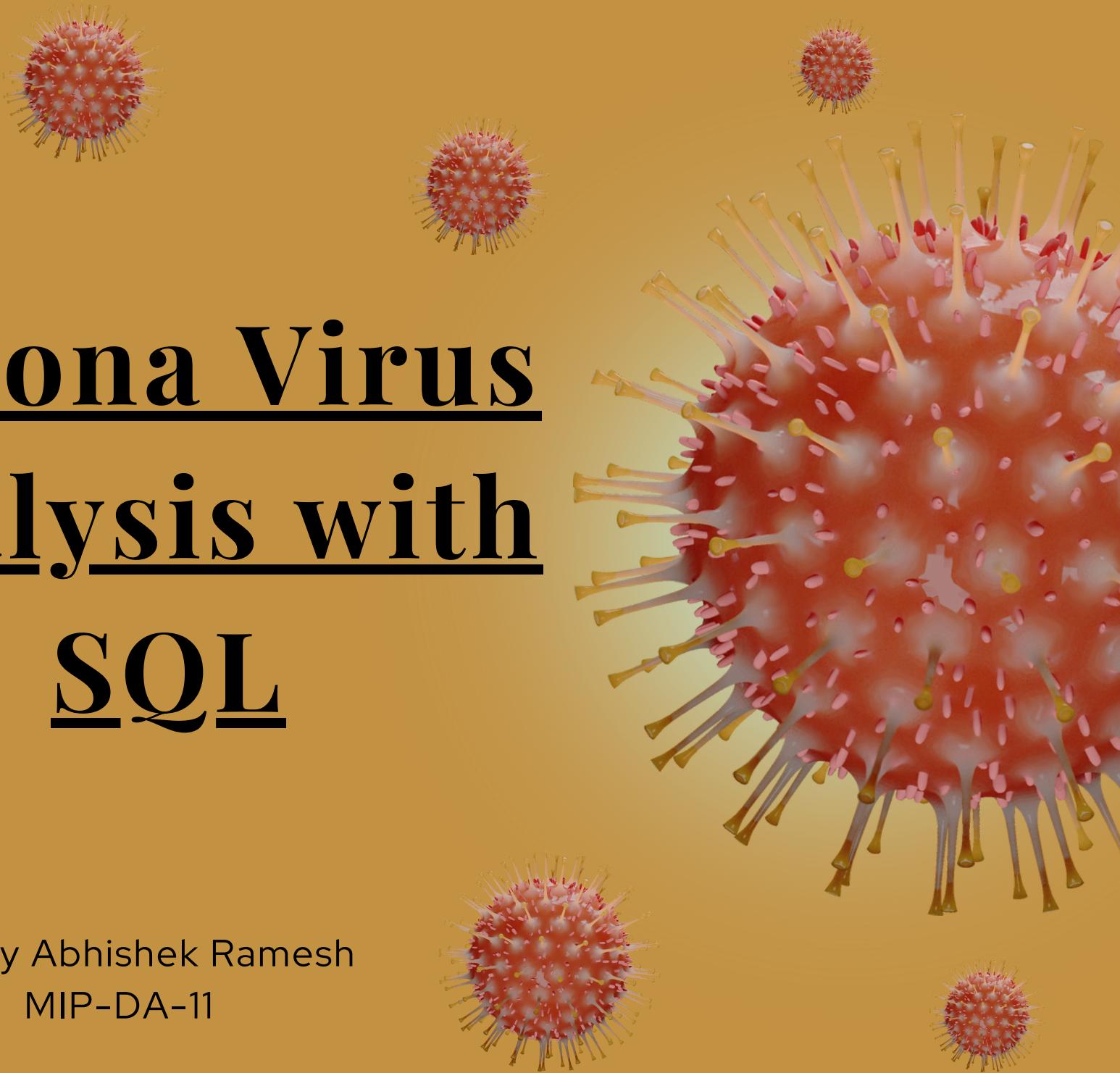


Corona Virus

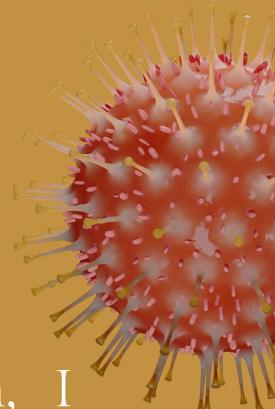
Analysis with

SQL

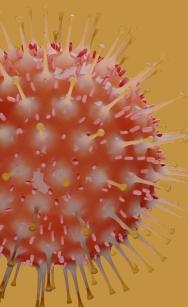
Pandey Abhishek Ramesh
MIP-DA-11



Introduction

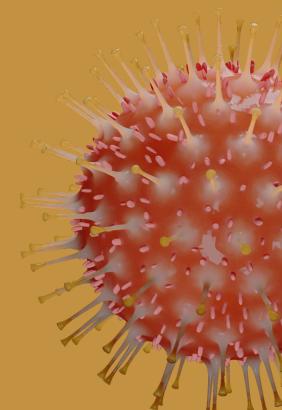
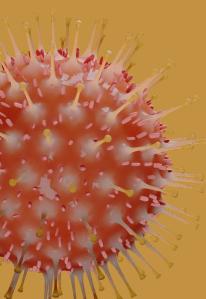


As a Data Analyst Intern at Mentorness Organization, I undertook an in-depth analysis of the COVID-19 pandemic's impact. Our work, spanning from January 22, 2020, to June 13, 2021, involved a vast dataset of 78,386 entries, capturing the pandemic's effects across different provinces and countries worldwide.



By applying rigorous SQL techniques, we explored this extensive data to uncover detailed trends and insights into the pandemic's spread and management. This presentation showcases our collective efforts and commitment to precise, data-driven solutions. Join us as we reveal the findings from our thorough analysis of COVID-19 data.

Columns



```
1 • CREATE DATABASE IF NOT EXISTS CORONA;
2 • USE CORONA;
3 • CREATE TABLE CORONAVIRUS
4 • (
5     Province VARCHAR(255),
6     Country VARCHAR(255),
