DOI: 10.58168/MoInSyTe2024_25-36
УДК 004.9

DATA ANALYSIS OF COVID-19 USING PYTHON

E.A. Anikeev[1], H.M. Firoz[1]

[1]Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The article concentrates on the data analysis of COVID-19 using the Python programming language. The goal of the work is to analyze the COVID-19 data and find out which countries and which continents are most affected by COVID-19. Besides, find out which vaccine for coronavirus is most commonly used. People have been more affected by the COVID-19 variant. Comparing the overall status of coronavirus by country and continent until May 25, 2022. As well as, we focused on time-series data analysis and visualization for COVID-19. We analyzed and visualized mortality, total recoveries, total cases, and active cases in our paper. Finally, this article provides a good analyze and visualization of COVID-19 very clearly through programming algorithms, calculations, statistics, and various libraries and codes. Through the observation, analysis, and testing of all the data, we have come to the conclusion that, through country-based, continent-based, time-based, variant, vaccination review and analysis on COVID-19, what is the overall condition of the COVID-19 epidemic from its onset until May 25, 2022?

Keywords: COVID-19, analysis, death, mortality, confirmed cases, active cases, recovered, vaccination, COVID-19 variant.

АНАЛИЗ ДАННЫХ COVID-19 С ИСПОЛЬЗОВАНИЕМ PYTHON

Е.А. Аникеев[1], Х.М. Фироз[1]

[1]ФГБОУ ВО «Воронежский государственный лесотехнический университет имени Г.Ф. Морозова»

Аннотация. Статья посвящена анализу данных о COVID-19 с использованием языка программирования Python. Цель работы — проанализировать данные о COVID-19 и выяснить, какие страны и какие континенты больше всего пострадали от COVID-19. Кроме того, узнайте, какая вакцина от коронавируса используется чаще всего. Люди больше пострадали от варианта COVID-19. Сравнение общего статуса коронавируса по странам и континентам до 25 мая 2022 года. Кроме того, мы сосредоточились на анализе и визуализации временных рядов данных о COVID-19. В нашей статье мы проанализировали и визуализировали смертность, общее количество выздоровлений, общее количество случаев и активных случаев. Наконец, в этой

статье представлен хороший анализ и очень четкая визуализация COVID-19 с помощью алгоритмов программирования, расчетов, статистики, различных библиотек и кодов. Путем наблюдения, анализа и тестирования всех данных мы пришли к выводу, что с помощью анализа и анализа вакцинации против COVID-19 в зависимости от страны, континента, времени, варианта, каково общее состояние? эпидемии COVID-19 с момента ее начала до 25 мая 2022 года?

Ключевые слова: COVID-19, анализ, смерть, смертность, подтвержденные случаи, активные случаи, выздоровевшие, вакцинация, вариант COVID-19.

Introduction

Many people in the world today are familiar with the COVID-19, which started its journey from Wuhan, City of China, in December 2019 and was temporarily named “2019-nCoV” [2, 5]. On February 11, 2020, the World Health Organization (WHO) named this disease “Coronavirus Disease 2019 (COVID-19)” and the International Committee on Taxonomy of Viruses (ICTV) named the virus “severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)”. On March 11, 2020, WHO declared the COVID-19 outbreaks as a pandemic [3, 4]. The virus can spread from an infected person’s mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe.

In this article, we implement an algorithm for analyzing COVID-19. Then, we used three data sources from those sources. For example, Worldometer, GitHub, and Kaggle. After collecting all the data, we have stored it in a folder for easy access. Then we have read all the data with Python. During the data analysis process, we prepared and cleaned the data for further analysis. The article highlights the development of in the structure of COVID-19 data analysis. Finally, described and analyze the data using the Python programming language.

Algorithm implement: To implement this algorithm implement, we used three different data sources and then collected all the data. After collecting all the data, we grouped and filtered it for better analysis. For any good data analysis, 'Static & Interactive’, ‘Multi-dimensional arrays & Statistical’, ‘Timeline Analysis," and ‘Map Analysis" are very important. So for my COVID-19 data analysis, we used those steps. Those steps provide me with a better visualization and graphical output. Then we were able to come to a conclusion by analyzing all this information (Figure 1).

Development the structure of COVID-19 data analysis: In Figure 2, I provide the development of the structure of the information system on COVID-19. First, we choose my operating system and hardware, then we install the programming language, IDE, and other necessary software. After successfully completing those steps, we collected

and prepared data for analysis. To do data analysis, we imported some libraries and finally showed the result.

