PROJECT - An analysis of the impact of the COVID-19 vaccine on the world - an emphasis on the stock market, the economy and covid-related deaths

Introduction/Abstract:

The COVID-19 pandemic brought unprecedented challenges to global public health, economies, and financial markets. This project aims to assess the correlations between vaccine rollout, stock performance of major vaccine companies, and the broader impacts on global economies and public health. By analyzing stock data of leading vaccine manufacturers, alongside vaccine rollout rates and COVID-19-related statistics, the study explores how vaccine availability influenced market trends, economic recovery, and the decline in death rates. The analysis involves comparing the stock performance of individual vaccine companies with economic indicators such as GDP growth and mortality rates across countries, providing insights into the interplay between vaccine distribution, stock market behavior, and the pandemic's overall impact.

Team Members:

- **Mansour, Ahmad**
- **Narwal, Muskan**
- **Kaur, Jasleen**
- **Bendjaballah, Nazim**

Question 1:Which Pharmaceutical company performed the most during Covid-19?

To get the stock performance for the companies ,I started with a dataset that included trading volumes and price performance(open,close) for several pharmaceutical companies during the pandemic.Two parameter considered to study the stock market performance are the trading volume and Percentage Change in the Opening and closing Price of stocks at the end of each year.The data is collected for year (2019-2023) to compare the stocks both pre-covid and the post-covid era.Once, the data is collected for different vaccines using the python library finance.The data is combined using merged function into a single csv file(combined_data.csv) which is used in the later part for all the analyses done.

Analyzes :

The First analysis is to graph the data -For Trading Volume and the corresponding stock performance as It gives me an insight into level of activity or interest in a particular stock.
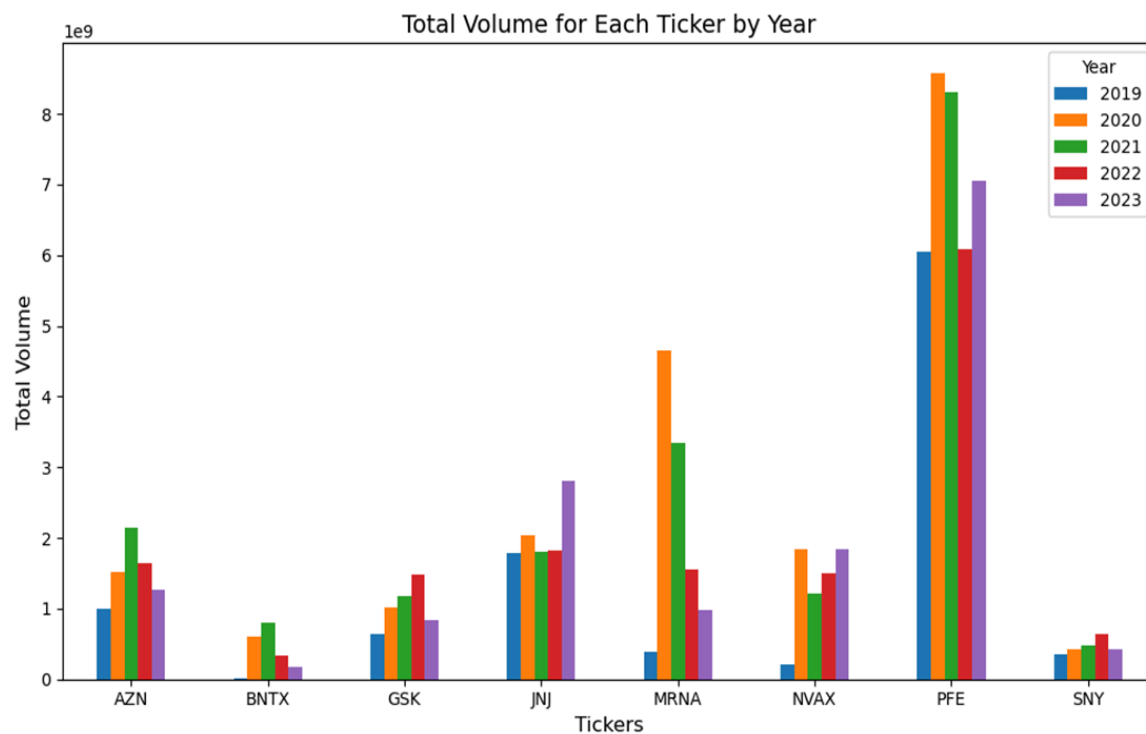
Figure:Trading Volume and Stock performance (for each Ticker)

