

Contents

1	Overview	2
2	Description	2
3	Preprocessing	3
3.1	NaN values	3
3.2	Date column	3
3.3	Vaccines column	3
4	Analysis of the data	3
4.1	Top 5 countries with respect to total vaccinations(Latest)	3
4.2	Progress of Vaccinations	5
4.3	US total vaccinations forecast	6
4.4	Heatmap of people fully vaccinated worldwide	8
4.5	Vaccines used by countries	8
5	Tools/Libraries used	9

Worldwide Covid-19 Vaccination Progress

Aja Sharma

April 2021

1 Overview

This project pertains to the analysis of vaccination efforts worldwide for Covid-19. The dataset was retrieved from <https://www.kaggle.com/gpreda/covid-world-vaccination-progress>. The data downloaded is a zip file. The CSV in the zip file with the name “country_vaccinations.csv” has been extracted to the parent directory. The data is collected and parsed from <https://github.com/owid/covid-19-data> on a regular basis. This analysis is performed on data collected on **11-04-2021**. The original sources of the information are also recorded in the dataset.

Analysis of this information should provide useful insights into how the world is recovering from the pandemic and the large scale vaccine programs involved.

2 Description

The data records the daily vaccination numbers of many countries. The scope of observation for individual vaccine usage is limited due to lack of data for countries utilising multiple types of vaccines.

The data contains the following information(1):

- 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 vaccinations 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.

3 Preprocessing

3.1 NaN values

Observing the dataset, we find that the columns with NA values are numeric. They are updated periodically and they are in increasing order with time for each country. We can interpolate the values by populating them with the data of the previous row since we are only provided with historical data for these columns. These values cannot be lower than the previous values recorded country-wise. The pandas forward fill function was used to populate these values. The initial rows may also contain NA values for which we have no historical data. We can replace these with zeroes.

3.2 Date column

The date column was read in as a string object. This column has been converted to datetime values using pandas for further processing.

3.3 Vaccines column

On further observation of the data, the vaccines column contains multiple comma-separated values in each row for countries that utilize multiple types of vaccines. Since there is a lack of data on the individual use of each vaccine country-wise, we can resort to storing the data as an array of values in the same column for further analysis.

4 Analysis of the data

4.1 Top 5 countries with respect to total vaccinations(Latest)

For this analysis, only a small subset of the data was chosen since there are numerous countries. The subset chosen includes the most recent estimates for countries at the top 5 on the list in terms of total

