-- To avoid any errors, check missing value / null value

-- Q1. Write a code to check NULL values

Ans.

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.

Ans.

UPDATE Corona\_Virus

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;

-- Q3. check total number of rows

Ans.

SELECT COUNT(\*) AS total\_rows

FROM Corona\_Virus;

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

Ans.

SELECT

MIN (Date) AS Start\_Date,

MAX (Date) AS End\_Date

FROM Corona\_Virus;

-- Q5. Number of month present in dataset

Ans.

SELECT

EXTRACT (MONTH FROM Date) AS Month,

COUNT(\*) AS Num\_of\_Months

FROM Corona\_Virus

GROUP BY EXTRACT (MONTH FROM Date)

ORDER BY EXTRACT (MONTH FROM Date);

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

Ans.

SELECT

EXTRACT (MONTH FROM Date) AS Month,

EXTRACT (YEAR FROM Date) AS Year,

Round (AVG (Confirmed), 3) AS Avg\_Confirmed,

ROUND (AVG (Deaths),3) AS Avg\_Deaths,

ROUND (AVG (Recovered), 3) AS Avg\_Recovered

FROM Corona\_Virus

GROUP BY

EXTRACT (YEAR FROM Date),

EXTRACT (MONTH FROM Date)

ORDER BY

EXTRACT (YEAR FROM Date),

EXTRACT (MONTH FROM Date);

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

Ans.

WITH AggregatedData AS (

SELECT

EXTRACT(YEAR FROM Date) AS Year,

EXTRACT(MONTH FROM Date) AS Month,

STRING\_AGG(Confirmed::TEXT, ',' ORDER BY Confirmed DESC) AS AggConfirmed,

STRING\_AGG(Deaths::TEXT, ',' ORDER BY Deaths DESC) AS AggDeaths,

STRING\_AGG(Recovered::TEXT, ',' ORDER BY Recovered DESC) AS AggRecovered

FROM

Corona\_Virus

GROUP BY

Year, Month

)

SELECT

Year,

Month,

SPLIT\_PART(AggConfirmed, ',', 1) AS MostFrequentConfirmed,

SPLIT\_PART(AggDeaths, ',', 1) AS MostFrequentDeaths,

SPLIT\_PART(AggRecovered, ',', 1) AS MostFrequentRecovered

FROM

AggregatedData

ORDER BY

Year, Month;

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

Ans.

SELECT

EXTRACT (YEAR FROM Date) AS Year,

MIN (Confirmed) AS MinConfirmed,

MIN (Deaths) AS MinDeaths,

MIN (Recovered) AS MinRecovered

FROM Corona\_Virus

GROUP BY

EXTRACT (YEAR FROM Date)

ORDER BY

Year;

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

Ans.

SELECT

EXTRACT (YEAR FROM Date) AS Year,

MAX (Confirmed) AS MaxConfirmed,

MAX (Deaths) AS MaxDeaths,

MAX (Recovered) AS MaxRecovered

FROM Corona\_Virus

GROUP BY

EXTRACT (YEAR FROM Date)

ORDER BY

Year;

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

Ans.

SELECT

EXTRACT (YEAR FROM Date) AS Year,

EXTRACT (MONTH FROM Date) AS Month,

SUM (Confirmed) AS TotalConfirmed,

SUM (Deaths) AS TotalDeaths,

SUM (Recovered) AS TotalRecovered

FROM Corona\_Virus

GROUP BY

EXTRACT (YEAR FROM Date),

EXTRACT (MONTH FROM Date)

ORDER BY

Year,

Month;

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

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

Ans.

SELECT

SUM (Confirmed) AS TotalConfirmedCases,

AVG (Confirmed) AS AverageConfirmedCases,

VARIANCE (Confirmed) AS VarianceConfirmedCases,

STDDEV (Confirmed) AS StdDevConfirmedCases

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 )

Ans.

SELECT

EXTRACT (MONTH FROM Date) AS Month,

SUM (Deaths) AS TotalDeathCases,

AVG (Deaths) AS AverageDeathCases,

VARIANCE (Deaths) AS VarianceDeathCases,

STDDEV (Deaths) AS StdDevDeathCases

FROM Corona\_Virus

GROUP BY

EXTRACT (MONTH FROM Date)

ORDER BY

Month;

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

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

Ans.

SELECT

SUM (Recovered) AS TotalRecoveredCases,

AVG (Recovered) AS AverageRecoveredCases,

VARIANCE (Recovered) AS VarianceRecoveredCases,

STDDEV (Recovered) AS StdDevRecoveredCases

FROM Corona\_Virus

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

Ans.

SELECT

Country\_Region,

SUM (Confirmed) AS TotalConfirmedCases

FROM Corona\_Virus

GROUP BY

Country\_Region

ORDER BY

TotalConfirmedCases DESC

LIMIT 5;

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

Ans.

SELECT

FROM

Country Region,

MIN (Deaths) AS Total DeathCases

Corona Virus

GROUP BY

Country Region

ORDER BY

TotalDeathCases ASC

LIMIT 5;

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

Ans.

SELECT

Country\_Region,

SUM (Recovered) AS TotalRecoveredCases

FROM Corona\_Virus

GROUP BY

Country\_Region

ORDER BY

TotalRecoveredCases DESC

LIMIT 5;