CORONA VIRUS ANALYSIS

BY

RIYA SINGH

TABLE CONTENTS

project Overview

The Coronavirus Disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus was first identified in December 2019 in Wuhan, Hubei Province, China, and has since spread globally, leading to an ongoing pandemic. This project aims to analyze the data related to COVID-19 to understand its spread, impact, and the effectiveness of various interventions. Coronaviruses are a large family of viruses known to cause illnesses ranging from the common cold to more severe diseases such as the Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). SARS-CoV-2 is a novel coronavirus that had not been previously identified in humans before the outbreak in Wuhan.
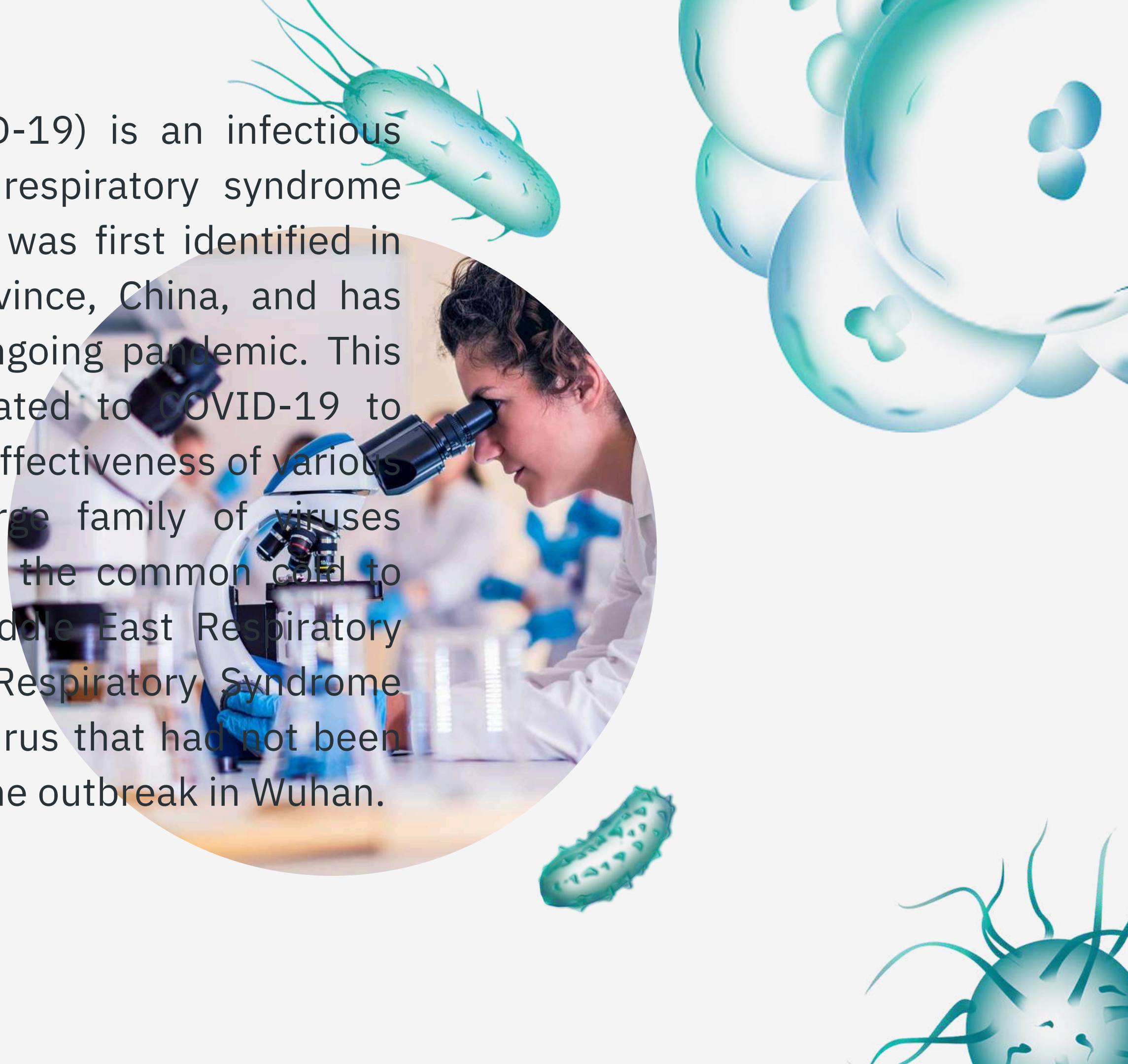
Data Description

01 Province:
Geographic subdivision within a country/region.

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

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

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

05 Date:
Recorded date of CORONA VIRUS data.

06 Confirmed:
Number of diagnosed CORONA VIRUS cases.

07 Deaths:
Number of CORONA VIRUS related deaths.s.

08 Recovered:
Number of recovered CORONA VIRUS cases..

01
02
03
04
05
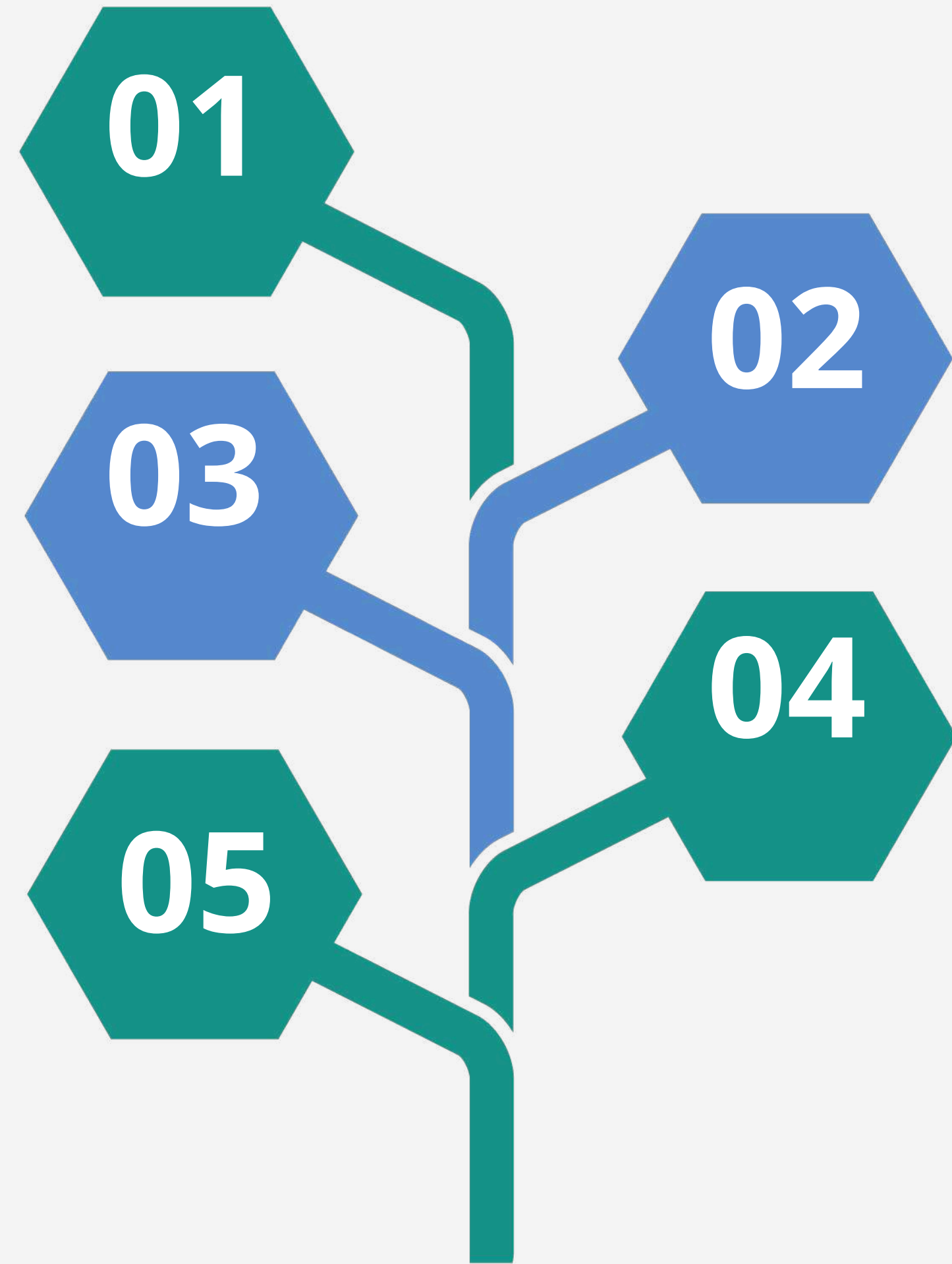
Exploratory Data Analysis