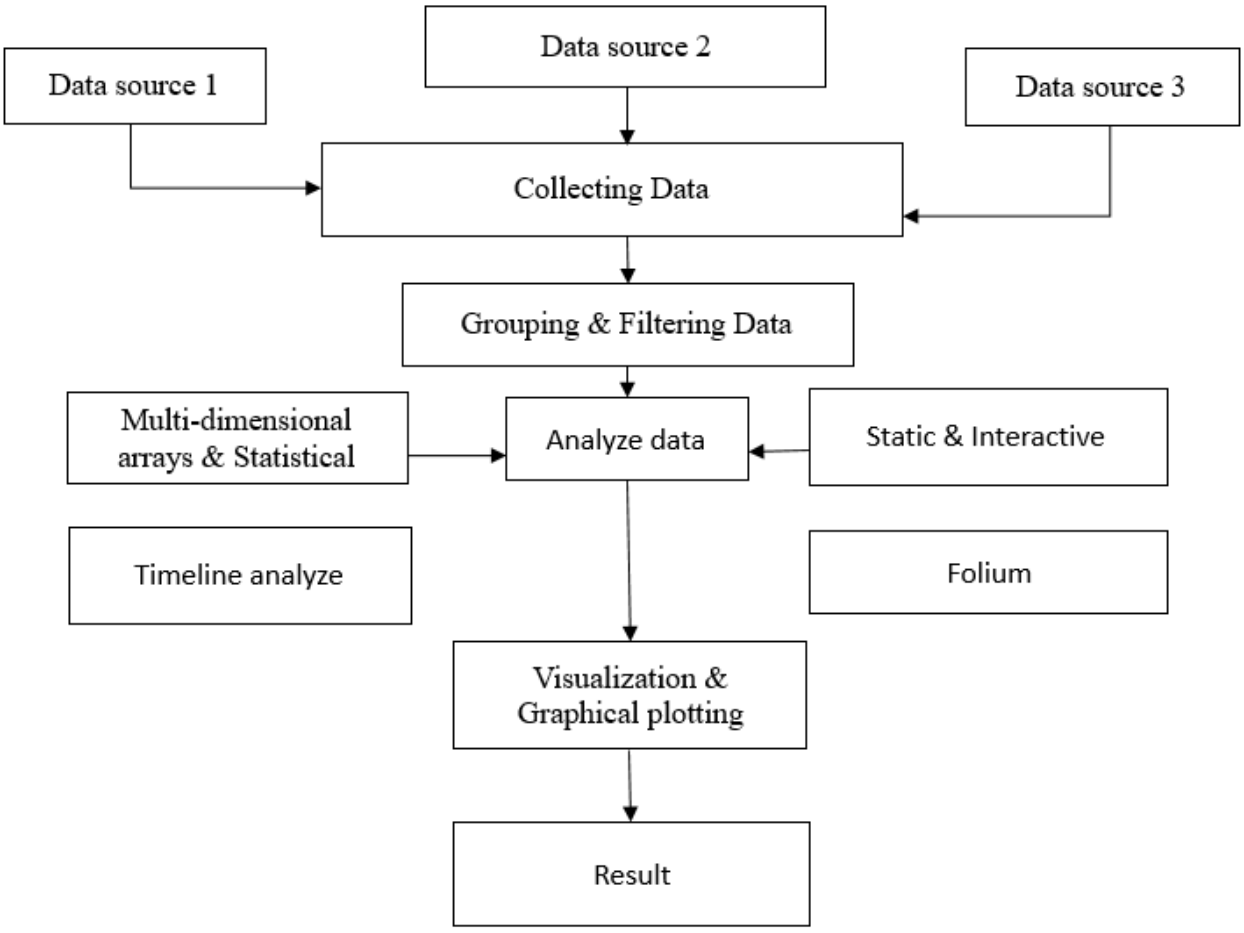

Figure 1 – Algorithm implement.

COVID-19 data analysis and visualization: Below, we will analyze and visualize all our data using Python and Jupiter notebooks.

If we analyze figure 3, as of May 2022, we can see that the highest number of people infected by COVID-19 in the United States is about 86 million people. The second and third places in the world are India and Brazil. Respectively, 43 and 31 million people are affected by the virus, and the least affected in Saint-Helena are only 2 people [1]. Developed countries have the highest number of countries in this list of 20 countries. The number of patients identified in Russia and South Korea was equal to 18 million.

