DATA EXPLORATION PROJECT IN MS SQL SERVER

An event that has changed our lives recently has been the COVID-19 epidemic. The portal from which we extract the data, ourworldindata.org compiles global information related to the epidemic and relies on data from Johns Hopkins University (https://ourworldindata.org/covid-deaths).

In this project, will be using MS SQL Server to do some Data Exploration, so we can see what they found.

The database contains one tables (CovidData). I have separated it into two (CovidDeaths and CovidVaccination) to be able to analyze the information more easily.

So, in this notebook, we are going to analyze Covid-19 global data. The dataset contains information about the impact of COVID-19 epidemic (until 8 April 2022) owed by all countries across several categories.

We are going to find the answers to questions like:

- Likelihood of dying if you contract covid for each country
- Percentage of population who got Covid-19
- Countries with Highest Infection Rate compared to the population
- Countries with the highest death count
- Continent with the highest death count

Skills used (SQL):

Joins, CTE's, Temp Tables, Windows Functions, Aggregate Functions, Creating Views, Converting Data Types

```sql
--We start by checking that the Excel files have been imported
correctly:

Select *
From PortfolioProject.dbo.CovidDeaths
order by 3,4

Select *
From PortfolioProject.dbo.CovidVaccinations
order by 3,4



--Selecting Data that we are going to be using

Select location, date, total_cases, new_cases, total_deaths,
population

From PortfolioProject.dbo.CovidDeaths

order by 1,2
```

Results | Messages

	location	date	total_cases	new_cases	total_deaths	population
1	Afghanistan	2020-02-24 00:00:00.000	5.0	5.0	NULL	39835428.0
2	Afghanistan	2020-02-25 00:00:00.000	5.0	0.0	NULL	39835428.0
3	Afghanistan	2020-02-26 00:00:00.000	5.0	0.0	NULL	39835428.0
4	Afghanistan	2020-02-27 00:00:00.000	5.0	0.0	NULL	39835428.0
5	Afghanistan	2020-02-28 00:00:00.000	5.0	0.0	NULL	39835428.0
6	Afghanistan	2020-02-29 00:00:00.000	5.0	0.0	NULL	39835428.0
7	Afghanistan	2020-03-01 00:00:00.000	5.0	0.0	NULL	39835428.0
8	Afghanistan	2020-03-02 00:00:00.000	5.0	0.0	NULL	39835428.0
9	Afghanistan	2020-03-03 00:00:00.000	5.0	0.0	NULL	39835428.0
10	Afghanistan	2020-03-04 00:00:00.000	5.0	0.0	NULL	39835428.0
11	Afghanistan	2020-03-05 00:00:00.000	5.0	0.0	NULL	39835428.0
12	Afghanistan	2020-03-06 00:00:00.000	5.0	0.0	NULL	39835428.0
13	Afghanistan	2020-03-07 00:00:00.000	8.0	3.0	NULL	39835428.0
14	Afghanistan	2020-03-08 00:00:00.000	8.0	0.0	NULL	39835428.0
15	Afghanistan	2020-03-09 00:00:00.000	8.0	0.0	NULL	39835428.0
16	Afghanistan	2020-03-10 00:00:00.000	8.0	0.0	NULL	39835428.0
17	Afghanistan	2020-03-11 00:00:00.000	11.0	3.0	NULL	39835428.0

Query executed successfully.

```sql
-- Looking at Total Cases vs Total Deaths
--Shows likelihood of dying if you contract covid in your country

Select location, date, total_cases, total_deaths, (CAST(total_deaths
AS decimal(11,4))*100/total_cases) as DeathPercentage

From PortfolioProject.dbo.CovidDeaths

Where location like '%spain%'

order by 1,2
```

