**COVID-19 DATA ANALYSIS**



**ABOUT**

There was a time when Covid-19 got out of hand. Even after the lockdown, this still resulted in a rapid increase in cases as in some countries cases were brought under control but the economy was sacrificed. In such a situation, only vaccines are seen as the only tool that can help the world fight covid-19. Many vaccines have been introduced so far to fight covid-19. No vaccine has guaranteed 100% accuracy so far, but most manufacturing companies claim their vaccine is not 100% accurate, but still, it will save your life by giving you immunity. Thus, each country tries to vaccinate a large part of its population so as not to depend on a single vaccine. That is what I’m going to analyse in this article, which is how many vaccines each country is using to fight covid-19. In the section below, I will take you through a data science tutorial on Covid-19 vaccines analysis with Python.

**DATASET**

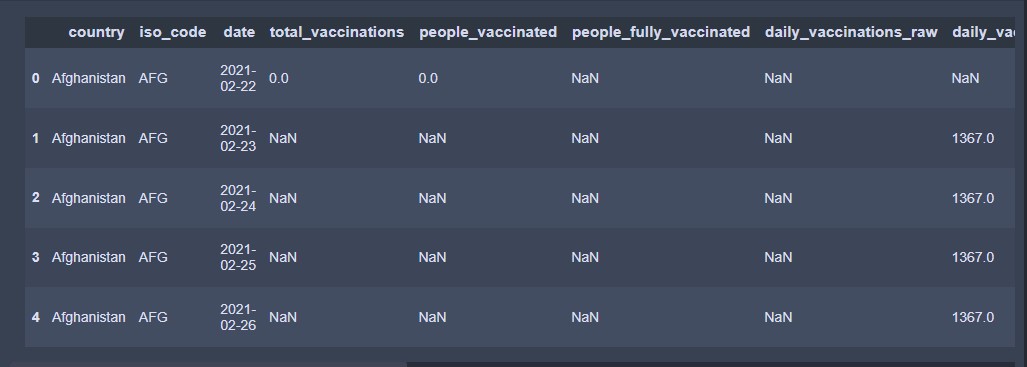
Data is collected daily from [**Our World in Data**](https://ourworldindata.org/) GitHub repository for [covid-19](https://github.com/owid/covid-19-data), merged and uploaded. Country level vaccination data is gathered and assembled in one single file. Then, this data file is merged with locations data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

The data (country vaccinations) contains the following information:

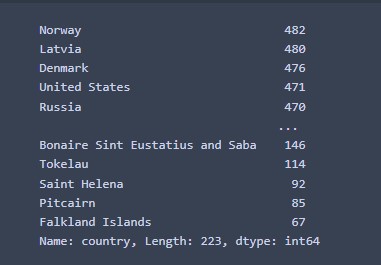
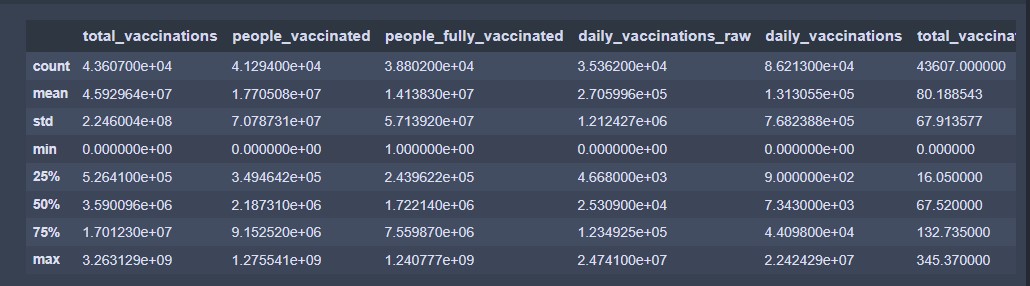
* **Country**- this is the country for which the vaccination information is provided;
* **Country ISO Code** - ISO code for the country;
* **Date** - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total;
* **Total number of vaccinations** - 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 organization, local organization etc.);
* **Source website** - website of the source of information;

**IMPLEMENTATION**

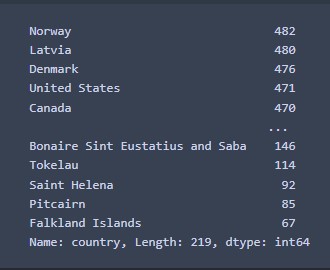
1. The dataset that I will be using here for the task of covid-19 vaccines analysis is taken from Kaggle.



