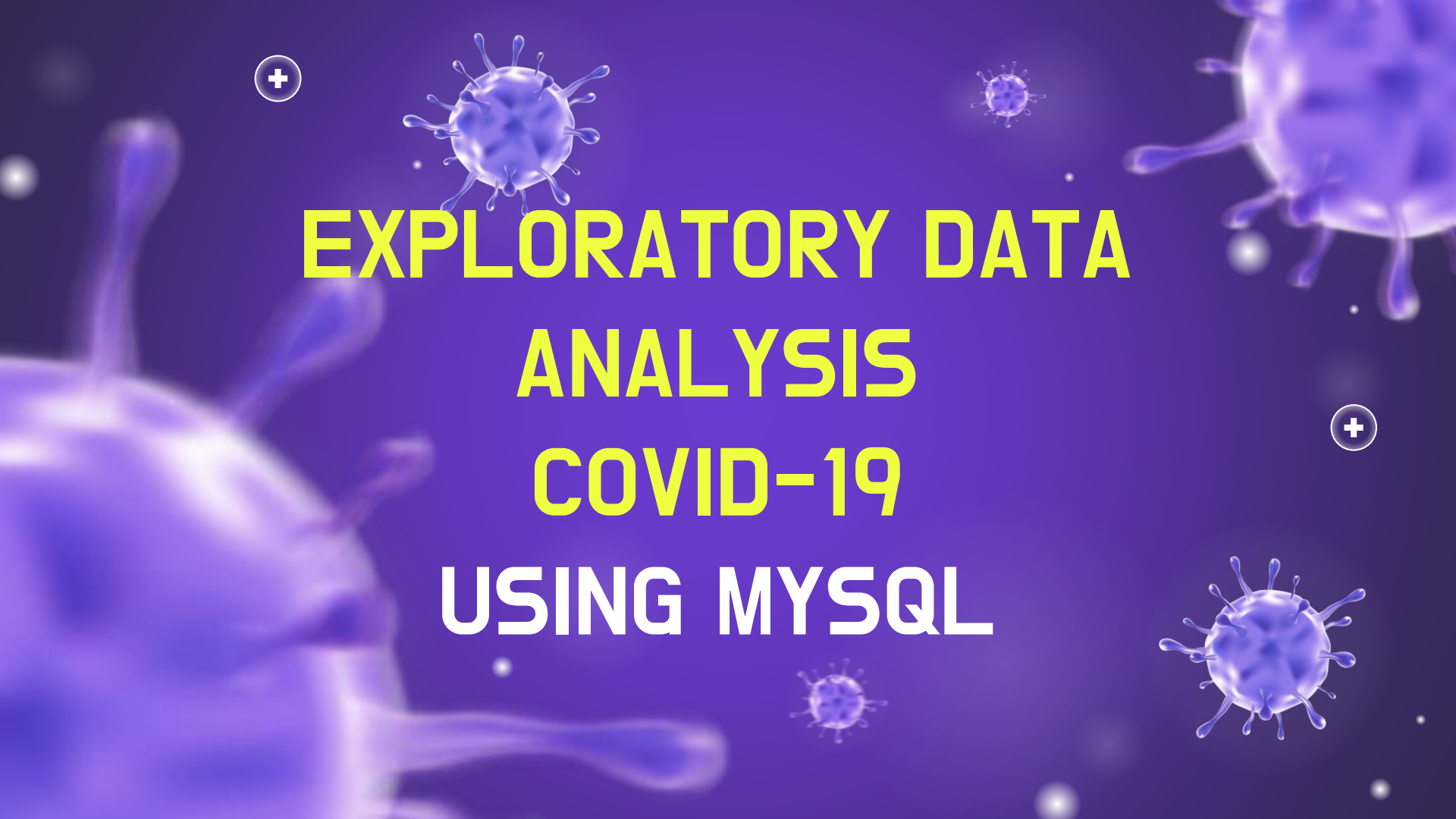
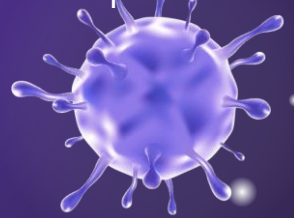

EXPLORATORY DATA
ANALYSIS
COVID-19
USING MYSQL

TABLE OF CONTENTS

Introduction

The COVID-19 pandemic has had a profound impact on global health, economies, and daily life since its emergence in late 2019. As the virus spread rapidly across the world, governments and health organizations collected vast amounts of data to monitor and control the outbreak. This dataset includes detailed information on cases, deaths, and vaccinations dated from 01-01-2020 to 04-07-2021 which provides a valuable resource for understanding the pandemic's progression and impact.



