

# COVID PANDEMIC ANALYSIS

## DATA EXPLORATION USING SQL

By Anurag Srivastava

# Contents

## 01 Introduction

This analysis uses SQL to explore the COVID-19 dataset, providing insights into the pandemic's global impact on various countries and regions.

## 02 Data Description

The data description section outlines the dataset being used and its source.

## 03 Data Exploration

In this section, insights will be derived through SQL queries.

## 04 Conclusion

Conclusion and Insights on COVID-19 Data Analysis



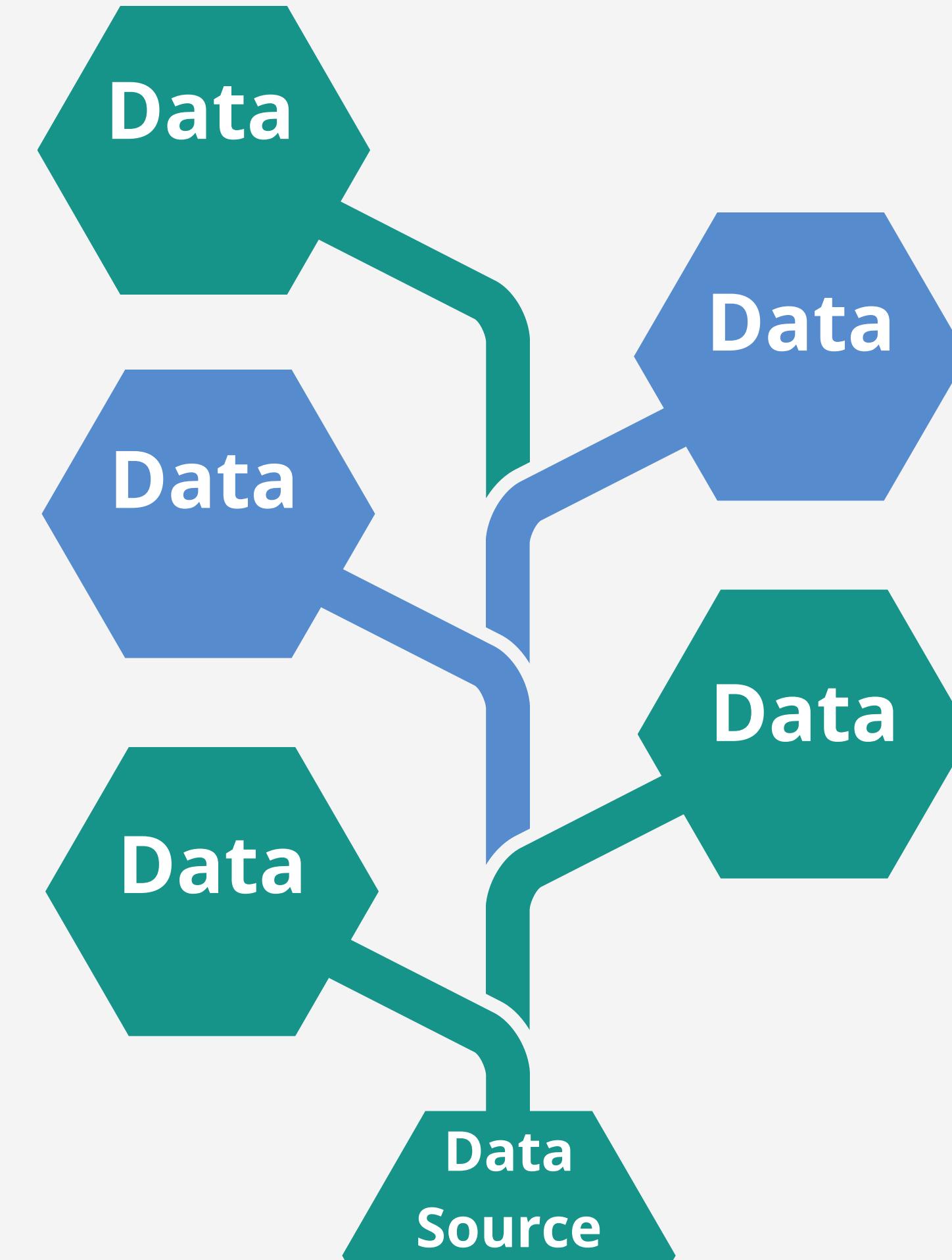
# Introduction

- The COVID-19 pandemic, caused by the novel coronavirus, has had a massive impact on the world. The COVID-19 dataset provides us with valuable data that can be used to understand the spread and impact of the pandemic. In this analysis, we use SQL to analyze the COVID-19 dataset and gain insights into the pandemic's effect on different countries and regions.



