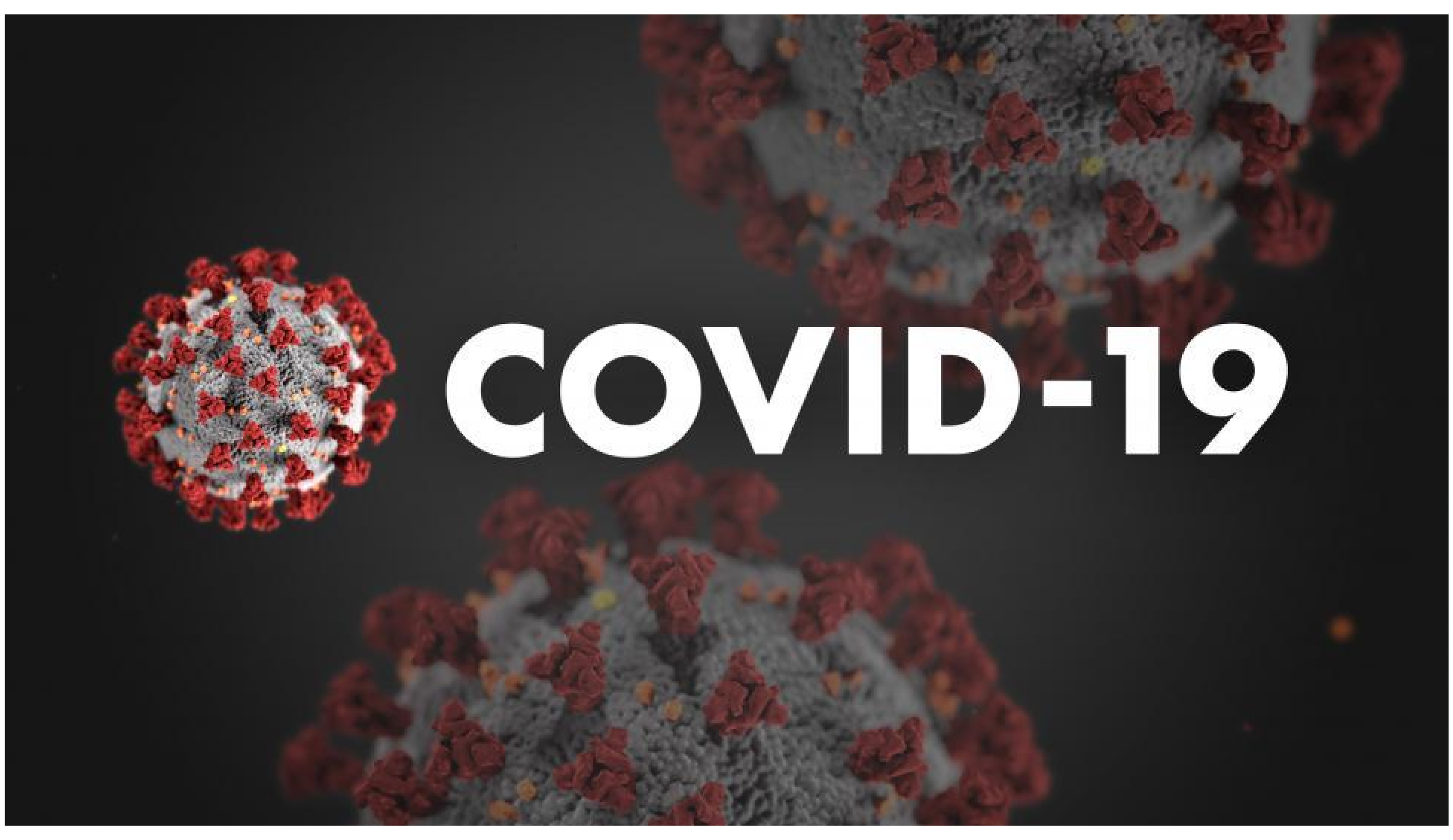

Exploratory Analysis Of COVID-19 Data Using SQL

The Covid-19 pandemic was a really serious outbreak that cost lives and affected the economy negatively. In this micro project, I have tried to analyse the COVID-19 data using SQL. The dataset used for this project have the recorded cases from 24th February 2020 to 30th April 2021. I got this daily updated dataset on this website; https://ourworldindata.org/covid-deaths. The data exploration methods and processes are documented in this article.

