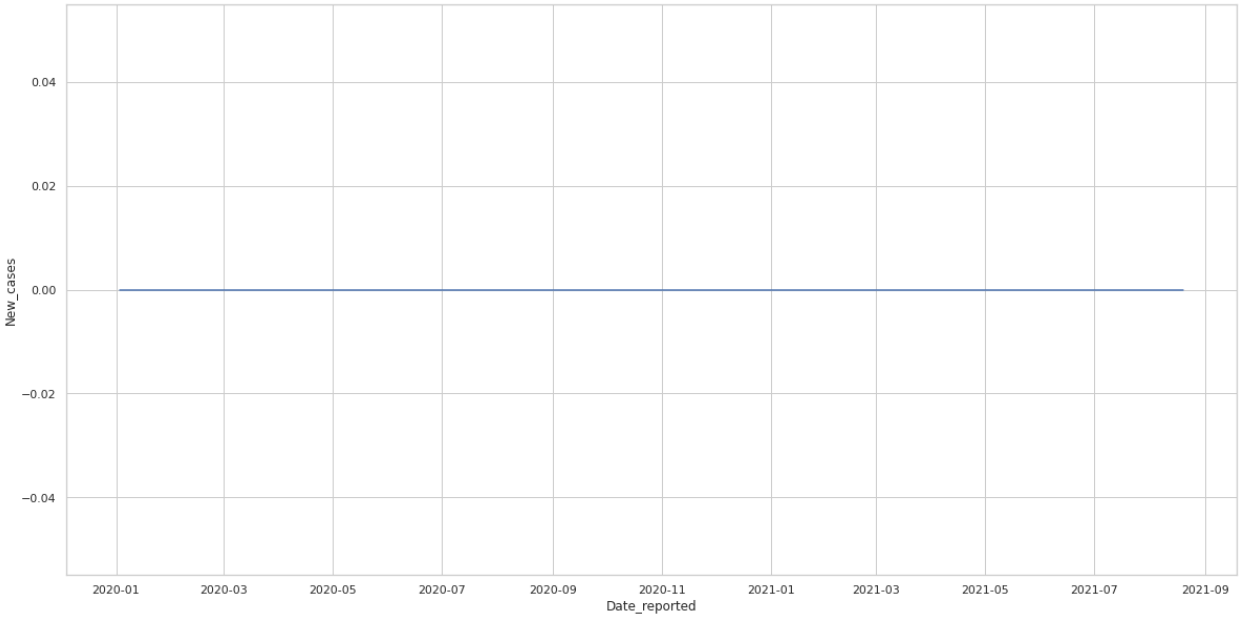


Fig Daily cases count

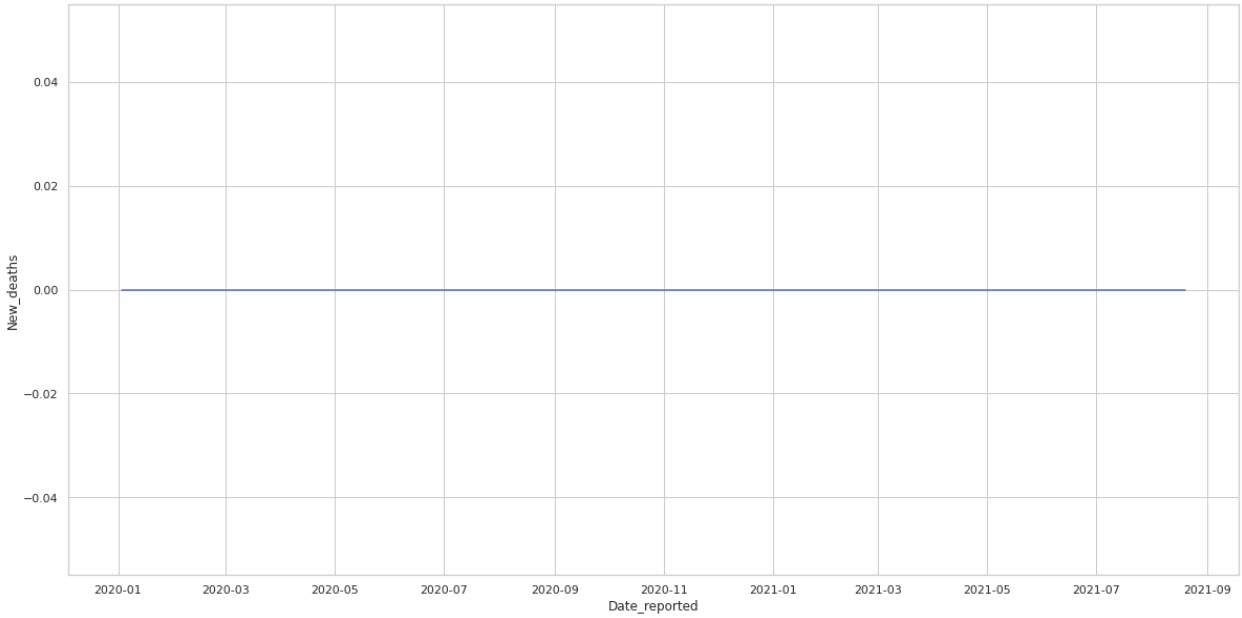


Fig daily deaths count

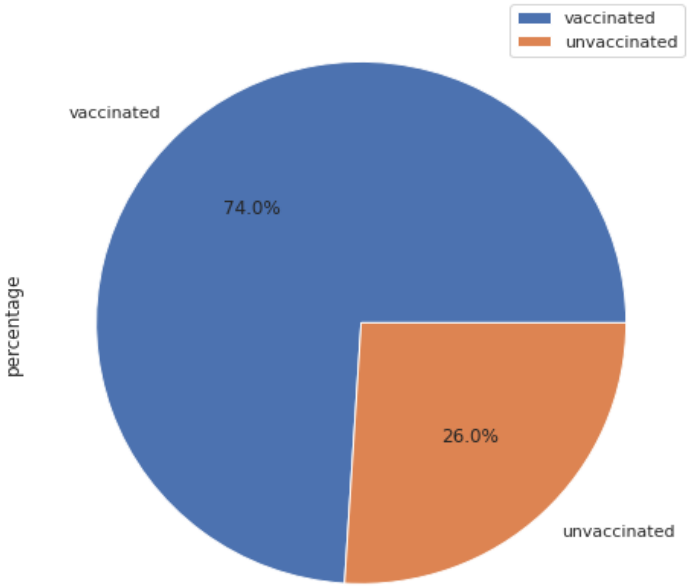


Fig Vaccination percentage

Much information of effect on cases by vaccine can't be gathered from these countries as they are not much populated

Code:

```
import pandas as pd
import numpy as np
from datetime import date
from datetime import timedelta
```

```
import seaborn as sns
import matplotlib.pyplot as plt


daily_cases_data=pd.read_csv('https://covid19.who.int/WHO-COVID-19-global-data.csv')
daily_cases_data


daily_cases_data.isnull().sum()
daily_cases_data[daily_cases_data['Country_code'].isnull()]
daily_cases_data[daily_cases_data['Country_code']==' ']
daily_cases_data.groupby('Date_reported').sum()
daily_cases_data['Date_reported']=pd.to_datetime(daily_cases_data['Date_reported'], format='%Y-%m-%d')
daily_cases_data

fig, ax = plt.subplots(1, 1, figsize=(10, 6))

g1 = sns.lineplot(data=daily_cases_data.groupby('Date_reported').sum(),x='Date_reported',y='New_cases')

#g1.set(xticklabels=[])  # remove the tick labels
g1.set(title='Daily cases count')  # add a title
#g1.set(xlabel='Date')  # remove the axis label

fig, ax = plt.subplots(1, 1, figsize=(10, 6))

f1 = sns.lineplot(data=daily_cases_data.groupby('Date_reported').sum(),x='Date_reported',y='New_deaths')

#g1.set(xticklabels=[])  # remove the tick labels
f1.set(title='Daily deaths count')  # add a title
#g1.set(xlabel='Date')  # remove the axis label

daily_cases_data['Month_Year']=daily_cases_data['Date_reported'].dt.strftime('%Y-%m')
daily_cases_data

monthly_data=daily_cases_data.groupby('Month_Year').sum()
monthly_data.reset_index(inplace=True)
monthly_data
fig, ax = plt.subplots(1, 1, figsize=(20, 10))
ax = sns.barplot(x="Month_Year", y="New_cases", data=monthly_data)
ax
```

```
fig, ax = plt.subplots(1, 1, figsize=(20, 10))
ax = sns.lineplot(x="Month_Year", y="New_deaths", data=monthly_data)
ax

today = date.today()

yesterday = today - timedelta(days=1)

print(yesterday)

def data_by_date(date_chosen):
  cases_by_date=daily_cases_data[daily_cases_data['Date_reported']==str(da
te_chosen)]
  print('The number of cases on ',date_chosen,' is ',cases_by_date['New_ca
ses'].sum())
  print('The number of cases on ',date_chosen,' is ',cases_by_date['New_de
aths'].sum())
  max_cases_country_data=cases_by_date[cases_by_date['New_cases']==cases_b
y_date['New_cases'].max()]
  print('The country with maximum cases on ',date_chosen,' is ',max_cases_
country_data['Country'].values,' with no. of cases: ',max_cases_country_da
ta['New_cases'].values)
  max_deaths_country_data=cases_by_date[cases_by_date['New_deaths']==cases
