CORONA VIRUS ANALYSIS

Year 2020 & 2021
SQL Analysis
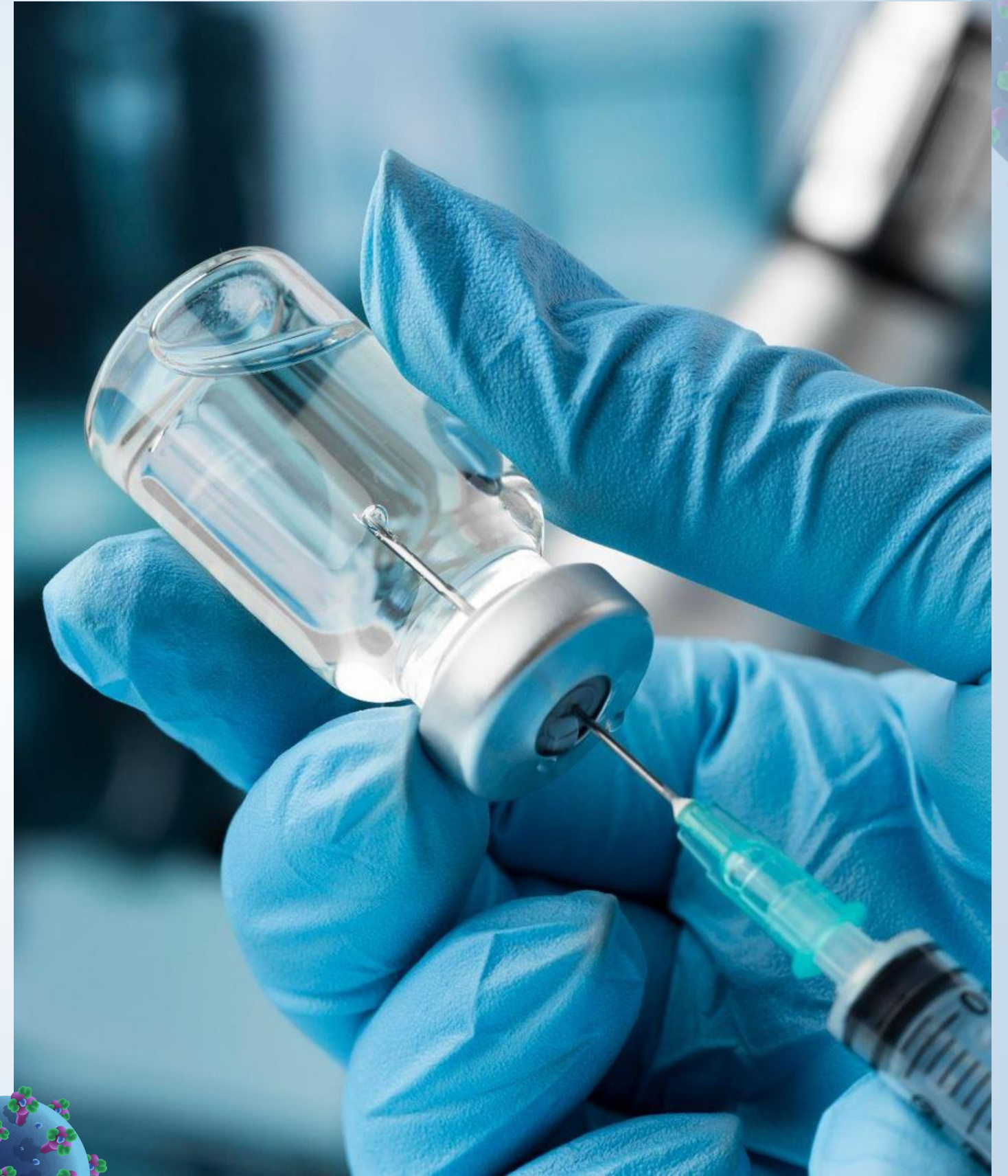
Surathni Gunawardena

TABLE OF CONTENT

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by the coronavirus SARS-CoV-2. The first known case was identified in Wuhan, China, in December 2019. The disease quickly spread worldwide, resulting in the COVID-19 pandemic.

