Development of a Data Manager for Analyzing the Global Distribution of COVID-19 Vaccines Using Data from 'Our World in Data'

Leticia Santos Almeida[1*], Vinicius Marques Rocha[1], Afrânio Ferreira Evangelista[1], Bruna Aparecida Souza Machado[1]

[1]SENAI CIMATEC University Center; Salvador, Bahia, Brazil

The COVID-19 pandemic has triggered a global race for vaccination as a crucial preventive measure. Considering socioeconomic disparities and public policies, this study investigated the number and proportion of vaccinations in different countries. Using data from the "Our World in Data" platform, we collected vaccination information since December 2020, including population estimates. We selected the top 10 countries regarding vaccination and stratified them by income. SQL Server and PHPStorm were used to manage the data. The results demonstrated the evolution of vaccination up until March 2023. Easy access to this data allows us to investigate the effectiveness of government interventions in public health, using comprehensive datasets better to understand the progress and impact of vaccination efforts.

Keywords: COVID-19. Vaccination. Database. Public Policy.

The COVID-19 pandemic not only represented an unprecedented public health challenge but also triggered a global race towards vaccination, which is a crucial measure for preventing the severity of the disease and death and for controlling its transmission (WHO, 2020) [1]. The multifaceted impact of COVID-19 necessitated different approaches from local health authorities in various countries concerning the vaccination schedules adopted and the outcomes of these measures [2]. To analyze these approaches, obtaining vaccination data from reliable sources such as public health organizations, government agencies, and academic institutions is essential.

The "Our World in Data" platform (OWD, 2024) [3] offers a comprehensive compilation of vaccination data from around the world since December 2020 and continues to update it. Therefore, this work aimed to organize and develop queries that can extract and facilitate research and analysis of the "Our World in Data" database, allowing for an in-depth examination of the distribution of vaccines administered in different countries from the onset of vaccinations to the present.

Materials and Methods

Data was collected on the online platform "Our World in Data", which combines official vaccination numbers since December 2020 with more recent data from government health agencies worldwide. This platform also provides population estimates for per capita metrics based on the United Nations World Population Prospects and categorizes population groups by income using the World Bank classification. To manage and organize the raw data spreadsheet from the "Our World in Data" platform, we utilized the software "SQL Server Management Studio" (Microsoft). The following SQL queries were created to facilitate the development and debugging of the code:

1. Total number of vaccinations per country.
2. The proportion of vaccinations with an initial and complete schedule.
3. Division of data by income group.

These queries were designed to extract relevant information for a comprehensive analysis of global

Received on 17 February 2024; revised 14 May 2024.
Address for correspondence: Leticia Santos Almeida. Avenida Orlando Gomes, 1845, Piatã. Salvador, bahia, Brazil. Zipcode 41650-010. E-mail: santosalmeida.leticia@outlook.com.

J Bioeng. Tech. Health 2024;7(2):202-204

vaccination distribution, considering various socioeconomic factors and public health policies.

Results and Discussion

Data collection on the "Our World in Data" platform was conducted on March 12, 2023.

This effort produced a comprehensive table containing vaccination data from December 2020 to the collection date, encompassing 243 countries, counties, and delimited territories. The dataset also included population estimates, income group classifications, and death counts, resulting in 12,816,911 entries or raw data points. Due to the substantial volume of data, analysis could not be performed directly in the Excel spreadsheet where the data was initially downloaded.

Consequently, it was necessary to employ a more robust tool capable of efficiently selecting data according to the parameters set for evaluation. Using SQL Server Management Studio, we significantly enhanced data processing efficiency. The generated SQL code was designed to handle large datasets, enabling rapid and effective data operations such as filtering, cleaning, transformation, and analysis.

This approach allowed us to extract and analyze the necessary information with greater accuracy and speed, providing valuable insights into the global distribution of COVID-19 vaccines. The processed data highlighted vital trends and disparities in vaccination rates across different countries and income groups, facilitating a better understanding of the impact of socioeconomic factors and public health policies on vaccination efforts.

The Excel spreadsheet was converted into a database using SQL Server Management Studio (Microsoft) to provide integrated data validation. This integration included all the spreadsheet information into a database table, facilitating efficient management of the high volume of data. The database table contained selected data entries such as the number of vaccinations per country, the number of people with initial and complete vaccination schedules, income groups, population numbers, and dates.