7     Latitude INT,
8     Longitude INT,
9     Date_ DATE,
10    Confirmed INT,
11    Deaths INT,
12    Recovered INT
13 )
```

1. Check NULL values



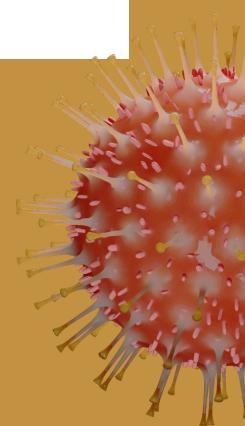
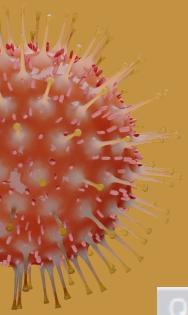
Query 1

```
16
17      -- Q1. Write a code to check NULL values
18      SELECT
19          SUM(CASE WHEN confirmed IS NULL THEN 1 ELSE 0 END) AS null_confirmed,
20          SUM(CASE WHEN deaths IS NULL THEN 1 ELSE 0 END) AS null_deaths,
21          SUM(CASE WHEN recovered IS NULL THEN 1 ELSE 0 END) AS null_recovered
22      FROM CORONAVIRUS;
23      |
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

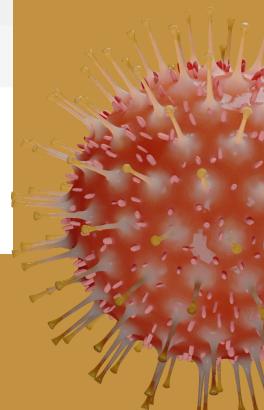
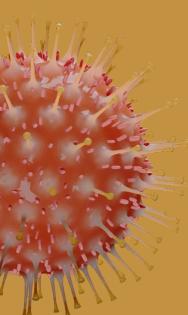
	null_confirmed	null_deaths	null_recovered
▶	0	0	0

2. Update the present NULL values with zeroes



```
Query 1 ×
Limit to 5000 rows
1 -- Q2. If NULL values are present, update them with zeros for all columns.
2 • UPDATE CORONAVIRUS
3 SET
4     confirmed = COALESCE(confirmed, 0),
5     deaths = COALESCE(deaths, 0),
6     recovered = COALESCE(recovered, 0);
7
```

3. Total number of rows



