# Covid-19 vaccine analysis

**This Project mainly aims to find out the trend of the vaccinations around the world for the prevention of the Covid 19 pandemic and how much has been achieved so far.**

****Methodology –****

****Step 1 –****

****Data Importing –****

**In power BI desktop with the help of the get data option import the CSV data which is named as country\_vaccinations and clicked load option**

****Step 2 –****

****Data Cleaning –****

**After loading the data and after analyzing the data | understood that there are 86512 rows and 15 columns. And in that some of the columns contained null values I have replaced the null values by 0 with the use of replace functions and started working on the data.**

****Step 3 –****

****Visualizations –****

**In visualization part with the help of power BI desktop software I have used different kinds of charts, graphs, cards and table to display the data in the format which will be easy to understand.**

# COVID-19 Vaccine Management Technology Solution

**The global COVID-19 pandemic presents the largest immunization and vaccination challenge our nation has ever faced. The final step—administering the vaccine to millions of people in an equitable and timely manner—is the responsibility of state and local governments. This enormous undertaking requires speed, accuracy, and precision—and a user - centric management solution that simplifies record-keeping, provides real-time data access, and coordinates recipient and vaccine inventory information.**

****Vaccine Management Solution Features****

**Guide house COVID-19 specialists have developed a vaccination management technology solution that addresses these key challenges. Our integrated vaccine management technology solution offers the following features critical to your COVID-19 vaccination efforts:**

* **Appointment scheduling (including booster vaccinations)**
* **Clinical vaccine administration**
* **Outcomes monitoring**
* **Vaccination site management**
* **Inventory management**
* **Communications, notifications, and public outreach**

**Our solution prioritizes the needs and concerns of each type of user who will interact with the system, addressing their most pressing questions:**

**The Vaccine Recipient**

* **When can I get vaccinated?**
* **What appointments are available locally for the first shot and booster shot?**
* **What do I do if I have side effects?**
* **I only have a smartphone; can I still make an appointment?**
* **Can I speak to a live person to get more information and help setting up an appointment?**

**The Vaccine Administrator**

* **Is this the right vaccine recipient and do they meet eligibility assessment requirements?**
* **I have a lot of paperwork and forms to complete; can the system pre populate this information?**
* **Different vaccines have different protocols; can the system match the protocols to the vaccine and recipient?**
* **I need to be sure that I’m administering the right booster vaccine to the right recipient; can the system perform checks such as these?**
* **We have a lot of vaccination stations and the space is cramped—sometimes just a chair and a small table. Can I use a tablet instead of a laptop?**

**The Vaccine Site Manager**

* **Will I have enough vaccines at each site, based on appointments? I can only accept appointments if I have inventory.**
* **If I need to move inventory, will the system tell me, or do I need to figure it out on my own? From time to time, I need to reallocate inventory between vaccination sites.**
* **Do we have any side-effect hot spots?**
* **Can I manage my inventory in one place or do I need to access several systems?**

**The State or Local Government Official**

* **Will we have a common communication platform or web portal where we can direct our citizens?**
* **Will the platform have the flexibility to provide various types of information to the general public, registered recipients, administrators, and managers?**
* **Will we have the reports we need on vaccine usage, inventory levels, and reported side effects? Is upload input to national vaccine registries available?**

**The cloud-based technology solution takes a user-centric approach to these questions:**

* **Relevant information and required actions are presented when a user logs in.**
* **Mobile-enabled means less typing and more selecting.**
* **QR codes and bar codes speed up the vaccination administration check-in process, minimize forms, and limit close contact.**
* **Step-by-step protocols guide the vaccine administrator—ensuring proper procedure.**
* **Dynamic dashboards provide insight and actions, while taking the worry out of information overload and misunderstandings.**

**DATASET DETAILS:**

**Dataset link:**

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

**ABOUT DATASETS:**

### **Context**

**Data is collected daily from [Our World in Data](https://ourworldindata.org/" \t "https://www.kaggle.com/datasets/gpreda/_blank) GitHub repository for [covid-19](https://github.com/owid/covid-19-data" \t "https://www.kaggle.com/datasets/gpreda/_blank), merged and uploaded. Country level vaccination data is gathered and assembled in one single file. Then, this data file is merged with locations data file to include vaccination sources information. A second file, with manufacturers information, is included**.

