College code:7177

Team members: Komathilakshmi D

Preethi G

Sheelajemi L

Megavannan M D

COVID VACCINES ANALYSIS USING DATA ANALYCTICS WITH COGNOS

PROJECT TITLE:

COVID vaccines analysis with IBM cognos.

PROJECT OVERVIEW:

 The problem is to conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. This project involves data collection, data preprocessing, exploratory data analysis, statistical analysis, and visualization.

ABSTRACT:

This project employs advanced data analytics techniques to assess the real-world effectiveness and societal impact of COVID-19 vaccines. Leveraging comprehensive datasets, our analysis examines vaccination rates, infection rates, hospitalizations, and mortality rates across diverse populations. Through statistical modeling, machine learning, and epidemiological insights, we aim to provide valuable insights for policymakers and public health officials to optimize vaccine distribution strategies and mitigate the ongoing pandemic's effects.

PROJECT OBJECTIVES :

* Effectiveness Assessment
* Vaccination Rate Analysis
* Infection Tracking
* Hospitalization Evaluation
* Mortality Rate Examination
* Population-Based Insights

KEY STEPS:

* Data Collection
* Data Preprocessing
* Exploratory Data Analysis
* Statistical Analysis
* Visualization
* Insights and Recommendations

DATA COLLECTION:

Collect data on COVID-19 vaccine efficacy, distribution, and adverse effects from a variety of sources, such as clinical trials, public health surveillance systems, and self-reported surveys.

DATA PREPROCESSING:

Clean and prepare the data for analysis by removing errors, inconsistencies, and missing values.

EXPLORATORY DATA ANALYSIS:

Use visualization and statistical methods to identify patterns and trends in the data.

STATISTICAL ANALYSIS:

Conduct hypothesis tests and develop predictive models to assess the effectiveness, distribution, and safety of COVID-19 vaccines.

VISUALIZATION:

Create informative and engaging visualizations to communicate the findings to policymakers and health organizations.

INSIGHTS AND RECOMMENDATIONS:

Insights into the effectiveness of different COVID-19 vaccines against different variants of the virus.Understanding of the factors that influence vaccine uptake and distribution.Identification of potential adverse effects of COVID-19 vaccines and their severity.Recommendations for optimizing vaccine deployment strategies to achieve maximum public health impact.

CONCLUSION:

This project provides a comprehensive analysis of COVID-19 vaccine data, offering insights into vaccine efficacy, distribution, and adverse effects. The findings are intended to assist policymakers and health organizations in optimizing vaccine deployment strategies to combat the pandemic effectively and ensure public safety.

**COVID VACCINE ANALYSIS REPORT**

**2.1 Short Explanation of the Question**

This analysis is aimed at comprehensively examining various facets of COVID-19 vaccinations. It encompasses factors such as vaccination rates, efficacy, distribution, and their impact on infection rates.

**2.2 Dataset and its details - source**

The dataset used in this analysis was sourced from Kaggle, a prominent platform for data science competitions and datasets. The specific dataset related to COVID-19 vaccines can be found at www.kaggle.com/data.

**2.3 Columns to be included**

1. Country
2. Date
3. Total vaccinations
4. Vaccinations per hundred
5. Total vaccinations percentage

**2.4 Details of libraries to be used and way to download**

The following libraries were employed for this analysis:

- Pandas for data manipulation

- Matplotlib for visualization

- Scikit-learn for machine learning

These libraries can be downloaded and installed using the `pip` package manager in the Python environment.

**2.5 Rest of the explanation**

Depending on the specific objectives of the analysis, various tasks will be undertaken. This will include data cleaning, exploratory data analysis (EDA), visualization, and, in certain cases, the application of machine learning techniques for predictions or classification.

**DATA ANALYTICS WITH COGNOS**

**Development Part-1**

**Identify reliable sources:**

As we look for trusted sources that provide COVID-19 vaccine data, It's available in kaggle.

**Download the data:**

Download the given datasetscontaining COVID-19 vaccine data. The dataset include information on vaccine distribution.

**Understand the data:**

Taking some time to understand the structure and format of the downloaded datasets and as we look for data dictionaries or documentation that explain the meaning of each variable and how the data is organized.

**Clean the data:**

We preprocessed the data to ensure it is in a suitable format for analysis. This involve tasks like removing duplicate entries, handling missing values, standardizing variable names, and converting data types if necessary.