```
Query 1 x
1 -- Q3. check total number of rows
2 • SELECT COUNT(*) AS total_rows
3 FROM CORONAVIRUS;
4
5
```

Result Grid | Filter Rows: _____ | Export: | Wrap Cell Content: |

total_rows
78386

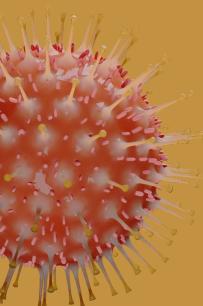
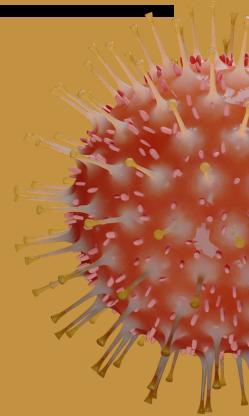
4. Start_date and End_date

Query 1 x

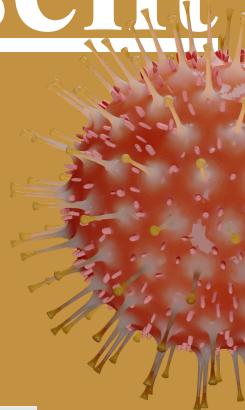
1 -- Q4. Check what is start_date and end_date
2 • **SELECT**
3 MIN(date_) **AS** start_date,
4 MAX(date_) **AS** end_date
5 **FROM** CORONAVIRUS;
6

Result Grid | Filter Rows: Export: Wrap Cell Content:

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



5. Number of month present in dataset



Query 1 ×

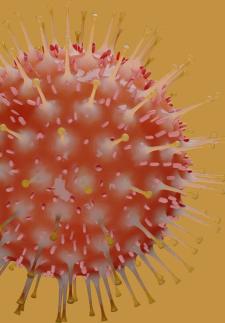
1 -- Q5. Number of month present in dataset
2 • **SELECT COUNT(DISTINCT DATE_FORMAT(date_, '%Y-%m')) AS num_months**
3 FROM CORONAVIRUS;
4
5
6
7

Result Grid | Filter Rows: _____ | Export: | Wrap Cell Content:

num_months
18

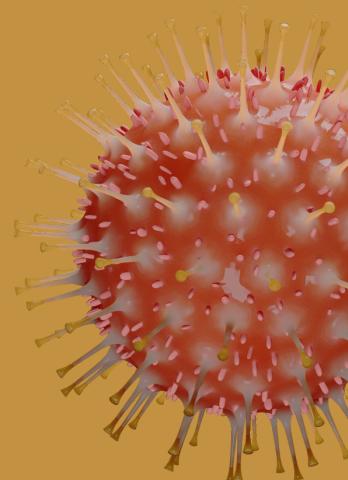


6. Monthly average for Confirmed, Deaths, Recovered

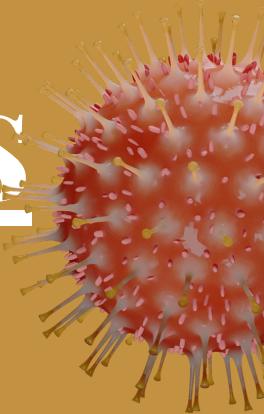


