



Course : Exploratory Data Analysis

Class & Div : IV E

Course Project Title : Covid 19 Vaccination Analysis

Team members

Anand Sagar-564

Ateeth KJ -546

Shakti Kumar Singh- 548

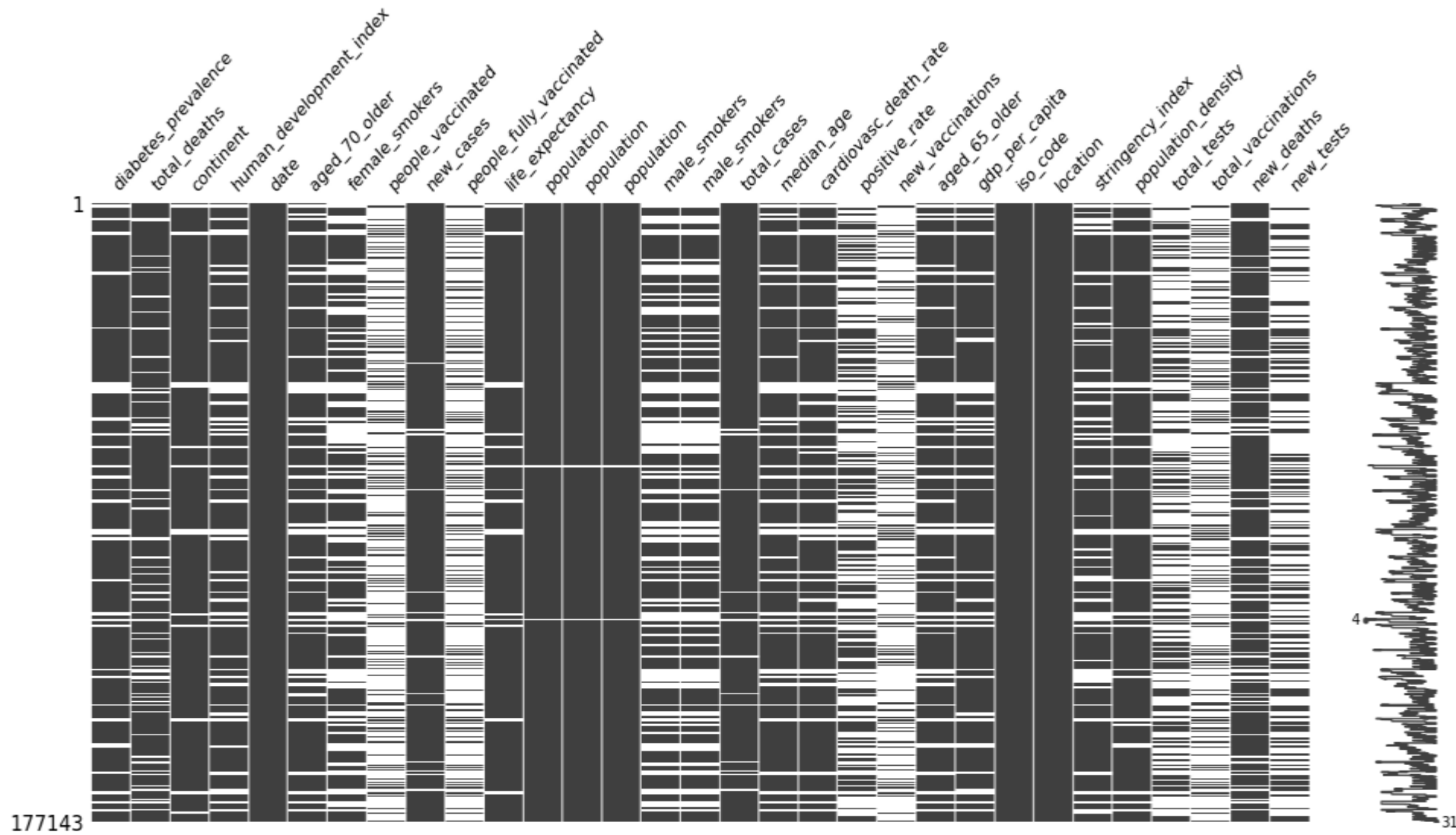
Shivam Singh-557

DATA PREPROCESSING

Important Attributes of Our World in Covid Dataset

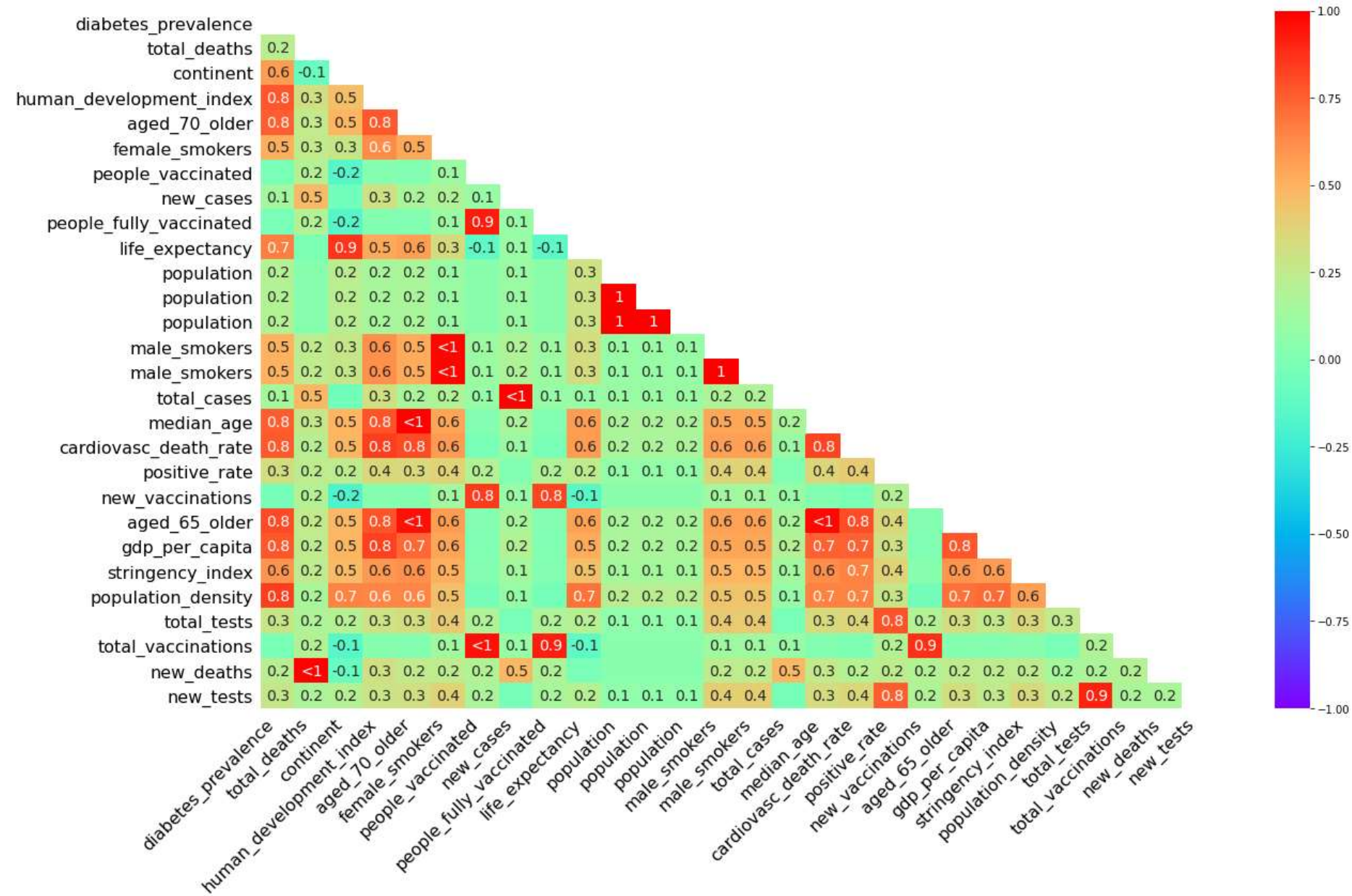
- | | |
|--|----------------------------------|
| 1.Total vaccination per hundred | 6.Population Density |
| 2.Total cases | 7.GDP per Capita |
| 3.Positive rate | 8.Human Development Index |
| 4.Total boosters | 9.ICU Patients |
| 5.Stringency Index | 10. Excess Mortality |

