Analyzing COVID-19 Vaccination Patterns And Comprehensive Analysis of Vaccination Data: From Understanding to Insights

Mrs. Uma Maheswari.M,
Assistant Professor/IT
Velammal College of Engineering and Technology Madurai, India.

Mr. Harish.R
Student/ECE
Velammal College of Engineering and Technology Madurai, India.
haribashyam122@gmail.com

Mr. Imayavaramban.C
Student/ECE
Velammal College of Engineering and Technology Madurai, India.
imayaelango9486@gmail.com

Mr. Janarthanan..M
Student/ECE
Velammal College of Engineering and Technology Madurai, India.
janarthananm089@gmail.com

Mr. Mani sankar.P
Student/ECE
Velammal College of Engineering and Technology Madurai, India.
sankarmani687@gmail.com

Mr. Yasar Aarafath.S
Student/ECE
Velammal College of Engineering and Technology Madurai, India.
ya0387288@gmail.com

***Abstract*— This case study project delves into a comprehensive analysis of vaccination data, covering various stages from data understanding to insights and interpretation. It begins with importing necessary libraries and loading datasets, followed by an exploratory data analysis (EDA) section that explores data structure, content, and preprocessing steps. Visualizations such as heatmaps and bar plots are employed to represent data graphically, facilitating better comprehension of complex information. Statistical analysis provides quantitative insights into key attributes, aiding in understanding data distribution and central tendencies. The documentation concludes with insights and interpretations drawn from the data, supporting decision-making processes. Additionally, the documentation is designed to be customizable, enabling users to adapt it to their specific analysis goals or share it for collaboration.**

Keywords*— COVID-19, vaccination analysis, regression modeling, country-level trends, vaccine-type-specific patterns, vaccination dynamics, public health, global immunization, healthcare policy, data analysis, epidemiology, vaccine preferences, age demographics, pandemic response.

Introduction

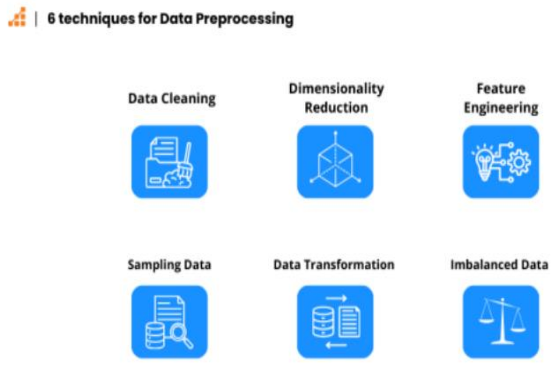
. The COVID-19 pandemic has posed unprecedented challenges to the global health system, affecting millions of lives and livelihoods. As of February 2024, more than 300 million cases and 6 million deaths have been reported worldwide, with new variants of the virus emerging and spreading rapidly[1]. To combat the pandemic, several vaccines have been developed and authorized for emergency use, offering hope and protection to the population. However, the availability, accessibility, and acceptance of COVID-19 vaccines vary significantly across countries, regions, and groups, creating disparities and inequities in the immunization coverage and outcom[2],[3].

Understanding the factors and patterns that influence the COVID-19 vaccination dynamics is crucial for designing and implementing effective and equitable strategies to enhance the vaccine uptake and impact. However, the existing literature on this topic is limited, fragmented, and often inconclusive. There is a need for a comprehensive and rigorous analysis of the COVID-19 vaccination dynamics using advanced statistical and computational methods, incorporating multiple dimensions and perspectives.

This research endeavors to provide such an analysis, using regression models to explore the COVID-19 vaccination dynamics across countries, vaccine types, and age groups. Leveraging extensive datasets from various sources, our study employs regression techniques to unveil intricate relationships and predictive insights across diverse parameters. The investigation unfolds insights into vaccination uptake variations among countries, preferences for specific vaccine types,and the impact of age demographics on immunization trends. By employing a multifaceted regression approach, our findings aim to contribute nuanced perspectives to the ongoing discourse surrounding COVID-19 vaccine distribution and adoption strategies. The study not only identifies key determinants influencing vaccination outcomes but also offers valuable implications for policymakers, healthcare practitioners, and public health initiatives aiming to optimize immunization efforts in diverse global contexts.

