CORONA VIRUS ANALYSIS

BY: MARJORIE CALIMOSO

PROJECT OVERVIEW

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.

In this project, we aimed to analyze corona virus dataset all over the world using MySQL. The data set includes Province, Country/Region, Longitude, Latitude, Date, Confirmed, Deaths and Recovered Cases.

The goal is to answer 16 major issue statements using SQL queries related to the dataset that will guide strategic decisions and optimize the status of corona virus around the world.

PROJECT OBJECTIVE

To analyze corona virus status in the world by writing SQL queries to extract relevant insights from the dataset.

DATA ANALYSIS

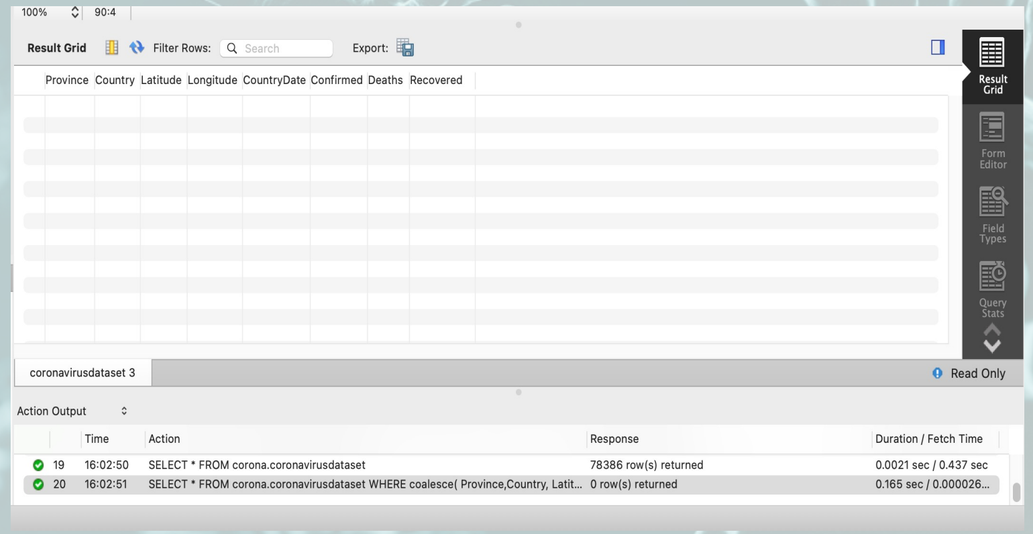
Question No. 1

Write a code to check NULL values

Query:

```
SELECT * FROM corona.coronavirusdataset
WHERE coalesce( Province, Country, Latitude,
Longitude, CountryDate, Confirmed,
 Deaths, Recovered) IS NULL;
```

Result Preview



DATA ANALYSIS

Question No. 2

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

Query

```
UPDATE corona.coronavirusdataset
SET Province = coalesce(Province, 0),
    Country = coalesce(Country, 0),
    Latitude = coalesce(Latitude, 0),
    Longitude = coalesce(Longitude, 0),
    CountryDate = coalesce(CountryDate, 0),
    Confirmed = coalesce(Confirmed, 0),
    Deaths = coalesce(Deaths, 0),
    Recovered = coalesce(Recovered, 0)
WHERE Province IS NULL OR Country IS NULL OR Latitude IS NULL OR
Longitude IS NULL OR CountryDate IS NULL OR Confirmed IS NULL OR
Deaths IS NULL OR Recovered IS NULL;
```

100% 41:13

Result Preview

Action Output

	Time	Action	Response	Duration / Fetch Time
170	00:37:21	UPDATE corona.coronavirusdataset...	0 row(s) affected Rows matched: 0 Changed: 0 War...	0.158 sec

DATA ANALYSIS

Question No. 3

Check total number of rows

Query

Result Preview

```
21 •  SELECT COUNT(*) AS total_rows FROM corona.coronavirusdataset;
22
```

100% 37:19

Result Grid | Filter Rows: Search | Export:

total_rows
78386

Result Grid | Form Editor

Result 51 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
171	00:40:15	SELECT COUNT(*) AS total_rows FR...	1 row(s) returned	0.014 sec / 0.000009...