### **Content**

**The data (country vaccinations) contains the following information:**

* ****Country**- this is the country for which the vaccination information is provided;**
* ****Country ISO Code** - ISO code for the country;**
* ****Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;**
* ****Total number of vaccinations** - this is the absolute number of total immunizations in the country;**
* ****Total number of people vaccinated** - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccination might be larger than the number of people;**
* ****Total number of people fully vaccinated** - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme;**
* ****Daily vaccinations (raw)** - for a certain data entry, the number of vaccination for that date/country;**
* ****Daily vaccinations** - for a certain data entry, the number of vaccination for that date/country;**
* ****Total vaccinations per hundred** - ratio (in percent) between vaccination number and total population up to the date in the country;**
* ****Total number of people vaccinated per hundred** - ratio (in percent) between population immunized and total population up to the date in the country;**
* ****Total number of people fully vaccinated per hundred** - ratio (in percent) between population fully immunized and total population up to the date in the country;**
* ****Number of vaccinations per day** - number of daily vaccination for that day and country;**
* ****Daily vaccinations per million** - ratio (in ppm) between vaccination number and total population for the current date in the country;**
* ****Vaccines used in the country** - total number of vaccines used in the country (up to date);**
* ****Source name** - source of the information (national authority, international organizat**ion, local organization etc.);
* ****Source website**** - website of the source of information;

There is a second file added recently (country vaccinations by manufacturer), with the following columns:

* ****Location**** - country;
* ****Date**** - date;
* ****Vaccine**** - vaccine type;
* ****Total number of vaccinations**** - total number of vaccinations / current time and vaccine type.

### **Acknowledgements**

**I would like to specify that I am only making available **Our World in Data** collected data about vaccinations to Kagglers. My contribution is very small, just daily collection, merge and upload of the updated version, as maintained by **Our World in Data** in their GitHub repository.**

### **Inspiration**

**Track COVID-19 vaccination in the World, answer instantly to your questions:**

* **Which country is using what vaccine?**
* **In which country the vaccination programme is more advanced?**
* **Where are vaccinated more people per day? But in terms of percent from entire population ?**

**Combine this dataset with [COVID-19 World Testing Progress](https://www.kaggle.com/gpreda/covid19-world-testing-progress" \t "https://www.kaggle.com/datasets/gpreda/_blank) and [COVID-19 Variants Worldwide Evolution](https://www.kaggle.com/gpreda/covid19-variants" \t "https://www.kaggle.com/datasets/gpreda/_blank) to get more insights on the dynamics of the pandemics, as reflected in the interdependence of amount of testing performed, results of sequencing and vaccination campaigns.**

**COLUMN DETAILS:**

**15 columns are:**

1. **Country**
2. **iso\_code**
3. **Date**
4. **Total\_vaccir**
5. **people\_vac**
6. **People\_full**
7. **Daily\_vaccir**
8. **Daily\_vaccir**
9. **Total\_vaccir**
10. **People\_vac**
11. **People\_full**
12. **Daily\_vaccir**
13. **Vaccines**
14. **Source\_nan**

**15. Source\_website**

**IMPORT LIBRARIES**

**For analyzing data, we need some libraries. In this section, we are importing all the required libraries like pandas, NumPy, matplotlib, plotly, seaborn, and word cloud that are required for data analysis.**

**NUMPY:**

**Numpy (Numerical Python) is an open source Python library that's used in almost every field of science and engineering. It's the universal standard for working with numerical data in Python, and it's at the core of the scientific Python and Data ecosystems.**

