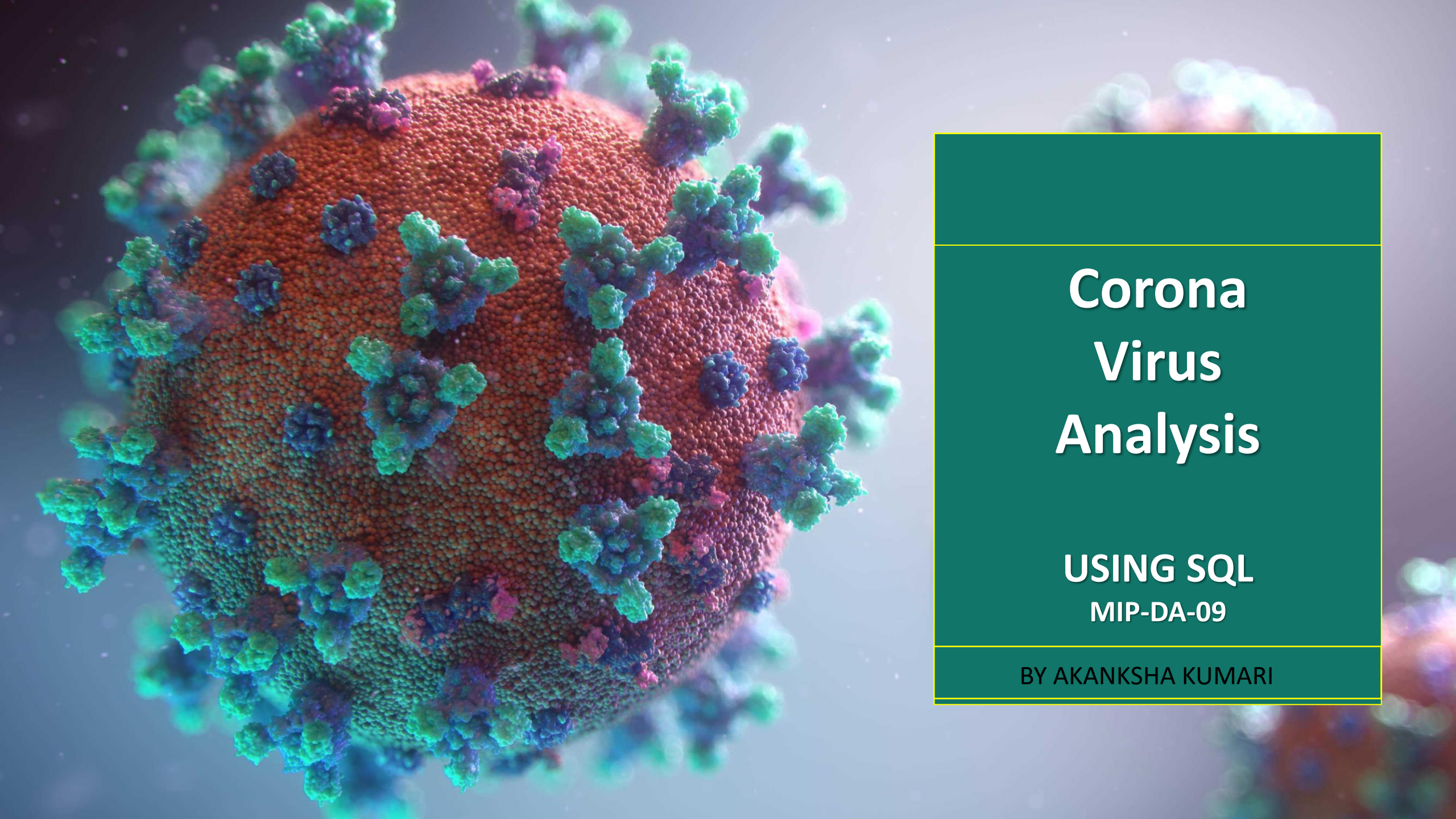
