

CORONA DATASET ANALYSIS USING SQL

By Lakshmi Priya

Introduction

- 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.



Dataset Description

- 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.



check whether data have null values

```
3 •   SELECT *
4     FROM corona.corona
5     WHERE NOT( Province IS NOT NULL AND
6                 Country IS NOT NULL AND
7                 Latitude IS NOT NULL AND
8                 Longitude IS NOT NULL AND
9                 Date IS NOT NULL AND
10                Confirmed IS NOT NULL AND
11                Deaths IS NOT NULL AND
12                Recovered IS NOT NULL);
```

update null values with 0

```
5 • Update corona.corona
6   set
7     Province = COALESCE(Province,'0'),
8     Country= COALESCE(Country,'0'),
9     Latitude =COALESCE(Latitude , '0'),
10    Longitude = COALESCE(Longitude,'0'),
11    Date=COALESCE(Date,'0'),
12    Confirmed = COALESCE(Confirmed,'0'),
13    Deaths =COALESCE(Deaths,'0'),
14    Recovered = COALESCE(Recovered,'0');
```

check total number of rows in dataset

```
# check total number of rows  
  
• SELECT COUNT(*) AS total_rows  
FROM corona.corona;
```

	Result Grid	Filter Rows:
	total_rows	
▶	78386	

Check what is start_date and end_date

```
#Check what is start_date and end_date  
  
SELECT min(DATE) AS START_DATE,  
       max(DATE) AS END_DATE  
FROM corona.corona;
```

	Result Grid	Filter Rows:
	START_DATE	END_DATE
▶	2020/01/22	2021/06/13

Number of month present in dataset

- **SELECT**

```
    EXTRACT(MONTH FROM Date) AS MONTH,
```

```
    COUNT(*) AS NUM_OF_MONTHS
```

```
FROM
```

```
corona.corona
```

```
GROUP BY
```

```
    extract(MONTH FROM Date);
```

	MONTH	NUM_OF_MONTHS
1	6314	
2	8778	
3	9548	
4	9240	
5	9548	
6	6622	
7	4774	
8	4774	
9	4620	
10	4774	
11	4620	
12	4774	

Find monthly average for confirmed, deaths, recovered

```
SELECT  
    YEAR(Date) AS year,  
    MONTH(Date) AS month,  
    AVG(Confirmed) AS avg_confirmed,  
    AVG(Deaths) AS avg_deaths,  
    AVG(Recovered) AS avg_recovered  
FROM  
    corona.corona  
GROUP BY  
    YEAR(Date), MONTH(Date);
```

Result

	year	month	avg_confirmed	avg_deaths	avg_recovered
▶	2020	1	4.1455	0.1234	0.0929
	2020	2	15.2960	0.5936	7.0320
	2020	3	161.1303	8.6607	27.8739
	2020	4	505.8004	41.5223	171.6422
	2020	5	574.8498	30.2809	318.2964
	2020	6	859.2281	29.8175	548.7916
	2020	7	1432.3611	35.1096	983.0582
	2020	8	1611.8429	37.5367	1299.2947
	2020	9	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	1	3911.2285	84.1837	1919.6370
	2021	2	2433.3636	69.1649	1558.3917
	2021	3	2916.7972	59.1998	1652.2859
	2021	4	4699.3552	78.4387	3074.7851
	2021	5	4005.2541	76.7803	4007.5078
	2021	6	2508.6324	66.2622	2769.4496

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

```
SELECT MONTH(date) AS Month, YEAR(date) AS Year,  
       MAX(confirmed) AS most_frequent_confirmed,  
       MAX(deaths) AS most_frequent_deaths,  
       MAX(recovered) AS most_frequent_recovered  
FROM corona.corona  
GROUP BY Month, Year  
ORDER BY Year, Month;
```

Result

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

Find minimum values for confirmed,
deaths, recovered per year

```
SELECT  
    YEAR(date) AS year,  
    MIN(confirmed) AS min_confirmed,  
    MIN(deaths) AS min_deaths,  
    MIN(recovered) AS min_recovered  
FROM  
    corona.corona  
GROUP BY  
    YEAR(date);
```

Find maximum values of confirmed,
deaths, recovered per year

```
SELECT  
    YEAR(date) AS year,  
    MAX(confirmed) AS max_confirmed,  
    MAX(deaths) AS max_deaths,  
    MAX(recovered) AS max_recovered  
FROM  
    corona.corona  
GROUP BY  
    YEAR(date);
```