I. DATA PRE-PROCESSING

Data preprocessing is a critical step in data analysis and machine learning, encompassing techniques to clean, transform, and prepare raw data for analysis. This process involves handling missing values, duplicates, and outliers, transforming variables into suitable formats, engineering features for better model performance, reducing dimensionality, addressing imbalanced data distributions, splitting data for model evaluation, and normalizing features. By performing these preprocessing steps, the data is refined and optimized for subsequent analysis or modeling tasks, ensuring more accurate and reliable results.



PEOPLE FULLY VACCINATED STATISTICAL ANALYSIS:

df_manufacturer['people_fully_vaccinated'].describe()

Output:

```
count    7.400000e+01
mean     3.046394e+07
std      8.904082e+07
min      2.607400e+04
25%      1.296678e+06
50%      5.586156e+06
75%      2.915660e+07
max      7.247684e+08
Name: people_fully_vaccinated, dtype: float64
```

MOST USED VACCINE IN THE WORLD:

df_manufacturer['vaccine'].value_counts()

Output:

```
Pfizer/BioNTech        39
Moderna                35
Johnson&Johnson        33
Oxford/AstraZeneca     30
Novavax                 8
Sinovac                 6
Sinopharm/Beijing       4
CanSino                 2
Sputnik V               2
Covaxin                 1
Name: vaccine, dtype: int64
```

DAILY VACCINATIONS PER MILLION TOP COUNTRIES

df_vaccinations.groupby("country")["daily_vaccinations_per_million"].mean().sort_values(ascending=False).head(20)

Output:

```
country
Finland         9154.0
New Zealand     8800.0
Australia       8621.0
Italy           6733.0
Peru            6609.0
Saudi Arabia    6330.0
Iran            6280.0
Argentina       5814.0
Canada          5165.0
Brazil          4864.0
Pakistan        4841.0
India           4281.0
Greece          4055.0
Singapore       3951.0
Luxembourg      3845.0
Mexico          3792.0
Thailand        3718.0
South Korea     3447.0
Kenya           3315.0
Hong Kong       3293.0
Name: daily_vaccinations_per_million, dtype: float64
```

PREFERRED VACCINE IN INDIA

x = df_vaccinations[df_vaccinations["country"] == "India"]
z = x.vaccines.value_counts() c = list(z.index)
print(c)

Output:

```
['Covaxin, Oxford/AstraZeneca, Sputnik V']
```

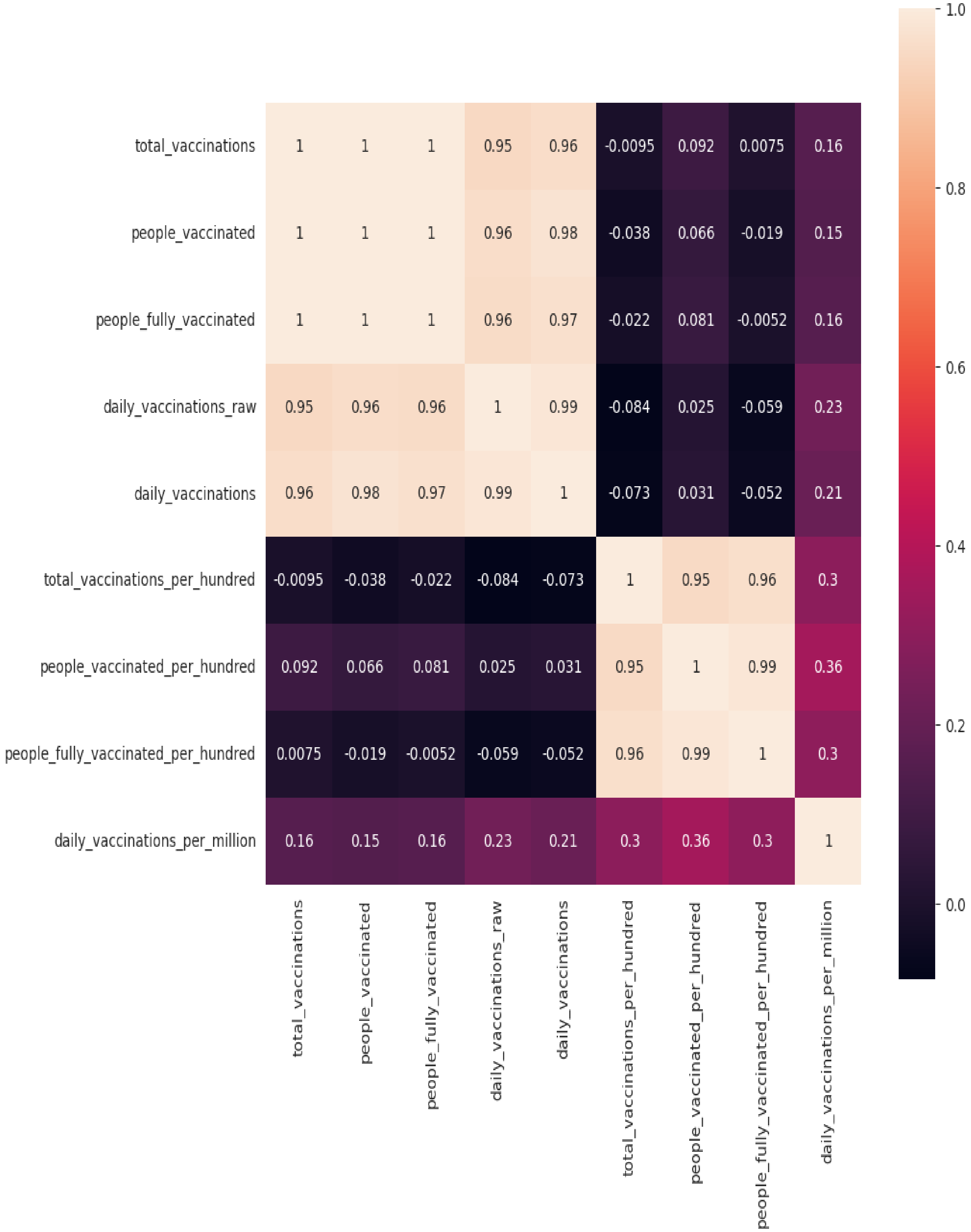
VACCINES MANUFACTURED ON A PARTICULAR DATE:

df_manufacturer = df_manufacturer[df_manufacturer.date == '2022-02-04']
df_manufacturer.head()

	location	date	vaccine	total_vaccinations
2305	Argentina	2022-02-04	CanSino	468481
2306	Argentina	2022-02-04	Moderna	5318406
2307	Argentina	2022-02-04	Oxford/AstraZeneca	25606912
2308	Argentina	2022-02-04	Pfizer/BioNTech	11225368
2309	Argentina	2022-02-04	Sinopharm/Beijing	27396208

VISUALIZATION OF DATA:

plt.subplots(figsize = (10,10))
sns.heatmap(df_vaccinations.corr(), annot = True, square = True)
plt.show()


total_vaccinations 1 1 1 0.95 0.96 -0.0095 0.092 0.0075 0.16
people_vaccinated 1 1 1 0.96 0.98 -0.038 0.066 -0.019 0.15
people_fully_vaccinated 1 1 1 0.96 0.97 -0.022 0.081 -0.0052 0.16
daily_vaccinations_raw 0.95 0.96 0.96 1 0.99 -0.084 0.025 -0.059 0.23
daily_vaccinations 0.96 0.98 0.97 0.99 1 -0.073 0.031 -0.052 0.21
total_vaccinations_per_hundred -0.0095 -0.038 -0.022 -0.084 -0.073 1 0.95 0.96 0.3
people_vaccinated_per_hundred 0.092 0.066 0.081 0.025 0.031 0.95 1 0.99 0.36
people_fully_vaccinated_per_hundred 0.0075 -0.019 -0.0052 -0.059 -0.052 0.96 0.99 1 0.3
daily_vaccinations_per_million 0.16 0.15 0.16 0.23 0.21 0.3 0.36 0.3 1
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
1.0
0.8
0.6
0.4
0.2
0.0

TOP COUNTRIES IN VACCINATIONS UTILIZATION

df_vaccinations["Total_vaccinations_count"]= df_vaccinatios.groupby("country").total_vaccinations.tail(1) df_vaccinatios.groupby("country")["Total_vaccinations_count"].mean().sort_values(ascending= False).head(10)

```
country
India              1.687048e+09
United States      5.469684e+08
Brazil             3.677782e+08
Pakistan           1.823960e+08
Vietnam            1.816654e+08
Mexico             1.685357e+08
Germany            1.666940e+08
Russia             1.553786e+08
Turkey             1.427355e+08
United Kingdom     1.384598e+08
Name: Total_vaccinations_count, dtype: float64
```