Q1. Write a code to check NULL values

```sql
select * from coronavirus
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_Region	Latitude	Longitude	Date	Confirmed	Deaths	Recovered

Q.2 . If null value are present ,update them with Zeros for all columns

ther is no null values present in this dataset
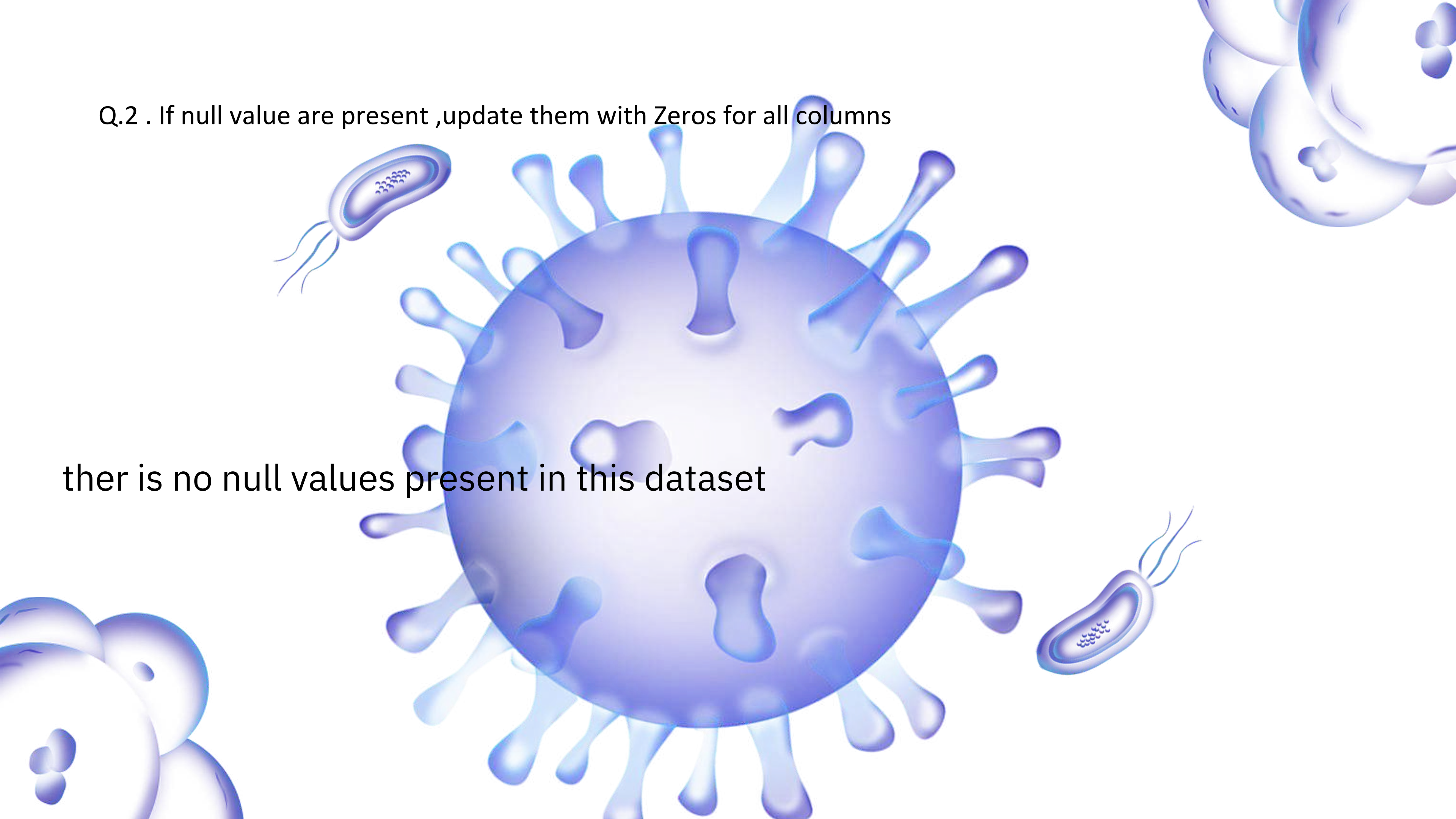
3. check total number of rows.

```sql
select count(*) 'Total_rows'from CoronaVirus
```

	Total_rows
1	78396
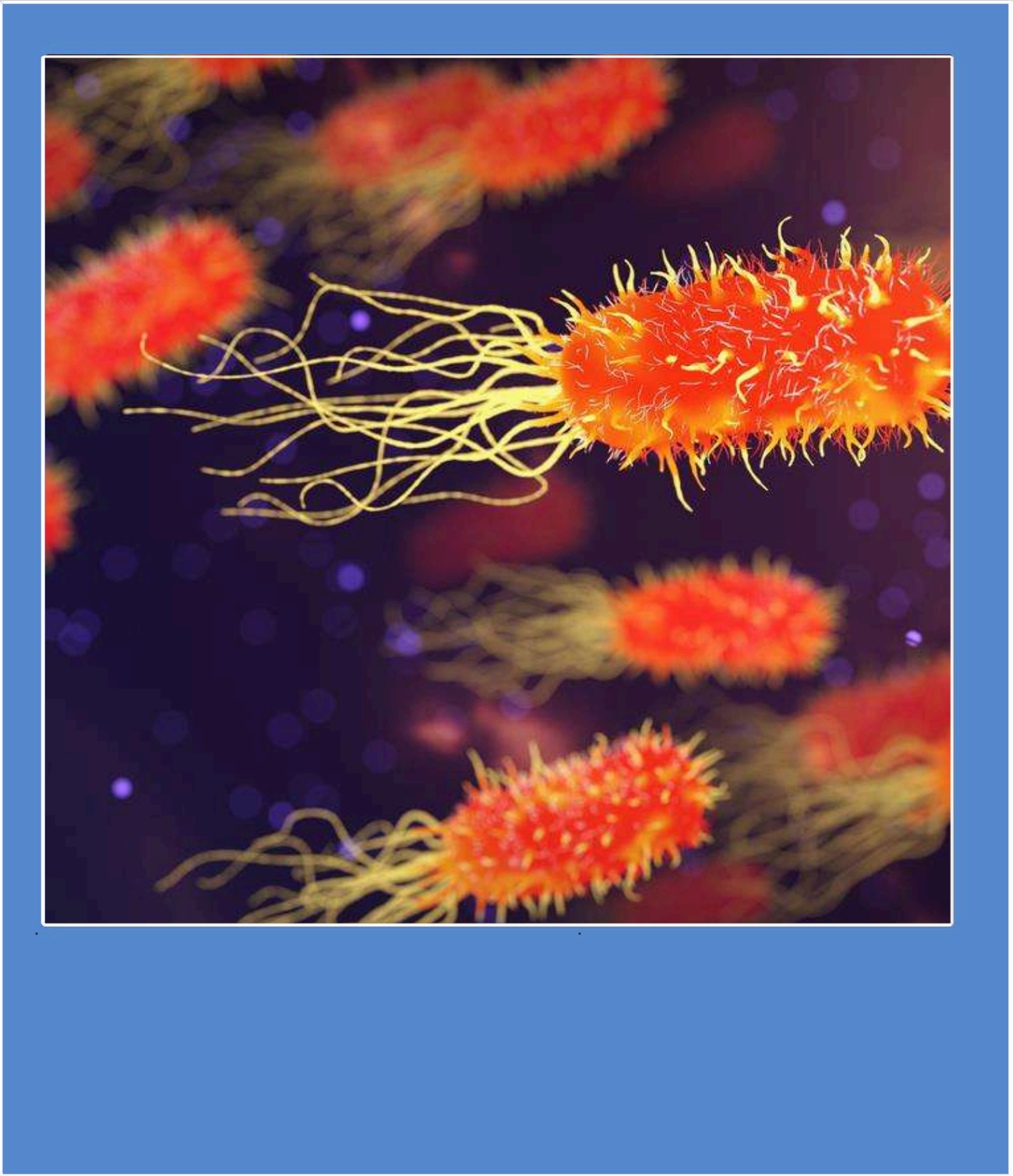
4. Check what is start_date and end_date

```sql
select min(date) 'start_date', max(date) 'end date' from [CoronaVirus ]
```

	start_date	end date
1	2020-01-22	2021-06-13

J

Q5. Number of month present in dataset

