National College of Ireland

Project Submission Sheet

Student Name: Zuu Zuu Kyaw Shwe

Student ID: 24106585

Programme: MSCDAD_C   **Year:** 2024

Module: Analytics Programming and Data Visualisation

Lecturer: Furqan Rustam

Submission Due Date: 16th Dec 2024

Project Title: Data Analytics of COVID-19's Global Impact

Word Count: 3540

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

__ALL__ internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: *Zuu*

Date: 16th Dec 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Data Analytics of COVID-19's Global Impact

Lin Myat Zaw Latt
School of Computing (MSCDAD-C)
National College of Ireland
Dublin, Ireland
x24121614@student.ncirl.ie

Zin Win Phyo
School of Computing (MSCDAD-C)
National College of Ireland
Dublin, Ireland
x23402849@student.ncirl.ie

Zuu Zuu Kyaw Shwe
School of Computing (MSCDAD-C)
National College of Ireland
Dublin, Ireland
x24106585@student.ncirl.ie

Abstract—In 2020, the WHO announced the discovery of a new disease, which was later labeled COVID-19. The COVID-19 pandemic has had severe impacts on the global economy and people's lives, causing millions of deaths worldwide and forcing the need to find ways to prevent its infection. The agenda of creating a vaccine was also one of the preventive measures. The aim of this project is to show insights into how the COVID-19 cases, the vaccination data, and the stocks of healthcare companies have had an effect on social and economic impacts, which include global GDP, unemployment, and government debts.

I. INTRODUCTION

The advent of the COVID-19 pandemic necessitated a great deal of attention towards comprehending its effects across populations, economies and entire societies. This research intends to showcase the effect the pandemic had on the resultant, the population epidemiology, the markets and even economic growth by discussing the relevant quantitative modeling techniques. Furthermore, the project emphasizes COVID-19 cases, vaccination, economic variables and stock markets and with the information, the projects aims to explore the different outcomes that were generated as a result of the pandemic's repercussions. The key indicator to indicate is the research question: What has been the effect of the COVID-19 pandemic on the economic activities and the market behaviors and the mainstream media narratives that influence these performance metrics? This analysis provides a novel perspective on the pandemic's interplay across different sectors while providing great emphasis on data driven techniques to derive broad conclusions.

II. RELATED WORK

This current review aims to incorporate existing scholarly works in order to evaluate the level of research which has examined the association between the COVID-19 pandemic, distribution of vaccines against it and the stock markets, and economic repercussions like GDP, a nation's debt, and unemployment rate.

The outbreak of the COVID-19 pandemic altered the boundaries of vulnerability on global public health care, economy and vaccine distribution. Mathieu et al. [1] pointed out the world's inequalities in vaccine usage, high income areas as North America and Europe had better vaccination coverage compared to Africa and other low income regions, which had limited healthcare facilities and faced logistical challenges.

Similarly, vaccine hesitancy was identified by Khubchandani et al. [2] as a significant obstacle in the United States, with resistance stemming from socioeconomic considerations, political affiliations, and disinformation. These results are consistent with our observations of differences in immunization rates and booster doses among WHO regions.

According to Oanh [3], COVID-19 vaccination campaigns increased investor confidence, especially in less developed countries, stabilizing financial markets around the world. This is consistent with our data, which shows that investors are quite optimistic about top vaccine manufacturers like Moderna and BioNTech, whose innovations were essential to the success of vaccinations.

Gagnon et al. [4] quantified the pandemic's effect on global GDP, demonstrating that lockdown measures and global trade disruptions were primary drivers of the economic downturn. Their conclusions are consistent with our evaluation where the level of global GDP expansion shrank drastically in 2020 since majority of the countries were in lockdown and trade was suppressed.

Blustein et al. [5] focused their research on the relation between unemployment and the COVID–19 pandemic, explaining how marginalized groups and precarious employment had worse standings. They however indicate that loss of employment and medical-related problems have compounded effects, which tallies with the increase of unemployment rate internationally during the pandemic. They also emphasize the necessity of public policies, e.g. economic stimulus and workforce re-education programs, for employment maintenance. This goes hand-in-hand with our analysis for high unemployment rates which indicates a large increase in public debt as possible explanations for economic shocks.

III. DATA PROCESSING METHODOLOGY

A. Dataset Selection

1) Dataset 1 - COVID Cases/Deaths Data: The first dataset is on COVID-19 cases and deaths at a global scale, which even specifies the daily counts, the total cumulative, and demographic breakdowns. The data spans from the onset of the pandemic in 2020 to the present. It was collected using the COVID Act Now API and aggregated into JSON format with variables not limited to country, date, number of cases, number of deaths, and so on.