**Explore the data:**

Once the data is cleaned, we explored it to gain insights and identify any patterns or trends. This can be done by performing basic descriptive statistics, visualizations, or more advanced analysis techniques depending on the specific research questions or goals of analysis.

Python packages:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.graph\_objs as go

import plotly.figure\_factory as ff

from plotly import tools

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

import plotly.express as px

init\_notebook\_mode(connected=True)

import warnings

warnings.filterwarnings("ignore")

data\_df = pd.read\_csv(“https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress ")

country\_vaccine = data\_df.groupby(["country", "iso\_code", "vaccines"])

['total\_vaccinations', 'total\_vaccinations\_per\_hundred','daily\_vaccinations', 'daily\_vaccinations\_per\_million', 'people\_vaccinated','people\_vaccinated\_per\_hundred', 'people\_fully\_vaccinated', 'people\_fully\_vaccinated\_per\_hundred' ].max().reset\_index()

country\_vaccine.columns = ["Country", "iso\_code", "Vaccines", "Total vaccinations", "Percent", "Daily vaccinations","Daily vaccinations per million", "People vaccinated", "People vaccinated per hundred", 'People fully vaccinated', 'People fully vaccinated percent

Vaccines are used in each country

vaccines = country\_vaccine.Vaccines.unique()

for v in vaccines:

countries = country\_vaccine.loc[country\_vaccine.Vaccines==v, 'Country'].values

print(f"Vaccines: {v}: \nCountries: {list(countries)}\n")

fig = px.choropleth(locations=country\_vaccine['Country'],

locationmode="country names",

color=country\_vaccine['Vaccines'],

title="Countries using each vaccine (different colors for each vaccine)",

height = 800

)

fig.update\_layout({'legend\_orientation':'v'})

fig.update\_layout({'legend\_title':'Vaccine scheme'})

fig.show()

Vaccination used the most

vaccine = data\_df.groupby(["vaccines"])['total\_vaccinations','total\_vaccinations\_per\_hundred',

'daily\_vaccinations','daily\_vaccinations\_per\_million'].max().reset\_index()

vaccine.columns = ["Vaccines", "Total vaccinations", "Percent", "Daily vaccinations",

"Daily vaccinations per million"]

def draw\_trace\_bar\_vaccine(data, feature, title, xlab, ylab,color='Blue'):

data = data.sort\_values(feature, ascending=False)

trace = go.Bar(

x = data['Vaccines'],

y = data[feature],

marker=dict(color=color),

text=data['Vaccines']

)

data = [trace]

layout = dict(title = title,

xaxis = dict(title = xlab, showticklabels=True, tickangle=45,

zeroline=True, zerolinewidth=1, zerolinecolor='grey',

showline=True, linewidth=2, linecolor='black', mirror=True,

tickfont=dict(

size=10,

color='black'),),

yaxis = dict(title = ylab, gridcolor='lightgrey', zeroline=True, zerolinewidth=1, zerolinecolor='grey',

showline=True, linewidth=2, linecolor='black', mirror=True),

plot\_bgcolor = 'rgba(0, 0, 0, 0)', paper\_bgcolor = 'rgba(0, 0, 0, 0)',

hovermode = 'closest',

height = 800

)

fig = dict(data = data, layout = layout)

iplot(fig, filename='draw\_trace')

Vaccinated people

def draw\_trace\_bar(data, feature, title, xlab, ylab,color='Blue'):

data = data.sort\_values(feature, ascending=False)

trace = go.Bar(

x = data['Country'],

y = data[feature],

marker=dict(color=color),

text=data['Country']

)

data = [trace]

layout = dict(title = title,

xaxis = dict(title = xlab, showticklabels=True, tickangle=45,

zeroline=True, zerolinewidth=1, zerolinecolor='grey',

showline=True, linewidth=2, linecolor='black', mirror=True,

tickfont=dict(

size=10,

color='black'),),

yaxis = dict(title = ylab, gridcolor='lightgrey', zeroline=True, zerolinewidth=1, zerolinecolor='grey',

showline=True, linewidth=2, linecolor='black', mirror=True),

plot\_bgcolor = 'rgba(0, 0, 0, 0)', paper\_bgcolor = 'rgba(0, 0, 0, 0)',

hovermode = 'closest'

)

fig = dict(data = data, layout = layout)