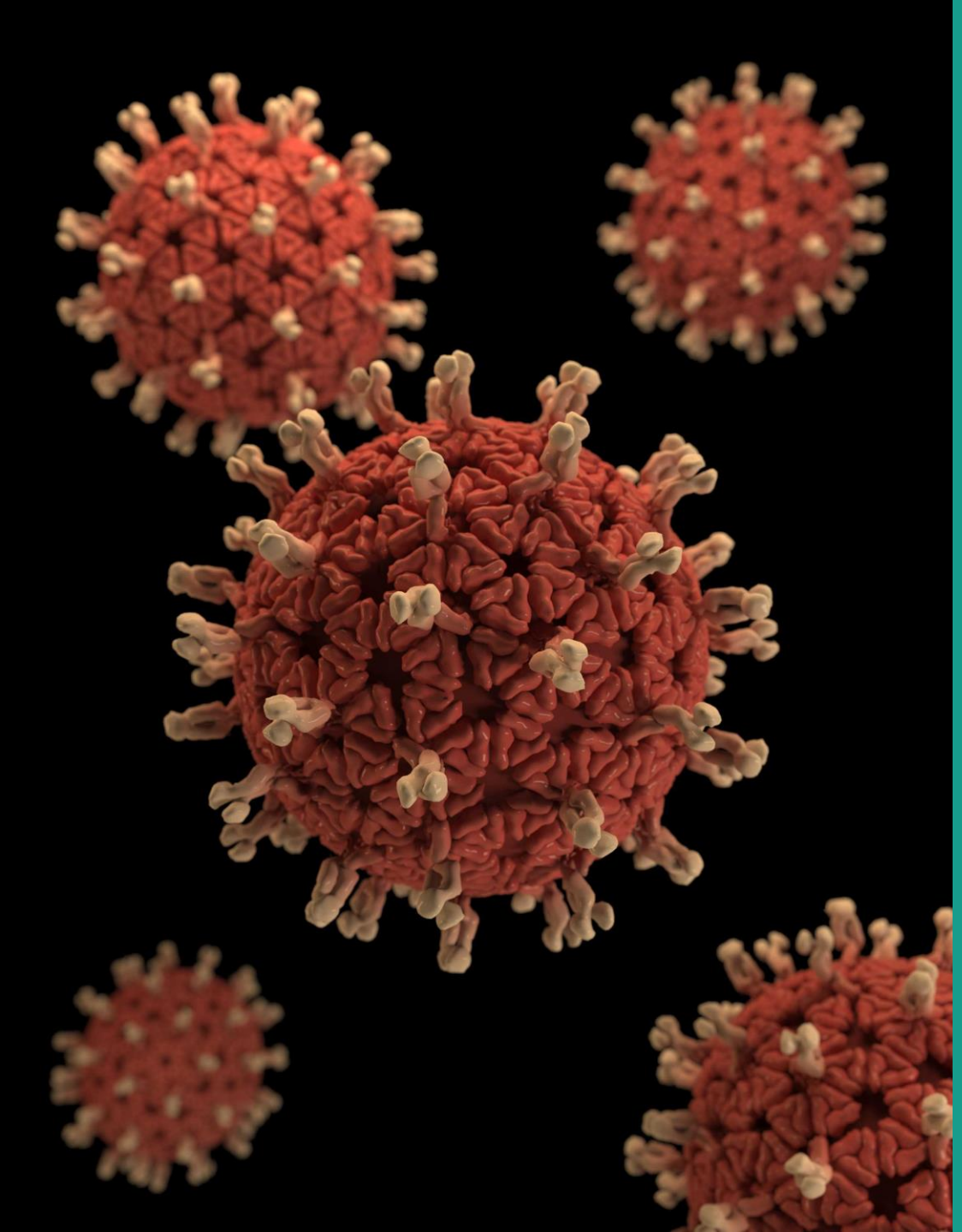


