



# Corona Virus Analysis using SQL

# CONTENT

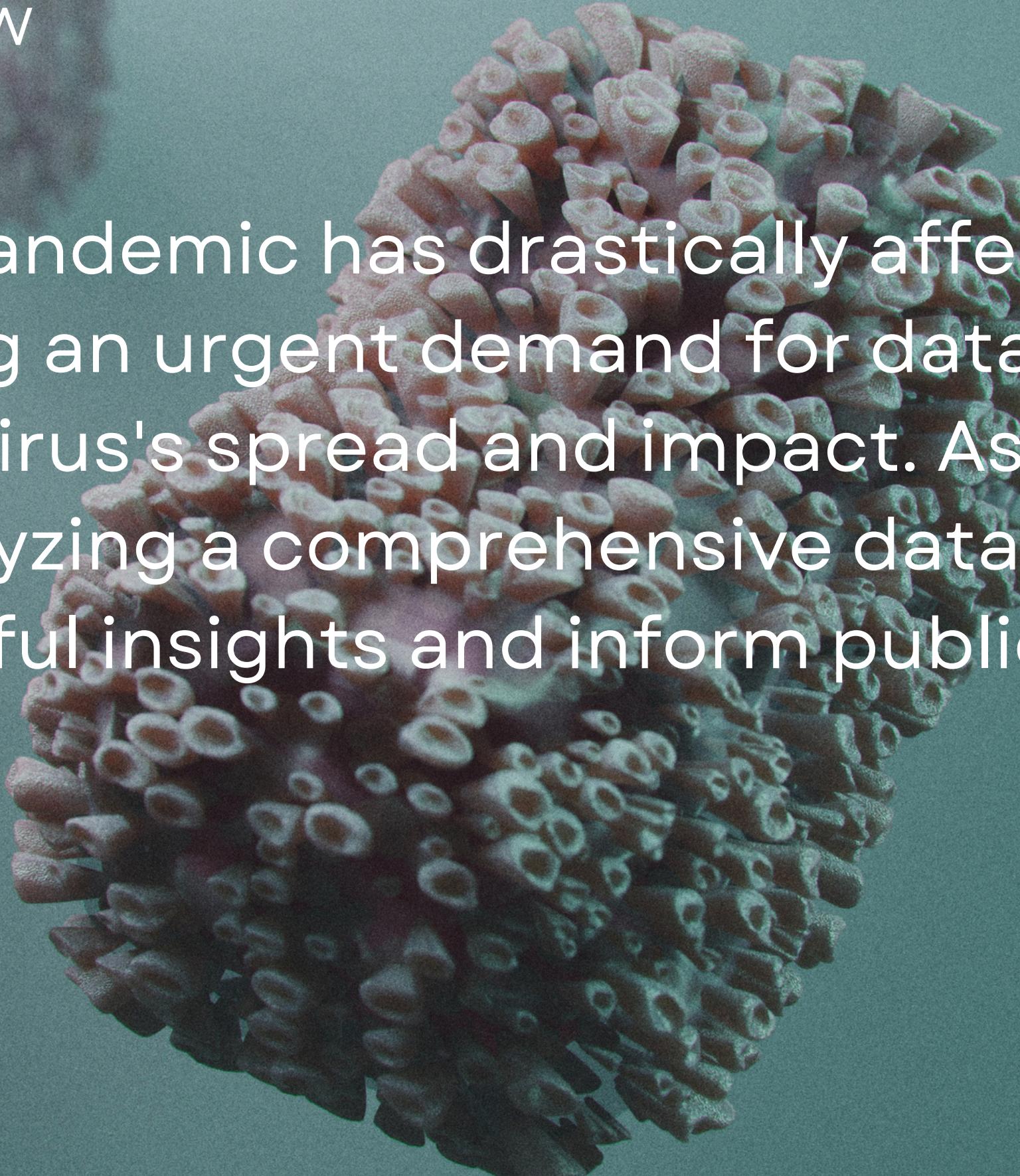
---

1.Overview

2.Description

3 Data Exploration and queries

## PROJECT OVERVIEW



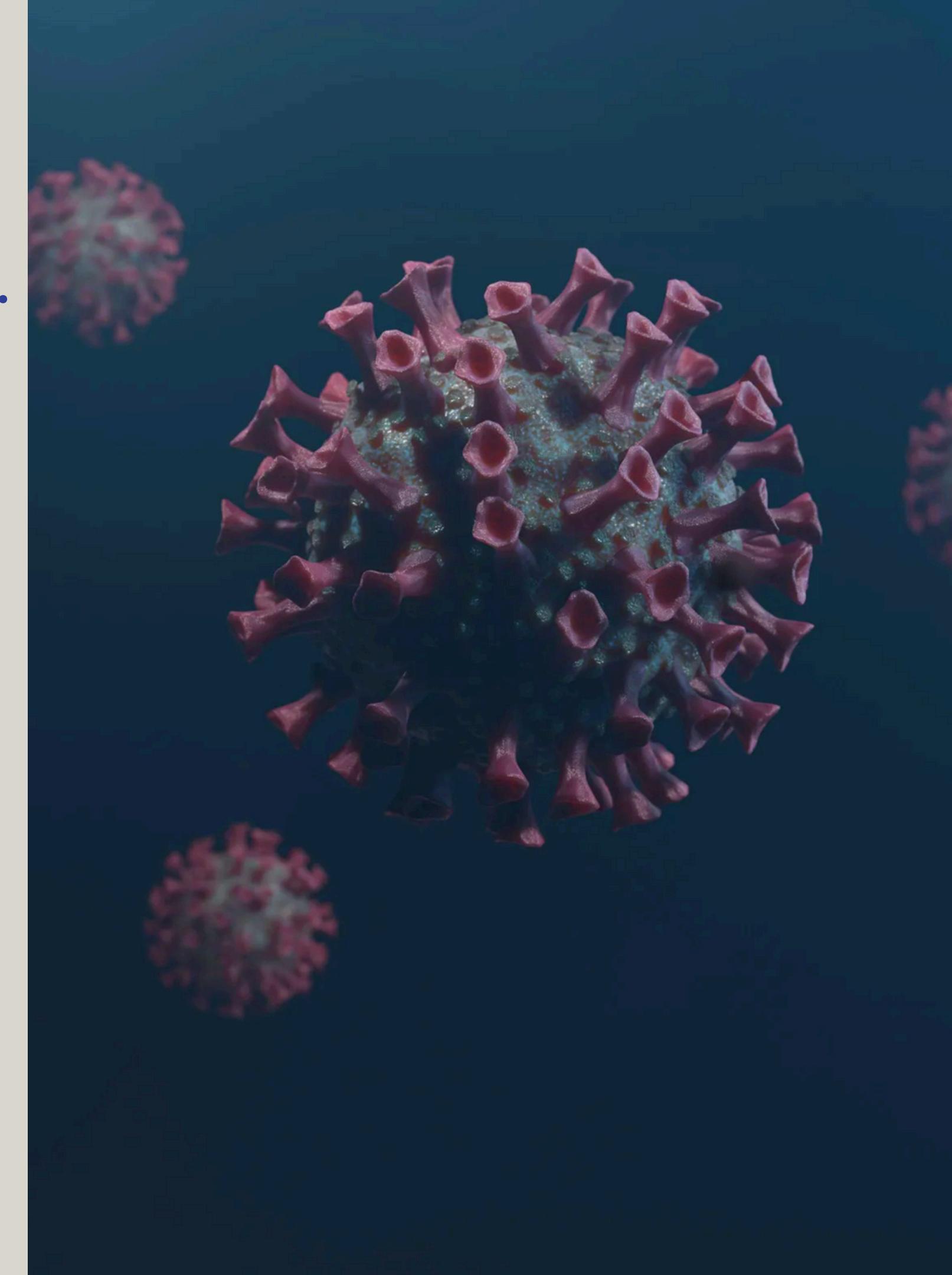
The COVID-19 pandemic has drastically affected public health globally, creating an urgent demand for data-driven insights to comprehend the virus's spread and impact. As a data analyst, I was tasked with analyzing a comprehensive dataset on COVID-19 to extract meaningful insights and inform public health strategies.

# DATASET DESCRIPTION

## Description of each column in the dataset:

- **Province:** Geographic subdivision within a country/region.
- **Country/Region:** Geographic entity where data is recorded.
- **Latitude:** North-south position on Earth's surface.
- **Longitude:** East-west position on Earth's surface.
- **Date:** Recorded date of CORONA VIRUS data.
- **Confirmed:** Number of diagnosed CORONA VIRUS cases.
- **Deaths:** Number of CORONA VIRUS related deaths.
- **Recovered:** Number of recovered CORONA VIRUS cases.