# Data Description

The COVID-19 dataset used in this analysis is from Our World in Data, available under the Creative Commons BY license. It includes daily COVID-19 cases, deaths, and vaccination data for various countries. The dataset is divided into two tables: CovidDeath, containing details on daily cases, deaths, total cases, and population, and CovidVaccinations, containing daily vaccination data and people vaccinated. It can be accessed at [this link](#).





# ANALYSIS

# Query

## Query 1: Select Data that we are going to use.

```
-- Select Data that we are going to be using
```

```
Select Location, date, total_cases, new_cases, total_deaths, population  
From PortfolioProject..CovidDeaths
```

50 %

Results Messages

	Location	date	total_cases	new_cases	total_deaths	population
1	Grenada	2020-07-17 00:00:00.000	23	0	NULL	112519
2	Grenada	2020-07-18 00:00:00.000	23	0	NULL	112519
3	Grenada	2020-07-19 00:00:00.000	23	0	NULL	112519
4	Grenada	2020-07-20 00:00:00.000	23	0	NULL	112519
5	Grenada	2020-07-21 00:00:00.000	23	0	NULL	112519
6	Grenada	2020-07-22 00:00:00.000	23	0	NULL	112519
7	Grenada	2020-07-23 00:00:00.000	23	0	NULL	112519
8	Grenada	2020-07-24 00:00:00.000	23	0	NULL	112519
9	Grenada	2020-07-25 00:00:00.000	23	0	NULL	112519
10	Grenada	2020-07-26 00:00:00.000	23	0	NULL	112519
11	Grenada	2020-07-27 00:00:00.000	23	0	NULL	112519
12	Grenada	2020-07-28 00:00:00.000	23	0	NULL	112519
13	Grenada	2020-07-29 00:00:00.000	23	0	NULL	112519
14	Grenada	2020-07-30 00:00:00.000	24	1	NULL	112519
15	Grenada	2020-07-31 00:00:00.000	24	0	NULL	112519
16	Grenada	2020-08-01 00:00:00.000	24	0	NULL	112519

## Query 2: Looking at Total Cases vs Total Deaths

```
Select Location, date, total_cases, total_deaths, (total_deaths/total_cases)*100 as DeathPercentage  
From PortfolioProject..CovidDeaths  
order by 1,2
```

	Location	date	total_cases	total_deaths	DeathPercentage
25	Afghanistan	2020-03-19 00:00:00.000	25	NULL	NULL
26	Afghanistan	2020-03-20 00:00:00.000	29	NULL	NULL
27	Afghanistan	2020-03-21 00:00:00.000	30	NULL	NULL
28	Afghanistan	2020-03-22 00:00:00.000	34	1	2.94117647058824
29	Afghanistan	2020-03-23 00:00:00.000	41	1	2.4390243902439
30	Afghanistan	2020-03-24 00:00:00.000	43	1	2.32558139534884
31	Afghanistan	2020-03-25 00:00:00.000	76	2	2.63157894736842
32	Afghanistan	2020-03-26 00:00:00.000	80	3	3.75
33	Afghanistan	2020-03-27 00:00:00.000	91	3	3.2967032967033
34	Afghanistan	2020-03-28 00:00:00.000	107	4	3.73831775700935
35	Afghanistan	2020-03-29 00:00:00.000	118	4	3.38983050847458
36	Afghanistan	2020-03-30 00:00:00.000	146	4	2.73972602739726
37	Afghanistan	2020-03-31 00:00:00.000	175	4	2.28571428571429
38	Afghanistan	2020-04-01 00:00:00.000	197	4	2.03045685279188
39	Afghanistan	2020-04-02 00:00:00.000	240	4	1.6666666666666667
40	Afghanistan	2020-04-03 00:00:00.000	275	8	2.9090909090909091
41	Afghanistan	2020-04-04 00:00:00.000	300	10	3.333333333333333

