COVID 19 VACCINES

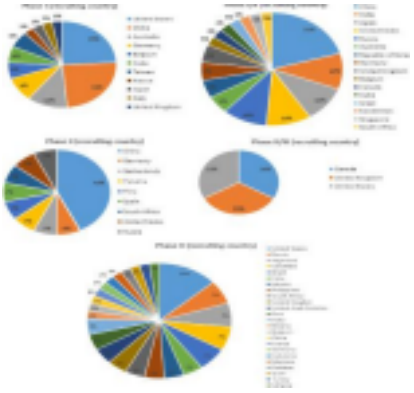
ANALYSIS Team member

723721205018 : GOKULA KANNAN M

Phase 5 submission document

Project Title: Analysis for covid 19 vaccines

Phase 5: Project Documentation and Submission



INTRODUCTION

❖The emergence of the COVID-19 pandemic has presented humanity with one of its most formidable challenges. As the virus spread across the globe, scientists, healthcare professionals, and pharmaceutical companies rallied to develop effective vaccines to combat the virus and protect the world's population.

❖The development and distribution of COVID-19 vaccines have been monumental milestones in our collective fight against the pandemic.

❖Understanding the impact and effectiveness of these

vaccines is a critical aspect of the ongoing battle against COVID-19. To this end, rigorous analysis of COVID-19 vaccines has become an essential pursuit.

❖This analysis seeks to provide insights, inform public health decisions, and guide policymakers and healthcare organizations in their strategies to achieve widespread vaccination and curb the spread of the virus.

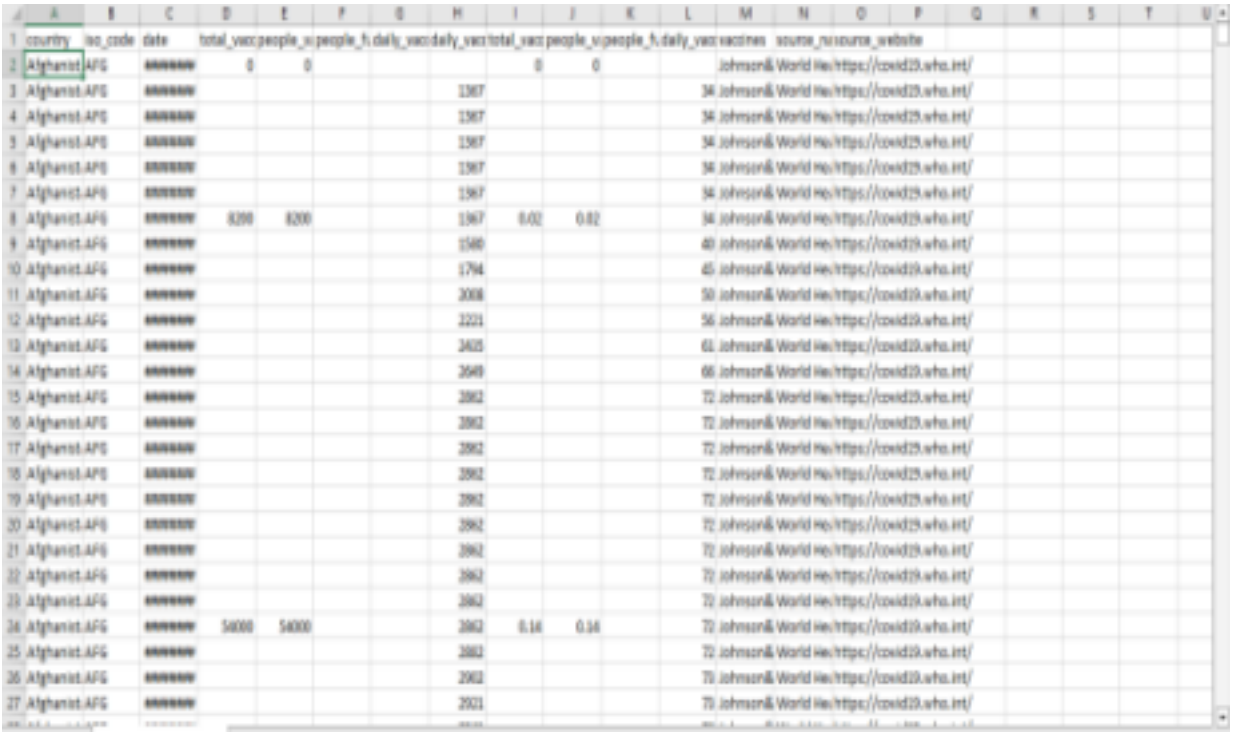
❖In this comprehensive analysis, we delve into various aspects of COVID-19 vaccines, ranging from vaccination coverage and efficacy to the safety of these life-saving doses. We explore data trends, patterns, and correlations, and we consider the implications of our findings in the broader context of public health.