Result Grid | Filter Rows: Export:

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

Result Grid | Filter Rows: Export:

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

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

```
SELECT  
    YEAR(date) AS year,  
    MONTH(date) AS month,  
    SUM(confirmed) AS total_confirmed,  
    SUM(deaths) AS total_deaths,  
    SUM(recovered) AS total_recovered  
FROM  
    corona.corona  
GROUP BY  
    YEAR(date), MONTH(date);
```

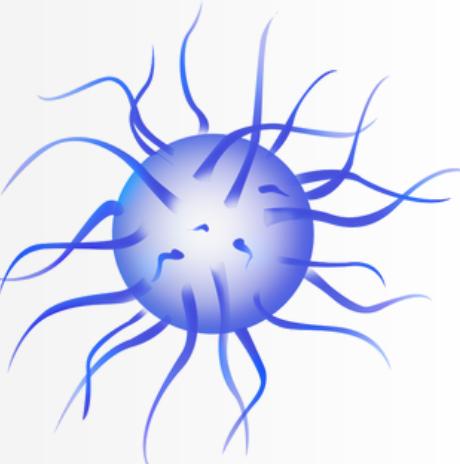
	year	month	total_confirmed	total_deaths	total_recovered
▶	2020	1	6384	190	143
	2020	2	68312	2651	31405
	2020	3	769236	41346	133070
	2020	4	2336798	191833	792987
	2020	5	2744333	144561	1519547
	2020	6	3969634	137757	2535417
	2020	7	6838092	167613	4693120
	2020	8	7694938	179200	6202833
	2020	9	8244794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

Check how corona virus spread out with respect to confirmed case

(Eg.: total confirmed cases, their average, variance & STDEV)

```
SELECT  
    YEAR(DATE) AS YEAR,  
    MONTH(DATE) AS MONTH,  
    SUM(confirmed) AS TOTALConfirmed,  
    AVG(confirmed) AS AVGConfirmed,  
    VARIANCE(confirmed) AS VARIANCEConfirmed,  
    STDDEV(confirmed) AS STDDEVConfirmed  
FROM  
    CORONA.CORONA  
GROUP BY  
    YEAR(DATE),MONTH(DATE)  
ORDER BY  
    YEAR(DATE),MONTH(DATE);
```

	YEAR	MONTH	TOTALConfirmed	AVGConfirmed	VARIANCEConfirmed	STDDEVConfirmed
▶	2020	1	6384	4.1455	4832.911310507683	69.51914348226452
	2020	2	68312	15.2960	78489.44618653238	280.15967980159525
	2020	3	769236	161.1303	1026414.1786678263	1013.1210088966798
	2020	4	2336798	505.8004	7012063.272294189	2648.030073902898
	2020	5	2744333	574.8498	6063580.339194553	2462.4338243279863
	2020	6	3969634	859.2281	13779211.570030754	3712.036041046848
	2020	7	6838092	1432.3611	46914022.88886134	6849.381204814151
	2020	8	7694938	1611.8429	54408583.16090782	7376.217401955275
	2020	9	8244794	1784.5874	69314698.6016606	8325.544943225073
	2020	10	11515841	2412.1996	68988159.04707989	8305.91109072809
	2020	11	16595938	3592.1944	195815877.8115701	13993.4226625072
	2020	12	19336799	4050.4397	459885446.66320175	21444.93988481203
	2021	1	18672205	3911.2285	316304694.1457216	17784.95696215545
	2021	2	10492664	2433.3636	79587921.44429906	8921.206277421179
	2021	3	13924790	2916.7972	83725265.48297499	9150.151118040345
	2021	4	21711021	4699.3552	501013206.38703865	22383.324292585286
	2021	5	19121083	4005.2541	628647609.3315454	25072.84605567436



Q12. Check how corona virus spread out with respect to death case per month -- (Eg.: total confirmed cases, their average, variance & STDEV)

SELECT

```
YEAR(DATE) AS YEAR,  
MONTH(DATE) AS MONTH,  
SUM(DEATHS) AS TOTALDeaths,  
AVG(DEATHS) AS AVGDeaths,  
VARIANCE(DEATHS) AS VARIANCEDeaths,  
STDDEV(DEATHS) AS STDDEVDeaths
```