As we can clearly see in the graph , it became clear that Pfizer consistently had the highest trading volume from 2019 to 2023. This surge in interest was likely due to its pivotal role in the COVID-19 vaccine rollout.
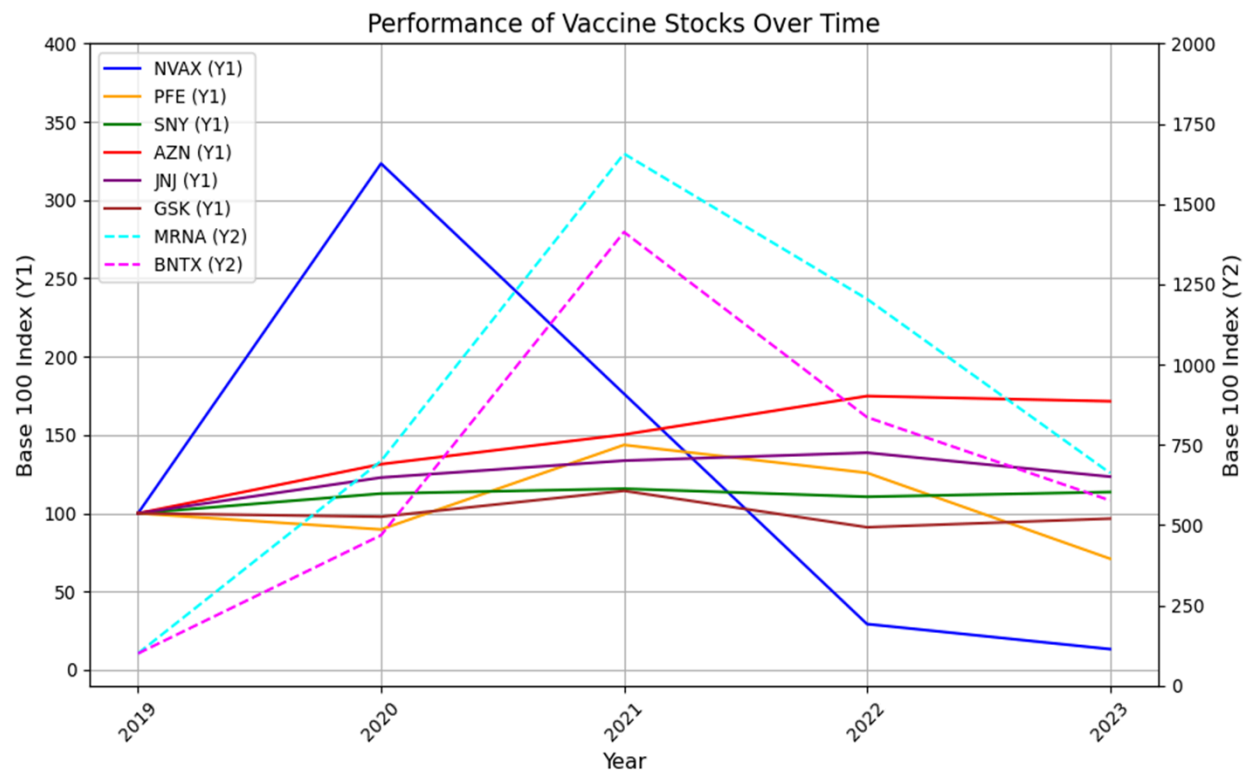
Figure:Price Growth and Stock performance (2019-2023)

However, despite this high trading volume, PFE didn't experience the same dramatic price spike as its competitors, Moderna and BioNTech. This suggests that while Pfizer was a stable investment with strong market interest, it didn't deliver the same level of price appreciation.

On the other hand, Moderna saw a significant increase in trading volume starting in 2020, coinciding with the development and rollout of its groundbreaking mRNA vaccine. The stock's trading volume remained robust into 2021, and its price performance reflected this momentum, skyrocketing to over 1750% of its 2019 base value. This incredible rise indicated that Moderna not only attracted substantial trading activity but also delivered immense returns to its investors during the pandemic.

BioNTech, while not matching the trading volume of Pfizer or Moderna, also experienced notable increases in volume during 2020 and 2021,likely due to its collaboration with Pfizer on the vaccine. Price-wise,BioNTech saw a massive rise, reaching over 1250% of its base value in 2021. This sharp increase in stock value underscored its strong performance, albeit on a slightly smaller scale than Moderna.

Novavax had a different story. It experienced a sharp spike in trading volume in 2020, fueled by high expectations for its vaccine development. However, this initial excitement did not translate into long-term gains, as both its trading volume and stock performance declined by 2023.

Despite a promising start, Novavax struggled to maintain its momentum.

In summary, the data painted a clear picture of performance during the pandemic. **Moderna emerged as the best-performing stock overall, achieving the highest price increase and capturing investor attention with its innovative mRNA vaccine.** BioNTech followed closely, benefiting from its partnership with Pfizer, while Pfizer itself, despite its high trading volume,saw more stable and moderate growth compared to the explosive performance of the mRNA stocks.

QUESTION 2 - ANALYZING THE CORRELATION BETWEEN NUMBER OF VACCINES ADMINISTERED FROM DIFFERENT MANUFACTURERS AND THE STOCK VOLUMES OF THEIR COMPANIES

In my section, I will build onto the work that my teammate Muskan did in section 1 on the stock data for the manufacturer companies. I will be taking the data provided on the stock volumes for each company and attempt to see if there is correlation between this and the vaccine rollout per manufacturer worldwide. Stock volume is a measure of the number of shares traded in a stock. In other words, it indicates the stock activity. Rising volume indicates rising interest in the stock, which is necessary for stock price to increase. Thus, stock volume is one of the measures that investors must consider when choosing which stock to invest in.

Logically speaking, an investor can assume that if one particular manufacturer is dominating the market, there should be an associated increase in the volume of the stock for the company as there will be much interest surrounding this company. This is where my data comes in to converge with the work in part 1 to produce the analysis. I collected data from Our World in Data (owid) covid-19-data Github repository, specifically the file vaccinations-by-manufacturers.csv in the vaccinations folder - https://github.com/owid/covid-19-data.git. This data breaks down vaccine rollout by manufacturer for a list of countries around the world. These countries are grouped by region, allowing us to get an idea of how many vaccines per manufacturer were administered in regions around the world.

Note/limitation: the list of countries providing this data is limited. Thus, for certain regions, the following analysis will not be entirely complete and will be more of a particular snapshot. Unfortunately, this is real-world data collected on a real-world pandemic and therefore the dataset will not be perfect. This limitation means that the analysis that will follow will be specific to the data that we have, which may not be complete for certain regions.