```
1 -- Q6. Find monthly average for confirmed, deaths, recovered
2 • SELECT
3     DATE_FORMAT(date_, '%Y-%m') AS month,
4     AVG(confirmed) AS avg_confirmed,
5     AVG(deaths) AS avg_deaths,
6     AVG(recovered) AS avg_recovered
7 FROM CORONAVIRUS
8 GROUP BY DATE_FORMAT(date_, '%Y-%m');
9
10
```

month	avg_confirmed	avg_deaths	avg_recovered
2020-01	4.1455	0.1234	0.0929
2020-02	15.2960	0.5936	7.0320
2020-03	161.1303	8.6607	27.8739
2020-04	505.8004	41.5223	171.6422
2020-05	574.8498	30.2809	318.2964
2020-06	859.2281	29.8175	548.7916
2020-07	1432.3611	35.1096	983.0582
2020-08	1611.8429	37.5367	1299.2947
2020-09	1784.5874	34.7773	1438.9067
2020-10	2412.1996	36.7583	1420.6431
2020-11	3592.1944	56.7634	1985.3446
2020-12	4050.4397	71.2183	2497.8850
2021-01	3911.2285	84.1837	1919.6370
2021-02	2433.3636	69.1649	1558.3917
2021-03	2916.7972	59.1998	1652.2859
2021-04	4699.3552	78.4387	3074.7851
2021-05	4005.2541	76.7803	4007.5078



7. Minimum Values

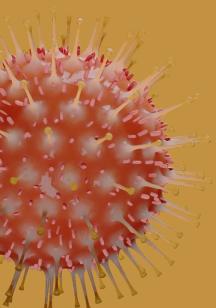


Query 1

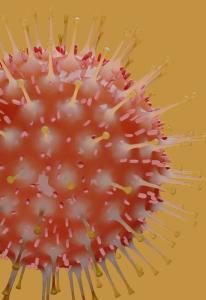
```
1 -- Q7. Find minimum values for confirmed, deaths, recovered per year
2 • SELECT
3     YEAR(date_) AS year,
4     MIN(confirmed) AS min_confirmed,
5     MIN(deaths) AS min_deaths,
6     MIN(recovered) AS min_recovered
7 FROM CORONAVIRUS
8 GROUP BY YEAR(date_);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

year	min_confirmed	min_deaths	min_recovered
2020	0	0	0
2021	0	0	0



8. Maximum Values



```
Query 1 ×
 1   -- Q8. Find maximum values of confirmed, deaths, recovered per year
 2 • SELECT
 3     YEAR(date_) AS year,
 4     MAX(confirmed) AS max_confirmed,
 5     MAX(deaths) AS max_deaths,
 6     MAX(recovered) AS max_recovered
 7   FROM CORONAVIRUS
 8   GROUP BY YEAR(date_);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	year	max_confirmed	max_deaths	max_recovered
▶	2020	823225	3752	1123456
	2021	414188	7374	422436

9. Total number of case of confirmed, death, recovered each month

Query 1 ×

```
1 -- Q9. The total number of case of confirmed, deaths, recovered each month
2 • SELECT
3     DATE_FORMAT(date_, '%Y-%m') AS month,
4     SUM(confirmed) AS total_confirmed,
5     SUM(deaths) AS total_deaths,
6     SUM(recovered) AS total_recovered
7 FROM CORONAVIRUS
8 GROUP BY DATE_FORMAT(date_, '%Y-%m');
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

month	total_confirmed	total_deaths	total_recovered
2020-01	6384	190	143
2020-02	68312	2651	31405
2020-03	769236	41346	133070
2020-04	2336798	191833	792987
2020-05	2744333	144561	1519547
2020-06	3969634	137757	2535417
2020-07	6838092	167613	4693120
2020-08	7694938	179200	6202833
2020-09	8244794	160671	6647749
2020-10	11515841	175484	6782150
2020-11	16595938	262247	9172292
2020-12	19336799	339996	11924903
2021-01	18672205	401893	9164347
2021-02	10492664	298239	6719785
2021-03	13924790	282620	7888013
2021-04	21711021	362387	14205507
2021-05	19121083	366549	19131842
2021-06	5022282	132657	5544438

10. Corona virus spread out with respect to confirmed case



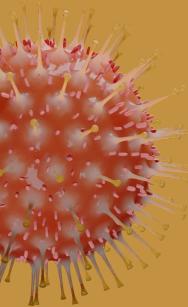
Query 1 x

1 -- Q10. Check how corona virus spread out with respect to confirmed case
2 -- (Eg.: total confirmed cases, their average, variance & STDEV)
3 • SELECT
4 SUM(confirmed) AS total_confirmed,
5 AVG(confirmed) AS average_confirmed,
6 VARIANCE(confirmed) AS variance_confirmed,
7 STDDEV(confirmed) AS stddev_confirmed
8 FROM CORONAVIRUS;

