# Covid­-19 Vaccine Analysis

Phase-4

**Dataset:**

Dataset consist of weekly confirmed cases and weekly cumulative confirmed cases for 35 weeks. Then the distribution of the data was examined using the most up-to-date Covid-19 weekly case data and its parameters were obtained according to the statistical distributions.

**Performing exploratory data analysis:**

The goal of this research is to analyze data on vaccinations, vaccination administration, and forecasting vaccination rates on a country-by-country basis for the general public, policymakers, vaccine manufacturers, national governments, and international governments to better understand the current state of COVID-19 vaccination. In this study, two public datasets were used: the Johns Hopkins University corona virus 2019 dataset and Our World in Data - Corona virus Pandemic dataset. With datasets, two approaches have been used: visual data analysis for COVID-19 vaccine administration and the auto regressive integrated moving average (ARIMA) model for forecasting vaccination rates. The findings confirm that Oxford/AstraZeneca is the top vaccine used across the globe with 26.54%, the United States is the top in vaccination, with 277,290,173, India is the top in number of daily vaccinations with 3.659357M, and in total vaccinations per hundred people, the United States has the highest count with 82.91, among the top five countries. It is also estimated that the vaccination rate in the United States will reach almost 60%, while India, Brazil, France, and Turkey will reach about 15%, 28%, 60%, and 23%, respectively, in the following 50 days beginning 20 May 2021. This exploratory study of COVID-19 vaccination data was carried out to effectively show the current state of COVID-19 vaccine administration and to anticipate vaccination rates in the United States, India, Brazil, France, and Turkey.

## **Introduction:**

The COVID-19 outbreak, officially identified as the corona virus disease outbreak, would be a continuing major worldwide public health problem of coronavirus disease 2019 (COVID-19) impacted by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The epidemic happened in Wuhan, China, in December 2019. The World Health Organization classified COVID-19 a Public Health Emergency of International Concern on January 30, 2020 and a pandemic on March 11, 2020. More than 167 million cases had occurred as of May 24, 2021, with more than 3.46 million authenticated fatalities attributed to COVID-19, making it one of the worst pandemics in history.

Since this virus's specific source is uncertain, the very first epidemic occurred in late 2019 in Wuhan, Hubei, China (To et al. [2021](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR23" \o )). Several previous instances of COVID-19 were connected to persons who may have travelled the Huanan Seafood Wholesale Market in Wuhan (Sun et al. [2020](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR21" \o )), though living person dissemination may have occurred prior to this (Hu et al. [2021](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR7" \o )). On February 11, 2020, the World Health Organization (WHO) called the sickness "COVID-19," short for coronavirus disease 2019 (World Health Organization [2020](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR24" \o )). The infection which sparked the pandemic is identified as SARS-CoV-2, a recently found virus which is directly connected to bat coronaviruses (Perlman [2020](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR18" \o )), pangolin coronaviruses (Zhang et al. [2020](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR25" \o )), and SARS-CoV (European Centre for Disease Prevention and Control: Risk assessment [2020](https://link.springer.com/article/10.1007/s43538-022-00064-7" \l "ref-CR5" \o )).

## **Importing Libraries:**

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

from plotly.offline import download\_plotlyjs,init\_notebook\_mode,plot,iplot

import plotly.graph\_objects as go

import plotly.figure\_factory as ff

from plotly.colors import n\_colors

from wordcloud import WordCloud,ImageColorGenerator

init\_notebook\_mode(connected=True)

from plotly.subplots import make\_subplots

from pywaffle import Waffle

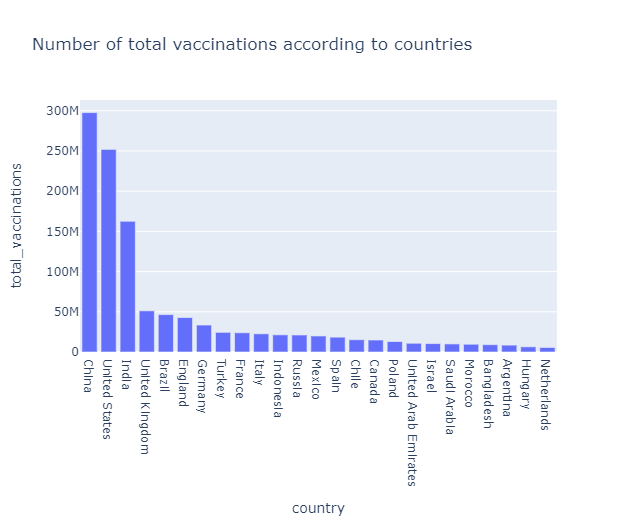
import warnings

warnings.filterwarnings("ignore")

The number of total vaccinations & daily vaccinations according to countries?