The dataset contains details of columns on cases, deaths, and vaccinations. For simplicity, I decided to split the dataset into two forms with Excel: CovidDeaths, and CovidVaccinations. The CovidDeaths dataset contained columns and records on cases, and deaths while the CovidVaccinations dataset contains only columns and rows regarding vaccinations and tests. This whole SQL project was executed on the Microsoft SQL Server Management Studio, Below is a snapshot of a few of the columns and entries in the CovidDeaths, CovidVaccinations dataset, the SQL query file will be available on my Github Repo:

Results | Messages

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	new_cases_per_million	new_cases_smoothed_per_million
1	AFG	Asia	Afghanistan	2020-02-24 00:00:00.000	38928341	1	1	NULL	NULL	NULL	NULL	0.026	0.026	NULL
2	AFG	Asia	Afghanistan	2020-02-25 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026	0	NULL
3	AFG	Asia	Afghanistan	2020-02-26 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026	0	NULL
4	AFG	Asia	Afghanistan	2020-02-27 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026	0	NULL
5	AFG	Asia	Afghanistan	2020-02-28 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026	0	NULL
6	AFG	Asia	Afghanistan	2020-02-29 00:00:00.000	38928341	1	0	0.143	NULL	NULL	0	0.026	0	0.004
7	AFG	Asia	Afghanistan	2020-03-01 00:00:00.000	38928341	1	0	0.143	NULL	NULL	0	0.026	0	0.004
8	AFG	Asia	Afghanistan	2020-03-02 00:00:00.000	38928341	1	0	0	NULL	NULL	0	0.026	0	0
9	AFG	Asia	Afghanistan	2020-03-03 00:00:00.000	38928341	2	1	0.143	NULL	NULL	0	0.051	0.026	0.004
10	AFG	Asia	Afghanistan	2020-03-04 00:00:00.000	38928341	4	2	0.429	NULL	NULL	0	0.103	0.051	0.011
11	AFG	Asia	Afghanistan	2020-03-05 00:00:00.000	38928341	4	0	0.429	NULL	NULL	0	0.103	0	0.011
12	AFG	Asia	Afghanistan	2020-03-06 00:00:00.000	38928341	4	0	0.429	NULL	NULL	0	0.103	0	0.011
13	AFG	Asia	Afghanistan	2020-03-07 00:00:00.000	38928341	4	0	0.429	NULL	NULL	0	0.103	0	0.011
14	AFG	Asia	Afghanistan	2020-03-08 00:00:00.000	38928341	5	1	0.571	NULL	NULL	0	0.128	0.026	0.015
15	AFG	Asia	Afghanistan	2020-03-09 00:00:00.000	38928341	7	2	0.857	NULL	NULL	0	0.18	0.051	0.022

Query - 1

1. Get the Daily Total Cases, Deaths, date wise PercentageDeathRate for country 'India'.

```
-- Get the day_wise_death_percentage_for_Country_India
SELECT location,date,total_cases,total_deaths,(total_deaths/total_cases)*100 AS deathpercentage
FROM COVID19.dbo.CovidDeaths
WHERE location = 'India'
```