Goal

The goal of this exploratory analysis is to understand the patterns and trends in COVID-19 deaths and vaccination across different countries and regions. By analyzing the data, we aim to derive key insights that can help in understanding the impact of the pandemic and the effectiveness of vaccination campaigns. Additionally, we aim to identify key performance indicators (KPIs) that can be used to monitor the situation effectively.

DATASET DESCRIPTION

The analysis is based on two tables:

COVID_DEATHS

- ISO_CODE : Country Code
- CONTINENT : Continent Name
- LOCATION : Country Name
- DATE : Date of the record
- NEW_CASES : New COVID-19 cases on the recorded date
- TOTAL_CASES: Total COVID-19 cases up to the recorded date
- NEW_DEATHS : New COVID-19 deaths on the recorded date
- TOTAL_DEATHS : Total COVID-19 deaths up to the recorded date
- POPULATION : Population of the country
- NEW_CASES_PER_MILLION : New cases per million people

COVID_DEATHS

- ISO_CODE : Country Code
- CONTINENT : Continent Name
- LOCATION : Country Name
- DATE : Date of the record
- NEW_VACCINATIONS: New COVID-19 cases on the recorded date
- TOTAL_VACCINATIONS: Total COVID-19 cases up to the recorded date
- PEOPLE_VACCINATED: New COVID-19 deaths on the recorded date
- PEOPLE_FULLY_VACCINATED Total COVID-19 deaths up to the recorded date
- POPULATION : Population of the country

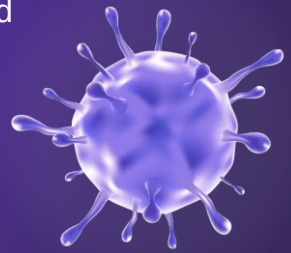
METHADOLOGY

Data Preparation:
•Load the data from the **covid_deaths** and **covid_vaccination** tables.
•Clean the data by handling missing values and ensuring data consistency.

Trend Analysis:
•Use SQL queries to calculate daily and cumulative cases, deaths, and vaccinations.

Correlation Analysis:
•Use SQL queries to join the **covid_deaths** and **covid_vaccination** tables on location and date.
•Calculate correlation coefficients to understand the relationship between vaccination rates and cases/deaths.

Regional Comparison:
•Group data by continent and country.
•Calculate aggregated metrics such as total cases, deaths, vaccinations, and rates per million.



TOTAL ROWS IN THE TABLE

```
SELECT COUNT(*) FROM COVID_DEATHS;
```

TOTAL_ROWS
100181

```
SELECT COUNT(*) FROM COVID_VACCINATION;
```

TOTAL_ROWS
100181

DESCRIBING THE TABLE

DESC COVID_DEATHS;

Field	Type	Null	Key	Default	Extra
iso_code	text	YES		NULL	
continent	text	YES		NULL	
location	text	YES		NULL	
date	date	YES		NULL	
population	bigint	YES		NULL	
total_cases	bigint	YES		NULL	
new_cases	bigint	YES		NULL	
new_cases_smoothed	double	YES		NULL	
total_deaths	bigint	YES		NULL	
new_deaths	bigint	YES		NULL	
new_deaths_smoothed	double	YES		NULL	
total_cases_per_million	double	YES		NULL	
new_cases_per_million	double	YES		NULL	

DESC COVID_VACCINATION;

Field	Type	Null	Key	Default	Extra
iso_code	text	YES		NULL	
continent	varchar(255)	YES		NULL	
location	varchar(255)	YES		NULL	
date	date	YES		NULL	
new_tests	int	YES		NULL	
total_tests	bigint	YES		NULL	
total_tests_per_thousand	double	YES		NULL	
new_tests_per_thousand	double	YES		NULL	
new_tests_smoothed	bigint	YES		NULL	
new_tests_smoothed_per_thousand	double	YES		NULL	
positive_rate	double	YES		NULL	
tests_per_case	double	YES		NULL	

