Comparing the reach and effect of different vaccines on the most recent wave of COVID-19

Saloni Gupta
Masters in Applied Computing
University of Windsor
gupta66@uwindsor.ca

Surya Chandiramouli Subhashini
Masters in Applied Computing
University of Windsor
chandir@uwindsor.ca

Vaishnavi Venkatesh
Masters in Applied Computing
University of Windsor
venkatev@uwindsor.ca

***Abstract*— Vaccines have been a great boon to human life, especially in the ongoing Covid-19 pandemic. Vaccines such as Pfizer, Moderna have given a big sigh of relief to people and have helped to curb the death rates in countries. This project is research on such vaccines and the study of their effectiveness and reach on different countries by querying the datasets from web using Google Bigquery.**

Keywords—Vaccine, Covid-19, Google Bigquery, Death rates

I. INTRODUCTION

Since the first wave of COVID 19 in 2020, our lives have been impacted in ways we could not have imagined. Economies around the world took a hit and life came to a halt. Vaccines have been a boon to humanity in restricting the spread of the virus. The medical fraternity has done an amazing job in developing the vaccines. During the first wave of Covid when there were no vaccines available, the death rate throughout the world was rapidly increasing. Countries then came up with their vaccines like AstraZeneca, Moderna, Pfizer which were administered to the people dose wise. We got the motivation for this idea from a paper where they had compared the impact of COVID-19 on countries with robust and poor health care systems. We decided to build on that and talk about the vaccines and its effects. We will discuss and analyze the reach, and performance of different vaccines across the world. The world's ongoing war against COVID-19 makes this problem interesting and important. To give our analysis, we will query three separate and huge databases which are being uploaded every day. We have used Google Bigquery which is a tool for writing queries and is helpful for analyzing the results as well.

II. RELATED WORK

A. Background study

We dug deep to look for research papers providing a similar insight to what we are trying to achieve. The few papers that we went through are: Efficacy and Safety of NVX-CoV2373 in Adults in the United States and Mexico[1], Preliminary Findings of mRNA Covid-19 Vaccine Safety in Pregnant Persons[2], Assessing the Effectiveness of BNT162b2 and ChAdOx1nCoV-19 COVID-19 Vaccination in Prevention of Hospitalisations in Elderly and Frail Adults[3], Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting[4], A Global Map of COVID-19 Vaccine Acceptance Rates per Country: An Updated Concise Narrative Review[5].

III. METHODOLOGIES

A. Google Big query

We have queried two large databases for the analysis. This will require efficient and complex queries. We will use Google Big Query to make our conclusions. Big Query is Google's serverless, highly scalable, enterprise data warehouse designed to make all our data analysts productive at an unmatched price-performance. Its server-less model lets us focus on writing efficient queries and making analysis rather than having to worry about the infrastructure. The reason for its drastically fast performance can be found in the following two core technologies: Columnar Storage allows very high compression ratio and scan throughput. Tree Structure for processing queries and fetching required data.

We will use Big Query to load our data sets, write queries and perform the analysis. We have the following reasons to do the same. The data can be easily populated with the Big Query and schema structure is automatically detected. Analysis is easily, and the processed data can be exported. We can design dashboards, generate reports and graphs to visualize the data and gain a better understanding. We created a new project on the Google Console and loaded our first dataset to the Big Query Dashboard. Big Query autodetected the schema of the table so we could directly start executing our queries.

B. Datasets

We have taken the two datasets from Kaggle[6] and WHO[7]. The first dataset consists of the different kinds of vaccines administered to people in different countries of the world and the second dataset consists of the deathrates in the countries. We have used the datasets to determine the vaccine efficiency.

C. Datapreprocessing

We have used the two tools Winpure Clean and Match and Google Dataprep clean our dataset. These tools helped us remove any unnecessary tabs, spaces, commas etc. in the values of our table. To further improve efficiency, we executed manually curated queries to remove rows that would not add value to our results. For instance, we deleted the rows which had 0 count of vaccinations as that would not affect the total sum.

