**COVID-19 VACCINE ANALYSIS**

**PHASE 3: DEVELOPMENT**

**PART 1**

**TOPIC :Building your Project by Loading and Preprocessing The Dataset.**



**Dataset Explanation:**

A dataset for COVID-19 vaccine analysis is a structured collection of information related to COVID-19 vaccination efforts. These datasets are valuable for researchers, epidemiologists, public health officials, and data analysts who want to study and understand various aspects of COVID-19 vaccine distribution, effectiveness, safety, and impact. Here's an explanation of the key components typically found in such datasets:

**Demographic Information:**

Age: The age of individuals receiving the vaccine.

Gender: Gender information of vaccine recipients.

Ethnicity/Race: Data on the racial and ethnic backgrounds of individuals vaccinated.

Vaccination Details:

Vaccine Type:

Information about the specific COVID-19 vaccine administered (e.g., Pfizer, Moderna, Johnson & Johnson).

Dose Number:

Whether the individual received the first, second, or booster dose.

Date of Vaccination:

The date when the vaccine was administered.

Vaccination Site:

Location where the vaccine was given (e.g., healthcare facility, pharmacy, mobile clinic).

Geographical Information:

Location:

Data on the geographic location of vaccination sites, including country, state, city, or zip code.

Population Density:

Information about the population density of the area where vaccination occurred.

Vaccine Efficacy:

Vaccine Effectiveness:

Data on the effectiveness of the vaccine in preventing COVID-19 infection and its variants.

Immunization Coverage:

The percentage of the population that has received the vaccine.

Adverse Reactions:

Adverse Events: Information on adverse reactions or side effects reported after vaccination, including the type and severity.

Hospitalizations:

Data on hospitalizations or emergency room visits related to vaccination.

Distribution and Supply Chain:

Vaccine Distribution: Details on the distribution process, including the number of doses distributed and received by different regions.

Vaccine Inventory:

Information on the available vaccine doses and their expiration dates.

Supply Chain Data:

Data on the logistics and supply chain management for vaccine distribution.

Time Series Data:

Daily/Monthly Counts: The number of vaccinations administered over time, which allows for tracking trends and fluctuations.

Cumulative Counts:

Total vaccinations administered up to a specific date.

Variants and Strains:

Information on the prevalence of COVID-19 variants and strains within the vaccinated population.

Healthcare Provider Data:

Data on healthcare providers administering the vaccine, including their locations and performance.

Research Variables:

Any other variables that researchers believe are relevant for analyzing the impact of COVID-19 vaccination, such as comorbidities, pre-existing conditions, or travel history.

These datasets can be sourced from various public health authorities, government agencies, research institutions, and healthcare organizations. Analyzing these datasets can help identify trends, assess the impact of vaccination campaigns, evaluate vaccine safety, and guide public health policy decisions. It's essential to ensure the privacy and security of individual data in these datasets and adhere to ethical and legal data handling practices during analysis.

**GIVEN DATASET:**

The given dataset for covid-19 vaccine analysis

<https://www.kaggle.com/datasets/imdevskp/corona-virus-report/code>

**Import Libraries:**

import numpy as np

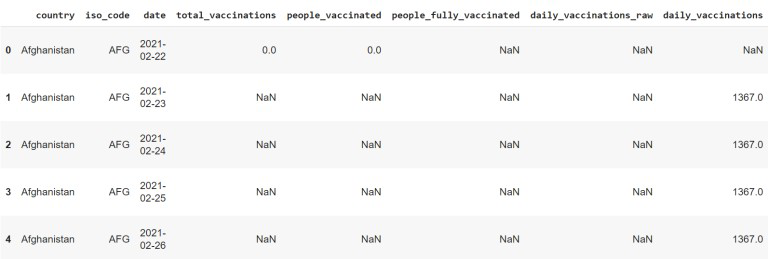
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data = pd.read\_csv("country\_vaccinations.csv")

data.head()



df= pd.read\_csv('/kaggle/input/complete-tweet-sentiment-extraction-data/tweet\_dataset.csv')

df.head()

