

SQL PROJECT

CORONA VIRUS

ANALYSIS

-Purva Phalak.

OBJECTIVE

- The objective of coronavirus analysis is to monitor how COVID-19 spreads and affects people by looking at data on cases, recoveries, and deaths in different areas over time.
- This helps health officials and governments make informed decisions about how to respond, like where to send medical supplies and when to implement safety measures.
- It also helps scientists predict future outbreaks and develop vaccines and treatments. Overall, the goal is to reduce the virus's impact, save lives, and help society return to normal.

DATASET

Description of each column in dataset:

Province : Geographic subdivision within a country/region.

Country : Geographic entity where data is recorded.

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

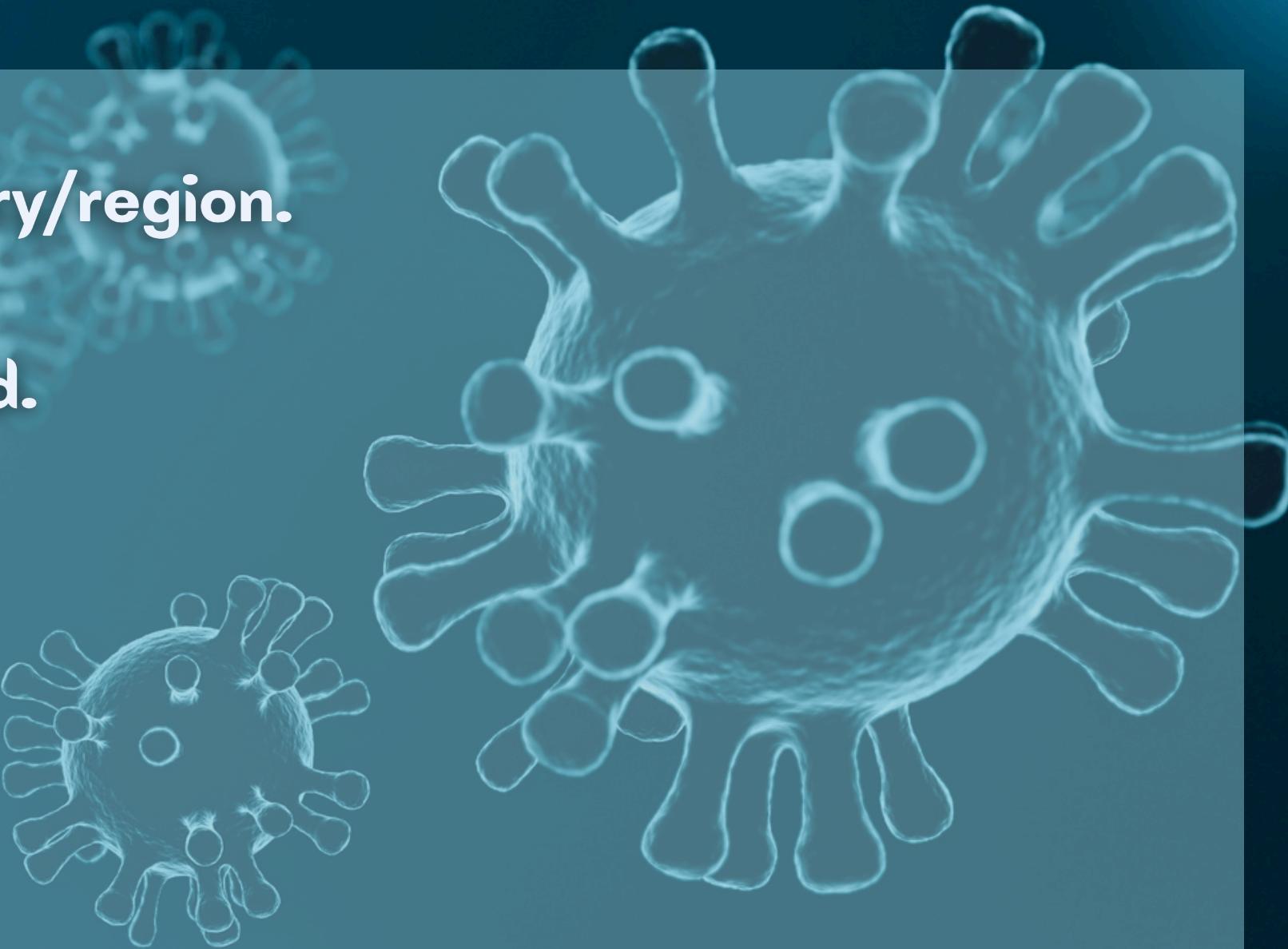
Longitude : East-west position on Earth's surface.