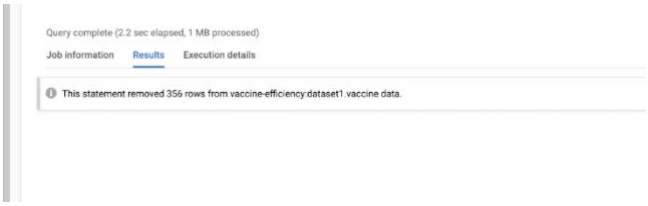

Fig 1

D. *Google Data Studio*

Google Data Studio is a part of the Google Big Query platform that is used for effective data visualization and information extraction. Through this platform, we can create and share reports with others, easily connect and import a variety of different data sources, visualize, and connect to highly configurable charts and tables.

IV. PROPOSED MODEL

We executed a few queries to get a sense of the data available to us. We first used the vaccine manufacturer dataset and executed a query to give us a count of how many doses of each vaccine was administered in each country. Below is a screenshot of the result of our query. Here, the result shows the number of doses of the vaccines namely, Moderna, Johnson&Johnson, Pfizer, and Modern, were administered in Austria.

Row	location	vaccine	Dose
1	Austria	Moderna	41434823
2	Austria	Johnson&Johnson	9621065
3	Austria	Pfizer/BioNTech	323178470
4	Austria	Oxford/AstraZeneca	60870114
5	Belgium	Moderna	54826390
6	Belgium	Johnson&Johnson	12123373
7	Belgium	Pfizer/BioNTech	412221661
8	Belgium	Oxford/AstraZeneca	99928485

Fig 2

We used Google Data studio to explore and visualize the results that we got from the queries.

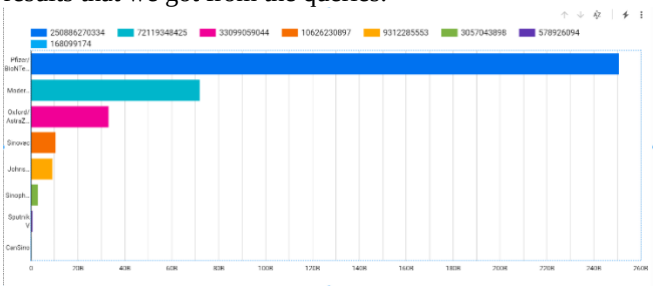

Fig 3

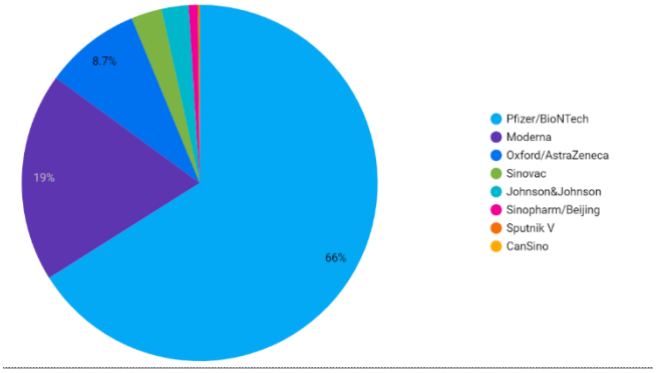

Fig 4

From the above pie chart and bar graph we could see that Pfizer dominated in its reach with over 66% doses administered all over the world. Moderna comes in second by a huge difference with 19% of the total doses. Sinovac and Johnson&Johnson had almost the same number of doses. This is our findings from our first dataset. For the next part of our project, we have taken a dataset from WHO, which consists of number of deaths due to covid for each country in the world. The dataset covers the data from 2020 through 2021 till 2022 March. Since it's a live dataset, the values are registered only until the day we took the dataset. The first step was to clean the dataset and bring the data in co-relation with the previous dataset of total vaccinations. So, we wrote a manual query to remove the data which is not needed. The below query depicts the data removed by comparing the two datasets. We got back the result of the query and were left with countries that were common in both the datasets. The other countries as the data were not sufficient.

