COVID-19 ANALYSIS USING SQL

Data Modeling and Insights
with Power BI & PostgreSQL

INTRODUCTION

Overview

- The Covid-19 pandemic has had a profound impact globally, affecting millions of lives and economies.
- Data analysis is crucial in understanding the spread, impact, and control measures of the virus.
- This project focuses on analyzing Covid-19 data using SQL, with visualizations and data modeling in Power BI.

PROJECT OBJECTIVES

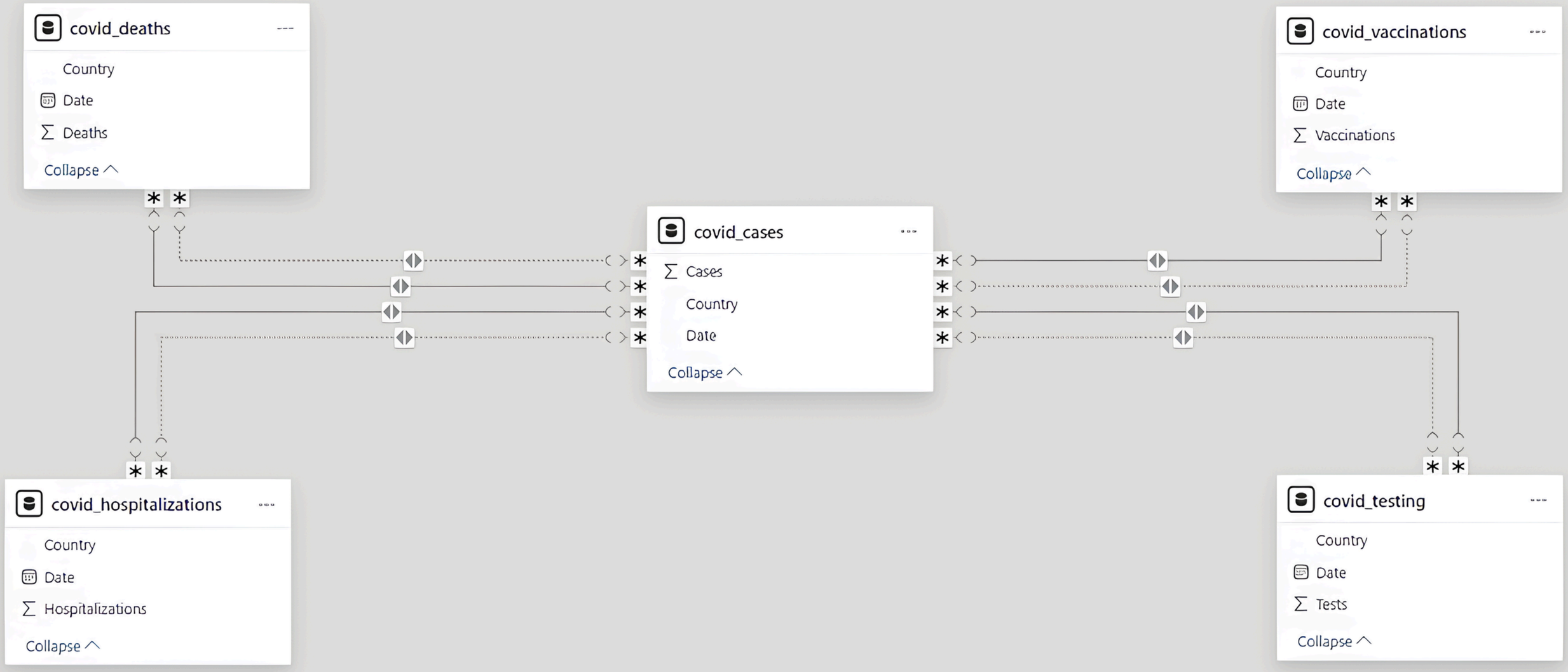
Main Objectives

- **Data Integration:** Combine multiple Covid-19 datasets for comprehensive analysis.
- **Data Modeling:** Use Power BI to automatically generate a data model from the integrated datasets.
- **SQL Querying:** Formulate and execute SQL queries to answer critical questions related to the pandemic.
- **Insights Generation:** Derive insights from the data to better understand Covid-19 trends and impacts.