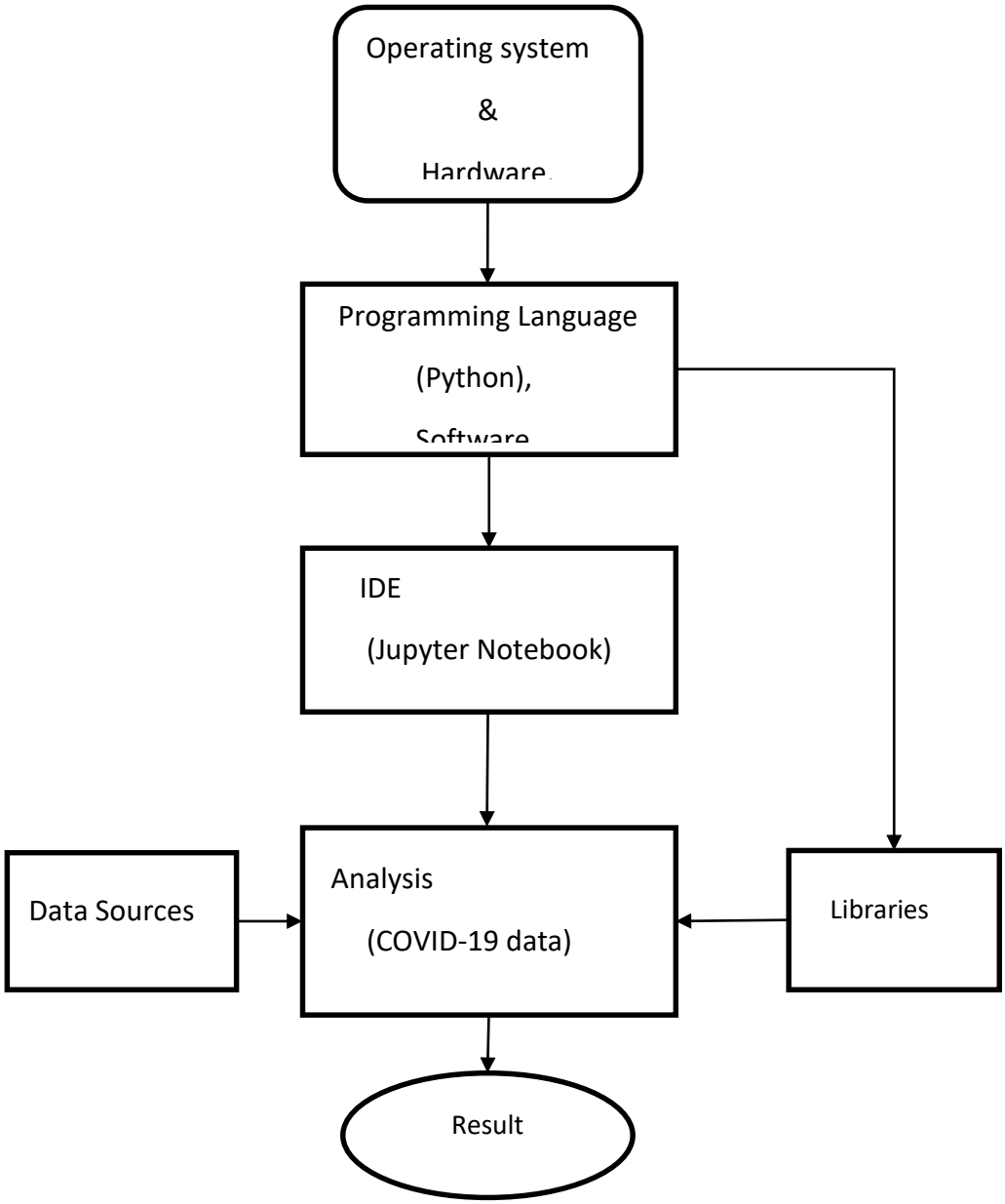

Figure 2 – Development the structure of COVID-19

On the other hand, in addition to being the most infected, the highest number of deaths is in the United States. More than 1 million people have died here. The deaths in Brazil, India, and Russia were 665955, 524607, and 378516, respectively (Figgure 4). If we compare top 25 country the least people dead in Canada and between the world only 5-10 people dead in Western Sahara [1].