FULLY VACCINATED COUNT:

df_vaccinations["Full_vaccinations_count"]=df_vaccinations.groupby ("coutry").people_fully_vaccinated.tail(1) df_vaccinations.groupby("country")["Full_vaccinations_count"].mean().sort_values(ascending= False).head(10)

```
country
India             724768356.0
United States     213893460.0
Brazil            150682483.0
Pakistan           84731497.0
Mexico             77478070.0
Vietnam            74187748.0
Russia             70232028.0
Germany            61873548.0
Iran               54405243.0
Turkey             52489431.0
Name: Full_vaccinations_count, dtype: float64
```

MOST COMMONLY USED VACCINES IN THE WORLD:

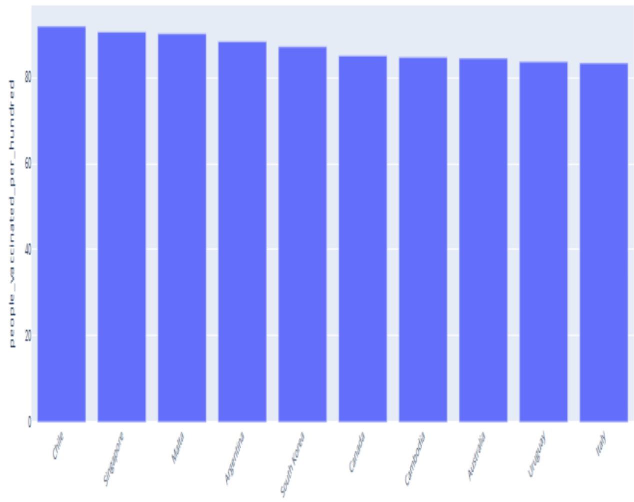
total = df_manufacturer.groupby('vaccine').sum() px.bar(x=total.index, y=total['total_vaccinations'],title='Most Used Vaccine in the World')

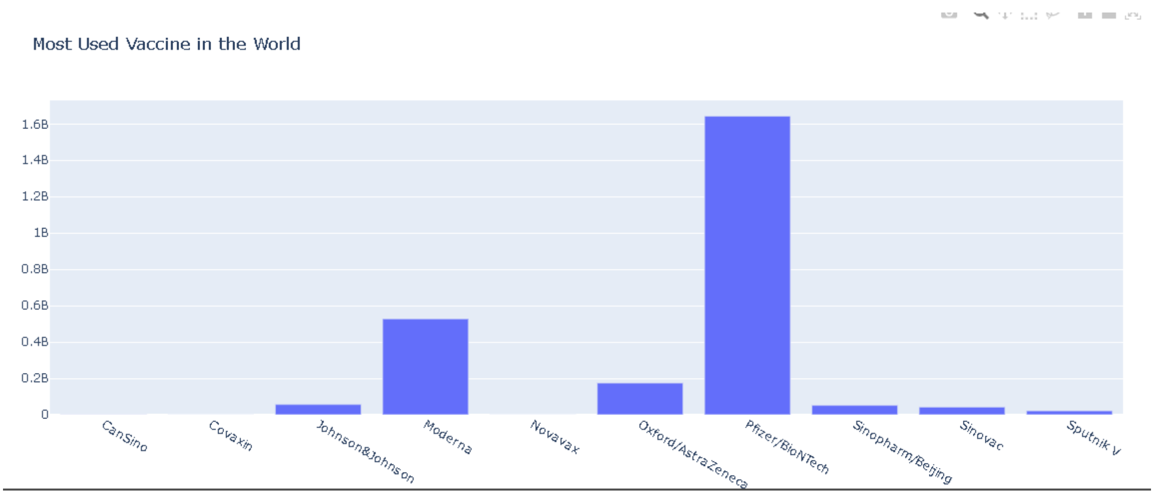
FULLY VACCINATED COUNT:

df_vaccinations.groupby("country")["Full_vaccinations_count"].mean().sort_values(ascending= False).head(10)