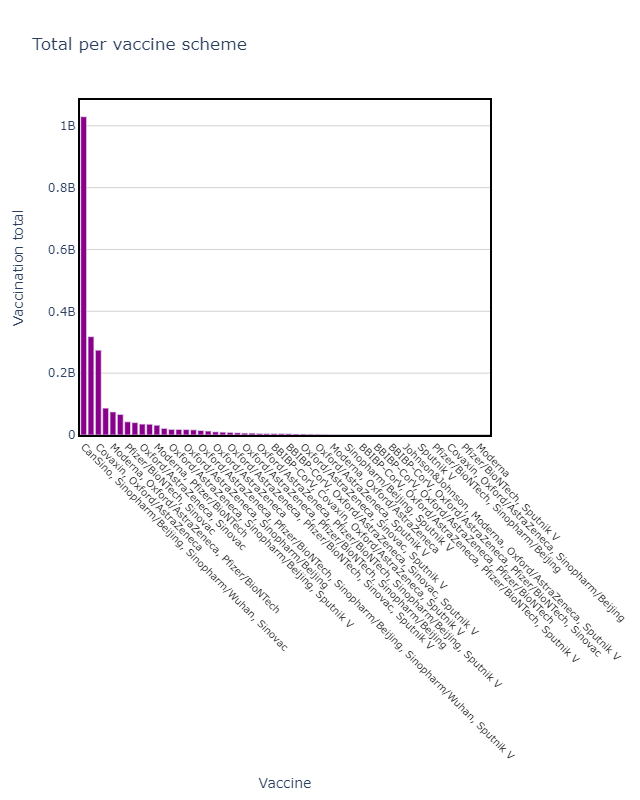
iplot(fig, filename='draw\_trace')

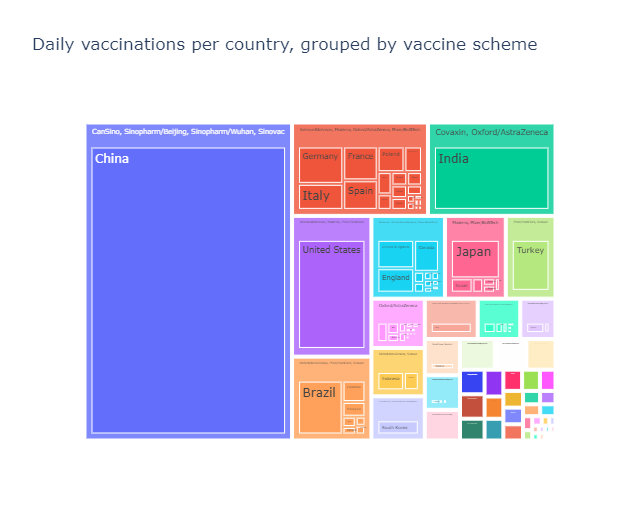
draw\_trace\_bar\_vaccine(vaccine, 'Total vaccinations', 'Total per vaccine scheme', 'Vaccine', 'Vaccination total', "darkmagenta" )

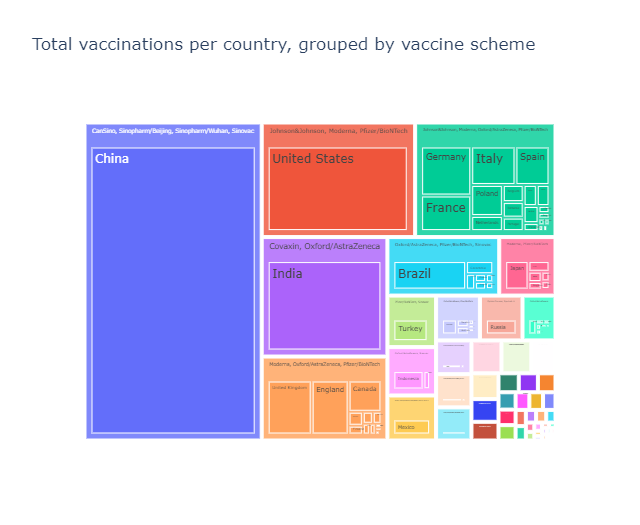
list of countries where each vaccine is used:

****

vaccination scheme used overall:

****

****

****

**countries statistics, irrespective to the vaccine scheme:**

Total number of vaccinations;

