COVID VACCINE ANALYSIS

ABSTRACT

In this analysis, we explored a comprehensive dataset on COVID-19 vaccination progress across different countries. The dataset includes information on daily vaccination numbers, total vaccinations administered, vaccine types, and more. Our goal was to gain insights into the global vaccination effort, understand vaccination trends over time, and identify factors influencing vaccination rates.

Key Findings and Insights

Global Vaccination Trends

Vaccine Types

Temporal Trends

Regional Disparities

Correlation Analysis

OBJECTIVE

1. Some math to understand it better, and making visuals to explain it clearly. The hope is that by doing this, we can give a good picture of how the vaccines are doing and help in the fight against Covid-19.
2. The project aims to thoroughly analyze Covid-19 vaccine data with key objectives:

    - Evaluating vaccine efficacy
    - Scrutinizing distribution strategies
    - Investigating adverse effects
    - Providing actionable insights

    By achieving these goals, the project seeks to enhance decision-making for policymakers and health organizations, fostering optimized deployment strategies in the ongoing battle against the Covid-19 pandemic.

DESIGN & THINKING

1. Data Preprocesing
2. Exploratory Data Analysis(EDA)
3. Statistical Analysis
4. Virtualization
5. Insights and Recommendation
6. Data Collection

EXPLORATORY DATA ANALYSIS

- Calculate summary statistics for relevant columns (mean, median, standard deviation, etc.).
- Create various visualizations to explore trends and patterns, such as:
  - Time series plots of vaccination progress over time.
  - Bar charts to compare vaccination rates among countries.
  - Heatmaps to identify correlations between variables.
- Analyze the geographical distribution of vaccination progress using world maps.

STATISTICAL ANALYSIS

- Conduct hypothesis testing to answer specific research questions (e.g., comparing vaccination rates between countries using t-tests).
- Use regression analysis to model the impact of variables (e.g., vaccine type or GDP) on vaccination rates.

VISUALIZATION

- Develop informative and visually appealing charts and graphs.
- Consider creating interactive visualizations for online sharing or presentations.
- Ensure that your visualizations are well-labeled and easy to interpret.

Data Source

Data set link https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress

country	iso_code	date	total_vaccin	people_vacc	people_fully	daily_vaccin	daily_vaccin	total_vaccin	people_vacc	people_fully	daily_vaccin	vaccines	source_nam	source_website
Afghanistan	AFG	22-02-2021	0	0				0	0			Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	23-02-2021					1367				34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	24-02-2021					1367				34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	25-02-2021					1367				34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	26-02-2021					1367				34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	27-02-2021					1367				34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	28-02-2021	8200	8200			1367	0.02	0.02		34	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	01-03-2021					1580				40	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	02-03-2021					1794				45	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	03-03-2021					2008				50	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	04-03-2021					2221				56	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	05-03-2021					2435				61	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	06-03-2021					2649				66	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	07-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	08-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	09-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	10-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	11-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	12-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	13-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	14-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	15-03-2021					2862				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	16-03-2021	54000	54000			2862	0.14	0.14		72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	17-03-2021					2882				72	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	18-03-2021					2902				73	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	19-03-2021					2921				73	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	20-03-2021					2941				74	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	21-03-2021					2961				74	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	22-03-2021					2980				75	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	23-03-2021					3000				75	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	24-03-2021					3000				75	Johnson&Jo	World Healt	https://covid19.who.int/
Afghanistan	AFG	25-03-2021					3000				75	Johnson&Jo	World Healt	https://covid19.who.int/

PROGRAM FOR EDA:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
url = "data.csv"
data = pd.read_csv(url)
print("Basic Info:")
print(data.info())
print("\nSummary Statistics:")
print(data.describe())
print("\nMissing Values:")
print(data.isnull().sum())
print("\nData Types:")
print(data.dtypes)
```

```
categorical_columns = data.select_dtypes(include=['object'])
print("\nUnique Values in Categorical Columns:")
for col in categorical_columns.columns:
    unique_values = data[col].nunique()
 print(f"{col}: {unique_values} unique values")
numeric_data = data.select_dtypes(include=['number'])
for col in numeric_data.columns:
    plt.figure(figsize=(6, 6))
    sns.histplot(data=data, x=col, kde=True, bins=20)
    plt.title(f"Distribution of {col}")
    plt.xlabel(col)
    plt.ylabel("Frequency")
plt.show()
for col in categorical_columns.columns:
    plt.figure(figsize=(6, 6))
    sns.boxplot(data=data[0:2500], x=col, y='total_vaccinations')
```

```
 plt.title(f"Box Plot of Total Vaccinations by {col}")
    plt.xticks(rotation=10)
    plt.xticks(fontsize=6)
plt.show()
plt.figure(figsize=(10, 6))
sns.lineplot(data=data, x=data.index, y='total_vaccinations')
plt.title("Total Vaccinations Over Time")
plt.xlabel("Date")
plt.ylabel("Total Vaccinations")
plt.xticks(rotation=45)
plt.show()
```

OUTPUT:

```
IDLE Shell 3.11.0
File  Edit  Shell  Debug  Options  Window  Help
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: D:\naan\New folder\eda.py ==========
Basic Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   country                            86512 non-null  object
 1   iso_code                           86512 non-null  object
 2   date                               86512 non-null  object
 3   total_vaccinations                 86512 non-null  float64
 4   people_vaccinated                  86512 non-null  float64
 5   people_fully_vaccinated            86512 non-null  float64
 6   daily_vaccinations_raw             86512 non-null  float64
 7   daily_vaccinations                 86512 non-null  float64
 8   total_vaccinations_per_hundred     86512 non-null  float64
 9   people_vaccinated_per_hundred      86512 non-null  float64
 10  people_fully_vaccinated_per_hundred 86512 non-null  float64
 11  daily_vaccinations_per_million     86512 non-null  float64
 12  vaccines                           86512 non-null  object
 13  source_name                        86512 non-null  object
 14  source_website                     86512 non-null  object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None

