

# Corona Virus Analysis

**Under Mentorship**

**Data-Driven Insights on the Spread and Impact of  
the Coronavirus Pandemic**

**Presented by: Bushra Khanam**



# Introduction

- The coronavirus pandemic has significantly impacted public health globally.
- This project analyzes coronavirus data to derive meaningful insights.
- The dataset includes information on confirmed cases, deaths, and recoveries by date and location.



# Dataset Overview

## Columns:

- **Province: Geographic subdivision within a country/region.**
- **Country/Region: Geographic entity where data is recorded.**
- **Latitude & Longitude: Position on Earth's surface.**
- **Date: Recorded date of data.**
- **Confirmed, Deaths, Recovered: Metrics related to coronavirus cases.**



# Checking for Null Values

## Checking for Null Values

```
SELECT  
    SUM(CASE WHEN Province IS NULL THEN 1 ELSE 0 END) AS Province_Null_Count,  
    SUM(CASE WHEN `Country/Region` IS NULL THEN 1 ELSE 0 END) AS CountryRegion_Null_Count,  
    SUM(CASE WHEN Latitude IS NULL THEN 1 ELSE 0 END) AS Latitude_Null_Count,  
    SUM(CASE WHEN Longitude IS NULL THEN 1 ELSE 0 END) AS Longitude_Null_Count,  
    SUM(CASE WHEN Date IS NULL THEN 1 ELSE 0 END) AS Date_Null_Count,  
    SUM(CASE WHEN Confirmed IS NULL THEN 1 ELSE 0 END) AS Confirmed_Null_Count,  
    SUM(CASE WHEN Deaths IS NULL THEN 1 ELSE 0 END) AS Deaths_Null_Count,  
    SUM(CASE WHEN Recovered IS NULL THEN 1 ELSE 0 END) AS Recovered_Null_Count  
FROM corona_virus;
```

**Result:**

**No null values were found.**

# Dataset Time Range

**Query:**

```
SELECT  
    MIN(Date) AS Start_Date,  
    MAX(Date) AS End_Date  
FROM corona_virus;
```

**Result:**

- **Start Date: 2020-01-22**
- **End Date: 2021-06-13**

# Total Rows and Number of Months

## Total Rows

Query:

```
SELECT COUNT(*) AS  
Total_Rows FROM  
corona_virus;
```

Result:

- Total Rows: 78,386

## Number Of Months

Query:

```
SELECT COUNT(DISTINCT  
DATE_FORMAT(Date, '%Y-%m'))  
AS Total_Months  
FROM corona_virus;
```

Result:

- 17 months

# Monthly Averages (Confirmed, Deaths, Recovered)

Query:

```
SELECT  
    DATE_FORMAT(Date, '%Y-%m') AS Month,  
    AVG(Confirmed) AS Avg_Confirmed,  
    AVG(Deaths) AS Avg_Deaths,  
    AVGRecovered) AS Avg_Recovered  
FROM corona_virus  
GROUP BY Month  
ORDER BY Month;
```

Result:

	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

# Dataset Time Range

## Query:

```
-- Most frequent confirmed cases per month
WITH Confirmed_Frequencies AS (
    SELECT
        DATE_FORMAT(Date, '%Y-%m') AS Month,
        Confirmed,
        COUNT(*) AS Frequency
    FROM corona_virus
    GROUP BY DATE_FORMAT(Date, '%Y-%m'), Confirmed
),
Max_Confirmed_Frequencies AS (
    SELECT
        Month,
        MAX(Frequency) AS Max_Frequency
    FROM Confirmed_Frequencies
    GROUP BY Month
)
SELECT
    c.Month,
    c.Confirmed AS Most_Frequent_Confirmed
FROM Confirmed_Frequencies c
JOIN Max_Confirmed_Frequencies m
ON c.Month = m.Month AND c.Frequency = m.Max_Frequency;

-- Most frequent deaths per month
WITH Deaths_Frequencies AS (
    SELECT
        DATE_FORMAT(Date, '%Y-%m') AS Month,
        Deaths,
        COUNT(*) AS Frequency
    FROM corona_virus
    GROUP BY DATE_FORMAT(Date, '%Y-%m'), Deaths
),
```

```
Max_Deaths_Frequencies AS (
    SELECT
        Month,
        MAX(Frequency) AS Max_Frequency
    FROM Deaths_Frequencies
    GROUP BY Month
)
SELECT
    d.Month,
    d.Deaths AS Most_Frequent_Deaths
FROM Deaths_Frequencies d
JOIN Max_Deaths_Frequencies m
ON d.Month = m.Month AND d.Frequency = m.Max_Frequency;

-- Most frequent recovered cases per month
WITH Recovered_Frequencies AS (
    SELECT
        DATE_FORMAT(Date, '%Y-%m') AS Month,
        Recovered,
        COUNT(*) AS Frequency
    FROM corona_virus
    GROUP BY DATE_FORMAT(Date, '%Y-%m'), Recovered
),
Max_Recovered_Frequencies AS (
    SELECT
        Month,
        MAX(Frequency) AS Max_Frequency
    FROM Recovered_Frequencies
    GROUP BY Month
)
SELECT
    r.Month,
    r.Recovered AS Most_Frequent_Recovered
FROM Recovered_Frequencies r
JOIN Max_Recovered_Frequencies m
ON r.Month = m.Month AND r.Frequency = m.Max_Frequency;
```

