Understanding the Impact of COVID-19: A Comprehensive Analysis

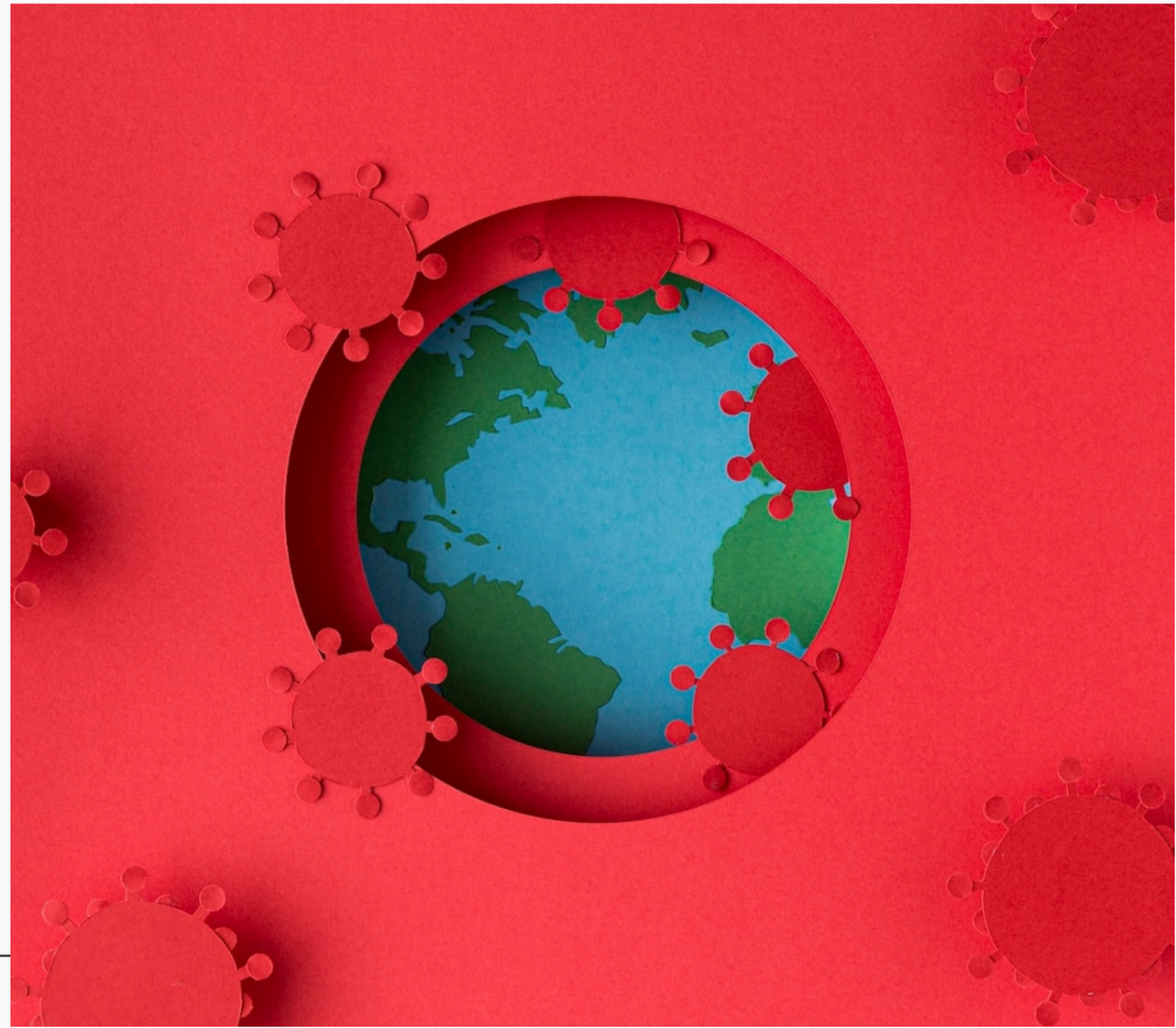
- by Shubham Nevgi

TABLE OF CONTENTS

POJECT OVERVIEW

The coronavirus, causes COVID-19 which can bring mild flu-like symptoms or severe respiratory illness. It spreads through coughs and sneezes.