|  | textID | sentiment | author | text | old\_text | aux\_id | new\_sentiment | selected\_text |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1956967341 | empty | xoshayzers | i know i was listenin to bad habit earlier a... | @tiffanylue i know i was listenin to bad habi... | p1000000000 | NaN | NaN |
| 1 | 1956967666 | sadness | wannamama | Layin n bed with a headache ughhhh...waitin o... | Layin n bed with a headache ughhhh...waitin o... | c811396dc2 | negative | headache |
| 2 | 1956967696 | sadness | coolfunky | Funeral ceremony...gloomy friday... | Funeral ceremony...gloomy friday... | 9063631ab1 | negative | gloomy |
| 3 | 1956967789 | enthusiasm | czareaquino | wants to hang out with friends SOON! | wants to hang out with friends SOON! | 2a815f151d | positive | wants to hang out with friends SOON! |
| 4 | 1956968416 | neutral | xkilljoyx | We want to trade with someone who has Houston... | @dannycastillo We want to trade with someone w... | 82565a56d3 |  |  |

**LOAD DATASET:**

vaccinations\_df = pd.read\_csv('../input/covid-world-vaccination-progress/country\_vaccinations.csv')

vaccinations\_df

| country | iso\_code | date | total\_vaccinations | people\_vaccinated | people\_fully\_vaccinated | daily\_vaccinations\_raw | daily\_vaccinations | total\_vaccinations\_per\_hundred | people\_vaccinated\_per\_hundred | people\_fully\_vaccinated\_per\_hundred | daily\_vaccinations\_per\_million | vaccines | source\_name | source\_website |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Afghanistan | AFG | 2021-02-22 | 0.0 | 0.0 | NaN | NaN | NaN | 0.00 | 0.00 | NaN | NaN | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... | World Health Organization | https://covid19.who.int/ |
| 1 | Afghanistan | AFG | 2021-02-23 | NaN | NaN | NaN | NaN | 1367.0 | NaN | NaN | NaN | 34.0 | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... | World Health Organization | https://covid19.who.int/ |
| 2 | Afghanistan | AFG | 2021-02-24 | NaN | NaN | NaN | NaN | 1367.0 | NaN | NaN | NaN | 34.0 | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... | World Health Organization | https://covid19.who.int/ |
| 3 | Afghanistan | AFG | 2021-02-25 | NaN | NaN | NaN | NaN | 1367.0 | NaN | NaN | NaN | 34.0 | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... | World Health Organization | https://covid19.who.int/ |
| 4 | Afghanistan | AFG | 2021-02-26 | NaN | NaN | NaN | NaN | 1367.0 | NaN | NaN | NaN | 34.0 | Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi... | World Health Organization | https://covid19.who.int/ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 86507 | Zimbabwe | ZWE | 2022-03-25 | 8691642.0 | 4814582.0 | 3473523.0 | 139213.0 | 69579.0 | 57.59 | 31.90 | 23.02 | 4610.0 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac... | Ministry of Health | https://www.arcgis.com/home/webmap/viewer.html... |
| 86508 | Zimbabwe | ZWE | 2022-03-26 | 8791728.0 | 4886242.0 | 3487962.0 | 100086.0 | 83429.0 | 58.25 | 32.38 | 23.11 | 5528.0 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac... | Ministry of Health | https://www.arcgis.com/home/webmap/viewer.html... |
| 86509 | Zimbabwe | ZWE | 2022-03-27 | 8845039.0 | 4918147.0 | 3493763.0 | 53311.0 | 90629.0 | 58.61 | 32.59 | 23.15 | 6005.0 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac... | Ministry of Health | https://www.arcgis.com/home/webmap/viewer.html... |
| 86510 | Zimbabwe | ZWE | 2022-03-28 | 8934360.0 | 4975433.0 | 3501493.0 | 89321.0 | 100614.0 | 59.20 | 32.97 | 23.20 | 6667.0 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac... | Ministry of Health | https://www.arcgis.com/home/webmap/viewer.html... |
| 86511 | Zimbabwe | ZWE | 2022-03-29 | 9039729.0 | 5053114.0 | 3510256.0 | 105369.0 | 103751.0 | 59.90 | 33.48 | 23.26 | 6874.0 | Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac... | Ministry of Health | https://www.arcgis.com/home/webmap/viewer.html... |