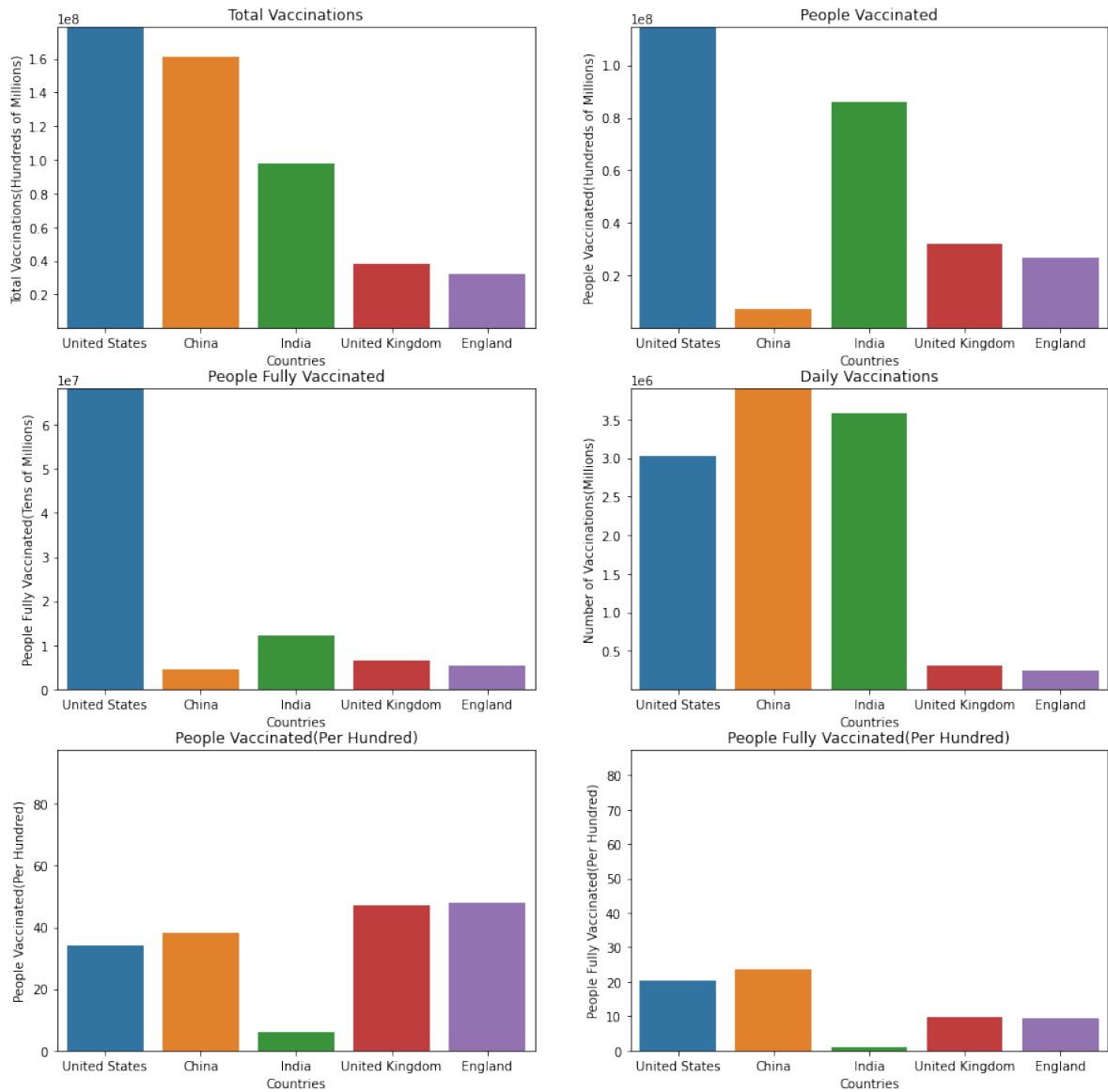


Figure 1: Analysis of the Top 5 Countries in terms of total vaccinations

vaccinations. These are the **United States, China, India, United Kingdom and England**. The analysis considers the latest numbers for these countries obtained on the most recent date. The plots in Figure 1 measure the total numbers for various metrics for these countries.

- **United States** - The US leads in total vaccination numbers at nearly 180 million, with people vaccinated at close to 120 million, people fully vaccinated close to 70 million. It's third in the daily vaccination numbers with about 300,000 doses a day. With respect to the proportion of the population, the US has vaccinated close to 30 people per hundred and fully vaccinated

about 20 people per hundred, second only to China.

- **China** - China leads in terms of the daily vaccinations with close to 4 million doses a day. The number of people vaccinated and fully vaccinated remain low in China even though it is only second in terms of total vaccinations with over 160 million doses. The number of people fully vaccinated per hundred in China has the highest proportion compared to the other countries with 30 people fully vaccinated per 100.
- **India** - India is third in terms of total vaccinations with close to 100 million doses. The number of people vaccinated is at 80 million with less than 10 million fully vaccinated. The daily vaccination numbers are second to China's at approx 3.5 million doses a day. People vaccinated and fully vaccinated per hundred remain low for India indicating that a majority of the population is yet to be vaccinated.
- **United Kingdom** - The United Kingdom leads in people vaccinated at about 40 people vaccinated per 100. Total vaccinations and the number of people vaccinated are lower but this maybe due to the lower population of the United Kingdom. Compared to the US and China, there appears to be a lower proportion of people fully vaccinated in the United Kingdom at 10 people per hundred.
- **England** - The numbers for England are close to those of the United Kingdom. England seems to have a slightly higher proportion of people vaccinated compared to the United Kingdom.

4.2 Progress of Vaccinations

Figure 2 looks at how the vaccination efforts have progressed over time for the countries at the top 10 on the list in terms of total vaccinations. The vaccination programs start mid-December with the sharper slopes for **US, China and India** over time. This indicates faster progress in the vaccination efforts compared to other nations. Since the data represents a time series of total vaccinations taken on a daily basis, we can consider fitting a polynomial regression model to make forecasts on the vaccination progress.