FROM

CORONA.CORONA

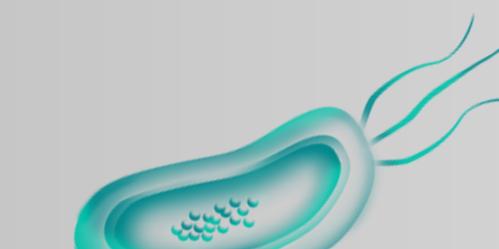
GROUP BY

YEAR(DATE), MONTH(DATE)

ORDER BY

YEAR(DATE), MONTH(DATE);

	YEAR	MONTH	TOTALDeaths	AVGDeaths	VARIANCEDeaths	STDDEVDeaths
▶	2020	1	190	0.1234	4.245817169843138	2.0605380777464752
	2020	2	2651	0.5936	68.3218488238449	8.265703165723101
	2020	3	41346	8.6607	3900.7922648320746	62.45632285711411
	2020	4	191833	41.5223	40504.26811767955	201.25672191924312
	2020	5	144561	30.2809	20684.911671085658	143.8225005730524
	2020	6	137757	29.8175	16929.445709928572	130.11320344195886
	2020	7	167613	35.1096	21140.154944373826	145.39654378414167
	2020	8	179200	37.5367	23272.99645685882	152.55489653517785
	2020	9	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	1	401893	84.1837	102758.43231925515	320.55956126631935
	2021	2	298239	69.1649	68478.87146663864	261.6846794648832
	2021	3	282620	59.1998	54385.969702527414	233.20799665218905
	2021	4	362387	78.4387	94611.47092309907	307.58977701331213
	2021	5	366549	76.7803	131769.4693132085	363.00064643635073



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

```
SELECT
    YEAR(date) AS YEAR,
    MONTH(date) AS MONTH,
    SUM(recovered) AS TOTALRecoveredCases,
    AVG(recovered) AS AVGRecoveredCases,
    VARIANCE(recovered) AS VARIANCERecoveredCases,
    STDDEV(recovered) AS STDDEVRecoveredCases
FROM
    CORONA.CORONA
GROUP BY
    YEAR(date), MONTH(date)
ORDER BY
    YEAR(date), MONTH(date);
```

	YEAR	MONTH	TOTALRecoveredCases	AVGRecoveredCases	VARIANCERecoveredCases	STDDEVRecoveredCases
▶	2020	1	143	0.0929	2.633585343228195	1.6228325062150422
	2020	2	31405	7.0320	12446.6619841429	111.5646090126385
	2020	3	133070	27.8739	40113.18979639222	200.28277458731247
	2020	4	792987	171.6422	769893.0319414467	877.4354859141763
	2020	5	1519547	318.2964	1978206.4175948969	1406.4872617961732
	2020	6	2535417	548.7916	6530172.493132202	2555.420218502664
	2020	7	4693120	983.0582	24843877.85333294	4984.363334803447
	2020	8	6202833	1299.2947	40170422.19780621	6338.014057873824
	2020	9	6647749	1438.9067	57023566.44389443	7551.394999858929
	2020	10	6782150	1420.6431	73731702.50184019	8586.716631043568
	2020	11	9172292	1985.3446	50727618.87346639	7122.332404027938
	2020	12	11924903	2497.8850	326694724.1038681	18074.69845125689
	2021	1	9164347	1919.6370	31493700.11602569	5611.924813825083
	2021	2	6719785	1558.3917	24427411.60468984	4942.409493828879
	2021	3	7888013	1652.2859	34897391.64112127	5907.401428811256
	2021	4	14205507	3074.7851	224419585.15055594	14980.640345143993
	2021	5	19131842	4007.5078	755175531.756431	27480.45726978412

Find Country having highest number of the Confirmed case

```
#Q14. Find Country having highest number of the Confirmed case
select country,max(confirmed) as Highest_confirmed_case
from corona.corona
group by Country
order by Highest_confirmed_case desc
limit 1;
```

	country	Highest_confirmed_case
▶	Turkey	823225

Find Country having lowest number of the death case

```
select country, min(deaths) as lowest_num_deaths
from corona.corona
group by country
order by lowest_num_deaths desc
limit 1;
```