2) Dataset 2 - COVID Vaccine Data: The second dataset provides a detailed record of vaccine administration worldwide. It features important statistics such as the total number of vaccinations given, how many people have received at least one dose, and how many have completed their vaccination series or received booster doses. Data were downloaded in CSV format from the WHO website.

3) Dataset 3 - COVID Vaccine Metadata: The third collection consists of COVID vaccination metadata, such as country code and manufacturing information. Additionally apparent are several vaccine producers, including Pfizer, Moderna, and AstraZeneca. The information was gathered from the WHO website and is kept in CSV format.

4) Dataset 4 - GDP Data: The fourth dataset contains the GDP and other variables like the year, country, country code, and GDP growth rate. Numerous factors contributed to the GDP performance of countries over the long period of years. The data was collected from the world bank group API and stored in the form of CSV.

5) Dataset 5 - Unemployment Data: The fifth dataset provides global unemployment rates, thereby showing the changes in unemployment over the years. Like the forth data set, the source of the data was the world bank group API and it was stored in csv files.

6) Dataset 6 - Central Govenment Debt Data: The sixth dataset contains temporal information about the central government debt statistics, which was also obtained from the world bank group API. Variables include country, year, total debt, etc. The dataset was stored in CSV format as well.

7) Dataset 7 - Stock Market Data: The seventh and final dataset provides insights into the stock markets affected by COVID-19, specifically examining pandemic-related businesses including vaccine development and other healthcare sectors. The company names include Pfizer, Moderna, Johnson and Johnson, and so on. The data, which covers a peroid of 10 years, was extracted via the yfinance API and saved in the CSV format. These data provide insights into stock performance of entities related to COVID-19.

B. *Data Processing Algorithms*

Figures 1 and 2 clearly show the ETL process, which stands for extract, transform, and load. The data extraction and data fetching were completed first followed by data transformation to tailor the needs of the research. Lastly, the transformed data was uploaded into MongoDB and Postgres databases. This integration of pandemic data, economic disruption data, and stock performance data utilized Tableau to assist in analyzing these information and gain insight into the impact of COVID-19 on the economy, and stock market.
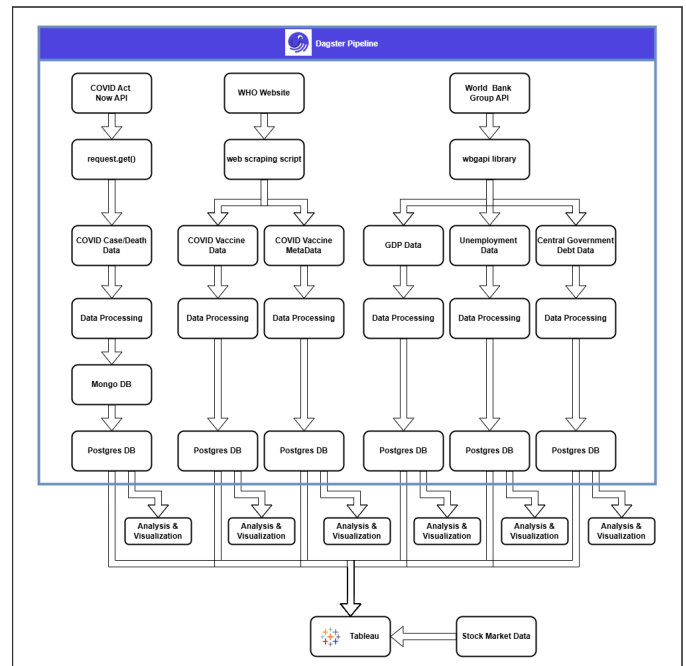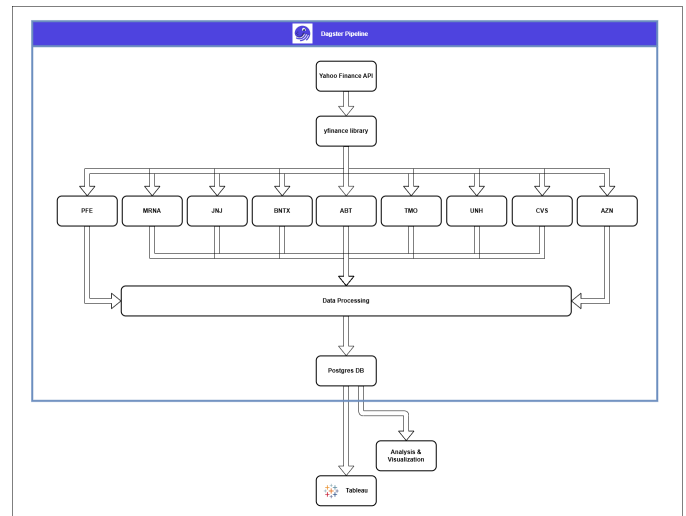