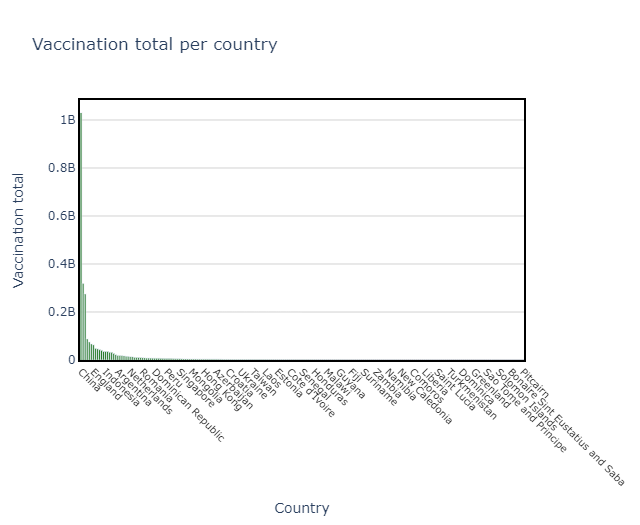
Percent of vaccinations from entire population;

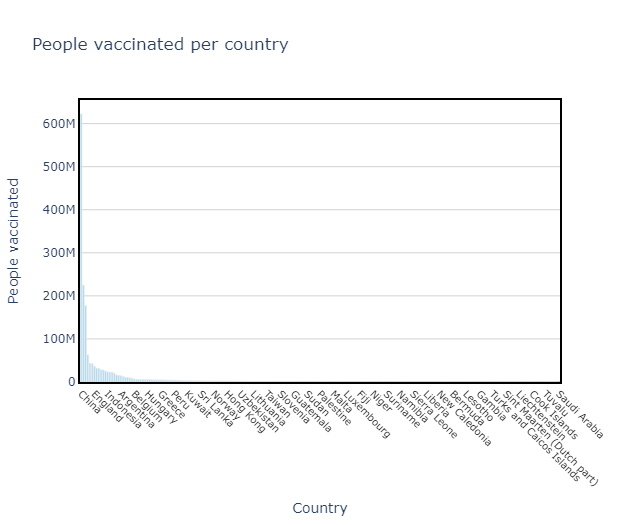
Daily number of vaccinations;

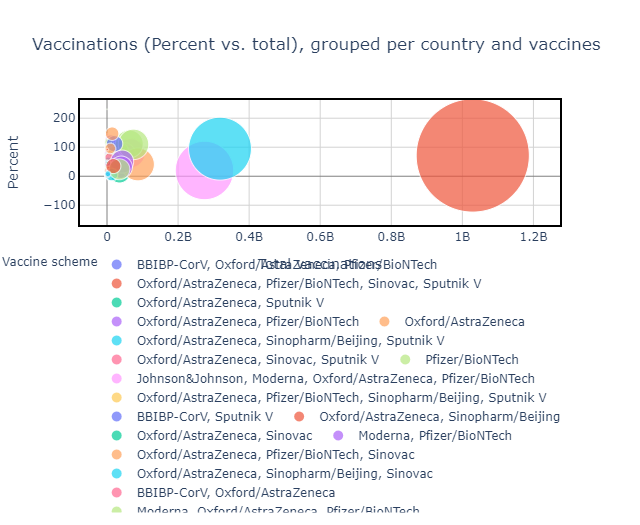
Daily number of vaccination per million population;

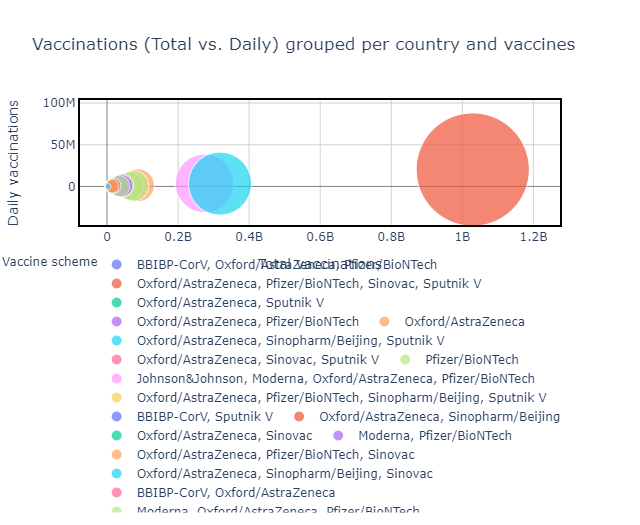
People vaccinated;