Corona Virus Analysis

USING SQL

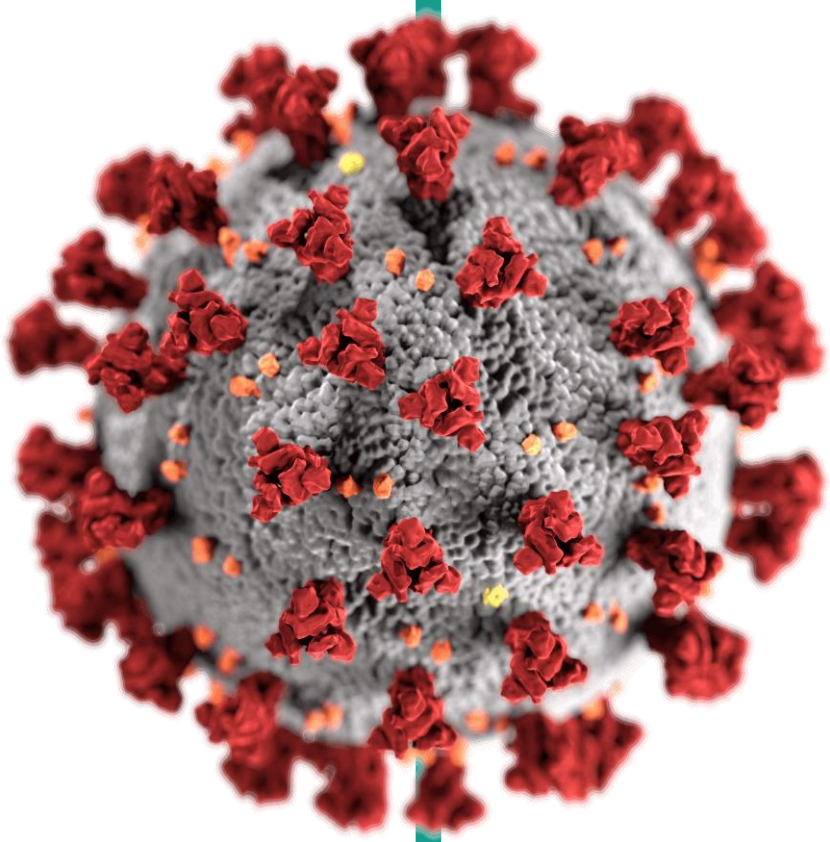
MIP-DA-09

BY AKANKSHA KUMARI



OVERVIEW

The COVID-19 pandemic has brought unprecedented challenges worldwide, affecting every aspect of society, from healthcare systems to economies. Analyzing the vast amount of data generated during this crisis is crucial for understanding its impact, identifying patterns, and informing decision-making processes. In this project, we propose to leverage SQL (Structured Query Language) to analyze COVID-19 data and gain insights into various aspects of the pandemic.



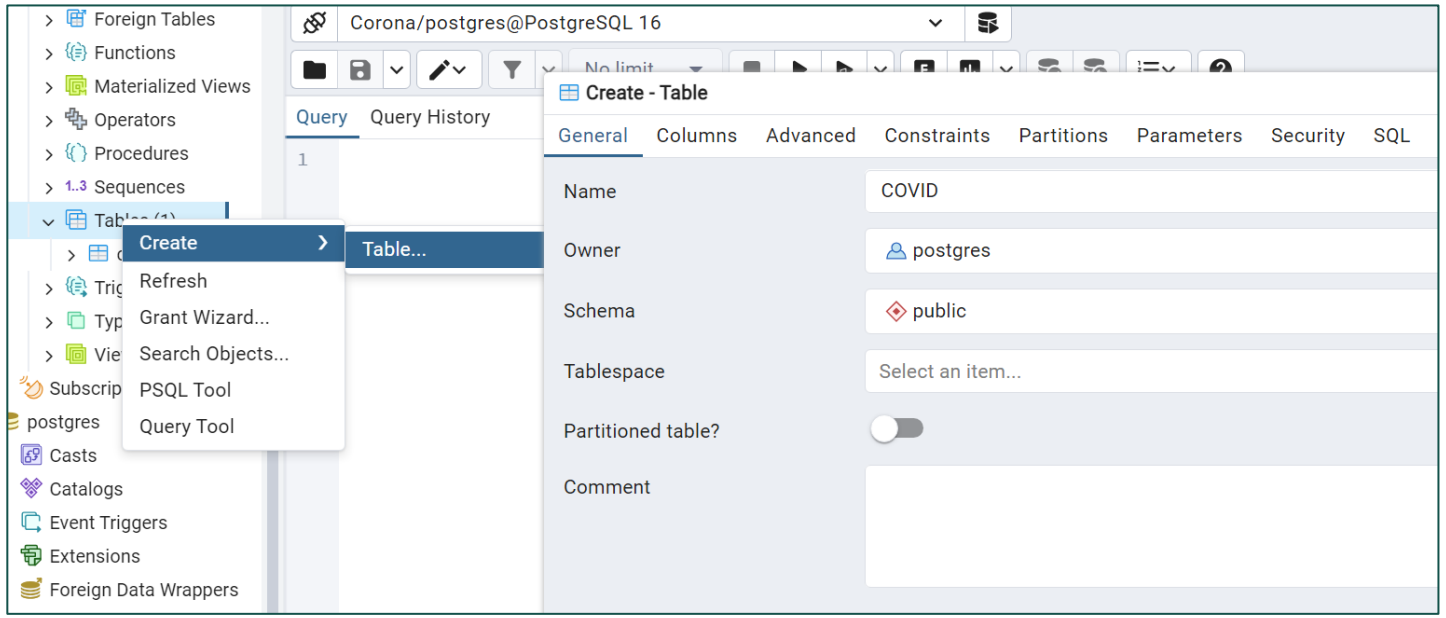
DATASET

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.