Fig. 1: Diagram of Data Gathering Activities

Fig. 2: Diagram of Data Gathering Activities - Stock Market Data

C. *Data Extraction*

The datasets used in this research were primarily sourced through Application Programming Interfaces (APIs), which allowed for efficient and automated retrieval of data from trusted sources. The APIs employed in this project provided real-time, reliable data essential for the analysis.

- **COVID Act Now API**: This API was employed to acquire the most current data concerning COVID-19 cases and deaths around the globe, including daily and demographic reports. It was chosen due to its reliability in providing real-time, accurate information on the pandemic's progression, which is essential for our analysis.

– Collected Data: COVID Cases/Deaths Data
- **World Bank Group API (wbgapi)**: This API provided information such as GDP increases, unemployment rate, and national government debt. It was included in order to ascertain the economic wider impact the pandemic as these indicators help in evaluating the impact of the crisis on the welfare of countries and enhancement of financial stability.
  – Collected Data: GDP Data, Unemployment Data, Central Government Debt Data
- **Yahoo Finance API (yfinance)**: The yfinance API was employed to gather stock market data for companies engaged in vaccine development and healthcare. Those companies are Pfizer (PFE), Moderna (MRNA), Johnson & Johnson (JNJ), BioNTech (BNTX), Abbott Laboratories (ABT), Thermo Fisher Scientific (TMO), UnitedHealth Group (UNH), CVS Health (CVS), and AstraZeneca (AZN). This API was selected due to its capability to deliver both historical and current stock data, enabling an analysis of how the pandemic has impacted market performance.
  – Collected Data: Stock Market Data
- **Web Scraping from WHO**: To obtain data from World Health Organization (WHO) regarding COVID vaccine related websites, a web-scraping script has been used. This approach was chosen to rapidly obtain relevant and current data from the WHO page which is vital for this particular study to evaluate the interaction between the trends in population vaccination coverage and the outcome indicators related to COVID-19.
  – Collected Data: COVID Vaccine Data; COVID Vaccine Metadata Data

D. Data Storage

Once the data was collected through extraction, it was stored in two kinds of databases, Mongo DB and Postgre SQL. The purpose of having both was determined by the type of data as well as the analysis to be conducted.

COVID Cases and Deaths raw data files in JSON format were brought into the MongoDB database. This is mainly due to the nature of Mongo DB which is very flexible and can support import of semi-structured and structured data.

On the contrary, the PostgreSQL database was utilized for the purpose of storing transformed data which has adopted a more organized approach. Enabling the performance of intricate analysis and data consistency with a multitude of manipulations and transformations, PostgreSQL delivered complex relational database structures.

E. Technologies Used

In our analysis, Python and Dagster were chosen for their flexibility and efficiency in handling data processing workflows.

Python served as the primary programming language due to its widespread use in data science, its simplicity, and the vast array of libraries available for data processing and analysis. Python's extensive ecosystem made it a suitable choice for handling data extraction, transformation, cleaning, and analysis.

We included various Python libraries, including at pandas and numpy for data manipulation , visualization using matplotlib and seaborn, web scraping using BeautifulSoup and urllib, etc. With such tools, data was processed effectively and smoothly in a more efficient way.

Dagster served as the data orchestrator in charge of running the ETL (Extract, Transform, Load). This automated and kept track of the data pipeline such that data movement from extraction to analysis was done without hiccups. Working with dagster integrated with python and databases made processes effective while ensuring data processing was uniform and scalable.

IV. DATA VISUALISATION METHODOLOGY

The COVID-19 Tableau dashboard visualizes case trends, vaccination progress, and economic indicators through bar charts, line graphs, and maps. Line graphs highlight trends over time, while maps show regional variations. By doing this, we can guarantee clear spatial insights. In this way, we are able to provide distinct geographic context. For instance, red denotes a surge in cases while green indicates levels of vaccination The use of interactivity such as filters and tooltips facilitates ease of exploration by region and time line. Visualizations were combined into a single, cohesive dashboard to enable a seamless, holistic analysis of COVID-19's impact.

