**CORONA VIRUS DATASET ANALYSIS**

**SQL QUERIES**

To avoid any errors, check missing value / null value

-- Q1. Write a code to check NULL values

SELECT \*

FROM corona\_virus

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;

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

UPDATE corona\_virus

SET

Province = COALESCE(Province, ''),

"country\_region" = COALESCE("country\_region", ''),

Latitude = COALESCE(Latitude, 0),

Longitude = COALESCE(Longitude, 0),

"date" = COALESCE("date", NULL),

Confirmed = COALESCE(Confirmed, 0),

Deaths = COALESCE(Deaths, 0),

Recovered = COALESCE(Recovered, 0);

-- Q3. check total number of rows

select

count(\*) as total\_rows

from corona\_virus

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

select

min(date) as start\_date

from corona\_virus

select

max(date) as last\_date

from corona\_virus

-- Q5. Number of month present in dataset

SELECT COUNT(DISTINCT DATE\_TRUNC('month', "date")) AS number\_of\_months

FROM corona\_virus

-- Q6. Find monthly average for confirmed, deaths, recovered

SELECT

DATE\_TRUNC('month', "date") AS month,

AVG(confirmed) AS avg\_confirmed,

AVG(deaths) AS avg\_deaths,

AVG(recovered) AS avg\_recovered

FROM corona\_virus

GROUP BY DATE\_TRUNC('month', "date")

ORDER BY month;

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

WITH monthly\_mode AS (

SELECT

DATE\_TRUNC('month', "date") AS month,

mode() WITHIN GROUP (ORDER BY confirmed) AS mode\_confirmed,

mode() WITHIN GROUP (ORDER BY deaths) AS mode\_deaths,

mode() WITHIN GROUP (ORDER BY recovered) AS mode\_recovered

FROM corona\_virus

GROUP BY DATE\_TRUNC('month', "date")

)

SELECT

month,

mode\_confirmed AS most\_frequent\_confirmed,

mode\_deaths AS most\_frequent\_deaths,

mode\_recovered AS most\_frequent\_recovered

FROM monthly\_mode

ORDER BY month;

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

SELECT

EXTRACT(year FROM "date") AS year,

MIN(confirmed) AS min\_confirmed,

MIN(deaths) AS min\_deaths,

MIN(recovered) AS min\_recovered

FROM corona\_virus

GROUP BY EXTRACT(year FROM "date")

ORDER BY year;

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

SELECT

EXTRACT(year FROM "date") AS year,

MAX(confirmed) AS max\_confirmed,

MAX(deaths) AS max\_deaths,

MAX(recovered) AS max\_recovered

FROM corona\_virus

GROUP BY EXTRACT(year FROM "date")

ORDER BY year;

-- Q10. The total number of case of confirmed, deaths, recovered each month

SELECT

DATE\_TRUNC('month', "date") AS month,

SUM(confirmed) AS total\_confirmed,

SUM(deaths) AS total\_deaths,

SUM(recovered) AS total\_recovered

FROM corona\_virus

GROUP BY DATE\_TRUNC('month', "date")

ORDER BY month;

Q11. 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 average\_confirmed\_cases,

VARIANCE(Confirmed) AS variance\_confirmed\_cases,

STDDEV(Confirmed) AS stdev\_confirmed\_cases

FROM corona\_virus

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

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

SELECT

SUM(deaths) AS total\_death\_cases,

AVG(deaths) AS average\_death\_cases,

VARIANCE(deaths) AS variance\_death\_cases,

STDDEV(deaths) AS stdev\_death\_cases

FROM corona\_virus

-- Q13. 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 average\_recovered\_cases,

VARIANCE(recovered) AS variance\_recovered\_cases,

STDDEV(recovered) AS stdev\_recovered\_cases

FROM corona\_virus

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

SELECT

"country\_region" AS country,

SUM(confirmed) AS total\_confirmed\_cases

FROM corona\_virus

GROUP BY "country\_region"

ORDER BY total\_confirmed\_cases DESC

LIMIT 1;

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

SELECT

"country\_region" AS country,

SUM(deaths) AS total\_death\_cases

FROM corona\_virus

GROUP BY "country\_region"

ORDER BY total\_death\_cases ASC

LIMIT 1;

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

SELECT

"country\_region" AS country,

SUM(recovered) AS total\_recovered\_cases

FROM corona\_virus

GROUP BY "country\_region"

ORDER BY total\_recovered\_cases DESC

LIMIT 5;

By – Mehakpreet kaur