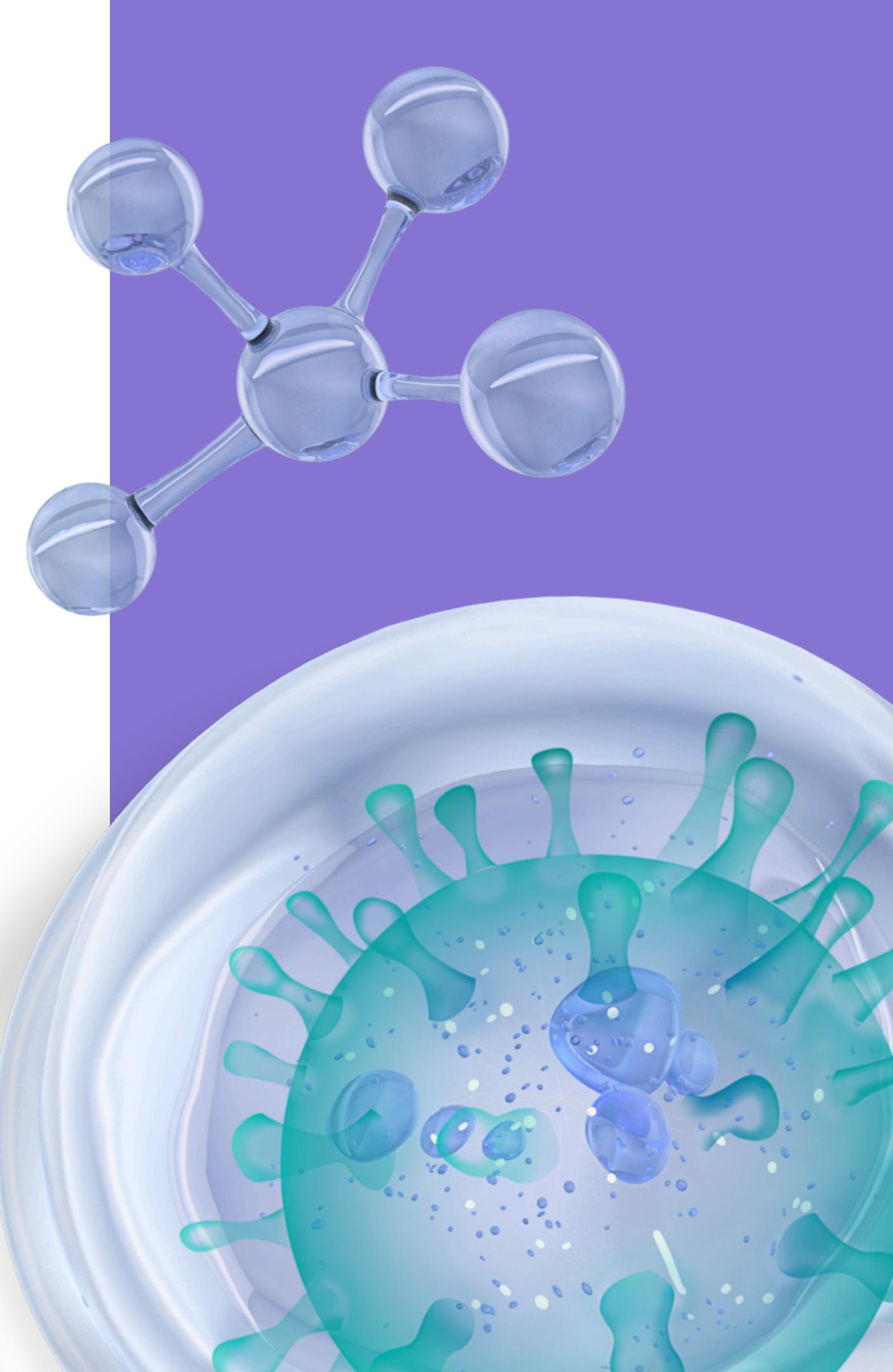
	country	lowest_num_deaths
▶	Afghanistan	0

Find top 5 countries having highest recovered case

```
select country,max(recovered) as highest_recoverd_case  
from corona.corona  
group by country  
order by highest_recoverd_case desc  
limit 5;
```

Result Grid | Filter Rows:

	country	highest_recoverd_case
▶	Turkey	1123456
	India	422436
	Brazil	388340
	US	150267
	Colombia	89557



THANK YOU

Contact us



Lakshmi Priya

- +91 8639851790
- palagullalakshmipriya@gmail.com
- www.linkedin.com/in/palagulla-lakshmi-priya-51697b257/