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.



QUESTION - 1

Corona Virus Analysis

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;
```

Province	Country	Latitude	Longitude	Date	Confirmed	Deaths	Recovered

- There are no null values in given dataset.

QUESTION - 2

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

```
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;SELECT
COUNT(DISTINCT EXTRACT(MONTH FROM str_to_date(Date, '%d-%m-%Y')))) AS num_months
FROM
corona_virus;
```

- There are no null values in given dataset.

QUESTION - 3

Check total number of rows.

```
SELECT  
    COUNT(*) AS Total_rows  
FROM  
    corona_virus;
```

	Total_rows
▶	78386

QUESTION - 4

Corona Virus Analysis

Check what is `start_date` and `end_date`.

```
SELECT  
    MIN(DATE) AS start_date, MAX(Date) AS end_date  
FROM  
    corona_virus;
```

	start_date	end_date
▶	2020-01-22	2021-06-13

QUESTION - 5

Corona Virus Analysis

Number of month present in dataset.

```
SELECT  
    COUNT(DISTINCT EXTRACT(MONTH FROM STR_TO_DATE(date, '%Y-%m-%d')))) AS num_months  
FROM  
corona_virus;
```

	num_months
▶	12

QUESTION - 6

Corona Virus Analysis

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

GROUP BY Month , Year;

	Month	Year	Avg_Confirmed	Avg_Deaths	Avg_Recovered
1	2020	4.1455	0.1234	0.0929	
2	2020	15.2960	0.5936	7.0320	
3	2020	161.1303	8.6607	27.8739	
4	2020	505.8004	41.5223	171.6422	
5	2020	574.8498	30.2809	318.2964	
6	2020	859.2281	29.8175	548.7916	
7	2020	1432.3611	35.1096	983.0582	
8	2020	1611.8429	37.5367	1299.2947	
9	2020	1784.5874	34.7773	1438.9067	
10	2020	2412.1996	36.7583	1420.6431	
11	2020	3592.1944	56.7634	1985.3446	
12	2020	4050.4397	71.2183	2497.8850	
1	2021	3911.2285	84.1837	1919.6370	
2	2021	2433.3636	69.1649	1558.3917	
3	2021	2916.7972	59.1998	1652.2859	
4	2021	4699.3552	78.4387	3074.7851	
5	2021	4005.2541	76.7803	4007.5078	
6	2021	2508.6324	66.2622	2769.4496	

QUESTION - 7

Corona Virus Analysis

Find 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  
GROUP BY Year , Month  
ORDER BY Year , Month;
```

Month	Year	Most_frequent_confirmed	Most_frequent_deaths	Most_frequent_recovered
1	2020	2131	49	51
2	2020	14840	242	3418
3	2020	26314	1085	4289
4	2020	50740	2607	33227
5	2020	34907	2309	51717
6	2020	54771	2003	94305
7	2020	75866	1595	140050
8	2020	85687	1505	95881
9	2020	97894	1703	101468
10	2020	99264	3351	388340
11	2020	207933	2259	139292
12	2020	823225	3752	1123456
1	2021	300462	4475	87090
2	2021	134975	3907	98389
3	2021	100158	3869	102138
4	2021	401993	4249	299988
5	2021	414188	4529	422436
6	2021	134154	7374	231456

QUESTION - 8

Corona Virus Analysis

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,  
MINRecovered) AS Min_Recovered  
FROM  
corona_virus  
GROUP BY Year  
ORDER BY Year;
```

	Year	Min_Confirmed	Min_Deaths	Min_Recovered
▶	2020	0	0	0
	2021	0	0	0

QUESTION - 9

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,  
    MAXRecovered) AS Max_Recovered  
FROM  
corona_virus  
GROUP BY Year  
ORDER BY Year;
```

	Year	Max_Confirmed	Max_Deaths	Max_Recovered
▶	2020	823225	3752	1123456
	2021	414188	7374	422436

QUESTION - 10

Corona Virus Analysis

Find minimum values for confirmed, deaths, recovered per year

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

GROUP BY Year , Month

ORDER BY Year , Month;

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

QUESTION - 11

Corona Virus Analysis

**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;
```

	Total_confirmed_cases	Avg_confirmed_cases	Variance_confirmed_cases	Stdev_confirmed_cases
▶	169065144	2156.8283	157288925.07796532	12541.488152446875

QUESTION - 12

Corona Virus Analysis

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_death_cases,  
ROUND(VARIANCE(Deaths), 2) AS Variance_death_cases,  
ROUND(STDDEV(Deaths), 2) AS Stdev_death_cases
```