(Analyis below)

Goal 1: Generate visualizations breaking down this data, to determine the top 2 manufacturers per region
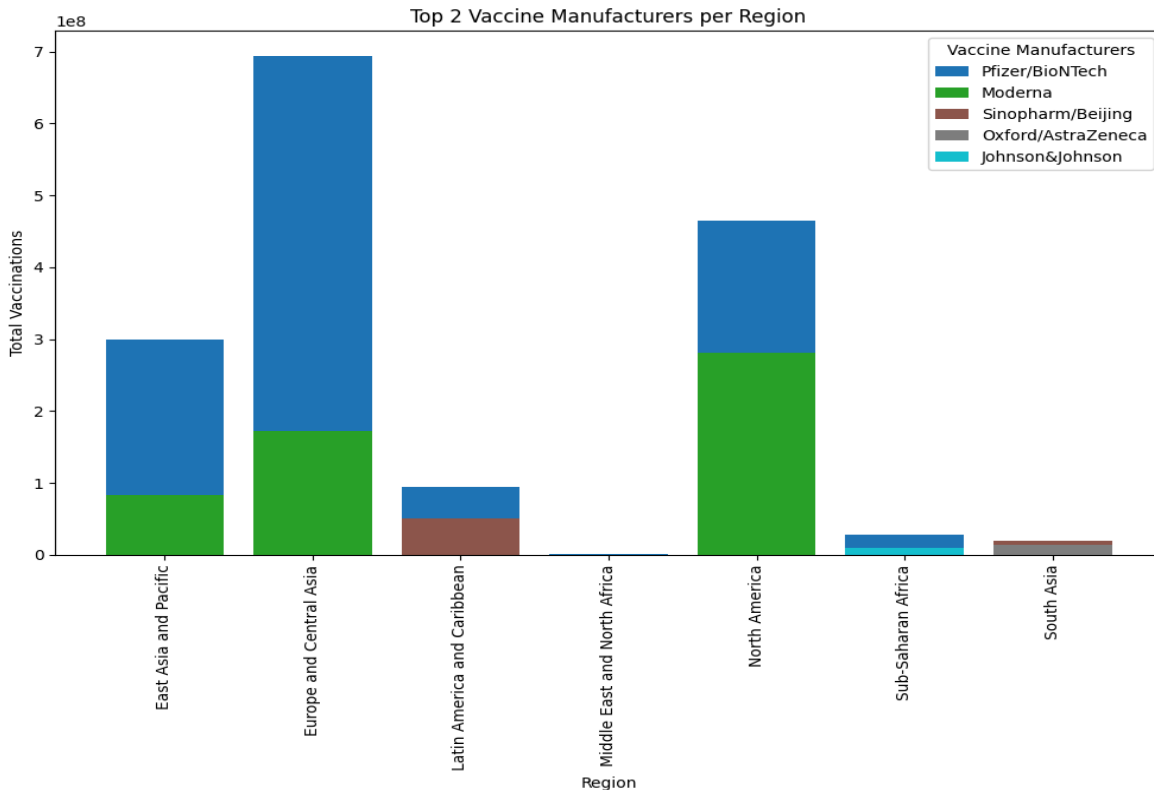
Analysis: As seen on the graph above, Pfizer dominates the market as the top manufacturer for every region aside from South Asia. In South Asia, Sinopharm is the top manufacturer and it is the second top manufacturer in Latin America. In East Asia and Pacific, Europe and Central Asia and North America, Moderna is the second largest manufacturer. The data in the csv files was cumulative day by day so a code was produced to retrieve the final value for each manufacturer for each country. Then, for each region, the final value for each manufacturer in each country was summed to give us the total vaccinations per manufacturer per region. The purpose of collecting this data is because the hypothesis is that if one particular company is dominating, like Pfizer, then we could expect an increase in the stock volume of the company. Thus, the goal is to collect this data into a dataframe to allow us to do a linear regression to see if there is a correlation between the domination of the vaccine market and the stock volume of the company. If a company is dominating, it can be expected that there would be a lot of interest in the stock of the company. As such, this portion collected this necessary data to allow us to know which stock volumes to do the linear regression with.

(Analysis below)

Goal 2: Generate a linear regression and get an r-value that allows us to determine if there is in fact a correlation

Top manufacturer per region vs Stock Volume
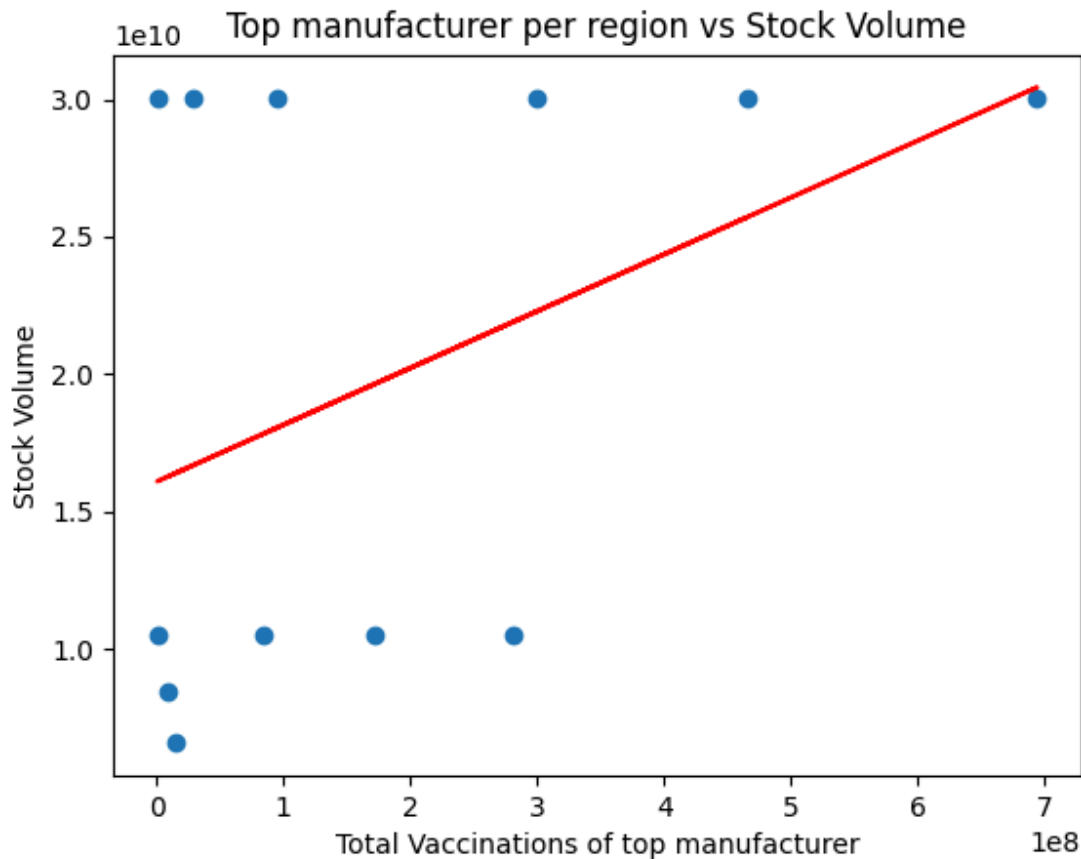
R-value = 0.42
Analysis: Seeing as the r-value is 0.42, it cannot be said that there is a strong correlation (0.6+). Rather, this value falls under the range of moderate. As such, it can only be said that there is a moderate correlation between domination of vaccine market and stock volume of the company. For starters, we would expect there to be a lot of interest surrounding a particular company as people keep up with the trends of the vaccine rollout. In other words, we would expect high stock volumes for Pfizer and Moderna as they are dominating the worldwide markets (Sinopharm is not considered as no stock data was provided for it). That said, this statistical measure shows that the correlation is only moderate, so we cannot say with strong conviction that the high stock volume (in comparison to the rest of the companies) is associated with the large amount of vaccine doses administered worldwide. Nonetheless, a moderate correlation is not necessarily negligible. So, it can be said that the amount of doses that are being administered by the companies should be moderately considered by investors when analyzing stock volumes to make investment decisions. Of course other metrics are also important but this is the one being considered for this section.
I would like to conclude by re-iterating that if all of the countries in the world provided the data on vaccine rollout by manufacturer, perhaps there would be a stronger correlation. Thus, this is the major limitation for this section and should be taken into account when looking through this analysis and its conclusions. Our moderate correlation has to do with the data that is available.

