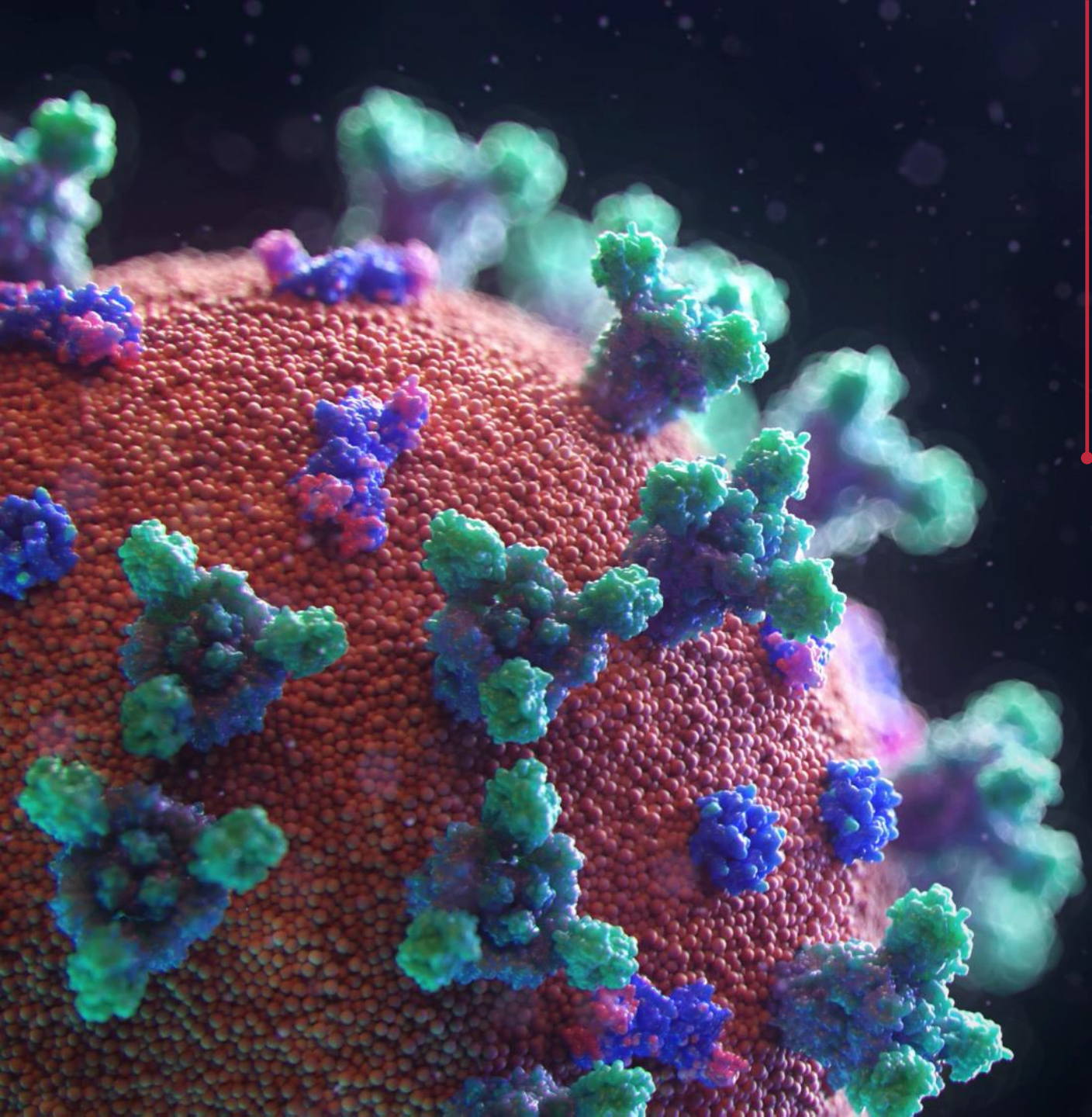


CORONAVIRUS / COVID-19

COVID-19

Corona Virus Analysis

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.