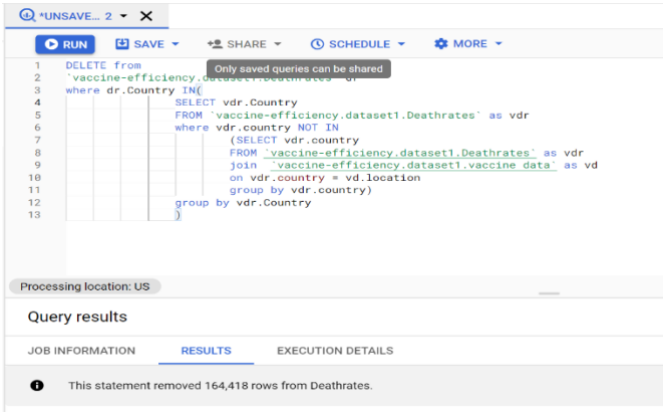

Fig 5

First, we grouped each vaccine by the location and found the total number of vaccinations for each vaccine across all the countries in the dataset1 which is the vaccine dataset. For each of the vaccine we sorted them by their count in descending order (Highest number of vaccines in a particular country would be shown first).

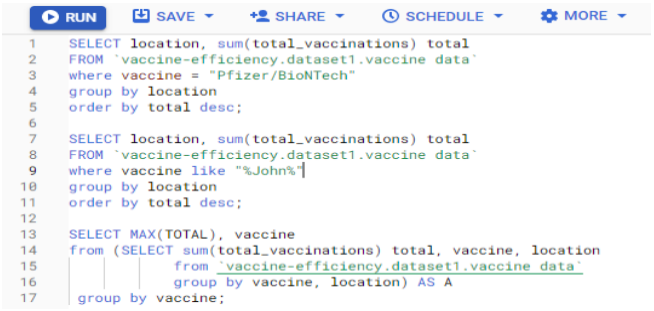

Fig 6

Next, we summed up the Total vaccination for each vaccine and queried only the Maximum of does for each vaccine, location, and total vaccination count.

Fig 7

For each country, we grouped the corresponding vaccine and its total count (Vaccination count) in both the datasets (Vaccine and deathrates)

```
SELECT sum(vd.total_vaccinations) total, vd.vaccine, vd.location
        FROM `vaccine-efficiency.dataset1.Deathrates` as dr
        JOIN `vaccine-efficiency.dataset1.vaccine data` as vd
        ON dr.Country = vd.location
        group by vd.vaccine, vd.location
        order by total desc
```

Fig 8

Finally, we queried only the Top 5 countries in which the death rates were more, and we extracted the dates in which the deaths were reported (The wave of 2020, 2021 and 2022) and we have compared the death rates in all the 3 years.

```
SELECT
  EXTRACT(year FROM dr.date_reported) AS year,
  SUM(dr.new_deaths) AS deaths,
  dr.country
FROM `vaccine-efficiency.dataset1.Deathrates` as dr
JOIN `vaccine-efficiency.dataset1.vaccine data` as vd
ON dr.Country = vd.location
where vd.location IN ('Germany', 'Italy', 'France', 'Chile', 'Peru', 'Romania')
GROUP BY country, EXTRACT(year FROM date_reported);
```

Fig 9

V. RESULTS AND CONCLUSION

From the two datasets that we have queried, we visualized the results shown as below in the graphs that depicts the number of deaths in Millions in 2020, 2021 and 2022 and how was the death rate in each year.

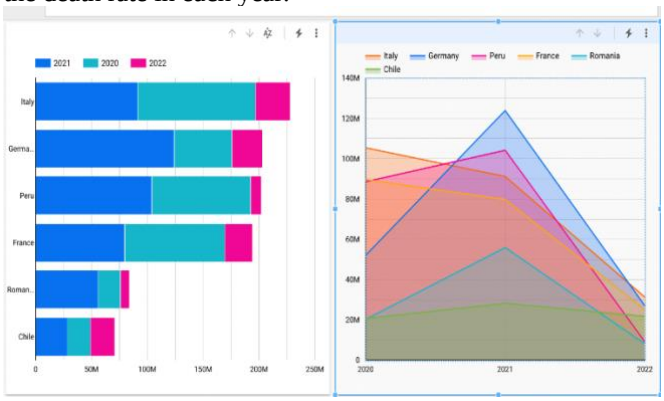

Fig 10

The above graph depicts the top 6 countries that were the worst affected in the recent wave of Covid-19. This is in accordance with our dataset and the countries that have been included in it. As we can clearly see from the graph, the death rates in 2020 were more in number as then the vaccines were not discovered.

```
SELECT location,
       vaccine,
       total_vaccinations
from `vaccine-efficiency.dataset1.vaccine data`
where (vaccine, total_vaccinations) = ANY(
          SELECT MAX(total_vaccinations) maximum, vaccine
          from `vaccine-efficiency.dataset1.vaccine data`
          group by vaccine
          );
```