Figure 3 – Total confirm cases measurement of top 20 countries

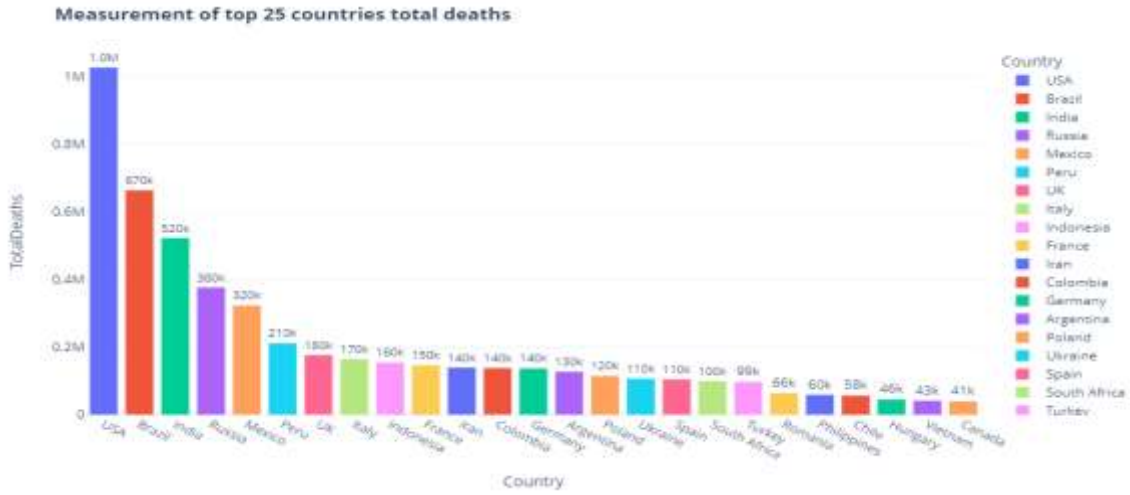

Figure 4 – Number of confirmed dead by 25 countries

A comparison between Italy and Hungary, two of the largest countries in Europe, shows that there is a big difference in the number of deaths. More than 150,000 people have died in France, Indonesia, and the UK from the deadly virus. Mexico and Peru, two of the most important countries in North America and South America, have the highest mortality rates. The death toll in South Africa and Turkey was almost equal to 100,000 (Figure 4).