Results | Messages

	location	date	total_cases	total_deaths	deathpercentage
1	India	2020-01-30 00:00:00.000	1	NULL	NULL
2	India	2020-01-31 00:00:00.000	1	NULL	NULL
3	India	2020-02-01 00:00:00.000	1	NULL	NULL
4	India	2020-02-02 00:00:00.000	2	NULL	NULL
5	India	2020-02-03 00:00:00.000	3	NULL	NULL
6	India	2020-02-04 00:00:00.000	3	NULL	NULL
7	India	2020-02-05 00:00:00.000	3	NULL	NULL
8	India	2020-02-06 00:00:00.000	3	NULL	NULL
9	India	2020-02-07 00:00:00.000	3	NULL	NULL
10	India	2020-02-08 00:00:00.000	3	NULL	NULL
11	India	2020-02-09 00:00:00.000	3	NULL	NULL
12	India	2020-02-10 00:00:00.000	3	NULL	NULL
13	India	2020-02-11 00:00:00.000	3	NULL	NULL
14	India	2020-02-12 00:00:00.000	3	NULL	NULL
15	India	2020-02-13 00:00:00.000	3	NULL	NULL
16	India	2020-02-14 00:00:00.000	3	NULL	NULL
17	India	2020-02-15 00:00:00.000	3	NULL	NULL
18	India	2020-02-16 00:00:00.000	3	NULL	NULL
19	India	2020-02-17 00:00:00.000	3	NULL	NULL
20	India	2020-02-18 00:00:00.000	3	NULL	NULL

	location	date	total_cases	total_deaths	deathpercentage
436	India	2021-04-09 00:00:00.000	13205926	168436	1.2754576998...
437	India	2021-04-10 00:00:00.000	13358805	169275	1.2671417840...
438	India	2021-04-11 00:00:00.000	13527717	170179	1.2580023665...
439	India	2021-04-12 00:00:00.000	13689453	171058	1.2495605193...
440	India	2021-04-13 00:00:00.000	13873825	172085	1.2403572915...
441	India	2021-04-14 00:00:00.000	14074564	173123	1.2300416552...
442	India	2021-04-15 00:00:00.000	14291917	174308	1.2196264503...
443	India	2021-04-16 00:00:00.000	14526609	175649	1.2091534920...
444	India	2021-04-17 00:00:00.000	14788003	177150	1.1979305116...
445	India	2021-04-18 00:00:00.000	15061805	178769	1.1869028977...
446	India	2021-04-19 00:00:00.000	15320972	180530	1.1783194956...
447	India	2021-04-20 00:00:00.000	15616130	182553	1.1690028195...
448	India	2021-04-21 00:00:00.000	15930774	184657	1.1591213333...
449	India	2021-04-22 00:00:00.000	16263695	186920	1.1493083213...
450	India	2021-04-23 00:00:00.000	16610481	189544	1.1411108444...
451	India	2021-04-24 00:00:00.000	16960172	192311	1.133897698679
452	India	2021-04-25 00:00:00.000	17313163	195123	1.1270210994...
453	India	2021-04-26 00:00:00.000	17636186	197894	1.1220906833...
454	India	2021-04-27 00:00:00.000	17997113	201187	1.1178848518...
455	India	2021-04-28 00:00:00.000	18376421	204832	1.1146457735...
456	India	2021-04-29 00:00:00.000	18762976	208330	1.1103249292...
457	India	2021-04-30 00:00:00.000	19164969	211853	1.1054179111...

Query - 2

Show the country wise Population,TotalCases,TotalDeaths,percentage of population infected & percentage of population dead.

```
-- Top countries with the highest total cases
SELECT location,population,
MAX(CAST(total_cases as INT)) as TotalCases,
MAX(CAST(total_deaths as INT)) as TotalDeaths,
(MAX(CAST(total_cases as INT))/population)*100 as PercentagePopulationInfected,
(MAX(CAST(total_deaths as INT))/population)*100 as PercentagePopulationDead
FROM COVID19.dbo.CovidDeaths
WHERE continent IS NOT NULL
GROUP BY location,population
ORDER BY TotalCases DESC
```