In 2021 the rates were comparatively lower as the vaccines had been rolled out, but still the death rates were high. In 2022 (till March), the rates have dropped down drastically low as more people are being vaccinated. As for the country's comparison, according to this graph, Germany was the worst affected in the last wave of Covid-19 with more than 120M deaths followed by 102M in Peru [8], 90M in Italy, 80M in France, 55M in Romania, and 29M in Chile. The year 2021 was the worst hit, but after vaccine administration in the year 2022, the death rates too came down as low as 25M for Germany, 34M for France, 31M for Italy, 25M for Germany [9], 17M for Chile and, 8M for Peru and Romania respectively.

Long term changes in environmental pressure will depend crucially on their economic drivers and the regional impacts. For instance, some sectors like manufacturing and construction have been more affected than sectors like agriculture. The pandemic has triggered the largest and most widespread crisis in labour markets in recent history. It has caused major disruptions in the way we consume goods and services, with significant effects on occupations which is the social impact of this pandemic. Thus, to curb this pandemic and to bring things under control, vaccines are playing a major role and will continue to be the important factor in the hope of completely eradicating this deadly virus from the world.

VI. CHALLENGES AND FUTURE WORK

The first drawback with our results is that even though we have visualized our results using Google Data Studio, which is quite comprehensive, and it can be clearly understood by anyone [10], we will try to create a more clearer visualization platform like a live dashboard which would have easier for understanding. We would implement that feature in the future using Tableau visualization tool. Since the data keeps updating everyday it is difficult to feed in live data and query it on a regular basis. For the future we would try to implement the same project with a live database updating the data every day, so that more countries can be used for querying and the results can be more broadened and not just restricted to 6 countries.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Dr. Kalyani Selvarajah for her support and guidance throughout this semester and the project. And we would also thank all our team members for their efforts and cooperation, so our project could be completed and meet our initial goal.

REFERENCES

[1] Lisa M. Dunkle, M.D., Karen L. Kotloff, M.D., Cynthia L. Gay, M.D. “Efficacy and Safety of NVX-CoV2373 in Adults in the United States and Mexico”

[2] Tom T. Shimabukuro, M.D., Shin Y. Kim, M.P.H.,I. S. Jacobs and C. P. Bean, “Preliminary Findings of mRNA Covid-19 Vaccine Safety in Pregnant Persons”

[3] Dr Catherine Hyams PhD, “Assessing the Effectiveness of BNT162b2 and ChAdOx1nCoV-19 COVID-19 Vaccination in Prevention of Hospitalisations in Elderly and Frail Adults”

[4] Noam Barda, M.D., Noa Dagan, M.D., B.Sc., Eldad Kepten, Ph.D.,“Safety of the BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting”

[5] Choudhary Sobhan Shakeel and Saad Jawaid Khan,“A Global Map of COVID-19 Vaccine Acceptance Rates per Country: An Updated Concise Narrative Review“

[6] Covid19Dataset,”https://www.kaggle.com/datasets/imdevskp/corona-virus-report”

[7] WHO Mortality Database,”https://www.who.int/data/data-collection-tools/who-mortality-database”

[8] Covid: Why has Peru been so badly hit?,” https://www.bbc.com/news/world-latin-america-53150808”

[9] Pfizer vaccine is 95 per cent effective,” https://www.newscientist.com/article/2244804-covid-19-news-archive-pfizer-vaccine-is-95-per-cent-effective”

[10] Google Data Studio is a Good Start, But Still Have a Long Way to Go in Democratizing Analytics,” https://towardsdatascience.com/a-free-google-data-studio-is-a-good-start-but-we-still-have-a-long-way-to-go-in-democratizing-f22f461444e”