Heat-Map of Null values



- The above heat map tells us if the occurrence of missing values are sparsely located or whether they are located as big chunk
- For example in female_smokers we see that there are bigger chunks of missing values where as in total_cases we see that data is missing in very less amount that too very far from each other.
- In total_vaccinations we see that data is missing in larger chunks very frequently, We can also see if two columns have almost same number of missing values or not if they are at same index then we can also verify it using Heatmap-null-coorelation Matrix

Correlation heat-map of the attributes



- The above Heat-Map shows the correlation between two attributes where they both are simultaneously null at same index.
- For example if two values are null at same index for most of the times then those two attributes will have higher correlation.
- The Column which doesn't show any score in many of its columns have data Missing Completely at Random for ex : Total_Vaccination

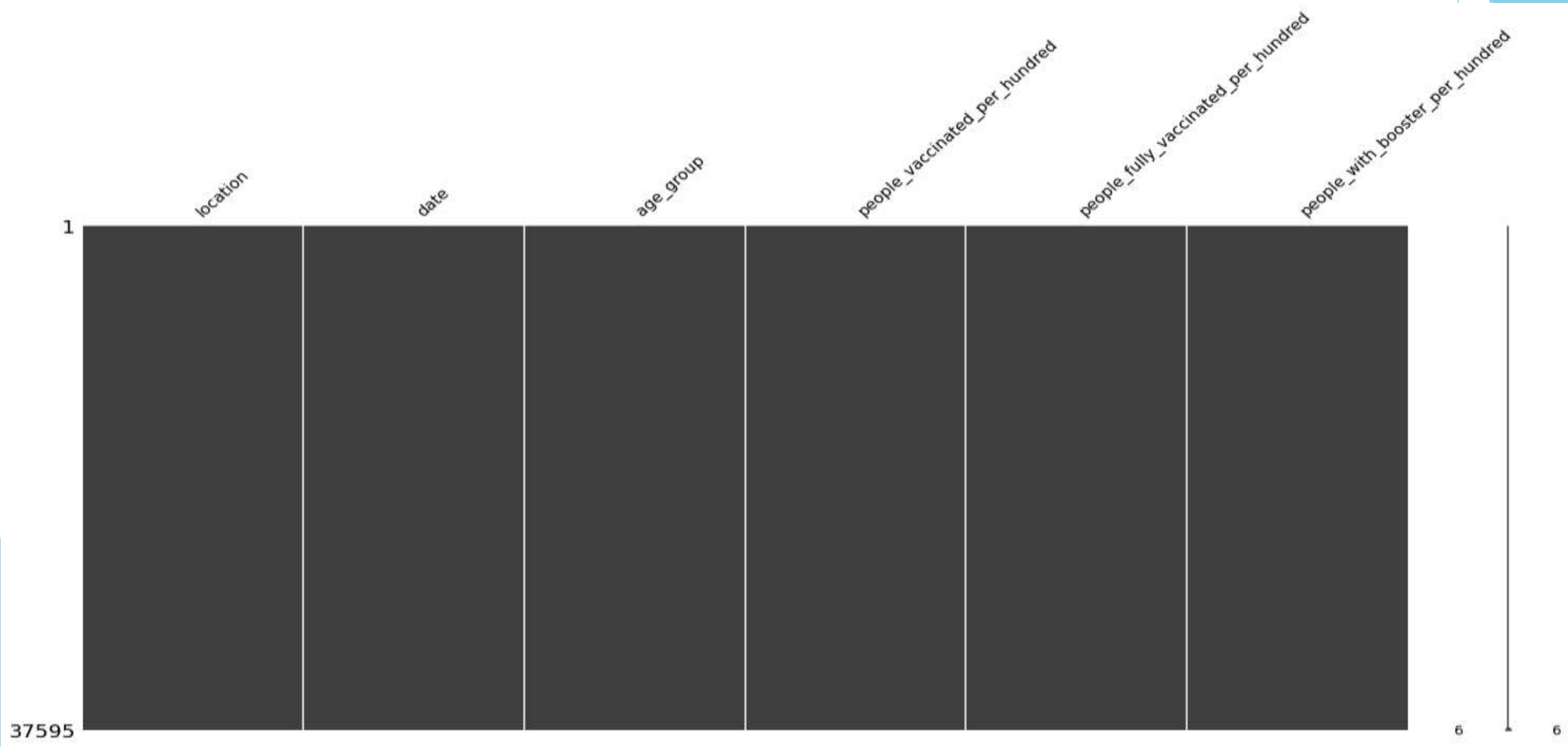
Handling Null values

1. Depending on the type of missing values we replaced it accordingly with next or previous valid data, or mean/median/mode or different values.
2. Since most of our data is given according to date in increasing order so the suitable value to fill in place of null would be next or previous valid value as it will be closest to the real or actual value.
3. Or the valid solution may be to go the same day of previous week and the next week and take an average of the two values.