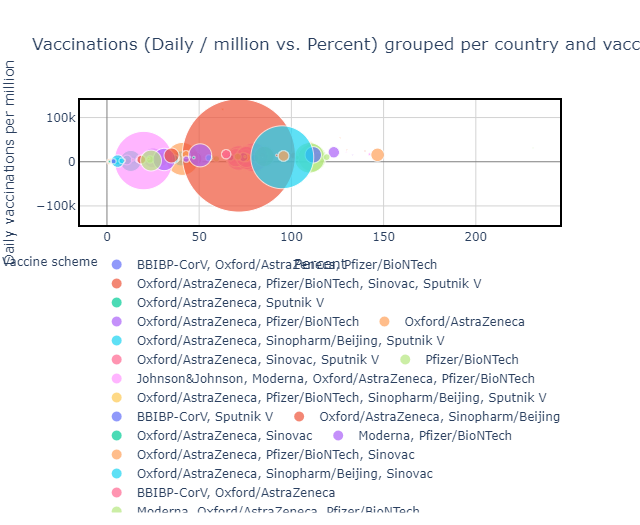
Percent of vaccinated people from entire population.

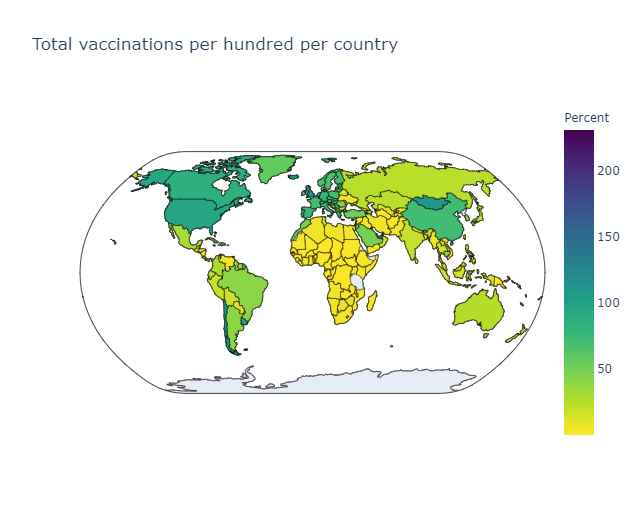


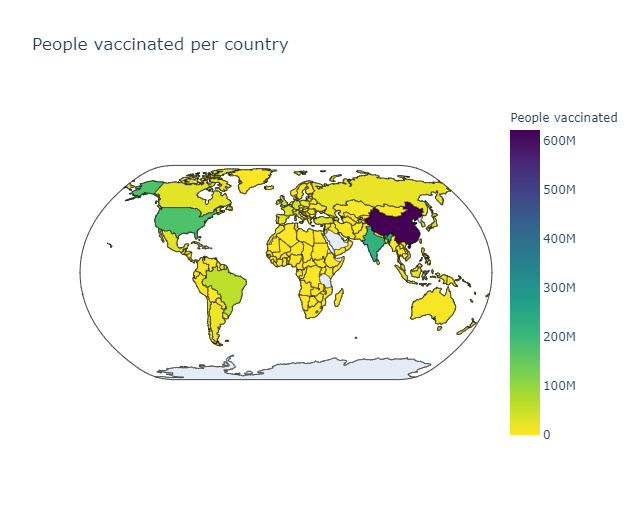




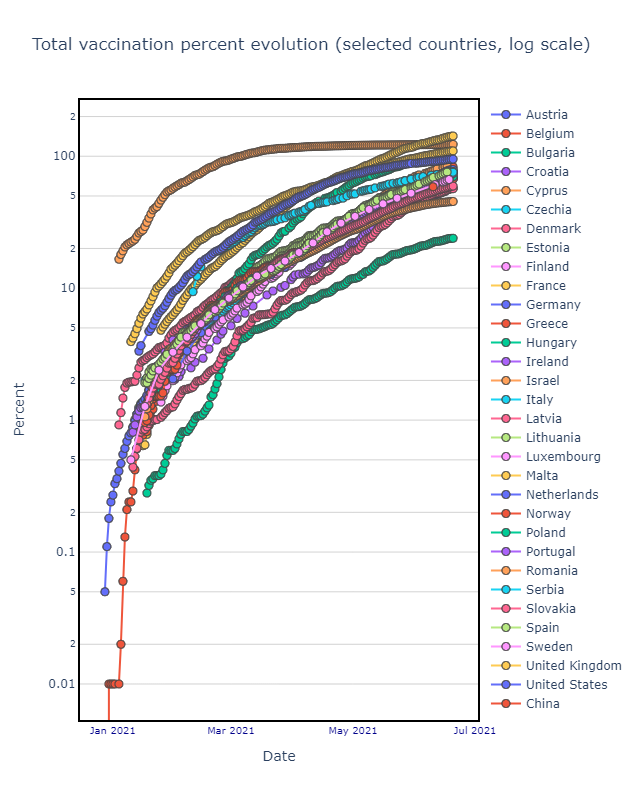