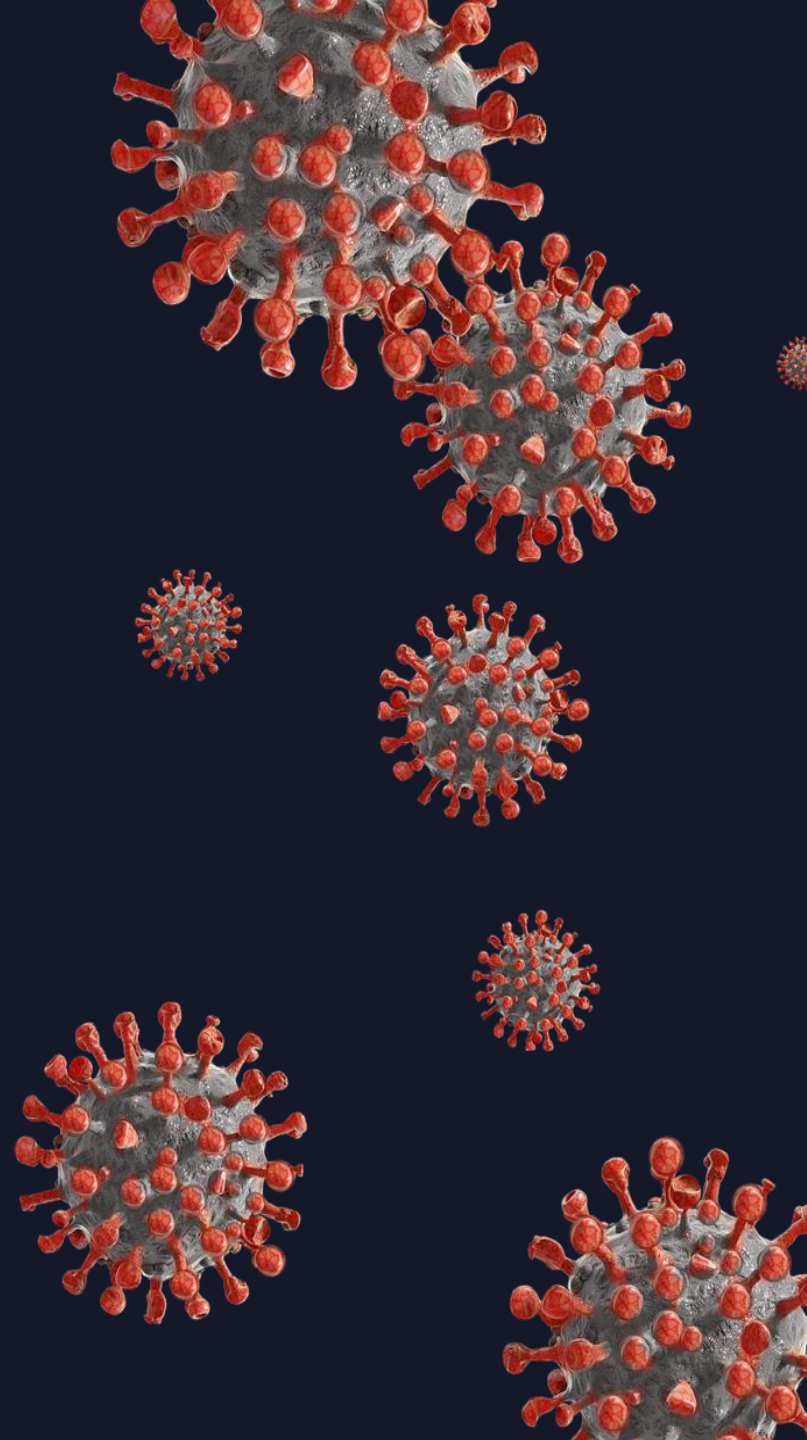
COVID - 19

DataSet

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 CORONA VIRUS data.
- Confirmed: Number of diagnosed CORONA VIRUS cases.
- Deaths: Number of CORONA VIRUS related deaths.
- Recovered: Number of recovered CORONA VIRUS cases.



COVID - 19

DataSet

Analysis Questions

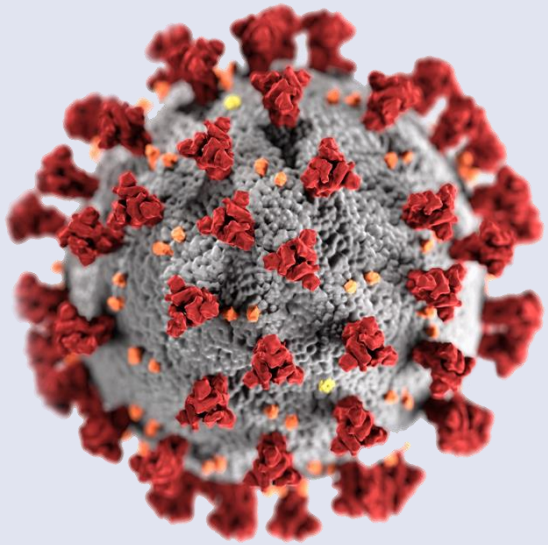
Dataset Contain 16 questions for performing detailed analysis:

1. Write a code to check NULL values
2. If NULL values are present, update them with zeros for all columns.
3. Check total number of rows
4. Check what is start_date and end_date
5. Number of month present in dataset
6. Find monthly average for confirmed, deaths, recovered
7. Find most frequent value for confirmed, deaths, recovered each month
8. Find minimum values for confirmed, deaths, recovered per year
9. Find maximum values of confirmed, deaths, recovered per year
10. The total number of case of confirmed, deaths, recovered each month
11. Check how corona virus spread out with respect to confirmed case (Eg.: total confirmed cases, their average, variance & STDEV)
12. Check how corona virus spread out with respect to death case per month (Eg.: total confirmed cases, their average, variance & STDEV)
13. Check how corona virus spread out with respect to recovered case (Eg.: total confirmed cases, their average, variance & STDEV)
14. Find Country having highest number of the Confirmed case
15. Find Country having lowest number of the death case
16. Find top 5 countries having highest recovered case

Dataset Columns

- Province
- Country/Region
- Latitude
- Longitude
- Date
- Confirmed
- Deaths
- Recovered