**DATA PREPROCESSING:**

Data preprocessing for COVID-19 vaccine analysis refers to the set of procedures and techniques used to clean, transform, and organize raw data related to COVID-19 vaccines in order to make it suitable for analysis. This process is critical for ensuring the quality, integrity, and readiness of the data before any meaningful analysis or insights can be derived. Data preprocessing may involve tasks such as data cleaning, handling missing values, feature engineering, standardization, and more, specifically tailored to data related to COVID-19 vaccines. The ultimate goal is to prepare the data in a way that facilitates meaningful and accurate analysis, allowing researchers, healthcare professionals, and policymakers to make informed decisions about vaccine efficacy, safety, distribution, and other critical factors in the fight against COVID-19.

vaccinations\_df.describe()

|  | total\_vaccinations | people\_vaccinated | people\_fully\_vaccinated | daily\_vaccinations\_raw | daily\_vaccinations | total\_vaccinations\_per\_hundred | people\_vaccinated\_per\_hundred | people\_fully\_vaccinated\_per\_hundred | daily\_vaccinations\_per\_million |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| count | 4.360700e+04 | 4.129400e+04 | 3.880200e+04 | 3.536200e+04 | 8.621300e+04 | 43607.000000 | 41294.000000 | 38802.000000 | 86213.000000 |
| mean | 4.592964e+07 | 1.770508e+07 | 1.413830e+07 | 2.705996e+05 | 1.313055e+05 | 80.188543 | 40.927317 | 35.523243 | 3257.049157 |
| std | 2.246004e+08 | 7.078731e+07 | 5.713920e+07 | 1.212427e+06 | 7.682388e+05 | 67.913577 | 29.290759 | 28.376252 | 3934.312440 |
| min | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 5.264100e+05 | 3.494642e+05 | 2.439622e+05 | 4.668000e+03 | 9.000000e+02 | 16.050000 | 11.370000 | 7.020000 | 636.000000 |
| 50% | 3.590096e+06 | 2.187310e+06 | 1.722140e+06 | 2.530900e+04 | 7.343000e+03 | 67.520000 | 41.435000 | 31.750000 | 2050.000000 |
| 75% | 1.701230e+07 | 9.152520e+06 | 7.559870e+06 | 1.234925e+05 | 4.409800e+04 | 132.735000 | 67.910000 | 62.080000 | 4682.000000 |
| max | 3.263129e+09 | 1.275541e+09 | 1.240777e+09 | 2.474100e+07 | 2.242429e+ |  |  |  |  |

# Checking null values in dataset

vaccinations\_df.isnull().sum()

country 0

iso\_code 0

date 0

total\_vaccinations 42905

people\_vaccinated 45218

people\_fully\_vaccinated 47710

daily\_vaccinations\_raw 51150

daily\_vaccinations 299

total\_vaccinations\_per\_hundred 42905

people\_vaccinated\_per\_hundred 45218

people\_fully\_vaccinated\_per\_hundred 47710

daily\_vaccinations\_per\_million 299

vaccines 0

source\_name 0

source\_website 0

dtype: int64

vaccinations\_df.columns

Output:

Index(['country', 'iso\_code', 'date', 'total\_vaccinations',

'people\_vaccinated', 'people\_fully\_vaccinated',

'daily\_vaccinations\_raw', 'daily\_vaccinations',

'total\_vaccinations\_per\_hundred', 'people\_vaccinated\_per\_hundred',

'people\_fully\_vaccinated\_per\_hundred', 'daily\_vaccinations\_per\_million',

'vaccines', 'source\_name', 'source\_website'],dtype='object')

vaccinations\_df.shape

Output:

(86512, 15)

vaccinations\_df.mean()

total\_vaccinations 2.315117e+07

people\_vaccinated 8.451007e+06

people\_fully\_vaccinated 6.341251e+06

daily\_vaccinations\_raw 1.106083e+05

daily\_vaccinations 1.308517e+05

total\_vaccinations\_per\_hundred 4.041962e+01

people\_vaccinated\_per\_hundred 1.953547e+01

people\_fully\_vaccinated\_per\_hundred 1.593274e+01

daily\_vaccinations\_per\_million 3.245792e+03

year 2.021199e+03

month 6.165711e+00

day 1.571936e+01

dtype: float64

vaccinations\_df.min()

country Afghanistan

iso\_code ABW

date 2020-12-02 00:00:00

total\_vaccinations 0.0

people\_vaccinated 0.0

people\_fully\_vaccinated 0.0

daily\_vaccinations\_raw 0.0

daily\_vaccinations 0.0

total\_vaccinations\_per\_hundred 0.0

people\_vaccinated\_per\_hundred 0.0

people\_fully\_vaccinated\_per\_hundred 0.0

daily\_vaccinations\_per\_million 0.0

vaccines Abdala, Johnson&Johnson, Oxford/AstraZeneca, P...