Summary Statistics:
       total_vaccinations  ...  daily_vaccinations_per_million
count        8.651200e+04  ...                   86512.000000
mean         2.315117e+07  ...                    3245.792248
std          1.611037e+08  ...                    3932.156455
min          0.000000e+00  ...                       0.000000
25%          0.000000e+00  ...                     629.000000
50%          1.008000e+03  ...                    2036.000000
75%          3.697554e+06  ...                    4667.000000
max          3.263129e+09  ...                  117497.000000

[8 rows x 9 columns]

Missing Values:
country                              0
iso_code                             0
date                                 0
total_vaccinations                   0
people_vaccinated                    0
people_fully_vaccinated              0
daily_vaccinations_raw               0
daily_vaccinations                   0
total_vaccinations_per_hundred       0
people_vaccinated_per_hundred        0
people_fully_vaccinated_per_hundred  0
daily_vaccinations_per_million       0
vaccines                             0
source_name                          0
source_website                       0
dtype: int64

Data Types:
country                               object
iso_code                              object
date                                  object
total_vaccinations                   float64
people_vaccinated                    float64
people_fully_vaccinated              float64
daily_vaccinations_raw               float64
daily_vaccinations                   float64
total_vaccinations_per_hundred       float64
people_vaccinated_per_hundred        float64
people_fully_vaccinated_per_hundred  float64
daily_vaccinations_per_million       float64
vaccines                              object
source_name                           object
source_website                        object
dtype: object