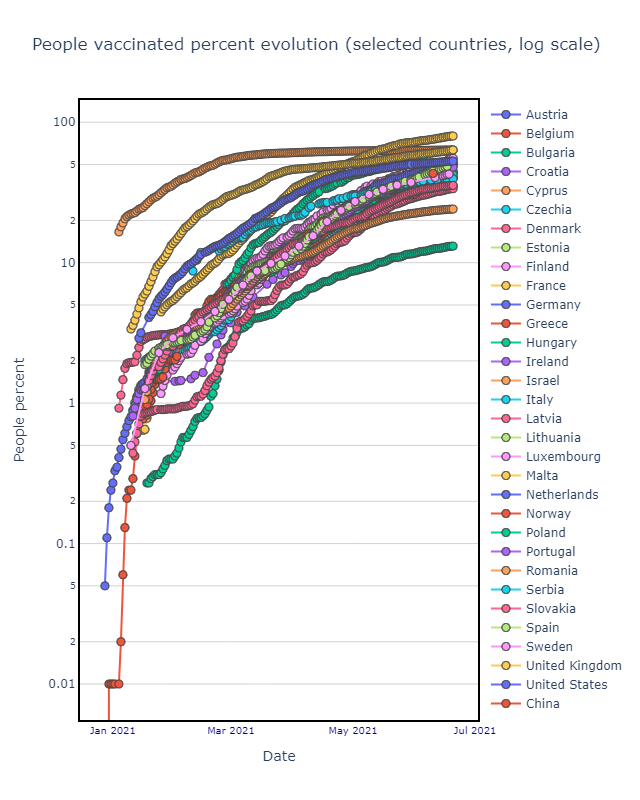


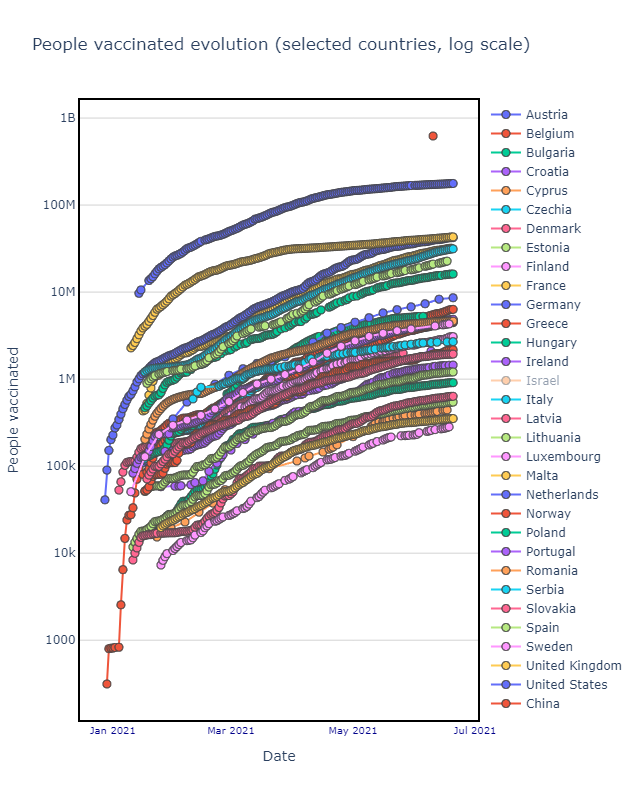
****

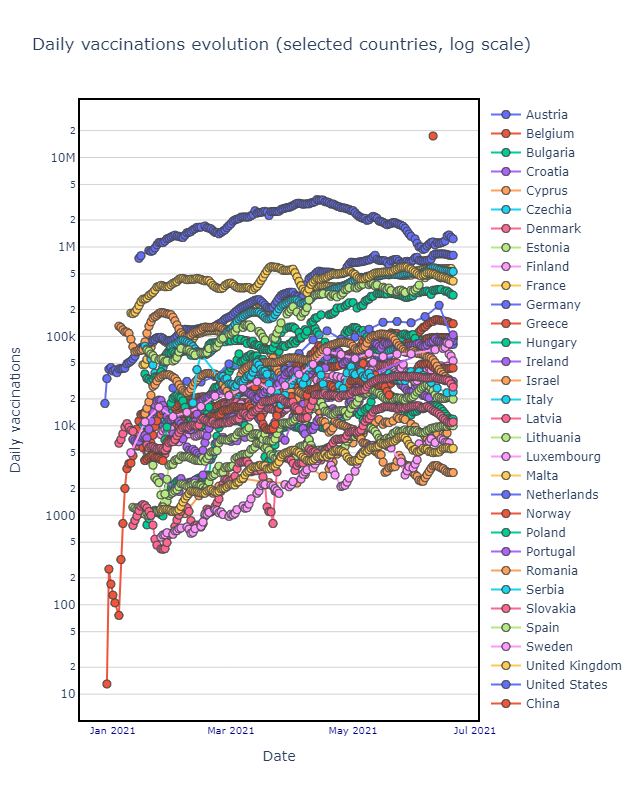
****

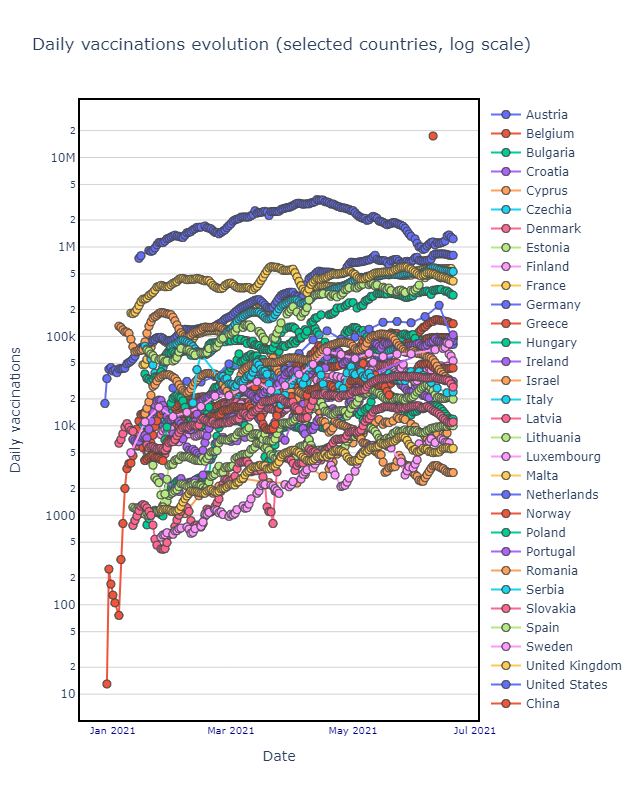