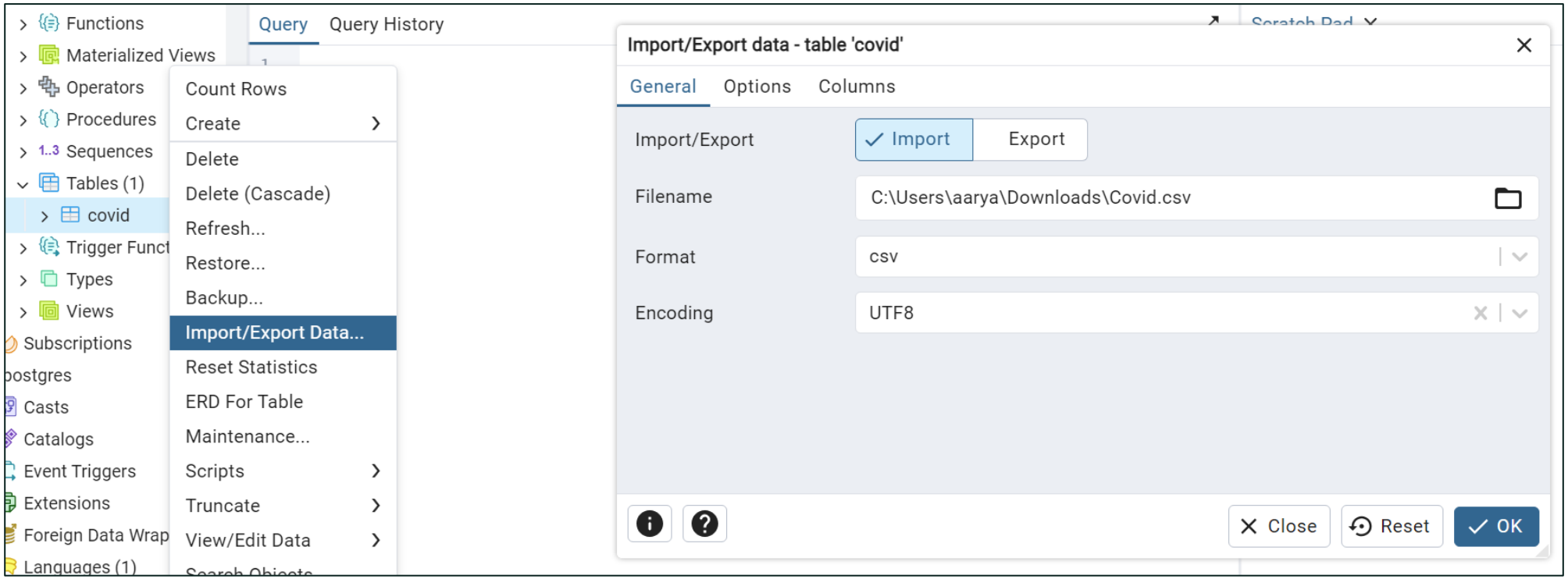
CREATE DB

```
1  CREATE DATABASE CORONA;
2
3  USE CORONA;
4
5
```

CREATE TABLE



IMPORT DATASET TO SQL SERVER



Q 1. Write a query to check for null values.

```
Select * from covid
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
```

Data Output | Messages | Notifications

SQL

province character varying (50)	country_region character varying (50)	latitude double precision	longitude double precision	date date	confirmed integer	deaths integer	recovered integer

Q 2. If Null values are present ,update them with zeros for all columns.

```
1   Update covid
2  set province = coalesce (province , 'NOT AVAILABLE'),
3  country_region  = coalesce (country_region , 'NOT AVAILABLE'),
4  latitude = coalesce (latitude,0.0),
5  longitude = coalesce(longitude ,0.0),
6  date = coalesce (date,'1970-01-01'::date),
7  confirmed = coalesce (confirmed,0),
8  deaths = coalesce (deaths,0),
9  recovered = coalesce(recovered,0);
10
```

Data Output | Messages | Notifications

```
UPDATE 78386

Query returned successfully in 472 msec.
```

Q 3. Check Total number of rows.