Unique Values in Categorical Columns:
country: 223 unique values
iso_code: 223 unique values
date: 483 unique values
vaccines: 84 unique values
source_name: 81 unique values
source_website: 119 unique values
>>>
```

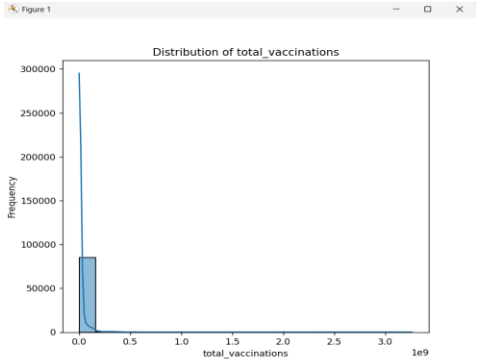


Figure 1
Distribution of total_vaccinations
Frequency
total_vaccinations
1e9

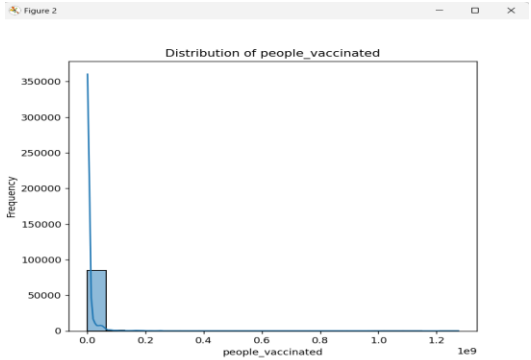


Figure 2
Distribution of people_vaccinated
Frequency
people_vaccinated
1e9

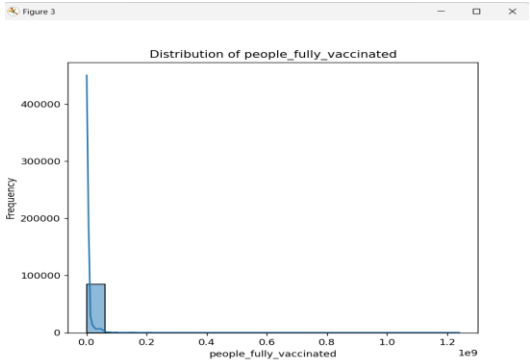


Figure 3
Distribution of people_fully_vaccinated
Frequency
people_fully_vaccinated
1e9

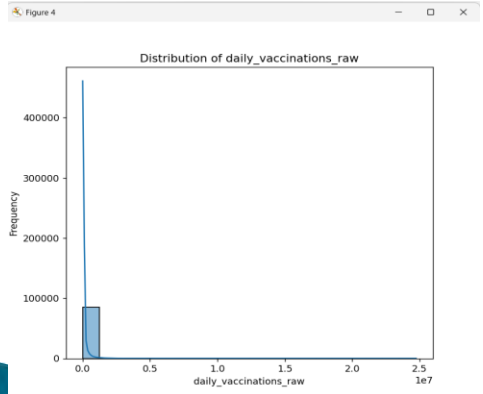


Figure 4
Distribution of daily_vaccinations_raw
Frequency
daily_vaccinations_raw
1e7

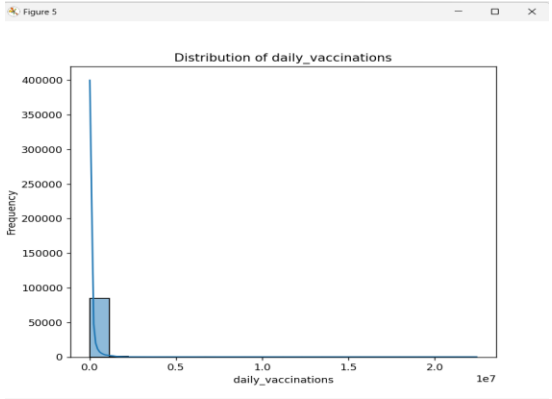


Figure 5
Distribution of daily_vaccinations
Frequency
daily_vaccinations
1e7

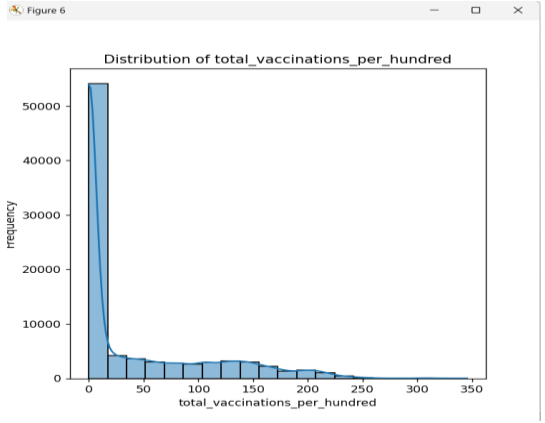


Figure 6
Distribution of total_vaccinations_per_hundred
Frequency
total_vaccinations_per_hundred

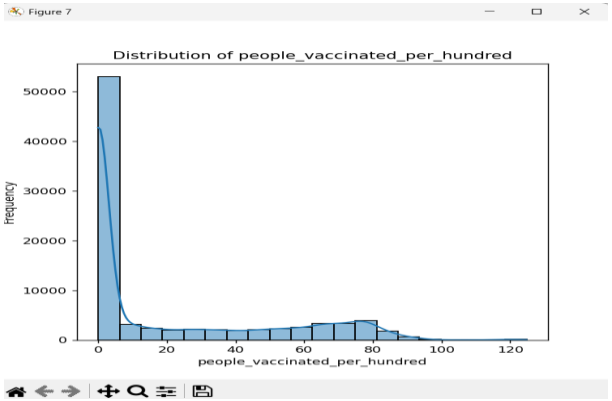


Figure 7
Distribution of people_vaccinated_per_hundred
Frequency
people_vaccinated_per_hundred

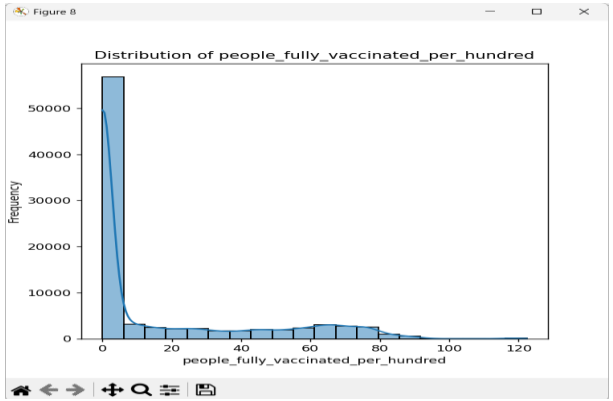


Figure 8
Distribution of people_fully_vaccinated_per_hundred
Frequency
people_fully_vaccinated_per_hundred

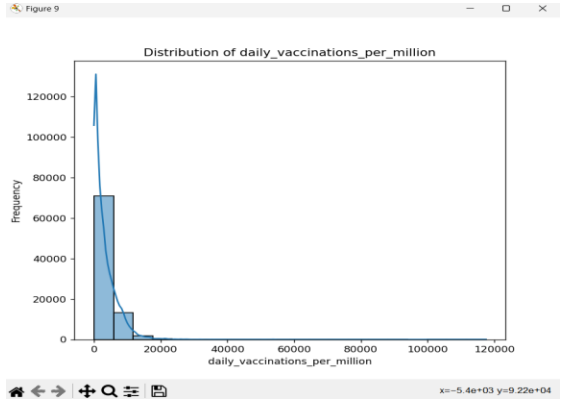


Figure 9
Distribution of daily_vaccinations_per_million
Frequency
daily_vaccinations_per_million
x=-5.4e+03 y=9.22e+04

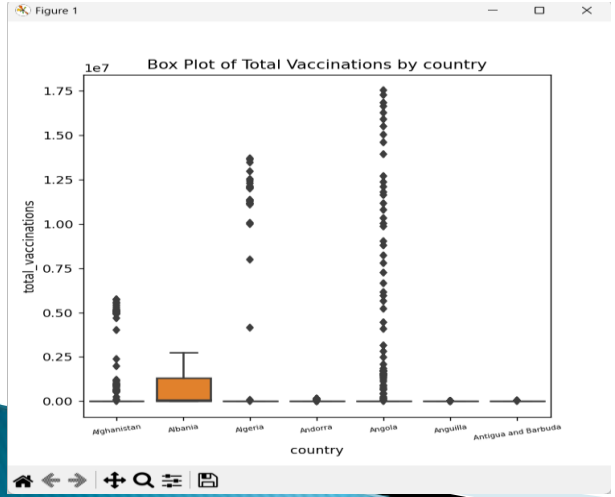


Figure 1
Box Plot of Total Vaccinations by country
1e7
total_vaccinations
country
Afghanistan
Albania
Algeria
Andorra
Angola
Anguilla
Antigua and Barbuda

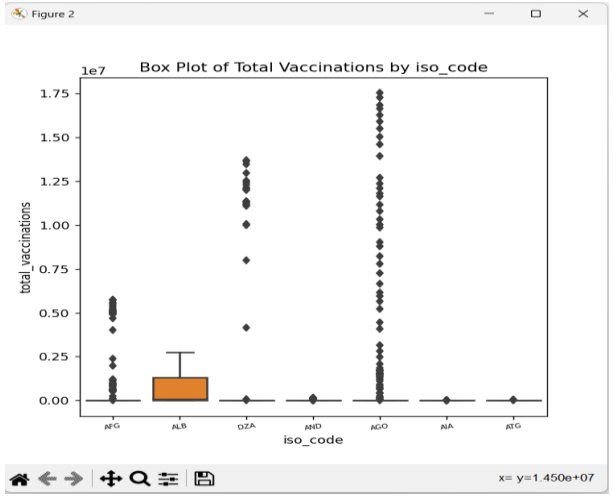


Figure 2
Box Plot of Total Vaccinations by iso_code
1e7
total_vaccinations
iso_code
AFG
ALB
DZA
AND