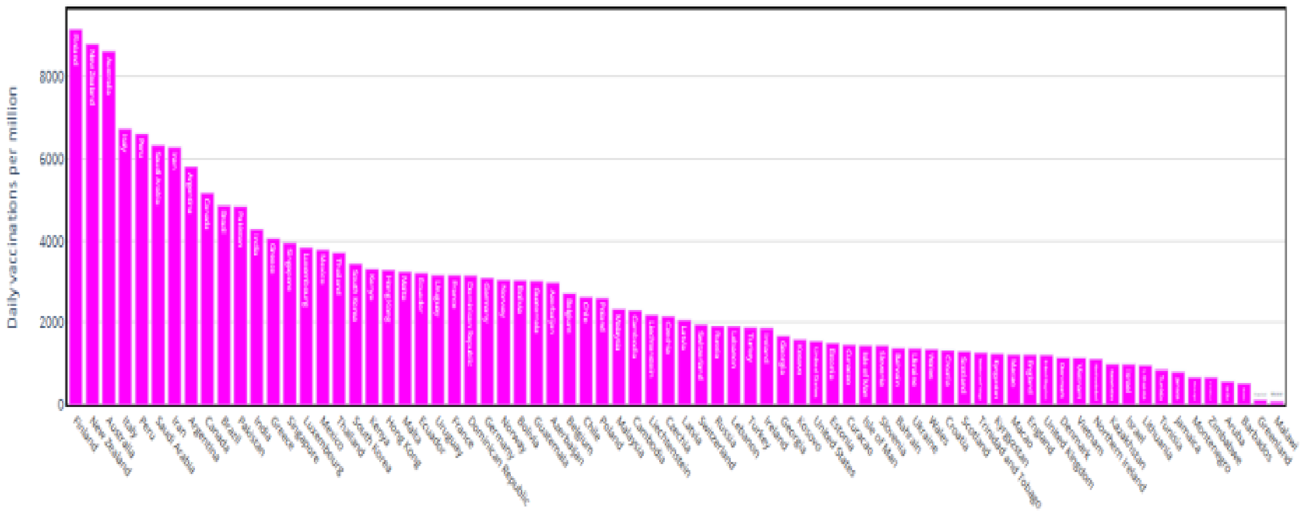
```
country
India             724768356.0
United States     213893460.0
Brazil            150682483.0
Pakistan           84731497.0
Mexico             77478070.0
Vietnam            74187748.0
Russia             70232028.0
Germany            61873548.0
Iran               54405243.0
Turkey             52489431.0
Name: Full_vaccinations_count, dtype: float64
```

