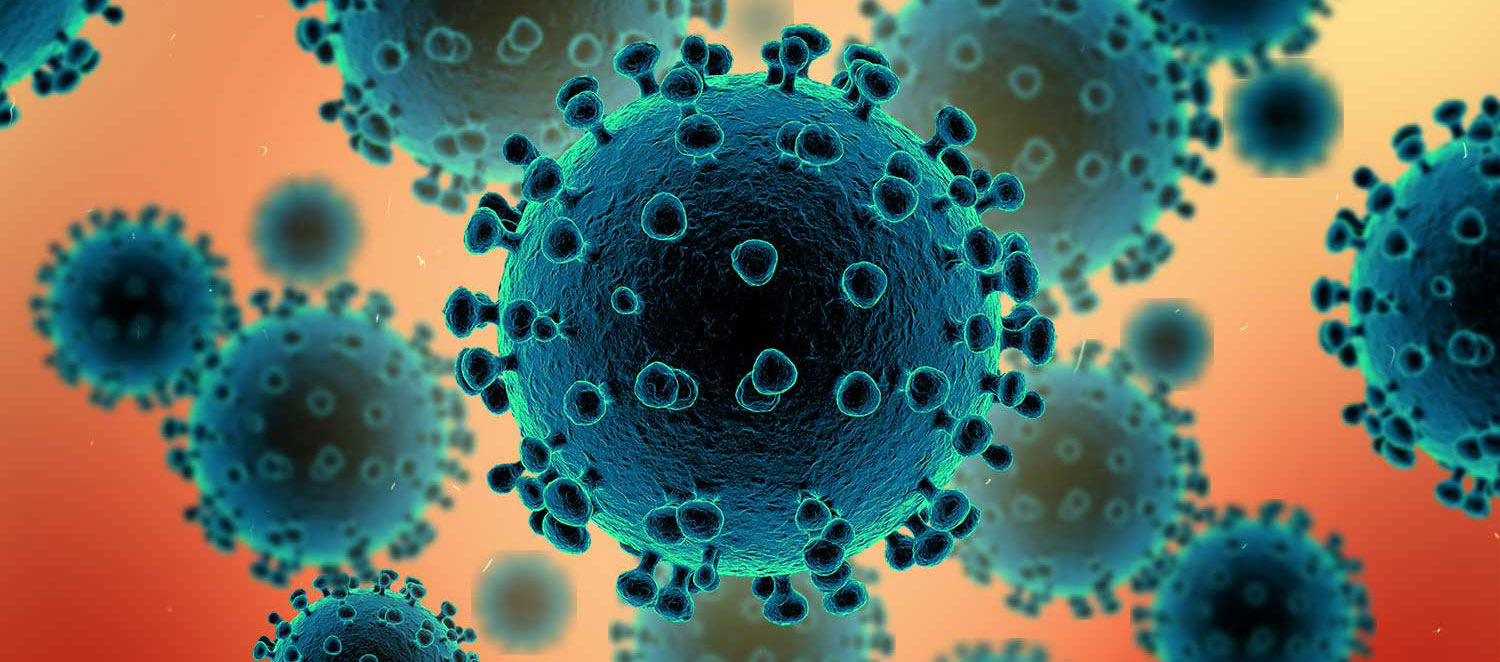
**CORONA VIRUS DATASET ANALYSIS BY ( In the video explanation first we will explain this ppt then we will run all the query in SQL workbench)**

**INTERN NAME - SAGAR SHUKLA-**



**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. As a data analyst, you have been tasked with analyzing a CORONA VIRUS dataset to derive meaningful insights and present your findings.