**MATPLOTLIB:**

**Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update. Customize visual style and layout.**

**PLOTLY:**

**The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.**

**SEABORN:**

**Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.**

**TRAIN AND TEST:**

**import seaborn as** **sns**

**import** **matplotlib**

**import matplotlib.pyplot as** **plt**

**%matplotlib inline**

**sns.set\_style('darkgrid')**

**matplotlib.rcParams['font.size'] = 14**

**matplotlib.rcParams['figure.figsize'] = (9, 5)**

**matplotlib.rcParams['figure.facecolor'] = '#00000000'**

**REST OF EXPLANATION:**

**We implemented a two-dose vaccination campaign achieving 40% coverage of the entire population within 284 days. We assumed that 70% was the maximum achievable coverage in any age group, with an age-dependent distribution similar to seasonal influenza vaccination in the US [[40](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R40)]. Vaccines were prioritized to the following groups sequentially: (i) healthcare workers (5% of the total population [[41](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R41)]), adults with comorbidities, and those aged 65 and older (i.e., protection cohort); and (ii) all other individuals aged 18–64 (i.e., disruption minimization cohort) [[42](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R42)]. Comorbidities included cardiovascular disease, diabetes, asthma, chronic obstructive pulmonary disease, hypertension, and cancer [[20](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R20)]. Pre-existing immunity or contemporaneous infection with COVID-19 was not a factor in vaccine allocation. The age-specific coverage resulting from this prioritization was 48% of those aged 18–49, 48% of those aged 50–64, and 70% of those aged 65+. We specified a roll-out strategy in which 30 individuals per 10,000 population would be vaccinated per day, corresponding to 6.93 million vaccine doses per week for the entire US population, for approximately 41 weeks. Vaccination occurred during this time period to reach 40% coverage and outcomes were evaluated for 300 days. Infection dynamics continued during the simulations for susceptible and vaccinated individuals.**

**We included a 21-day interval between the first and second vaccine doses [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R6)]. The vaccine efficacy (Ve) against symptomatic and severe disease was assumed to be 52%, 14 days after the first dose, and 95%, 1 week after the second dose [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709178/" \l "R6)]. In the absence of data for vaccine efficacy against infection or transmission, we assumed that the vaccine protection against infection was 50% lower than its efficacy against disease (base-case), with additional scenarios of (i) 0%; and (ii) the same efficacy against disease, after each dose of the vaccine. We further simulated the model for these scenarios with a 28-day interval between the two doses.**

**METRICS USED FOR ACCURACY CHECK:**

**Kaggle datasets [[37](https://link.springer.com/article/10.1007/s12553-022-00712-4" \l "ref-CR37" \o "Kaggle. COVID-19 World Vaccination Progress. Daily and Total Vaccination for COVID-19 in the World from Our World in Data. 2022. Available at                   https://www.kaggle.com/gpreda/covid-world-vaccination-progress                                  . Accessed 5 Jan 2022.)] were utilized to evaluate the prediction outcomes of the daily COVID-19 vaccination to decrease pandemic risk. The dataset consists of all the countries that have been vaccinated, fully vaccinated with COVID-19, the types of vaccines used and the date vaccinated. Training datasets and testing datasets are the two types of data sets utilized in this study. COVID-19 daily data from the 13th of December 2020 to the 13th of June, 2021 make up the training dataset, whereas COVID-19 daily data from the 14th of June, 2021 to the 14th of October, 2021 make up the testing dataset. Table [2](https://link.springer.com/article/10.1007/s12553-022-00712-4" \l "Tab2) displayed the attributes of COVID-19 vaccination dataset which consists of country, iso\_code, date, total\_vaccinations, people\_vaccinated, people\_fullt\_vaccinated, daily\_vaccination\_raw, daily\_vaccinations, total\_vaccinations\_per\_hundred, people\_vaccinated\_hundred, people\_fully\_vaccinated\_per\_hundred, daily\_vaccinations\_per\_million and vaccines.**