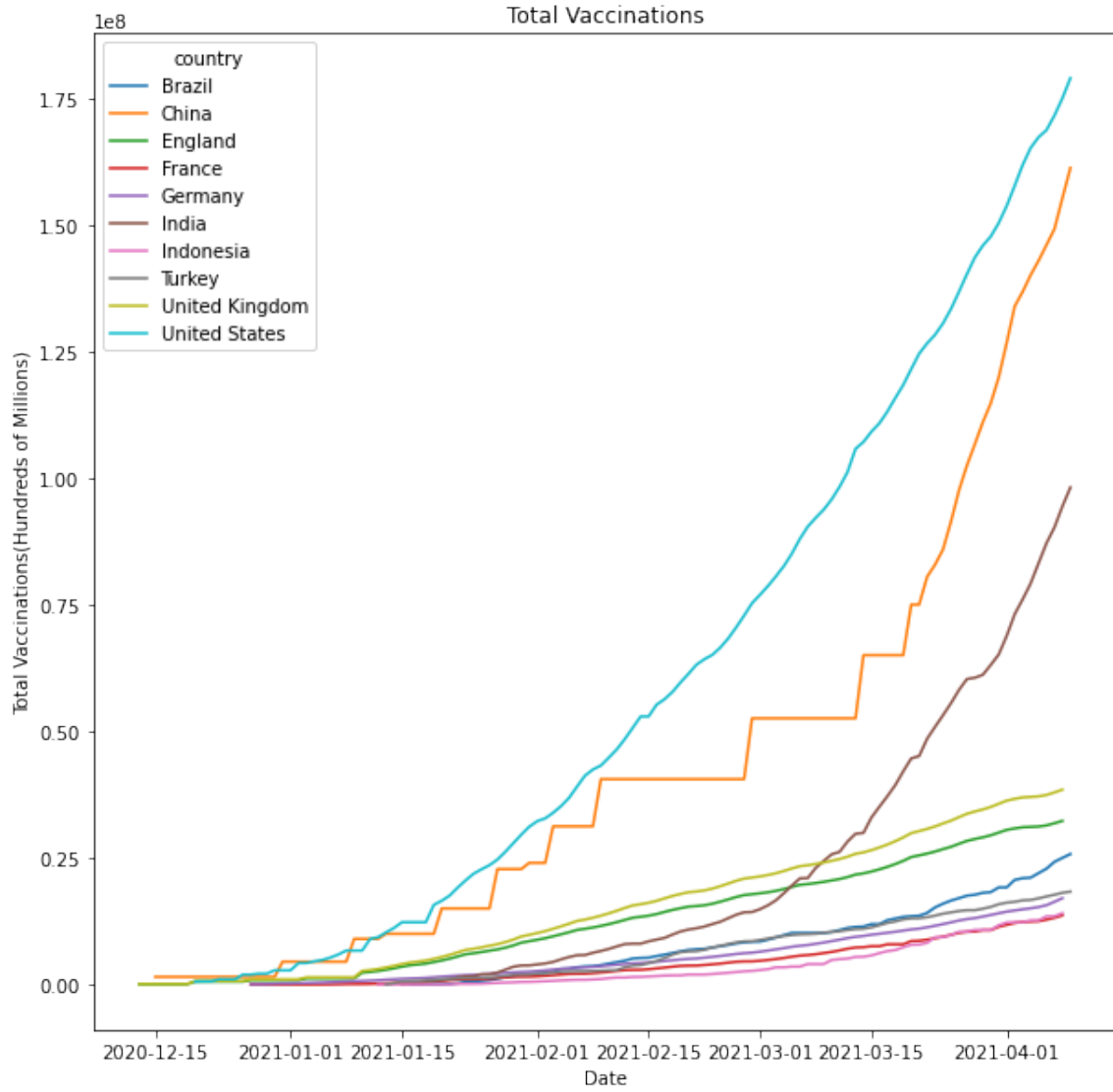


Figure 2: Total Vaccinations of countries(Top 10)

4.3 US total vaccinations forecast

A **polynomial regression** model of two degrees is fitted on the total vaccinations curve of the United States. This is depicted in Figure 3. The regression line fits the data points well. We can further analyse the residuals to check if the model is a good fit for the data.