**Dataset:**

Description of each column in dataset:

**Province:** Geographic subdivision within a country/region.

**Country/Region:** Geographic entity where data is recorded.

**Latitude:** North-south position on Earth's surface.

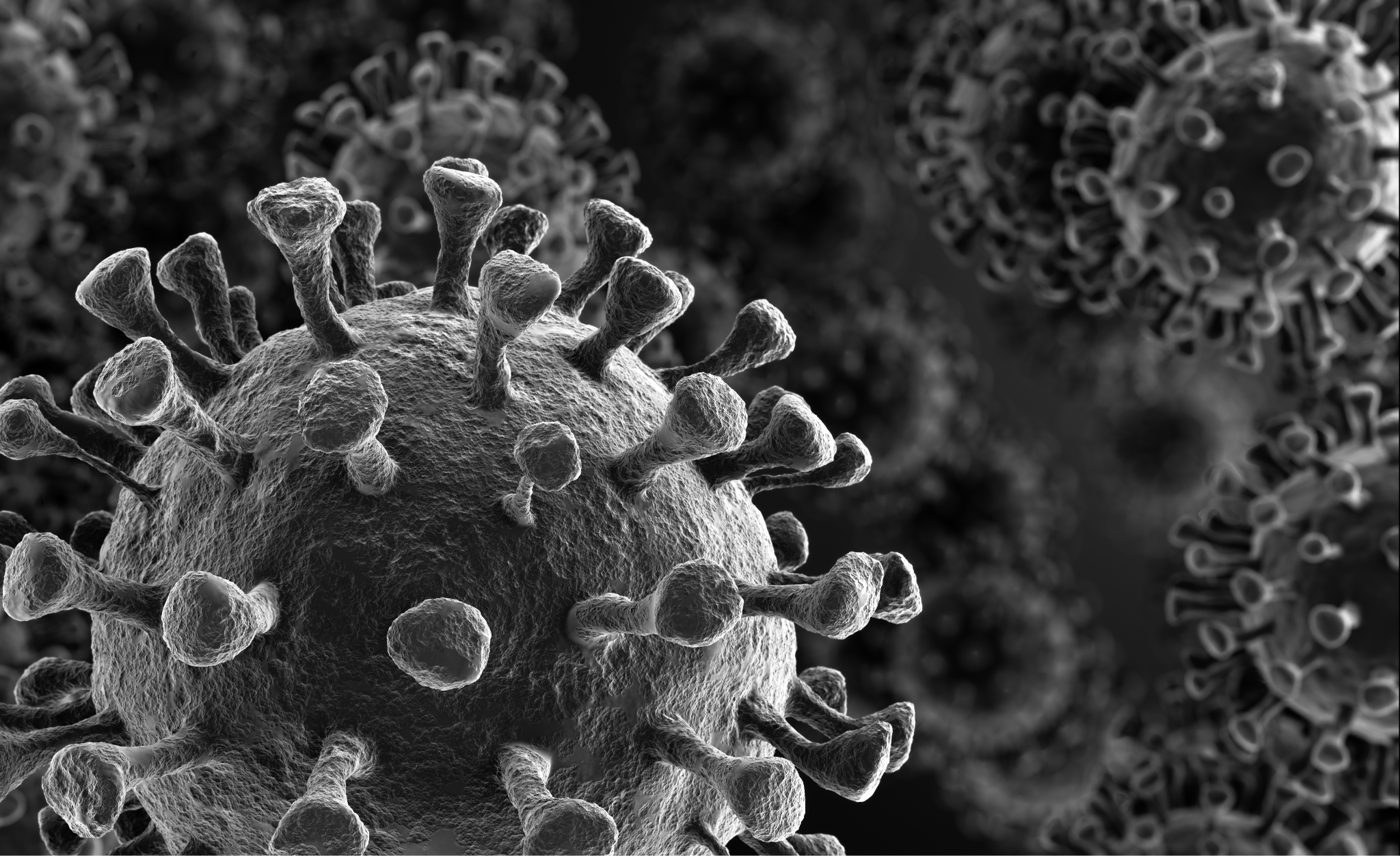
**Longitude:** East-west position on Earth's surface.