**Vaccine progressed:**

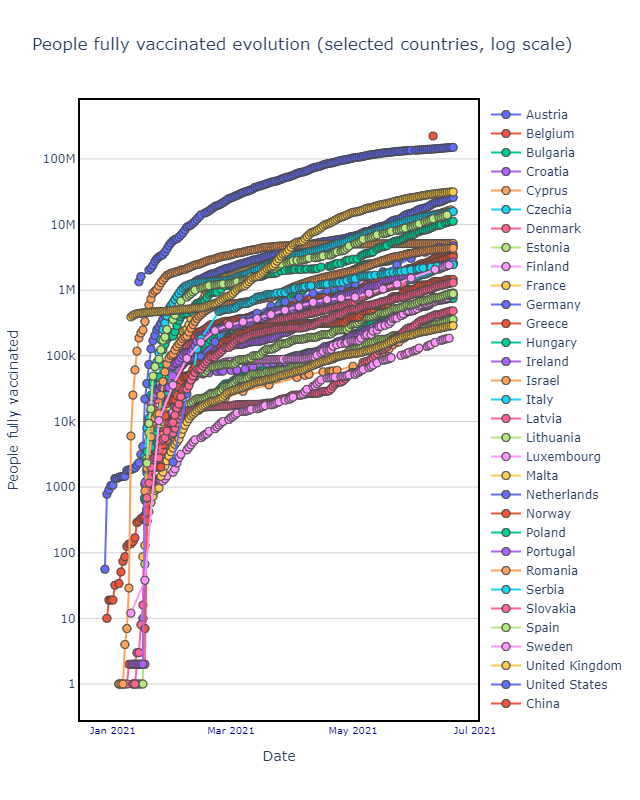
****

****

****

****

****

****

Exploratory Data Analysis (EDA) is a critical step in data analysis process, with the analysis of COVID-19 vaccine distribution data.