DATA ANALYSIS

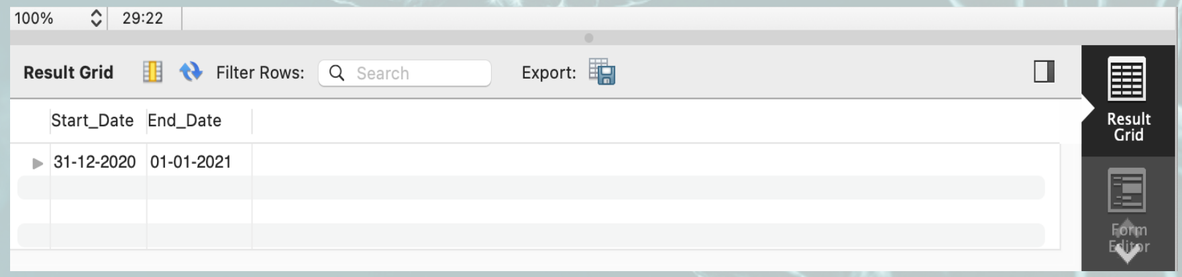
Question No. 4

Check what is start_date and end_date

Query:

```
SELECT MIN(CountryDate) AS Start_Date,
MAX(CountryDate) AS End_Date FROM
corona.`coronavirusdataset`;
```

Result Preview



DATA ANALYSIS

Question No. 5

Number of month present in dataset

Query

Result Preview

```
24  SELECT
25      COUNT(DISTINCT EXTRACT(MONTH FROM STR_TO_DATE(CountryDate, '%d-%m-%Y'))) AS total_month
26   FROM
27  corona.`coronavirusdataset`;
28
```

100% 1:29

Result Grid | Filter Rows: Search | Export:

total_months
12

Result 16 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
110	21:23:33	SELECT COUNT(DISTINCT EXTRAC...	1 row(s) returned	0.105 sec / 0.000011...

DATA ANALYSIS

Question No. 6

Find monthly average for confirmed, deaths, recovered

Query

Result Preview

```
31 •  SELECT
32        AVG(Confirmed) AS ave_confirmed,
33        AVG(Deaths) AS ave_deaths,
34        AVG(Recovered) AS ave_recovered
35     FROM
36    coronavirusdataset GROUP BY EXTRACT(MONTH FROM CountryDate);
37
```

100% 7:31

Result Grid | Filter Rows: Search | Export:

ave_confirmed	ave_deaths	ave_recovered
2156.8283	46.5376	1442.7264

Result Grid | Form Editor

Result 20 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
115	21:44:45	SELECT	AVG(Confirmed) AS av...	1 row(s) returned	0.150 sec / 0.000020...

DATA ANALYSIS

Question No. 7

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

Query

Result Preview

```
39 •  SELECT
40        MAX(Confirmed) AS most_frequent_confirmed,
41        MAX(Deaths) AS most_frequent_deaths,
42        MAX(Recovered) AS most_frequent_recovered
43     FROM
44    coronavirusdataset GROUP BY EXTRACT(MONTH FROM CountryDate);
45
46
```

100% 1:45

Result Grid | Filter Rows: Search | Export:

most_frequent_confirmed	most_frequent_deaths	most_frequent_recovered
823225	7374	1123456

Result 21 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
116	21:49:30	SELECT	MAX(Confirmed) AS mos...	1 row(s) returned	0.129 sec / 0.000014...

DATA ANALYSIS

Question No. 8

Find minimum values for confirmed, deaths, recovered per year

Query

Result Preview

```
47 •  SELECT
48        MIN(Confirmed) AS min_frequent_confirmed,
49        MIN(Deaths) AS min_frequent_deaths,
50        MIN(Recovered) AS min_frequent_recovered
51     FROM
52    coronavirusdataset GROUP BY EXTRACT(YEAR FROM CountryDate);
53
```

100% 1:53

Result Grid | Filter Rows: Search | Export:

min_frequent_confirmed	min_frequent_deaths	min_frequent_recovered
0	0	0

Result 26 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
122	22:01:32	SELECT MIN(Confirmed) AS min_...	1 row(s) returned	0.173 sec / 0.000020...