PEOPLE VACCINATED PER HUNDRED FOR THE DATE 2022-02-04



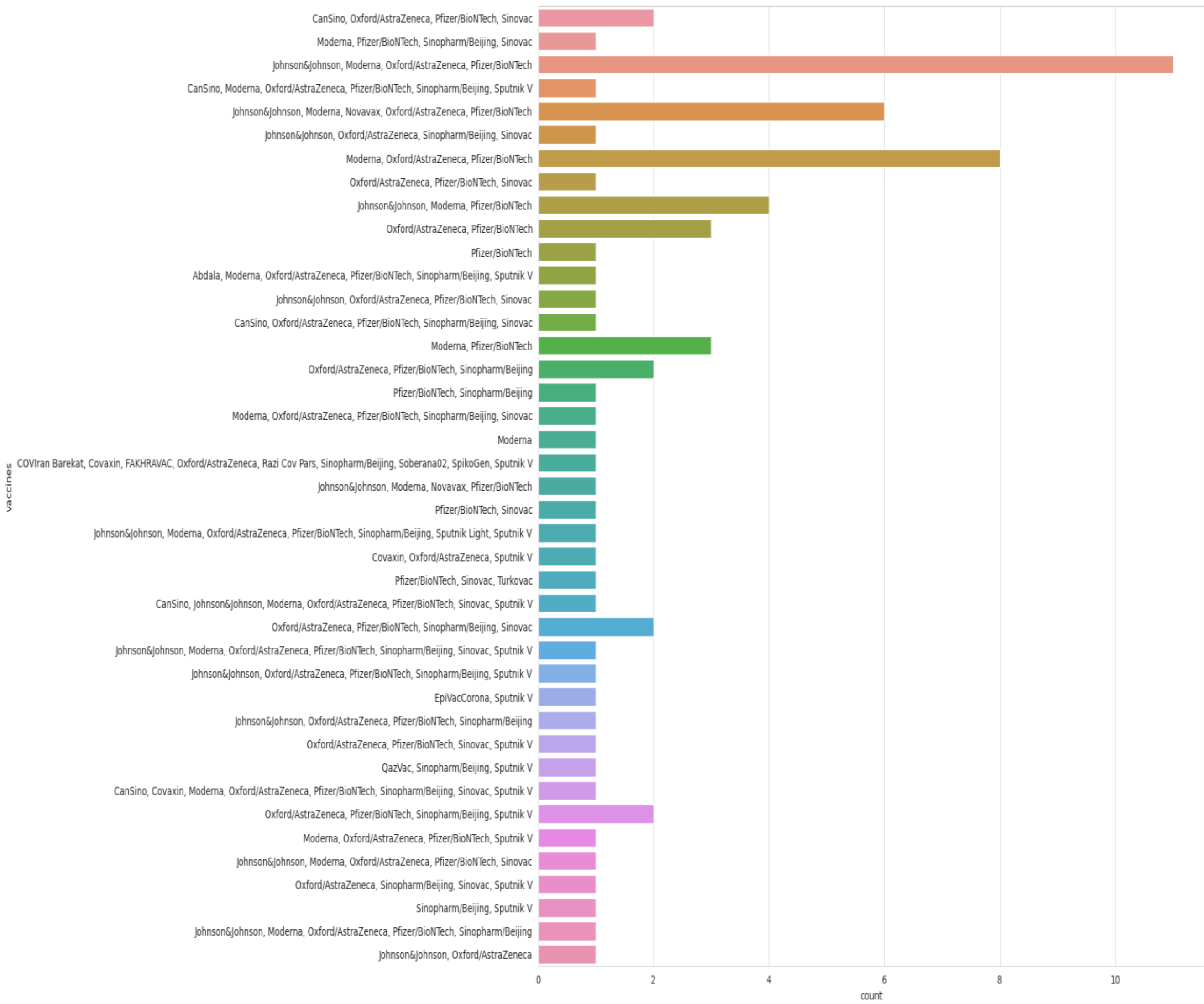


COUNTRYWISE DAILY VACCINATION PER-MILLION:



TYPE OF VACCINE UTILIZED VS COUNT:

```
plt.figure(figsize=(15,15))
sns.countplot(y= "vaccines",data= df_vaccinations)
plt.show()
```



Outcome:

1.Data Understanding:

- The documentation begins with importing necessary libraries and loading the data sets, providing insight into the tools and datasets used.

2. Exploratory Data Analysis (EDA):

- The EDA section explores the structure and content of the data, presenting key information such as columns, shapes, and data types.
- It also includes the preprocessing steps, checking and handling missing values, and ensuring data quality.

3. Visualization:

- Various visualizations are included to provide a graphical representation of the data. This includes heatmaps, bar plots, and charts showcasing top countries in vaccination, preferred vaccines, and more.
- The visualizations aim to make complex data more understandable and highlight trends and patterns.

4. Statistical Analysis:

- Statistical analysis is performed on key attributes, providing summary statistics such as mean, standard deviation, and quartiles.
- This section enables a quantitative understanding of the data distribution and central tendencies.

5. Insights and Interpretation:

- Throughout the documentation, insights are provided, such as the top countries in vaccination, most commonly used vaccines, and statistical summaries.
- These insights aid in drawing meaningful conclusions from the data, supporting decision-making processes.

6. Customization:

- The documentation is designed to be customizable based on specific requirements. Users can adapt and extend the documentation to suit their analysis goals or share it with others to facilitate collaboration.

REFERENCES:

1:World Health Organization. Coronavirus disease (COVID-2019 situation reports.

2: COVID-19 Vaccine Hesitancy: Demographic Factors, Geographic Patterns, and Changes Over Time.

3: COVID-19 vaccine challenges.

4.The Comparison Study of Regression Models (Multiple Linear Regression, Ridge, Lasso, Random Forest, and Polynomial Regression) for House Price Prediction in West Nusa Tenggara

5.Real-time prediction of COVID-19 related mortality using electronic health records

6.World Health Organization. Coronavirus disease (COVID-2019) situation reports.

7. D. L. Heymann and N. Shindo, "COVID-19: What is next for public health?", Lancet, vol. 395, no. 10224, pp. 542-545, Feb. 2020.

8. J. Liu et al., "Data mining and information retrieval in the 21st century: A bibliographic review", Comput. Sci. Rev., vol. 34, Nov. 2019.

9. Q. Zhou, W. Tao, Y. Jiang and B. Cui, "A comparative study on the prediction model of COVID-19", Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), vol. 9, pp. 1348-1352, Dec. 2020.

10. W. Li et al., "Progression of mental health services during the COVID-19 outbreak in China", Int. J. Biol. Sci., vol. 16, no. 10, pp. 1732, 2020.