Corona Virus Analysis

Description: The Coronavirus pandemic has significantly impacted public health, creating an urgent need for data-driven insights to understand its spread. As a data analyst, you will analyze a Coronavirus dataset to uncover key patterns and trends, such as infection rates, mortality rates, and recovery statistics across different regions and demographics.

Query for Dataset:

```
SELECT * FROM `CoronaVirusDataset` LIMIT 10;
```

Output:

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

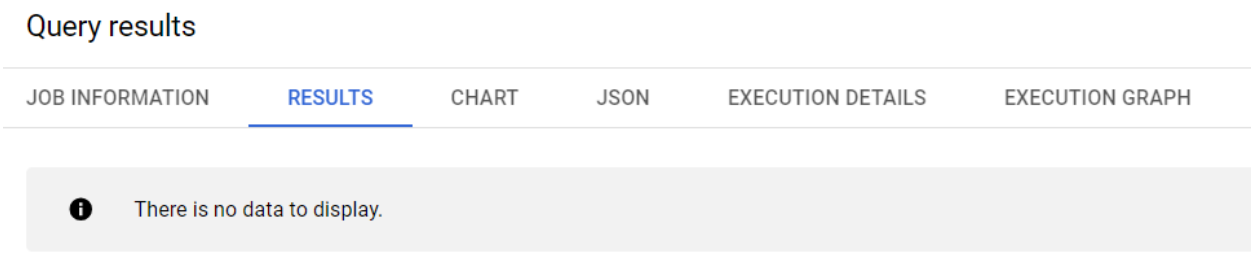
Row	Province	Country	Latitude	Longitude	Date	Confirmed	Deaths	Recovered
1	Afghanistan	Afghanistan	33.93911	67.709953	3/6/2020	758	24	72
2	Afghanistan	Afghanistan	33.93911	67.709953	6/6/2020	582	18	68
3	Afghanistan	Afghanistan	33.93911	67.709953	7/6/2020	791	30	45
4	Afghanistan	Afghanistan	33.93911	67.709953	10/6/2020	684	21	362
5	Afghanistan	Afghanistan	33.93911	67.709953	11/6/2020	748	21	313
6	Afghanistan	Afghanistan	33.93911	67.709953	12/6/2020	656	20	602
7	Afghanistan	Afghanistan	33.93911	67.709953	14-06-2020	664	20	524
8	Afghanistan	Afghanistan	33.93911	67.709953	18-06-2020	658	42	1502
9	Afghanistan	Afghanistan	33.93911	67.709953	20-06-2020	546	21	330
10	Afghanistan	Afghanistan	33.93911	67.709953	23-06-2020	324	20	419

Insights: Our dataset contains 8 Columns named as :

- 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.

-- Q1. Write a code to check NULL values

```
SELECT Province,Country,Latitude,Longitude,Date,Confirmed,Deaths, Recovered
FROM ` CoronaVirusDataset`
WHERE Province||Country||Latitude||Longitude||Date||Confirmed||Deaths|| Recovered
IS NULL;
```



Insights: There is no missing value in our dataset

--Q2. If NULL values are present, update them with zeros for all columns.

```
SELECT IFNULL(Province, 0 )
FROM `CoronaVirusDataset`;
```

Insights: There are no missing values are present. If present then we can replace it with zero by using below query

-- Q3. check total number of rows

```
select count(*) as Total_raows
from ` CoronaVirusDataset`;
```

Insights: There are 78386 rows in our dataset.

-- Q4. Check what is start_date and end_date

```
Select MIN(Date) as start_date, MAX(Date) as end_date,
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	start_date	end_date
1	9/9/2020	1/1/2021

Insights: Start Date of Corona in our dataset is 9th Sept 2020 and End date is 1st Jan 2021.

-- Q5. Number of month present in dataset

```
SELECT COUNT(DISTINCT EXTRACT(MONTH FROM Date)) AS num_months
FROM `CoronaVirusDataset`;
```

Insights: There are 12 Distinct months are present.

-- Q6. Find average for confirmed, deaths, recovered rate

```
select Round(avg(Confirmed),2) as avg_confirmed_rate, Round(avg(Deaths),2) as
avg_death_rate , round(avg(Recovered),2) as avg_recovery_rate
from `CoronaVirusDataset` ;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	avg_confirmed_rate	avg_death_rate	avg_recovery_rate
1	2156.83	46.54	1442.73

Insights: In our dataset Average rate for confirmed is “2156.83”, deaths is “46.54”, recovered rate is “1442.73”.