AGO
AIA
ATG
x= y=1.450e+07

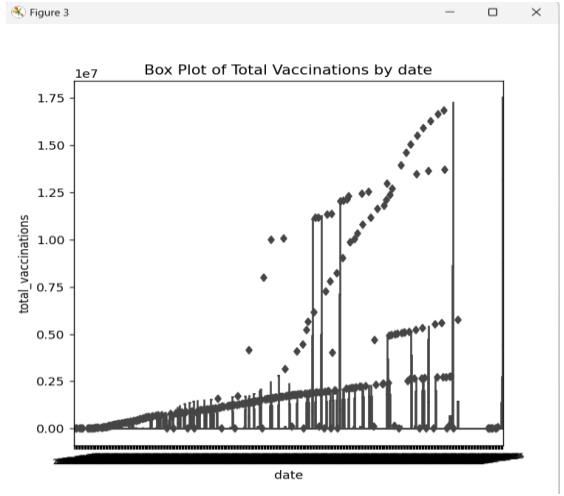


Figure 3
Box Plot of Total Vaccinations by date
1e7
total_vaccinations
date

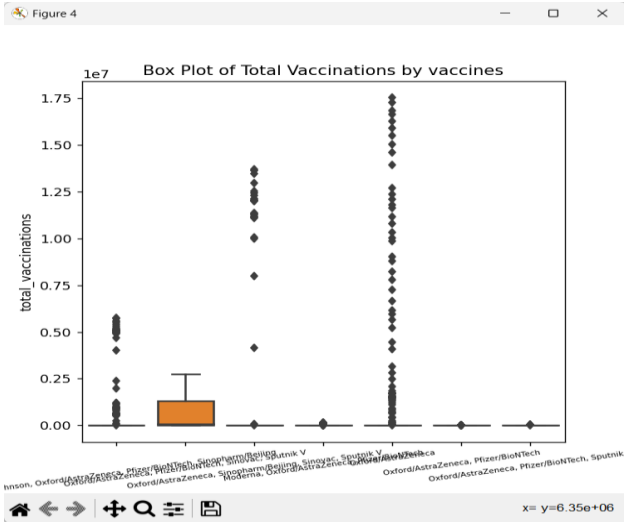


Figure 4
Box Plot of Total Vaccinations by vaccines
1e7
total_vaccinations
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Oxford/AstraZeneca, Pfizer/BioNTech
x= y=6.35e+06

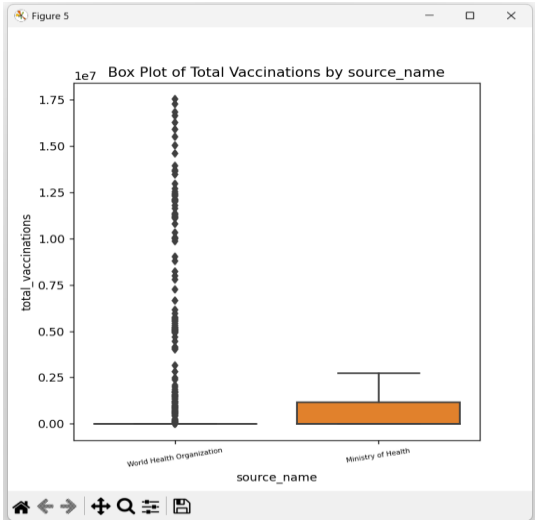


Figure 5
Box Plot of Total Vaccinations by source_name
1e7
total_vaccinations
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
World Health Organization
Ministry of Health
source_name

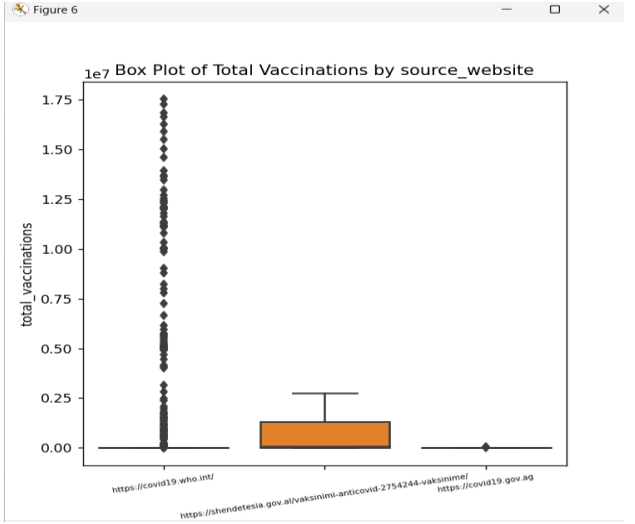


Figure 6
1e7 Box Plot of Total Vaccinations by source_website
total_vaccinations
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
https://covid19.who.int/
https://shendetesia.gov.al/vaksinimi-anticovid-2754244-vaksinime/
https://covid19.gov.ag

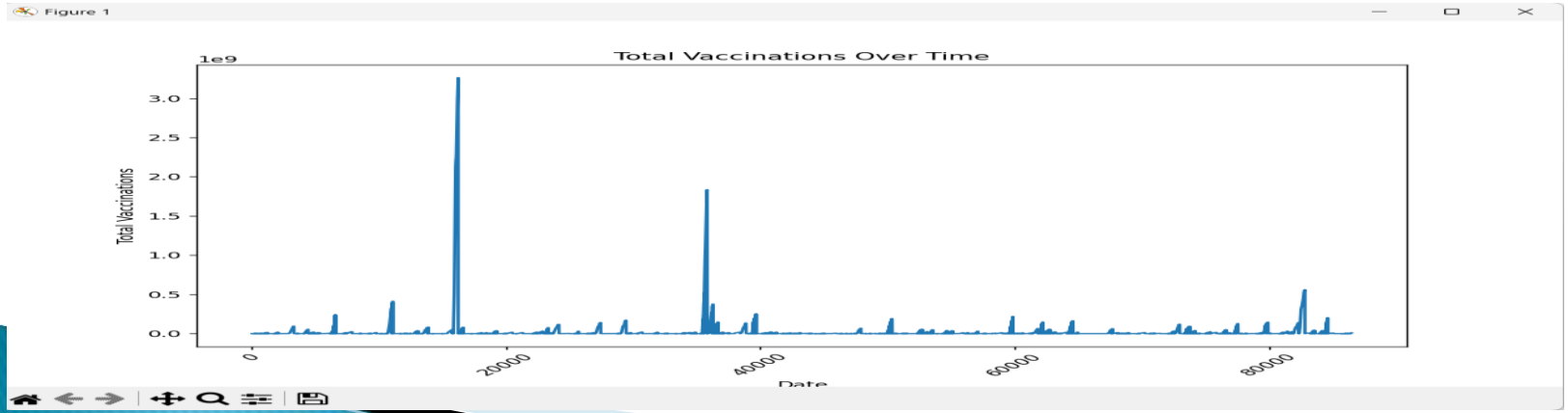


Figure 1
Total Vaccinations Over Time
1e9
Total Vaccinations
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
20000
40000
60000
80000
Date

PROGRAM FOR STATISTICAL ANALYSIS

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('data.csv')
print(df.head())
print(df.info())
print(df.describe())
df.fillna(0, inplace=True)
afghanistan_data = df[df['country'] == 'Afghanistan']
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
sns.lineplot(x='date', y='total_vaccinations', data=afghanistan_data)
```

```
plt.title('Total Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.subplot(1, 2, 2)
sns.lineplot(x='date', y='daily_vaccinations', data=afghanistan_data)
plt.title('Daily Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Daily Vaccinations')
plt.tight_layout()
plt.show()
df.to_csv('data.csv', index=False)
```