Additionally, data visualizations were created using Python libraries such as Matplotlib and Seaborn, which provided further flexibility for generating customized plots and in-depth analysis.

V. RESULTS AND EVALUATION

The world map cases visualization (Figure 3) depicts the global perspective showing how the pandemic affected different countries and regions. The area of the pandemic as the visualization indicates, is determined to varying degrees in several parts of the globe. The orange color was used and the indication of more cases were shown with a brighter color whereas areas with less cases are colored in light orange. This map is important for considering the global effects of COVID-19.
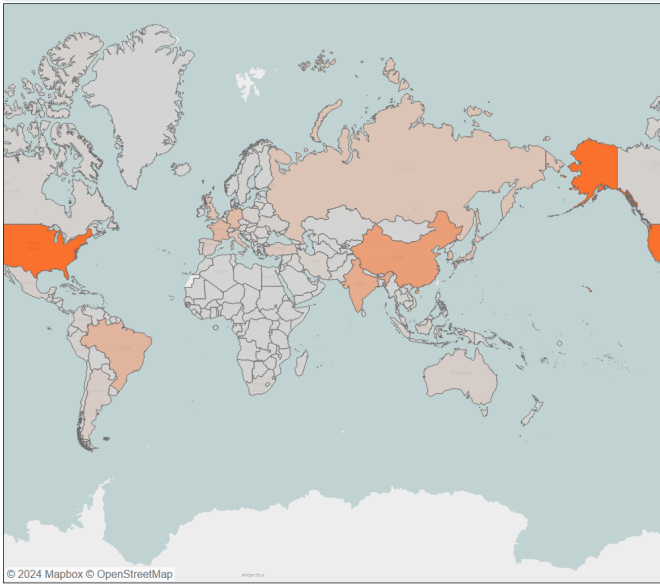
Fig. 3: COVID-19 Cases World Map

The line graph (Figure 4) depicts the trends in COVID-19 infections and deaths; specifically it shows the occurrence of death in early 2021 because there were no vaccines nor treatment options available. With the emergence of the Omicron variation, there was a spike in cases in early 2022, but the number of deaths continued to be low. The years 2023 and 2024 show how both the number of cases and fatalities dropped which indicates good controlling measures in place in the nation and the existence of herd immunity.
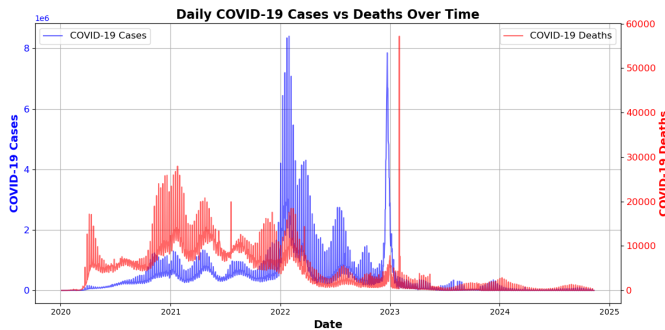
Fig. 4: Daily COVID-19 Cases vs Deaths Over Time

Further comparing COVID-19 cases and deaths within WHO regions (Figure 5) shows some important trends. The Europe and the Americas region had the greatest number of infected persons and deaths perhaps due to healthcare outcome differences or outbreak intensity. The Western Pacific Region had more cases but less deaths probably due to better containment or health care interventions. Sout-East Asia (SEAR); Eastern Mediterranean Area (EMR); and Africa (AFR) had lower cases and deaths in general.
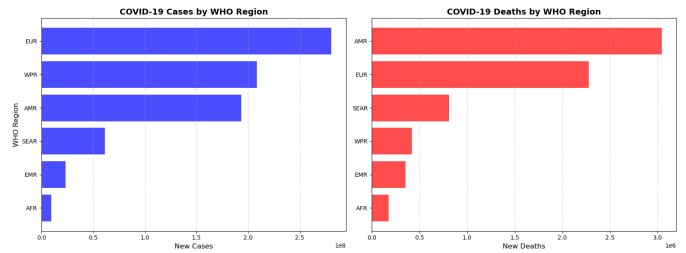