PROBLEM STATEMENT

The CORONA VIRUS pandemic has had a significant impact on public health and has created an urgent need for data-driven insights to understand the spread of the virus. This analysis uses SQL to showcase the key insights of its spread globally in year 2020 & 2021.

DATASET OVERVIEW

DATASET EXPLANATION

- 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

DATA EXPLORATION USING SQL

CODE TO CHECK NULL VALUES

```sql
SELECT *
FROM [Corona Virus Dataset]
WHERE Province is null or Country_Region is null or
Latitude is null or Longitude is null or Date is null or
Confirmed is null or Deaths is null
or Recovered is null;


SELECT
    SUM(CASE WHEN Province IS NULL THEN 1 ELSE 0 END) AS Province_Null_Count,
    SUM(CASE WHEN Country_Region IS NULL THEN 1 ELSE 0 END) AS Countryregion_Null_Count,
    SUM(CASE WHEN Latitude IS NULL THEN 1 ELSE 0 END) AS Latitude_Null_Count,
    SUM(CASE WHEN Longitude IS NULL THEN 1 ELSE 0 END) AS Longitude_Null_Count,
    SUM(CASE WHEN Date IS NULL THEN 1 ELSE 0 END) AS Date_Null_Count,
    SUM(CASE WHEN Confirmed IS NULL THEN 1 ELSE 0 END) AS Confirmed_Null_Count,
    SUM(CASE WHEN Deaths IS NULL THEN 1 ELSE 0 END) AS Deaths_Null_Count,
    SUM(CASE WHEN Recovered IS NULL THEN 1 ELSE 0 END) AS Recovered_Null_Count
FROM [Corona Virus Dataset];
```

110 %

Results | Messages

Province	Country_Region	Latitude	Longitude	Date	Confirmed	Deaths	Recovered

	Province_Null_Count	Countryregion_Null_Count	Latitude_Null_Count	Longitude_Null_Count	Date_Null_Count	Confirmed_Null_Count	Deaths_Null_Count	Recovered_Null_Count
1	0	0	0	0	0	0	0	0

UPDATING THE NULL VALUE WITH 0

TOTAL NUMBER OF ROWS

```
--Q3--
SELECT COUNT(*) AS Total_Rows
FROM [Corona Virus Dataset];
```

110 %

Results | Messages

	Total_Rows
1	78386

TOTAL
78386

START DATE & END DATE

Start Date

2020/01/22

End Date

2021/06/13

NUMBER OF MONTHS
IN THE DATASET

```sql
--Q5--
SELECT COUNT(DISTINCT CONCAT(YEAR(TRY_CONVERT(date,Date,105)),'-',
MONTH(TRY_CONVERT(date,Date,105)))) AS NumberOfMonths
FROM [Corona Virus Dataset];
```

	NumberOfMonths
1	18

Total

Months 18

MONTHLY AVERAGE

MINIMUM / MAXIMUM ?
VALUES PER YEAR

TOTAL NUMBER OF CASES IN EACH MONTH

```sql
SELECT
    YEAR(TRY_CONVERT(date, Date,105)) AS Year,
    MONTH(TRY_CONVERT(date, Date,105)) AS Month,
    SUM (TRY_CONVERT (int, Confirmed)) AS Total_confirmed,
    SUM (TRY_CONVERT (int, Deaths)) AS Total_deaths,
    SUM (TRY_CONVERT (INT, Recovered)) AS Total_recovered
FROM
    [Corona Virus Dataset]
GROUP BY
    YEAR(TRY_CONVERT(date, Date,105)),
    MONTH(TRY_CONVERT(date, Date,105))
ORDER BY
    Year, Month;
```
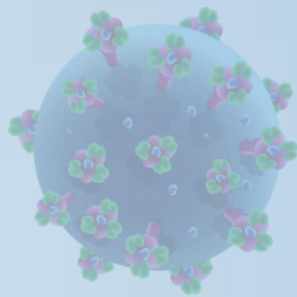
110 %