Data Description



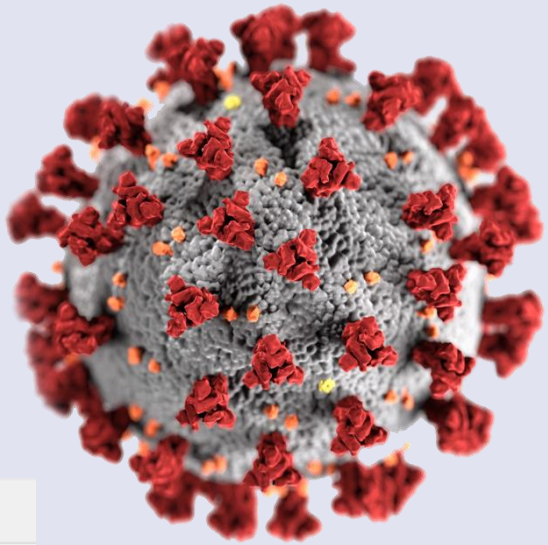
SQL Query

```
SELECT * FROM [Corona Virus Analysis]..['Corona Virus Dataset']
ORDER BY Date,[Country/Region]
```

Results | Messages

	Province	Country/Region	Latitude	Longitude	Date	Confirmed	Deaths	Recovered
1	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-22 00:00:00.000	0	0	0
2	Algeria	Algeria	28.0339	1.6596	2020-01-22 00:00:00.000	0	0	0
3	Argentina	Argentina	-38.4161	-63.6167	2020-01-22 00:00:00.000	0	0	0
4	Queensland	Australia	-27.4698	153.0251	2020-01-22 00:00:00.000	0	0	0
5	Australian Capital Territory	Australia	-35.4735	149.0124	2020-01-22 00:00:00.000	0	0	0
6	South Australia	Australia	-34.9285	138.6007	2020-01-22 00:00:00.000	0	0	0
7	New South Wales	Australia	-33.8688	151.2093	2020-01-22 00:00:00.000	0	0	0
8	Tasmania	Australia	-42.8821	147.3272	2020-01-22 00:00:00.000	0	0	0
9	Northern Territory	Australia	-12.4634	130.8456	2020-01-22 00:00:00.000	0	0	0
10	Victoria	Australia	-37.8136	144.9631	2020-01-22 00:00:00.000	0	0	0
11	Western Australia	Australia	-31.9505	115.8605	2020-01-22 00:00:00.000	0	0	0
12	Austria	Austria	47.5162	14.5501	2020-01-22 00:00:00.000	0	0	0
13	Bangladesh	Bangladesh	23.685	90.3563	2020-01-22 00:00:00.000	0	0	0
14	Barbados	Barbados	13.1939	-59.5432	2020-01-22 00:00:00.000	0	0	0

01 Write a code to check NULL values.



SQL Query

```
SELECT COUNT(*) AS Null_Values FROM
[Corona Virus Analysis]..['Corona Virus
Dataset'] AS cvd
WHERE cvd.Province IS NULL
OR cvd.[Country/Region] IS NULL
OR cvd.Latitude IS NULL
OR cvd.Longitude IS NULL
OR cvd .Date IS NULL
OR cvd.Confirmed IS NULL
OR cvd.Deaths IS NULL
OR cvd.Recovered IS NULL
```

Results | Messages

	Null_Values
1	0

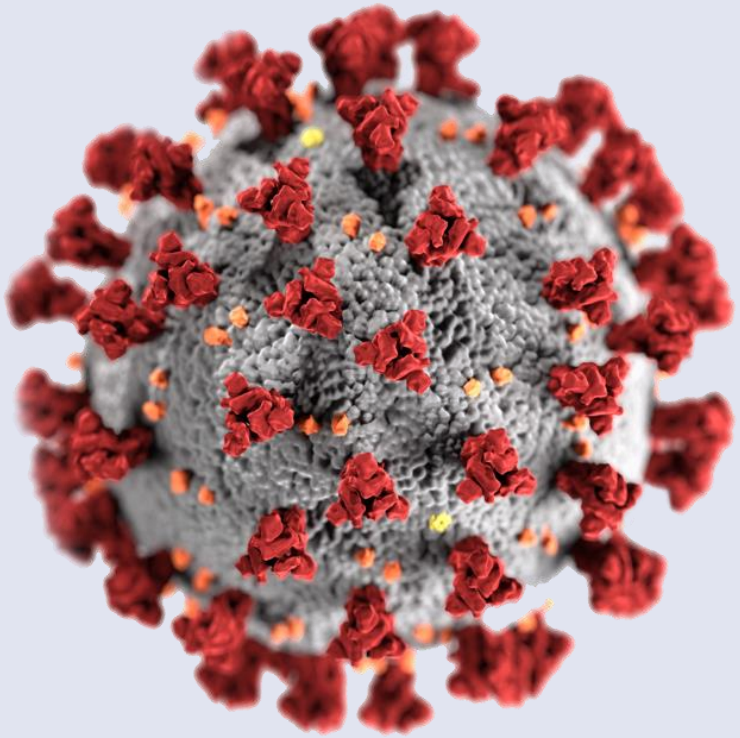
02 Check total number of rows.