Question 3:

Research Question: How did the vaccine rollout affect the economy?

While doing our analysis we first looked at each vaccine from a stocks standpoint to the vaccine rollout. The two indicators we used for measuring the economy are CPI and GDP. CPI is the consumer price index. It's the rate at which individuals in that region of the world purchased consumer goods. GDP is the measure of goods and services produced in that region**.**

The analysis is the comparison between the region that has the most vaccinated individuals and the GDP. **As below(shows that Europe and Central Asia had the most individuals vaccinated):**
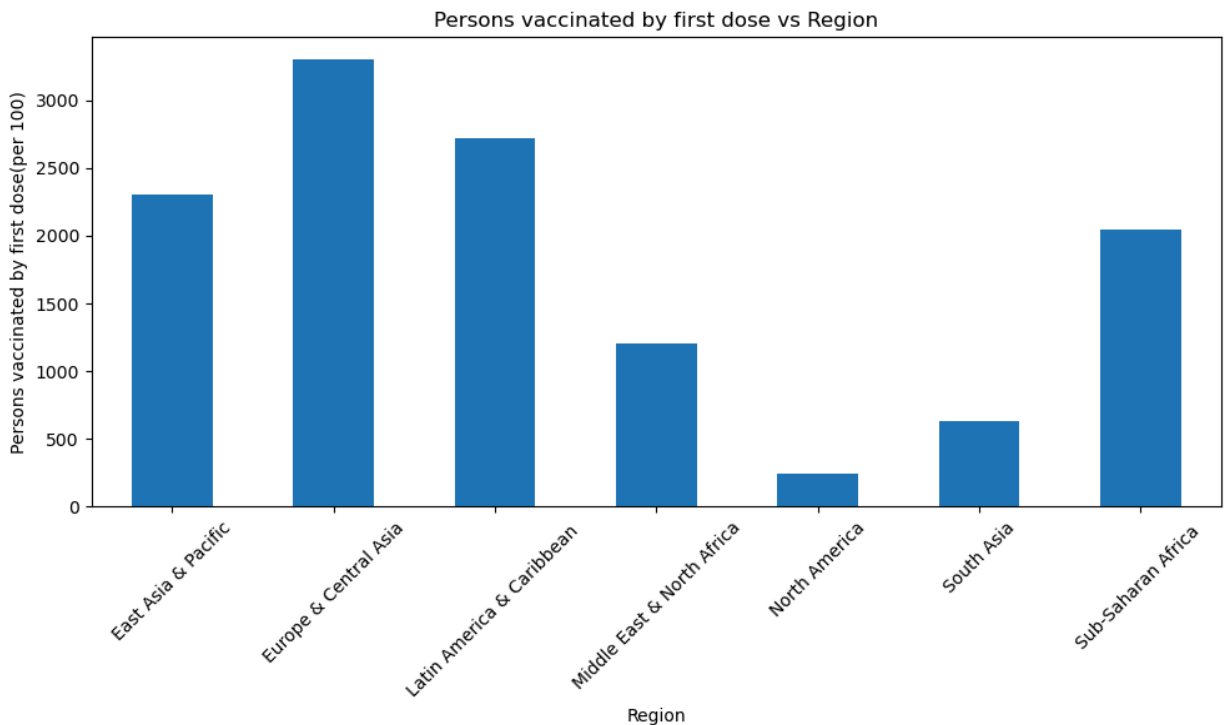

GDP in 2020(at the beginning of the pandemic):