Age Grouped-COVID-19 Vaccination Dataset

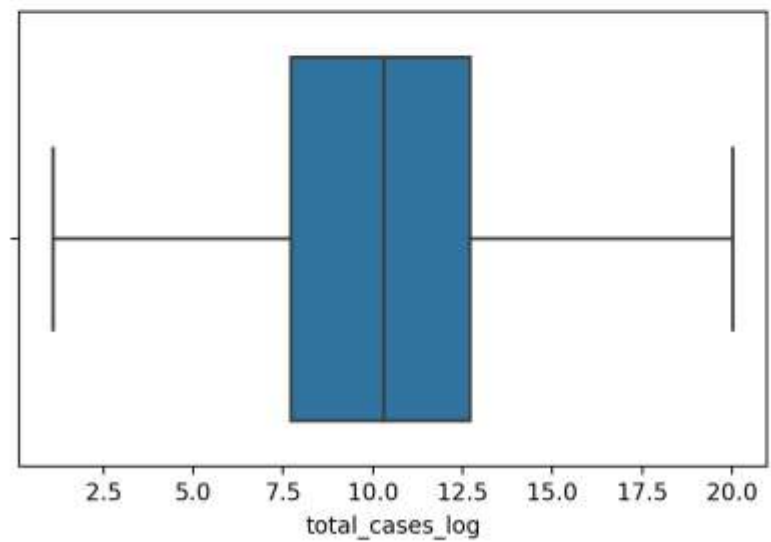
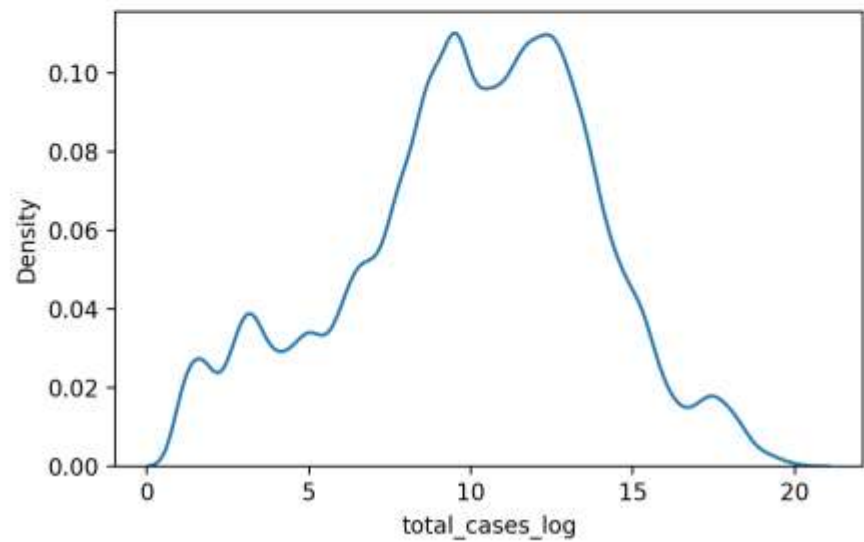
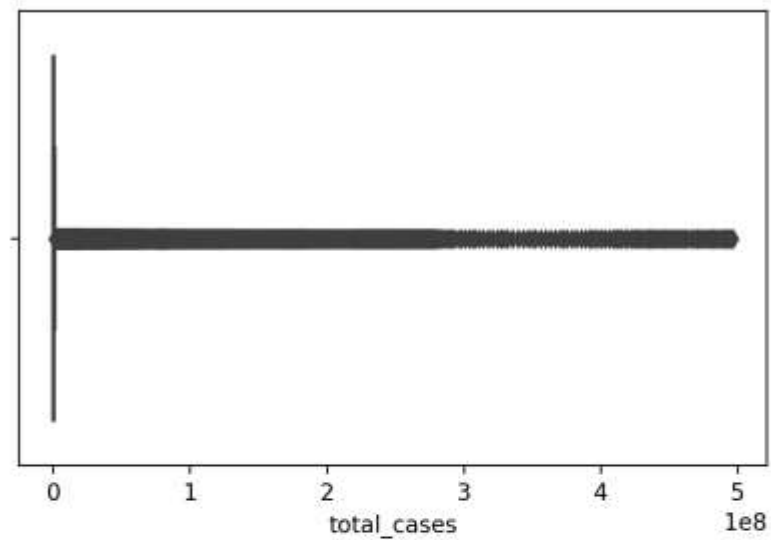
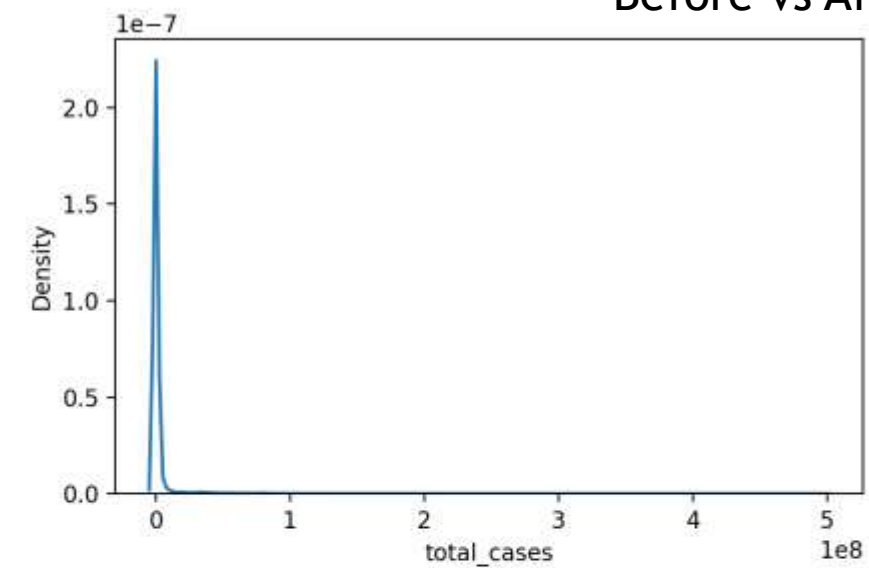
- Location
- Date
- Age-group
- People Vaccinated per hundred
- People fully vaccinated per hundred
- People with booster per hundred