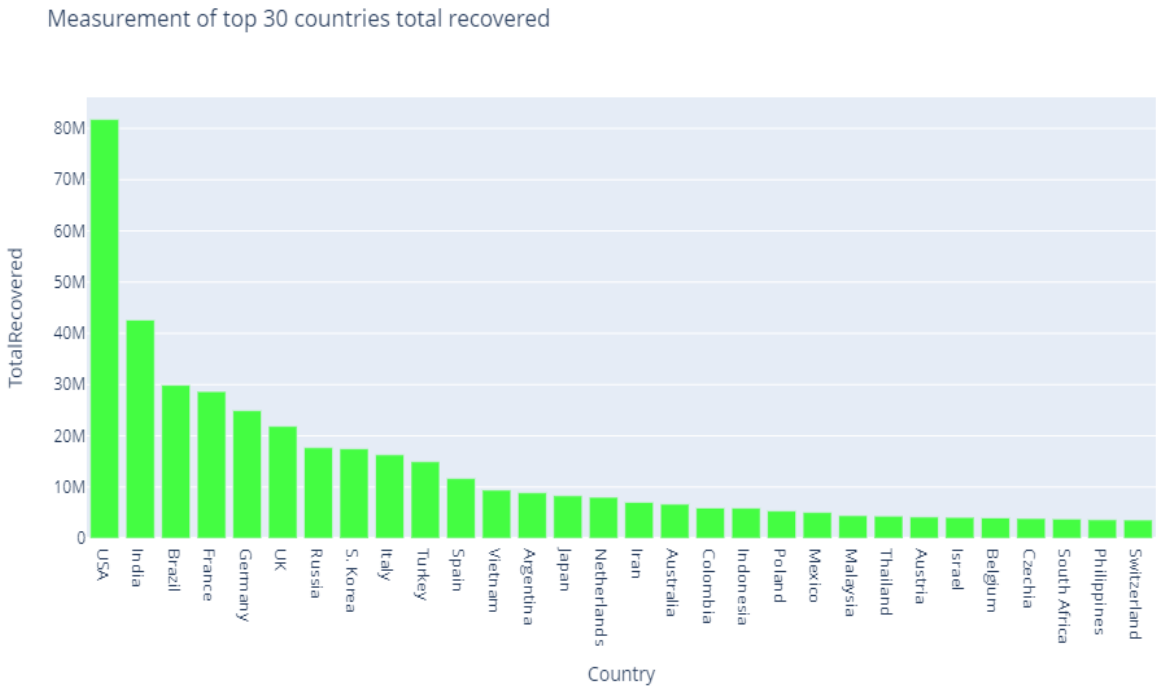

Figure 5 – Recovered case by country-wise

As of May 25, 2022, approximately 82 million people in the United States have recovered from the COVID-19 virus. That is a milestone. In many European countries, including Russia and India, the recovery rate is relatively satisfactory. In Japan, Netherland and Australia, mortality and the total number of confirmed cases were low, but the number of recovered patients was very high. By this figure, over 8 million people in Japan, the Netherlands, and over 6.5 million in Australia have recovered from the coronavirus. In India, Brazil, France, Germany, and the UK had 42.60, 29.88, 28.65, and 24.93 million recovery patients, respectively. Interestingly, the number of infected patients and deaths in the USA, India, and Brazil is as high as the number of recovered cases. Recovery in Malaysia, Thailand, Austria, and Israel was satisfactory, although the number of infected patients was much lower (Figure 5).

With the number 6 figure, we can see that, although mortality, infection, and recovery rates were relatively similar in early 2020, differences were seen towards the end of the same year. Towards the end of 2021, the confirmed cases surpassed the 500 million mark. Between 2020 and 2021, the mortality rate increased dramatically and became widespread. But at the present time, it is much less.