data

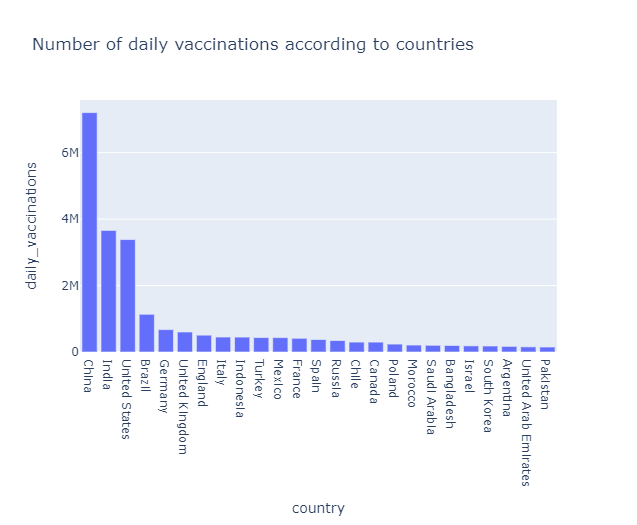
new\_df[['country','total\_vaccinations']].nlargest(25,'total\_vaccinations')

fig = px.bar(data, x = 'country',y = 'total\_vaccinations',title="Number of total vaccinations according to countrie

data = new\_df[['country','daily\_vaccinations']].nlargest(25,'daily\_vaccinations')

fig = px.bar(data, x = 'country',y = 'daily\_vaccinations',title="Number of daily vaccinations according to countries",)

fig.show()



**STATISTICAL ANALYSIS:**

Statistical analysis is a set of methods used to make sense of data by identifying patterns, relationships, and trends. It involves collecting, organizing, and interpreting data to draw meaningful conclusions. Key components of statistical analysis include:

**Descriptive Statistics:** Summarizing data through measures like mean, median, mode, standard deviation, and range to provide an overview of the data's characteristics.

**Inferential Statistics:** Making predictions or inferences about a population based on a sample. This includes hypothesis testing, confidence intervals, and regression analysis.

**Data Visualization:** Creating graphs, charts, and plots to visualize data, helping to understand patterns and trends.

**Probability:** Evaluating the likelihood of events occurring and using probability distributions like the normal distribution, binomial distribution, etc.

**Hypothesis Testing:** Testing hypotheses to determine whether observed differences or relationships in data are statistically significant.

**Regression Analysis:** Assessing relationships between and variables making predictions based on those relationships.

Statistical analysis is crucial in various fields, including science, business, healthcare, and social sciences, formaking data-driven decisions and drawing reliable conclusions fromdata

**CODING:**

path="../input/covid19-daily-data-updated/Covid-19\_data\_version\_8.csv"covid\_data=pd.read\_csv(path, parse\_dates=True)*#vaccine\_data.tail()*covid\_data=covid\_data.sort\_values(by='date')

covid\_data.people\_vaccinated\_per\_hundred=covid\_data.people\_vaccinated\_per\_hundred/10covid\_data=covid\_data.rename(columns={'people\_vaccinated\_per\_hundred':'people\_vaccinated\_per\_thousand'})

covid\_data.columns

**OUTPUT:**

Index(['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',

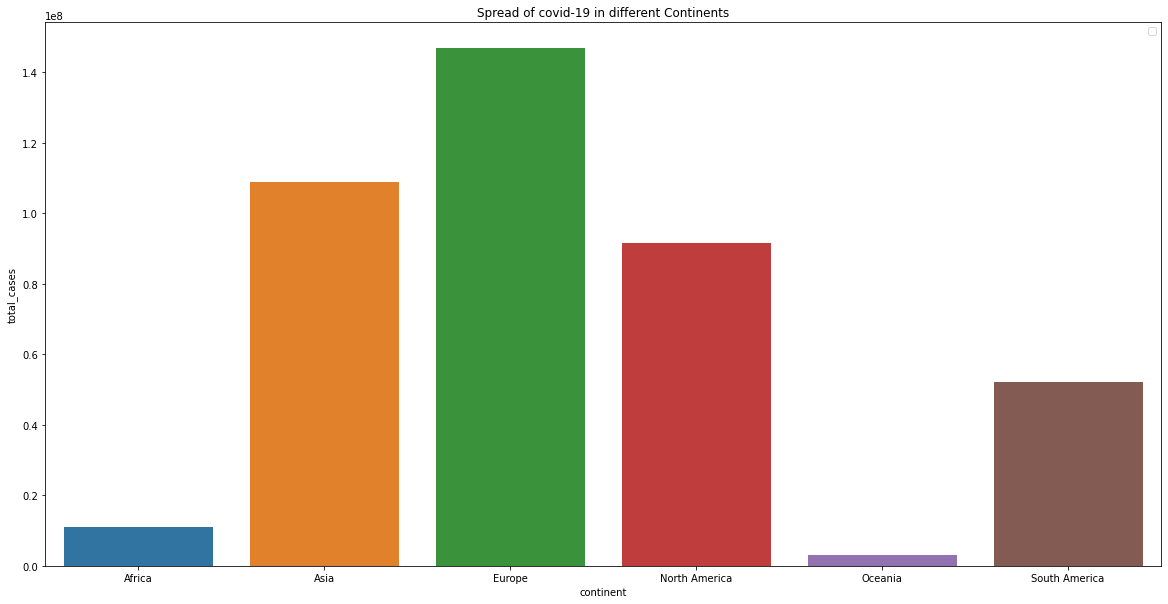
covid\_data\_asia=covid\_data.continent=='Asia'plt.figure(figsize=(20,10))*#covid\_data\_cont.isnull=0*plt.title('Spread of covid-19 in different Continents')

sns.barplot(x=cases.index , y= cases.total\_cases)plt.legend()