-- Q7. Find most frequent value for confirmed, deaths, recovered each month

Confirmed Cases

```
SELECT Confirmed, COUNT(Confirmed) AS `most_confirmed`
FROM `CoronaVirusDataset`
GROUP BY 1
ORDER BY 1 LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	Confirmed	most_confirmed
1	0	26050
2	1	2728
3	2	1620
4	3	1138
5	4	945

Death cases

```
SELECT Deaths, COUNT(Deaths) AS `most_Deaths`
FROM `CoronaVirusDataset`
GROUP BY 1
ORDER BY 1 LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	Deaths	most_Deaths
1	0	43082
2	1	4088
3	2	2457
4	3	1932
5	4	1644

Recovered cases

```
SELECT Recovered, COUNT(Recovered) AS `most_Recovered`
FROM `CoronaVirusDataset`
GROUP BY 1
ORDER BY 1 LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	Recovered	most_Recovered
1	0	36609
2	1	2173
3	2	1303
4	3	957
5	4	760

Insights: In our dataset for Confirmed, Death and recovered cases, most frequent value is 0 followed by 1.

-- Q8. Find minimum values for confirmed, deaths, recovered.

```
select min(Confirmed) as min_confirmed, min(Deaths) as min_death , min(Recovered)
as min_recovery
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	min_confirmed	min_death	min_recovery
1	0	0	0

Insights: In our dataset for Confirmed, Death and recovered cases, minimum value is 0.

-- Q9. Find maximum values of confirmed, deaths, recovered.

```
select max(Confirmed) as max_confirmed, max(Deaths) as max_death ,
max(Recovered) as max_recovery
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	max_confirmed	max_death	max_recovery
1	823225	7374	1123456

Insights: In our dataset for Confirmed, Death and recovered cases, maximum value is 823225, 7374, 1123456 respectively.

-- Q10. The total number of case of confirmed, deaths, recovered.

```
select sum(Confirmed) as total_confirmed, sum(Deaths) as total_death ,
sum(Recovered) as total_recovery
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	total_confirmed	total_death	total_recovery
1	169065144	3647894	113089548

Insights: Total number of Confirmed, Death and recovered cases, is 169065144, 3647894, 113089548.

-- Q11. Check how corona virus spread out with respect to confirmed case (Eg.: total confirmed cases, their average, variance & STDEV).

```
select sum(Confirmed) as total_confirmed, round(avg(Confirmed),2) as
avg_confirmed,round(stddev(distinct Confirmed),2) as standard_deviation,
round(variance(distinct Confirmed),2) as confirm_variance
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	total_confirmed	avg_confirmed	standard_deviation	confirm_variance
1	169065144	2156.83	33315.33	1109911026.16

Insights: There are total “169065144” confirmed cases for which average rate of confirmation is “2156.83”, standard deviation is “33315.33” and variance is calculated as “1109911026.16” .

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

```
select sum(Deaths) as total_Deaths, round(avg(Deaths),2) as
avg_Deaths,round(stddev(distinct Deaths),2) as standard_deviation,
round(variance(distinct Deaths),2) as Deaths_variance
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	total_Deaths	avg_Deaths	standard_deviation	Deaths_variance
1	3647894	46.54	972.38	945515.27

Insights: There are total “3647894” death cases for which average rate of death is “46.54”, standard deviation is “972.38” and variance is calculated as “945515.27”.

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

```
select sum(Recovered) as total_Recovered, round(avg(Recovered),2) as
avg_Recovered,round(stddev(distinct Recovered),2) as standard_deviation,
round(variance(distinct Recovered),2) as Recovered_variance
from `CoronaVirusDataset`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	total_Recovered	avg_Recovered	standard_deviation	Recovered_variance
1	113089548	1442.73	30652.1	939551283.3

Insights: There are total “113089548” recovered cases for which average rate of recovery is “1442.73”, standard deviation is “30652.1” and variance is calculated as “939551283.3”.

-- Q14. Find Country having highest number of the Confirmed case