Result Grid | Filter Rows: Export: Wrap Cell Content:

	total_confirmed	average_confirmed	variance_confirmed	stddev_confirmed
▶	169065144	2156.8283	157288925.07796532	12541.488152446875

11. Corona virus spread out with respect to death case per month



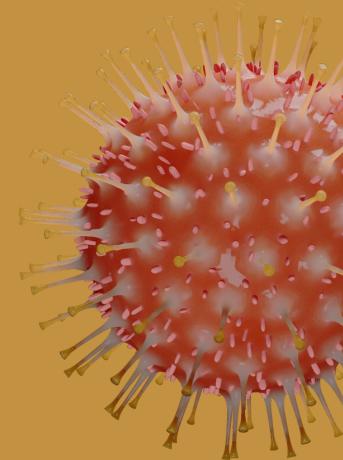
Query 1

```
1  -- Q11. Check how corona virus spread out with respect to death case per month
2  --          (Eg.: total confirmed cases, their average, variance & STDEV )
3  • SELECT
4      DATE_FORMAT(date_, '%Y-%m') AS month,
5      SUM(deaths) AS total_deaths,
6      AVG(deaths) AS average_deaths,
7      VARIANCE(deaths) AS variance_deaths,
8      STDDEV(deaths) AS stddev_deaths
9  FROM CORONAVIRUS
10 GROUP BY DATE_FORMAT(date_, '%Y-%m');
```

Result Grid

month	total_deaths	average_deaths	variance_deaths	stddev_deaths
2020-01	190	0.1234	4.24581716984138	2.0605380777464752
2020-02	2651	0.5936	68.3218488238449	8.265703165723101
2020-03	41346	8.6607	3900.7922648320746	62.45632285711411
2020-04	191833	41.5223	40504.26811767955	201.25672191924312
2020-05	144561	30.2809	20684.911671085658	143.8225005730524
2020-06	137757	29.8175	16929.445709928572	130.11320344195886
2020-07	167613	35.1096	21140.154944373826	145.39654378414167
2020-08	179200	37.5367	23272.99645685882	152.55489653517785
2020-09	160671	34.7773	20102.7692237308	141.78423475030925
2020-10	175484	36.7583	17580.07101972725	132.589860169348
2020-11	262247	56.7634	27773.793596962234	166.6547136955995
2020-12	339996	71.2183	65345.36920134891	255.6274030720277
2021-01	401893	84.1837	102758.43231925515	320.55956126631935
2021-02	298239	69.1649	68478.87146663864	261.6846794648832
2021-03	282620	59.1998	54385.969702527414	233.20799665218905
2021-04	362387	78.4387	94611.47092309907	307.58977701331213
2021-05	366549	76.7803	131769.4693132085	363.00064643635073
2021-06	132657	66.2622	112963.67298959807	336.1006887669201

Filter Rows: Export: Wrap Cell Content:



12. Corona virus spread out with respect to recovered case



Query 1 x

File Edit View Insert Cell Help

Limit to 5000 rows

1 -- Q12. Check how corona virus spread out with respect to recovered case
2 -- (Eg.: total confirmed cases, their average, variance & STDEV)
3 • SELECT
4 SUM(recovered) AS total_recovered,
5 AVG(recovered) AS average_recovered,
6 VARIANCE(recovered) AS variance_recovered,
7 STDDEV(recovered) AS stddev_recovered
8 FROM CORONAVIRUS;

Result Grid | Filter Rows: _____ | Export: | Wrap Cell Content: AA

	total_recovered	average_recovered	variance_recovered	stddev_recovered
▶	113089548	1442.7264	107029523.26229636	10345.507395110999



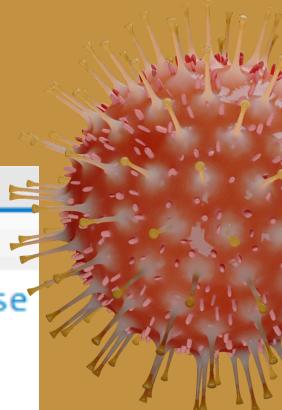
13. Country having highest number of the Confirmed case