DATASETS OVERVIEW

- **Covid-19 Cases:** Daily reported cases by country.
- **Covid-19 Deaths:** Daily reported deaths by country.
- **Covid-19 Vaccinations:** Vaccination data by country.
- **Covid-19 Testing:** Testing data by country.
- **Covid-19 Hospitalizations:** Hospitalization records by country.

DATA MODEL IN POWER BI



SQL QUERYING

Querying Strategy

- Divided into three levels:

1. Easy
2. Intermediate
3. Advanced

- Queries are designed to answer specific questions related to the Covid-19 pandemic.

Query # 1

What is the total number of Covid-19 cases reported in 2020?

```
SELECT SUM(Cases) AS Total_Cases_2020
FROM covid_cases
WHERE Date BETWEEN '2020-01-01' AND '2020-12-31';
```

total_cases_2020 bigint
18343396

Query # 2

Which country had the highest number of Covid-19 cases on a single day?

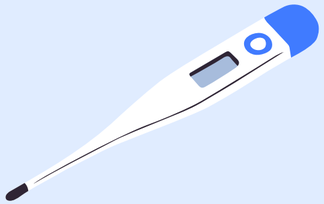
```
SELECT Country, Date, MAX(Cases) AS Max_Cases
FROM covid_cases
GROUP BY Country, Date
ORDER BY Max_Cases DESC
LIMIT 1;
```

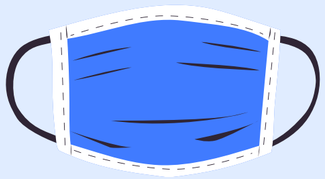
country character varying (50)	date date	max_cases integer
Italy	2021-12-07	9998

Query # 3

What is the average number of deaths per day in the month of April 2021?

```sql
SELECT AVG(Deaths) AS Avg_Deaths_April_2021
FROM covid_deaths
WHERE Date BETWEEN '2021-04-01' AND '2021-04-30';
```





avg_deaths_april_2021
numeric
252.4333333333333333

Query # 4

How many countries reported more than 1,000,000 cases by December 2021?

```sql
SELECT COUNT(DISTINCT Country) AS Countries_With_1M_Cases
FROM (
    SELECT Country, SUM(Cases) AS Total_Cases
    FROM covid_cases
    WHERE Date <= '2021-12-31'
    GROUP BY Country
    HAVING SUM(Cases) > 1000000
) AS Subquery;
```

countries_with_1m_cases bigint
10

Query # 5

Which country had the highest vaccination rate by the end of 2021?

```sql
SELECT Country, SUM(Vaccinations) AS Total_Vaccinations
FROM covid_vaccinations
WHERE Date <= '2021-12-31'
GROUP BY Country
ORDER BY Total_Vaccinations DESC
LIMIT 1;
```

country character varying (50)	total_vaccinations bigint
India	7613626

INTERMEDIATE LEVEL QUERIES

Query # 6

Calculate the average number of tests conducted per day across all countries in 2021.

```sql
SELECT AVG(Tests) AS Avg_Tests_Per_Day
FROM covid_testing
WHERE Date BETWEEN '2021-01-01' AND '2021-12-31';
```

avg_tests_per_day numeric
15046.623013698630

Query # 7

Identify the top 5 countries with the highest hospitalization rates in 2020.

```sql
SELECT Country, SUM(Hospitalizations) AS Total_Hospitalizations
FROM covid_hospitalizations
WHERE Date BETWEEN '2020-01-01' AND '2020-12-31'
GROUP BY Country
ORDER BY Total_Hospitalizations DESC
LIMIT 5;
```

country character varying (50)	total_hospitalizations bigint
UK	385324
India	378822
Argentina	370959
Germany	368431
USA	367898