## Query 3 : Looking at Countries with Highest Infection Rate compared to population

```
-- Looking at Countries with Highest Infection Rate compared to Population

Select Location, Population, MAX(total_cases) as HighestInfectionCount, Max((total_cases/population))*100 as
    PercentPopulationInfected
From PortfolioProject..CovidDeaths
--Where location like '%states%'
Group by Location, Population
order by PercentPopulationInfected desc
```

	Location	Population	HighestInfectionCount	PercentPopulationInfected
1	Andorra	77265	13232	17.1254772536077
2	Montenegro	628062	97389	15.5062716738156
3	Czechia	10708982	1630758	15.2279460363273
4	San Marino	33938	5066	14.9272202251164
5	Slovenia	2078932	240292	11.5584348117206
6	Luxembourg	625976	67205	10.7360346083556
7	Bahrain	1701583	176934	10.3981997939566
8	Serbia	6804596	689557	10.1336949320724
9	United States	331002647	32346971	9.77242064169958
10	Israel	8655541	838481	9.68721654718059
11	Sweden	10099270	973604	9.64034034143062
12	Estonia	1326539	122019	9.19829722307448
13	Lithuania	2722291	247269	9.0831215325621
14	Netherlands	17134873	1522973	8.88814874787808
15	Belgium	11589616	990229	8.54410534395618
16	Panama	4314768	364576	8.44949253354989
17	France	68147687	5677835	8.33166208561121

## Query 4 : Showing Countries with Highest Death Count Per Population

```
-- Showing Countries with Highest Death Count per Population

Select Location, MAX(Total_deaths) as TotalDeathCount
From PortfolioProject..CovidDeaths
--Where location like '%states%'
Group by Location
order by TotalDeathCount desc
```

	Location	TotalDeathCount
1	Austria	9997
2	Belgium	9996
3	Iran	9996
4	Egypt	9994
5	France	99936
6	Tunisia	9993
7	Pakistan	9992
8	Dominican Republic	999
9	Montenegro	999
10	Switzerland	9988
11	Europe	998761
12	Bangladesh	9987
13	World	998635
14	Netherlands	9985
15	Bolivia	9985
16	Africa	99814
17	Russia	99800

## Query 5 : Let's Break things down by Continent

```
-- LET'S BREAK THINGS DOWN BY CONTINENT

Select continent, MAX(cast(Total_deaths as int)) as TotalDeathCount
From PortfolioProject..CovidDeaths
--Where location like '%states%'
Where continent is not null
Group by continent
order by TotalDeathCount desc
```

Results    Messages

	continent	TotalDeathCount
1	North America	576232
2	South America	403781
3	Asia	211853
4	Europe	127775
5	Africa	54350
6	Oceania	910

## Query 6 : Global Numbers

```
Select SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths, SUM(cast  
    (new_deaths as int))/SUM(New_Cases)*100 as DeathPercentage  
From PortfolioProject..CovidDeaths  
--Where location like '%states%'  
where continent is not null  
--Group By date  
order by 1,2
```

Results    Messages

	total_cases	total_deaths	DeathPercentage
1	150574977	3180206	2.11204149810363

## Query 7: Look at Total Population vs Vaccinations

```
-- Looking at Total Population vs Vaccinations

Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations
From PortfolioProject..CovidDeaths dea
Join PortfolioProject..CovidVaccinations vac
    On dea.location = vac.location
        and dea.date = vac.date
where dea.continent is not null
order by 1,2,3
```