source\_name Africa Centres for Disease Control and Prevention

source\_website http://103.247.238.92/webportal/pages/covid19-...

year 2020

1dtype: object

**DATA ANALYSIS:**

Data analysis is the process of inspecting, cleaning, transforming, and interpreting data to discover meaningful insights, patterns, and trends. It involves using various techniques and tools to make data more understandable and valuable for decision-making, problem-solving, and knowledge discovery.

**df** = **pd**.**read\_csv("../input/covid-world-vaccination-progress/country\_vaccinations.csv")**

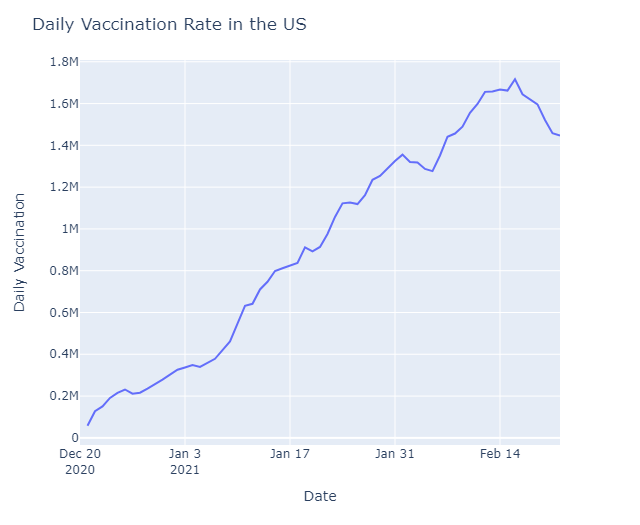
**df**.**head()**

|  | country | iso\_code | date | total\_vaccinations | people\_vaccinated | people\_fully\_vaccinated | daily\_vaccinations\_raw | daily\_vaccinations | total\_vaccinations\_per\_hundred | people\_vaccinated\_per\_hundred | people\_fully\_vaccinated\_per\_hundred | daily\_vaccinations\_per\_million | vaccines | source\_name | source\_website |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Albania | ALB | 2021-01-10 | 0.0 | 0.0 | NaN | NaN | NaN | 0.00 | 0.00 | NaN | NaN | Pfizer/BioNTech | Ministry of Health | https://shendetesia.gov.al/covid19-ministria-e... |
| 1 | Albania | ALB | 2021-01-11 | NaN | NaN | NaN | NaN | 64.0 | NaN | NaN | NaN | 22.0 | Pfizer/BioNTech | Ministry of Health | https://shendetesia.gov.al/covid19-ministria-e... |
| 2 | Albania | ALB | 2021-01-12 | 128.0 | 128.0 | NaN | NaN | 64.0 | 0.00 | 0.00 | NaN | 22.0 | Pfizer/BioNTech | Ministry of Health | https://shendetesia.gov.al/covid19-ministria-e... |
| 3 | Albania | ALB | 2021-01-13 | 188.0 | 188.0 | NaN | 60.0 | 63.0 | 0.01 | 0.01 | NaN | 22.0 | Pfizer/BioNTech | Ministry of Health | https://shendetesia.gov.al/covid19-ministria-e... |
| 4 | Albania | ALB | 2021-01-14 | 266.0 | 266.0 | NaN | 78.0 |  |  |  |  |  |  |  |  |

**px**.**line(df[df['country']**==**'United States'],x**=**'date',y**=**'daily\_vaccinations')**.**update\_layout(title** = **'Daily Vaccination Rate in the US',xaxis\_title**=**'Date',yaxis\_title**=**'Daily Vaccination')**