Null Data Visualisation Using Heat-Map



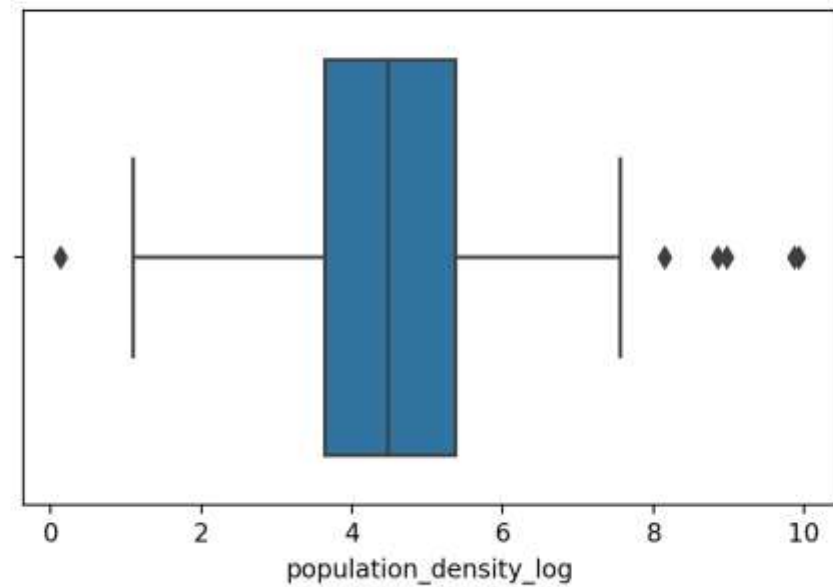
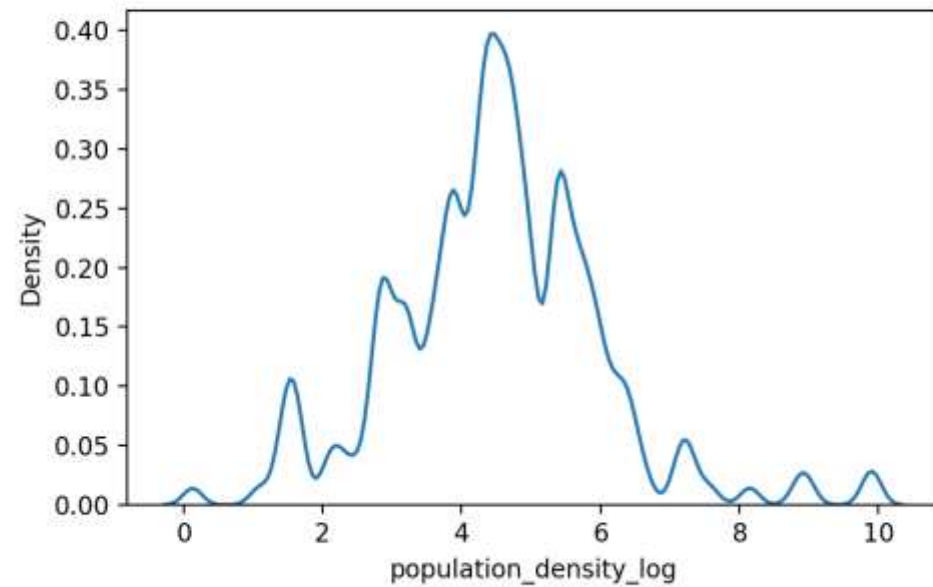
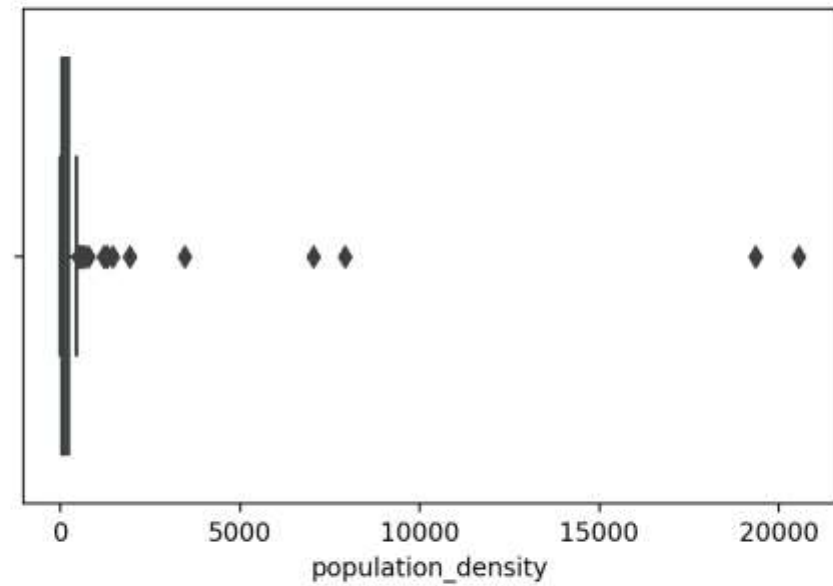
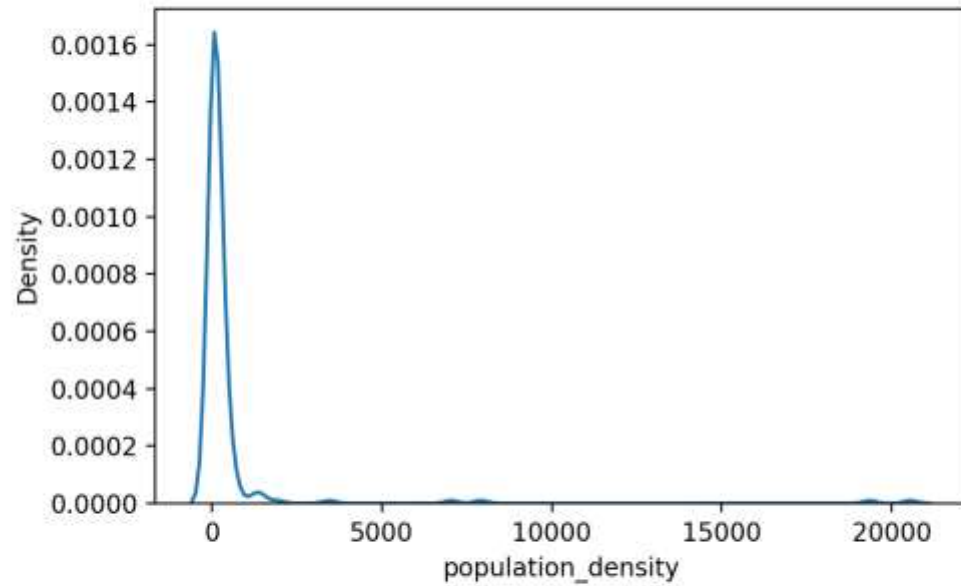
DATA TRANSFORMATION

Before vs After transformation



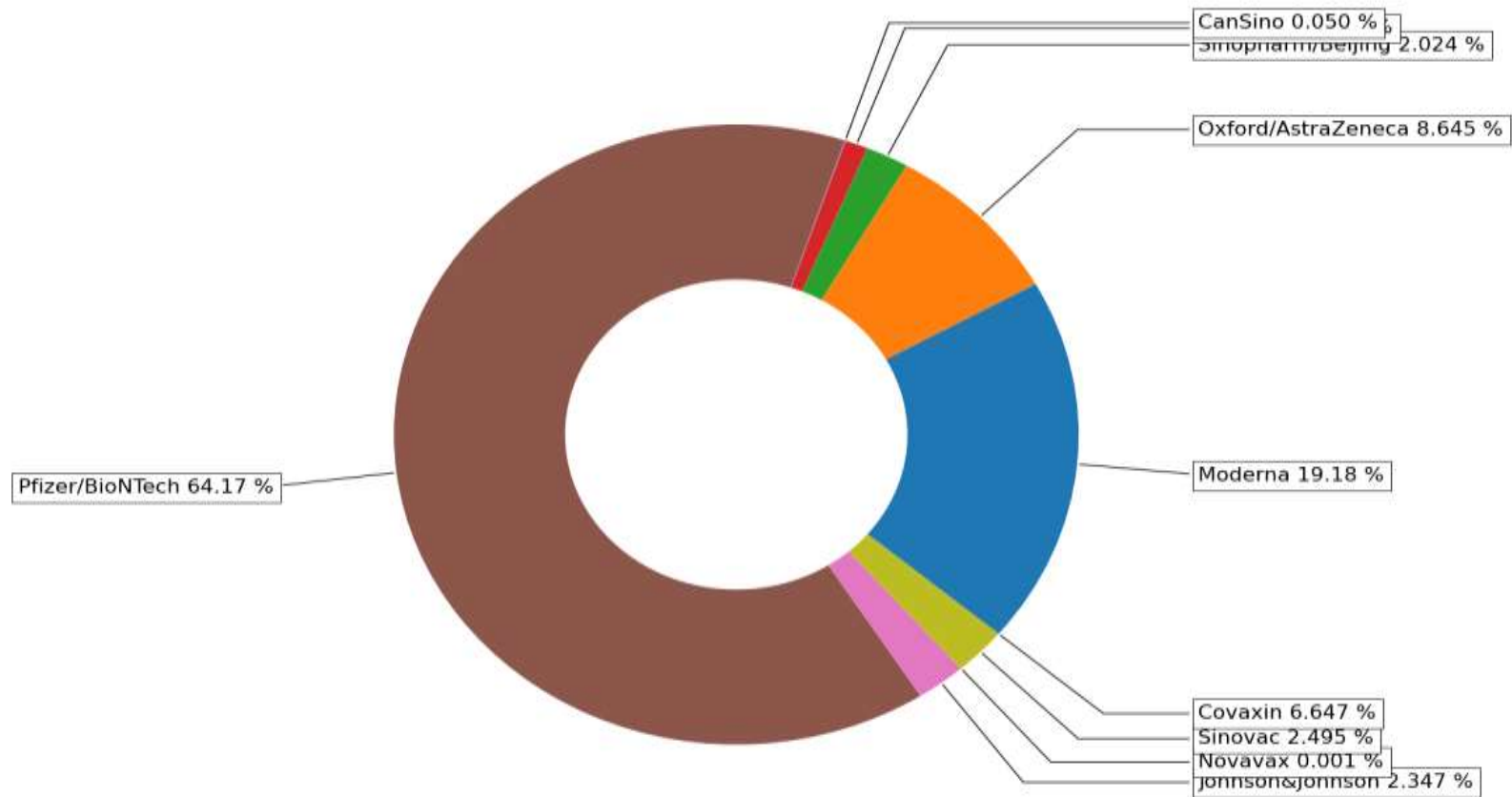
- As we can see in the above figure before, the data was highly skewed (13.50), due to which a very high % of data was segregated towards a particular value
- Due to that many values were treated as outliers, So we took log of the column after adding 1 to each value of the column to remove skewness.
- Now we can see that the data has better distribution now (skewness=-0.3)