OUTPUT:

IDLE Shell 3.11.0

File Edit Shell Debug Options Window Help

```
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: D:\naan\New folder\eda.py ==========
Basic Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   country                              86512 non-null  object
 1   iso_code                             86512 non-null  object
 2   date                                 86512 non-null  object
 3   total_vaccinations                   86512 non-null  float64
 4   people_vaccinated                    86512 non-null  float64
 5   people_fully_vaccinated              86512 non-null  float64
 6   daily_vaccinations_raw               86512 non-null  float64
 7   daily_vaccinations                   86512 non-null  float64
 8   total_vaccinations_per_hundred       86512 non-null  float64
 9   people_vaccinated_per_hundred        86512 non-null  float64
 10  people_fully_vaccinated_per_hundred  86512 non-null  float64
 11  daily_vaccinations_per_million       86512 non-null  float64
 12  vaccines                             86512 non-null  object
 13  source_name                          86512 non-null  object
 14  source_website                       86512 non-null  object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None

Summary Statistics:
       total_vaccinations  ...  daily_vaccinations_per_million
count        8.651200e+04  ...                   86512.000000
mean         2.315117e+07  ...                    3245.792248
std          1.611037e+08  ...                    3932.156455
min          0.000000e+00  ...                       0.000000
25%          0.000000e+00  ...                     629.000000
50%          1.008000e+03  ...                    2036.000000
75%          3.697554e+06  ...                    4667.000000
max          3.263129e+09  ...                  117497.000000

[8 rows x 9 columns]

Missing Values:
country                                0
iso_code                               0
date                                   0
total_vaccinations                     0
people_vaccinated                      0
people_fully_vaccinated                0
daily_vaccinations_raw                 0
daily_vaccinations                     0
total_vaccinations_per_hundred         0
people_vaccinated_per_hundred          0
people_fully_vaccinated_per_hundred    0
daily_vaccinations_per_million         0
vaccines                               0
source_name                            0
source_website                         0
dtype: int64

Data Types:
country                                 object
iso_code                                object
date                                    object
total_vaccinations                     float64
people_vaccinated                      float64
people_fully_vaccinated                float64
daily_vaccinations_raw                 float64
daily_vaccinations                     float64
total_vaccinations_per_hundred         float64
people_vaccinated_per_hundred          float64
people_fully_vaccinated_per_hundred    float64
daily_vaccinations_per_million         float64
vaccines                                object
source_name                             object
source_website                          object
dtype: object

Unique Values in Categorical Columns:
country: 223 unique values
iso_code: 223 unique values
date: 483 unique values
vaccines: 84 unique values
source_name: 81 unique values
source_website: 119 unique values
>>>
```

