Corona Virus Analysis using SQL.

Ashley Dcunha

**Q1: Null Value Counts**

SELECT

SUM(IF(country IS NULL, 1, 0)) AS country\_null\_count,

SUM(IF(Province IS NULL, 1, 0)) AS province\_null\_count,

SUM(IF(latitude IS NULL, 1, 0)) AS latitude\_null\_count,

SUM(IF(longitude IS NULL, 1, 0)) AS longitude\_null\_count,

SUM(IF(date\_of IS NULL, 1, 0)) AS date\_of\_null\_count,

SUM(IF(recovered IS NULL, 1, 0)) AS recovered\_null\_count,

SUM(IF(Deaths IS NULL, 1, 0)) AS deaths\_null\_count,

SUM(IF(confirmed IS NULL, 1, 0)) AS confirmed\_null\_count

FROM

corona;

A screenshot of a computer

Description automatically generated

**Purpose: This query counts the number of null values for each column in the corona table. It helps identify missing data in each of these key fields.**

**Q2. Fill Null Values**

UPDATE corona SET country = COALESCE(country, '0') WHERE country IS NULL;

UPDATE corona SET Province = COALESCE(Province, '0') WHERE Province IS NULL;

UPDATE corona SET latitude = COALESCE(latitude, 0) WHERE latitude IS NULL;

UPDATE corona SET longitude = COALESCE(longitude, 0) WHERE longitude IS NULL;

UPDATE corona SET date\_of = COALESCE(date\_of, '0') WHERE date\_of IS NULL;

UPDATE corona SET recovered = COALESCE(recovered, 0) WHERE recovered IS NULL;

UPDATE corona SET Deaths = COALESCE(Deaths, 0) WHERE Deaths IS NULL;

UPDATE corona SET confirmed = COALESCE(confirmed, 0) WHERE confirmed IS NULL;

**Purpose: These queries update the corona table to replace null values with default values. This ensures that the dataset does not contain nulls, which can be crucial for subsequent analysis.**

**Q3. Total Row Count**

SELECT COUNT(\*) AS total\_rows FROM corona;

A screenshot of a computer

Description automatically generated

**Purpose: This query returns the total number of rows in the corona table. It gives an overview of the dataset's size.**

**Q4. Date Range**

SELECT

MIN(date\_of) AS start\_date,

MAX(date\_of) AS end\_date

FROM

corona;

A screenshot of a computer

Description automatically generated

**Purpose: This query finds the earliest and latest dates in the dataset, indicating the time span covered by the data.**

**Q5. Monthly Case Count**

SELECT DATE\_FORMAT(date\_of, '%Y-%m') AS month,COUNT(\*) AS month\_count

FROM corona

GROUP BY month

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This query counts the number of entries per month, providing a monthly breakdown of the data.**

**Q6. Monthly Averages**

SELECT DATE\_FORMAT(date\_of, '%Y-%m') AS month,AVG(confirmed) AS avg\_confirmed,AVG(deaths) AS avg\_deaths,

AVG(recovered) AS avg\_recovered

FROM corona

GROUP BY month

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This query calculates the average number of confirmed cases, deaths, and recoveries per month.**

**Q7. Most Frequent Monthly Values**

SELECT month,

(

SELECT confirmed FROM corona c2 WHERE DATE\_FORMAT(c2.date\_of, '%Y-%m') = month

GROUP BY confirmed ORDER BY COUNT(\*) DESC LIMIT 1

) AS most\_frequent\_confirmed,

(

SELECT deaths FROM corona c2 WHERE DATE\_FORMAT(c2.date\_of, '%Y-%m') = month

GROUP BY deaths ORDER BY COUNT(\*) DESC LIMIT 1

) AS most\_frequent\_deaths,

(

SELECT recovered FROM corona c2 WHERE DATE\_FORMAT(c2.date\_of, '%Y-%m') = month

GROUP BY recovered ORDER BY COUNT(\*) DESC LIMIT 1

) AS most\_frequent\_recovered

FROM (

SELECT DISTINCT DATE\_FORMAT(date\_of, '%Y-%m') AS month FROM corona

) AS months

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This complex query identifies the most frequently occurring values for confirmed cases, deaths, and recoveries each month.**