Fig. 5: COVID-19 Cases & Deaths by WHO Region

Figure 6 shows that the United States had the highest number of COVID-19 cases and fatalities, indicating that the pandemic had a significant impact on the nation. China and India were next in line for cases, and Brazil, India, Russia, and Mexico were among the top nations in terms of deaths. This implies that the pandemic was controlled in relation to the country's demography, healthcare's level and containment policies.
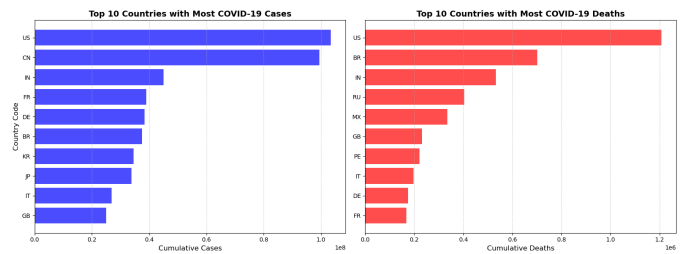

Fig. 6: Top 10 Countries with Most COVID-19 Cases & Deaths

The heatmaps (Figure 7) confirm that the numbers of daily COVID-19 infections aggravated in 2022 possibly due to the emergence of Omicron while the deaths peaked in early 2021, this being prior to significant vaccination of the population. Both the number of infections and the number of deaths significantly reduced from the year 2023 onwards implying success of vaccination programs, better medical treatment and efficient containment measures taken around the world.
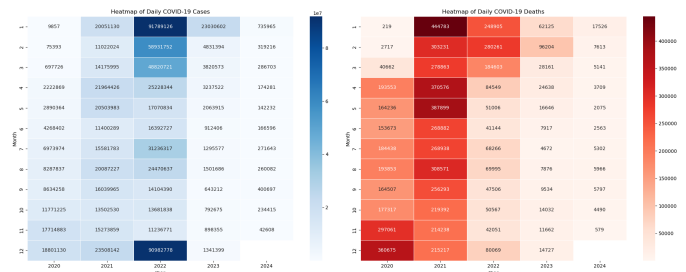

Fig. 7: Heatmap of COVID-19 Cases & Deaths

According to the bubble chart in Figure 8, there seem to be differences in vaccine coverage level across WHO regions. The Western Pacific and South-East Asia regions are at the top with 81% of their populations having received at least one dose, while those in the Americas trail at 68%.

According to the report Africa has a coverage of hardly 44% while Europe and Eastern Mediterranean regions cover 64% and 56% respectively. These variations are a consequence of the regional health care structures and differential access to vaccination.
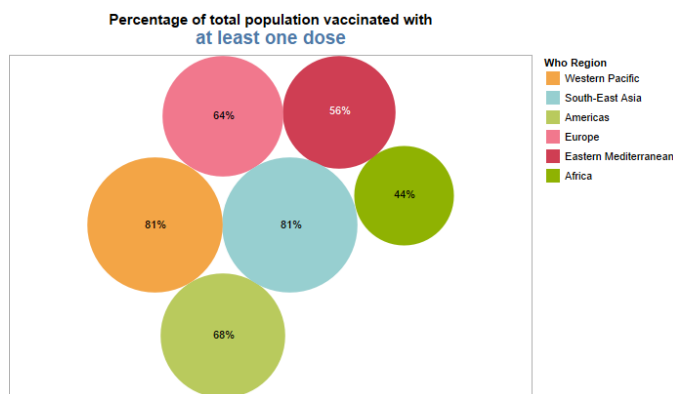

Fig. 8: Percentage of total population vaccinated with at least one dose

According to Figure 9, there are large disparities in the scope of the booster vaccination. The Western Pacific region leads this race with 49%, while followed by Europe with 44%. South-East Asia and The Americas come next with moderate coverage (37% and 30% respectively). On the contrary, Booster doses are the least common in Africa (12%) as well as the Eastern Mediterranean (24%). This gap emphasizes the difficulties with access, resource availability, and trust in public policies that have resulted in the rate of booster doses being taken up at a much slower rate than that of the primary vaccinations.
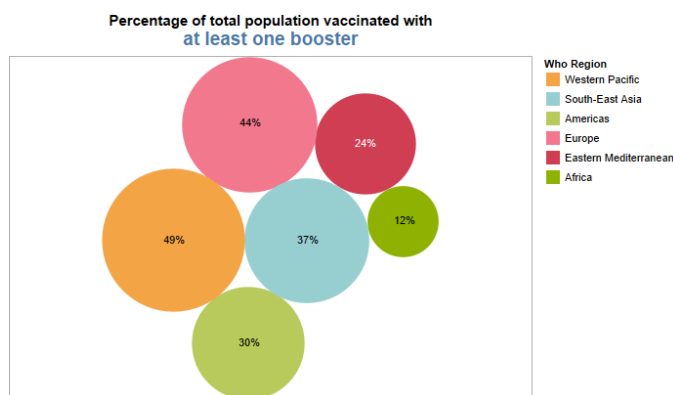