**Date:** Recorded date of CORONAVIRUS data.

**Confirmed:** Number of diagnosed CORONAVIRUS cases.

**Deaths:** Number of CORONAVIRUS related deaths.

**Recovered:** Number of recovered CORONAVIRUS cases.



select \* from corona\_virus\_dataset;

-- 1)-Code for check null values

select \* from corona\_virus\_dataset

where 'Province' is null

or 'Country/Regi' 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 ;

-- 2)- If null value are present,update them with zero for all columns

SET SQL\_SAFE\_UPDATES = 0;

UPDATE corona\_virus\_dataset

SET

Province = 0,

`Country/Region` = 0,

Latitude = 0,

Longitude = 0,

Date = 0,

Confirmed = 0,

Deaths = 0,

Recovered = 0

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;

SET SQL\_SAFE\_UPDATES = 1;

-- 3)-check total number of rows

select count(\*) from corona\_virus\_dataset;

-- 4)- Check what is start\_date start\_date and end\_date

SELECT

MIN(Date) AS start\_date,

MAX(Date) AS end\_date

FROM

corona\_virus\_dataset;

-- 5)- Number of month present in dataset

select count(distinct extract(month from str\_to\_date(Date, '%d-%m-%y'))) as total\_month

from corona\_virus\_dataset;

-- 6)-Find monthly average for confirmed , deaths , recovered

SELECT

AVG(Confirmed) AS avg\_confirmed,

AVG(Deaths) AS avg\_deaths,

AVG(Recovered) AS avg\_recovered,

MONTH(STR\_TO\_DATE(Date, '%d-%m-%y')) AS month\_number,

MONTHNAME(STR\_TO\_DATE(Date, '%d-%m-%y')) AS month\_name

FROM

corona\_virus\_dataset

GROUP BY

month\_name , month\_number

order by month\_number;

-- 7)-Find most frequent value for confirmed,death,recovered each month

SELECT

YEAR(STR\_TO\_DATE(Date, '%d-%m-%y')) AS year,

MONTH(STR\_TO\_DATE(Date, '%d-%m-%y')) AS month\_number,

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

month\_number,year

order by month\_number;



-- 8)-minimum value of confirmed , death , recover per year

select min(Confirmed),min(Deaths),min(Recovered) from corona\_virus\_dataset

group by YEAR(STR\_TO\_DATE(Date, '%d-%m-%y'))

order by YEAR(STR\_TO\_DATE(Date, '%d-%m-%y')) ;

-- 9)- maximum value of confirmed , death , recover per year

select max(Confirmed),max(Deaths),max(Recovered) from corona\_virus\_dataset

group by YEAR(STR\_TO\_DATE(Date, '%d-%m-%y'))

order by YEAR(STR\_TO\_DATE(Date, '%d-%m-%y')) DESC;

-- 10)- total number of case confirmed , death , recover per month

select sum(Confirmed),sum(Deaths),sum(Recovered) from corona\_virus\_dataset

group by MONTH(STR\_TO\_DATE(Date, '%d-%m-%y'))

order by MONTH(STR\_TO\_DATE(Date, '%d-%m-%y')) ;

-- 11)- Check how corona virus spread out with respect to confirmed case

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

select sum(Confirmed) as toatl\_confirmed\_case ,

avg(Confirmed) as avg\_confirmed\_case ,

variance(Confirmed) as varrince\_confirmed\_case,

stddev(Confirmed) as std\_confirmed\_case

from corona\_virus\_dataset;

