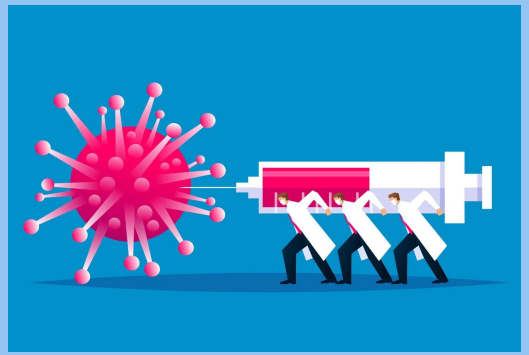
CORONA VIRUS ANALYSIS

MYSQL PROJECT



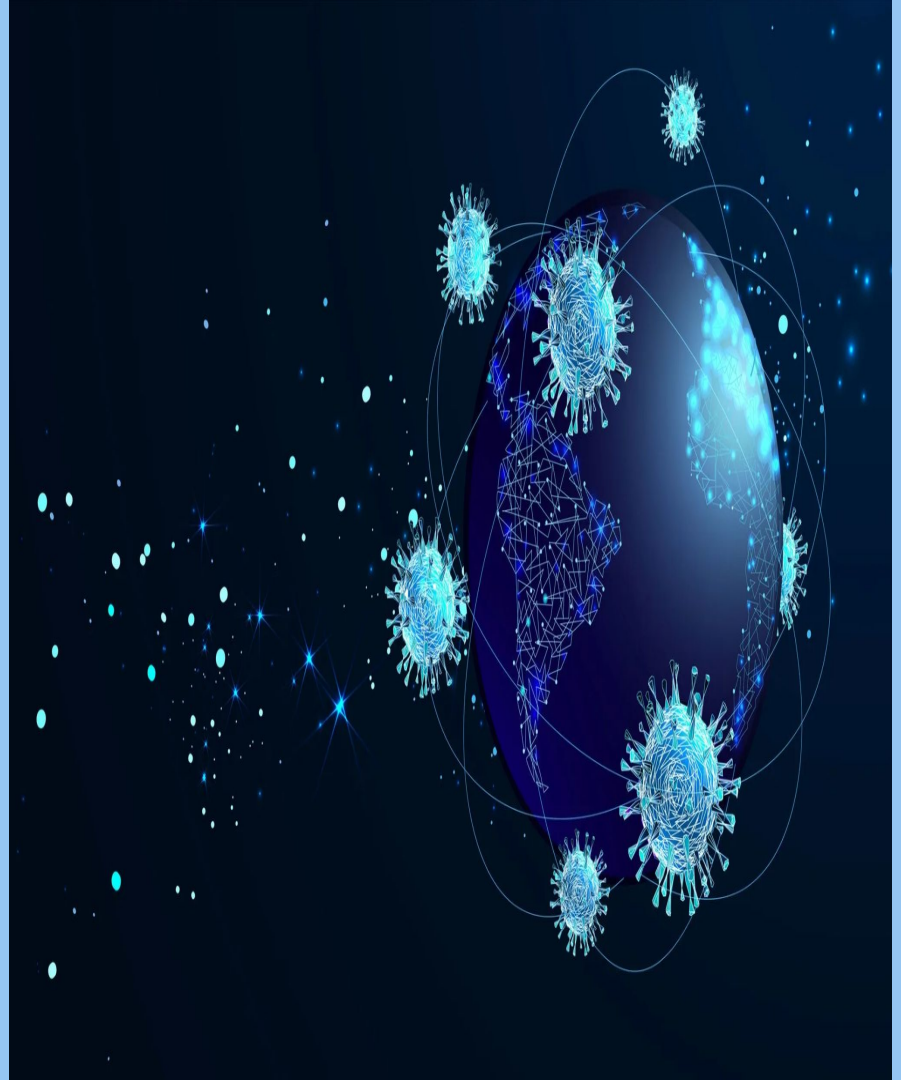
BY - JANVI SRIVASTAVA

CONTENT

1) INTRODUCTION
2) OVERVIEW
3) DATASET
4) DATA EXPLORATION USING SQL

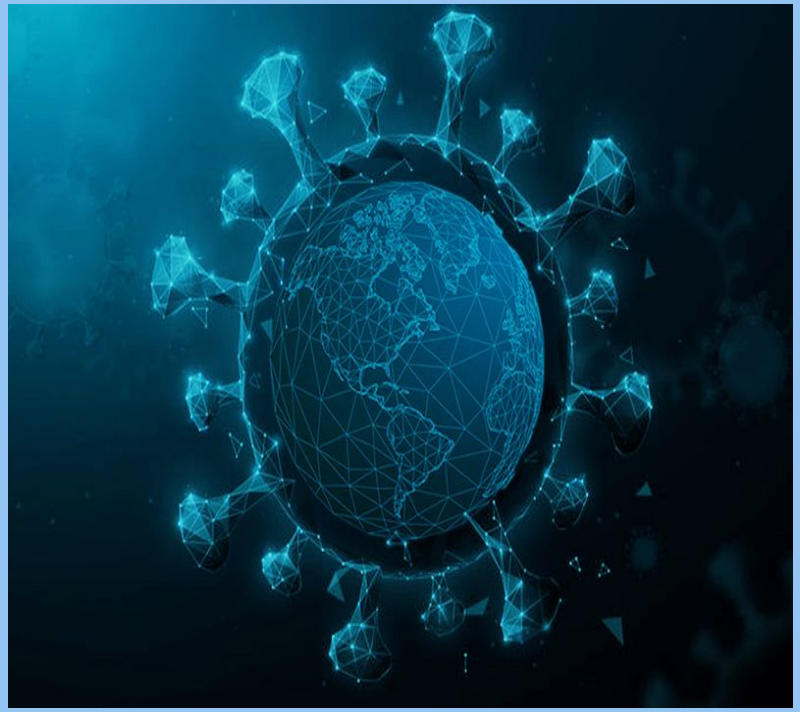
INTRODUCTION

Corona Virus disease 2019 also known as **COVID 19** is an illness caused by a virus. The virus is known as **SARS-CoV-2.** It started spreading at the end of 2019 and became a pandemic disease in 2020.



OVERVIEW

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.



DATASET

1) **PROVINCE:** Geographic subdivision within a country/region.
2) **COUNTRY/REGION:** Geographic entity where data is recorded.
3) **LATITUDE:** North-south position on Earth's surface.
4) **LONGITUDE:** East-west position on Earth's surface.
5) **DATE:** Recorded date of CORONA VIRUS data.
6) **CONFIRMED:** Number of diagnosed CORONA VIRUS cases.
7) **DEATHS:** Number of CORONA VIRUS related deaths.
8) **RECOVERED:** Number of recovered CORONA VIRUS cases.

DATA EXPLORATION USING SQL

CODE TO CHECK NULL VALUES

```