Fig. 9: Percentage of total population vaccinated with at least one booster

The pie chart (Figure 10) depicts successful doses where the lead is taken by Pfizer BioNTech with 25.7% of vaccinations, followed by AstraZeneca with 21.7%, Moderna at 19.1%, and Janssen Pharmaceuticals at. 18.5%. Beijing Bio-Institute Biological Products (CNBG) accounts for 14.9%. The overall vaccination scheme was led by Western companies, with Pfizer, AstraZeneca, and Moderna holding the largest shares.
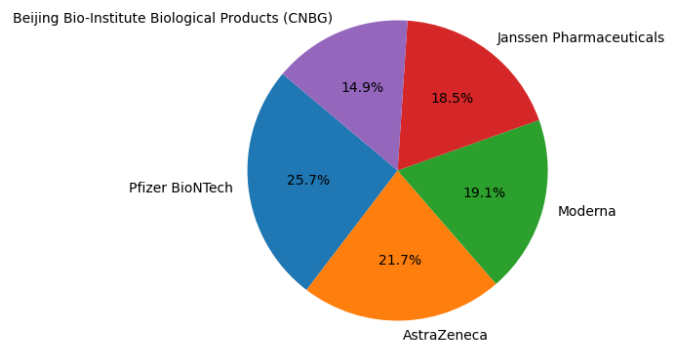

Fig. 10: Top 5 Vaccinations Companies

As we can see in Figure 11, the stock prices of leading pharmaceutical companies increased and fluctuated greatly during the Covid outbreak. Moderna (MRNA) recorded the highest level because of its major participation in the creation of the COVID-19 vaccine. Pfizer BioNTech (BNTX) also had significant growth, while Johnson & Johnson (JNJ) and Astrazeneca (AZN) exhibited slight but steady growth. These trends show how investors had trust in the companies which rendered vaccine and services during the outbreak.
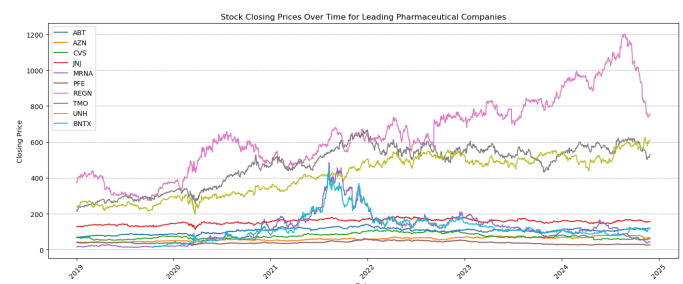

Fig. 11: Stock Closing Prices Over Time for Leading Pharmaceutical Companies

As noted in Figure 12, stock prices during the outbreak of COVID 19 as well as after the outbreak, stock prices of BioNTech (BNTX) and Moderna (MRNA) surged heavily as they were the primary manufacturers of vaccine. Another company that fared well was Regeneron (REGN which possibly due to its role in the treatment of COVID 19. At the same time, Johnson and Johnson (JNJ) and AstraZeneca (AZN) improved moderately. These developments illustrate the fact that the pharmaceutical sector, and especially those companies that produce vaccines and treatments, was greatly affected by the pandemic.
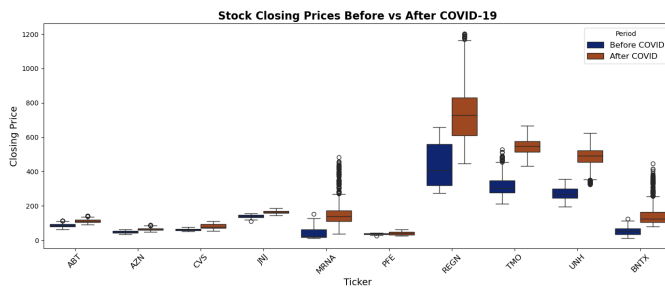
Fig. 12: Stock Closing Prices Before vs After COVID-19