EXPLORING SOME IMPORTANT COLUMN OF THE TABLES

```
SELECT
CONTINENT,LOCATION,`DATE`,TOTAL_CASES,TOTAL_DEATHS,POPULATI
ON FROM COVID_DEATHS;
```

continent	location	date	total_cases	total_deaths	population
Asia	Afghanistan	2020-04-06	368	15	38928341
Asia	Afghanistan	2020-04-07	424	16	38928341
Asia	Afghanistan	2020-04-08	445	16	38928341
Asia	Afghanistan	2020-04-09	485	17	38928341
Asia	Afghanistan	2020-04-10	532	18	38928341
Asia	Afghanistan	2020-04-11	556	18	38928341
Asia	Afghanistan	2020-04-12	608	19	38928341

```
SELECT
CONTINENT,LOCATION,`DATE`,TOTAL_TESTS,TOTAL_VACCINATIONS
FROM COVID_VACCINATION;
```

continent	location	date	total_tests	total_vaccinations
Europe	Albania	2021-04-29	612944	453248
Europe	Albania	2021-04-30	615752	476903
Europe	Albania	2021-05-01	618714	494028
Europe	Albania	2021-05-02	620249	0
Europe	Albania	2021-05-03	622011	517501
Europe	Albania	2021-05-04	624006	537387
Europe	Albania	2021-05-05	626857	556584

CHECKING FOR DUPLICATE VALUES

```
SELECT LOCATION,`DATE`,CONTINENT, COUNT(*) AS COUNT_DUP FROM
COVID_DEATHS GROUP BY LOCATION,`DATE`,CONTINENT HAVING
COUNT_DUP>1;
```

	location	date	continent	count_dup

```
SELECT LOCATION,`DATE`,CONTINENT, COUNT(*) AS COUNT_DUP FROM
COVID_VACCINATION GROUP BY LOCATION,`DATE`,CONTINENT HAVING
COUNT_DUP>1;
```

	location	date	continent	count_dup

TOTAL CONTINENT & COUNTRIES

```
SELECT
        COUNT(DISTINCT CONTINENT)
        AS TOTAL_CONTINENT

FROM
        COVID_DEATHS;
```

TOTAL_CONTINENT
6

```
SELECT
        COUNT(DISTINCT LOCATION)
        AS  TOTAL_COUNTRIES

FROM
        COVID_DEATHS;
```

TOTAL_COUNTRIES
230

TOTAL COVID-19 CASES ACROSS THE WORLD

```
SELECT
        MAX(TOTAL_CASES)TOTAL_CASES,LOCATION
FROM
        COVID_DEATHS
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION ORDER BY TOTAL_CASES DESC;
```

TOTAL_CASES	LOCATION
33717567	United States
30545433	India
18769808	Brazil
5978650	France
5544209	Russia
5440368	Turkey
4920168	United Kingdom
4535473	Argentina
4350495	Colombia
4263317	Italy
3833868	Spain
3738470	Germany

TOTAL COVID-19 DEATHS ACROSS THE WORLD

```
SELECT
        MAX(TOTAL_DEATHS)TOTAL_DEATHS,LOCATION
FROM
        COVID_DEATHS
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION ORDER BY TOTAL_DEATHS DESC;
```

LOCATION	TOTAL_DEATHS
United States	605526
Brazil	524417
India	402005
Mexico	233622
Peru	193230
Russia	135637
United Kingdom	128486
Italy	127649
France	111323
Colombia	108896
Argentina	95904
Germany	91040

TOTAL COVID-19 VACCINATIONS ACROSS THE WORLD

```
SELECT
        LOCATION, MAX(TOTAL_VACCINATIONS) AS
        TOTAL_VACCINATIONS
FROM
        COVID_VACCINATION
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION
ORDER BY
        TOTAL_VACCINATIONS DESC;
```

LOCATION	TOTAL_VACCINATIONS
China	1296037000
India	344300590
United States	329970551
Brazil	102780096
United Kingdom	78537908
Germany	75781404
France	54483343
Italy	53203327
Turkey	52479924
Mexico	46945511
Japan	46248972
Indonesia	45495972