❖As we collectively work to achieve global vaccination coverage and transition to a post-pandemic world, this analysis serves as a vital resource for informed decision making and a testament to the power of science, medicine, and the collective determination to overcome adversity.

Data set link:

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

GIVEN DATASET



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	country	iso_code	date	total_vacc	people_v	people_fi	daily_vacc	daily_vacc	total_vacc	people_v	people_fi	daily_vacc	vaccines	source_na	source_website						
2	Afghanist	AFG	######	0	0				0	0			Johnson&	World He	https://covid19.who.int/						
3	Afghanist	AFG	######					1367				34	Johnson&	World He	https://covid19.who.int/						
4	Afghanist	AFG	######					1367				34	Johnson&	World He	https://covid19.who.int/						
5	Afghanist	AFG	######					1367				34	Johnson&	World He	https://covid19.who.int/						
6	Afghanist	AFG	######					1367				34	Johnson&	World He	https://covid19.who.int/						
7	Afghanist	AFG	######					1367				34	Johnson&	World He	https://covid19.who.int/						
8	Afghanist	AFG	######	8200	8200			1367	0.02	0.02		34	Johnson&	World He	https://covid19.who.int/						
9	Afghanist	AFG	######					1560				40	Johnson&	World He	https://covid19.who.int/						
10	Afghanist	AFG	######					1794				45	Johnson&	World He	https://covid19.who.int/						
11	Afghanist	AFG	######					2008				50	Johnson&	World He	https://covid19.who.int/						
12	Afghanist	AFG	######					2221				56	Johnson&	World He	https://covid19.who.int/						
13	Afghanist	AFG	######					2435				61	Johnson&	World He	https://covid19.who.int/						
14	Afghanist	AFG	######					2649				66	Johnson&	World He	https://covid19.who.int/						
15	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
16	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
17	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
18	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
19	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
20	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
21	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
22	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
23	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
24	Afghanist	AFG	######	54000	54000			2862	0.14	0.14		72	Johnson&	World He	https://covid19.who.int/						
25	Afghanist	AFG	######					2862				72	Johnson&	World He	https://covid19.who.int/						
26	Afghanist	AFG	######					2903				73	Johnson&	World He	https://covid19.who.int/						
27	Afghanist	AFG	######					2921				73	Johnson&	World He	https://covid19.who.int/						

OBJECTIVES

This project unveils the important features of covid data. The distribution of COVID-19 vaccines have been pivotal in the global response to the ongoing pandemic caused by the SARS CoV-2 virus. Since the emergence of the virus in late 2019, researchers, healthcare professionals, and policymakers have been working tirelessly to understand, evaluate, and optimize various COVID-19 vaccines. Using this data we can determine how the vaccines are distributed on timely basis and also helps to increase vaccine production depending on demand. Let's see what're all the steps and processes involved in this covid analysis.

IMPORTING LIBRARIES

A library in the context of computer programming refers to a collection of pre-written code or modules that provide specific functions or features. These libraries are designed to be reusable, helping developers save time and effort by offering ready-made solutions to common programming tasks. Libraries can cover a wide range of functionalities, such as data manipulation, web development, machine learning, or graphics, and they are an integral part of many programming languages, including Python.

Since our project is analysis we may need only few libraries and we won't need any other algorithms. They are

Pandas: Used for data manipulation and analysis, particularly with structured data in tables.

Matplotlib: A popular library for creating static, animated, or interactive visualizations in Python.

NumPy: A fundamental library for numerical operations, providing support for arrays and matrices.

Seaborn: is a Python data visualization library based on Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics.

Plotly: Plotly is a Python graphing library and interactive

visualization tool that allows you to create a wide variety of interactive, web-based plots and data visualizations.

LOADING DATA AND PREPROCESSING

Loading data is a fundamental step in many data analysis and machine learning projects. It involves bringing external data into your Python environment for further processing.

For our analysis the dataset is "COVID-19 World Vaccination Progress" . We use pandas framework to read this dataset from local disk or directly from Kaggle.

Data = pd.readcsv("COVID-19 World Vaccination Progress")

GETTING COLUMN NAMES:

To return the columns present in our dataset we use Data.columns which will give the each column’s name as below

Index(['country', 'iso_code', 'date',

'total_vaccinations "people_vaccinated'.