Results | Messages

	location	date	total_cases	total_deaths	DeathPercentage
44	Spain	2020-03-15 00:00:00.000	7798.0	289.0	3.70607848166...
45	Spain	2020-03-16 00:00:00.000	9942.0	342.0	3.43995171997...
46	Spain	2020-03-17 00:00:00.000	11748.0	533.0	4.53694245829...
47	Spain	2020-03-18 00:00:00.000	13910.0	623.0	4.47879223580...
48	Spain	2020-03-19 00:00:00.000	17963.0	830.0	4.62060902967...
49	Spain	2020-03-20 00:00:00.000	20410.0	1043.0	5.11024007839...
50	Spain	2020-03-21 00:00:00.000	25374.0	1375.0	5.41893276582...
51	Spain	2020-03-22 00:00:00.000	28768.0	1772.0	6.15962180200...
52	Spain	2020-03-23 00:00:00.000	35136.0	2311.0	6.57729963570...
53	Spain	2020-03-24 00:00:00.000	39885.0	2808.0	7.04024069198...
54	Spain	2020-03-25 00:00:00.000	49515.0	3647.0	7.36544481470...
55	Spain	2020-03-26 00:00:00.000	57786.0	4365.0	7.55373273803...
56	Spain	2020-03-27 00:00:00.000	65719.0	5138.0	7.81813478598...
57	Spain	2020-03-28 00:00:00.000	73235.0	5982.0	8.16822557520...
58	Spain	2020-03-29 00:00:00.000	80110.0	6803.0	8.49207339907...
59	Spain	2020-03-30 00:00:00.000	87956.0	7716.0	8.77256810223...
60	Spain	2020-03-31 00:00:00.000	95923.0	8464.0	8.82374404470...

Query executed successfully.

```sql
--Looking at Total Cases vs Population
--Shows what percentage of population got Covid

Select location, date, total_cases, population, (CAST(total_cases AS
decimal(17,4))*100/population) as PercentPopulationInfected

From PortfolioProject.dbo.CovidDeaths

Where location like '%spain%'

order by 1,2
```

Results | Messages

	location	date	total_cases	population	PercentPopulationInfected
1	Spain	2020-02-01 00:00:00.000	1.0	46745211.0	0.00000213925657539
2	Spain	2020-02-02 00:00:00.000	1.0	46745211.0	0.00000213925657539
3	Spain	2020-02-03 00:00:00.000	1.0	46745211.0	0.00000213925657539
4	Spain	2020-02-04 00:00:00.000	1.0	46745211.0	0.00000213925657539
5	Spain	2020-02-05 00:00:00.000	1.0	46745211.0	0.00000213925657539
6	Spain	2020-02-06 00:00:00.000	1.0	46745211.0	0.00000213925657539
7	Spain	2020-02-07 00:00:00.000	1.0	46745211.0	0.00000213925657539
8	Spain	2020-02-08 00:00:00.000	1.0	46745211.0	0.00000213925657539
9	Spain	2020-02-09 00:00:00.000	2.0	46745211.0	0.00000427851315079
10	Spain	2020-02-10 00:00:00.000	2.0	46745211.0	0.00000427851315079
11	Spain	2020-02-11 00:00:00.000	2.0	46745211.0	0.00000427851315079
12	Spain	2020-02-12 00:00:00.000	2.0	46745211.0	0.00000427851315079
13	Spain	2020-02-13 00:00:00.000	2.0	46745211.0	0.00000427851315079
14	Spain	2020-02-14 00:00:00.000	2.0	46745211.0	0.00000427851315079
15	Spain	2020-02-15 00:00:00.000	2.0	46745211.0	0.00000427851315079
16	Spain	2020-02-16 00:00:00.000	2.0	46745211.0	0.00000427851315079
17	Spain	2020-02-17 00:00:00.000	2.0	46745211.0	0.00000427851315079

Query executed successfully.