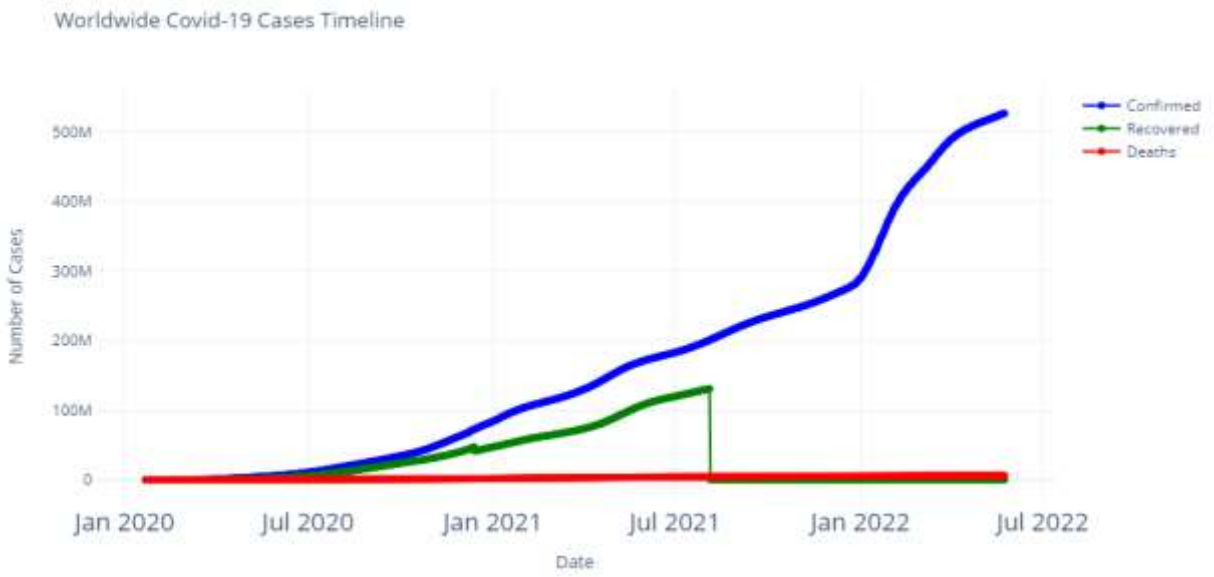

Figure 6 – Time series analysis of COVID-19

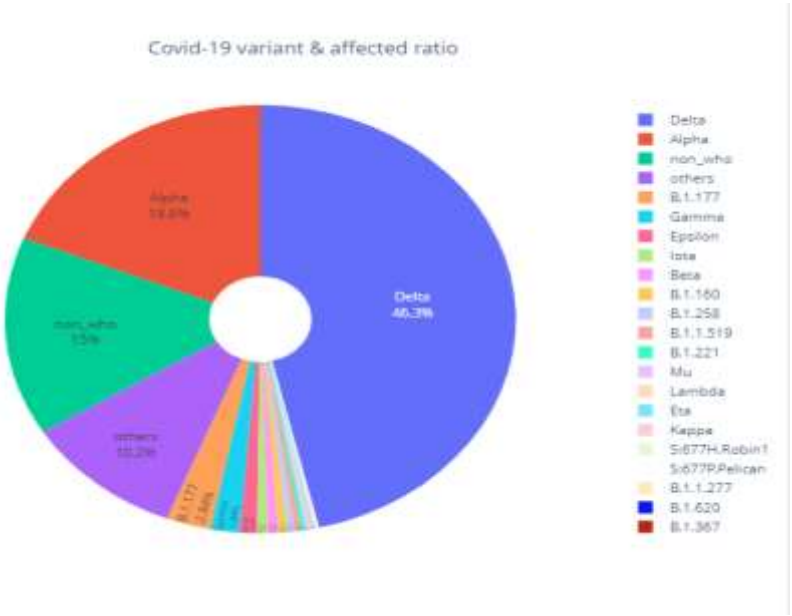

Figure 7 – Covid-19 variant & affected ratio by the variant

The most dangerous variant of COVID-19 is the Delta variant, which has affected 46.3 percent of the world's population. On the other hand, by the Alpha and non-WHO variants, 18.8% and 15% of people are affected, respectively.

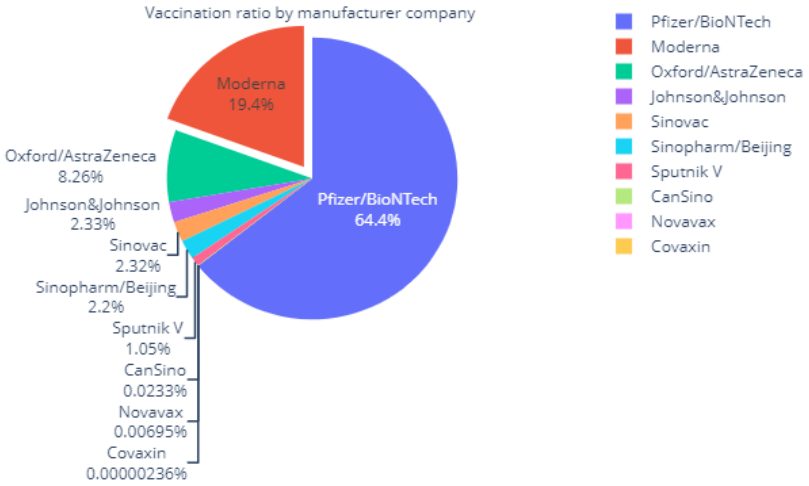

Figure 8 – COVID-19 vaccine manufacturer company

We can see that in Figure 8, there are several vaccine manufacturing companies. For example, Pfizer–BioNTech, Oxford–AstraZeneca, Sinopharm, Moderna, Sputnik V, and so on. Pfizer–BioNTech and Moderna are the largest vaccine manufacturers.

Pfizer–BioNTech alone produced more than half the vaccines, while Moderna and Oxford–AstraZeneca produced 19% and 8.26%. Sputnik V produced only 1.05% of the total vaccine manufacturer company. About 2.30% vaccination has been complete by the sinovac and johoson&johoson.

Observing this figure, we will be able to get an overall idea of the COVID-19 epidemic. From these 3 pie charts, we can easily observe that the African continent is in a better position than Europe, America, and Asia in spite of low vaccination. As of May 25, 2022, In Africa only 3.98% people are dead by the virus whereas Europe and America was 72.9% of total deaths. In the total cases Europe are 31.3% and Asia are 28.7% whereas Africa only 2.99%. On the other hand, Asia and the American continent are far ahead in terms of vaccination. 53.9% of the population's total population is fully vaccinated in Asia. In Europe and North America, 17.5% and 15.8% of people are fully vaccinated (figure 9).