Query # 8

Find the country with the lowest death rate among those with more than 500,000 cases.

```sql
SELECT cc.Country, SUM(cd.Deaths) / SUM(cc.Cases) AS Death_Rate
FROM covid_cases cc
JOIN covid_deaths cd ON cc.Country = cd.Country AND cc.Date = cd.Date
GROUP BY cc.Country
HAVING SUM(cc.Cases) > 500000
ORDER BY Death_Rate ASC
LIMIT 1;
```

country character varying (50)	death_rate bigint
Argentina	0

ADVANCED LEVEL QUERIES

Query # 9

Analyze the correlation between the number of tests conducted and the number of positive cases in 2021.

```sql
SELECT
    CORR(Tests, Cases) AS Correlation
FROM
    covid_testing ct
JOIN
    covid_cases cc ON ct.Country = cc.Country AND ct.Date = cc.Date
WHERE
    ct.Date BETWEEN '2021-01-01' AND '2021-12-31';
```

correlation double precision
-0.012929605661204472

Query # 10

Determine the month with the highest increase in vaccination rates compared to the previous month for each country.

```sql
SELECT Country, Month, MAX(Vaccination_Change) AS Max_Vaccination_Increase
FROM (
    SELECT
        Country,
        DATE_TRUNC('month', Date) AS Month,
        SUM(Vaccinations) - LAG(SUM(Vaccinations))
        OVER (PARTITION BY Country ORDER BY DATE_TRUNC('month', Date))
        AS Vaccination_Change
    FROM
        covid_vaccinations
    GROUP BY
        Country, Month
) AS Subquery
GROUP BY Country, Month
ORDER BY Max_Vaccination_Increase DESC;
```

RESULT of QUERY # 10

Italy	2020-07-01 00:00:00+05	146601
India	2020-03-01 00:00:00+05	129883
UK	2021-05-01 00:00:00+05	120906
Spain	2020-10-01 00:00:00+05	93809
France	2021-12-01 00:00:00+05	89816
UK	2021-02-01 00:00:00+05	88593
Argentina	2020-08-01 00:00:00+05	87728
Italy	2021-08-01 00:00:00+05	87626
USA	2020-03-01 00:00:00+05	86032
Argentina	2021-05-01 00:00:00+05	84500
Brazil	2020-08-01 00:00:00+05	83618
Spain	2021-12-01 00:00:00+05	82358
Italy	2020-05-01 00:00:00+05	81924

Query # 11

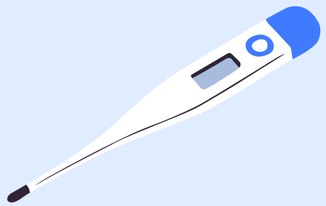
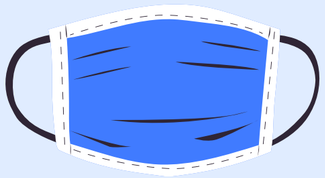
Analyze the relationship between Covid-19 testing rates and case detection rates. Identify the top 5 countries with the highest ratio of detected cases to tests conducted, indicating possible underreporting or low testing efficiency.

```
WITH Case_Test_Ratio AS (
    SELECT
ct.Country, SUM(ct.Cases) AS Total_Cases, SUM(tt.Tests) AS Total_Tests,
CASE WHEN SUM(tt.Tests) > 0 THEN SUM(ct.Cases) * 1.0 / SUM(tt.Tests)
ELSE 0 END AS Case_Test_Ratio
FROM covid_cases ct JOIN
covid_testing tt ON ct.Country = tt.Country AND ct.Date = tt.Date
GROUP BY ct.Country)
SELECT
    Country, Total_Cases, Total_Tests, Case_Test_Ratio
FROM Case_Test_Ratio
ORDER BY Case_Test_Ratio DESC
LIMIT 5;
```

RESULT of QUERY # 11

country character varying (50)	total_cases bigint	total_tests bigint	case_test_ratio numeric
UK	3775018	10766811	0.35061616666253359514
USA	3769473	10807937	0.34876896488201217309
Germany	3756998	11072327	0.33931422003703467212
Spain	3710368	11030850	0.33636283695272803093
France	3588871	10737173	0.33424729209448334306




THANK YOU !

+92 3176063654

waleedilyas99@gmail.com

linkedin.com/in/waleed-ilyas-664839213