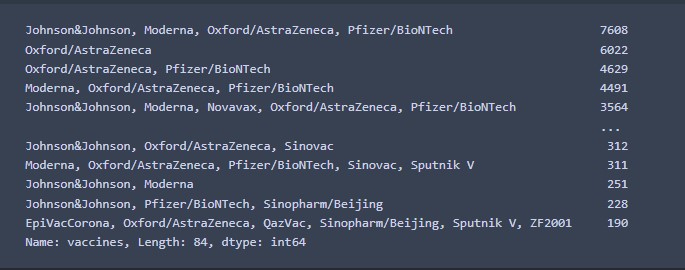
1. Now let us explore this data before we start analysing the vaccines taken by countries. The United Kingdom is made up of England, Scotland, Wales, and Northern Ireland.



1. But in the above data, these countries are mentioned separately with the same values as in the United Kingdom. So, this may be an error while recording this data. So, let us see how we can fix this error.



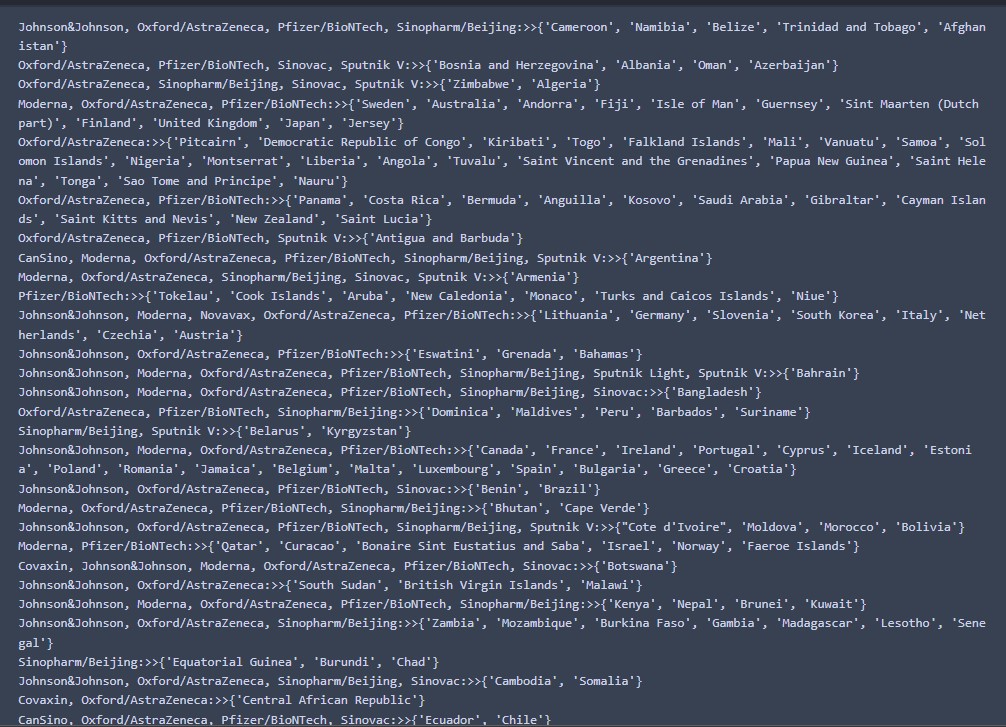
1. Now let us explore the vaccines available in this dataset.



1. So, we have almost all the Covid-19 vaccines available in this dataset. Now I will create a new Data Frame by only selecting the vaccine and the country columns to explore which vaccine is taken by which country.



1. Now let us see how many countries are taking each of the vaccines mentioned in this data.



1. Now let us visualize this data to have a look at what combination of vaccines every country is using.