DATA ANALYSIS

Question No. 9

Find maximum values of confirmed, deaths, recovered per year

Query

Result Preview

```
54 •  SELECT
55        MAX(Confirmed) AS max_per_year_confirmed,
56        MAX(Deaths) AS max_per_year_deaths,
57        MAX(Recovered) AS max_per_year_recovered
58     FROM
59    coronavirusdataset GROUP BY EXTRACT(YEAR FROM CountryDate);
60
```

100% 1:60

Result Grid | Filter Rows: Search | Export:

max_per_year_confirmed	max_per_year_deaths	max_per_year_recovered
823225	7374	1123456

Result 30 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
129	22:10:07	SELECT	MAX(Confirmed) AS max...	1 row(s) returned	0.145 sec / 0.000023...

DATA ANALYSIS

Question No. 10

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

Query

Result Preview

```
62 •  SELECT
63        sum(Confirmed) AS total_each_month_confirmed,
64        sum(Deaths) AS total_each_month_deaths,
65        sum(Recovered) AS total_each_month_recovered
66     FROM
67    coronavirusdataset GROUP BY EXTRACT(MONTH FROM CountryDate)
68    ORDER BY EXTRACT(MONTH FROM CountryDate);
69
```

100% 1:69

Result Grid | Filter Rows: Search | Export:

total_each_month_confirmed	total_each_month_deaths	total_each_month_recovered
169065144	3647894	113089548

Result 31 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
130	22:15:07	SELECT	sum(Confirmed) AS total...	1 row(s) returned	0.174 sec / 0.000021...

DATA ANALYSIS

Question No. 11

Check how corona virus spread out with respect to confirmed case

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

Query

Result Preview

```
70  -- Calculate total confirmed cases, average, variance, stdev
71
72  SELECT
73  sum(Confirmed) AS total_confirmed_cases,
74  avg(Confirmed) AS ave_confirmed_cases,
75  variance(Confirmed) AS variance_confirmed_cases,
76  stddev(Confirmed) AS stdev_confirmed_cases
77
78  FROM coronavirusdataset;
79
```

100% 2:79

Result Grid | Filter Rows: Search | Export:

total_confirmed_cases	ave_confirmed_cases	variance_confirmed_cases	stdev_confirmed_cases
169065144	2156.8283	157288925.07796532	12541.488152446875

Result 36 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
135	22:24:25	SELECT sum(Confirmed) AS total_c...	1 row(s) returned	0.061 sec / 0.000016...

DATA ANALYSIS

Question No. 12

Check how corona virus spread out with respect to death case per month

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

Query

Result Preview

```
79
80      -- Calculate total deaths cases, average, variance, stdev
81  •   SELECT
82      sum(Deaths) AS total_deaths_cases,
83      avg(Deaths) AS ave_deaths_cases,
84      variance(Deaths) AS variance_deaths_cases,
85      stddev(Deaths) AS stdev_deaths_cases
86
87      FROM coronavirusdataset GROUP BY EXTRACT(MONTH FROM CountryDate);
88
```

100% 1:88

Result Grid | Filter Rows: Search | Export:

total_deaths_cases	ave_deaths_cases	variance_deaths_cases	stdev_deaths_cases
3647894	46.5376	45892.01885355753	214.22422564583476

Result Grid

Result 42 Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
157	23:56:32	SELECT sum(Deaths) AS total_deat...	1 row(s) returned	0.143 sec / 0.000009...