TOTAL COVID-19 VACCINATIONS ACROSS THE WORLD

```
SELECT
        LOCATION, MAX(TOTAL_VACCINATIONS) AS
        TOTAL_VACCINATIONS
FROM
        COVID_VACCINATION
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION
ORDER BY
        TOTAL_VACCINATIONS DESC;
```

LOCATION	TOTAL_VACCINATIONS
China	1296037000
India	344300590
United States	329970551
Brazil	102780096
United Kingdom	78537908
Germany	75781404
France	54483343
Italy	53203327
Turkey	52479924
Mexico	46945511
Japan	46248972
Indonesia	45495972

WORST AFFECTED COUNTRIES WITH RESPECT TO POPULATION

```
SELECT
        LOCATION,  MAX(TOTAL_CASES),  POPULATION,
        (MAX(TOTAL_CASES) / POPULATION) * 100 AS
        AFFECTED_POPULATION_PERCENTAGE
FROM
        COVID_DEATHS

WHERE
        CONTINENT IS NOT NULL

GROUP BY
        LOCATION,POPULATION

ORDER BY
        AFFECTED_POPULATION_PERCENTAGE DESC ;
```

LOCATION	MAX(TOTAL_CASES)	POPULATION	AFFECTED_POPULATION_PERCENTAGE
Andorra	13918	77265	18.0133
Seychelles	15857	98340	16.1247
Montenegro	100327	628062	15.9741
Bahrain	266426	1701583	15.6575
Czechia	1667935	10708982	15.5751
San Marino	5091	33938	15.0009
Maldives	74351	540542	13.7549
Slovenia	257421	2078932	12.3824
Luxembourg	71031	625976	11.3472
Sweden	1090880	10099270	10.8016
Uruguay	372709	3473727	10.7294
Serbia	716873	6804596	10.5351

COUNTRIES WITH HIGHEST DEATH PERCENTAGE WITH RESPECT TO POPULATION

```
SELECT
        LOCATION,    MAX(TOTAL_DEATHS),
        POPULATION,    (MAX(TOTAL_DEATHS) /
        POPULATION) * 100 AS
        PERECNTDEATHPOPULATION
FROM
        COVID_DEATHS
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION,POPULATION
ORDER BY
        PERECNTDEATHPOPULATION DESC;
```

LOCATION	MAX(TOTAL_DEATHS)	POPULATION	PERECNTDEATHPOPULATION
Peru	193230	32971846	0.5860
Hungary	29992	9660350	0.3105
Bosnia and Herzegovina	9667	3280815	0.2947
Czechia	30310	10708982	0.2830
San Marino	90	33938	0.2652
North Macedonia	5486	2083380	0.2633
Bulgaria	18084	6948445	0.2603
Montenegro	1615	628062	0.2571
Brazil	524417	212559409	0.2467
Slovakia	12513	5459643	0.2292
Belgium	25185	11589616	0.2173
Colombia	108896	50882884	0.2140

COUNTRIES WITH HIGHEST AVREAGE OF DEATH CASES A DAY

```
SELECT
        LOCATION, ROUND(AVG(NEW_DEATHS)) AS
        AVG_DEATH_A_DAY
FROM
        COVID_DEATHS
WHERE
        CONTINENT IS NOT NULL
GROUP BY
        LOCATION
ORDER BY
        AVG_DEATH_A_DAY DESC;
```

location	avg_death_a_day
United States	1143
Brazil	1059
India	770
Mexico	424
Peru	354
Russia	260
United Kingdom	247
Italy	245
Colombia	224
France	211
Argentina	174
Germany	172

JOINING TWO TABLES

```
SELECT    *

FROM
                COVID_DEATHS CD

 JOIN

                 COVID_VACCINATION CV
ON
                CD.LOCATION = CV.LOCATION
AND
                CD.DATE = CV.DATE;
```