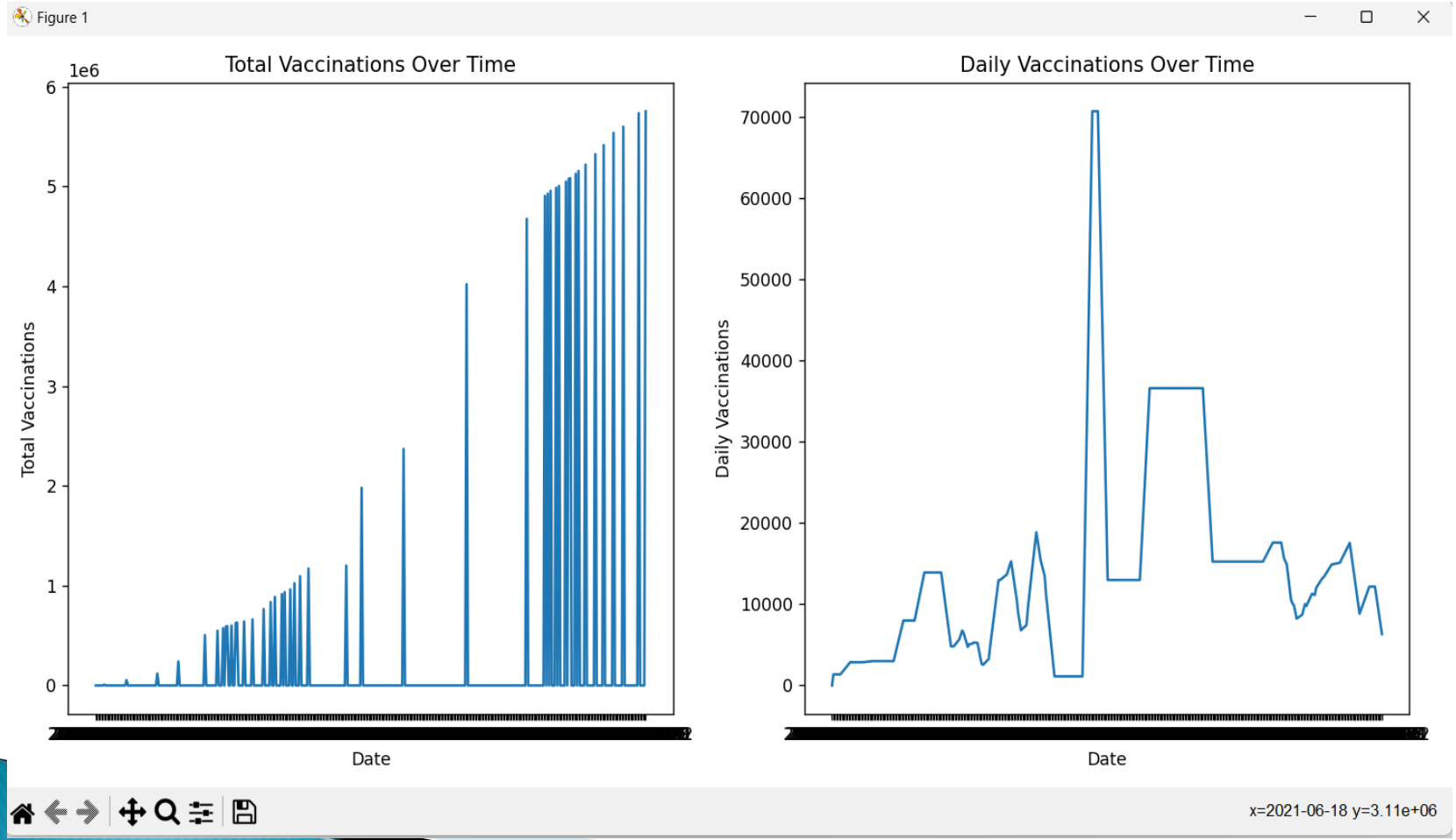


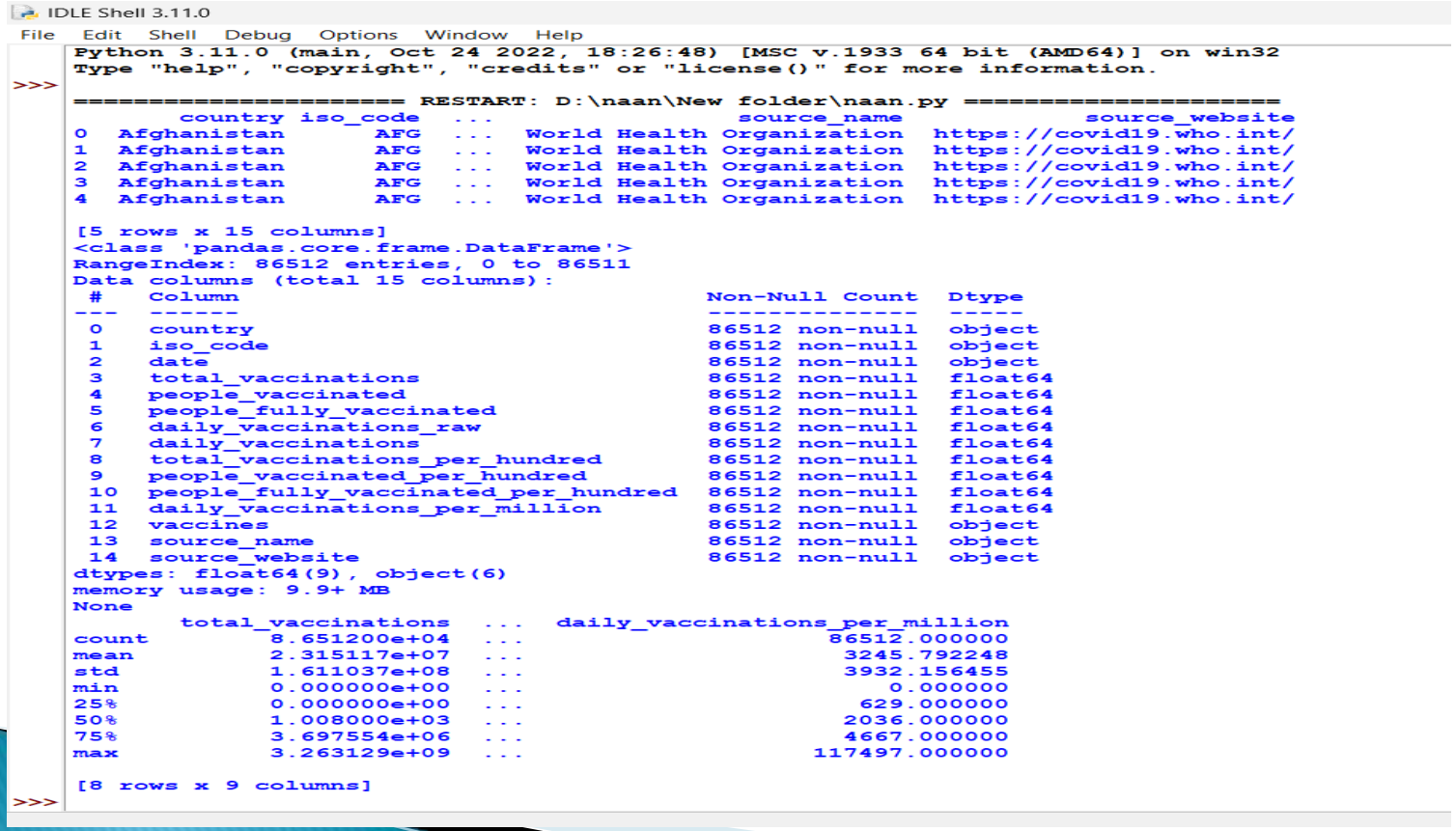
Figure 1
Total Vaccinations Over Time
1e6
Total Vaccinations
Date
Daily Vaccinations Over Time
Daily Vaccinations
Date
x=2021-06-18 y=3.11e+06

