# Insight into current COVID-19 variants and data Science Applications

**Department of Computer Science**

## ABSTRACT

COVID-19 has undergone several mutations resulting in the emergence of new variants such as Alpha, Beta, Gamma, Delta, and Omicron. Data science plays a vital role in understanding the spread of COVID-19 and its variants. The emergence of new COVID-19 variants has raised concerns about the effectiveness of existing vaccines and treatments. This poster provides insights into the current COVID-19 variants and the data science applications used to monitor and understand their spread. We will showcase the use of SQL to manage and analyze genomic data of SARS-CoV-2 variants, Power BI for visualization and tracking of COVID-19 cases and deaths, and machine learning algorithms for variant classification and prediction of transmission. The poster highlights the importance of integrating data science tools in COVID-19 research for effective variant surveillance, risk assessment, and mitigation strategies.

## INTRODUCTION

COVID-19 is a respiratory illness caused by the SARS-CoV-2 virus that was first identified in Wuhan, China, in December 2019, and has since spread globally, leading to a pandemic. The virus mainly spreads through respiratory droplets when an infected person talks, coughs, or sneezes. Symptoms of COVID-19 can range from mild to severe and include fever, cough, fatigue, loss of taste or smell, and difficulty breathing. The healthcare system has been strained by the influx of COVID-19 patients, leading to shortages of personal protective equipment (PPE), hospital beds, and medical staff. The global economy has suffered due to the closure of businesses, travel restrictions, and decreased consumer spending. Governments worldwide have implemented various measures to slow the spread of the virus, including social distancing, wearing masks, and widespread vaccination efforts. Vaccination is a critical tool in ending the pandemic and returning to normalcy. The COVID-19 pandemic has highlighted the importance of global cooperation in responding to public health crises. International organizations such as the World Health Organization have played a critical role in coordinating global responses to the pandemic. The pandemic has also highlighted existing health disparities, particularly in vulnerable communities, and has emphasized the need for equitable access to healthcare. Overall, COVID-19 is a global health crisis that has had a profound impact on the world and will continue to shape our lives for years to come.

| | VOC Omicron GRA (B.1.1.529+BA.*) | 2-11 | 2022-12-18 | 2022-12-25 | 2023-01-01 | 2023-01-08 | 2023-01-15 | 2023-01-22 | 2023-01-29 | 2023-02-05 | 2023-02-12 | 2023-02-19 | 2023-02-26 | 2023-03-05 | 2023-03-12 | 2023-03-19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aruba | count | 11 | 20 | 31 | 28 | 37 | 24 | 25 | 9 | 8 | | | | | | |
| | total | 11 | 20 | 31 | 28 | 37 | 24 | 25 | 9 | 8 | | | | | | |
| Australia | count | 2,034 | 1,736 | 1,117 | 1,037 | 1,305 | 1,244 | 1,016 | 743 | 866 | 793 | 838 | 793 | 673 | 396 | 7 |
| | total | 2,038 | 1,737 | 1,118 | 1,039 | 1,307 | 1,246 | 1,018 | 743 | 870 | 794 | 841 | 812 | 702 | 418 | 8 |
| Austria | count | 2,712 | 2,934 | 2,689 | 2,491 | 2,328 | 2,306 | 2,067 | 2,262 | 2,551 | 2,650 | 2,955 | 2,866 | 2,357 | 1,555 | 13 |
| | total | 2,712 | 2,934 | 2,689 | 2,491 | 2,328 | 2,306 | 2,068 | 2,264 | 2,551 | 2,650 | 2,955 | 2,867 | 2,357 | 1,555 | 13 |
| Azerbaijan | count | | | | | | | | | | | | | | | |
| | total | | | | | | | | | | | | | | | |
| Bahrain | count | 6 | 76 | 17 | 44 | 48 | 83 | 86 | 101 | 111 | 48 | | | | | |
| | total | 6 | 76 | 17 | 44 | 48 | 86 | 89 | 101 | 111 | 50 | | | | | |
| Bangladesh | count | | | | 5 | 2 | 1 | | | | | | | | | |
| | total | | | | 5 | 2 | 1 | | | | | | | | | |
| Barbados | count | 2 | 1 | | 1 | 4 | 12 | 7 | 3 | 3 | | | | | | |
| | total | 2 | 1 | | 1 | 4 | 12 | 7 | 3 | 3 | | | | | | |
| Belarus | count | | | | | | | | | | | | | | | |
| | total | | | | | | | | | | | | | | | |
| Belgium | count | 452 | 465 | 377 | 249 | 336 | 240 | 206 | 169 | 250 | 360 | 290 | 295 | 366 | 184 | 27 |
| | total | 470 | 466 | 378 | 251 | 336 | 240 | 206 | 170 | 250 | 360 | 291 | 298 | 368 | 187 | 27 |
| Belize | count | | | | | | | | | | | | | | | |