_by_date['New_deaths'].max()]
  print('The country with maximum cases on ',date_chosen,' is ',max_deaths
_country_data['Country'].values,' with no. of deaths: ',max_cases_country_
data['New_deaths'].values)



data_by_date(yesterday)

def data_by_month(given_month):
  cases_by_month=daily_cases_data[daily_cases_data['Month_Year']==str(give
n_month)]
  print('The number of cases on ',given_month,' is ',cases_by_month['New_c
ases'].sum())
  print('The number of cases on ',given_month,' is ',cases_by_month['New_d
eaths'].sum())
  #max_cases_country_data=monthly_data[cases_by_month['New_cases']==cases_
by_month['New_cases'].max()]
  #print('The country with maximum cases on ',given_month,' is ',max_cases
_country_data['Country'].values,' with no. of cases: ',max_cases_country_d
ata['New_cases'].values)
```

```
  #max_deaths_country_data=cases_by_date[cases_by_month['New_deaths']==cas
es_by_month['New_deaths'].max()]
  #print('The country with maximum cases on ',given_month,' is ',max_death
s_country_data['Country'].values,' with no. of deaths: ',max_cases_country
_data['New_deaths'].values)

data_by_month('2021-01')

daily_cases_data[daily_cases_data['New_cases']==daily_cases_data['New_case
s'].max()]

daily_cases_data[daily_cases_data['New_deaths']==daily_cases_data['New_dea
ths'].max()]

daily_cases_data.groupby('Date_reported').sum().max()
x=daily_cases_data.groupby('Date_reported').sum().max()
x

gen=daily_cases_data.groupby('Date_reported').sum()
gen[gen['New_cases']==x['New_cases']]

gen[gen['New_deaths']==x['New_deaths']]

WHO_Rregs_cases_count=daily_cases_data.groupby(['WHO_region'])['New_cases'
].sum()
WHO_Rregs_cases_count

WHO_Rregs_cases_count.plot.pie(autopct="%.1f%%",figsize=(8, 8));
WHO_Rregs_deaths_count=daily_cases_data.groupby(['WHO_region'])['New_death
s'].sum()
WHO_Rregs_deaths_count

WHO_Rregs_deaths_count.plot.pie(autopct="%.1f%%",figsize=(8, 8));

daily_cases_data['Year']=daily_cases_data['Date_reported'].dt.strftime('%Y
')
daily_cases_data

monthly_data['dummy']=pd.to_datetime(monthly_data['Month_Year'], format='%
Y-%m')
monthly_data['Year']=monthly_data['dummy'].dt.strftime('%Y')
monthly_data['Month'] = monthly_data['dummy'].dt.month_name()
monthly_data

fig, ax = plt.subplots(1, 1, figsize=(20, 10))
```

```
ax=sns.lineplot(x='Month',y='New_cases',hue='Year',data=monthly_data)
ax.set(title='Monthly cases variation for 2020 and 2021')

fig, ax = plt.subplots(1, 1, figsize=(20, 10))
ax=sns.lineplot(x='Month',y='New_deaths',hue='Year',data=monthly_data)
ax.set(title='Monthly deaths variation for 2020 and 2021')

latest_cases_death_transmission_data=pd.read_csv('https://covid19.who.int/