SQL Query

```
SELECT COUNT(*) AS Toal_Rows
FROM [Corona Virus Analysis]..['Corona Virus Dataset']
```

Results | Messages

	Toal_Rows
1	78386



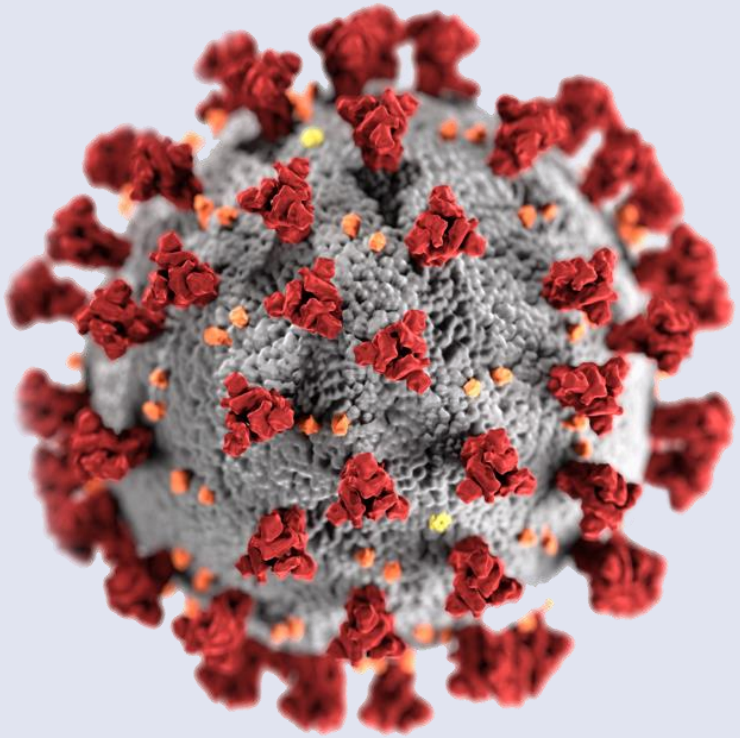
03 Check what is start_date and end_date.

SQL Query

```
SELECT MIN(Date) AS Start_Date , MAX(Date) AS End_Date
FROM [Corona Virus Analysis]..['Corona Virus Dataset']
```

Results | Messages

	Start_Date	End_Date
1	2020-01-22 00:00:00.000	2021-06-13 00:00:00.000



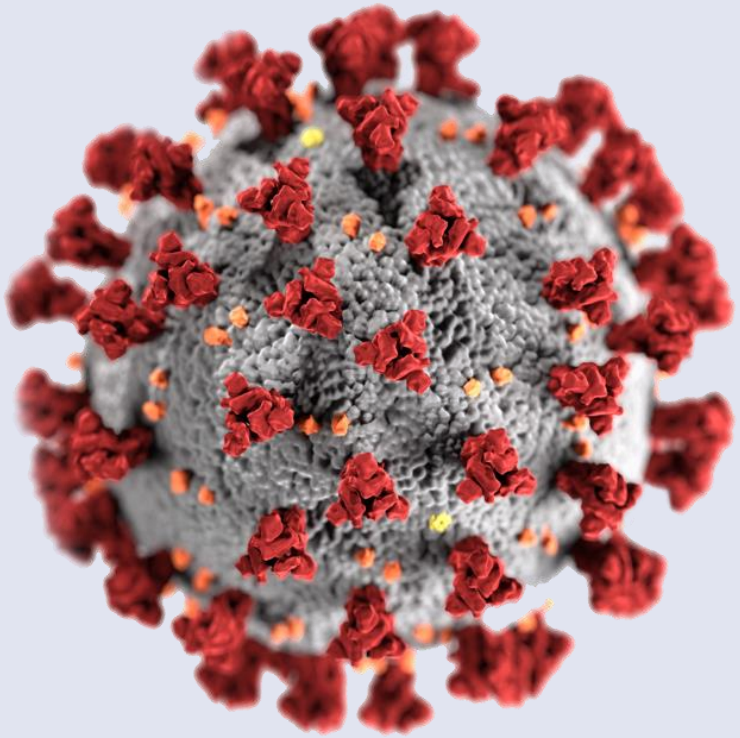
04 Number of month present in dataset

SQL Query

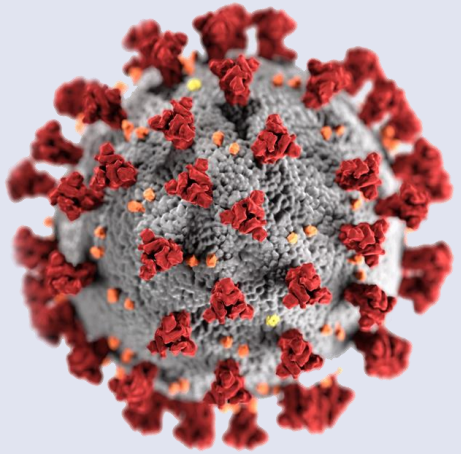
```
SELECT COUNT(DISTINCT CONCAT(YEAR(Date),'-',MONTH(Date))) AS
Total_Month
FROM [Corona Virus Analysis]..['Corona Virus Dataset']
```