This analysis aims to explore the impact of the coronavirus pandemic by examining key metrics such as confirmed cases, deaths, and recoveries across different regions over time.



DATA DESCRIPTION

- **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.

EXPLORATORY DATA ANALYSIS

To avoid any errors, check missing value / null value

Q1. Write a code to check NULL values

```
--  Q1. Write a code to check NULL values
SELECT *
FROM CoronaReport
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;
```

100 %

Results | Messages

Province	Country_Region	Latitude	Longitude	Date	Confirmed	Deaths	Recovered

EXPLORATORY DATA ANALYSIS

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

No NULL Values in the Dataset

EXPLORATORY DATA ANALYSIS

Q3. check total number of rows

```sql
-- Q3. check total number of rows
SELECT COUNT(*) AS Total_Rows
FROM CoronaReport;
```

100 %

Results | Messages

	Total_Rows
1	78386

EXPLORATORY DATA ANALYSIS

Q4. Check what is start_date and end_date

```sql
-- Q4. Check what is start_date and end_date
SELECT
    MIN(Date) AS Start_Date,
    MAX(Date) AS End_Date
FROM CoronaReport;
```

100 %

Results | Messages

	Start_Date	End_Date
1	2020-01-22	2021-06-13

EXPLORATORY DATA ANALYSIS

Q5. Number of month present in dataset

```sql
-- Q5. Number of month present in dataset
SELECT COUNT(DISTINCT FORMAT(Date, 'yyyy-MM')) AS Number_of_Months
FROM CoronaReport;
```

100 %

Results | Messages

	Number_of_Months
1	18

EXPLORATORY DATA ANALYSIS

Q6. Find monthly average for confirmed, deaths, recovered

```sql
-- Q5. Number of month present in dataset
SELECT COUNT(DISTINCT FORMAT(Date, 'yyyy-MM')) AS Number_of_Months
FROM CoronaReport;
```

100 %

Results | Messages

	Month	Avg_Confirmed	Avg_Deaths	Avg_Recovered
1	2020-05	574.849811478844	30.280896522832	318.296397151236
2	2021-05	4005.25408462505	76.7802681189778	4007.5077503142
3	2021-03	2916.79723502304	59.1998324256389	1652.28592375367
4	2020-09	1784.58744588745	34.7772727272727	1438.90670995671
5	2021-02	2433.36363636364	69.1648886827458	1558.39169758813
6	2020-04	505.800432900433	41.5222943722944	171.642207792208
7	2021-06	2508.63236763237	66.2622377622378	2769.44955044955

EXPLORATORY DATA ANALYSIS

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

```sql
-- Q7. Find most frequent value for confirmed, deaths, recovered each month
WITH RankedData AS (
    SELECT
        FORMAT(Date, 'yyyy-MM') AS Month,
        Confirmed,
        Deaths,
        Recovered,
        ROW_NUMBER() OVER (PARTITION BY FORMAT(Date, 'yyyy-MM') ORDER BY COUNT(*) DESC) AS Rank
    FROM CoronaReport
    GROUP BY FORMAT(Date, 'yyyy-MM'), Confirmed, Deaths, Recovered
)
SELECT Month, Confirmed, Deaths, Recovered
FROM RankedData
WHERE Rank = 1;
```

100 %

Results | Messages

	Month	Confirmed	Deaths	Recovered
1	2020-01	0	0	0
2	2020-02	0	0	0
3	2020-03	0	0	0
4	2020-04	0	0	0
5	2020-05	0	0	0

EXPLORATORY DATA ANALYSIS

Q8. Find minimum values for confirmed, deaths, recovered per year

```sql
-- Q8. Find minimum values for confirmed, deaths, recovered per year
SELECT
FORMAT(Date, 'yyyy') AS year,
        MIN(Confirmed) AS Minimum_Confirmed,
        MIN(Deaths) AS Minimum_Deaths,
        MIN(Recovered) AS Minimum_Recovered
FROM CoronaReport
GROUP BY FORMAT(Date, 'yyyy');
```