GDP Rate in 2020 vs Regions
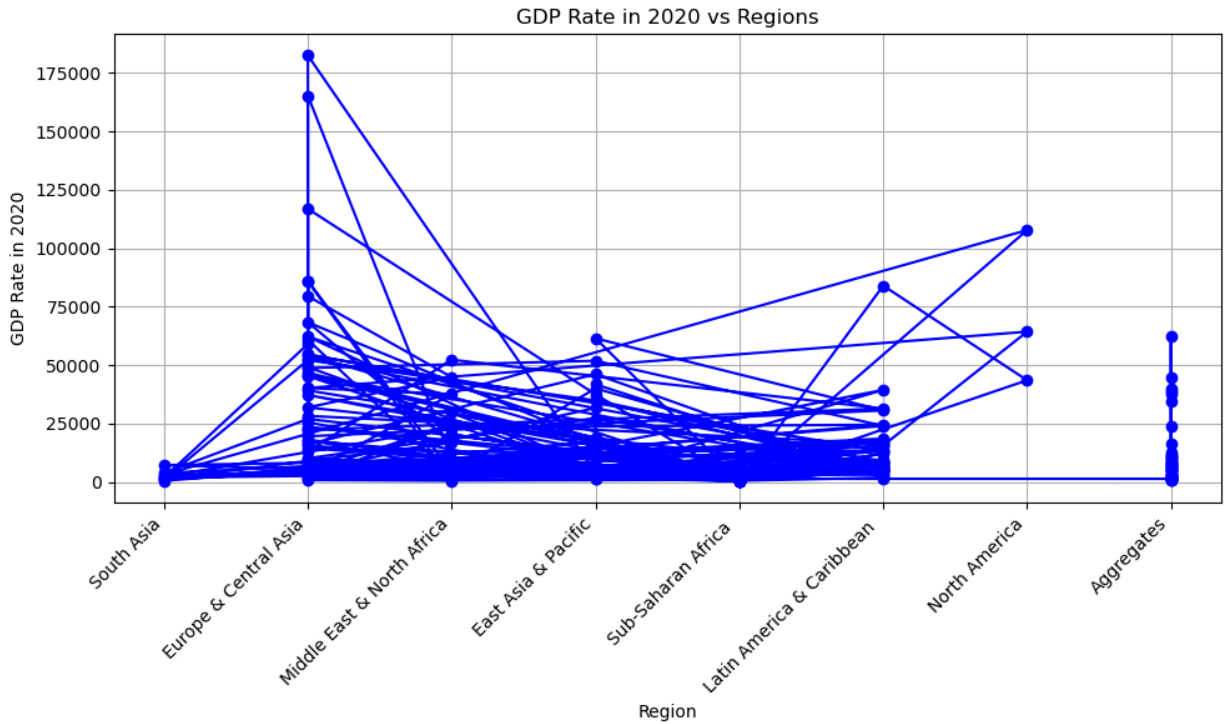
There is a direct correlation between the GDP and people vaccinated in each region. There is a positive correlation between the region with the most people vaccinated and the GDP. Europe and Central Asia have the highest number of people vaccinated and have the highest GDP around the time the first dose was initiated. We can conclude they had the greatest production of goods and services.

**This similar trend can be seen in the GDP during the booster:**

Persons vaccinated by Booster vs Region
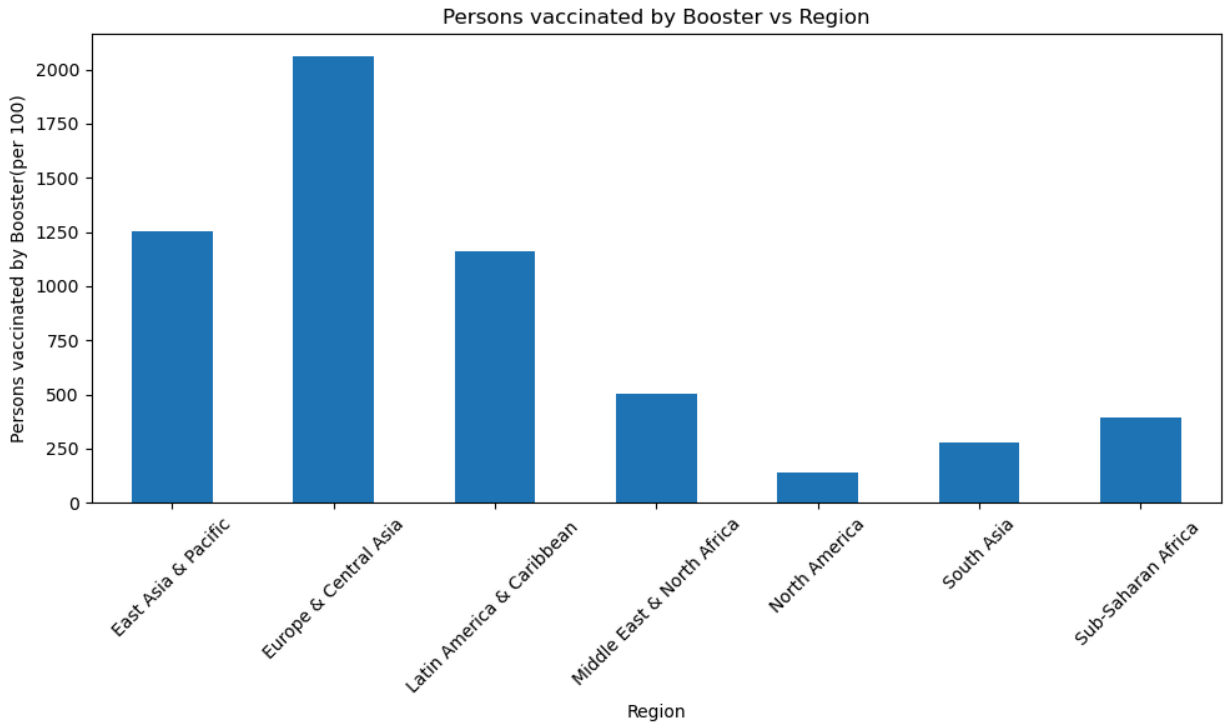
GDP in 2022(at the post of the pandemic):