POPULATION VS TOTAL VACCINATION
DAY BY DAY

```
SELECT
        CD.CONTINENT,CD.LOCATION,CD.DATE,
        CD.POPULATION,CV.NEW_VACCINATIONS,
        SUM(CV.NEW_VACCINATIONS)
        OVER(PARTITION BY CD.LOCATION ORDER BY
        CD.LOCATION, CD.DATE) AS TOTAL_VACCINATION
FROM
        COVID_DEATHS CD
JOIN
         COVID_VACCINATION CV
ON
        CD.LOCATION = CV.LOCATION
 WHERE
        CD.CONTINENT IS NOT NULL
AND
        CD.DATE = CV.DATE;
```

continent	location	date	population	new_vaccinations	total_vaccination
Europe	Albania	2021-03-23	2877800	3461	6358
Europe	Albania	2021-03-24	2877800	2302	8660
Europe	Albania	2021-03-25	2877800	5356	14016
Europe	Albania	2021-03-26	2877800	2900	16916
Europe	Albania	2021-03-27	2877800	1827	18743
Europe	Albania	2021-03-28	2877800	13925	32668
Europe	Albania	2021-03-29	2877800	0	32668
Europe	Albania	2021-03-30	2877800	0	32668
Europe	Albania	2021-03-31	2877800	19525	52193
Europe	Albania	2021-04-01	2877800	16617	68810
Europe	Albania	2021-04-02	2877800	17023	85833
Europe	Albania	2021-04-03	2877800	13010	98843

POPULATION VS PERCENTAGE OF PEOPLE VACCINATED

```
WITH POPULATION_VS_VACCINATION AS
                (SELECT       CD.CONTINENT,CD.LOCATION,CD.DATE,
                CD.POPUL     ATION,CV.NEW_VACCINATIONS,
                SUM(CV.NEW_VACCINATIONS)
                OVER(PARTITION              BY CD.LOCATION ORDER BY
                CD.LOCATION, CD.DATE) AS
                TOTAL_PEOPLE_VACCINATED
FROM
                COVID_DEATHS CD
JOIN
                COVID_VACCINATION CV
ON
                CD.LOCATION = CV.LOCATION
 WHERE
                CD.CONTINENT IS NOT NULL
AND
                CD.DATE = CV.DATE )
 SELECT   *,
                (ROLLING_PEOPLE_VACCINATION/POPULATION)*
                100 AS PERCENTPOPULATIONVACCINATED
FROM
POPULATION_VS_VACCINATION ;
```

continent	location	date	population	new_vaccinations	rolling_People_vaccination	PercentPopulationVaccinated
Europe	Albania	2021-03-30	2877800	0	32668	1.1352
Europe	Albania	2021-03-31	2877800	19525	52193	1.8136
Europe	Albania	2021-04-01	2877800	16617	68810	2.3911
Europe	Albania	2021-04-02	2877800	17023	85833	2.9826
Europe	Albania	2021-04-03	2877800	13010	98843	3.4347
Europe	Albania	2021-04-04	2877800	7386	106229	3.6913
Europe	Albania	2021-04-05	2877800	13826	120055	4.1718
Europe	Albania	2021-04-06	2877800	14880	134935	4.6888
Europe	Albania	2021-04-07	2877800	10791	145726	5.0638
Europe	Albania	2021-04-08	2877800	10163	155889	5.4170
Europe	Albania	2021-04-09	2877800	8134	164023	5.6996
Europe	Albania	2021-04-10	2877800	6134	170157	5.9127

CREATED A VIEW ‘SUMMARY’ WITH ALL MAJOR CALCULATIONS5

```
CREATE VIEW SUMMARY AS
SELECT
        CD.LOCATION,    CD.DATE,
        SUM(CD.NEW_CASES),    SUM(CD.NEW_DEATHS),    SUM(CV.NEW_VACCINATIONS),
        SUM(CD.TOTAL_DEATHS)AS TOTAL_DEATHSS,
        SUM(CD.TOTAL_CASES) AS TOTAL_CASESS,
        SUM(CV.TOTAL_VACCINATIONS)AS TOTAL_VACCINATIONS,
        (SUM(CD.TOTAL_DEATHS) / CD.POPULATION) * 100 AS
        DEATH_PERCENTAGE,
        (SUM(CD.TOTAL_CASES) / CD.POPULATION) * 100 AS
        AFFECTED_PERCENTAGE,
        (SUM(CV.TOTAL_VACCINATIONS)/CD.POPULATION)*100 AS
        VACCCINATION_PERCENTAGE,      (SUM(TOTAL_DEATHS)/SUM(TOTAL_CASES))*100
        AS MORTALITY_RATE
FROM
        COVID_DEATHS CD
JOIN
        COVID_VACCINATION CV
ON
        CD.LOCATION=CV.LOCATION
WHERE
        CD.CONTINENT IS NOT NULL
 AND
         CD.DATE=CV.DATE
GROUP BY
        CD.LOCATION,    CD.DATE,    CD.POPULATION
ORDER BY
         CD.LOCATION,    CD.DATE;
```