WHO-COVID-19-global-table-data.csv')
cols=latest_cases_death_transmission_data.columns
latest_cases_death_transmission_data.reset_index(inplace=True)
latest_cases_death_transmission_data.drop(columns=['Deaths - newly reporte
d in last 24 hours'],inplace=True)
latest_cases_death_transmission_data.columns=cols
latest_cases_death_transmission_data
latest_cases_death_transmission_data.nlargest(11, 'Cases - cumulative tota
l per 100000 population')
latest_cases_death_transmission_data.nlargest(11, 'Cases - cumulative tota
l')
latest_cases_death_transmission_data.isnull().sum()

latest_cases_death_transmission_data[latest_cases_death_transmission_data[
'Cases - newly reported in last 7 days per 100000 population'].isnull()]
latest_cases_death_transmission_data.nlargest(11, 'Cases - cumulative tota
l')
vaccination_data=pd.read_csv('https://covid19.who.int/who-
data/vaccination-data.csv')
vc=vaccination_data
vaccination_data.isnull().sum()
vaccination_data['FIRST_VACCINE_DATE'].fillna('Not available',inplace=True
)
vaccination_data['VACCINES_USED'].fillna('Not available',inplace=True)
vaccination_data['PERSONS_FULLY_VACCINATED_PER100'].fillna('Not available'
,inplace=True)

vaccination_data['PERSONS_FULLY_VACCINATED'].fillna('Not available',inplac
e=True)
vaccination_data['PERSONS_VACCINATED_1PLUS_DOSE_PER100'].fillna('Not avail
able',inplace=True)
vaccination_data['TOTAL_VACCINATIONS_PER100'].fillna('Not available',inpla
ce=True)
vaccination_data['TOTAL_VACCINATIONS'].fillna('Not available',inplace=True
)
def country_data(country):
  cdf=daily_cases_data[daily_cases_data['Country']==country]
```

```
  cdf['Month_Year']=cdf['Date_reported'].dt.strftime('%Y-%m')
  cdf['Year'] = cdf['Date_reported'].dt.strftime('%Y')
  cdf['day']=cdf['Date_reported'].dt.strftime('%m-%d')
  total_t_cou=latest_cases_death_transmission_data[latest_cases_death_tran
smission_data['Name']==country]
  vac_dat=vaccination_data[vaccination_data['COUNTRY']==country]
  print('Total number of cases for ',country,' is ',total_t_cou['Cases - c
umulative total'].values)
  print('Total number of deaths for ',country,' is ',total_t_cou['Deaths -
 cumulative total'].values)
  print('Total number of cases in the last seven days for ',country,' is '
,total_t_cou['Cases - newly reported in last 7 days'].values)
  print('Total number of deaths in the last seven days for ',country,' is
',total_t_cou['Deaths - newly reported in last 7 days'].values)
  print('Total number of cases in the last seven days per 100000 people fo
r ',country,' is ',total_t_cou['Cases - newly reported in last 7 days per
100000 population'].values)