In figure 13, it shows the growth of GDP within a certain year range. GDP growth of the world plunged around 5.5%, from a low growth of 2% in 2019 to -3.5% . This was the result of economic disruptions such as lockdowns and less movement activities during the pandemic. Although the drop in 2008 is not related to this project, it is due to the global financial crisis.
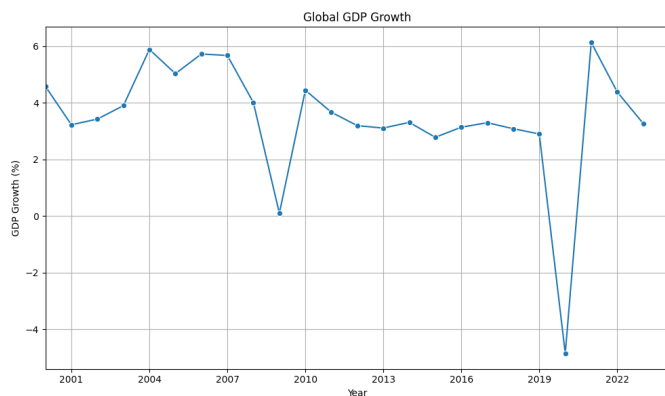

Fig. 13: Global GDP growth over the years

The graph (Figure 14) illustrates how the unemployment rate changed overtime globally, there was an increase in the year 2002 to 2021 unemployment rate when it rose from 7.1% to 8.1% with the increased joblessness caused by business shutdowns, loss and economic disruption in the global market due to the COVID-19 pandemic, however unemployment rate fell drastically in subsequent years and reached 6.8% in this case, pandemic economy impact made rounds of drastic effect on economy when lockdown ended and businesses started opening again with various operating capacities.
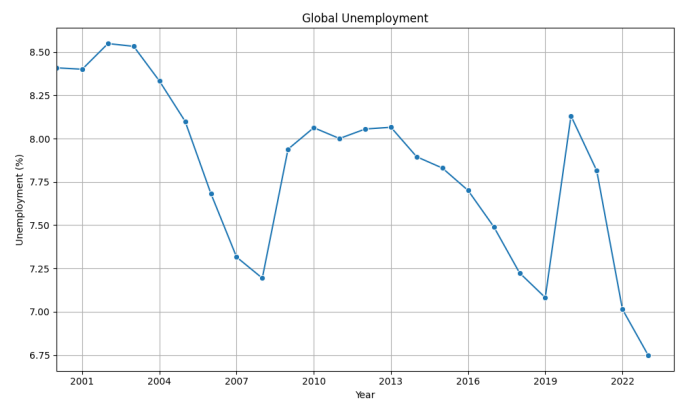

Fig. 14: Global unemployment over the years

In figure 15, it shows the average percentage of government debt of all countries throughout the years. Global government debt spiked significantly, jumping by 15% of GDP from 65% of GDP in 2019 to 80% in 2020, because of boosted spending on health systems and economic-boosting measures. By 2022, debt slightly declined to 70% as economies began to recover, reflecting the fiscal impact of the pandemic.
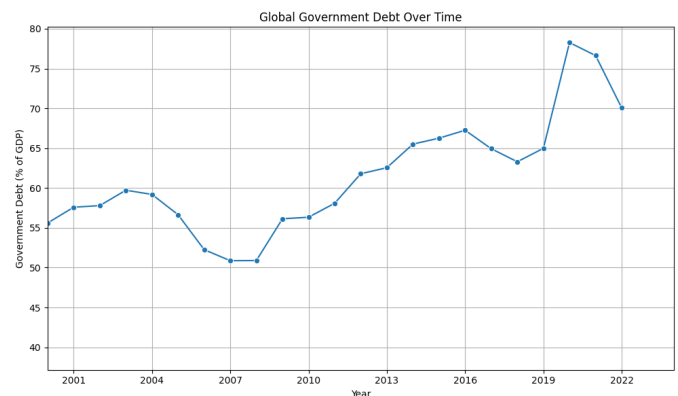

Fig. 15: Global governemnt debt over the years

Figure 16 shows the correlation between GDP growth, unemployment, and government debt. The correlation of GDP growth and unemployment indicators is very low, a value of 0.03, implying that there are no connections at all. On the other hand, the correlation between GDP growth and government debt, -0.24, does indicate a slight negative relationship, i.e. having a larger government debt correlates with lower GDP growth.
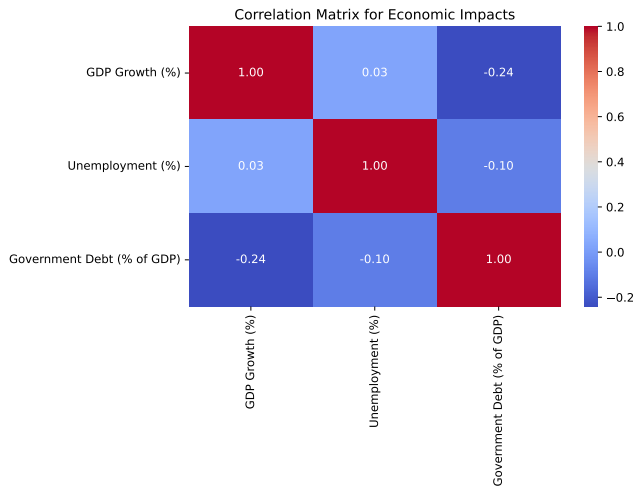
Fig. 16: Correlation Matrix for impacts