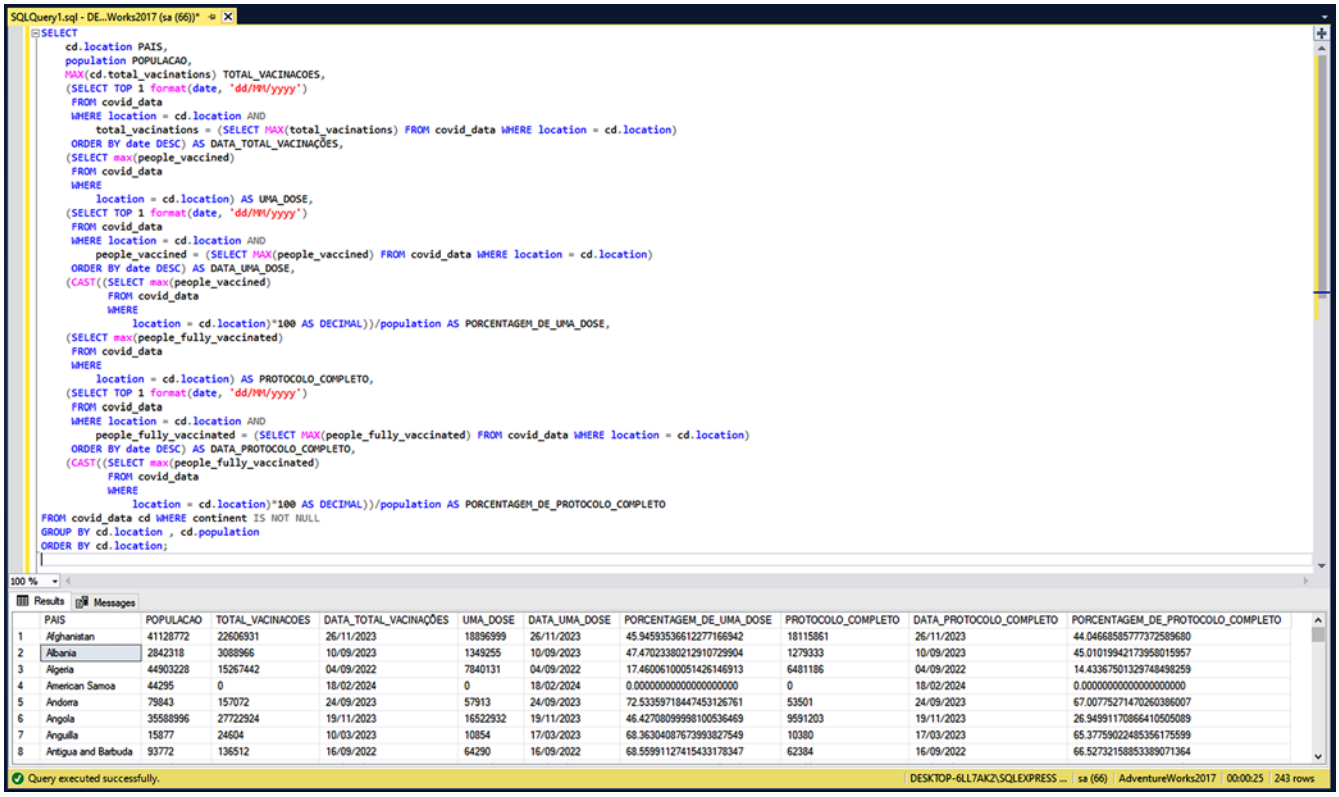
For the development of the code, selected SQL queries were added. These queries included commands to retrieve, modify, or manipulate specific data such as the total number of vaccinations per country, the proportion of vaccinations with initial and complete schedules, and divisions by income group[4].

After stratifying the data, they select the top 10 countries regarding the number and proportion of vaccinations [(#vaccines/population) × 100] with initial and complete schedules as possible. Additionally, the data was stratified by income group, covering the period from December 2020 to the latest consultation on March 12, 2023 (Figure 1). This approach allows for ongoing updates; the same code can generate updated analyses if the spreadsheet is updated with more recent data.

This method enabled us to handle the large dataset efficiently, ensuring accurate and timely data analysis. The processed data provided valuable insights into the global distribution of COVID-19 vaccines, highlighting key trends and disparities across different countries and income groups. This facilitated a deeper understanding of the impact of socioeconomic factors and public health policies on vaccination efforts, contributing to more informed decision-making in public health management.

All this analysis enabled efficient data extraction on a vast scale. Another significant advantage was the feasibility of data manipulation and the potential for integrating automation into the process. Once processed, the data could be easily exported for later presentation in tables. This step was crucial for preparing the collected data for subsequent analysis of the vaccination situation in different countries. The generated tables may aid future studies in comparing public policy adoption for vaccination and the resulting vaccination rates. Additionally, they serve as tools for visualizing the progression of vaccination about the implementation or lack of public policies promoting vaccine availability and uptake. Furthermore, new queries can be added to the code to complement the study.