  print('Total number of deaths in the last seven days per 100000 people f
or ',country,' is ',total_t_cou['Deaths - cumulative total per 100000 popu
lation'].values)
  print('First vaccination date for ',country,' is ',vac_dat['FIRST_VACCIN
E_DATE'].values)
  fig, ax = plt.subplots(1, 1, figsize=(20, 10))
  ax=sns.lineplot(x='Date_reported',y='New_cases',data=cdf)
  figd, axd = plt.subplots(1, 1, figsize=(20, 10))
  axd=sns.lineplot(x='Date_reported',y='New_deaths',data=cdf)
  #fig1, ax1 = plt.subplots(1, 1, figsize=(20, 10))
  #ax1=sns.lineplot(x='Month',y='New_cases',hue='Year',data=cdf)
  #fig2, ax2 = plt.subplots(1, 1, figsize=(20, 10))
  #ax2=sns.lineplot(x='Month',y='New_deaths',hue='Year',data=cdf)
  if country=='China':
    print('No vaccine data available')
  elif country=='Gibraltar':
    print('Reported vaccination percentage is more than 100')
  else:
    vr=vac_dat['PERSONS_FULLY_VACCINATED_PER100'].values
    uvr=100-int(vr[0])
    cnt_vac=pd.DataFrame({'percentage':[int(vr[0]),uvr]},index=['vaccinate
d','unvaccinated'])
    cnt_vac.plot.pie(y='percentage',autopct="%.1f%%",figsize=(8, 8));
country_data('India')
country_data('United States of America)
country_data('Pakistan')
country_data('Bangladesh')
country_data('China')
```

```
country_data('Russian Federation')
country_data('Iran')
country_data('Indonesia')
country_data('Mexico')
country_data('Nigeria')
vc

vc=pd.read_csv('https://covid19.who.int/who-data/vaccination-data.csv')
vc['PERSONS_FULLY_VACCINATED'] = vc['PERSONS_FULLY_VACCINATED'].astype(str
).astype(float)
vc.dtypes
vc['PERSONS_FULLY_VACCINATED'].sum()
vc['TOTAL_VACCINATIONS'].sum()
ctr=latest_cases_death_transmission_data
import plotly.express as px
figure = px.choropleth(ctr,locations='Name', locationmode='country names',
 color='Cases - cumulative total', hover_name='Name', color_continuous_sca
le='tealgrn', range_color=[1,1000000],title='Countries with Confirmed case
s',width=1000,height=600)
figure.show()
deathfigure = px.choropleth(ctr,locations='Name', locationmode='country na
mes', color='Deaths - cumulative total', hover_name='Name', color_continuo
us_scale='reds', range_color=[1,1000000],title='Countries with Confirmed d
eaths',width=1000,height=600)
deathfigure.show()
vacc=pd.read_csv('https://covid19.who.int/who-data/vaccination-data.csv')

vacc.fillna(value = 0, inplace = True)
vacc.TOTAL_VACCINATIONS = vacc.TOTAL_VACCINATIONS.astype(int)
vacc.PERSONS_VACCINATED_1PLUS_DOSE = vacc.PERSONS_VACCINATED_1PLUS_DOSE.as
type(int)
vacc.PERSONS_FULLY_VACCINATED = vacc.PERSONS_FULLY_VACCINATED.astype(int)
date = vacc.DATE_UPDATED.str.split('-', expand =True)
date
vacc['year'] = date[0]
vacc['month'] = date[1]