-- 12)- Check how corona virus spread out with respect to death case per month

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

select MONTH(STR\_TO\_DATE(Date, '%d-%m-%y')) as month\_num ,

YEAR(STR\_TO\_DATE(Date, '%d-%m-%y')) as year\_num,

sum(Deaths) as toatl\_death\_case,

avg(Deaths) as avg\_death\_case ,

variance(Deaths) as varrince\_death\_case,

stddev(Deaths) as std\_death\_case

from corona\_virus\_dataset

group by

month\_num

order by

month\_num

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

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

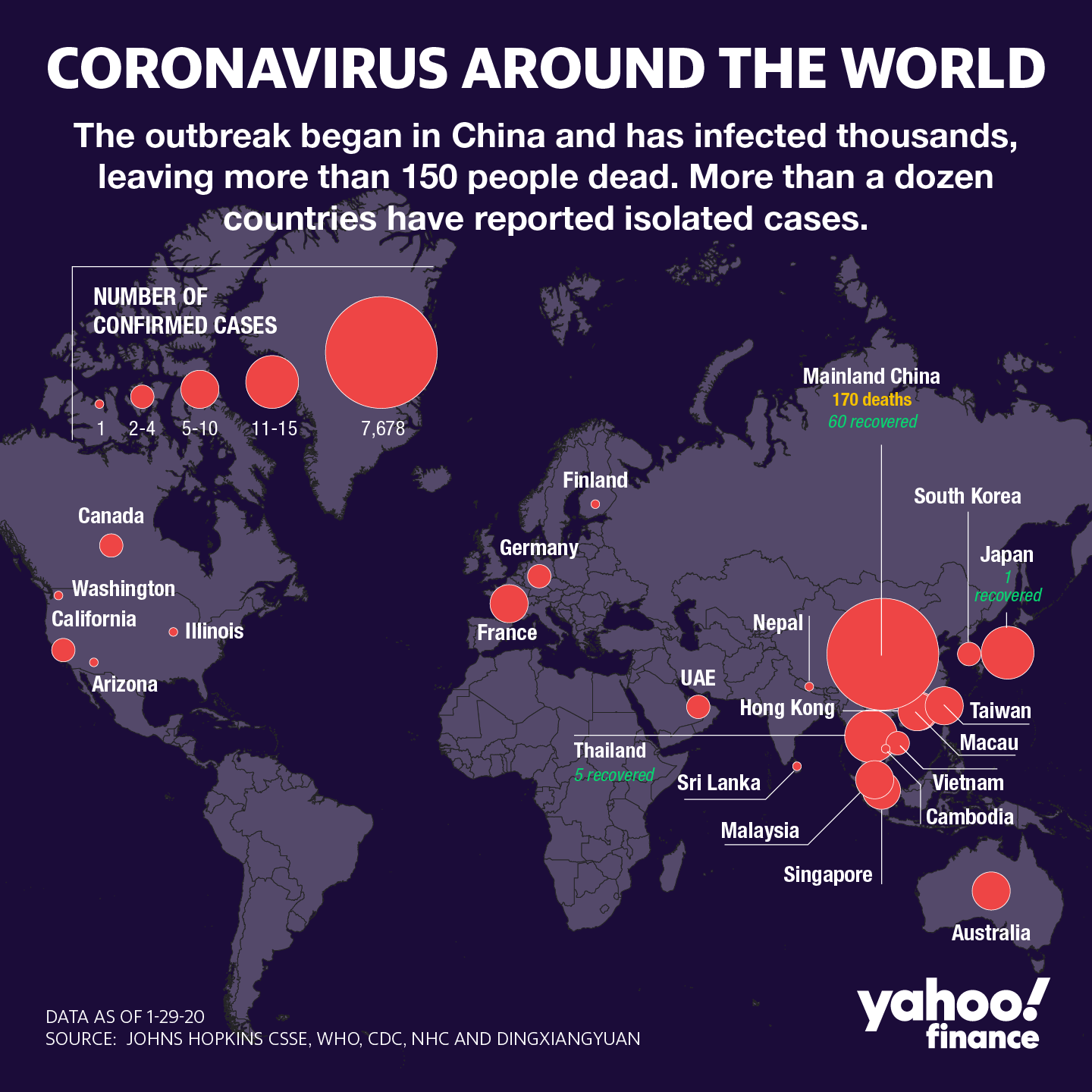
select sum(Recovered) as toatl\_death\_case,

avg(Recovered) as avg\_recovered\_case ,

variance(Recovered) as varrince\_recovered\_case,

stddev(Recovered) as std\_recovered\_case

from corona\_virus\_dataset;



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

select `Country/Region`,sum(Confirmed) as total\_confirmed

from corona\_virus\_dataset

group by `Country/Region`

order by total\_confirmed DESC limit 1;

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

select `Country/Region`,sum(Deaths) as total\_deaths

from corona\_virus\_dataset

group by `Country/Region`

order by total\_deaths ASC limit 1;

-- Q16. 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;