```sql
select count(distinct concat(year(date),' ',month(date)))
'Total_month' from [CoronaVirus ]
```

	Total_month
1	18

Q6. Find monthly average for confirmed, deaths, recovered

```sql
select datepart(year,date) 'year',
datepart(month,date)'month_count', DATENAME(month,date)'month_name',
avg(confirmed) 'Average_confirmed case',
avg(Deaths) 'Average_deaths case',
avg(Recovered)'Average_Recovered case' from [CoronaVirus ]
group by datepart(year,date) , datepart(month,date),
DAT ENAME(mon th,dat e)
order by year ,month_count
```
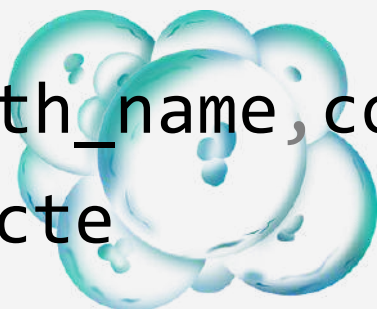
|	year	month_count	month_name	Average_confirmed case	Average_deaths case	Average_Recovered case
1	2020	1	January	4	0	0
2	2020	2	February	15	0	7
3	2020	3	March	161	8	27
4	2020	4	April	505	41	171
5	2020	5	May	574	30	318
6	2020	6	June	859	29	548
7	2020	7	July	1432	35	983
8	2020	8	August	1611	37	1299
9	2020	9	September	1784	34	1438
10	2020	10	October	2412	36	1420
11	2020	11	November	3592	56	1985
12	2020	12	December	4050	71	2497
13	2021	1	January	3911	84	1919
14	2021	2	February	2433	69	1558
15	2021	3	March	2916	59	1652
16	2021	4	April	4699	78	3074
17	2021	5	May	4005	76	4007
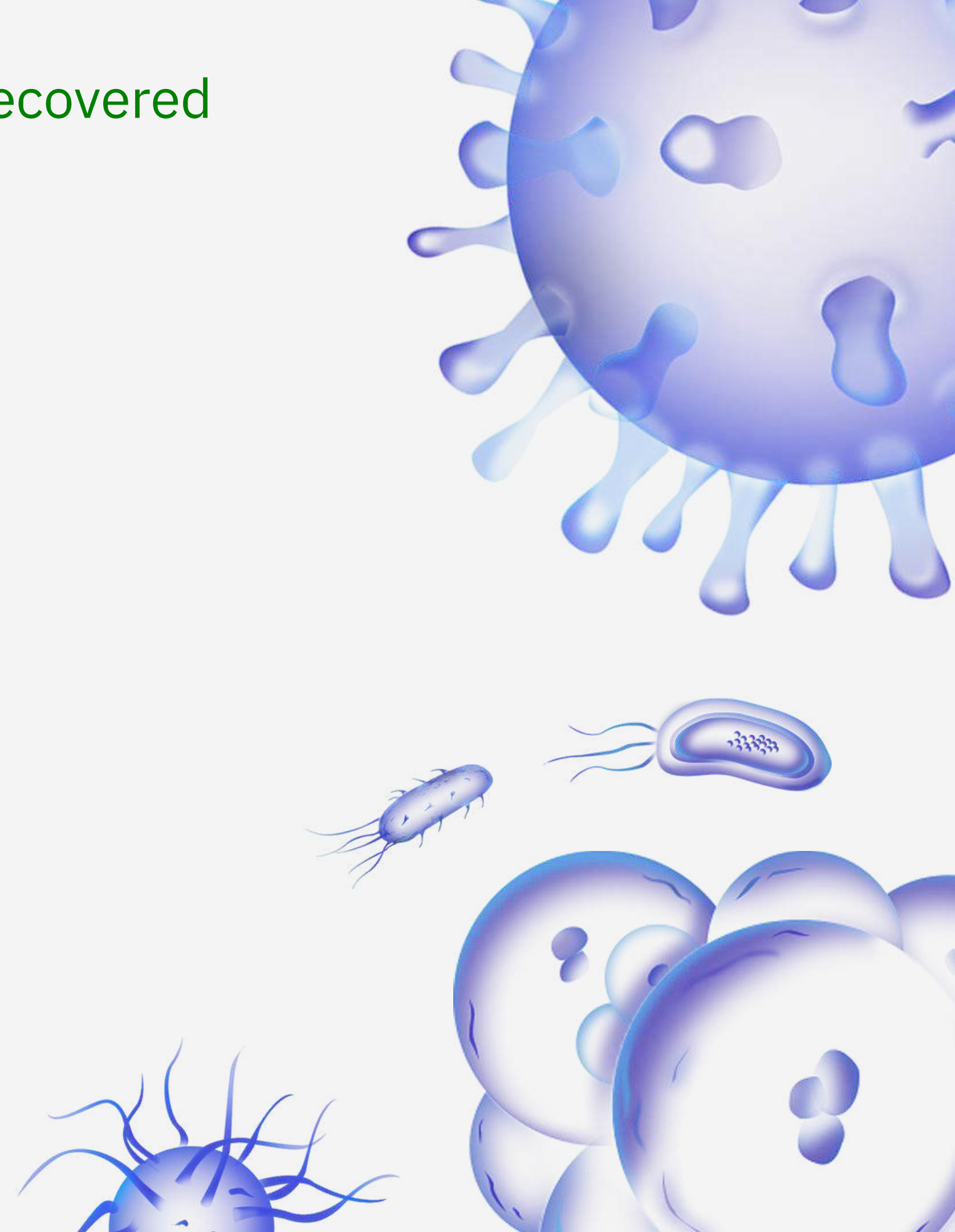
Q7. Find most frequent value for confirmed, deaths, recovered each month

```sql
with cte as (select
datepart(month,date)'month_count',
DATENAME(month,date)'month_name',
confirmed,
D ea th s,
Recovered, rank() over(partition by
datepart(month,date),datename(month,date)
order by count(*) desc) as rnk
from [CoronaVirus ]
group by datepart(year,date) ,
datepart(month,date),
DATENAME(month,date),confirmed,
Deaths, Recovered)

select month_count,month_name,confirmed,
Deaths,Recovered from cte
where rnk =1
```

	month_count	month_name	confirmed	Deaths	Recovered
1	1	January	0	0	0
2	2	February	0	0	0
3	3	March	0	0	0
4	4	April	0	0	0
5	5	May	0	0	0
6	6	June	0	0	0
7	7	July	0	0	0
8	8	August	0	0	0
9	9	September	0	0	0
10	10	October	0	0	0
11	11	November	0	0	0
12	12	December	0	0	0
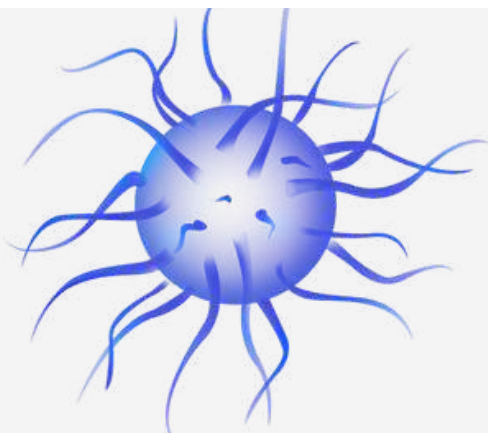
Q8. Find minimum values for confirmed, deaths, recovered per year