Out[11]:

**OUTPUT:**

<matplotlib.legend.Legend at 0x7fbf2cc44d90>



covid\_data\_asia=covid\_data.continent=='Asia'plt.figure(figsize=(20,10))*#covid\_data\_cont.isnull=0*plt.title('Total deaths due to covid-19 of smokers ')

sns.barplot(x=smoker\_deaths.index, y= smoker\_deaths.total\_deaths)plt.legend

**Output:**

<function matplotlib.pyplot.legend(\*args, \*\*kwargs)>

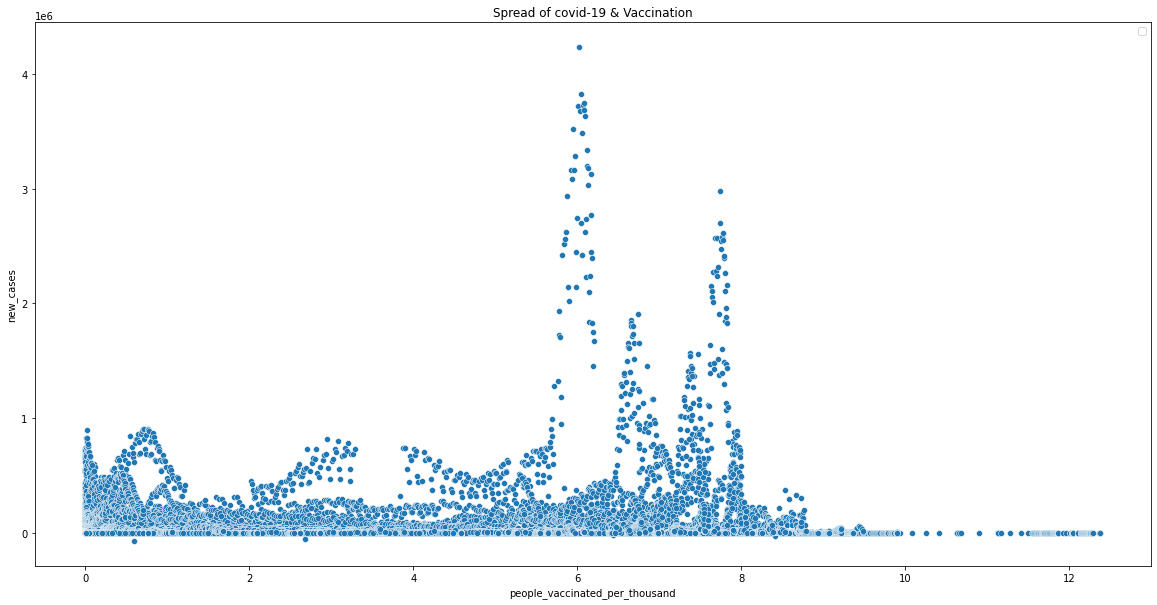


covid\_data\_asia=covid\_data.continent=='Asia'plt.figure(figsize=(20,10))*#covid\_data\_cont.isnull=0*plt.title('Spread of covid-19 & Vaccination')

sns.scatterplot(x=covid\_data.people\_vaccinated\_per\_thousand , y= covid\_data.new\_cases)plt.legend()

**OUTPUT:**

<matplotlib.legend.Legend at 0x7fbf15d1f49b0>



**VISUALIZATION:**

Visualization is the process of representing data, information, or concepts in a graphical or visual format. It helps individuals understand complex ideas, identify patterns, and make data-driven decisions by presenting information in a more intuitive and accessible manner. Various types of visualizations, such as charts, graphs, maps, and diagrams, can be used to convey information effectively. Visualizations are commonly used in fields like data analysis, science, business, and design to communicate ideas and insights in a way that is easy to comprehend.

**CODING:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data = pd.read\_csv('case\_time\_series.csv')

Y = data.iloc[61:,1].values

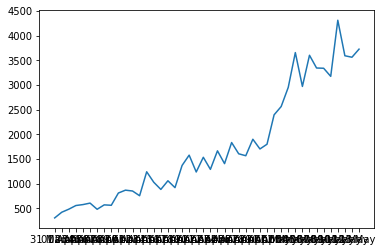
R = data.iloc[61:,3].values

D = data.iloc[61:,5].values

X = data.iloc[61:,0]

plt.plot(X,Y)

**OUTPUT:**



 fig=px.bar(x=datewise.index,y=datewise["Confirmed"]-datewise["Recovered"]-datewise["Deaths"])fig.update\_layout(title="Distribution of Number of Active Cases",

xaxis\_title="Date",yaxis\_title="Number of Cases",)fig.show()