vacc['day'] = date[2]
vacc.year = pd.to_numeric(vacc.year)
vacc.month = pd.to_numeric(vacc.month)
vacc.day = pd.to_numeric(vacc.day)
vacc.date = pd.to_datetime(vacc.DATE_UPDATED)
vacc.head()
print('Data point starts from ',vacc.date.min())
print('Data point ends at ',vacc.date.max())
print('Total no of countries in the data set ',len(vacc.COUNTRY.unique()))
```

```
print('Total no of unique vaccines in the data set ',len(vacc.VACCINES_USE
D.unique()))
from wordcloud import WordCloud
import plotly.express as px
sns.set(color_codes = True)
sns.set(style="whitegrid")
wordCloud = WordCloud(
    background_color='white',
    max_font_size = 50).generate(' '.join(vacc.COUNTRY))
plt.figure(figsize=(15,7))
plt.axis('off')
plt.imshow(wordCloud)
plt.show()
country_wise_total_vaccinated = {}
for COUNTRY in vacc.COUNTRY.unique() :
    vaccinated = 0
    for i in range(len(vacc)) :
        if vacc.COUNTRY[i] == COUNTRY :
            vaccinated += int(vacc.PERSONS_FULLY_VACCINATED[i])
    country_wise_total_vaccinated[COUNTRY] = vaccinated

    country_wise_total_vaccinated_df = pd.DataFrame.from_dict(country_wise
_total_vaccinated,
                                                              orient='index',
                                                              columns = ['total
_vaccinted_till_date'])

country_wise_total_vaccinated_df.sort_values(by = 'total_vaccinted_till_da
te', ascending = False, inplace = True)
country_wise_total_vaccinated_df
fig = px.bar(country_wise_total_vaccinated_df,
             y = 'total_vaccinted_till_date',
             x = country_wise_total_vaccinated_df.index,
             color = 'total_vaccinted_till_date',
             color_discrete_sequence= px.colors.sequential.Viridis_r
            )
fig.update_layout(
    title={
            'text' : "Vaccination till date in various countries",
            'y':0.95,
            'x':0.5
        },
    xaxis_title="Countries",
    yaxis_title="Total vaccinated")
fig.show()
```

```
vaccination_meta_data=pd.read_csv('https://covid19.who.int/who-
data/vaccination-metadata.csv')
vaccination_meta_data.isnull().sum()
vaccm=vaccination_meta_data
del vaccm['END_DATE']
del vaccm['COMMENT']
vaccm.isnull().sum()
def plot_pie(value, title, color) :
    new_dict = {}
    for v in vaccm[value].unique() :
        value_count = 0
        for i in range(len(vaccm)) :
            if vaccm[value][i] == v :
                value_count += 1
        new_dict[v] = value_count

    new_df = pd.DataFrame.from_dict(new_dict, orient = 'index', columns =
['Total'])
    if color == 'plasma' :
        fig = px.pie(new_df, values= 'Total',
                     names = new_df.index,