Results | Messages

	Total_Month
1	18



05 Find monthly average for confirmed, deaths, recovered.



SQL Query

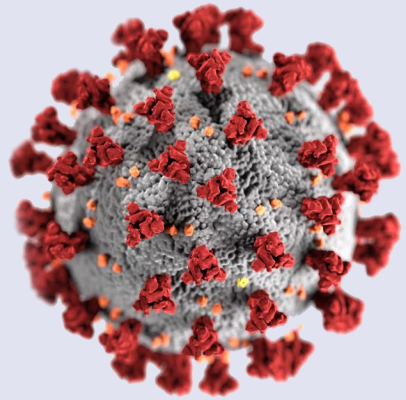
```
WITH Average_Calculation AS (
SELECT YEAR(Date) AS Year,MONTH(Date) AS
Month,
SUM(Confirmed) AS
Total_Confirmed,SUM(Deaths) AS Total_Deaths,
SUM(Recovered) AS Total_Recovers,COUNT(*) AS
Dayscount
FROM [Corona Virus Analysis]..['Corona Virus
Dataset']
GROUP BY YEAR(Date),MONTH(Date))
SELECT Year,Month,
Total_Confirmed/Dayscount AS
Average_Confirmed,
Total_Deaths/Dayscount AS Average_Deaths,
Total_Recovers/Dayscount AS
Average_Recovered
FROM Average_Calculation
ORDER BY Year,Month
```

Results | Messages

	Year	Month	Average_Confirmed	Average_Deaths	Average_Recovered
1	2020	1	4.14545454545455	0.123376623376623	0.0928571428571429
2	2020	2	15.2960143304971	0.5935960591133	7.0320197044335
3	2020	3	161.130289065773	8.66066191872643	27.8739002932551
4	2020	4	505.800432900433	41.5222943722944	171.642207792208
5	2020	5	574.849811478844	30.280896522832	318.296397151236
6	2020	6	859.228138528139	29.8175324675325	548.791558441558
7	2020	7	1432.36112274822	35.109551738584	983.05823209049
8	2020	8	1611.84289903645	37.5366568914956	1299.29472140762
9	2020	9	1784.58744588745	34.7772727272727	1438.90670995671
10	2020	10	2412.19962295769	36.7582739840804	1420.64306661081
11	2020	11	3592.19437229437	56.7634199134199	1985.34458874459
12	2020	12	4050.43967323	71.2182656053624	2497.88500209468
13	2021	1	3911.22852953498	84.1837033933808	1919.63699204022
14	2021	2	2433.36363636364	69.1648886827458	1558.39169758813
15	2021	3	2916.79723502304	59.1998324256389	1652.28592375367
16	2021	4	4699.35519480519	78.4387445887446	3074.78506493506
17	2021	5	4005.25408462505	76.7802681189778	4007.5077503142
18	2021	6	2508.63236763237	66.2622377622378	2769.44955044955

06

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



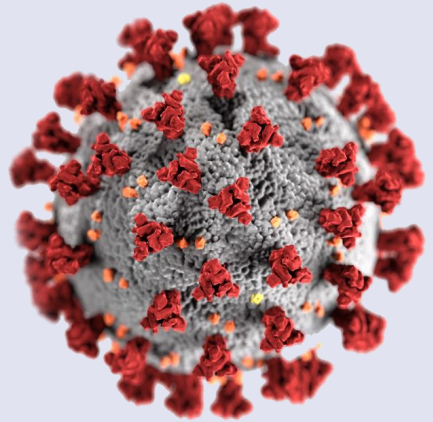
SQL Query

```
WITH MOSTFREQUENTVALUE AS(
SELECT YEAR(Date) AS
Year,MONTH(Date) AS
Month,MAX(Confirmed) AS
Max_Confirmed_Value,
MAX(Deaths) AS
Max_Death_Value,MAX(Recovered) AS
Max_Recovered_Value
FROM [Corona Virus
Analysis]..['Corona Virus Dataset']
GROUP BY YEAR(Date),MONTH(Date))
SELECT
Year,Month,Max_Confirmed_Value,Max_D
eath_Value,Max_Recovered_Value
FROM MOSTFREQUENTVALUE
ORDER BY Year,Month
```

Results | Messages

	Year	Month	Max_Confirmed_Value	Max_Death_Value	Max_Recovered_Value
1	2020	1	2131	49	51
2	2020	2	14840	242	3418
3	2020	3	26314	1085	4289
4	2020	4	50740	2607	33227
5	2020	5	34907	2309	51717
6	2020	6	54771	2003	94305
7	2020	7	75866	1595	140050
8	2020	8	85687	1505	95881
9	2020	9	97894	1703	101468
10	2020	10	99264	3351	388340
11	2020	11	207933	2259	139292
12	2020	12	823225	3752	1123456
13	2021	1	300462	4475	87090
14	2021	2	134975	3907	98389
15	2021	3	100158	3869	102138
16	2021	4	401993	4249	299988
17	2021	5	414188	4529	422436
18	2021	6	134154	7374	231456

07 Find minimum values for confirmed, deaths, recovered per year



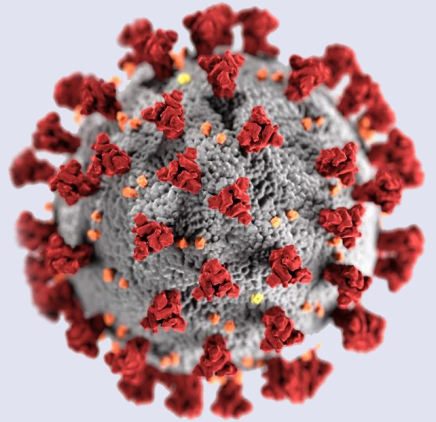
SQL Query

```
WITH MINFREQUENTVALUE AS(
SELECT YEAR(Date) AS Year,
MIN(CASE WHEN Confirmed > 0 THEN Confirmed ELSE NULL
END) AS Min_Confirmed_Value,
MIN(CASE WHEN Deaths > 0 THEN Deaths ELSE NULL END)
AS Min_Death_Value,
MIN(CASE WHEN Recovered > 0 THEN Recovered ELSE NULL
END) AS Min_Recovered_Value
FROM [Corona Virus Analysis]..['Corona Virus
Dataset']
GROUP BY YEAR(Date))
SELECT Year,Min_Confirmed_Value,Min_Death_Value,
Min_Recovered_Value
FROM MINFREQUENTVALUE
ORDER BY Year
```

Results | Messages

	Year	Min_Confirmed_Value	Min_Death_Value	Min_Recovered_Value
1	2020	1	1	1
2	2021	1	1	1

08 Find maximum values of confirmed, deaths, recovered per year



SQL Query

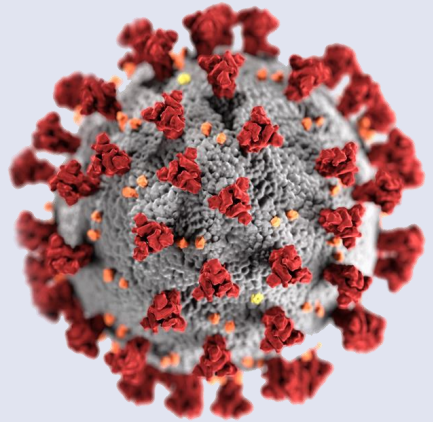
```
WITH MAXVALUE AS(
SELECT YEAR(Date) AS Year,MAX(Confirmed) AS
Max_Confirmed_Value,
MAX(Deaths) AS Max_Death_Value,MAX(Recovered) AS
Max_Recovered_Value
FROM [Corona Virus Analysis]..['Corona Virus
Dataset']
GROUP BY YEAR(Date))
SELECT Year,Max_Confirmed_Value,Max_Death_Value,
Max_Recovered_Value
FROM MAXVALUE
ORDER BY Year
```

Results | Messages

	Year	Max_Confirmed_Value	Max_Death_Value	Max_Recovered_Value
1	2020	823225	3752	1123456
2	2021	414188	7374	422436

09

The total number of case of confirmed, deaths, recovered each month



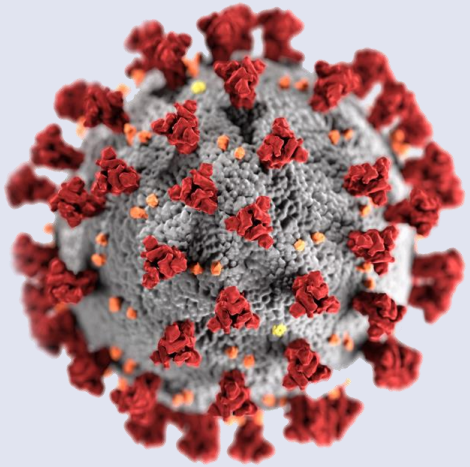
SQL Query

```
WITH TOTALCASES AS(
SELECT YEAR(Date) AS Year,MONTH(Date) AS
Month,SUM(Confirmed) AS Total_Confirmed_Value,
SUM(Deaths) AS
Total_Death_Value,SUM(Recovered) AS
Total_Recovered_Value
FROM [Corona Virus Analysis]..['Corona Virus
Dataset']
GROUP BY YEAR(Date),MONTH(Date))
SELECT
Year,Month,Total_Confirmed_Value,Total_Death_V
alue,Total_Recovered_Value
FROM TOTALCASES
ORDER BY Year,Month
```

Results | Messages

	Year	Month	Total_Confirmed_Value	Total_Death_Value	Total_Recovered_Value
1	2020	1	6384	190	143
2	2020	2	68312	2651	31405
3	2020	3	769236	41346	133070
4	2020	4	2336798	191833	792987
5	2020	5	2744333	144561	1519547
6	2020	6	3969634	137757	2535417
7	2020	7	6838092	167613	4693120
8	2020	8	7694938	179200	6202833
9	2020	9	8244794	160671	6647749
10	2020	10	11515841	175484	6782150
11	2020	11	16595938	262247	9172292
12	2020	12	19336799	339996	11924903
13	2021	1	18672205	401893	9164347
14	2021	2	10492664	298239	6719785
15	2021	3	13924790	282620	7888013
16	2021	4	21711021	362387	14205507
17	2021	5	19121083	366549	19131842
18	2021	6	5022282	132657	5544438