'people_fully_vaccinated'

"dally-vaccinations_raw', 'dally_vaccinations

'total_vaccinations_per_hundred`.
'people_vaccinated_per_hundred'

people_fully_vaccinated_per_hundred',
'dally_vaccinations_per_million',

'vaccines', 'source_name'

dtype='object')

'source-website'].

Finding metrics

Data.describe() returns count, mean , standard deviation, minimum, 25%,50%,75% and maximum. It helps to gain insights.

Shape

The dataset has a total of 15 columns and 86512 records. Once the dataset is loaded we will

take a look at the dataset entries. head() is used to get the starting records from the dataset and tail() is used to get the ending records from the dataset. We can also specify how much records are to be displayed.

DATA PREPROCESSSING

Data preprocessing, on the other hand, involves cleaning, transforming, and organizing the data to make it suitable for analysis and modeling. This may include handling missing values, normalizing or scaling features, encoding categorical variables, and splitting the data into training and testing sets. Data preprocessing is critical as it helps improve the quality of the data and ensures that machine learning algorithms can

work effectively.

We can get sources from Github, Kaggle etc.

Steps:

In the above figure data.tail(5) has returned the last 5 records.

Preprocessing Steps:

Here are common data preprocessing steps:

1. Data Cleaning:

- Handling Missing Values: Fill in or remove missing data points.

- Outlier Detection: Identify and deal with outliers that may skew analysis.

2. Data Transformation:

- Feature Scaling: Normalize or standardize numerical features to a common scale.

- Encoding Categorical Data: Convert categorical variables into numerical representations (e.g., one-hot encoding).

- Feature Engineering: Create new features or transform existing ones to improve model performance.

3. Data Reduction:

- Dimensionality Reduction: Reduce the number of features while retaining meaningful information (e.g., using techniques

like Principal Component Analysis).

4. Data Splitting:

- Split the dataset into training and testing sets to assess model performance.

5. Data Normalization:

- Normalize the data to have a standard distribution (e.g., using Z-score normalization).

6. Data Balancing:

- Address class imbalances, especially in classification tasks, to prevent bias in the model.

7. Text Data Preprocessing:

- Tokenization: Split text into words or tokens.

- Stop Word Removal: Eliminate common, non-informative words.

- Lemmatization or Stemming: Reduce words to their base form.

8. Time Series Data Preprocessing:

- Resampling: Aggregate or interpolate time series data into

different time intervals.

- Feature Extraction: Create meaningful features from time based data.

9. Handling Duplicates:

- Identify and remove duplicate records to avoid bias in analysis.

In our project we only need few of the above preprocessing steps .

Missing values:

isnull() returns true or false in presence or absence of NaN values in the records. Using sum() we can get the total number of null values in the records.

Removing null values:

dropna() is used to remove NaN values from columns mentioned. Here we have removed

null values from columns 'total_vaccinations' and 'people_vaccinated'.

Removing unecessary columns:

Since there is no use of 'daily_vaccinations_raw' and 'source_website' we have dropped these

columns using drop() by defining axis=1 which denotes the columns.

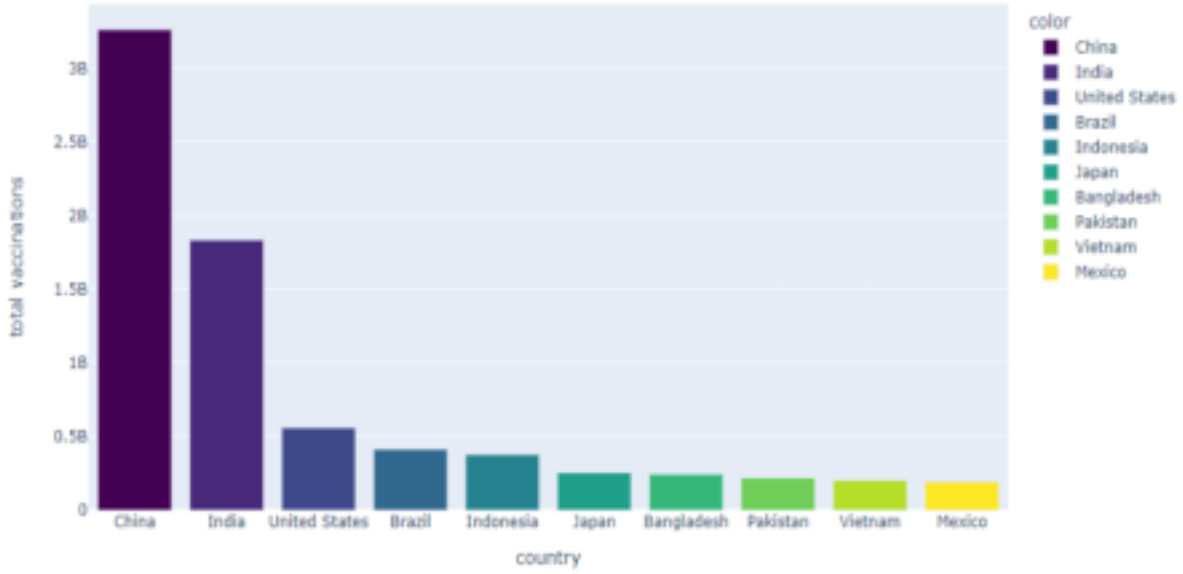
No transfornation is used as we need to plot actual data rather than Normalised data.