Figure 1. Data stratified by income group (December 2020 to March 12, 2023).



```
SELECT
    cd.location PAIS,
    population POPULACAO,
    MAX(cd.total_vacinations) TOTAL_VACINACOES,
    (SELECT TOP 1 format(date, 'dd/MM/yyyy')
     FROM covid_data
     WHERE location = cd.location AND
         total_vacinations = (SELECT MAX(total_vacinations) FROM covid_data WHERE location = cd.location)
     ORDER BY date DESC) AS DATA_TOTAL_VACINAÇÕES,
    (SELECT max(people_vaccined)
     FROM covid_data
     WHERE
         location = cd.location) AS UMA_DOSE,
    (SELECT TOP 1 format(date, 'dd/MM/yyyy')
     FROM covid_data
     WHERE location = cd.location AND
         people_vaccined = (SELECT MAX(people_vaccined) FROM covid_data WHERE location = cd.location)
     ORDER BY date DESC) AS DATA_UMA_DOSE,
    (CAST((SELECT max(people_vaccined)
           FROM covid_data
           WHERE
               location = cd.location)*100 AS DECIMAL))/population AS PORCENTAGEM_DE_UMA_DOSE,
    (SELECT max(people_fully_vaccinated)
     FROM covid_data
     WHERE
         location = cd.location) AS PROTOCOLO_COMPLETO,
    (SELECT TOP 1 format(date, 'dd/MM/yyyy')
     FROM covid_data
     WHERE location = cd.location AND
         people_fully_vaccinated = (SELECT MAX(people_fully_vaccinated) FROM covid_data WHERE location = cd.location)
     ORDER BY date DESC) AS DATA_PROTOCOLO_COMPLETO,
    (CAST((SELECT max(people_fully_vaccinated)
           FROM covid_data
           WHERE
               location = cd.location)*100 AS DECIMAL))/population AS PORCENTAGEM_DE_PROTOCOLO_COMPLETO
FROM covid_data cd WHERE continent IS NOT NULL
GROUP BY cd.location , cd.population
ORDER BY cd.location;
```

	PAIS	POPULACAO	TOTAL_VACINACOES	DATA_TOTAL_VACINAÇÕES	UMA_DOSE	DATA_UMA_DOSE	PORCENTAGEM_DE_UMA_DOSE	PROTOCOLO_COMPLETO	DATA_PROTOCOLO_COMPLETO	PORCENTAGEM_DE_PROTOCOLO_COMPLETO
1	Afghanistan	41128772	22606931	26/11/2023	18896999	26/11/2023	45.945935366122777166942	18115861	26/11/2023	44.046685877737258968O
2	Albania	2842318	3088966	10/09/2023	1349255	10/09/2023	47.470233802129107299O4	1279333	10/09/2023	45.010199421739580159S7
3	Algeria	44903228	15267442	04/09/2022	7840131	04/09/2022	17.460061000514261469I3	6481186	04/09/2022	14.433675013297484982S9
4	American Samoa	44295	0	18/02/2024	0	18/02/2024	0.000000000000000000O0	0	18/02/2024	0.000000000000000000O0
5	Andorra	79843	157072	24/09/2023	57913	24/09/2023	72.533597184474531267G1	53501	24/09/2023	67.007752714702603860O7
6	Angola	35588996	27722924	19/11/2023	16522932	19/11/2023	46.427080999981005364G9	9591203	19/11/2023	26.949911708664105050B9
7	Anguilla	15877	24604	10/03/2023	10854	17/03/2023	68.363040876739938275G9	10380	17/03/2023	65.377590224853561755B9
8	Antigua and Barbuda	93772	136512	16/09/2022	64290	16/09/2022	68.559911274154331783A7	62384	16/09/2022	66.527321588533890713G4

Query executed successfully. DESKTOP-6LL7AK2\SQLEXPRESS ... sa (66) AdventureWorks2017 00:00:25 243 rows

Conclusion

Data stratification facilitated easier access, enabling the following analysis stage concerning vaccination coverage and political responses in different sociopolitical contexts. By leveraging comprehensive datasets from "Our World in Data" containing epidemiological metrics and vaccination progress, future work can continue to link policy interventions and elucidate the differential impacts of government strategies on public health outcomes.

References

1. WHO. Como funcionam as vacinas - Organização Mundial da Saúde. 2020. Available at: <https://www.who.int/pt/news-room/feature-stories/detail/how-do-vaccines-work? gad_source=1&gclid=EAIaIQobChMIn9z177j3hAMVZkVIAB1aFAJ9EAAYASAAEgJzqvD_BwE>. Acessed on February 15, 2024.
2. Ma M et al. Comparison of COVID-19 vaccine policies and their effectiveness in Korea, Japan, and Singapore. International Journal for Equity in Health 2023;22(1). Available at: <https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-023-02034-x>. Acessed on February 15, 2024.
3. OWD – Our World in Data. Coronavirus (COVID-19) Vaccinations. Available at: <https://ourworldindata.org/covid-vaccinations>. Acessed on February 17, 2024.
4. Afrati F, Chirkova R. Answering queries using views. Cham: Springer International Publishing, 2018.