Before vs After Transformation



Analysis

Vaccine Distribution



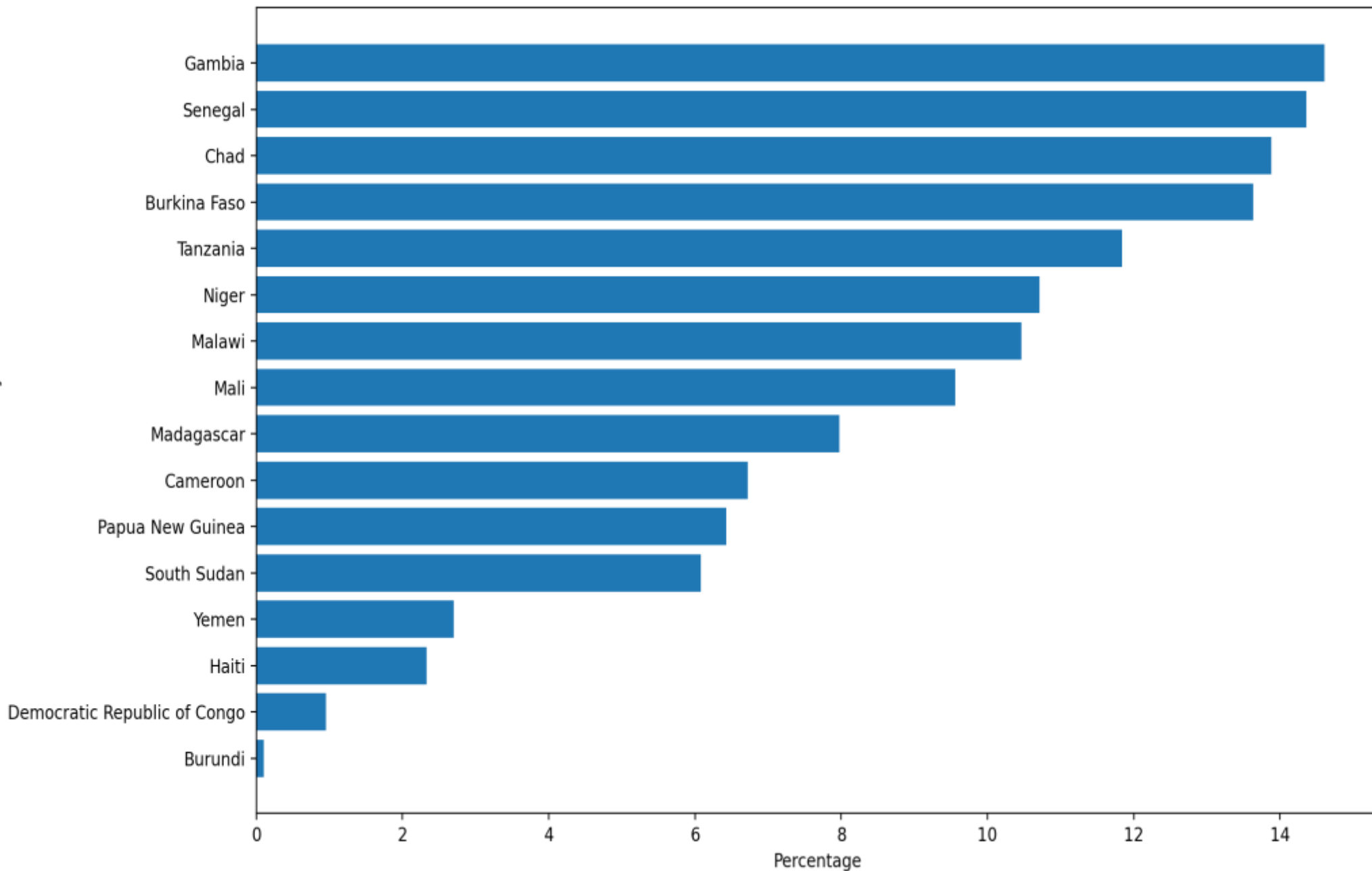
- We can clearly see from the above graph that Pfizer/BioNtech is the most distributed vaccine in the world covering 64% of total vaccinations in the world

Some of the reasons could be

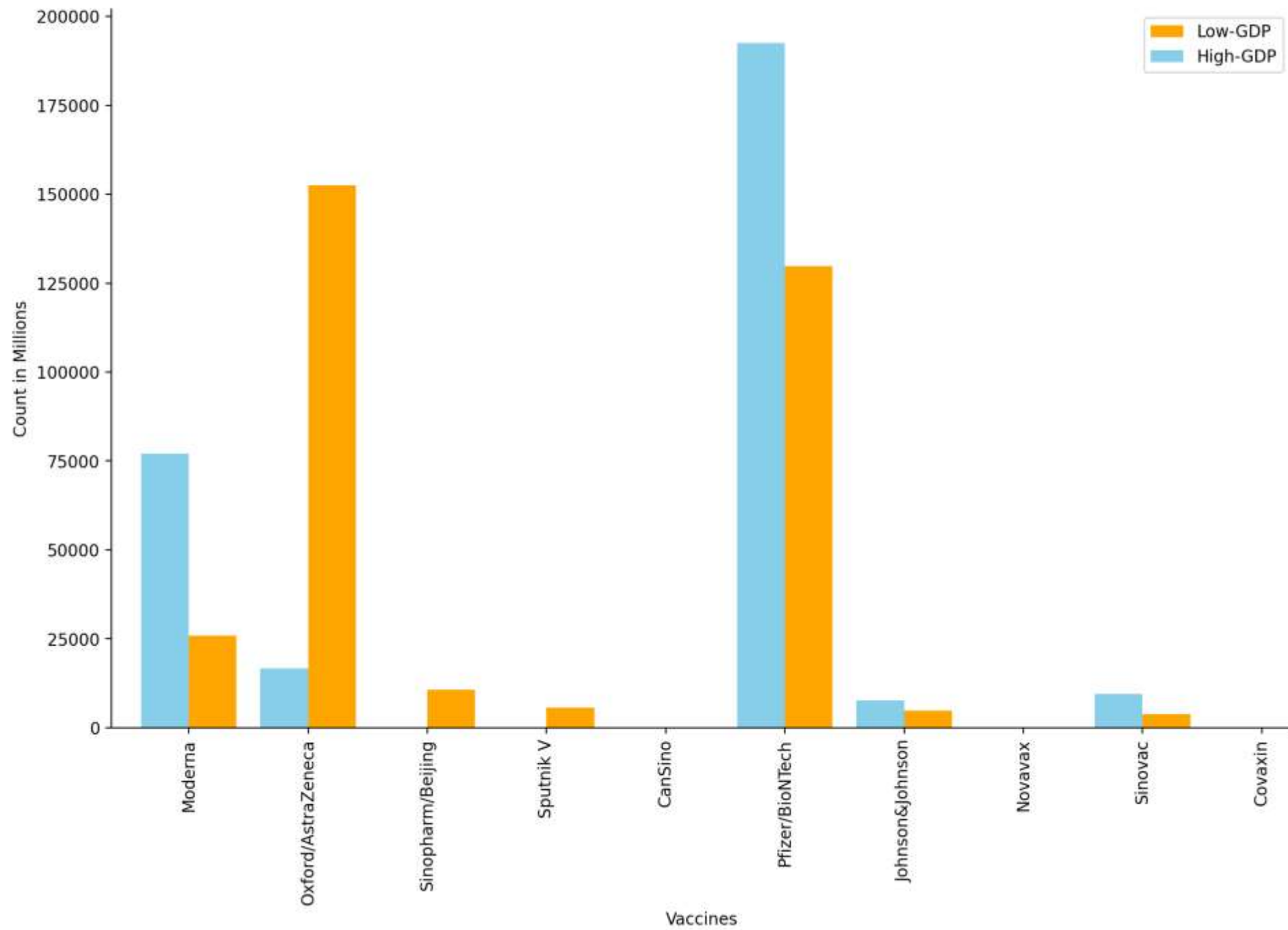
- Pfizer was the first to market a coronavirus vaccine -- and the first to gain authorization
- It has higher efficacy i.e 95%
- As it is a MRNA vaccine it doesn't have much side effects