Query | Query History

```
1  SELECT count(*) as total_rows
2  from covid;
3
```

Data Output | Messages | Notifications

	total_rows bigint
1	78386

Q 4. Check what is start_date and what is the end_date.

Query | Query History

```
1  SELECT MIN(date) AS START_DATE , MAX(date) AS END_DATE
2  from covid;
3
```

Data Output | Messages | Notifications

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

Q 5. Number of month present in Dataset.

Query | Query History

```
1 SELECT extract(month from date) as month_number , count(*) as month_count
2 from covid
3 group by month_number
4 order by month_count;
```

Data Output | Messages | Notifications

	month_number numeric	month_count bigint
1	11	4620
2	9	4620
3	10	4774
4	7	4774
5	8	4774
6	12	4774
7	1	6314
8	6	6622
9	2	8778
10	4	9240
11	5	9548

Total rows: 12 of 12 | Query complete 00:00:00.135

Q 6. Monthly average for confirmed ,deaths and recovered cases.

INPUT

Query | Query History

```
1  SELECT
2      extract(year from date) as year,
3      extract(month from date) as month,
4      round(avg(confirmed),2) as confirmed_avg,
5      round(avg(deaths),2) as deaths_avg,
6      round(avg(recovered),2) as recovered_avg
7  from covid
8  group by year , month
9  order by year , month asc;
10
```

OUTPUT

Data Output | Messages | Notifications

	year numeric	month numeric	confirmed_avg numeric	deaths_avg numeric	recovered_avg numeric
1	2020	1	4.15	0.12	0.09
2	2020	2	15.30	0.59	7.03
3	2020	3	161.13	8.66	27.87
4	2020	4	505.80	41.52	171.64
5	2020	5	574.85	30.28	318.30
6	2020	6	859.23	29.82	548.79
7	2020	7	1432.36	35.11	983.06
8	2020	8	1611.84	37.54	1299.29
9	2020	9	1784.59	34.78	1438.91
10	2020	10	2412.20	36.76	1420.64
11	2020	11	3592.19	56.76	1985.34
12	2020	12	4050.44	71.22	2497.89
13	2021	1	3911.23	84.18	1919.64
14	2021	2	2433.36	69.16	1558.39

Total rows: 18 of 18 | Query complete 00:00:00.178

Q 7. Find most frequent value for confirmed , deaths , recovered each month.

INPUT

```
SELECT
    extract(year from date) as year,
    extract(month from date) as month,
    max(confirmed) as most_confirmed_value,
    max(deaths) as most_deaths_value,
    max(recovered) as most_recovered_value
from covid
group by year,month
order by year,month
```

OUTPUT

Data Output | Messages | Notifications

	year numeric	month numeric	most_confirmed_value integer	most_deaths_value integer	most_recovered_value integer
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

Total rows: 18 of 18 | Query complete 00:00:00.117

Q 8. Find minimum values for confirmed , deaths , recovered per year.

Query | Query History

```
SELECT
    extract(year from date) as year,
    min(confirmed) as min_confirmed_value,
    min(deaths) as min_deaths_value,
    min(recovered) as min_recovered_value
from covid
group by year
order by year
```

Data Output | Messages | Notifications

	year numeric	min_confirmed_value integer	min_deaths_value integer	min_recovered_value integer
1	2020	0	0	0
2	2021	0	0	0

Q 9. Find maximum values for confirmed , deaths , recovered per year.

Query | Query History

```
1  SELECT
2      extract(year from date) as year,
3      max(confirmed) as max_confirmed_value,
4      max(deaths) as max_deaths_value,
5      max(recovered) as max_recovered_value
6  from covid
7  group by year
8  order by year
```

Data Output | Messages | Notifications

	year numeric	max_confirmed_value integer	max_deaths_value integer	max_recovered_value integer
1	2020	823225	3752	1123456
2	2021	414188	7374	422436

Q 10. Find maximum values for confirmed , deaths , recovered per year.