GDP Rate in 2022 vs Regions
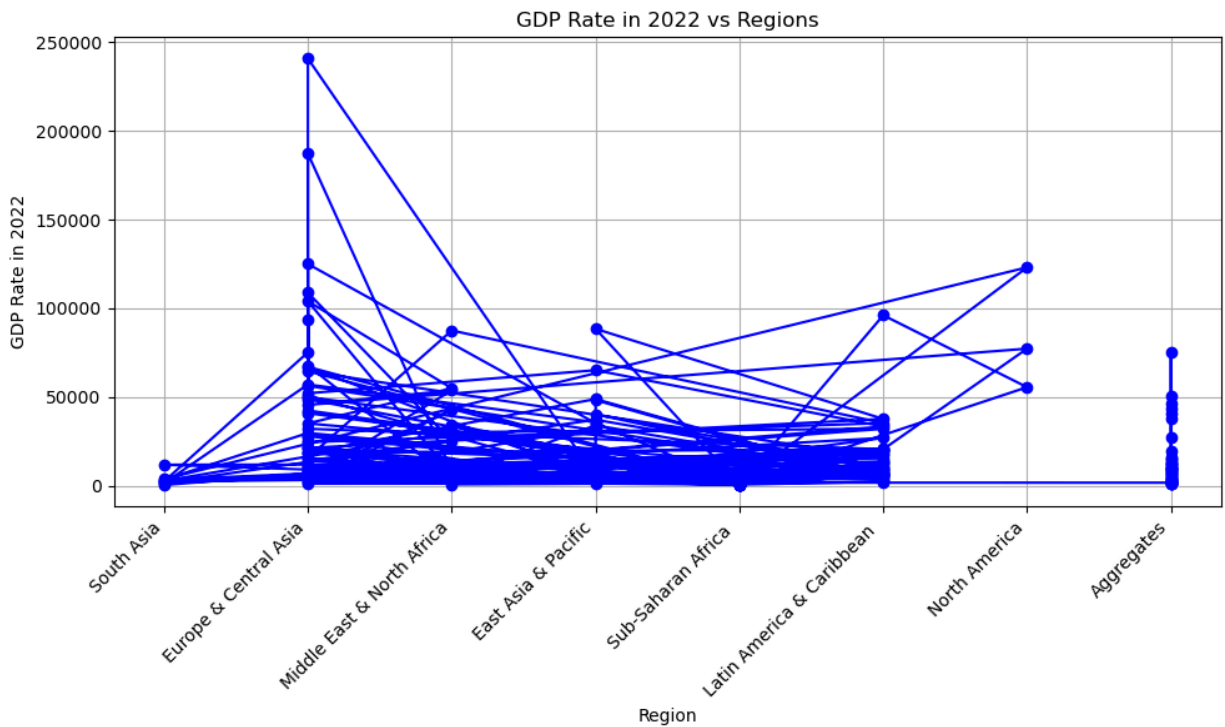
The CPI rate had an inverse trend compared to GDP. The region that had the least number of people vaccinated, had the greatest CPI rate. Showing that due to a lack of resources in the economy of the region, people were spending more on consumer goods. Before that pandemic

there was an increase in consumer goods in Sub-Saharan Africa, yet Sub-Saharan Africa has smaller GDP compared to other regions.
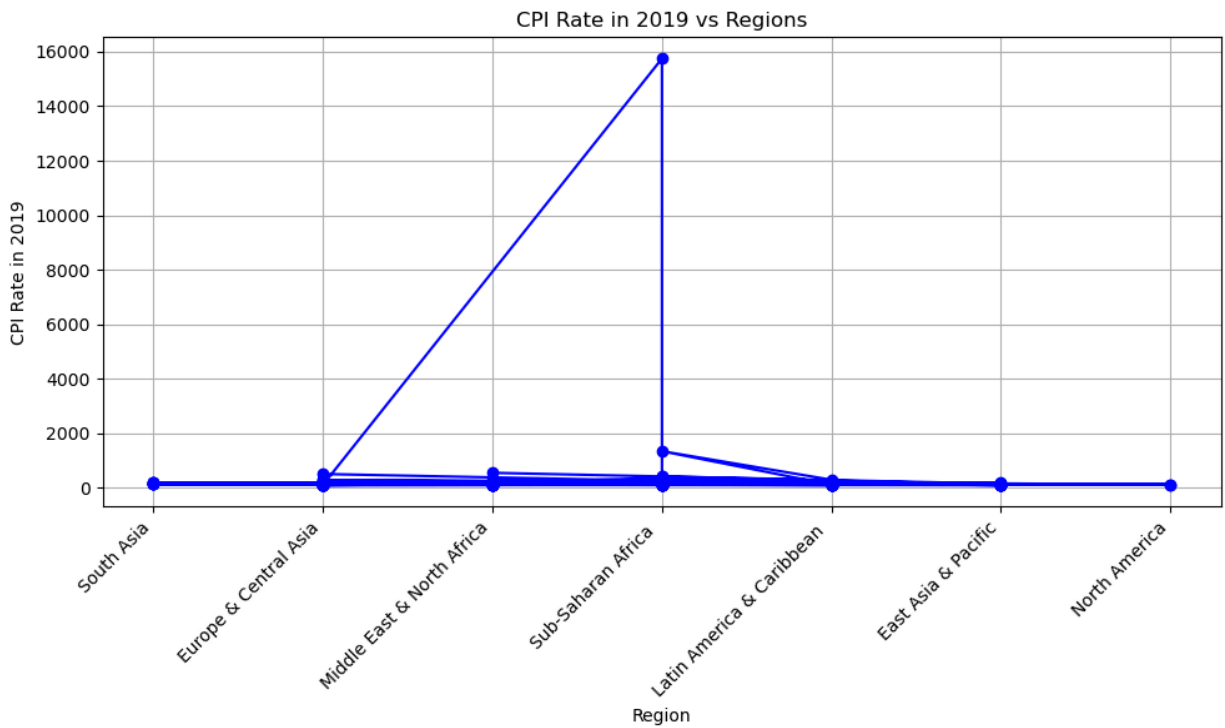
After seeing the results of the CPI rates post pandemic. It can be concluded that the GDP was one of the main indicators affected by COVID. Whereas individuals purchasing power of consumer good relatives stayed consistent.