Least vaccinated Countries

Country



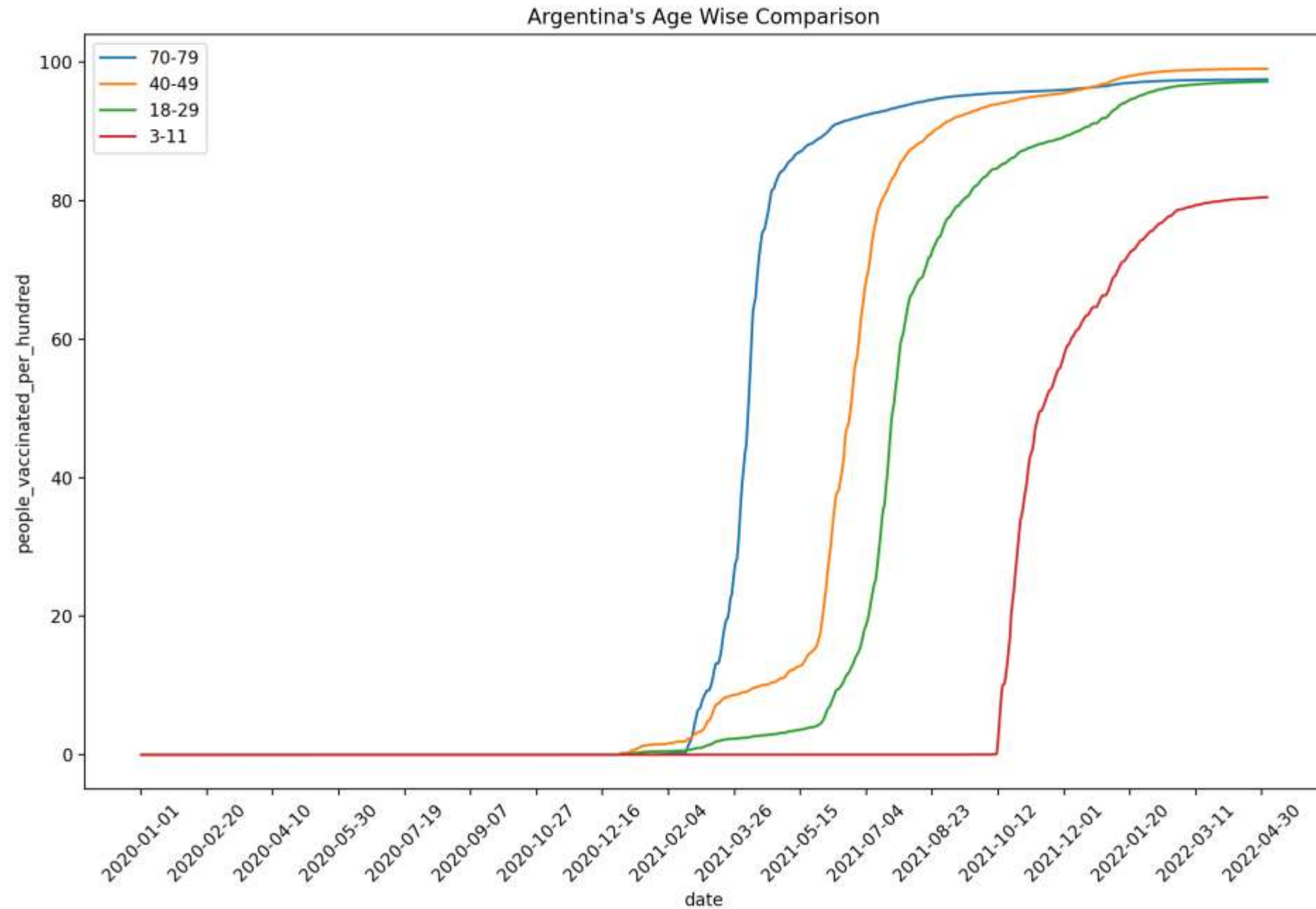
- From the above graph we can observe that out of 15 least vaccinated countries 13 are from Africa
- On analysing we found that these are low income countries with low GDP.
- The Covax and partners are providing concerted support to 34 countries that were at or below 10% coverage in January 2022 and off track to achieve the WHO's 70% global coverage target by June 2022.
- Countries are supported to access urgent operational funding, technical assistance, political engagement and demand and supply planning required to plan, implement and scale their vaccination response and monitor