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;
```

IF NULL VALUES ARE PRESENT, UPDATE THEM WITH ZEROS FOR ALL COLUMNS.

```
update corona_virus_dataset
set Confirmed = coalesce(Confirmed, 0),
    Deaths = coalesce(Deaths, 0),
    Recovered = coalesce(Recovered, 0)
where Confirmed is null
and Deaths is null
and Recovered is null;
```

CODE TO CHECK TOTAL NUMBER OF ROWS

```
select count(*) as Total_Rows from
corona_virus_dataset;
```

TOTAL ROWS = 78386

CODE TO CHECK WHAT IS START DATE AND END DATE

```
select Min(Date) as start_date, Max(Date)
as end_date from corona_virus_dataset;
```

NUMBER OF MONTHS PRESENT IN DATASET

```
select count(Distinct Extract(Month from
str_to_date(Date, '%y/%m/%d'))) as
Num_months from corona_virus_dataset;
```

CODE TO FIND MONTHLY AVERAGE FOR CONFIRMED, DEATHS, RECOVERED

```
select
  Extract(Month from str_to_date(Date, '%y/%m/%d')) as Month,
  Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
  Avg(Confirmed) as Avg_Confirmed,
  Avg(Deaths) as Avg_Deaths,
  Avg(Recovered) as Avg_Recovered
from corona_virus_dataset group by month,year;
```