INPUT

Query | Query History

```
1  SELECT
2      extract(year from date) as year,
3      extract(month from date) as month,
4      sum(confirmed) as total_confirmed_value,
5      sum(deaths) as total_deaths_value,
6      sum(recovered) as total_recovered_value
7  from covid
8  group by year,month
9  order by year ,month
```

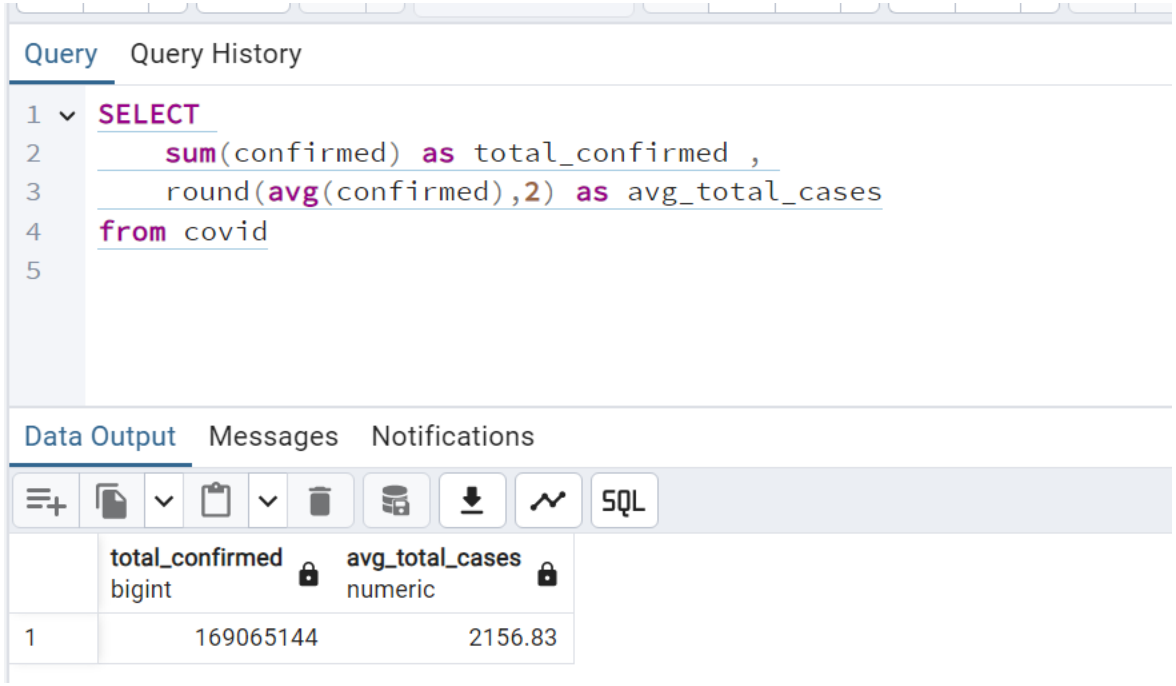
OUTPUT

Data Output | Messages | Notifications

	year numeric	month numeric	total_confirmed_value bigint	total_deaths_value bigint	total_recovered_value bigint
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

Total rows: 18 of 18 | Query complete 00:00:00.121

Q 11. Check how corona virus spread out with respect to confirmed cases.



Q 12. Check how corona virus spread out with respect to death cases per month.

INPUT

Query | Query History

```
1  SELECT
2      extract(year from date) as year,
3      extract(month from date) as month,
4      sum(deaths) as total_death,
5      round(avg(deaths),2) as avg_death_cases
6  from covid
7  group by year,month
8  order by year,month
9
```