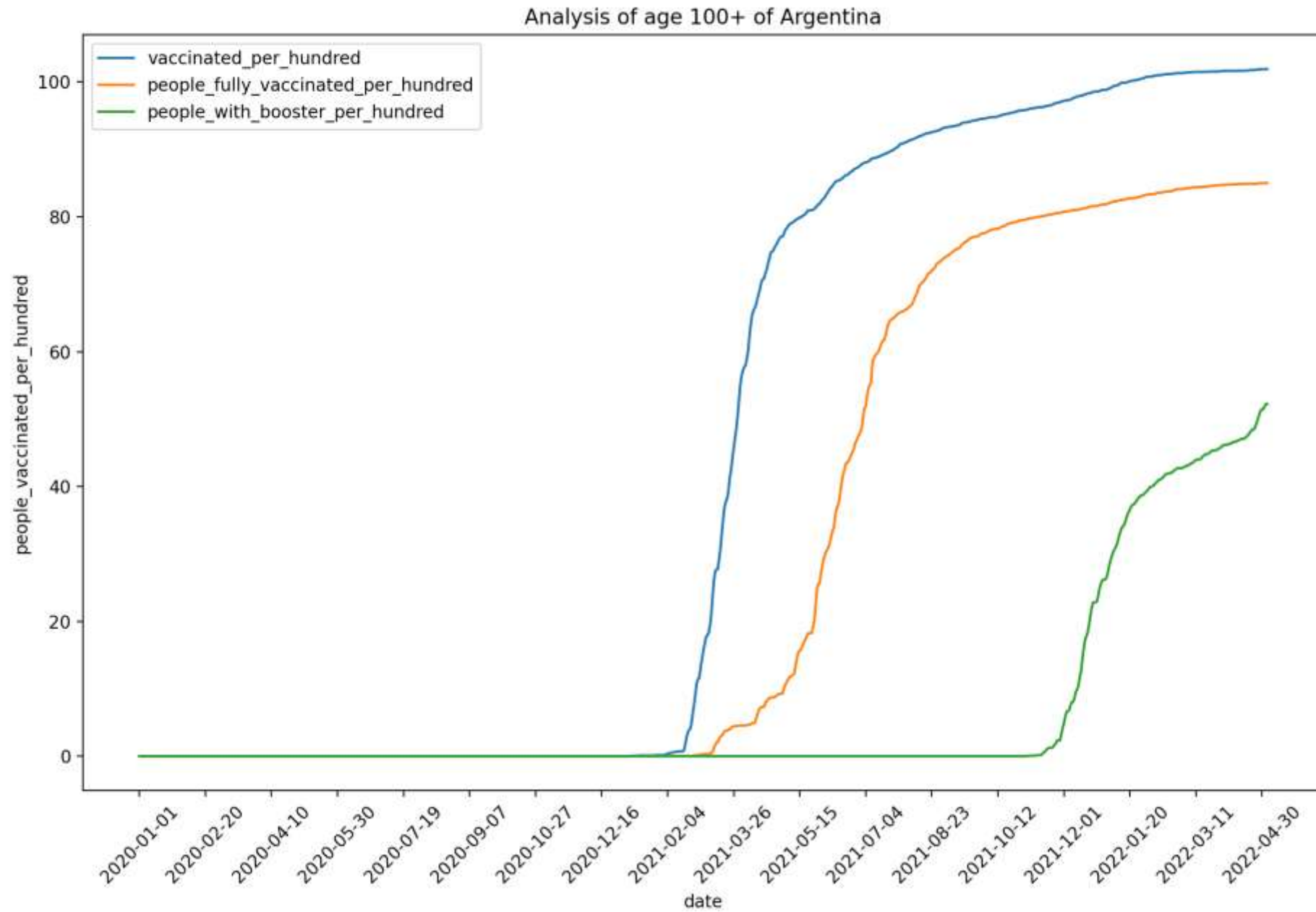
We can observe from the above graph that Oxford and Astrazeneca vaccines are used the most in countries with low GDP

The possible reasons could be

- It's efficacy is 76%
- It is 18 times cheaper than Pfizer.
- It can be kept in a freezer just between 2 to 8 degree celsius and can be used upto six months. This is a huge relief for Countries with huge population and low GDP.



- Above graph is the analysis for different age groups.
- We have divided the people in 4 different age groups i.e. old age people of age 70-79, mid age people of age 40-49, youth like us of age 18-29 and the kids of age 3-11.
- We can see from the graph that the old age people got vaccinated first and there is rapid rise in graph due to rapid vaccination for old age group.
- After that mid age people got vaccinated then teenagers.
- Here you can see from the graph that old age people got preference for vaccination because mostly old age people were affected by disease and heart problems and youth age group have high immunity power so they have got vaccinated later.



- This is the analysis for 100+ age groups.
- We can see from the graph that approximately 100% of the people got partially vaccinated.
- Approximately 85% of the 100+ age group got completely vaccinated.
- Approximately 50% of 100+ age group got booster dose.

Important Questions Identified

1. How GDP and Human development Index of the country affect the vaccination drive of the country
Which one of the above mentioned factors affect the most and why?

- Higher GDP can support the production, testing, and distribution of vaccines whereas countries with lower GDP may suffer in the distribution and production of vaccines.
- According to the United Nations Development Program 2020, 43.6 % of participants were from low Human Development Index (HDI) regions, 48.3 % high and very high, and 8.1 % from medium. The overall vaccination hesitancy rate was 37 %.

This implies that Human development index is not a good measure to analyse vaccination drive of the country.

 The picture can't be displayed.

- Above correlation table indicates a positive relationship between GDP per Capita, HDI, and the percentage of people vaccinated.
- Both of these correlations are not high; however, GDP has a higher correlation than HDI (0.23 Vs. 0.17). Lastly, this table shows a high positive relationship between GDP per Capita and HDI among the countries in the analysis (0.66).
- This shows that higher GDP per capita is significantly associated with greater numbers of vaccinations; however, there was no significant relationship between HDI and vaccine distribution.
- The reasoning would be that GDP may support the production, test, and distribution greater than HDI.

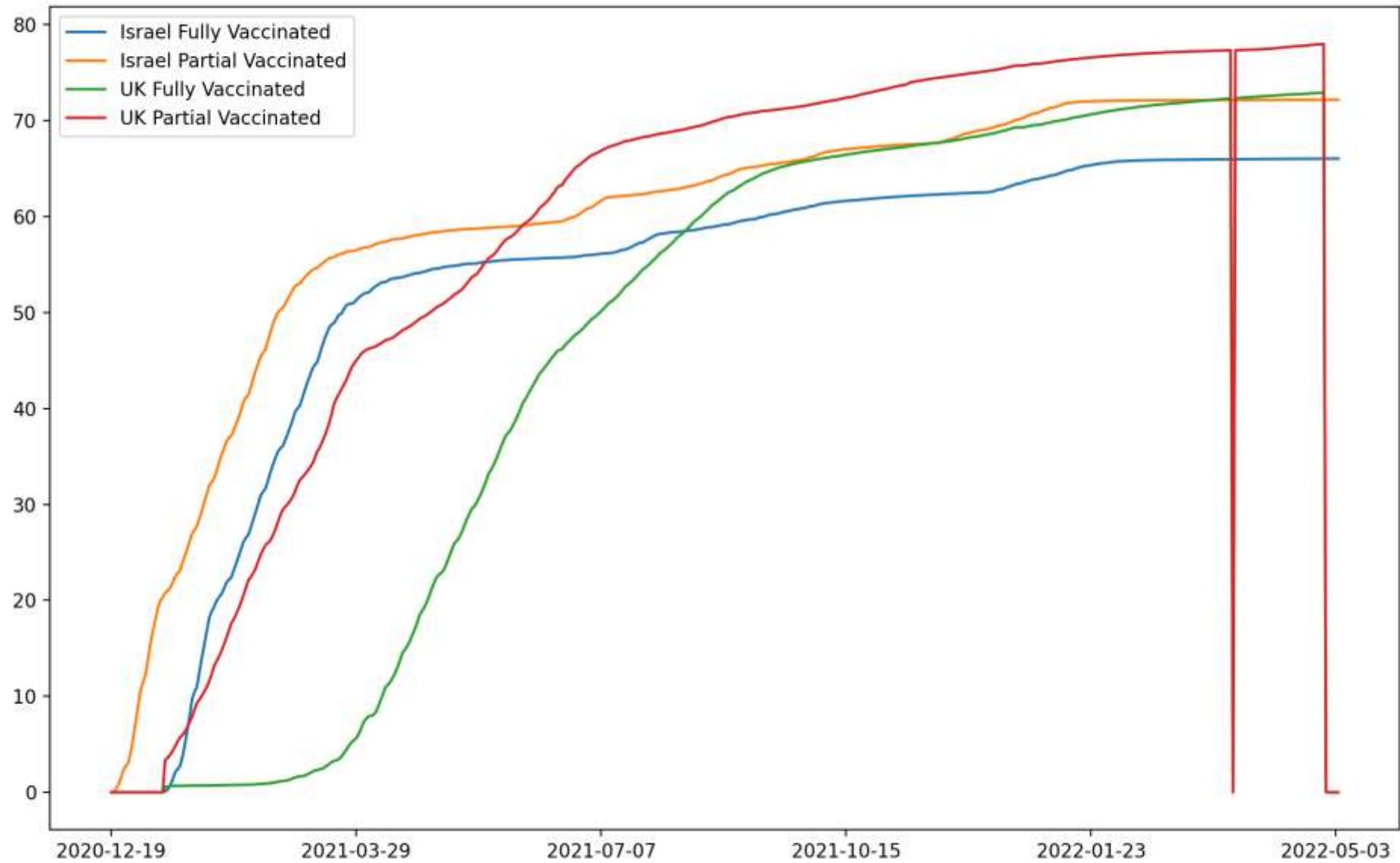
We have seen how GDP of the country helps in improving vaccination drive of the country, Similarly higher vaccination rate also helps in improving GDP of the country