PROGRAM FOR VISUALIZATION:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
url = "data.csv"
data = pd.read_csv(url)
sns.set(style="whitegrid")
plt.figure(figsize=(8, 8))
sns.barplot(x='country', y='total_vaccinations', data=data.head(200))
plt.xticks(rotation=45)
plt.title('Total Vaccinations in Top 10 Countries')
plt.xlabel('Country')
plt.ylabel('Total Vaccinations')
plt.show()
afghanistan_data = data[data['country'] == 'Afghanistan']
plt.figure(figsize=(8, 8))
sns.lineplot(x='date', y='daily_vaccinations', data=afghanistan_data[0:100])
plt.xticks(rotation=90)
plt.xticks(fontsize=6)
```

```
plt.title('Daily Vaccinations Trend in Afghanistan')
plt.xlabel('Date')
plt.ylabel('Daily Vaccinations')
plt.show()
plt.figure(figsize=(8, 8))
sns.scatterplot(x='total_vaccinations', y='people_vaccinated', data=data)
plt.title('Total Vaccinations vs. People Vaccinated')
plt.xlabel('Total Vaccinations')
plt.ylabel('People Vaccinated')
plt.show()
plt.figure(figsize=(8, 8))
sns.boxplot(x='vaccines', y='daily_vaccinations', data=data)
plt.xticks(rotation=90)
plt.xticks(fontsize=6)
plt.title('Distribution of Daily Vaccinations by Vaccine Type')
plt.xlabel('Vaccine Type')
plt.ylabel('Daily Vaccinations')
plt.show()
```

OUTPUT:



```
IDLE Shell 3.11.0
File  Edit  Shell  Debug  Options  Window  Help
Python 3.11.0 (main, Oct 24 2022, 18:26:48) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
========== RESTART: D:\naan\New folder\naan.py ==========
       country iso_code  ...                source_name              source_website
0  Afghanistan       AFG  ...  World Health Organization  https://covid19.who.int/
1  Afghanistan       AFG  ...  World Health Organization  https://covid19.who.int/
2  Afghanistan       AFG  ...  World Health Organization  https://covid19.who.int/
3  Afghanistan       AFG  ...  World Health Organization  https://covid19.who.int/
4  Afghanistan       AFG  ...  World Health Organization  https://covid19.who.int/

[5 rows x 15 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   country                               86512 non-null  object
 1   iso_code                              86512 non-null  object
 2   date                                  86512 non-null  object
 3   total_vaccinations                    86512 non-null  float64
 4   people_vaccinated                     86512 non-null  float64
 5   people_fully_vaccinated               86512 non-null  float64
 6   daily_vaccinations_raw                86512 non-null  float64
 7   daily_vaccinations                    86512 non-null  float64
 8   total_vaccinations_per_hundred        86512 non-null  float64
 9   people_vaccinated_per_hundred         86512 non-null  float64
 10  people_fully_vaccinated_per_hundred   86512 non-null  float64
 11  daily_vaccinations_per_million        86512 non-null  float64
 12  vaccines                              86512 non-null  object
 13  source_name                           86512 non-null  object
 14  source_website                        86512 non-null  object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None
       total_vaccinations  ...  daily_vaccinations_per_million
count        8.651200e+04  ...                   86512.000000
mean         2.315117e+07  ...                    3245.792248
std          1.611037e+08  ...                    3932.156455
min          0.000000e+00  ...                       0.000000
25%          0.000000e+00  ...                     629.000000
50%          1.008000e+03  ...                    2036.000000
75%          3.697554e+06  ...                    4667.000000
max          3.263129e+09  ...                  117497.000000