Analysing the residuals of the fit in Figure 5, the histogram of the residuals appears normal. The QQ plot for the residuals approximately fits a normal distribution. The box plot displays no outliers. This model appears to be a good fit for the data. It has been used to further forecast total vaccinations numbers for the next 7 days in Figure 3. It appears that the United States will be able

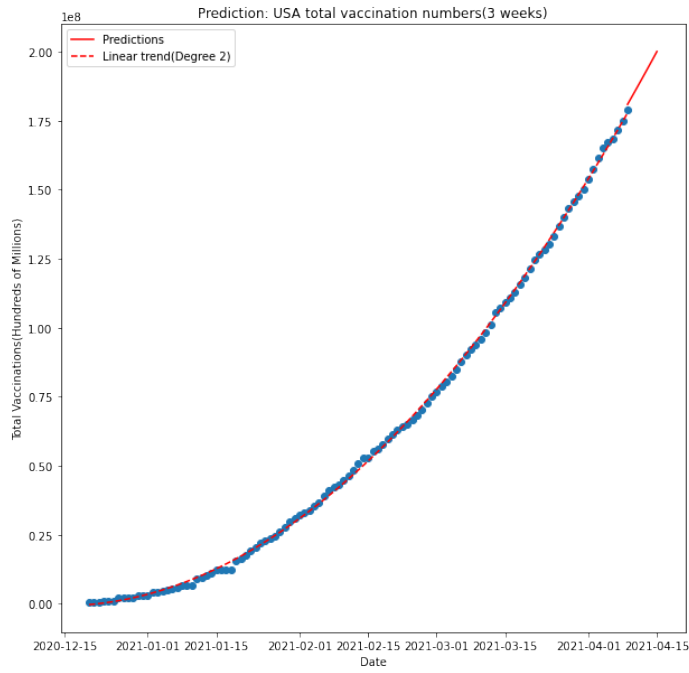


Figure 3: Model fit and forecasting: USA total vaccinations

date	total_vaccinations
2021-04-09	180942002.58
2021-04-10	184068421.21
2021-04-11	187221507.24
2021-04-12	190401260.67
2021-04-13	193607681.51
2021-04-14	196840769.75
2021-04-15	200100525.40

Figure 4: Forecast:
USA total vaccinations

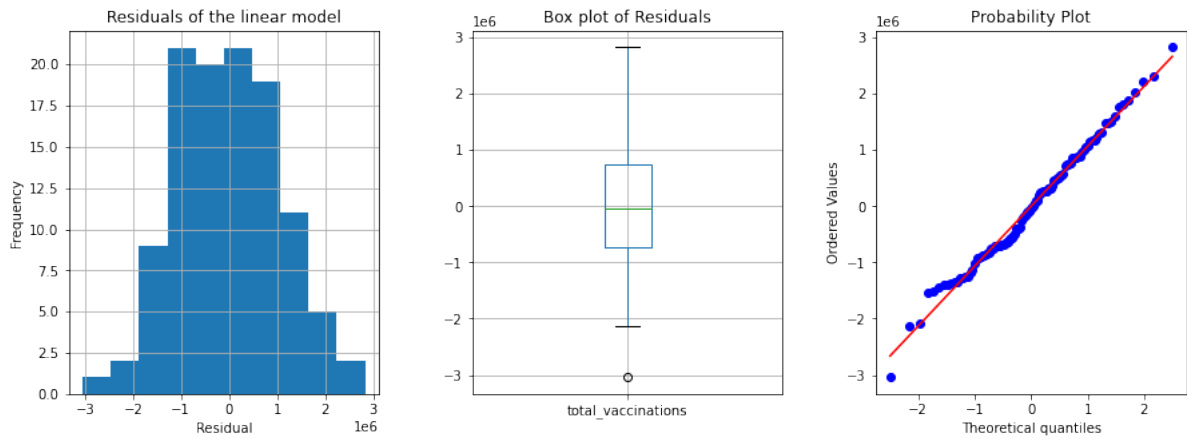


Figure 5: Analysis of Residuals

to administer over a total of 200 million doses next week.

4.4 Heatmap of people fully vaccinated worldwide

People fully vaccinated worldwide(Per hundred)