```sql
select DATEPART(year,date), min(confirmed)
'minimum confirmed case',
min(deaths) 'minimum deaths case',
min(recovered) 'minimum recoverd case'from
[CoronaVirus ]
group by DATEPART(year,date)
```

	(No column name)	minimum confirmed case	minimum deaths case	minimum recoverd case
1	2021	0	0	0
2	2020	0	0	0
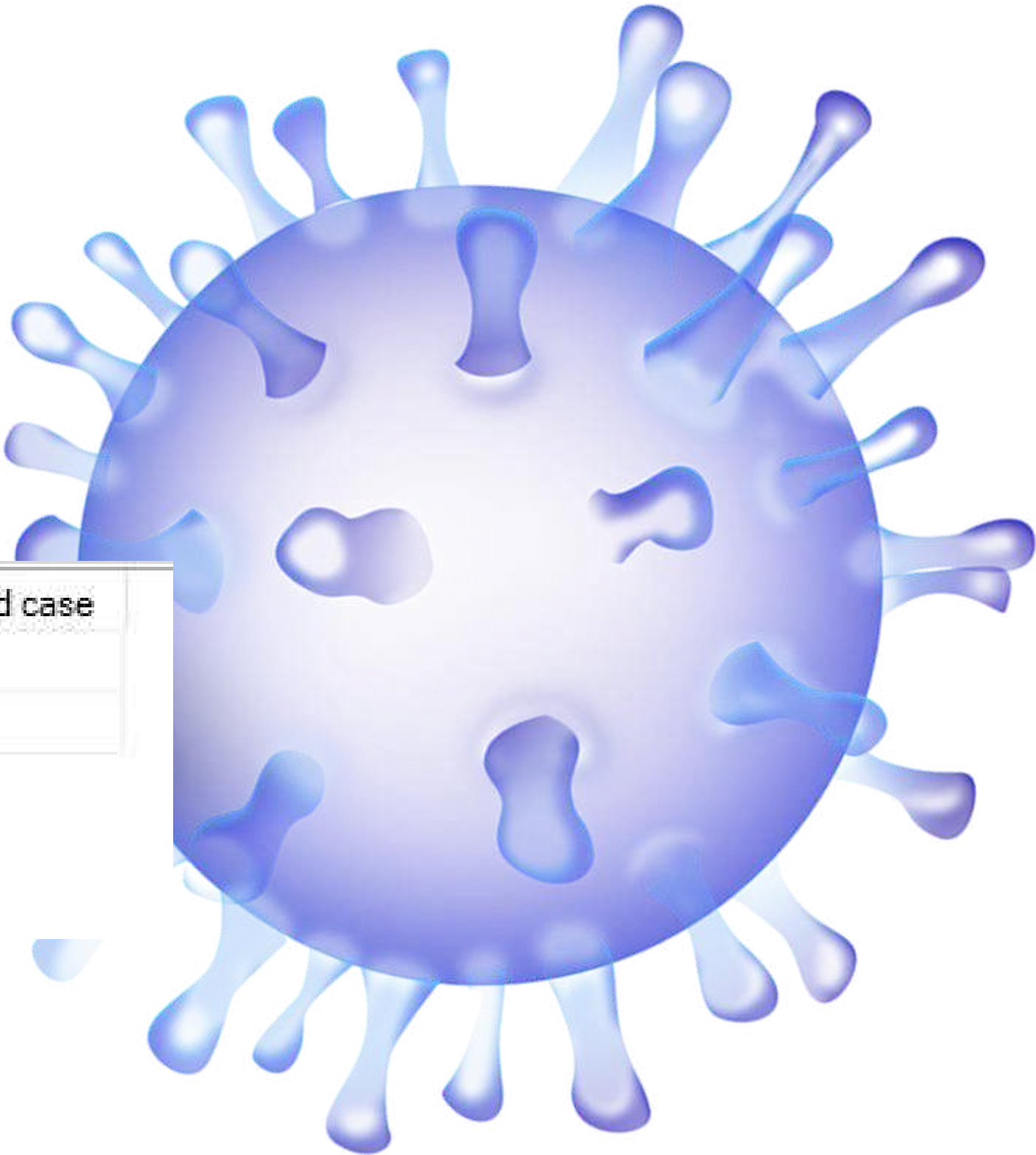
Q9. Find maximum values of confirmed, deaths, recovered per year

```sql
select DATEPART(year,date), max(confirmed)
'maximum confirmed case',
max(deaths) 'maximum deaths case',
max(recovered) 'maximum recoverd case'from
[CoronaVirus ]
group by DATEPART(year,date)
```

	(No column name)	maximum confirmed case	maximum deaths case	maximum recoverd case
1	2021	414188	7374	422436
2	2020	823225	3752	1123456
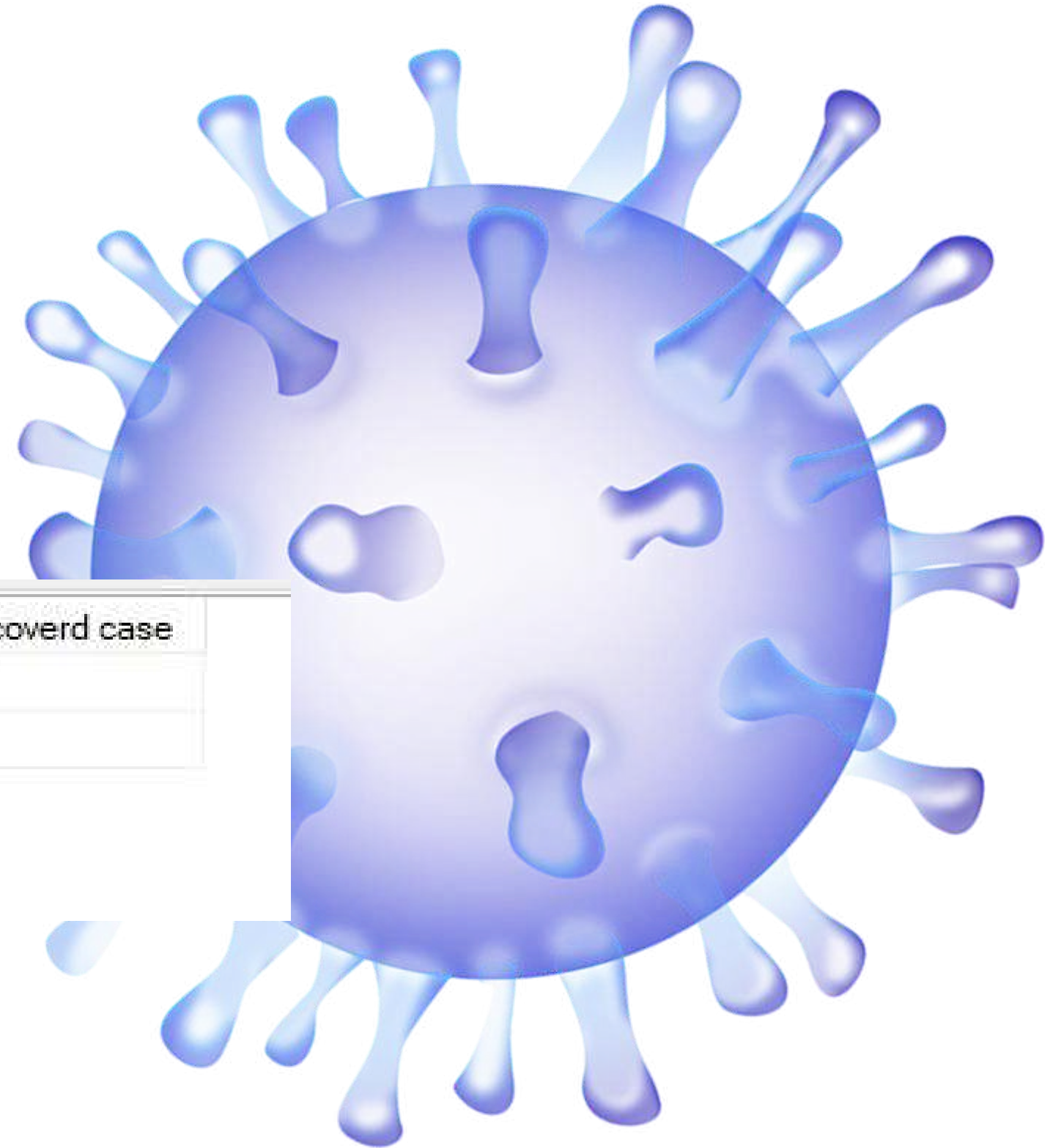
Q10. The total number of case of confirmed, deaths, recovered each month