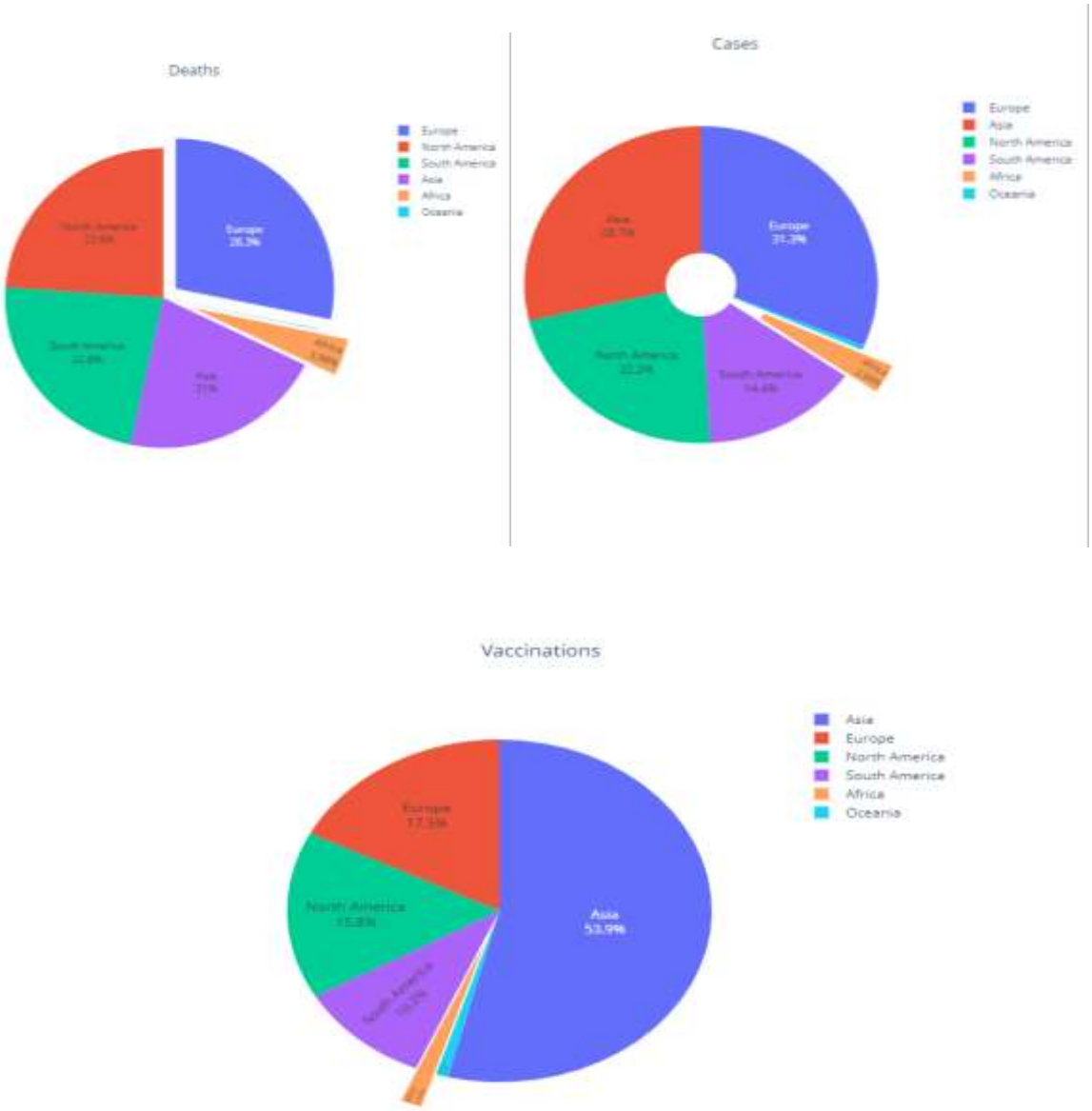

Figure 9 – Continent wise Covid-19 overall analysis

Conclusion

In conclusion, the result is 6.27 M people dead worldwide from COVID-19 as of May 25, 2022. From this data analysis, the total number of infected people is 525.28 million, and the number of active patients until May 2022 is about 519 million, which is increasing day by day. A total of 23491 million Corona patients have recovered with and without vaccination. Of the coronavirus variants, people are mostly affected by the Delta variant, accounting for about 46.3% of the total.

Furthermore, when it comes to vaccination, people prefer Pfizer/BioNtech, and it is approximately 64.4%, whereas Russian Sputnik V is only 1.05%. Moderna and Oxford/Astrazeneca are in the second and third positions, respectively, whereas

Moderna manufactures 19.4% and Oxford/Astrazeneca manufacter 8.26% (Figure 8). Only 16.2% of people in low-income countries have received at least one dose [2]. On the other hand, 65.8% of the global population has received at least one dose of a coronavirus vaccine, 6.97 million of which are administered every day, and 11.80 billion doses have been administered globally. In the world, a total of 61% of the population is fully vaccinated, and 6.97 million are now administered each day. On the other hand, without overpopulation, undeveloped health facility, low vaccination rates (1.61%), Africans have a morality rate only 3.89%, as well as a confirmed case ratio about 2.99.

References

1. Data design: visualising quantities, locations, connections by Per Mollerup. London: Bloomsbury, 2015.001.4226 M6D2 (190991).

2. Chotirmall SH, Martinez FJ, Schumacker PT, Cooke CR, Seam N, Brochard L, et al. Life at the editorial "COVID frontline". The American Thoracic Society Journal Family. Am J Respir Crit Care Med 2020;201:1457–1459.