- Countries that have been able to achieve higher rates of vaccine coverage have also been able to recover faster from the economic disruption caused by COVID-19.
- Prior to the war in Ukraine, while advanced economies had been projected to return to their pre-pandemic growth trends this year.
- Increasing vaccination levels will help drive a sustainable global economic recovery. Reducing the impact of COVID-19 on health systems will also allow countries to return to other crucial health and economic priorities.

Which approach or prioritization strategy was the best for vaccination drive ?

For ex UK has taken first dose first approach, delaying the delivery of second dose to achieve greater single dose coverage where as countries like Israel placed greater emphasis on giving two doses to smaller share of population first

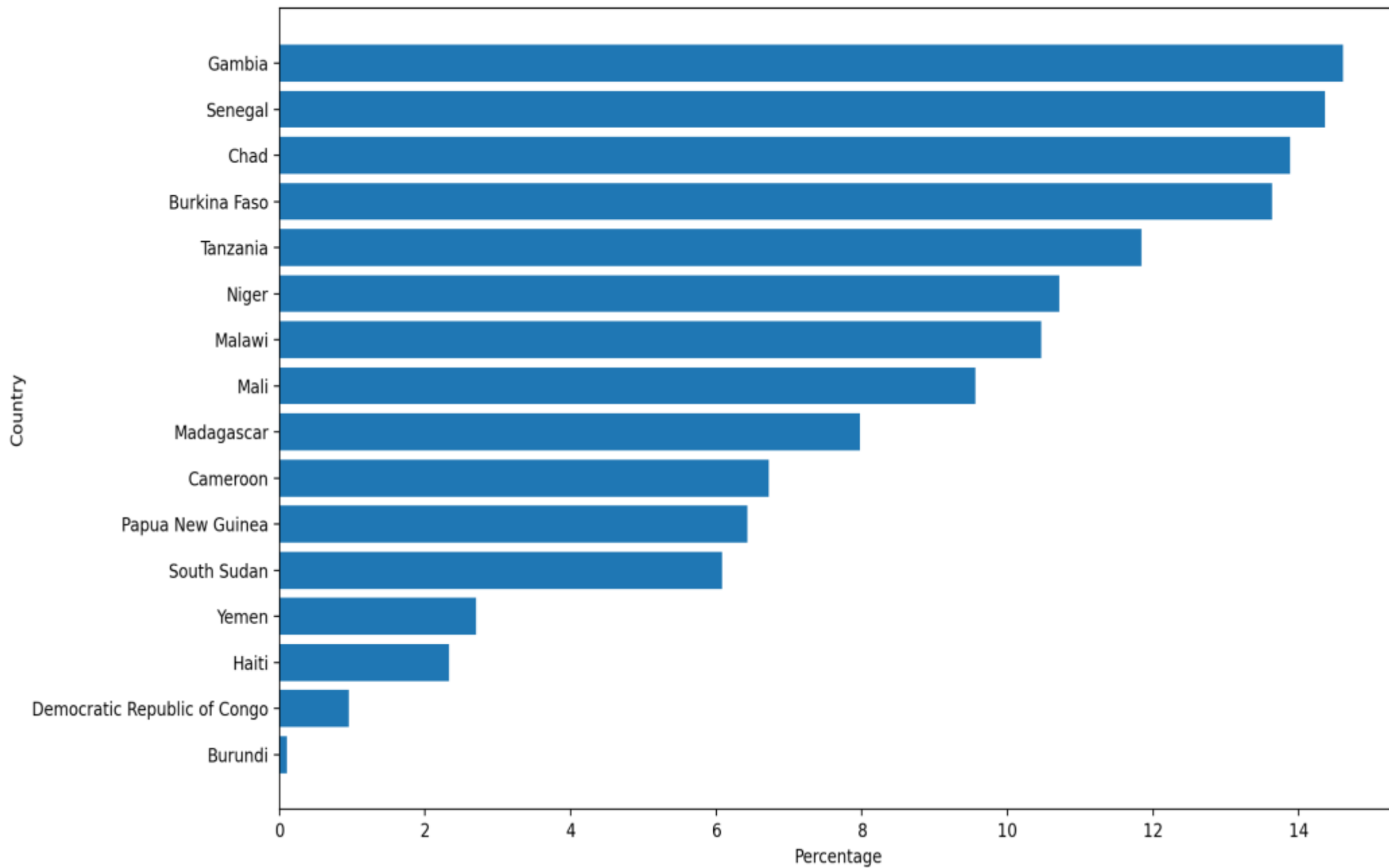
- The UK's new policy is to prioritise the first dose and to deliver the second one within three months rather than three weeks.
- The short-term benefit of “first dose first” may depend on how good a single dose is in the short term (pretty good in the case of the first vaccine from BioNTech/Pfizer), whether it will still be good enough for the elderly people? and whether a delay will ruin the booster effect of the second dose?



What is the effect on vaccination drives of underdeveloped and developing countries as developed nations are demanding for booster(3rd)doses

- Wealthy countries, such as Australia, Austria, Belgium, Canada etc have secured vaccine supplies and approved the booster shots (3rd doses) for their citizens by negotiating bilateral deals with vaccine manufacturers. As of January 12, 2022, 67.6% of the population have been vaccinated with at least one dose in high-income countries.
- Many under-developed and developing countries remain burdened by financial constraints, in low-income countries, only 11.4% of the population have been vaccinated with at least one dose. The booster shots in developed countries lead to further delay of vaccination in developing countries. This may cause serious problems, even for developed countries, because it facilitates further mutations of COVID-19 virus, which may lower the efficacy of currently available vaccines.

Least vaccinated Countries



Which Vaccine has performed the best in the given continent/country and why

What factors of the vaccine made it the better than the other vaccines

Based on the available data can we predict the most suitable vaccines for the respective countries

- Eight different Covid-19 vaccines are currently being used around the world and just over 172 million people have received their first dose, 2.2 for every 100 people
- All of the vaccines being used require two shots.
- Total 61 countries are using the Pfizer BioNTech vaccine including the United States, Canada, Mexico, the European Union and Saudi Arabia.

THANK YOU