CPI Rate in 2023 vs Regions
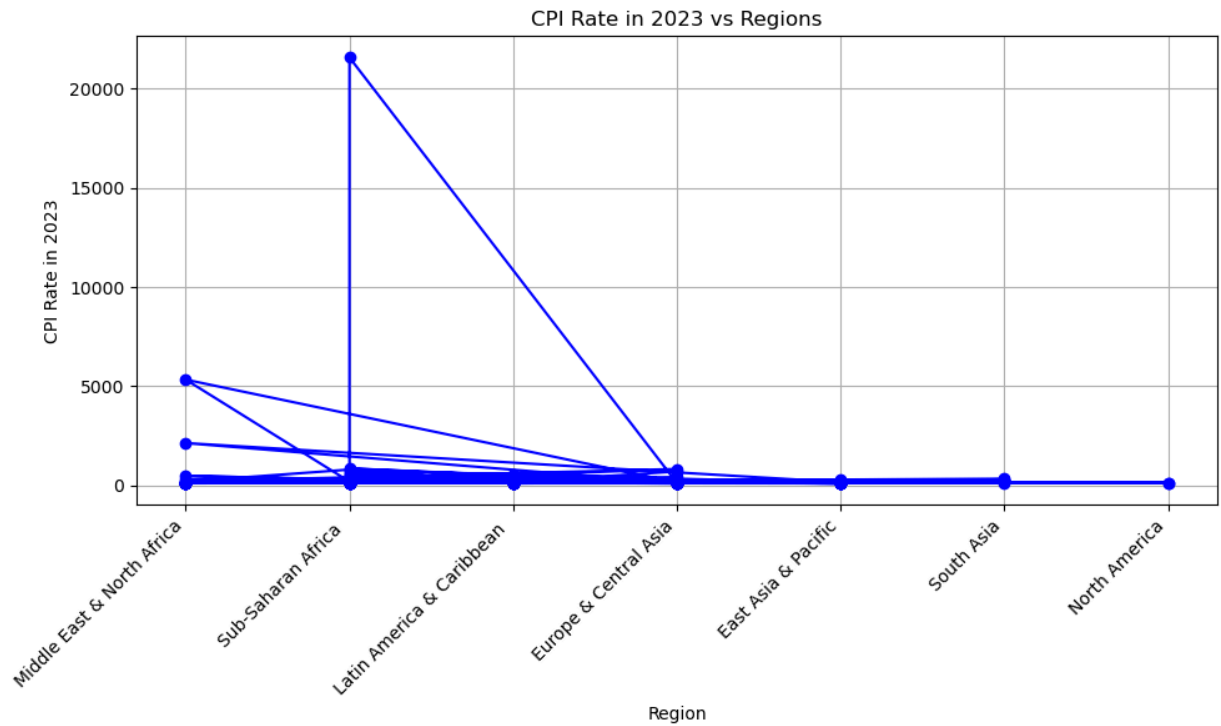
Question 4:

Question 4

Is there a correlation between Covid population vaccination and deaths number

Data collection
 the analysis is based on 4 sources of data

    CSV file

    CSV file : Deaths per day and per country

    CSV file : Vaccination End Of period

Source : World Heath Organisation :
https://data.who.int/dashboards/covid19/data?n=o

CSV file : Vaccination per day and per country

Source : KAGGLE :
https://www.kaggle.com/datasets/fedesoriano/coronavirus-covid19-vaccinations-data?resource=download

CSV file : API's

Countries referential + Population per country

API World Bank :

https://api.worldbank.org/v2/country?format=json

Data cleaning and transformation

- For the daily data : vaccination and deaths
- Outliers were removed using IQR method
- A moving average (7days) was used to smooth the data
- Data for vaccination was not available for the full period, this data was removed

The Data per region is directly originated from data per country : there was no significant transformation was applied on data per countr
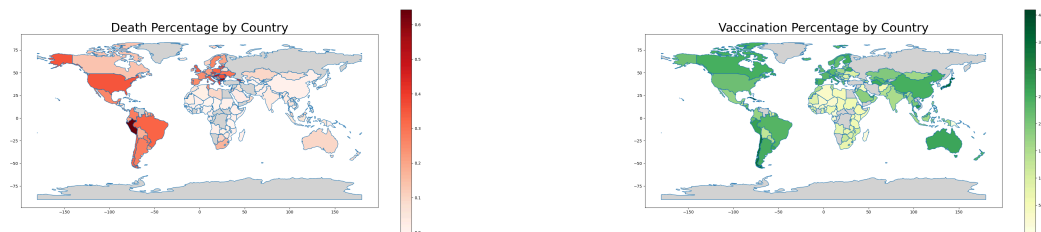

Daily analysis

This chart below shows how the daily diseases due to COVID evolved between the beginning of the pandemic till today,

There was several pics corresponding to the different waves

Since the launch of the vaccines late 2020, the number of deaths starts decreasing which would suggest a correlation between the vaccination campaign and the effect of the pandemic,