DATA ANALYSIS

Question No. 13

Check how corona virus spread out with respect to recovered case

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

Query

Result Preview

```
89     -- Calculate total recovered cases, average, variance, stdev
90   SELECT
91   sum(Recovered) AS total_recovered_cases,
92   avg(Recovered) AS ave_recovered_cases,
93   variance(Recovered) AS variance_recovered_cases,
94   stddev(Recovered) AS stdev_recovered_cases
95
96   FROM coronavirusdataset GROUP BY EXTRACT(MONTH FROM CountryDate);
97
```

100% 1:97

Result Grid | Filter Rows: Search | Export:

total_recovered_cases	ave_recovered_cases	variance_recovered_cases	stdev_recovered_cases
113089548	1442.7264	107029523.26229636	10345.507395110999

Result Grid

Result 43 Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
158	23:59:37	SELECT sum(Recovered) AS total_r...	1 row(s) returned	0.162 sec / 0.000020...

DATA ANALYSIS

Question No. 14

Find Country having highest number of the Confirmed case

Query

Result Preview

```
98    -- Country having the highest number or Confirmed cases
99
100   SELECT `Country`,
101    Confirmed FROM coronavirusdataset WHERE
102   Confirmed = (SELECT MAX(Confirmed) FROM coronavirusdataset);
103
```

100% 1:103

Result Grid | Filter Rows: Search | Export:

Country	Confirmed
Turkey	823225

Result Grid

coronavirusdataset 45 | Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
163	00:06:51	SELECT `Country`, Confirmed FRO...	1 row(s) returned	0.093 sec / 0.000010...

DATA ANALYSIS

Question No. 15

Find Country having lowest number of the death case

Query

Result Preview

```
104    -- Country having the minimum number of Deaths cases
105
106    SELECT `Country`,
107    MIN(Deaths) AS LowestDeaths
108    FROM coronavirusdataset GROUP BY `Country`
109     ORDER BY LowestDeaths ASC LIMIT 5;
110
111
112
```

100% 2:114

Result Grid | Filter Rows: Search | Export:

Country	LowestDeaths
Afghanistan	0
Algeria	0
Argentina	0
Australia	0
Austria	0

Result 47 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
166	00:13:23	SELECT `Country`, MIN(Deaths) AS...	5 row(s) returned	0.154 sec / 0.000019...

DATA ANALYSIS

Question No. 16

Find top 5 countries having highest recovered case

Query

Result Preview

```
113 •  SELECT `Country`, sum(Recovered) AS HighestRecovered
114    FROM coronavirusdataset GROUP BY `Country`
115     ORDER BY HighestRecovered DESC LIMIT 5;
116
117
118
```

100% 2:117

Result Grid | Filter Rows: Search | Export:

Country	HighestRecovered
India	28089649
Brazil	15400169
US	6303715
Turkey	5202251
Russia	4745756

Result Grid | Form Editor

Result 50 — Read Only

Action Output

	Time	Action	Response	Duration / Fetch Time
169	00:18:09	SELECT `Country`, sum(Recovered)...	5 row(s) returned	0.153 sec / 0.000019...

OVERALL INSIGHTS

- The dataset does not contain NULL values
- There are 78,386 rows in the dataset
- The start date of corona virus is December 31, 2020 and end date is January 01, 2021 according from the dataset
- There are 12 months in the dataset
- The monthly average for confirmed cases is 2,156.8283, average deaths is 46.5376 and 1,442.7284 for recovered cases
- The most frequent value each month in confirmed cases is 823,225, deaths 7,374 and in recovered cases is 1,123,456
- There are no minimum values per year
- The maximum values per year in confirmed cases is 823,225, deaths 7,374 and in recovered cases is 1,123,456
- The total number of cases each month confirmed is 169,065,144, deaths 3,647,894 and recovered is 113,089,548
- The total variance per month of confirmed cases is 157,288,925 and the STDEV is 12,541
- The total variance per month of deaths is 45,892 and STDEV is 214
- The total variance per month of recovered is 107,029,523 and STDEV is 10,345
- The country having the lowest death cases is Afghanistan, Algeria, Argentina, Australia, and Austria
- The top 5 countries having the highest recovered cases is India, Brazil, USA, Turkey, and Russia