sum(CD.NEW_CASES)	sum(CD.NEW_DEATHS)	SUM(CV.NEW_VACCINATIONS)	TOTAL_DEATHSS	TOTAL_CASESS	TOTAL_VACCINATIONS	death_percentage	affected_percentage	vacccination_percentage	MORTALITY_RATE
542	12	0	206	9219	0	0.0005	0.0237	0.0000	2.2345
782	11	0	217	10001	0	0.0006	0.0257	0.0000	2.1698
584	2	0	219	10585	0	0.0006	0.0272	0.0000	2.0690
591	1	0	220	11176	0	0.0006	0.0287	0.0000	1.9685
658	1	0	221	11834	0	0.0006	0.0304	0.0000	1.8675
625	7	0	228	12459	0	0.0006	0.0320	0.0000	1.8300
580	8	0	236	13039	0	0.0006	0.0335	0.0000	1.8100
623	11	0	247	13662	0	0.0006	0.0351	0.0000	1.8079
866	3	0	250	14528	0	0.0006	0.0373	0.0000	1.7208
680	8	0	258	15208	0	0.0007	0.0391	0.0000	1.6965
545	8	0	266	15753	0	0.0007	0.0405	0.0000	1.6886

CREATED A VIEW ‘SUMMARY’ WITH ALL MAJOR CALCULATIONS5

```
SELECT
        MAX(POPULATION) AS POPULATION,
       MAX(CD.TOTAL_CASES) AS TOTAL_CASES,
       MAX(CD.TOTAL_DEATHS) AS TOTAL_DEATHS,
       MAX(CV.TOTAL_VACCINATIONS) AS TOTAL_VACCINATION,
       MAX(CD.TOTAL_CASES) / MAX(POPULATION) * 100 AS
       AFFECTED_POPULATION_PERCENTAGE,
       MAX(CD.TOTAL_DEATHS) / MAX(POPULATION) * 100 AS
       DECEASED_POPULATION_PERCENTAGE,
       MAX(CD.TOTAL_DEATHS) / MAX(CD.TOTAL_CASES) * 100 AS
       MORTALITY_RATE
FROM
       COVID_DEATHS CD
  JOIN
       COVID_VACCINATION CV
ON
       CD.LOCATION = CV.LOCATION
AND
       CD.DATE = CV.DATE
WHERE
       CD.LOCATION = ‘World';
```

POPULATION	TOTAL_CASES	TOTAL_DEATHS	TOTAL_VACCINATION	AFFECTED_POPULATION_PERCENTAGE	DECEASED_POPULATION_PERCENTAGE	MORTALITY_RATE
7794798729	183742035	3976335	3192698685	2.3572	0.0510	2.1641

INSIGHTS

- The highest number of total cases is observed in Asia(56366162), likely due to its larger population and higher density, which facilitates virus transmission.
- Continents like Oceania(56856), have lower total cases, potentially due to better containment measures or less international travel.
- The distribution of total deaths closely follows the distribution of total cases, with Europe(1109009) experiencing the highest number of deaths.
- Andorra, a country in Europe recorded highest(18.01%) percentage of population affected by covid-19.
- China recorded highest number of people fully vaccinated(223299000)

KPI's

New Cases per Day

345787

Total Cases

183742035

Case Fatality Rate

2.16%

New Deaths per Day

18050

Total Deaths

3975335

*Global Metrics

CONCLUSION

This exploratory analysis provides a comprehensive understanding of the COVID-19 situation by analyzing the trends in cases, deaths, and vaccinations. The insights derived from this analysis can help in identifying key areas of concern and evaluating the effectiveness of vaccination campaigns. The defined KPIs enable continuous monitoring and provide a basis for data-driven decision-making.

THANK YOU!!!