Results | Messages

	location	population	TotalCases	TotalDeaths	PercentagePopulationInfected	PercentagePopulationDead
1	United States	331002647	32346971	576232	9.77242064169958	0.174086825353998
2	India	1380004385	19164969	211853	1.38876145672537	0.0153516178863446
3	Brazil	212559409	14659011	403781	6.89643007052207	0.189961480369001
4	France	68147687	5677835	104675	8.33166208561121	0.153600224171952
5	Turkey	84339067	4820591	40131	5.71572720860192	0.0475829309328262
6	Russia	145934460	4750755	108290	3.25540314467193	0.0742045436012851
7	United Kingdom	67886004	4432246	127775	6.52895403889143	0.188219945896359
8	Italy	60461828	4022653	120807	6.65321101439407	0.199807058430321
9	Spain	46754783	3524077	78216	7.53736147165949	0.167289836421656
10	Germany	83783945	3405365	83097	4.06446008241794	0.0991800994808731
11	Argentina	45195777	2977363	63865	6.58770176691508	0.141307450030121
12	Colombia	50882884	2859724	73720	5.62020816272914	0.144881724864495
13	Poland	37846605	2792142	67502	7.37752303013705	0.178356816945668
14	Iran	83992953	2499077	71758	2.97534127654733	0.0854333577246653
15	Mexico	128932753	2344755	216907	1.81858755470769	0.168232660013085

Query executed successfully.

The snapshot of the result above shows the top 15 countries with the highest total cases, United States emerges as the country with the highest sum of cases of COVID-19, with over an 32million cases and recorded almost a half million total deaths, followed by India recording over 19 million cases and so on in descending order of Total Cases.

Query-3

Find the country wise population,TotalCases,TotalVaccinations,VaccinationPercentage.

```
/* Create an Temp Table to store the Percentage of
population vaccinated for each country*/

DROP TABLE IF EXISTS
CREATE TABLE #PercentagePopulationVaccinated
(
location nvarchar(255),
population int,
Total_cases int,
Total_vaccinations int,
VaccinationPercentage float
)
INSERT INTO #PercentagePopulationVaccinated
SELECT deaths.location,deaths.population,
MAX(CAST(deaths.total_cases AS INT)) AS TotalCases,
MAX(CAST(vaccin.total_vaccinations AS INT)) AS TotalVaccinations,
MAX(CAST(vaccin.total_vaccinations AS INT))/population*100 as VaccinationPercentage
FROM COVID19.dbo.CovidDeaths AS deaths
JOIN COVID19.dbo.CovidVaccinations AS vaccin
    ON deaths.location = vaccin.location
    AND deaths.date = vaccin.date
WHERE deaths.continent is not null
GROUP BY deaths.location,deaths.population
ORDER BY VaccinationPercentage DESC

SELECT *
FROM #PercentagePopulationVaccinated
```

Results | Messages

	location	population	Total_cases	Total_vaccinations	VaccinationPercentage
1	Gibraltar	33691	NULL	70334	208.761983912618
2	Seychelles	98340	5873	126840	128.981086028066
3	Falkland Islands	3483	NULL	4407	126.528854435831
4	Israel	8655541	838481	10497355	121.279016528256
5	United Arab Emirates	9890400	520236	10547584	106.644665534255
6	Cayman Islands	65720	NULL	65694	99.9604382227632
7	Bermuda	62273	NULL	55904	89.7724535512983
8	Isle of Man	85032	NULL	72617	85.3996142628657
9	San Marino	33938	5066	28970	85.3615416347457
10	Jersey	101073	NULL	82738	81.8596459984368
11	Chile	19116209	1198245	14767327	77.2502905780116
12	Saint Helena	6071	NULL	4572	75.3088453302586
13	Malta	441539	30292	331438	75.0642638589117
14	Maldives	540542	29835	397875	73.6066762619741
15	Bahrain	1701583	176934	1235170	72.5894652215026
16	United States	331002647	32346971	240159677	72.5552134330817
17	Aruba	106766	NULL	76865	71.9938931869696

Query executed successfully.

Summary

This is an COVID-19 data exploration project using SQL, which focuses on cases, deaths, and vaccinations. I believe more explorations can be done, as the purpose of this project was to apply basic knowledge of SQL in exploring a real-life scenario(COVID-19 Outbreak).

In [ ]: