MS-SQL EDA PROJECT ON COVID DATA

THIS PROJECT HAS BEEN PERFORMED ON WORLDWIDE COVID DATA ACROSS ALL COUNTRIES.
SEVERAL QUESTIONS WERE POSED AND THEY WERE ANSWERED USING SQL QUERIES.

MY PROJECT IS INTENDED TO EXPLORE SIMILAR HEALTHCARE DATASETS IN THE INDUSTRY.

THE VARIOUS SQL COMMANDS OR FUNCTIONS EXPLORED ARE:

1. CREATE
2. WHERE
3. GROUP BY
4. HAVING
5. ORDER BY
6. CTE
7. CONVERT,ROUND,MAX
8. CASE, WHEN
9. JOIN

BELOW ARE THE OUTPUT RESULTS AND ATTACHED QUERIES.

Q1. DISPLAY SOME SPECIFIC PARAMETER INFORMATION FROM THE TABLE:
   ### 1. COUNTS OF DISTINCT DATES OF OBSERVATIONS
   ### 2. COUNTS OF DISTINCT COUNTRIES
   ### 3. COUNTS OF DISTINCT CONTINENTS

```sql
select count(distinct date) as Number_of_distinct_days,count(distinct location) as Number_of_countries,
count(distinct continent) as Number_of_continents
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
```

	Number_of_distinct_days	Number_of_countries	Number_of_continents
1	1286	243	6

Q2. DISPLAYING THE TOTAL COVID DEATHS, THE HIGHEST NUMBER OF CASUALTIES IN A SINGLE DAY AND THE HIGHEST NUMBER OF VACCINATIONS IN A DAY ORDER BY MOST DEATHS IN A COUNTRY

```sql
SELECT location as Country, max(total_deaths) as TOTAL_DEATHS_BY_COUNTRY,max(new_deaths) as Highest_deaths_in_a_single_day,
max(new_vaccinations) as Highes_vaccination_in_a_single_day
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by [location]
order by TOTAL_DEATHS_BY_COUNTRY desc
```

	Country	TOTAL_DEATHS_BY_COUNTRY	Highest_deaths_in_a_single_day	Highes_vaccination_in_a_single_day
1	United States	1127152	5061	4581777
2	Brazil	703964	4249	3976605
3	India	531908	6148	18627269
4	Russia	399649	1254	2389472
5	Mexico	334336	1495	7246123
6	United Kingdom	227739	1487	1062705
7	Peru	221043	843	603805
8	Italy	190868	993	724860
9	Germany	174807	6460	1562042
10	France	167923	5602	970641
11	Indonesia	161879	2069	2952259
12	Iran	146297	709	1706088
13	Colombia	142836	720	1036209
14	Argentina	130472	926	572308

Q3. DISPLAYING THE COUNTRIES ORDERED BY THE ONES WITH MOST PEOPLE INFECTED AS PROPORTION TO THE TOTAL POPULATION

```sql
SELECT location as Country, round(max(total_cases)/avg(population)*100,2) as Percentage_infected,
max(total_cases) as Total_Cases,avg(population) as Population
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by [location]
order by 2 desc
```

	Country	Percentage_infected	Total_Cases	Population
1	Cyprus	73.76	660854	896007
2	San Marino	72.21	24326	33690
3	Brunei	68.77	308777	449002
4	Austria	68.03	6081287	8939617
5	Faeroe Islands	65.25	34658	53117
6	Slovenia	63.43	1344559	2119843
7	Gibraltar	62.89	20550	32677
8	Martinique	62.68	230354	367512
9	South Korea	62.25	32256154	51815808
1…	Andorra	60.14	48015	79843
1…	Jersey	59.92	66391	110796
1…	Luxembourg	59.11	382785	647601
1…	Saint Pierre and Miquelon	58.22	3426	5885
1…	Denmark	58.04	3414333	5882259

Q4. APPLYING CTE TO THE TABLE TO DISPLAY THE MOST POPULOUS COUNTRIES IN EACH CONTINENT. THE QUERY CAN BE MODIFIED TO DISPLAY EVEN THE TOP N NUMBER OF COUNTRIES BY POPULATION AS WELL

```sql
WITH CTE AS (
    SELECT  location, continent, population,
          ROW_NUMBER() OVER (PARTITION BY continent ORDER BY population DESC) AS RowNum
    FROM Portfolio_Covid..[Covid-data]
    WHERE continent IS NOT NULL   -- Exclude rows where continent is NULL
)
SELECT location as Country, continent as CONTINENT, population AS POPULATION
FROM CTE
WHERE RowNum <= 1
ORDER BY POPULATION asc;
```

	Country	CONTINENT	POPULATION
1	Australia	Oceania	26177410
2	Russia	Europe	144713312
3	Brazil	South America	215313504
4	Nigeria	Africa	218541216
5	United States	North America	338289856
6	China	Asia	1425887360

Q5. SELECT SOME FUNDAMENTAL PARAMETERS RELATING TO MOST VACCINATIONS, CASES AND DEATHS BY A COUNTRY

```sql
select iso_code as Country_code, location as Country,CONVERT(int,MAX(total_deaths)) as Deaths,
(select MAX(total_cases) )as Total_Cases, MAX(people_fully_vaccinated) as Total_Vaccinations,avg(population) as Population,
round((max(people_fully_vaccinated)/max(total_cases)),2) as Vaccinations_over_cases
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by [location],continent,iso_code
order by Deaths desc
```

	Country_code	Country	Deaths	Total_Cases	Total_Vaccinations	Population	Vaccinations_over_ca
1	USA	United States	1127152	103436829	230637348	338289856	2.23
2	BRA	Brazil	703964	37671420	176164186	215313504	4.68
3	IND	India	531908	44994351	951979378	1417173120	21.16
4	RUS	Russia	399649	22963688	79702396	144713312	3.47
5	MEX	Mexico	334336	7633355	81849962	127504120	10.72
6	GBR	United Kingdom	227739	24639160	50762968	67508936	2.06
7	PER	Peru	221043	4512091	28671694	34049588	6.35
8	ITA	Italy	190868	25897801	47963720	59037472	1.85
9	DEU	Germany	174807	38435774	63563414	83369840	1.65
10	FRA	France	167923	38989382	53191743	67813000	1.36
11	IDN	Indonesia	161879	6812127	174893201	275501344	25.67

Q6. DISPLAYING THE DATES IN DESCENDING ORDER IN WHICH MOST DEATHS HAPPENED BY COVID IN A SINGLE DAY.

```sql
SELECT date, SUM(new_deaths) as Deaths_per_day, SUM(new_cases) as Cases_per_day
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by date
order by Deaths_per_day desc
```

	date	Deaths_per_day	Cases_per_day
1	2021-01-24	27941	1155210
2	2021-01-17	26731	1270403
3	2021-01-10	25952	1298406
4	2021-01-31	25899	1035421
5	2020-12-20	25888	1146866
6	2020-12-13	25790	1111685
7	2020-12-06	24029	1048521
8	2020-11-22	24001	1104066
9	2021-01-03	23455	976221
10	2021-02-07	23251	834372
11	2020-11-29	23073	999737

Q7. ORDERING THE COUNTRIES HAVING THE HIGHEST FATALITIES IN A SINGLE DAY

```sql
select top 20 date,location as Highest_fatality_country, MAX(new_deaths) as Deaths_in_a_day
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by date,location
having MAX(new_deaths) > 3000
order by Deaths_in_a_day desc
```

	date	Highest_fatality_country	Deaths_in_a_day
1	2022-03-22	Chile	11447
2	2021-07-21	Ecuador	8786
3	2020-12-20	Germany	6460
4	2021-06-10	India	6148
5	2020-04-05	Spain	5841
6	2020-12-13	Germany	5763
7	2021-01-10	Germany	5760
8	2021-01-03	Germany	5746
9	2020-12-27	Germany	5728
10	2020-03-29	Spain	5663
11	2020-11-22	France	5602

Q8. DISPLAYING THE PER CAPITA GDP AND CORRESPONDING MEDICAL INFRASTRUCTURE OF DIFFERENT COUNTRIES AND ALSO CLASSIFYING THEM FURTHER BY INCOME USING CASE WHEN CLAUSE.

```sql
select location, max(gdp_per_capita) as Per_Capita_GDP ,
round(avg(hospital_beds_per_thousand),2) as Per_thousand_Hospital_beds,
SUM(new_deaths) as Total_deaths ,
CASE when avg(gdp_per_capita) > 12000 then 'High Income Country'
when avg(gdp_per_capita) between 4000 and 12000 then 'Upper middle income country'
when avg(gdp_per_capita) between 1000 and 4000  then 'Lower middle income country'
else 'Low income country'
end as 'Country_by_development'
from Portfolio_Covid..[Covid-data]
where continent not like 'NULL'
group by location
order by location asc
```

	location	Per_Capita_GDP	Per_thousand_Hospital_beds	Total_deaths	Country_by_development
1	Afghanistan	1803.987	0.5	7928	Lower middle income country
2	Albania	11803.431	2.89	3604	Upper middle income country
3	Algeria	13913.839	1.9	6881	High Income Country
4	American Samoa	NULL	NULL	34	Low income country
5	Andorra	NULL	NULL	159	Low income country
6	Angola	5819.495	NULL	1934	Upper middle income country
7	Anguilla	NULL	NULL	12	Low income country
8	Antigua and Barbuda	21490.943	3.8	146	High Income Country
9	Argentina	18933.907	5	130472	High Income Country
10	Armenia	8787.58	4.2	8751	Upper middle income country
11	Aruba	35973.781	NULL	290	High Income Country
12	Australia	44648.71	3.84	21917	High Income Country

Q9. USING INNER JOIN TO DISPLAY THE NUMBER OF DEATHS HAPPENING IN A SINGLE DAY ACROSS ALL NATIONS ON A PARTICULAR DAY EXACTLY A YEAR APART. THIS SHOWS HOW COVID PROGRESSED OVER ONE YEAR.

```sql
select c.location as Country,c.new_deaths as Deaths_in_2020,c1.new_deaths as Deaths_in_2021,c.[date] as Day_in_2020,
c1.date as Day_in_2021
from Portfolio_Covid..[Covid-data] c
inner join Portfolio_Covid..[Covid-data] c1
on c.location=c1.[location]
where c.date= '2021-05-01' and c1.date='2022-05-01' and c.continent not like 'NULL' and c1.continent not like 'NULL'
```

	Country	Deaths_in_2020	Deaths_in_2021	Day_in_2020	Day_in_2021
1	Afghanistan	6	0	2021-05-01	2022-05-01
2	Algeria	9	0	2021-05-01	2022-05-01
3	Azerbaijan	27	0	2021-05-01	2022-05-01
4	Bangladesh	60	0	2021-05-01	2022-05-01
5	Belgium	46	15	2021-05-01	2022-05-01
6	Brazil	3001	185	2021-05-01	2022-05-01
7	Bulgaria	0	64	2021-05-01	2022-05-01
8	Burkina Faso	0	0	2021-05-01	2022-05-01
9	Central African Republic	0	0	2021-05-01	2022-05-01
1…	Cook Islands	0	0	2021-05-01	2022-05-01
1…	Costa Rica	15	0	2021-05-01	2022-05-01

CONCLUSION:

THIS PROJECT THROUGH ITS DIFFERENT QUESTIONS AND QUERIES ANSWER VARIOUS ASPECTS ABOUT THE COVID SPREAD ACROSS ALL COUNTRIES.

THE MAIN INFERENCES ARE : THE DEATHS AND CASES SUBSIDED IN ALMOST ALL COUNTRIES AFTER 2020;

UNITED STATES HAD THE MOST DEATHS WHILE INDIA HAS THE MOST VACCINATIONS OVERALL AND ALSO IN A SINGLE DAY.

CYPRUS SHOWS THE HIGHEST PERCENTAGE OF POPULATION TAHT GOT INFECTED WHILE CHILE RECORDED THE MOST DEATHS OUT OF ALL NATIONS IN A SINGLE DAY.