```sql
select DATEPART(year,date) 'year',DATEPART(MONTH,date)
'month_count',DATENAME(MONTH,date) 'month_name',
sum(confirmed) 'Totla_confirmed_case',sum(deaths)
'Totla_deaths_case',sum(recovered) 'Totla_recovered_case'
from [CoronaVirus ]
group by
DATEPART(year,date),DATEPART(MONTH,date),DATENAME(MONTH,date)
order by year,month_count
```

	year	month_count	month_name	Totla_confirmed_case	Totla_deaths_case	Totla_recovered_case
1	2020	1	January	6384	190	143
2	2020	2	February	68312	2651	31405
3	2020	3	March	769236	41346	133070
4	2020	4	April	2336798	191833	792987
5	2020	5	May	2744333	144561	1519547
6	2020	6	June	3969634	137757	2535417
7	2020	7	July	6838092	167613	4693120
8	2020	8	August	7694938	179200	6202833
9	2020	9	September	8244794	160671	6647749
10	2020	10	October	11515841	175484	6782150
11	2020	11	November	16595938	262247	9172292
12	2020	12	December	19336799	339996	11924903
13	2021	1	January	18672205	401893	9164347
14	2021	2	February	10492664	298239	6719785
15	2021	3	March	13924790	282620	7888013
16	2021	4	April	21711021	362387	14205507
17	2021	5	May	19121083	366549	19131842
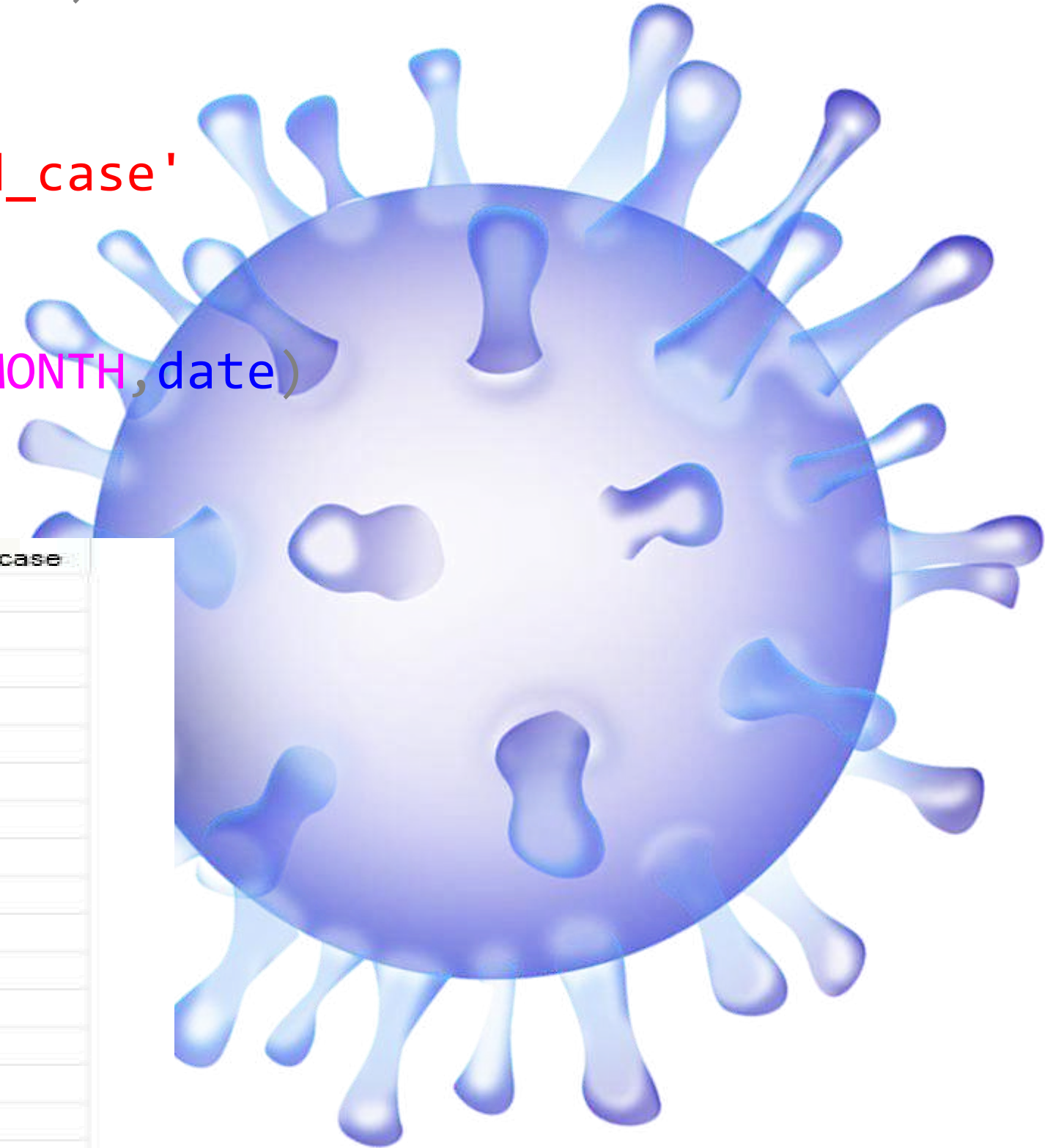
-Q11. Check how corona virus spread out with respect to confirmed case
--- (Eg.: total confirmed cases, their average, variance & STDEV)

```sql
select sum(confirmed) 'total confirmed case',
avg(confirmed) 'avg confirmed case',
var(confirmed)'varians confirmed case',
stdev(confirmed) 'standared deviation confirmed case'
from [CoronaVirus ]
```
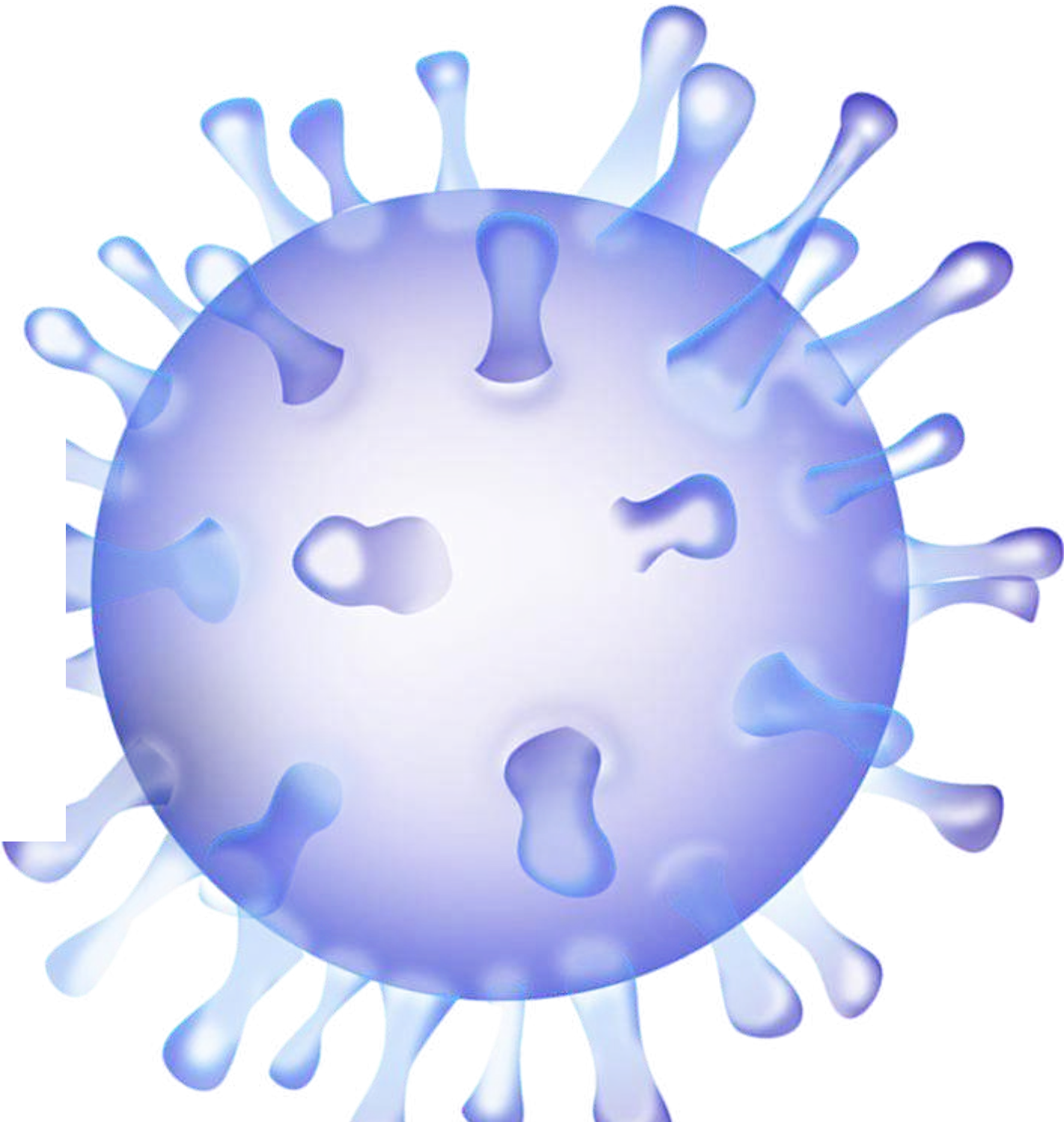
	total confirmed case	avg confirmed case	varians confirmed case	standared deviation confirmed case
1	169065144	2156	157290931.698175	12541.5681514783
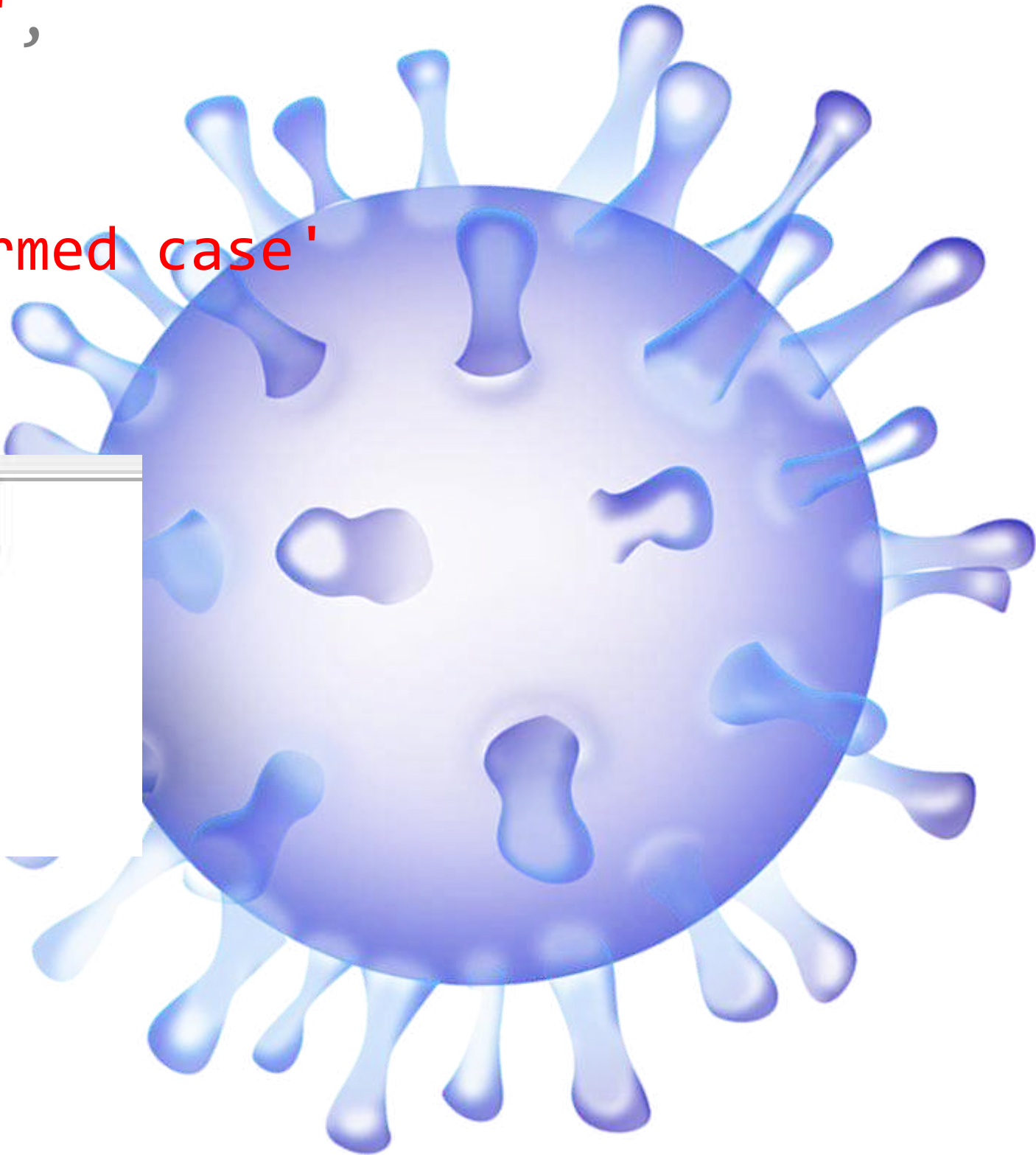
Q12. Check how corona virus spread out with respect to death case per month (Eg.: total confirmed cases, their average, variance & STDEV)