* EDA helps us get a better understanding of our dataset. For COVID vaccine analysis, this might involve examining variables such as vaccination rates, demographic information. EDA allows us to identify what data we have, its quality, and how it's structured.
* During EDA, we can identify and address data quality issues, missing values, outliers, and inconsistencies. Ensuring data cleanliness is crucial for reliable vaccine analysis.
* EDA techniques like data visualization can help us spot patterns and trends in the data. We can create various plots and charts to visualize vaccination trends, vaccine coverage over time, and variations among different regions or population groups.
* EDA allows us to explore relationships between variables. For COVID vaccine analysis, we can examine correlations between vaccination rates and COVID case numbers, hospitalizations, and mortality rates.
* EDA can inspire hypotheses for more in-depth analysis. For example, we might notice that vaccination rates are lower in certain areas and hypothesize that these areas could be potential hotspots for COVID outbreaks.
* EDA can help identify subpopulations that may require special attention. We can explore vaccination rates and outcomes among different age groups, ethnicities, or comorbidity categories.
* EDA can help you identify unusual or unexpected patterns in the data. For COVID vaccine analysis, this might involve detecting adverse events or side effects that occur at a higher rate than expected.
* If we plan to build predictive models for COVID-related outcomes, EDA can guide you in selecting relevant features from the dataset. This can help improve model performance and interpretability.
* EDA involves creating various visualizations, such as histograms, scatter plots, and time series plots, to provide clear insights into the data. Visualization can make complex data more understandable and accessible to a broader audience.
* EDA findings can be used to create informative reports or dashboards to communicate insights to stakeholders, public health officials, and the general public.

**Statistical analysis** is a powerful tool for conducting COVID vaccine analysis. It helps researchers and public health professionals make data-driven decisions, assess the effectiveness of vaccines, and gain insights into their impact on COVID-19 outcomes.

Certainly, here's how statistical analysis can be applied specifically to the distribution of COVID-19 vaccines:

* Statistical analysis can assess the efficiency of vaccine distribution by examining key metrics such as the time it takes to administer vaccines, the number of doses distributed versus administered, and the distribution coverage across different regions. This analysis can help identify bottlenecks and areas that need improvement.
* Statistical methods can be used to analyze the performance of the vaccine supply chain, including logistics and inventory management. This helps ensure a smooth and timely flow of vaccines to distribution centers, clinics, and vaccination sites.
* By employing statistical techniques like demand forecasting, inventory control models, and safety stock analysis, vaccine distribution centers can optimize their inventory management. This ensures that there are enough vaccine doses available without excessive overstock.
* Statistical analysis can help predict and understand the demand for vaccines in different geographic areas and demographic groups. This information is crucial for allocating vaccine doses effectively.
* Geographic information system (GIS) and spatial statistical methods can be used to analyze the geographic distribution of vaccines. This can help identify underserved areas and guide allocation decisions.
* Statistical analysis can be used to monitor the temperature and condition of vaccines throughout the supply chain. Deviations from recommended storage conditions can be detected, allowing for timely corrective actions.
* Statistical methods can assess the performance of vaccination sites, including wait times, appointment availability, and vaccination rates. This information is vital for optimizing site operations.
* Statistical analysis can evaluate vaccine distribution to ensure equity and fairness. This includes analyzing whether vulnerable or underserved populations have equitable access to vaccines.
* Statistical models can assist in resource allocation decisions, such as determining the number of vaccination sites needed in a given region or the allocation of vaccines to different states or communities.
* Statistical techniques can be applied to assess the quality of the distributed vaccines, ensuring that they meet safety and efficacy standards.
* Statistical surveillance systems are used to monitor adverse events following vaccination. Analyzing this data can help identify potential safety concerns and guide regulatory decisions.
* Statistical modeling can be used to assess the public health impact of vaccine distribution. This involves estimating the reduction in COVID-19 cases, hospitalizations, and deaths attributable to the vaccination campaign.
* Statistical simulations can help decision-makers explore different distribution strategies and their potential outcomes. For example, they can simulate the impact of prioritizing specific populations for vaccination or altering distribution routes.