Query 1 x

1 -- Q13. Find Country having highest number of the Confirmed case
2 • SELECT
3 country,
4 SUM(confirmed) AS total_confirmed
5 FROM CORONAVIRUS
6 GROUP BY country
7 ORDER BY total_confirmed DESC
8 LIMIT 1;

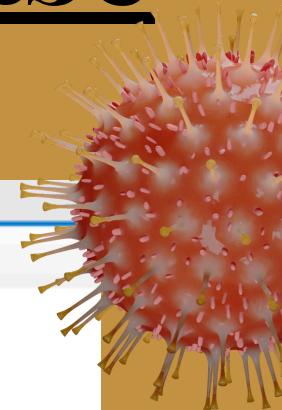
Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

	country	total_confirmed
▶	US	33461982



14. Country having lowest number of the Death case

Query 1 ×



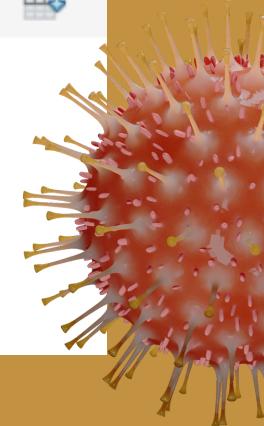
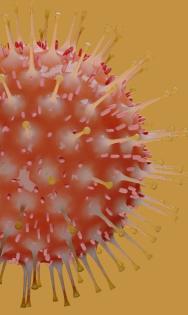
```
1 -- Q14. Find Country having lowest number of the death case
2 • SELECT
3     country,
4     SUM(deaths) AS total_deaths
5 FROM CORONAVIRUS
6 GROUP BY country
7 ORDER BY total_deaths ASC
8 LIMIT 1;
```

Result Grid | Filter Rows: _____ | Export: | Wrap Cell Content: | Fetch rows:

	country	total_deaths
▶	Dominica	0



15. Top 5 countries having highest recovered case



Query 1 ×

1 -- Q15. Find top 5 countries having highest recovered case

2 • SELECT

3 country,

4 SUM(recovered) AS total_recovered

5 FROM CORONAVIRUS

6 GROUP BY country

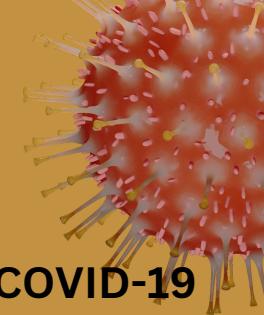
7 ORDER BY total_recovered DESC

8 LIMIT 5;

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

country	total_recovered
India	28089649
Brazil	15400169
US	6303715
Turkey	5202251
Russia	4745756

Conclusions:



1. **Pandemic Trajectory:** The presentation provided crucial insights into the **COVID-19 pandemic's trajectory**, facilitating informed decision-making and contributing to policy interventions aimed at addressing the **global health crisis**.
2. **Data Examination:** Key tasks in the analysis included checking for **NULL values**, assessing **row strength**, determining the **span duration** of data, calculating **monthly averages** for cases, and performing country-level analysis.
3. **Confirmed Cases:** Over the 18-month period from January 2020 to June 2021, the total number of **confirmed COVID-19 cases** reached **16,90,65,144**.
4. **Death Toll:** The highest number of **death cases** recorded was **401,893** in January 2021, with a total of **3,64,78,94** death records throughout the analysis period.
5. **Outbreak Origin:** The first reported outbreak of COVID-19 was on **January 22, 2020**.
6. **Case and Recovery Trends:** The **confirmed** and **recovered cases** were highest in the year 2020, reflecting the initial global impact of the pandemic.
7. **Recovery Performance:** **India** demonstrated the best **recovery performance** among countries, showcasing effective management and treatment strategies.
8. **Impact on US:** The **United States** was the hardest-hit country during the COVID-19 pandemic, facing severe challenges in managing the outbreak.
9. **Recovery Statistics:** A total of **11,30,89,548 recovery cases** were reported during the analysis period, highlighting significant global recovery efforts.
10. **Policy Implications:** These findings underscore the need for targeted **policy interventions** and continued vigilance to manage and mitigate the effects of such global health crises effectively.

Thank You !!