```sql
select sum(Deaths) 'total Deaths case',
avg(Deaths) 'avg Deaths case',
var(Deaths)'varians Deaths case',
stdev(Deaths) 'standared deviation Deaths case'
from [CoronaVirus ]
```

	total Deaths case	avg Deaths case	varians Deaths case	standared deviation Deaths case
1	3647894	46	45892.6043229562	214.22559212885

Q13. Check how corona virus spread out with respect to recovered case
--   (Eg.: total confirmed cases, their average, variance & STDEV)

```sql
select sum(Recovered) 'total Recovered case',
avg(Recovered) 'avg Recovered case',
var(Recovered)'varians Recovered case',
stdev(Recovered) 'standared deviation Recovered case'
from [CoronaVirus ]
```

	total Recovered case	avg Recovered case	varians Recovered case	standared deviation Recovered case
1	113089548	1442	107030888.69603	10345.5733865277
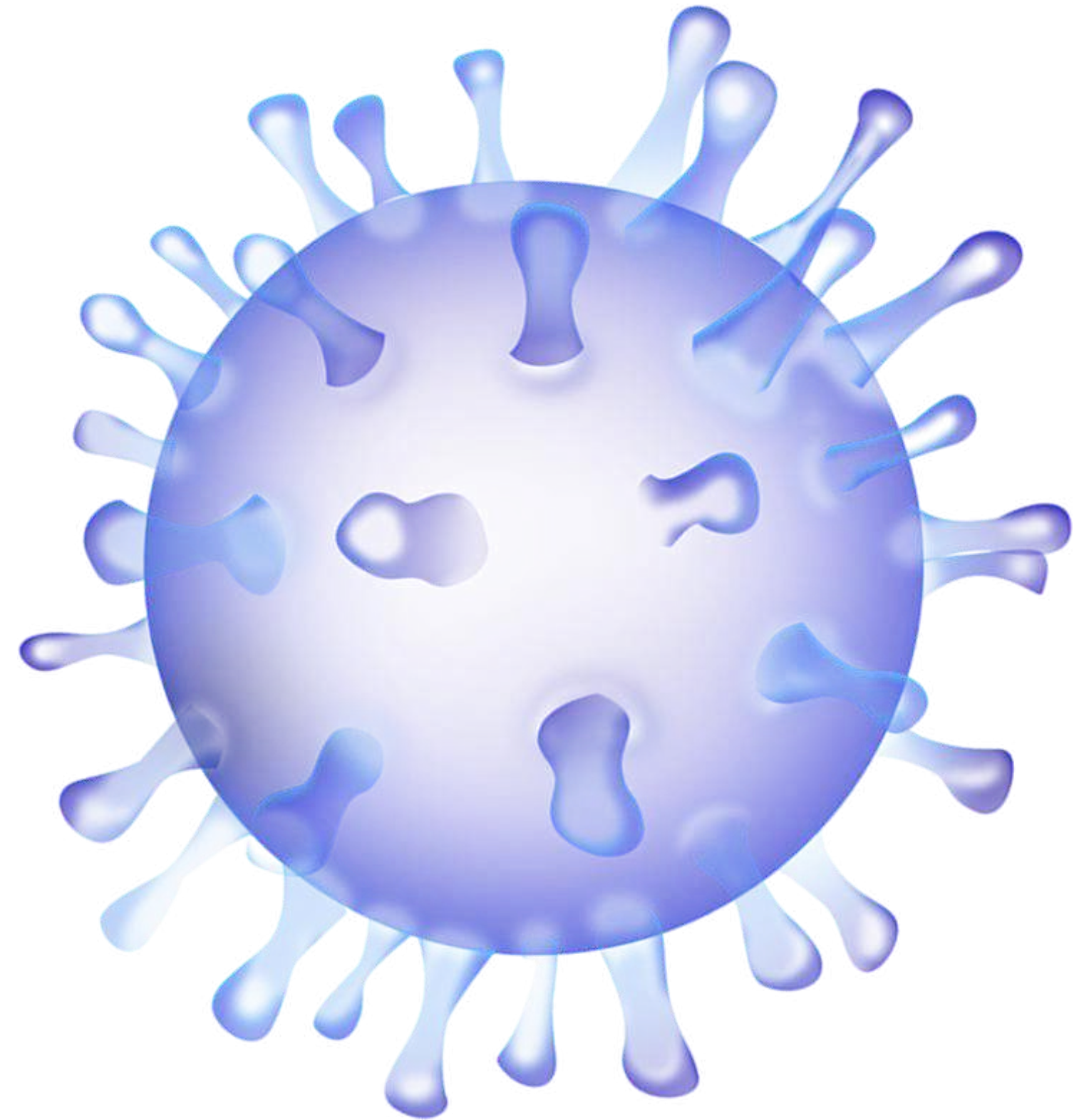
Q14. Find Country having highest number of the Confirmed case

select top 1 Country_Region 'counrty name',sum(Confirmed) 'highest confirmed case'
from [CoronaVirus]
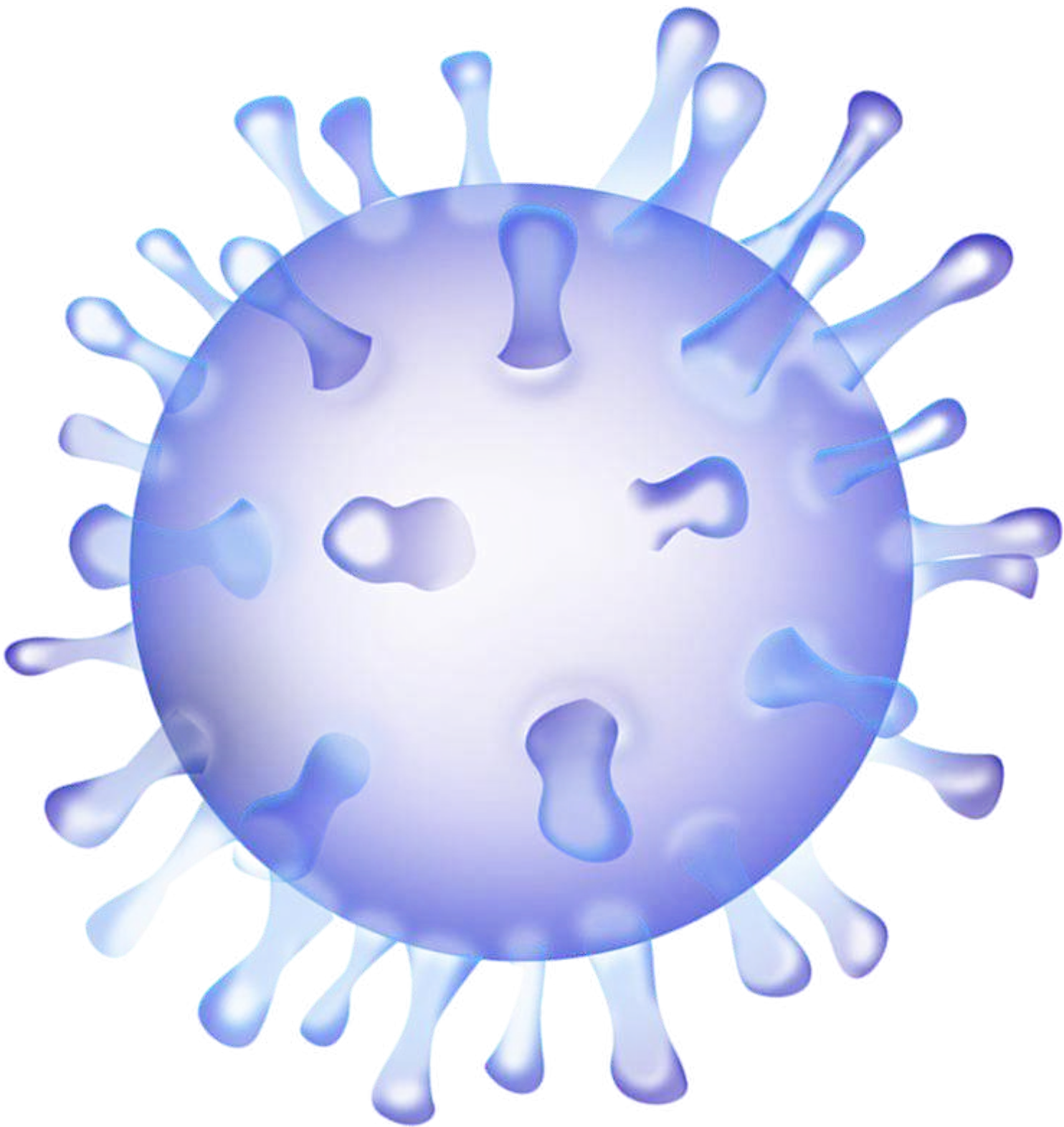group by Country_Region
order by [highest confirmed case] desc

counrty name	highest confirmed case
US	33461982
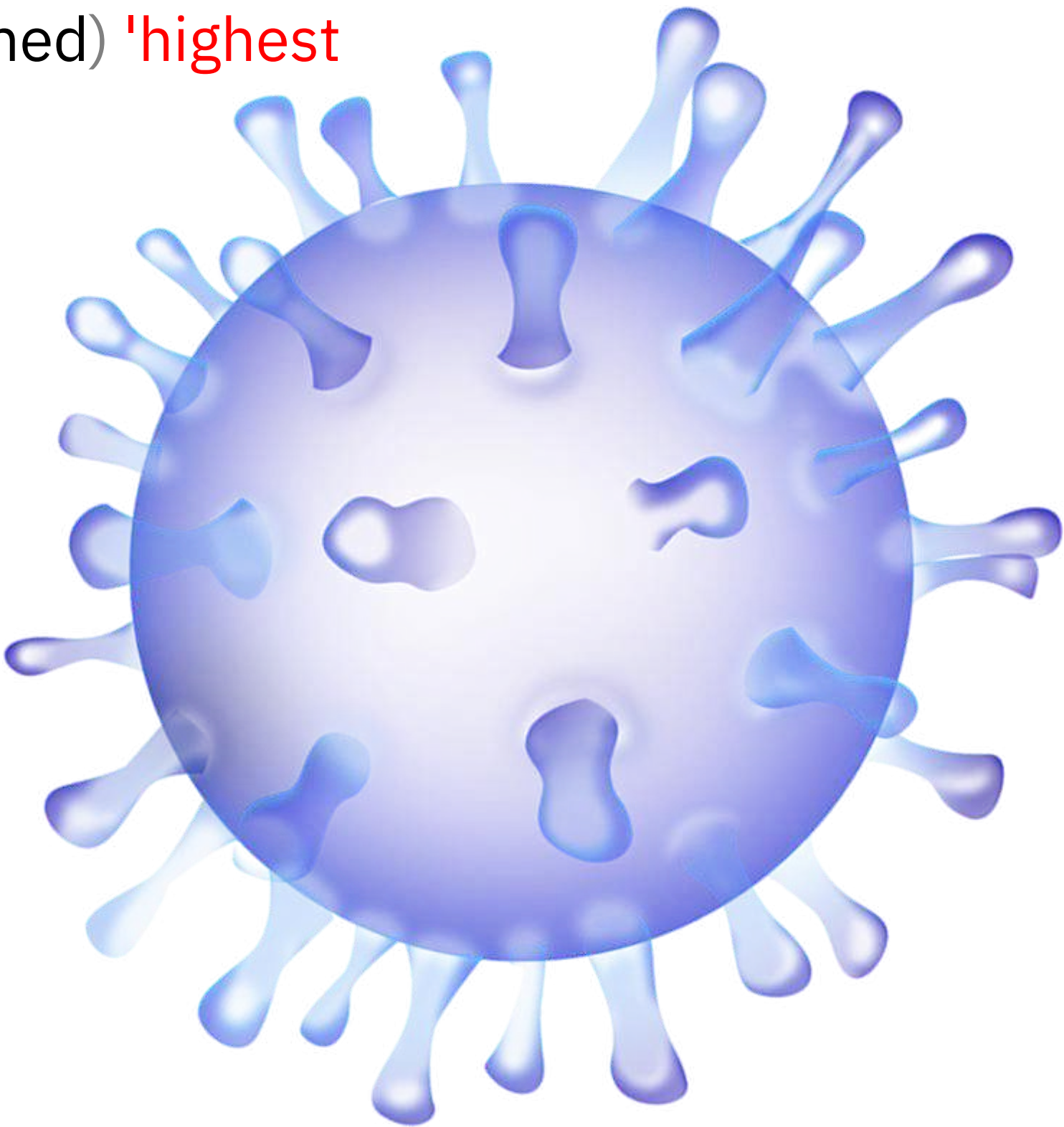
Q15. Find Country having lowest number of the death case