```
select Country,sum(Confirmed) as Confirmed,
from `CoronaVirusDataset`
group by 1
order by 2 desc;
```

Query results

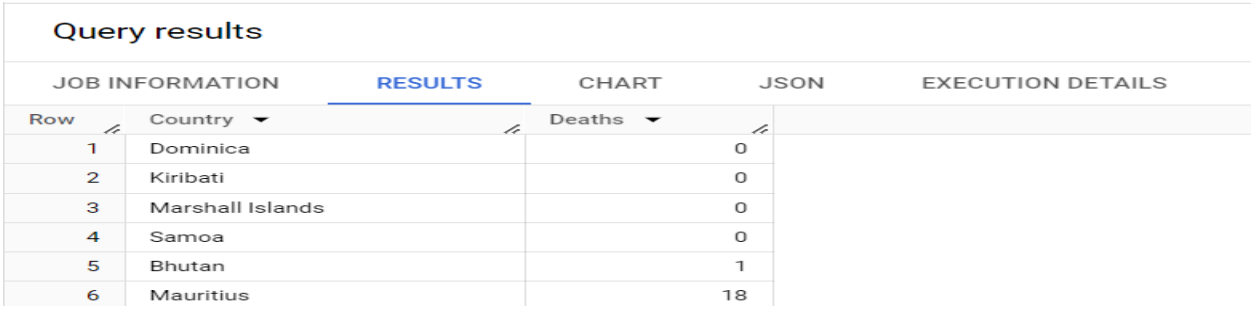
JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	Country	Confirmed
1	US	33461982
2	India	29460523
3	Brazil	17412766
4	France	6106009
5	Turkey	5330447

Insights: US is the country which is having the highest number of confirmed cases(33461982), followed by India(29461982) and Brazil(17412766).

-- Q15. Find Country having lowest number of the death case

```
select Country,sum(Deaths) as Deaths,
from `CoronaVirusDataset`
group by 1
order by 2 ;
```



Query results

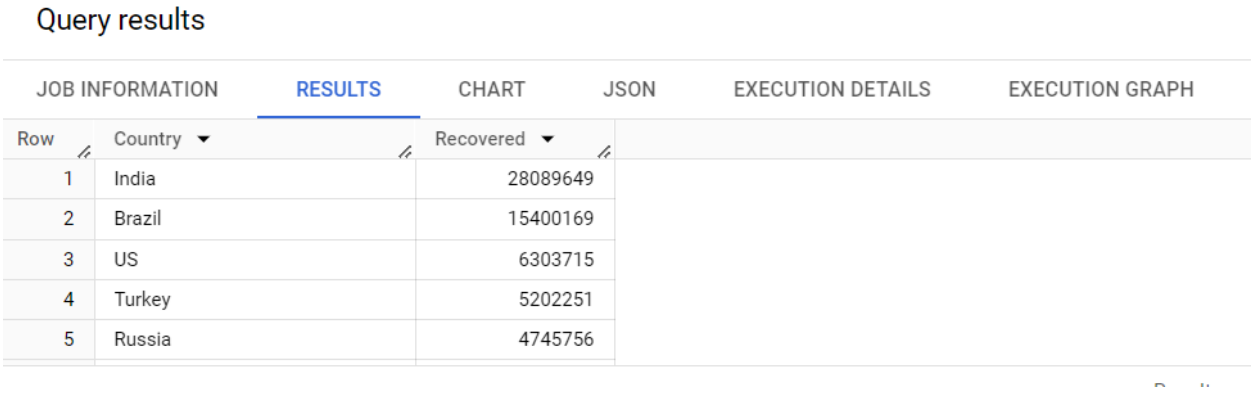
JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	Country	Deaths
1	Dominica	0
2	Kiribati	0
3	Marshall Islands	0
4	Samoa	0
5	Bhutan	1
6	Mauritius	18

Insights: Dominica, Kiribati, Marshall Islands, Samoa are the country in which 0 deaths happen, followed by Bhutan with only 1 death.

-- Q16. Find top 5 countries having highest recovered case

```
select Country,sum(Recovered) as Recovered,
from `CoronaVirusDataset`
group by 1
order by 2 desc,
```



Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	Country	Recovered
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756

Insights: India, Brazil, US, Turkey, Russia are the top 5 countries in which recovered cases are high.