3. Harker JA, Johansson C. Rapidly deployable mouse models of SARS-CoV-2 infection add flexibility to the COVID-19 toolbox. Am J Respir Cell Mol Biol 2021; 64:7–9.

4. Agusti A, Sibila O, Casas-Recasens S, Mendoza N, Perea L, Lopez-Giraldo A, et al. Molecular interactions of SARS-CoV-2 in lung tissue of patients with chronic obstructive pulmonary disease. Ann Am Thorac Soc 2021;18:1922–1924.

5. "COVID-19 CORONAVIRUS PANDEMIC". Worldometer. Archived from the original on 3 February 2020. Retrieved 2 December 2020.

6. Coronavirus disease (COVID-19): Variants of SARS-COV-2. - URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-%28covid-19%29-variants-of-sars-cov-2?gclid=CjwKCAiAgbiQBhAHEiwAuQ6BkpohLNMPN9q2HvuT-JogXwYMVsySdzRTsZohqrsrfoOue2NSin811yhoCWtkQAvD_BwE (date of access: 01.03.2022).

7. Coronavirus (COVID-19) Vaccinations.- URL: https://ourworldindata.org/covid-vaccinations?country=OWID_WRL (date of access: 03.03.2022).

8. Daniel, S. J. (2020). Education and the COVID-19 pandemic. Prospects, 49(1), 91-96.

9. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., ... & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama, 323(11), 1061-1069.

10. Wikipedia: COVID-19.- URL: https://en.wikipedia.org/wiki/COVID-19#:~:text=COVID%2D19%20transmits%20when,distances%2C%20particularly%20indoors. (date of access: 06.03.2022).

11. Boccia, S., Ricciardi, W., & Ioannidis, J. P. (2020). What other countries can learn from Italy during the COVID-19 pandemic. JAMA internal medicine, 180(7), 927-928.

12. Poluektov A.V., Makarenko F.V., Yagodkin A.S. The use of third-party libraries when writing programs for processing statistical data // Modeling of systems and processes. - 2022. – Vol. 15, No. 2. – pp. 33-41.

Список литературы

1. Data design: visualising quantities, locations, connections by Per Mollerup. London: Bloomsbury, 2015.001.4226 M6D2 (190991).

2. Chotirmall SH, Martinez FJ, Schumacker PT, Cooke CR, Seam N, Brochard L, et al. Life at the editorial “COVID frontline”. The American Thoracic Society Journal Family. Am J Respir Crit Care Med 2020;201:1457–1459.

3. Harker JA, Johansson C. Rapidly deployable mouse models of SARS-CoV-2 infection add flexibility to the COVID-19 toolbox. Am J Respir Cell Mol Biol 2021; 64:7–9.

4. Agusti A, Sibila O, Casas-Recasens S, Mendoza N, Perea L, Lopez-Giraldo A, et al. Molecular interactions of SARS-CoV-2 in lung tissue of patients with chronic obstructive pulmonary disease. Ann Am Thorac Soc 2021; 18:1922–1924.

5. "COVID-19 CORONAVIRUS PANDEMIC". Worldometer. Archived from the original on 3 February 2020. Retrieved 2 December 2020.

6. Coronavirus disease (COVID-19): Variants of SARS-COV-2. - URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-%28covid-19%29-variants-of-sars-cov-2?gclid=CjwKCAiAgbiQBhAHEiwAuQ6BkpohLNMPN9q2HvuT-JogXwYMVsySdzRTsZohqrsrfoOue2NSin811yhoCWtkQAvD_BwE (date of access: 01.03.2022).

7. Coronavirus (COVID-19) Vaccinations.- URL: https://ourworldindata.org/covid-vaccinations?country=OWID_WRL (date of access: 03.03.2022).

8. Daniel, S. J. (2020). Education and the COVID-19 pandemic. Prospects, 49(1), 91-96.

9. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., ... & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama, 323(11), 1061-1069.

10. Wikipedia: COVID-19.- URL: https://en.wikipedia.org/wiki/COVID-19#:~:text=COVID%2D19%20transmits%20when,distances%2C%20particularly%20indoors. (date of access: 06.03.2022).

11. Boccia, S., Ricciardi, W., & Ioannidis, J. P. (2020). What other countries can learn from Italy during the COVID-19 pandemic. JAMA internal medicine, 180(7), 927-928.

12. Poluektov A.V., Makarenko F.V., Yagodkin A.S. The use of third-party libraries when writing programs for processing statistical data // Modeling of systems and processes. - 2022. – Vol. 15, No. 2. – pp. 33-41.