## Result:

	Month	Most_Frequent_Confirmed
▶	2020-01	0
	2020-02	0
	2020-03	0
	2020-04	0
	2020-05	0
	2020-06	0
	2020-07	0
	2020-08	0
	2020-09	0
	2020-10	0
	2020-11	0
	2020-12	0
	2021-01	0
	2021-02	0
	2021-03	0
	2021-04	0
	2021-05	0

# Yearly Minimum and Maximum Values

## Query: MINIMUM

```
SELECT  
    YEAR(Date) AS Year,  
    MIN(Confirmed) AS Min_Confirmed,  
    MIN(Deaths) AS Min_Deaths,  
    MINRecovered() AS Min_Recovered  
FROM corona_virus  
GROUP BY Year  
ORDER BY Year;
```

Result:

	Year	Min_Confirmed	Min_Deaths	Min_Recovered
▶	2020	0	0	0
▶	2021	0	0	0

## Query: MAXIMUM

```
SELECTYEAR(Date) AS Year,  
    MAX(Confirmed) AS Max_Confirmed,  
    MAX(Deaths) AS Max_Deaths,  
    MAXRecovered() AS Max_Recovered  
FROM corona_virus GROUP BY Year;
```

	Year	Max_Confirmed	Max_Deaths	Max_Recovered
▶	2020	823225	3752	1123456
▶	2021	414188	7374	422436

Result:

# Total Cases Per Month (Confirmed, Deaths, Recovered)

**Query:**

```
SELECT  
    DATE_FORMAT(Date, '%Y-%m') AS Month,  
    SUM(Confirmed) AS Total_Confirmed,  
    SUM(Deaths) AS Total_Deaths,  
    SUMRecovered) AS Total_Recovered  
FROM corona_virus  
GROUP BY Month;
```

**Result:**

	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

# Spread of Coronavirus (Confirmed Cases)

**Query:**

```
SELECT  
    SUM(Confirmed) AS Total_Confirmed,  
    AVG(Confirmed) AS Avg_Confirmed,  
    VARIANCE(Confirmed) AS Var_Confirmed,  
    STDDEV(Confirmed) AS Stddev_Confirmed  
FROM corona_virus;
```

**Result:**

	Total_Confirmed	Avg_Confirmed	Var_Confirmed	Stddev_Confirmed
▶	169065144	2156.8283	157288925.07796532	12541.488152446875

# Spread of Coronavirus (Death Cases Per Month)

Query:

```
SELECT  
    DATE_FORMAT(Date, '%Y-%m') AS Month,  
    SUM(Deaths) AS Total_Deaths,  
    AVG(Deaths) AS Avg_Deaths,  
    VARIANCE(Deaths) AS Var_Deaths,  
    STDDEV(Deaths) AS Stddev_Deaths  
FROM corona_virus  
GROUP BY Month  
ORDER BY Month;
```

Result:

Month	Total_Deaths	Avg_Deaths	Var_Deaths	Stddev_Deaths
2020-01	190	0.1234	4.245817169843138	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

# Countries with the Highest and Lowest Cases

Query:

Lowest Cases

```
SELECT  
    `Country/Region`,  
    SUM(Deaths) AS Total_Deaths  
FROM corona_virus  
GROUP BY `Country/Region`  
ORDER BY Total_Deaths ASC  
LIMIT 1;
```

Result Grid		
	Country/Region	Total_Deaths
▶	Dominica	0

Result:



Highest Cases

```
SELECT  
    `Country/Region`,  
    SUM(Recovered) AS Total_Recovered  
FROM corona_virus  
GROUP BY `Country/Region`  
ORDER BY Total_Recovered DESC  
LIMIT 5;
```

Country/Region	Total_Confirmed
US	33461982

Result:



# Top 5 Countries with Highest Recovered Cases

Query:

Lowest Cases

```
SELECT  
    `Country/Region`,  
    SUM(Recovered) AS Total_Recovered  
FROM corona_virus  
GROUP BY `Country/Region`  
ORDER BY Total_Recovered DESC  
LIMIT 5;
```

Result:



Country/Region	Total_Recovered
India	28089649
Brazil	15400169
US	6303715
Turkey	5202251
Russia	4745756

# Conclusion

- The analysis provides a comprehensive view of the coronavirus pandemic's impact.
- Key insights include the progression of confirmed cases, deaths, and recoveries over time.
- The data highlights the varying impact across different regions and time periods.



# Conclusion

- The analysis provides a comprehensive view of the coronavirus pandemic's impact.
- Key insights include the progression of confirmed cases, deaths, and recoveries over time.
- The data highlights the varying impact across different regions and time periods.





Thank  
you

By: Bushra Khanam