100 %

Results | Messages

	year	(No column name)	(No column name)	(No column name)
1	2020	0	0	0
2	2021	0	0	0

EXPLORATORY DATA ANALYSIS

Q9. Find maximum values of confirmed, deaths, recovered per year

```sql
-- Q9. Find maximum values of confirmed, deaths, recovered per year
SELECT
FORMAT(Date, 'yyyy') AS year,
        MAX(Confirmed) AS Maximum_Confirmed,
        MAX(Deaths) AS Maximum_Deaths,
        MAX(Recovered) AS Maximum_Recovered
FROM CoronaReport
GROUP BY FORMAT(Date, 'yyyy');
```

100 %

Results | Messages

	year	Minimum_Confirmed	Minimum_Deaths	Minimum_Recovered
1	2020	823225	3752	1123456
2	2021	414188	7374	422436

EXPLORATORY DATA ANALYSIS

Q10. The total number of case of confirmed, deaths, recovered each month

```sql
-- Q10. The total number of case of confirmed, deaths, recovered each month
SELECT
FORMAT(Date, 'yyyy-MM') AS Month,
        SUM(Confirmed) AS Total_Confirmed,
        SUM(Deaths) AS Total_Deaths,
        SUM(Recovered) AS Total_Recovered
FROM CoronaReport
GROUP BY FORMAT(Date, 'yyyy-MM');
```

100 %

Results | Messages

	Month	Total_Confirmed	Total_Deaths	Total_Recovered
1	2020-05	2744333	144561	1519547
2	2021-05	19121083	366549	19131842
3	2021-03	13924790	282620	7888013
4	2020-09	8244794	160671	6647749
5	2021-02	10492664	298239	6719785
6	2020-04	2336798	191833	792987
7	2021-06	5022282	132657	5544438
8	2020-10	11515841	175484	6782150
9	2020-08	7694938	179200	6202833

EXPLORATORY DATA ANALYSIS

Q11. Check how corona virus spread out with respect to confirmed case

```sql
-- 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,
    AVG(CAST(Confirmed AS FLOAT)) AS Avg_Confirmed,
    VAR(Confirmed) AS Var_Confirmed,
    STDEV(Confirmed) AS Stdev_Confirmed
FROM CoronaReport;
```

100 %

Results | Messages

	Total_Confirmed	Avg_Confirmed	Var_Confirmed	Stdev_Confirmed
1	169065144	2156.82831117802	157290931.698175	12541.5681514783

EXPLORATORY DATA ANALYSIS

Q12. Check how corona virus spread out with respect to death case per month

```sql
-- Q12. Check how corona virus spread out with respect to death case per month
--      (Eg.: total confirmed cases, their average, variance & STDEV )
SELECT
    FORMAT(Date, 'yyyy-MM') AS Month,
    SUM(Deaths) AS Total_Deaths,
    AVG(CAST(Confirmed AS FLOAT)) AS Avg_Deaths,
    VAR(Deaths) AS Var_Deaths,
    STDEV(Deaths) AS Stdev_Deaths
FROM CoronaReport
GROUP BY FORMAT(Date, 'yyyy-MM');
```

100 %

Results | Messages

	Month	Total_Deaths	Avg_Deaths	Var_Deaths	Stdev_Deaths
1	2020-05	144561	574.849811478844	20689.2454049367	143.837566042174
2	2021-05	366549	4005.25408462505	131797.07657684	363.03867091102
3	2021-03	282620	2916.79723502304	54397.3642069696	233.232425290674
4	2020-09	160671	1784.58744588745	20107.1214145132	141.799581855918
5	2021-02	298239	2433.36363636364	68494.7561503472	261.715028514503
6	2020-04	191833	505.800432900433	40513.0371733448	201.278506486273
7	2021-06	132657	2508.63236763237	113020.126599288	336.184661457491
8	2020-10	175484	2412.19962295769	17583.7542527085	132.60374901453
9	2020-08	179200	1611.84289903645	23277.8724251087	152.570876726552
10	2020-11	262247	3592.19437229437	27779.8065421012	166.672752848512