10 Check how corona virus spread out with respect to confirmed case



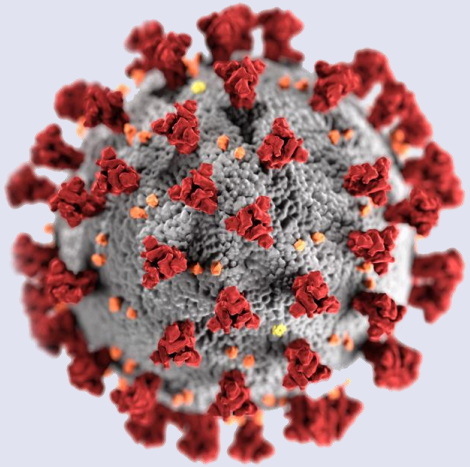
SQL Query

```
SELECT
    SUM(Confirmed) AS Total_Confirmed_Cases,
    AVG(Confirmed) AS Average_Confirmed_Cases,
    VAR(Confirmed) AS Variance_Confirmed_Cases,
    STDEV(Confirmed) AS Stdev_Confirmed_Cases
FROM [Corona Virus Analysis]..['Corona Virus Dataset']
WHERE Confirmed > 0;
```

Results | Messages

	Total_Confirmed_Cases	Average_Confirmed_Cases	Variance_Confirmed_Cases	Stdev_Confirmed_Cases
1	169065144	3230.37954753898	232115211.365141	15235.327740654

11 Check how corona virus spread out with respect to death case per month



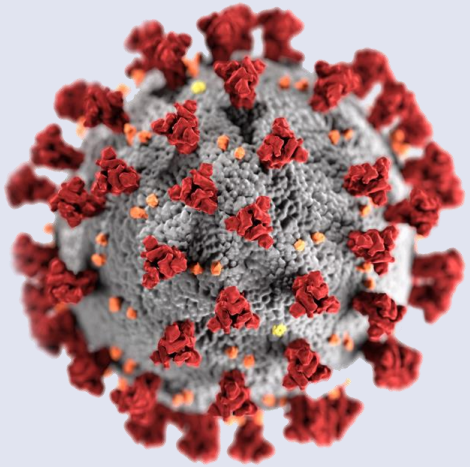
SQL Query

```
WITH DEATHPERMONTH AS(
SELECT YEAR(Date) AS Year, MONTH(Date) AS
Month,
    SUM(Deaths) AS
Total_Deaths_Cases,AVG(Deaths) AS
Average_Deaths_Cases,
    VAR(Deaths) AS
Variance_Deaths_Cases,STDEV(Deaths) AS
Stdev_Deaths_Cases
FROM [Corona Virus Analysis]..['Corona
Virus Dataset'] WHERE Deaths >0 GROUP BY
YEAR(Date),MONTH(Date))
SELECTYear,Month,Total_Deaths_Cases,Averag
e_Deaths_Cases,Variance_Deaths_Cases,
Stdev_Deaths_Cases FROM DEATHPERMONTH
ORDER BY Year,Month
```

Results | Messages

	Year	Month	Total_Deaths_Cases	Average_Deaths_Cases	Variance_Deaths_Cases	Stdev_Deaths_Cases
1	2020	1	190	19	328	18.1107702762748
2	2020	2	2651	28.8152173913043	2530.87756808409	50.3078280994528
3	2020	3	41346	43.1136600625652	17951.8670261515	133.984577568284
4	2020	4	191833	91.0023719165085	84308.4038861606	290.359094719212
5	2020	5	144561	72.4616541353384	46465.4342109034	215.558424124188
6	2020	6	137757	73.4704	38527.5085950907	196.284254577617
7	2020	7	167613	81.3655339805825	45250.1193554699	212.720754406969
8	2020	8	179200	80.0357302367128	46241.9352732713	215.039380749832
9	2020	9	160671	70.7490092470277	38367.9202358021	195.877309139681
10	2020	10	175484	72.5440264572137	32112.2547812995	179.198925167813
11	2020	11	262247	107.039591836735	47011.0866311114	216.820401787081
12	2020	12	339996	131.070161912105	112460.234180691	335.350912002176
13	2021	1	401893	151.429163526752	174724.190042496	418.000227323499
14	2021	2	298239	121.979141104294	114373.823214483	338.191991647471
15	2021	3	282620	104.790507971821	91525.8837674347	302.532450767574
16	2021	4	362387	136.749811320755	157030.510805929	396.270754416635
17	2021	5	366549	136.874159820762	226761.333803459	476.194638570678
18	2021	6	132657	119.726534296029	197886.955862594	444.844867186971

12 Check how corona virus spread out with respect to recovered case



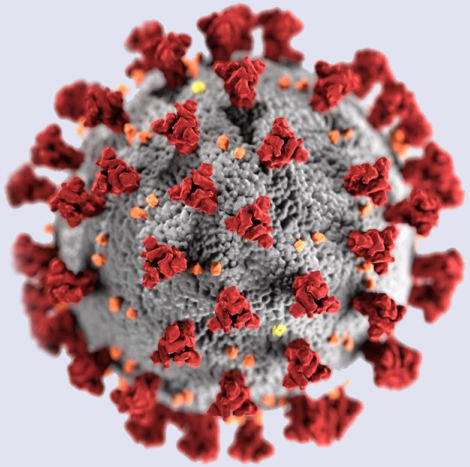
SQL Query

```
SELECT
    SUM(Recovered) AS Total_Recovered_Cases,
    AVG(Recovered) AS Average_Recovered_Cases,
    VAR(Recovered) AS Variance_Recovered_Cases,
    STDEV(Recovered) AS Stdev_Recovered_Cases
FROM [Corona Virus Analysis]..['Corona Virus Dataset']
WHERE Recovered > 0;
```