**Q8. Minimum Annual Values**

SELECT YEAR(date\_of) AS year,MIN(confirmed) AS min\_confirmed,MIN(deaths) AS min\_deaths,

MIN(recovered) AS min\_recovered

FROM corona

GROUP BY YEAR(date\_of)

ORDER BY YEAR(date\_of);

A screenshot of a computer

Description automatically generated

**Purpose: This query finds the minimum values for confirmed cases, deaths, and recoveries each year.**

**Q9. Maximum Annual Values**

SELECT YEAR(date\_of) AS year,Max(confirmed) AS max\_confirmed,Max(deaths) AS max\_deaths,

Max(recovered) AS max\_recovered

FROM corona

GROUP BY YEAR(date\_of)

ORDER BY YEAR(date\_of);

A screenshot of a computer

Description automatically generated

**Purpose: This query finds the maximum values for confirmed cases, deaths, and recoveries each year.**

**Q10. Monthly Totals**

SELECT DATE\_FORMAT(date\_of, '%Y-%m') AS month,SUM(confirmed) AS total\_confirmed,SUM(deaths) AS total\_deaths,

SUM(recovered) AS total\_recovered

FROM corona

GROUP BY DATE\_FORMAT(date\_of, '%Y-%m')

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This query calculates the total number of confirmed cases, deaths, and recoveries per month.**

**Q11. Statistical Summary for Confirmed Cases**

SELECT

SUM(confirmed) AS total\_confirmed\_cases,

AVG(confirmed) AS average\_confirmed\_cases,

VARIANCE(confirmed) AS variance\_confirmed\_cases,

STDDEV(confirmed) AS stddev\_confirmed\_cases

from corona;

A screenshot of a computer

Description automatically generated

**Purpose: This query provides a statistical summary (total, average, variance, and standard deviation) of confirmed cases.**

**Q12. Monthly Death Statistics**

SELECT

DATE\_FORMAT(date\_of, '%Y-%m') AS month,ROUND(SUM(deaths), 2) AS total\_death\_cases,ROUND(AVG(deaths), 2) AS average\_death\_cases,

ROUND(VARIANCE(deaths), 2) AS variance\_death\_cases,ROUND(STDDEV(deaths), 2) AS stddev\_death\_cases

FROM corona

GROUP BY DATE\_FORMAT(date\_of, '%Y-%m')

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This query calculates the monthly total, average, variance, and standard deviation for death cases, rounding the results to two decimal places.**

**Q13. Monthly Recovered Statistics**

SELECT

DATE\_FORMAT(date\_of, '%Y-%m') AS month,ROUND(SUM(recovered), 2) AS total\_recovered\_cases,

ROUND(AVG(recovered), 2) AS average\_recovered\_cases,ROUND(VARIANCE(recovered), 2) AS variance\_recovered\_cases,

ROUND(STDDEV(recovered), 2) AS stddev\_recovered\_cases

FROM corona

GROUP BY DATE\_FORMAT(date\_of, '%Y-%m')

ORDER BY month;

A screenshot of a computer

Description automatically generated

**Purpose: This query calculates the monthly total, average, variance, and standard deviation for recovered cases, rounding the results to two decimal places.**

**Q14. Country with Highest Confirmed Cases**

SELECT country,Sum(confirmed) AS highest\_confirmed\_cases

FROM corona

GROUP BY country

ORDER BY highest\_confirmed\_cases DESC

LIMIT 1;

A screenshot of a computer

Description automatically generated

**Purpose: This query identifies the country with the highest total number of confirmed cases.**

**Q15. Top 4 Countries by Death Cases**

SELECT country,SUM(deaths) AS total\_death\_cases

FROM corona

GROUP BY country

ORDER BY total\_death\_cases

LIMIT 4;

A screenshot of a computer

Description automatically generated

**Purpose: This query lists the top four countries with the highest total number of death cases.**

**Q16. Top 5 Countries by Recovered Cases**

SELECT country,SUM(recovered) AS total\_recovered\_cases

FROM corona

GROUP BY country

ORDER BY total\_recovered\_cases DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

**Purpose: This query lists the top five countries with the highest total number of recovered cases.**