                     title = title,
                     color_discrete_sequence=px.colors.sequential.Plasma)
    elif color == 'rainbow' :
        fig = px.pie(new_df, values= 'Total',
                     names = new_df.index,
                     title = title,
                     color_discrete_sequence=px.colors.sequential.Rainbow)
    else :
        fig = px.pie(new_df, values= 'Total',
                     names = new_df.index,
                     title = title)
    fig.update_layout(
        title={
            'y':0.95,
            'x':0.5
        })
    return fig.show()
plot_pie('VACCINE_NAME', 'Various vaccines and their gloabal usage', 'plas
ma')
plot_pie('COMPANY_NAME', 'Various companies for their vaccine share', 'rai
nbow')
wordCloud = WordCloud(
    background_color='white',
```

```
    max_font_size = 50).generate(' '.join(vaccm.VACCINE_NAME))
plt.figure(figsize=(12,5))
plt.axis('off')
plt.imshow(wordCloud)
plt.show()
fig = px.bar(vacc, x = 'COUNTRY', y ='PERSONS_FULLY_VACCINATED', color = '
COUNTRY',width=1500,height=800)
fig.update_layout(
    title={
            'text' : "vaccination trend",
        },
    yaxis_title="Fully Vaccinated people"
)
fig.show()
#IND VS USA

india_usa = [vacc[vacc.COUNTRY == 'United States of America'], vacc[vacc.C
OUNTRY == 'India']]
res = pd.concat(india_usa)
fig = px.bar(res, x = 'COUNTRY', y ='TOTAL_VACCINATIONS', color = 'TOTAL_V
ACCINATIONS',width=1000, height=500)
fig.update_layout(
    title={
            'text' : "Total vaccinated - India vs USA",
            'y':0.95,
            'x':0.5
        },
    yaxis_title="Total Vaccinations"
)
fig.show()
#top 5

top5 = [vacc[vacc.COUNTRY == 'United States of America'], vacc[vacc.COUNTR
Y == 'India'], vacc[vacc.COUNTRY == 'China'],vacc[vacc.COUNTRY == 'Brazil'
],vacc[vacc.COUNTRY == 'Japan']]
res = pd.concat(top5)
fig = px.bar(res, x = 'COUNTRY', y ='TOTAL_VACCINATIONS', color = 'TOTAL_V
ACCINATIONS',width=1200, height= 600)
fig.update_layout(
    title={
            'text' : "Total vaccinations of Top 5 countries",
            'y':0.95,
            'x':0.5
        },
    yaxis_title="Total Vaccinations"
```

```
)
fig.show()
latest_cases_death_transmission_data.nlargest(11,'Cases - newly reported i
n last 7 days')
latest_cases_death_transmission_data.nlargest(11,'Cases - newly reported i
n last 7 days per 100000 population')
vacc.nlargest(10,'PERSONS_FULLY_VACCINATED_PER100')
country_data('Palau')
country_data('Malta')
country_data('Gibraltar')
latest_cases_death_transmission_data.nsmallest(11,'Cases - cumulative tota
l')
latest_cases_death_transmission_data.nsmallest(20,'Cases - newly reported
in last 7 days')
```