Results | Messages

	Year	Month	Total_confirmed	Total_deaths	Total_recovered
1	2020	1	6384	190	143
2	2020	2	68312	2651	31405
3	2020	3	769236	41346	133070
4	2020	4	2336798	191833	792987
5	2020	5	2744333	144561	1519547
6	2020	6	3969634	137757	2535417
7	2020	7	6838092	167613	4693120
8	2020	8	7694938	179200	6202833
9	2020	9	8244794	160671	6647749
10	2020	10	11515841	175484	6782150
11	2020	11	16595938	262247	9172292
12	2020	12	19336799	339996	11924903
13	2021	1	18672205	401893	9164347
14	2021	2	10492664	298239	6719785
15	2021	3	13924790	282620	7888013
16	2021	4	21711021	362387	14205507
17	2021	5	19121083	366549	19131842
18	2021	6	5022282	132657	5544438

VIRUS SPREAD WITH RESPECT TO CONFIRMED CASES
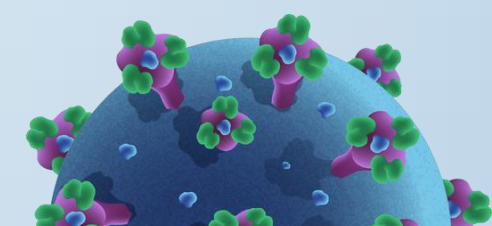
VIRUS SPREAD WITH RESPECT TO RECOVERED CASES

```sql
SELECT
    COUNT(Recovered) AS TotalRecoveredCases,
    AVG(TRY_CONVERT(BIGINT, Recovered)) AS AverageRecoveredCases,
    SUM(POWER(TRY_CONVERT(BIGINT, Recovered) - AVG_Cases.AverageRecoveredCases, 2))/COUNT(Recovered) AS VarianceRecoveredCases,
    SQRT(SUM(POWER(TRY_CONVERT(BIGINT, Recovered) - AVG_Cases.AverageRecoveredCases, 2))/COUNT(Recovered)) AS StdDevRecoveredCases
FROM
[Corona Virus Dataset],
    (SELECT AVG(TRY_CONVERT(BIGINT, Recovered)) AS AverageRecoveredCases FROM [Corona Virus Dataset]) AS AVG_Cases;
```

110 %

Results Messages

	TotalRecoveredCases	AverageRecoveredCases	VarianceRecoveredCases	StdDevRecoveredCases
1	78386	1442	107029523	10345.5073824342

COUNTRY HAVING HIGHEST NUMBER OF CONFIRMED CASES


```
SELECT TOP 1
    Country_Region AS Country,
    SUM(CAST(Confirmed AS INT)) AS TotalConfirmedCases

FROM
    [Corona Virus Dataset]
WHERE
    ISNUMERIC(Confirmed)=1
GROUP BY
    Country_Region
ORDER BY
    SUM(CAST(Confirmed AS INT)) DESC;
```

10 %

Results | Messages

	Country	TotalConfirmedCases
1	US	33461982

USA

COUNTRY HAVING LOWEST NUMBER OF DEATH CASES


```sql
    Country_Region AS Country,
    SUM(CAST(Deaths AS INT)) AS TotalDeathCases

FROM
    [Corona Virus Dataset]
WHERE
    ISNUMERIC(Deaths)=1
GROUP BY
    Country_Region
HAVING
    SUM(CAST(Confirmed AS INT)) = 0
ORDER BY
    TotalDeathCases ASC;
```

% ▼ ◀

Results | Messages

| Country | TotalDeathCases |

TOP 5 COUNTIRES HAVING HIGHEST RECOVERED CASES

```sql
SELECT TOP 5
    Country_Region AS Country,
    sum(cast(Recovered AS INT)) AS TotalRecoveredCases
FROM
    [Corona Virus Dataset]
WHERE
    ISNUMERIC(recovered) =1
GROUP BY
    Country_Region
ORDER BY |
    TotalRecoveredCases DESC;
```

110 %

Results   Messages

	Country	TotalRecoveredCases
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756

- **INDIA**
- **BRAZIL**
- **US**
- **TURKEY**
- **RUSSIA**

THANK YOU!