FROM

corona_virus

GROUP BY Year , Month

ORDER BY Year , Month;

	Month	Year	Total_death_cases	Avg_death_cases	Variance_death_cases	Stdev_death_cases
▶	1	2020	190	0.1234	4.25	2.06
	2	2020	2651	0.5936	68.32	8.27
	3	2020	41346	8.6607	3900.79	62.46
	4	2020	191833	41.5223	40504.27	201.26
	5	2020	144561	30.2809	20684.91	143.82
	6	2020	137757	29.8175	16929.45	130.11
	7	2020	167613	35.1096	21140.15	145.4
	8	2020	179200	37.5367	23273	152.55
	9	2020	160671	34.7773	20102.77	141.78
	10	2020	175484	36.7583	17580.07	132.59
	11	2020	262247	56.7634	27773.79	166.65
	12	2020	339996	71.2183	65345.37	255.63
	1	2021	401893	84.1837	102758.43	320.56
	2	2021	298239	69.1649	68478.87	261.68
	3	2021	282620	59.1998	54385.97	233.21
	4	2021	362387	78.4387	94611.47	307.59
	5	2021	366549	76.7803	131769.47	363
	6	2021	132657	66.2622	112963.67	336.1

QUESTION - 13

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,
    ROUND(VARIANCE(Recovered), 2) AS Variance_recovered_cases,
    ROUND(STDDEV(Recovered), 2) AS Stdev_recovered_cases
FROM
    corona_virus;
```

Total_recovered_cases	Avg_recovered_cases	Variance_recovered_cases	Stdev_recovered_cases
113089548	1442.7264	107029523.26	10345.51

QUESTION - 14

Corona Virus Analysis

Find Country having highest number of the Confirmed case.

```
SELECT  
    Country, SUM(Confirmed) AS Total_confirmed_cases  
FROM  
    corona_virus  
GROUP BY Country  
ORDER BY Total_confirmed_cases DESC  
LIMIT 1;
```

	Country	Total_confirmed_cases
▶	US	33461982

QUESTION - 14

Corona Virus Analysis

Find Country having highest number of the Confirmed case.

```
SELECT  
    Country, SUM(Confirmed) AS Total_confirmed_cases  
FROM  
    corona_virus  
GROUP BY Country  
ORDER BY Total_confirmed_cases DESC  
LIMIT 1;
```

	Country	Total_confirmed_cases
▶	US	33461982

QUESTION - 15

Corona Virus Analysis

Find Country having lowest number of the death case

```
WITH rankingCountry AS (
  SELECT
    Country,
    Sum(Deaths) AS Total_death_cases,
    RANK() OVER(ORDER BY SUM(Deaths) ASC) AS rank_no
  FROM
    corona_virus
  GROUP BY
    Country
)
SELECT
  Country , Total_death_cases
FROM
  rankingCountry
WHERE
  rank_no = 1;
```

	Country	Total_death_cases
▶	Dominica	0
	Kiribati	0
	Marshall Islands	0
	Samoa	0

QUESTION - 16

Corona Virus Analysis

Find top 5 countries having highest recovered case

```
SELECT Country,  
       SUM(Recovered) AS Total_Recovered  
  FROM corona_virus  
 GROUP BY Country  
 ORDER BY Total_Recovered DESC  
 LIMIT 5;
```

	Country	Total_Recovered
▶	India	28089649
	Brazil	15400169
	US	6303715
	Turkey	5202251
	Russia	4745756

-Purva Phalak

INSIGHTS AND FINDINGS

- The Total Number of Cases Confirmed were **16,90,65,144.**
- The Total Number of Cases Recovered were **11,30,89,548 .**
- The Total Number of Death Cases were **36,47,894**, With the Highest Death Cases Record of **4,01,893 in Jan 2021** as per the Analysis.
-
- The United States has the **highest number of confirmed cases.**
- The top five countries with the **highest number of recovered cases** are India, Brazil, the US, Turkey, and Russia.
- Countries like Dominica, Kiribati, Marshall Islands, and Samoa have the **lowest number of death cases.**

CONCLUSION

- The coronavirus analysis project provides valuable insights into the spread and impact of COVID-19 by examining detailed data on cases, recoveries, and deaths across various regions.
- These insights help health authorities and policymakers make informed decisions to manage and mitigate the pandemic's effects effectively.
- The project ultimately aims to save lives and guide efforts to safely return to normal activities.

**THANK
YOU !**