Daily Covid deaths vs vaccination evolution
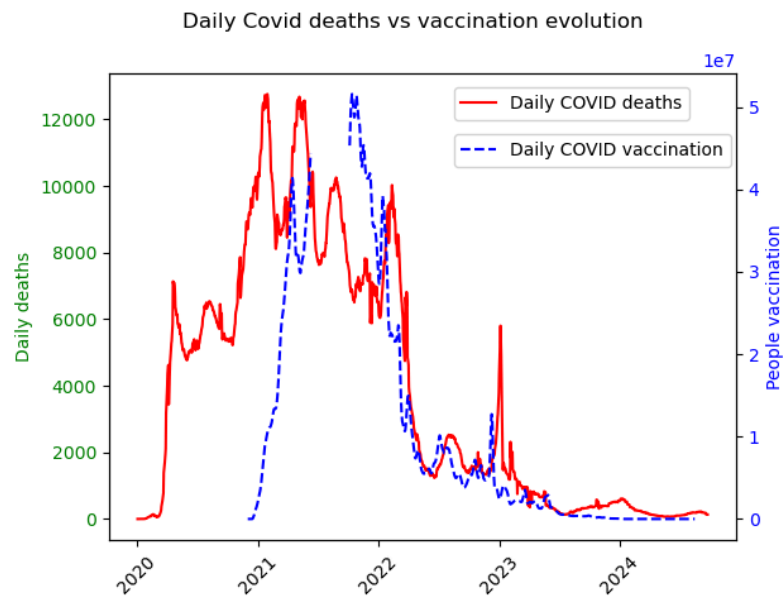
Correlation on a daily basis

When we compare the evolution of the vaccines with the daily deaths number, we do not observe a correlation between the two indicators.

The data is skewed and does not follow an obvious pattern

daily_vaccination_smooth vs daily_deaths_smooth

$y = 0.0x + 1295.09$
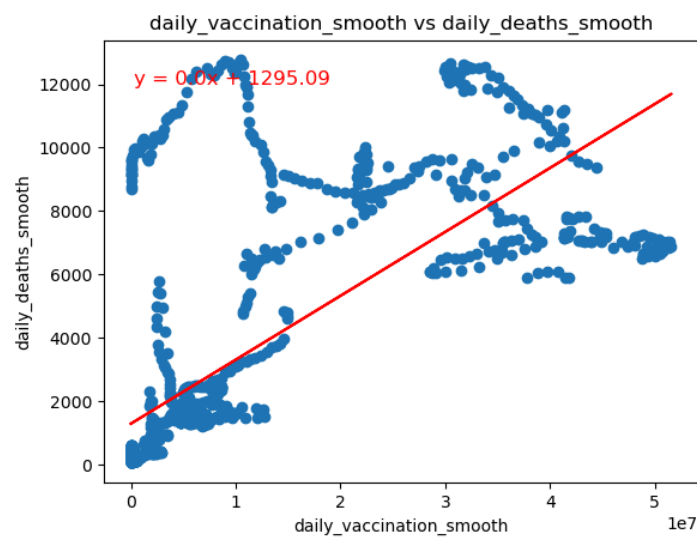
Correlation with a country aggregation

The previous observation is confirmed when we compare the vaccination% (number of vaccins compared to the total population) and the deaths% (numbers of deaths compared to the total population), and this country by country

In this case we also observe no correlation between the 2 sets of data collected for 200 countries.
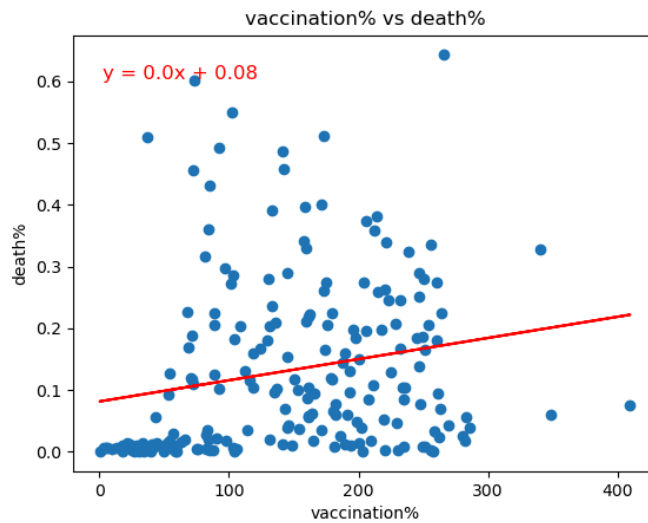
Analysis by geographical regions

If we aggregate the countries by region, this observation is also confirmed, However, we recognize some specifics that can explain the impact of the pandemic in the world

vaccination% vs deaths% per region
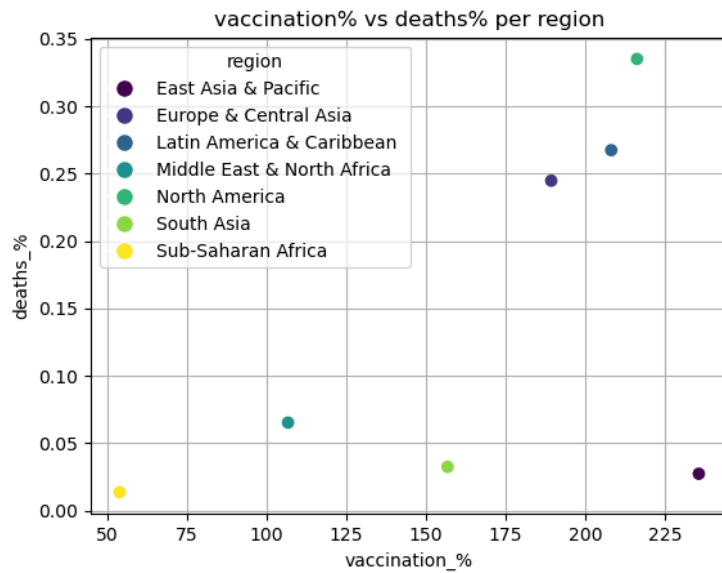
- Sub-Saharan Africa : the impact was limited and observed from the beginning (no explanation till now). The most probable is a natural immunization of the population to the severe consequences.
- Asia : the contingent policies that last longer could explain the limitation of deaths in these areas : longer and more strict lockdowns, population control.
- There are 3 regions highly impacted with severe consequences