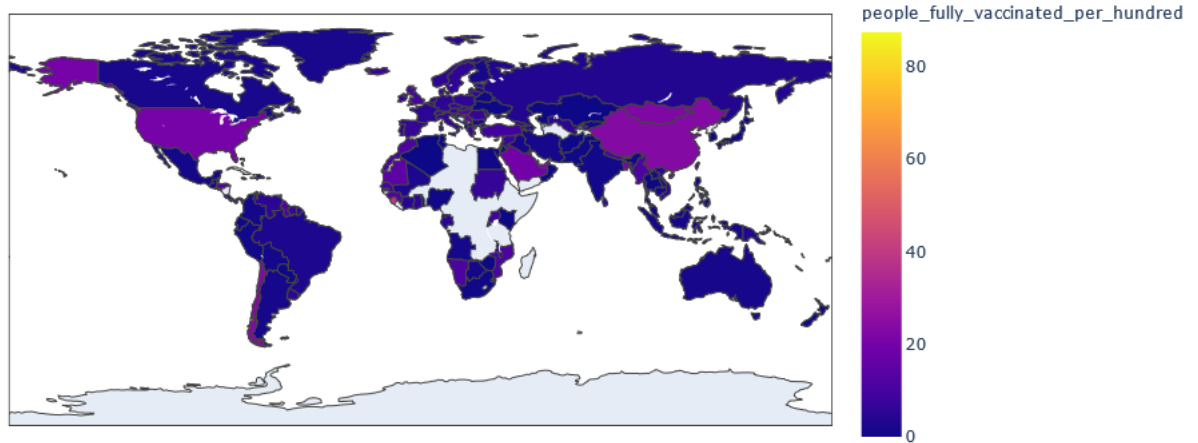


Figure 6: Heatmap showing people fully vaccinated worldwide

Figure 6 displays a heat map of the number of people fully vaccinated (per hundred). This should give us a clear picture on how many people are immunized in proportion to the population. The plotly library was used to generate the **Choropleth map**. Upon closer inspection, we find smaller nations like **Israel and Sierra Leone** have vaccinated higher proportions of the population compared to the world. Countries like the **USA, Saudi Arabia, Mongolia and China** have fully vaccinated approximately 30% of their populations. On the other hand, most of the world remains in the blue region with little to no data about many nations in the Africa. This indicates that a large proportion of the world's population is yet to be vaccinated.

4.5 Vaccines used by countries

Figure 6 shows the vaccines used by countries worldwide. We can use this to check which vaccine types are preferred. The histogram plots various vaccine types and their corresponding usage in terms of the number of countries. As of the most recent date on record for each country, 116 countries are using **Oxford/Astrazeneca** for vaccination and 83 countries are using **Pfizer/BioNTech**. **Moderna, Sinopharm/Beijing, Sinovac and Sputnik V** also see widespread use in many countries with **Covaxin, EpiVacCorona, Johnson&Johnson and Sinopharm/Wuhan** being used in 1 or 2 countries.

Vaccines used worldwide(Latest date)

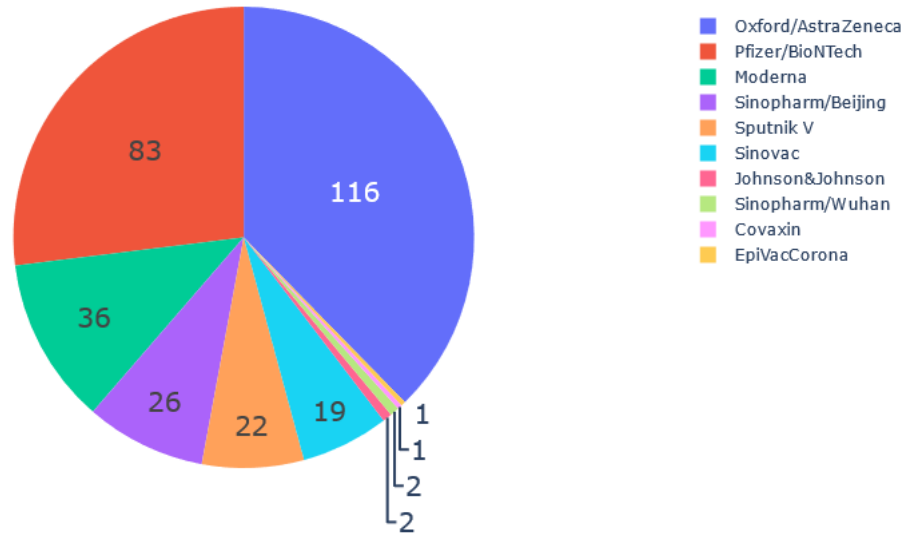


Figure 7: Vaccines used by the world

5 Tools/Libraries used

- Jupyter Notebook - For processing and visualising the data in Python
- Pandas - To load the dataset and perform operations on the dataframe
- Numpy - To fit a polynomial regression model and forecast values
- Plotly, Matplotlib, Seaborn- For plotting data
- Scipy - For analysis of the residuals from polynomial regression.

References

- [1] G. Preda, "Covid-19 world vaccination progress," 2021. [Online]. Available: <https://www.kaggle.com/gpreda/covid-world-vaccination-progress>