```sql
 select top 4 Country_Region 'counrty name',sum(Deaths)
'lowest deaths case'
from [CoronaVirus ]
group by Country_Region
order by [lowest deaths case] asc
```

|	counrty name	lowest deaths case
1	Marshall Islands	0
2	Samoa	0
3	Kiribati	0
4	Dominica	0

Q16. Find top 5 countries having highest recovered case

select top 5 Country_Region 'counrty name',sum(Recovered) 'highest Recovered case'
from [CoronaVirus]
group by Country_Region
order by [highest Recovered case] desc

|	counrty name	highest Recovered case
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756
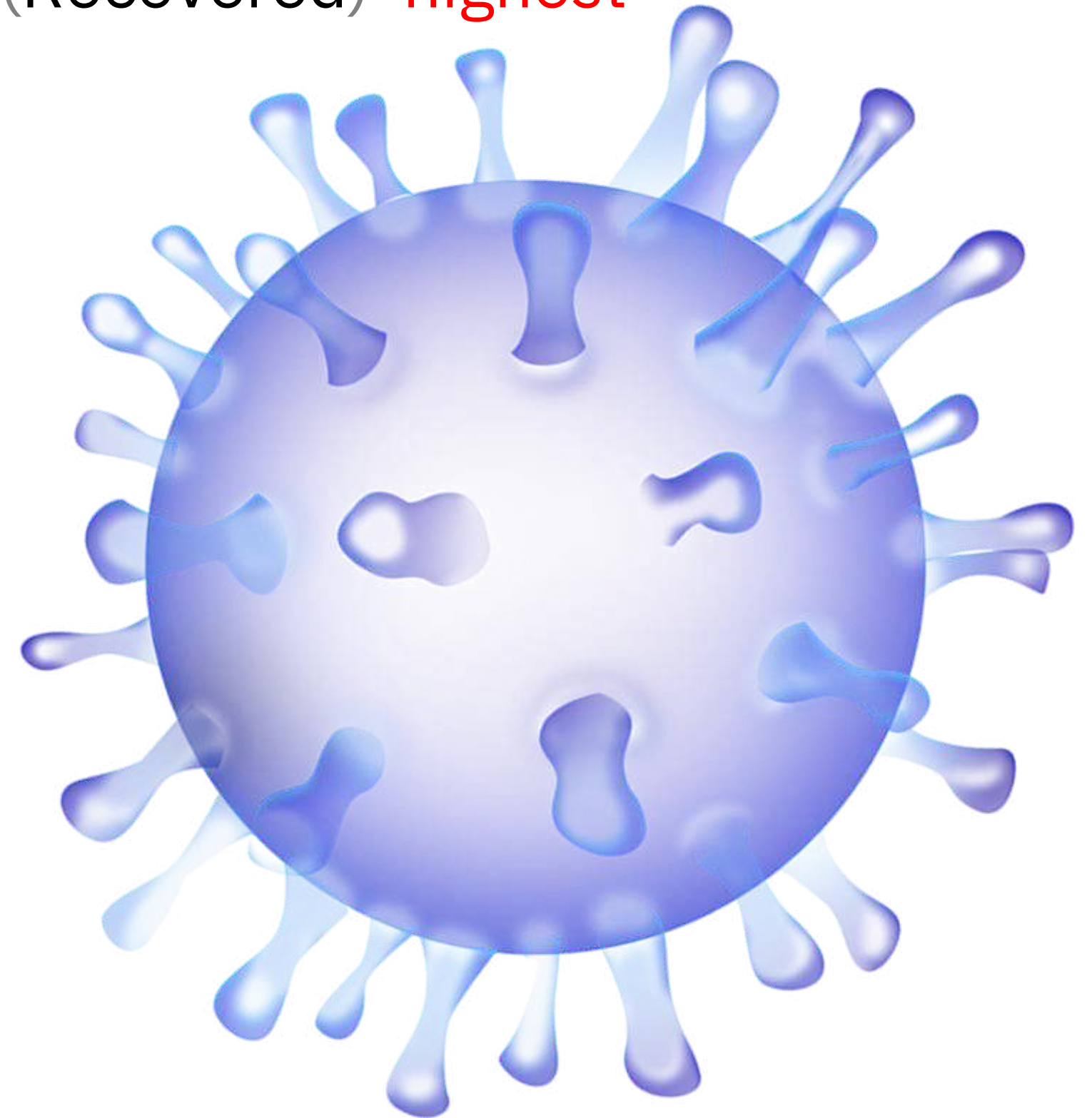
Insight

• The data records COVID-19 cases from January 2020 to May 2021.

• The United States has the highest number of confirmed cases, with 33.5 million people infe
with the virus. India recorded at most 29.6 million confirmed cases, and Brazil had a total of
million confirmed cases.

• The top five countries with the highest recovered cases are India, Brazil, United States, Turk
and Russia.

• The top three highest average confirmed cases occurred in April 2021 with a total of 4,699,
December 2020 with a total of 4,050, and May 2021 with a total of 4,005.

• The top three highest average recovered cases occurred in May 2021 with a total of 4,007,
2021 with a total of 3,074, and June 2021 with a total of 2,769.

• The year 2020 had the maximum number of confirmed cases and recovered cases, wherea
recorded more death cases.

• The highest number of deaths were reported in January 2021.
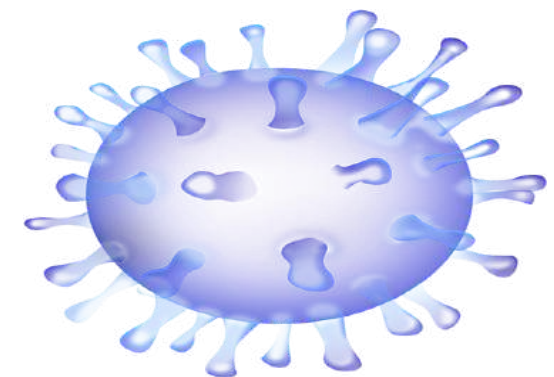
Insight

• The result from the analysis indicates a high variability in the number of confirmed cases and recovered cases, which proves that the spread of the virus is uneven, with some areas or time periods experiencing higher or lower cases compared to others.

• January 2020 recorded the least total death cases and least total confirmed cases of 190 and 6,384, respectively. The low variability indicates significant consistency and stability. This means the death cases and confirmed cases were relatively similar in different regions, proving a more uniform spread of the virus in terms of fatalities during the month. However, the regions witnessed a significant increase and decrease in the spread of the virus from February 2020 to May 2021.

• Marshall Islands, Samoa, Dominica, and Kiribati had the lowest count of death cases with a total of 0 respectively.
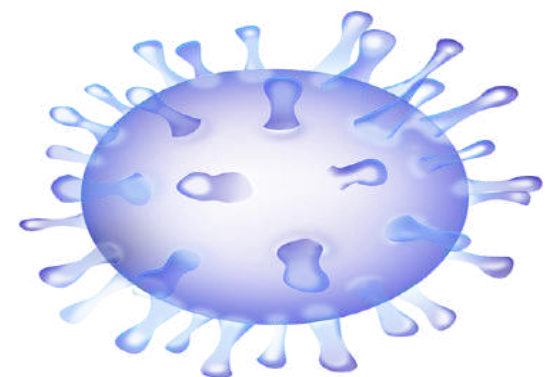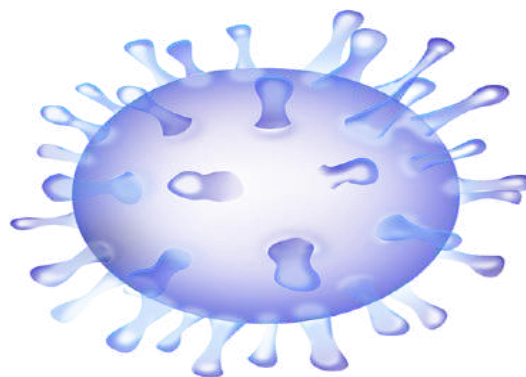
Recommendations
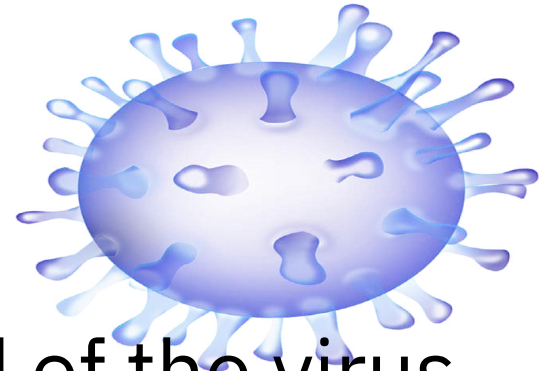
• Management should promote the benefits of vaccination to minimize the spread of the virus.

• Identify and focus on high-risk areas using regulations such as travel restrictions, and quarantine

policies.

• Incorporate the use of predictive modelling to forecast potential outbreaks and take adequate

measures to prevent them.

• Distribute resources such as medical equipment, healthcare workers, and hospital beds to areas

with greater number of cases in order to control and manage the spread of the virus effectively

THANK- YOU