Results | Messages

	Total_Recovered_Cases	Average_Recovered_Cases	Variance_Recovered_Cases	Stdev_Recovered_Cases
1	113089548	2706.98106613687	197401431.839047	14049.9619871033

13 Find Country having highest number of the Confirmed case



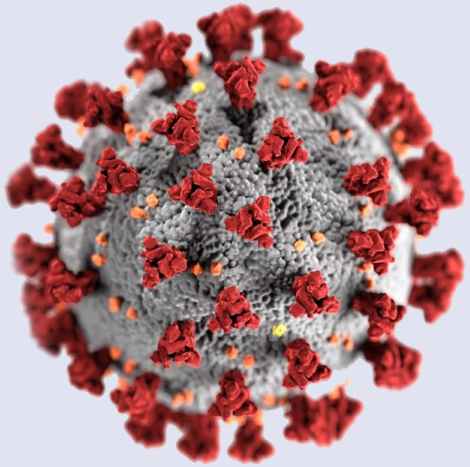
SQL Query

```
SELECT TOP 1 [Country/Region],
SUM(Confirmed) AS
MAXTotal_Confirmed
FROM [Corona Virus
Analysis]..['Corona Virus
Dataset']
Group BY [Country/Region]
ORDER BY MAXTotal_Confirmed DESC
```

Results | Messages

	Country/Region	MAXTotal_Confirmed
1	US	33461982

14 Find Country having lowest number of the death case



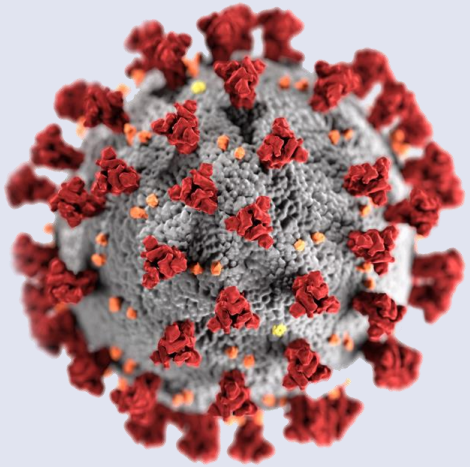
SQL Query

```
SELECT TOP 1 [Country/Region],
SUM(Deaths) AS Lowest_Death_Cases
FROM [Corona Virus
Analysis]..['Corona Virus
Dataset']
Group BY [Country/Region]
ORDER BY Lowest_Death_Cases
```

Results | Messages

	Country/Region	Lowest_Death_Cases
1	Marshall Islands	0

15 Find top 5 countries having highest recovered case

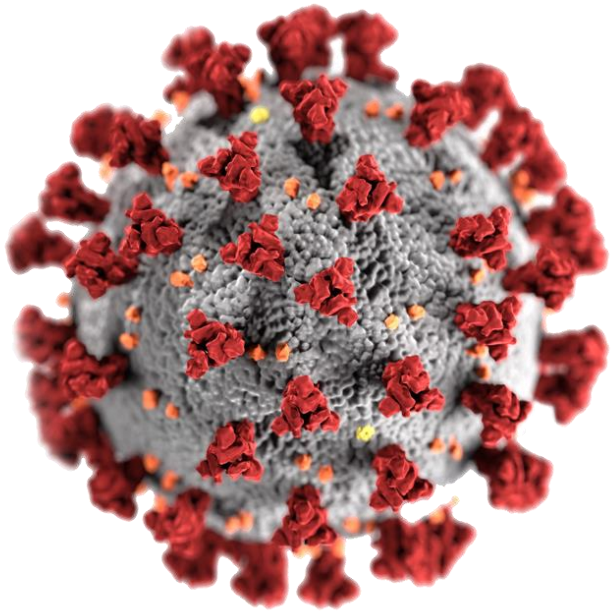


SQL Query

```
SELECT TOP 5 [Country/Region],
SUM(Recovered) AS
MAX_Recovered_Cases
FROM [Corona Virus
Analysis]..['Corona Virus
Dataset']
Group BY [Country/Region]
ORDER BY MAX_Recovered_Cases DESC
```

Results | Messages

	Country/Region	MAX_Recovered_Cases
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756



Thank YOU!

Key points:

- **Data Quality**: Ensure completeness by checking for **NULL** values and total rows.
- **Temporal Insights**: Determine the date range, number of months, and monthly trends.
- **Descriptive Statistics**: Calculate averages, most frequent, minimum, and maximum values.
- **Spread Analysis**: Evaluate how the virus spread over time using total, average, variance, and standard deviation.
- **Geographical Insights**: Identify countries most and least affected by confirmed, death, and recovered cases.