EXPLORATORY DATA ANALYSIS

Q13. Check how corona virus spread out with respect to recovered case

```sql
-- Q13. Check how corona virus spread out with respect to recovered case
--      (Eg.: total confirmed cases, their average, variance & STDEV )
SELECT
    FORMAT(Date, 'yyyy-MM') AS Month,
    SUM(Recovered) AS Total_Recovered,
    AVG(CAST(Recovered AS FLOAT)) AS Avg_Recovered,
    VAR(Recovered) AS Var_Recovered,
    STDEV(Recovered) AS Stdev_Recovered
FROM CoronaReport
GROUP BY FORMAT(Date, 'yyyy-MM');
```

100 %

Results | Messages

	Month	Total_Recovered	Avg_Recovered	Var_Recovered	Stdev_Recovered
1	2020-05	1519547	318.296397151236	1978620.87525624	1406.63459194499
2	2021-05	19131842	4007.5077503142	755333749.969666	27483.3358595653
3	2021-03	7888013	1652.28592375367	34904703.0577654	5908.0202316652
4	2020-09	6647749	1438.90670995671	57035911.8793661	7552.21238309451
5	2021-02	6719785	1558.39169758813	24433077.9029048	4942.98269296028
6	2020-04	792987	171.642207792208	770059.711532687	877.530461883054
7	2021-06	5544438	2769.44955044955	233150866.36452	15269.2785148651
8	2020-10	6782150	1420.64306661081	73747150.1663075	8587.61609332342
9	2020-08	6202833	1299.29472140762	40178838.3767708	6338.67796758684
10	2020-11	9172292	1985.34458874459	50738601.2546903	7123.10334437809

EXPLORATORY DATA ANALYSIS

Q14. Find Country having highest number of the Confirmed case

```sql
-- Q14. Find Country having highest number of the Confirmed case
SELECT TOP 1
    Country_Region,
    SUM(Confirmed) AS Total_Confirmed
FROM CoronaReport
GROUP BY Country_Region
ORDER BY Total_Confirmed DESC;
```

100 %

Results | Messages

	Country_Region	Total_Confirmed
1	US	33461982

EXPLORATORY DATA ANALYSIS

Q15. Find Country having lowest number of the death case

```sql
-- Q15. Find Country having lowest number of the death case
SELECT TOP 4
    Country_Region,
    SUM(Deaths) AS Lowest_Deaths
FROM CoronaReport
GROUP BY Country_Region
ORDER BY Lowest_Deaths;
```

100 %

Results | Messages

	Country_Region	Lowest_Deaths
1	Marshall Islands	0
2	Samoa	0
3	Kiribati	0
4	Dominica	0

EXPLORATORY DATA ANALYSIS

Q16. Find top 5 countries having highest recovered case

```sql
-- Q16. Find top 5 countries having highest recovered case
SELECT TOP 5
    Country_Region,
    SUM(Recovered) AS Highest_Recovered
FROM CoronaReport
GROUP BY Country_Region
ORDER BY Highest_Recovered DESC;
```

100 %

Results | Messages

	Country_Region	Highest_Recovered
1	India	28089649
2	Brazil	15400169
3	US	6303715
4	Turkey	5202251
5	Russia	4745756

INSIGHTS

- The data records COVID-19 cases from January 2020 to May 2021.
- The United States has the highest number of confirmed cases, with 33.5 million people infected with the virus. India recorded at most 29.6 million confirmed cases, and Brazil had a total of 17.4 million confirmed cases.
- The top five countries with the highest recovered cases are India, Brazil, United States, Turkey, and Russia.
- The top three highest average confirmed cases occurred in April 2021 with a total of 4,699, December 2020 with a total of 4,050, and May 2021 with a total of 4,005.
- The top three highest average recovered cases occurred in May 2021 with a total of 4,007, April 2021 with a total of 3,074, and June 2021 with a total of 2,769.
- The year 2020 had the maximum number of confirmed cases and recovered cases, whereas 2021 recorded more death cases.
- The highest number of deaths were reported in January 2021.
- Marshall Islands, Samoa, Dominica, and Kiribati had the lowest count of death cases with a total of 0 respectively.

Thanks!