TOOL USED: SQL SERVER MANAGEMENT STUDIO



# DATA CLEANING

CHECK FOR NULL  
VALUES

-- Q1. Write a code to check NULL values

```
SELECT *  
FROM Corona_Virus_Dataset  
WHERE Country_Region IS NULL  
OR Confirmed IS NULL  
OR Deaths IS NULL  
OR Recovered IS NULL;
```

## REPLACE NULL VALUES WITH 0

-- Q2. If NULL values are present, update them with zeros  
for all columns.

```
UPDATE Corona_Virus_Dataset  
SET  
province = CASE WHEN province IS NULL THEN '0' ELSE  
              province END,  
country_region = CASE WHEN country_region IS NULL  
                  THEN '0' ELSE country_region END,  
latitude = CASE WHEN latitude IS NULL THEN '0' ELSE  
              latitude END,  
longitude = CASE WHEN longitude IS NULL THEN '0' ELSE  
              longitude END,  
date = CASE WHEN date IS NULL THEN '0' ELSE date END,  
confirmed = CASE WHEN confirmed IS NULL THEN 0 ELSE  
              confirmed END,  
deaths = CASE WHEN deaths IS NULL THEN 0 ELSE deaths  
              END,  
recovered = CASE WHEN recovered IS NULL THEN 0 ELSE  
              recovered END;
```

Total number of  
rows

-- Q3. check total number of rows

```
SELECT COUNT(*) AS Total_number_ofrows  
FROM Corona_Virus_Dataset;
```

## QUERY RESULT

THERE ARE 49150 ROWS IN THE DATASET

## DATASET TIMELINE

-- Q4. Check what is start\_date and end\_date

```
SELECT  
    MIN(date) AS start_date,  
    MAX(date) AS end_date  
FROM Corona_Virus_Dataset;
```

## QUERY RESULT

--start\_date 22-01-2020 and end\_date 13-06-2021

## NUMBER OF MONTHS

```
-- Q5. Number of month present in dataset  
SELECT DATEDIFF(MONTH, MIN(datE), MAX(date)) + 1 AS  
    NumberOfMonths  
FROM Corona_Virus_Dataset;
```

## QUERY RESULT

---The Number of Months are 18months

## Monthly Average

```
-- Q6. Find monthly average for confirmed, deaths,  
     recovered  
     SELECT  
         YEAR(date) AS Year,  
         DATENAME (MONTH,(date))as Month,  
         AVG(Confirmed) AS Avg_Confirmed,  
         AVG(Deaths) AS Avg_Deaths,  
         AVG(Recovered) AS Avg_Recovered  
     FROM  
         Corona_Virus_Dataset  
     GROUP BY  
         YEAR(date),  
         DATENAME (MONTH,(date))  
     ORDER BY  
         YEAR(Date),  
         DATENAME (MONTH,(Date)) ASC;
```

## USING CTE AND WINDOW FUNCTION TO GET FREQUENT VALUES

-- Q7. Find most frequent value for confirmed, deaths, recovered each month.

```
SELECT * FROM Corona_Virus_Dataset;
WITH FrequentData AS (
    SELECT
        YEAR(Date) AS Year,
        DATENAME(MONTH, Date) AS Month_Name,
        MONTH(Date) AS Month_Number,
        Confirmed,
        Deaths,
        Recovered,
        RANK() OVER (
            PARTITION BY YEAR(Date), MONTH(Date)
            ORDER BY COUNT(*) DESC
        ) AS rank
    FROM Corona_Virus_Dataset
    GROUP BY
        YEAR(Date),
        DATENAME(MONTH, Date),
        MONTH(Date),
        Confirmed,
        Deaths,
        Recovered
    )
    SELECT
        Year,
        Month_Name,
        Confirmed,
        Recovered,
        Deaths
    FROM FrequentData
    WHERE rank = 1
    ORDER BY Year, Month_Number;
```

## MINIMUM VALUE FOR CASES

-- Q8. Find minimum values for confirmed, deaths, recovered per year

```
SELECT
    YEAR(Date) AS Year,
    MIN(Confirmed) AS Max_Confirmed,
    MIN(Deaths) AS Max_Deaths,
    MINRecovered() AS Max_Recovered,
    Province
FROM
    Corona_Virus_Dataset
GROUP BY
    YEAR(Date), Province
ORDER BY
    Year, Province;
```

## MAXIMUM VALUE FOR CASES

-- Q9. Find maximum values of confirmed, deaths, recovered per year

```
SELECT
    YEAR(Date) AS Year,
    Max(Confirmed) AS Max_Confirmed,
    Max(Deaths) AS Max_Deaths,
    MaxRecovered) AS Max_Recovered,
    Province
FROM
    Corona_Virus_Dataset
GROUP BY
    YEAR(Date),Province
ORDER BY
    Year,Province;
```

## TOTAL NUMBER OF CASES PER MONTH

```
SELECT
    YEAR(date) AS Year,
    DATENAME(MONTH,(date))as Month,
    MONTH(date) AS Month_Number,
    Sum(Confirmed) AS Avg_Confirmed,
    Sum(Deaths) AS Avg_Deaths,
    SumRecovered) AS Avg_Recovered
FROM
    Corona_Virus_Dataset
GROUP BY
    YEAR(date),
    DATENAME(MONTH,(date)),
    MONTH(date)
ORDER BY
    YEAR(Date),
    MONTH(date);
```