Dec 202020Jan 32021Jan 17Jan 31Feb1400.2M0.4M0.6M0.8M1M1.2M1.4M1.6M1.8M

*#Daily Vaccination Rate in the USDateDaily Vaccination*



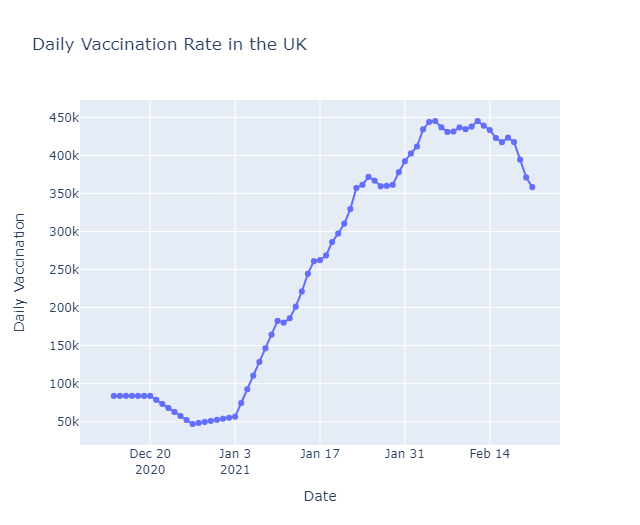
*#Daily Vaccination rate in the UK*

**fig**=**go**.**Figure()**

**df1**=**df[df['country']**==**'United Kingdom']**

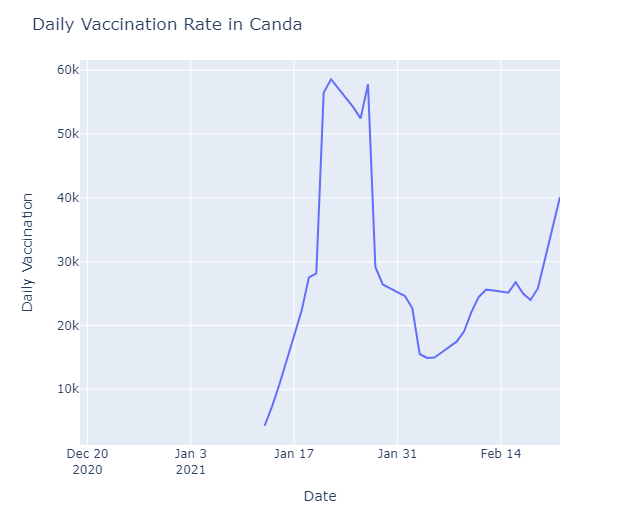
**fig**.**add\_trace(go**.**Scatter(x**=**df1**.**date,y**=**df1**.**daily\_vaccinations,mode**=**'lines+markers',name**=**'UK'))**

**fig**.**update\_layout(title** = **'Daily Vaccination Rate in the UK',xaxis\_title**=**'Date',yaxis\_title**=**'Daily Vaccination' )**



*#Daily Vaccination Rate in Canada*

**px**.**line(df[df['country']**==**'Canada'],x**=**'date',y**=**'daily\_vaccinations')**.**update\_layout(title** = **'Daily Vaccination Rate in Canda',xaxis\_title**=**'Date',yaxis\_title**=**'Daily Vaccination' )**



*#Daily Vaccination Rate Comparision*

**fig**=**go**.**Figure()**

**df1**=**df[df['country']**==**'United States']**

**df2**=**df[df['country']**==**'United Kingdom']**

**df3**=**df[df['country']**==**'Canada']**

**fig**.**add\_trace(go**.**Scatter(x**=**df1**.**date,y**=**df1**.**daily\_vaccinations,mode**=**'lines+markers',name**=**'US'))**

**fig**.**add\_trace(go**.**Scatter(x**=**df2**.**date,y**=**df2**.**daily\_vaccinations,mode**=**'lines+markers',line**=**dict(color**=**'firebrick',width**=**2),name**=**'UK'))**

**fig**.**add\_trace(go**.**Scatter(x**=**df3**.**date,y**=**df3**.**daily\_vaccinations,mode**=**'lines+markers',line**=**dict(dash**=**'dashdot'),name**=**'Canada'))**

**fig**.**update\_layout(title** = **'Daily Vaccination Rate Comparision in the US, UK & Canada',xaxis\_title**=**'Date',yaxis\_title**=**'Daily Vaccination' )**