FIND THE MOST FREQUENT VALUE FOR CONFIRMED, DEATHS, RECOVERED EACH MONTH

```
select
   Extract(Month from str_to_date(Date, '%y/%m/D')) as Month,
   Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
   Substring_Index(Group_Concat(Confirmed order by Confirmed Desc), ',', 1) as
Most_Frequent_Confirmed,
   Substring_Index(Group_Concat(Deaths order by Deaths Desc), ',',1) as Most_Frequent_Deaths,
   Substring_Index(Group_Concat(Recovered order by Recovered Desc), ',', 1) as
Most_Frequent_Recovered
from corona_virus_dataset
group by Year, month
Order by Year, Month;
```

FIND MINIMUM VALUES FOR CONFIRMED, DEATHS, RECOVERED PER YEAR

```
select
   Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
   Min(Confirmed) as Min_Confirmed,
   Min(Deaths) as Min_Deaths,
   Min(Recovered) as Min_Recovered
 from corona_virus_dataset group by Year order by Year;
```

FIND MAXIMUM VALUES OF CONFIRMED, DEATHS, RECOVERED PER YEAR

```
select
    Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
    Max(Confirmed) as Max_Confirmed,
    Max(Deaths) as Max_Deaths,
    Max(Recovered) as Max_Recovered
 from corona_virus_dataset group by Year order by Year;
```

THE TOTAL NUMBER OF CASE OF CONFIRMED, DEATHS, RECOVERED EACH MONTH

```
select
    Extract(Month from str_to_date(Date, '%y/%m/%d')) as Month,
    Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
    Sum(Confirmed) as Total_Confirmed,
    Sum(Deaths) as Total_Deaths,
    Sum(Recovered) as Total_Recovered
 from corona_virus_dataset group by Year, Month order by Year, Month;
```

CHECK HOW CORONA VIRUS SPREAD OUT WITH RESPECT TO CONFIRMED CASE (EG: TOTAL CONFIRMED CASES, THEIR AVERAGE, VARIANCE & STDEV)

```
select
    Sum(Confirmed) as Total_Confirmed_Cases,
    Avg(Confirmed) as Avg_Confirmed_Cases,
    Variance(Confirmed) as Variance_Confirmed_Cases,
    Stddev(Confirmed) as Stdev_Confirmed_Cases
 from corona_virus_dataset;
```

CHECK HOW CORONA VIRUS SPREAD OUT WITH RESPECT TO DEATH CASE PER MONTH (EG: TOTAL CONFIRMED CASES, THEIR AVERAGE, VARIANCE & STDEV)

```
select
    Extract(Month from str_to_date(Date, '%y/%m/%d')) as Month,
    Extract(Year from str_to_date(Date, '%y/%m/%d')) as Year,
    Sum(Deaths) as Total_Death_Cases,
    Avg(Deaths) as Avg_Deaths_Cases,
    Variance(Deaths) as Variance_Deaths_Cases,
    Stddev(Deaths) as Stdev_Deaths_Cases
 from corona_virus_dataset group by Year, Month order by Year,
Month;
```

CHECK HOW CORONA VIRUS SPREAD OUT WITH RESPECT TO RECOVERED CASE (EG: TOTAL CONFIRMED CASES, THEIR AVERAGE, VARIANCE & STDEV)

```
select
    Sum(Recovered) as Total_Recovered_Cases,
    Avg(Recovered) as Avg_Recovered_Cases,
    Variance(Recovered) as Variance_Recovered_Cases,
    Stddev(Recovered) as Stdev_Recovered_Cases
 from corona_virus_dataset;
```

FIND COUNTRY HAVING HIGHEST NUMBER OF THE CONFIRMED CASE

```
select
   Country_Region, Sum(Confirmed) as Total_Confirmed_Cases
from corona_virus_dataset
group by Country_Region
order by Total_Confirmed_Cases Desc
limit 1;
```

FIND COUNTRY HAVING LOWEST NUMBER OF THE DEATH CASE

```
with rankingCountry as(
 select
    Country_Region,
    Sum(Deaths) as Total_Deaths_Cases,
    rank() over(order by sum(Deaths) asc) as rank_no
    from corona_virus_dataset
    group by Country_Region
)
select
   Country_Region, Total_Deaths_Cases
from rankingCountry where rank_no = 1;
```

FIND TOP 5 COUNTRIES HAVING HIGHEST RECOVERED CASE

```
select Country_Region, sum(Recovered) as Total_Recovered
from corona_virus_dataset
group by Country_Region order by Total_Recovered desc
limit 5;
```


Thank you