[8 rows x 9 columns]
>>>
```

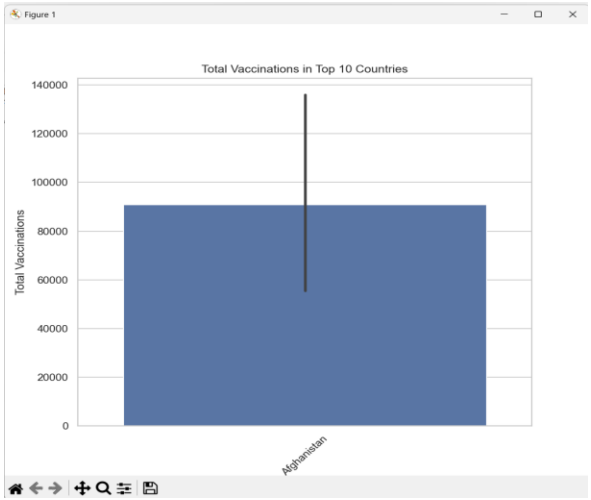


Figure 1
Total Vaccinations in Top 10 Countries
Total Vaccinations
140000
120000
100000
80000
60000
40000
20000
0
Afghanistan

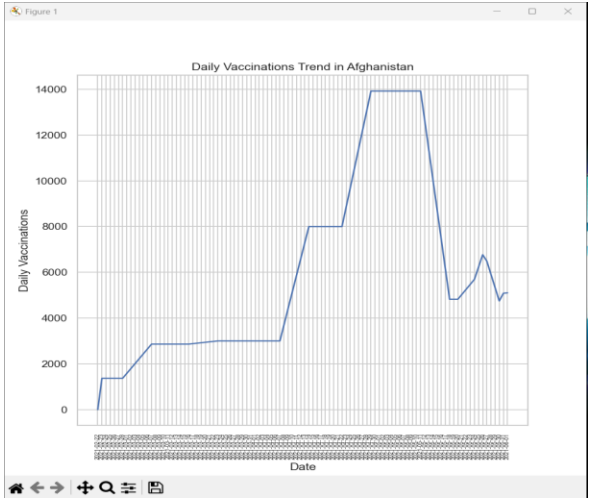


Figure 1
Daily Vaccinations Trend in Afghanistan
Daily Vaccinations
14000
12000
10000
8000
6000
4000
2000
0
Date

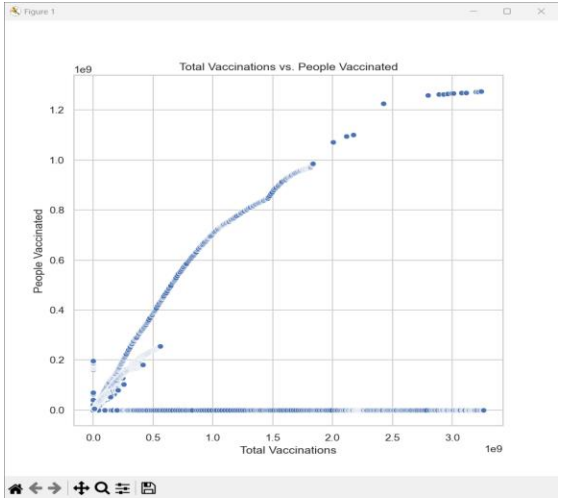


Figure 1
Total Vaccinations vs. People Vaccinated
1e9
People Vaccinated
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1e9
Total Vaccinations

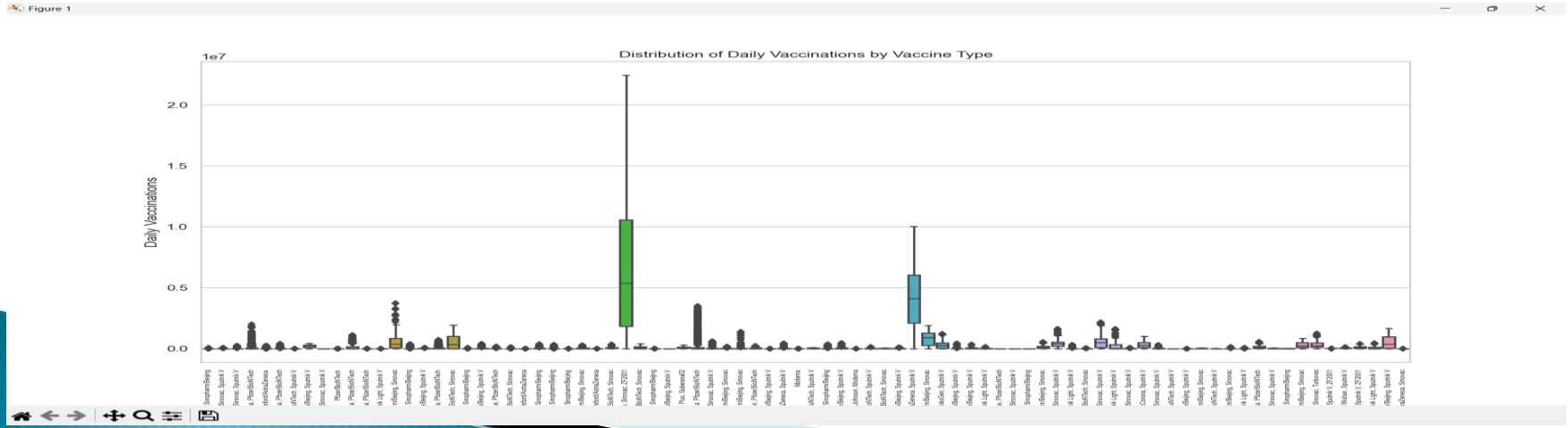


Figure 1
Distribution of Daily Vaccinations by Vaccine Type
1e7
Daily Vaccinations
2.0
1.5
1.0
0.5
0.0

CONCLUSION

The analysis of the COVID-19 vaccine dataset has provided valuable insights into the global vaccination effort. It is evident that vaccination progress is influenced by a combination of factors, including vaccine availability, distribution strategies, and regional disparities in healthcare resources.

To improve vaccination rates worldwide and ensure equitable access to vaccines, policymakers and public health officials should consider the following:

- Continuously monitor and adjust vaccination distribution strategies to address disparities.
- Promote public awareness and confidence in vaccines to encourage higher uptake.
- Collaborate with international organizations to ensure the availability of vaccines in underserved regions.
- Use data-driven insights to optimize vaccination campaigns and target high-risk populations.

This analysis serves as a foundation for further research and policy decisions aimed at effectively combatting the COVID-19 pandemic and achieving global vaccination goals.