VACCINATION BY COUNTRY:

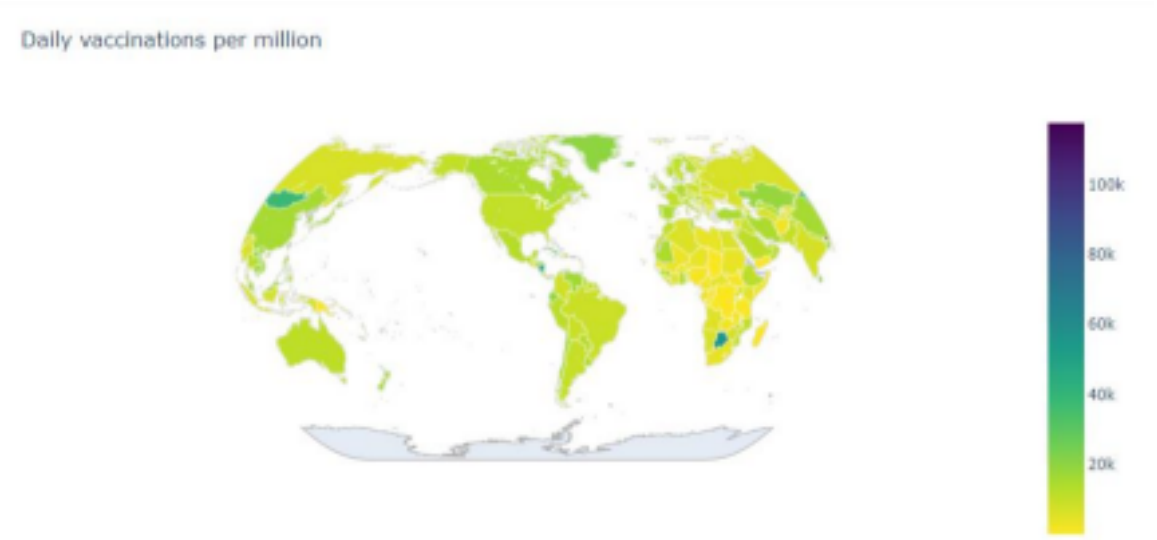
Visualization in Python is a crucial part of data analysis, and there are several libraries available to help you create a wide range of charts, graphs, and interactive visualizations.

A bar graph, also known as a bar chart or bar plot, is a common type of data visualization used to represent categorical data. It displays data using rectangular bars, where the length (or height) of each bar is proportional to the value it represents.

In this we have used plotly to show a bar graph as follows.

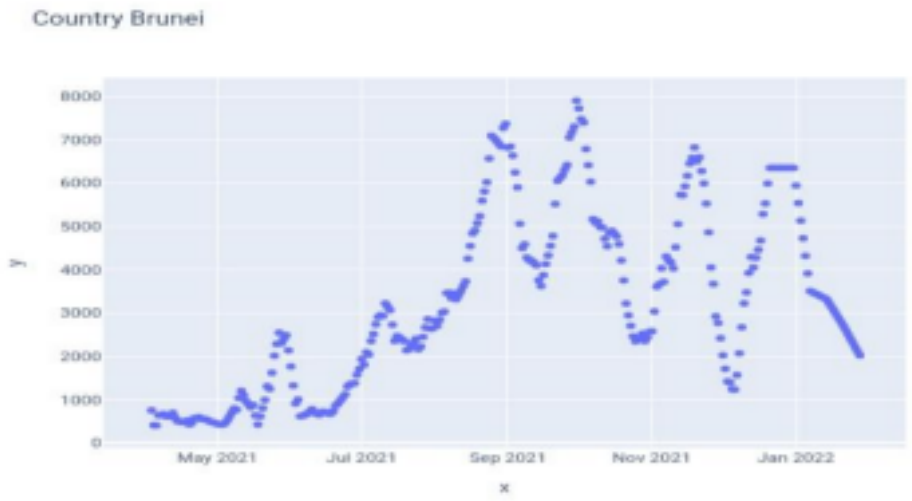


MAP VISUALISATION USING PLOTLY:



Example scatter plot of a country:(daily vaccination)

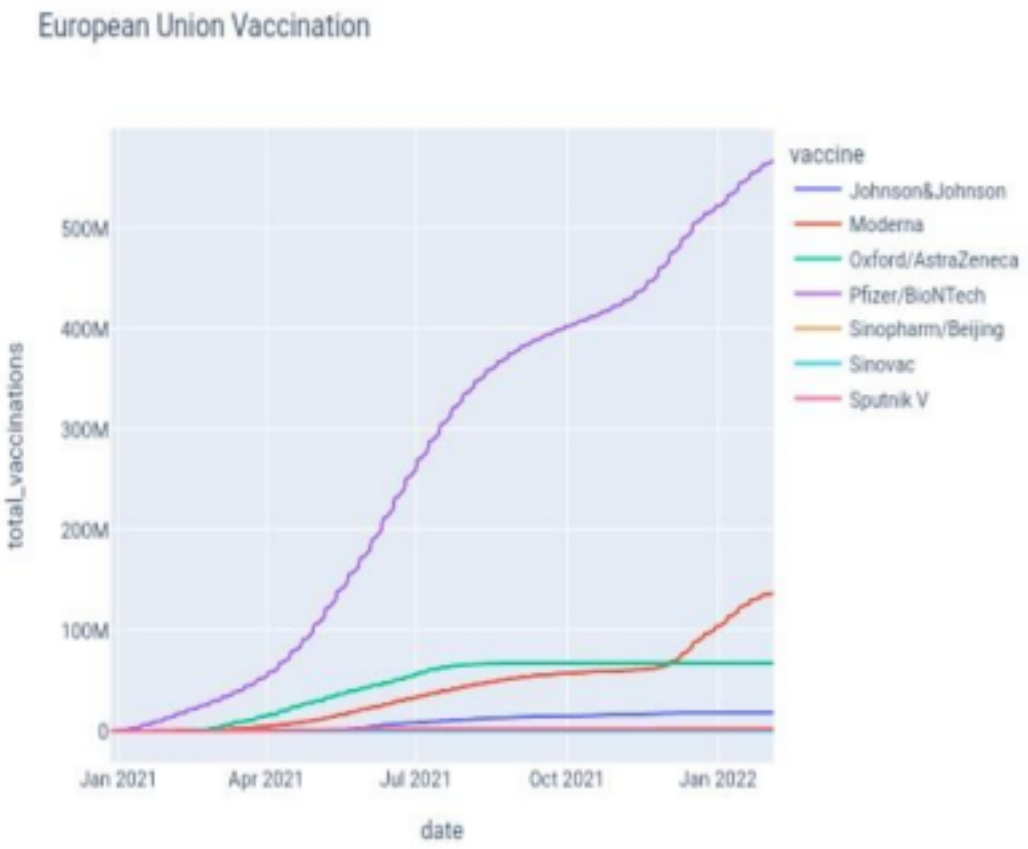
A scatter plot is a type of data visualization that displays individual data points on a two-dimensional graph. Each data point is represented by a dot or marker, and the position of the dot on the graph is determined by its x and y coordinates.



This scatter plot shows the everyday fluctuations in vaccinations in the country Brunei.

Most used vaccine:

A line graph, also known as a line chart or line plot, is a type of data visualization that represents data points with a series of connected line segments. It is used to show how data changes over a continuous interval or time, making it an effective tool
for displaying trends and variations in data. The different colors in the graph represents different vaccines and their magnitudes represent their usage.



CONCLUSION

Analyzing the COVID-19 vaccines is an ongoing process, and as of my last knowledge update in January 2022, several vaccines were developed and authorized for emergency use or full approval in various parts of the world. These vaccines have played a crucial role in the global effort to control and eventually end the COVID-19 pandemic.Please note that the situation regarding COVID-19 and its vaccines has likely evolved since my last update in January 2022. For the most current and accurate information, it is important to refer to reputable sources such as the World Health Organization (WHO), the Centers for Disease Control and Prevention

(CDC), and other relevant health authorities.