**Figure 1-Raw Data**

Results | Messages

| | Id | Country | Month | Total count |
|---|---|---|---|---|
| 1 | 1 | Algeria | January | 2 |
| 2 | 2 | Algeria | January | 3 |
| 3 | 3 | Algeria | January | 5 |
| 4 | 4 | Algeria | January | 7 |
| 5 | 5 | Algeria | February | 1 |
| 6 | 6 | Algeria | February | 2 |
| 7 | 7 | American Samoa | January | 4 |
| 8 | 8 | Andorra | January | 3 |
| 9 | 9 | Anguilla | January | 1 |
| 10 | 10 | Antigua and Barbuda | January | 1 |

**Figure 2 – Prepared SQL Data**

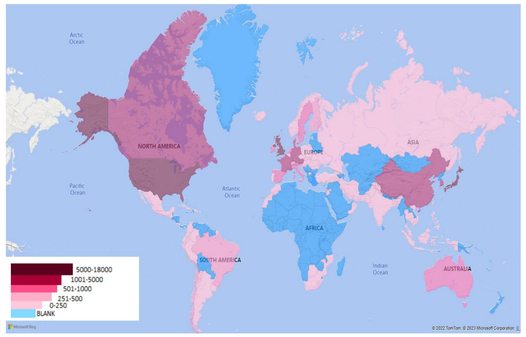


**Figure 3 – Map distribution of Omicron Variant**

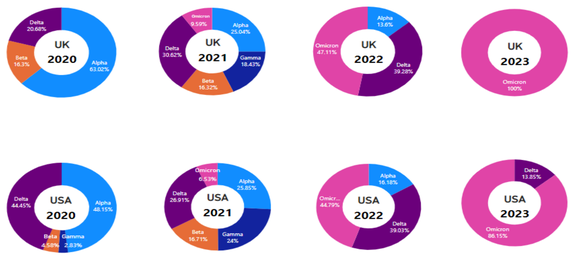


**Figure 4 –Illustration of Dominant Variant of UK and USA with respect to year 2020 -2023**

## Methodology

New variants of the COVID-19 virus have emerged, some of which are more contagious and potentially more deadly than the original strain. Analyzing the genetic changes of these variants is crucial in informing public health responses and vaccine development. SQL can be used to query databases of sequenced viral genomes to identify specific mutations associated with each variant and track the spread of the virus over time. Power BI can be used to visualize and analyze data on the spread of the virus and the effectiveness of public health interventions, integrating data from multiple sources to gain insights into how the virus is spreading and how well public health interventions are working. Machine learning techniques can be applied to COVID-19 data to predict future trends and identify factors that contribute to the spread of the virus. For example, by analyzing data on the number of COVID-19 cases and deaths over time, machine learning algorithms can predict future trends and identify patterns in the spread of the virus, helping public health officials prepare for future surges. These data science applications are crucial in informing public health responses and developing effective interventions to control the pandemic.

## CONCLUSION

As the COVID-19 pandemic continues, predicting future cases is difficult due to various factors affecting the virus spread. To address this, research on the mechanism of SARS-CoV-2 Spike protein binding with ACE2 using Machine Learning Techniques like Biotite, PyTorch, and Torch Drug can aid in understanding the binding behavior. GCN with mutual attention can be used to predict protein-protein interactions, leading to new treatments and drugs. Therefore, investing in further research on protein-protein interactions using Machine Learning Techniques is crucial for managing COVID-19 and other viral diseases. Data science also plays a critical role in tracking COVID-19 variants, guiding public health interventions and vaccine distribution.

KENNESAW STATE UNIVERSITY

Author(s): Ravi Potlapalli, Thanusha Sai Ande

Advisor(s): Dr.Chloe Yixin Xie