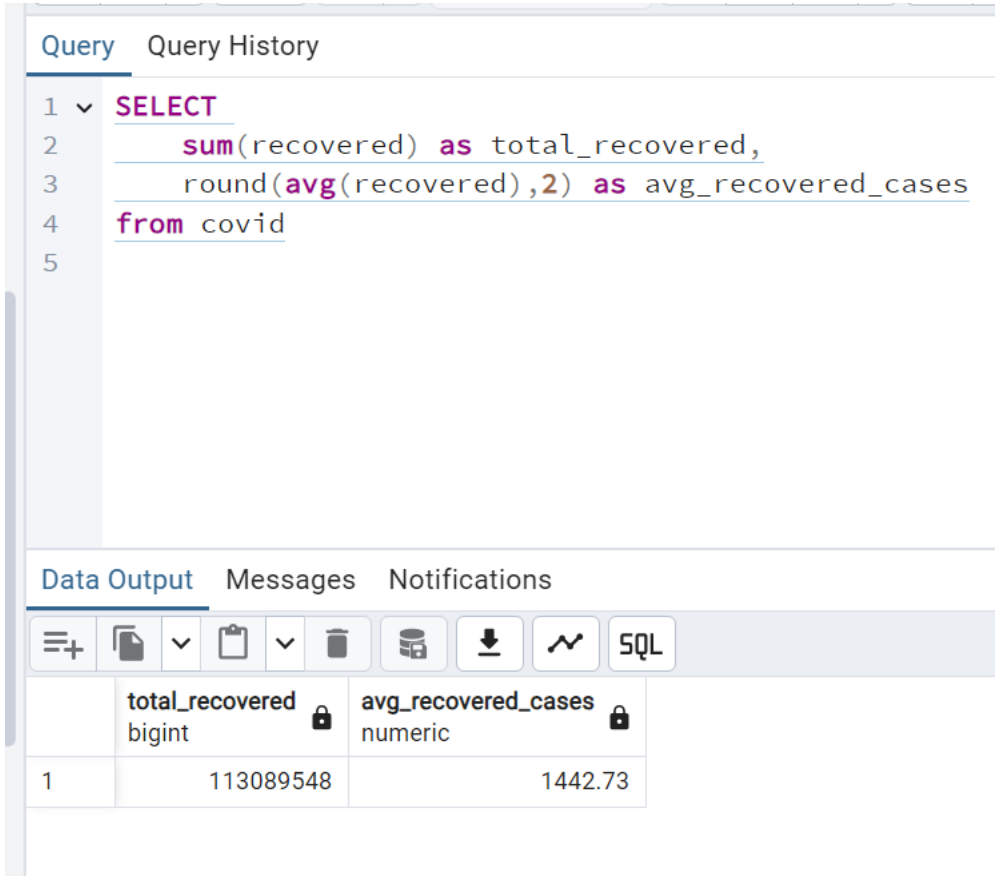
OUTPUT

Data Output | Messages | Notifications

	year numeric	month numeric	total_death bigint	avg_death_cases numeric
1	2020	1	190	0.12
2	2020	2	2651	0.59
3	2020	3	41346	8.66
4	2020	4	191833	41.52
5	2020	5	144561	30.28
6	2020	6	137757	29.82
7	2020	7	167613	35.11
8	2020	8	179200	37.54
9	2020	9	160671	34.78
10	2020	10	175484	36.76
11	2020	11	262247	56.76
12	2020	12	339996	71.22
13	2021	1	401893	84.18

Total rows: 18 of 18 Query complete 00:00:00.124

Q 13. Check how corona virus spread out with respect to recovered cases.



Q 14. Find the country have highest number of confirmed cases.

Query | Query History

```
SELECT country_region,
    sum(confirmed) as total
from covid
group by country_region
order by total desc
limit 1;
```

Data Output | Messages | Notifications

	country_region character varying (50)	total bigint
1	US	33461982

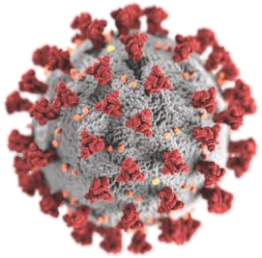
Q 15. Find top 5 countries have highest recovered cases.

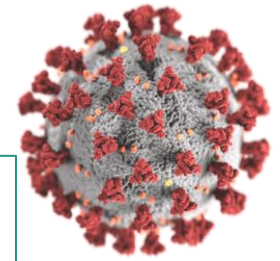
Query | Query History

```
1  SELECT country_region,
2      sum(recovered) as total_recovered
3  from covid
4  group by country_region
5  order by total_recovered desc
6  limit 5;
```

Data Output | Messages | Notifications

	country_region character varying (50)	total_recovered bigint
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756



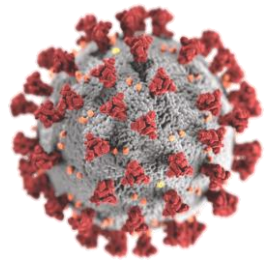


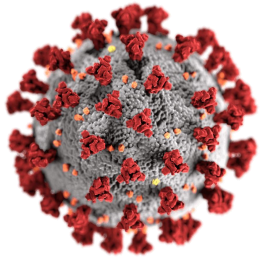
INSIGHTS

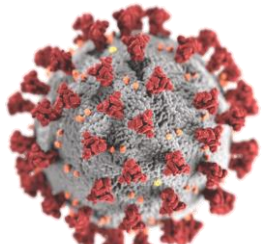
1. COVID-19 breakout on 22 January 2020 till 13 June 2021.
2. India has the highest recovery cases.
3. Most confirmed cases were in US.
4. Top regions having lowest number of deaths.