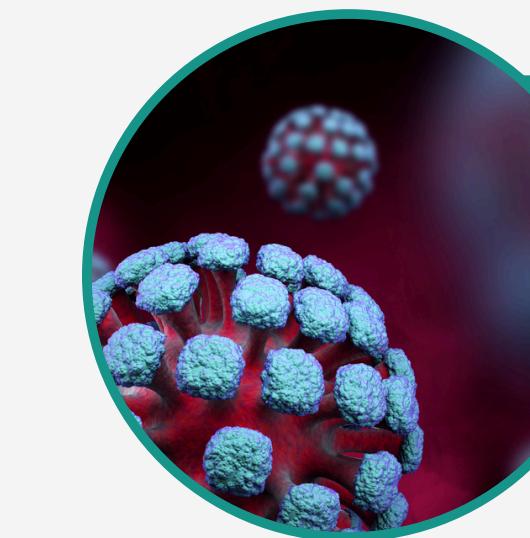
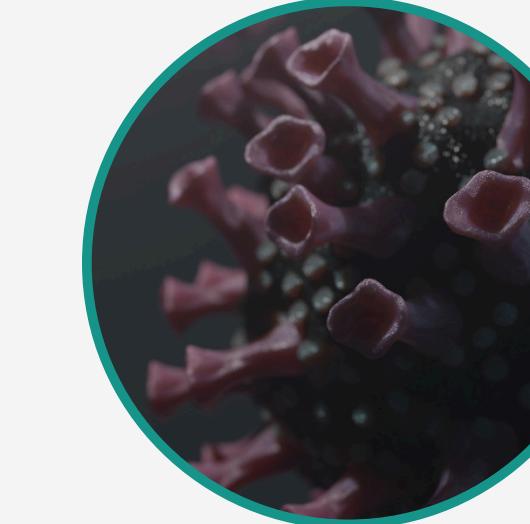
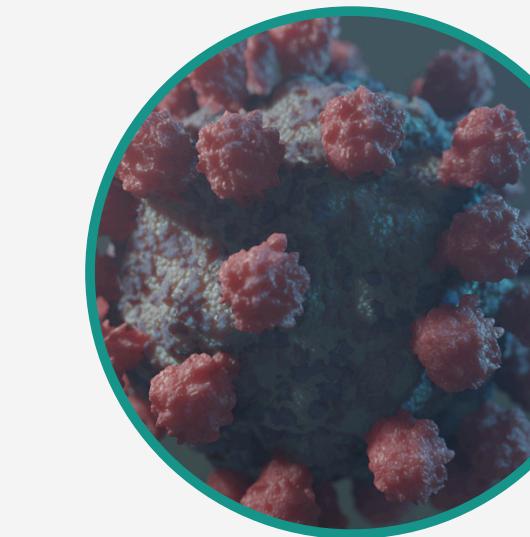
	continent	location	date	population	new_vaccinations
13...	North A...	Canada	2021-04-17 00:00:00.000	37742157	254177
13...	North A...	Canada	2021-04-18 00:00:00.000	37742157	200717
13...	North A...	Canada	2021-04-19 00:00:00.000	37742157	262792
13...	North A...	Canada	2021-04-20 00:00:00.000	37742157	238000
13...	North A...	Canada	2021-04-21 00:00:00.000	37742157	316732
13...	North A...	Canada	2021-04-22 00:00:00.000	37742157	360486
13...	North A...	Canada	2021-04-23 00:00:00.000	37742157	368302
13...	North A...	Canada	2021-04-24 00:00:00.000	37742157	286263
13...	North A...	Canada	2021-04-25 00:00:00.000	37742157	231840
13...	North A...	Canada	2021-04-26 00:00:00.000	37742157	262744
13...	North A...	Canada	2021-04-27 00:00:00.000	37742157	256540
13...	North A...	Canada	2021-04-28 00:00:00.000	37742157	261941
13...	North A...	Canada	2021-04-29 00:00:00.000	37742157	305605
13...	North A...	Canada	2021-04-29 00:00:00.000	37742157	305605

# Conclusions

**The COVID-19 data exploration reveals key insights: countries and regions experienced varying infection and death rates, with some facing higher fatality risks.**

**Vaccination efforts played a crucial role in controlling the spread and reducing mortality in many areas.**

**The global overview highlights the widespread impact of the pandemic, emphasizing the importance of timely interventions and healthcare capacity.**



## Conclusion 01

**Infection Rate vs Population:** Countries with the highest infection rates relative to their populations were identified, highlighting regions where COVID-19 spread more rapidly. This helps to understand which nations were most vulnerable or experienced widespread transmission.

## Conclusion 02

**Death Rate vs Total Cases:** The analysis of total deaths compared to total cases provides insight into the severity of the virus in different regions. This showcases the likelihood of death for those who contracted COVID-19, indicating regions that faced greater fatality risks.

## Conclusion 03

**Vaccination Progress:** By calculating the percentage of the population vaccinated, the analysis tracks vaccination progress across countries and regions. This is crucial for understanding which areas were quicker in their vaccine rollouts and how that might have impacted infection and death rates.

# THANK YOU

Anurag Srivastava

Let's Connect  
on