The stacked bar chart in figure 17 highlights the impact of COVID-19 in 2020. It is significant to note that debts incurred by the government takes the largest proportion which rather understandable considering the amount of resources spent by the countries during the pandemic on health systems and economic assistance. Moreover, it is also worth noting that the proportion of GDP growth rate decreased significantly, and the unemployment grew. Hence, this provides an additional sense of how the government debt was one of the hardest hit measurements during this time, showing the cost that countries had to bear to try and curb the impact of the pandemic.
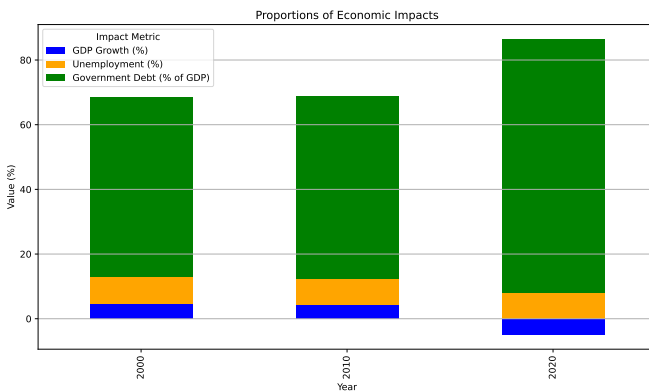

Fig. 17: Proportions of the three impacts

VI. Conclusions and Future Work

The case study underscores the deleterious consequences of COVID-19, the vaccination performances, and the economic indicators. World GDP experienced a significant drop in the year 2020 primarily due to the enforcing of lockdown policies, pandemic-induced export and import shortages, and a decrease in economic activities. The pandemic witnessed an increase in mass unemployment while the government debt escalated massively as the nations introduced fiscal policies to restore their economies. Vaccination campaigns turned out to be useful in saving lives after 2021, but regional disparities in the coverage of covid vaccines and boosters by the WHO showed that there are still many barriers to equitable healthcare distribution and infrastructure.

The results also pointed on a high level of confidence of the investors in pharmaceutical companies like Moderna and BioNTech whose shares were increasing in value because they were in the centre of vaccine developments. Other companies such as AstraZeneca and Johnson & Johnson registered a slow growth while Regeneron had a good performance owing to its role in the management of COVID-19 patients. Such observations underline the impact of creativeness, fields of medicine, and the stock market during times of global instability.

For future work, it would be appropriate to proceed further with the more regional approach towards the study of health system responses and the econometric modelling of sectoral effects. An example would be an evaluation on the effects of healthcare measures between American region and the European region to discover which nation did better. Further research is also needed to understand the long-term effects of rising government debt and unemployment post-pandemic. Furthermore, the implementation of machine learning models should allow forecasting the consequences of overseas pandemics on economic indicators, vaccine rates and the stock market.

References

[1] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, and L. Rodés-Guirao, "A global database of COVID-19 vaccinations," Nature Human Behaviour, vol. 5, no. 7, pp. 947–953, 2021. [Online]. Available: https://www.nature.com/articles/s41562-021-01122-8

[2] J. Khubchandani, S. Sharma, J. H. Price, M. J. Wiblishauser, M. Sharma, and F. J. Webb, "COVID-19 vaccination hesitancy in the United States: A rapid national assessment," Journal of Community Health, vol. 46, pp. 270–277, 2022. [Online]. Available: https://research.ebsco.com/c/x47ol5/viewer/pdf/qub32fuolf

[3] T. T. K. Oanh, "The impact of COVID-19 vaccination on stock market: Is there any difference between developed and developing countries?," Heliyon, vol. 8, no. 9, p. e10718, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405844022020060

[4] . E. Gagnon, S. B. Kamin, and J. Kearns, "The impact of the COVID-19 pandemic on global GDP growth," Journal of the Japanese and International Economies, vol. 68, p. 101252, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0889158323000138?

[5] D. L. Blustein et al., "Unemployment in the time of COVID-19: A research agenda," Journal of Vocational Behavior, vol. 119, p. 103436, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001879120300610