. Samoa
.Kiribati
.Dominica
.The Marshall Islands

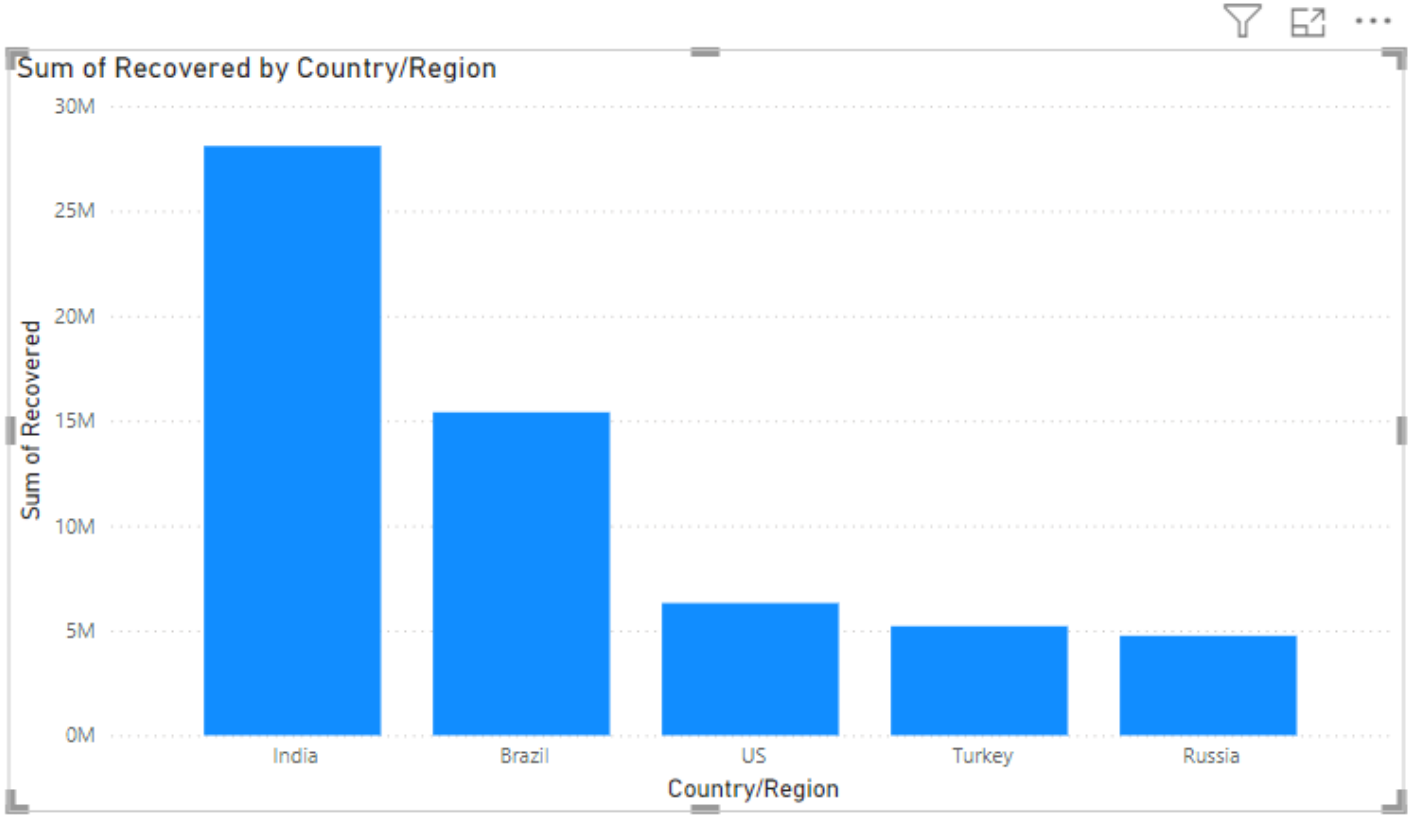
5. Maximum deaths were in 2021 .







TOP 5 COUNTRY RECOVERED VALUES



Thank You!