## CORONA VIRUS SPREAD ANALYSIS

----- Q11. Check how corona virus spread out with respect to confirmed case

-- (Eg.: total confirmed cases, their average, variance & STDEV )

SELECT

```
YEAR(Date) AS Year,  
MONTH(Date) AS Month_Number,  
DATENAME(MONTH,(Date)) AS Month_Name,  
SUM(Confirmed) AS Total_Confirmed,  
AVG(Confirmed) AS Avg_Confirmed,
```

```
ROUND (VAR(Confirmed),2) AS  
Variance_Confirmed,
```

```
Round(STDEV(Confirmed),2) AS  
StdDev_Confirmed
```

FROM

Corona\_Virus\_Dataset

GROUP BY

```
YEAR(Date),  
MONTH(Date),  
DATENAME(MONTH,(Date))
```

ORDER BY

```
Year,  
Month;
```

-- Q12. Check how corona virus spread out with respect to death case per month

-- (Eg.: total confirmed cases, their average, variance & STDEV )

SELECT

```
YEAR(Date) AS Year,  
MONTH(Date) AS Month,  
SUM(Confirmed) AS Total_Confirmed,  
Sum(Deaths) AS Total_deaths,  
VAR(Deaths) AS Vaiance_deaths,  
STDEV(Deaths) AS StdDev_deaths
```

FROM

Corona\_Virus\_Dataset

GROUP BY

```
YEAR(Date),  
MONTH(Date)
```

ORDER BY

```
Year,  
Month;
```

## CORONA VIRUS SPREAD ANALYSIS

-- Q13. Check how corona virus spread out with respect to recovered case

-- (Eg.: total confirmed cases, their average, variance & STDEV )

SELECT

YEAR(Date) AS Year,  
MONTH(Date) AS Month,  
SUM(Confirmed) AS Total\_Confirmed,  
SumRecovered) AS Total\_Recovered,  
VARRecovered) AS Variance\_Recovered,  
STDEVRecovered) AS Recovered

FROM

Corona\_Virus\_Dataset

GROUP BY

YEAR(Date),

MONTH(Date)

ORDER BY

Year,

Month;

## HIGHEST NUMBER OF CONFIRMED CASE

-- Q14. Find Country having highest number of the Confirmed case

```
SELECT * FROM Corona_Virus_Dataset;
```

```
SELECT TOP 5
    Country_Region AS Country,
    SUM(Confirmed ) AS Total_Confirmed
    FROM
        Corona_Virus_Dataset
    GROUP BY
        Country_Region
    ORDER BY
        Total_Confirmed DESC;
```

Country	Total Confirmed
Brazil	34825532
Argentina	8248380
France	5848812
Colombia	3753224
Chile	2952946

## HIGHEST NUMBER OF CONFIRMED CASE

-- Q16. Find top 5 countries having highest recovered case

```
SELECT TOP 5
    Country_Region AS Country,
    SUM(Recovered) AS Recovered_cases
    FROM Corona_Virus_Dataset
    GROUP BY
        Country_Region
    ORDER BY
        Recovered_cases DESC;
```

Country	Recovered_cases
Brazil	30800338
Argentina	7442700
Colombia	3543807
Chile	2798374
Bangladesh	1532532

## LOWEST DEATH CASES

Country	Death_rate
Dominica	0
Bhutan	2
Barbados	94
Central African Republic	196
Cyprus	382

## COUNTRIES WITH LOWEST DEATH CASES

- Q15. Find Country having lowest number of the death case

```
SELECT TOP 5
Country_Region AS
Country,
SUM(Deaths) AS
Death_rate
FROM
Corona_Virus_Dataset
GROUP BY
Country_Region
ORDER BY
Death_rate ASC;
```

# CONCLUSION

BRAZIL AND ARGENTINA, COLUMBIA AND CHILE HAS THE HIGHEST NUMBER OF CONFIRMED CASES AND CONVERSELY THE HIGHEST NUMBER OF RECOVERED CASES  
WHILE  
DOMINICA,BHUTAN RANKED LOWEST IN DEATH RATE



A detailed 3D rendering of a COVID-19 virus particle, also known as a SARS-CoV-2 virus. The particle is spherical with a complex, spike-covered surface. The spikes are colored in a gradient from light yellow-green at the base to bright lime green at the tips. The main body of the virus is a darker, mottled color with visible internal structures. The entire image is set against a dark, slightly blurred background.

**THANK YOU!**