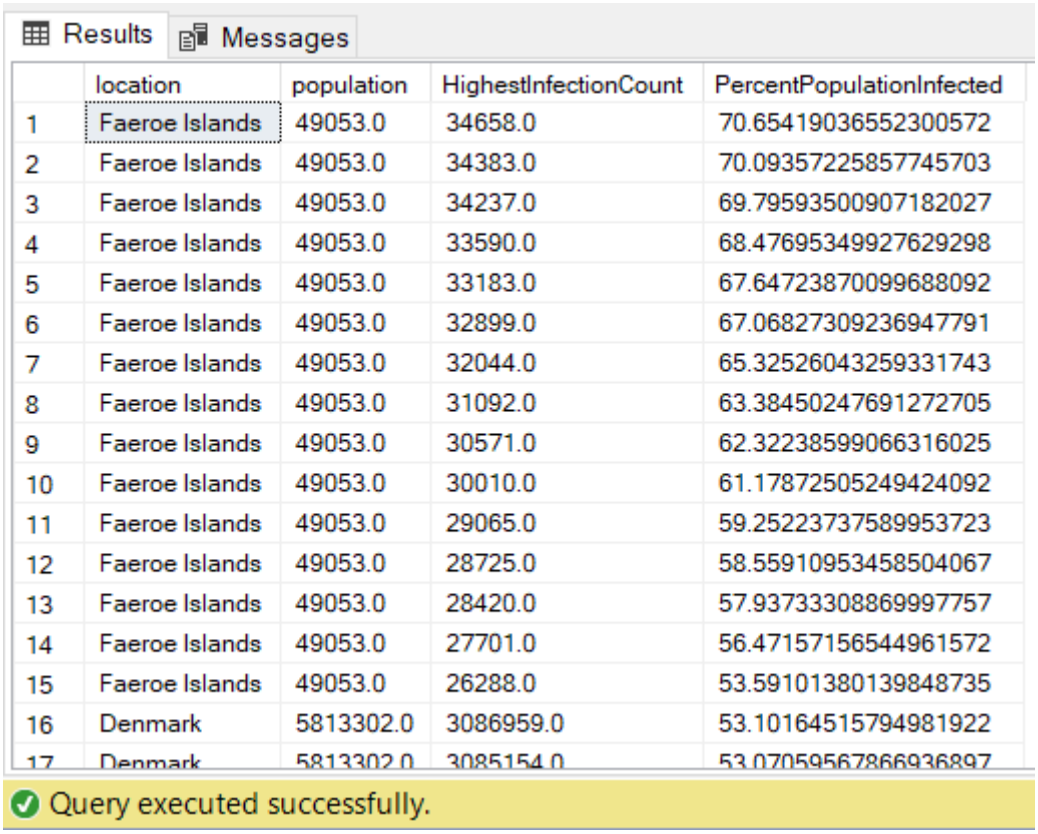
```sql
--Looking at Countries with Highest Infection Rate compared to
Population

Select location, population, MAX(total_cases) as
HighestInfectionCount, (CAST(total_cases AS
decimal(17,4))*100/population) as PercentPopulationInfected

From PortfolioProject.dbo.CovidDeaths

group by location, population, total_cases

order by PercentPopulationInfected desc
```



Results | Messages

	location	population	HighestInfectionCount	PercentPopulationInfected
1	Faeroe Islands	49053.0	34658.0	70.65419036552300572
2	Faeroe Islands	49053.0	34383.0	70.09357225857745703
3	Faeroe Islands	49053.0	34237.0	69.79593500907182027
4	Faeroe Islands	49053.0	33590.0	68.47695349927629298
5	Faeroe Islands	49053.0	33183.0	67.64723870099688092
6	Faeroe Islands	49053.0	32899.0	67.06827309236947791
7	Faeroe Islands	49053.0	32044.0	65.32526043259331743
8	Faeroe Islands	49053.0	31092.0	63.38450247691272705
9	Faeroe Islands	49053.0	30571.0	62.32238599066316025
10	Faeroe Islands	49053.0	30010.0	61.17872505249424092
11	Faeroe Islands	49053.0	29065.0	59.25223737589953723
12	Faeroe Islands	49053.0	28725.0	58.55910953458504067
13	Faeroe Islands	49053.0	28420.0	57.93733308869997757
14	Faeroe Islands	49053.0	27701.0	56.47157156544961572
15	Faeroe Islands	49053.0	26288.0	53.59101380139848735
16	Denmark	5813302.0	3086959.0	53.10164515794981922
17	Denmark	5813302.0	3085154.0	53.07059567866936897

Query executed successfully.

```sql
--Showing countries wih Highest Death Count per Population
Select location, MAX((CAST(total_deaths AS decimal(19,8)))) as
TotalDeathCount

From PortfolioProject.dbo.CovidDeaths

where continent is not null

group by location

order by TotalDeathCount desc
```

Results | Messages

	location	TotalDeathCount
1	United States	985826.00000000
2	Brazil	661576.00000000
3	India	521710.00000000
4	Russia	364506.00000000
5	Mexico	323727.00000000
6	Peru	212507.00000000
7	United Kingdom	170264.00000000
8	Italy	160863.00000000
9	Indonesia	155674.00000000
10	France	143535.00000000
11	Iran	140650.00000000
12	Colombia	139725.00000000
13	Germany	132017.00000000
14	Argentina	128233.00000000
15	Poland	115635.00000000
16	Ukraine	112459.00000000
17	Spain	103104.00000000

Query executed successfully.

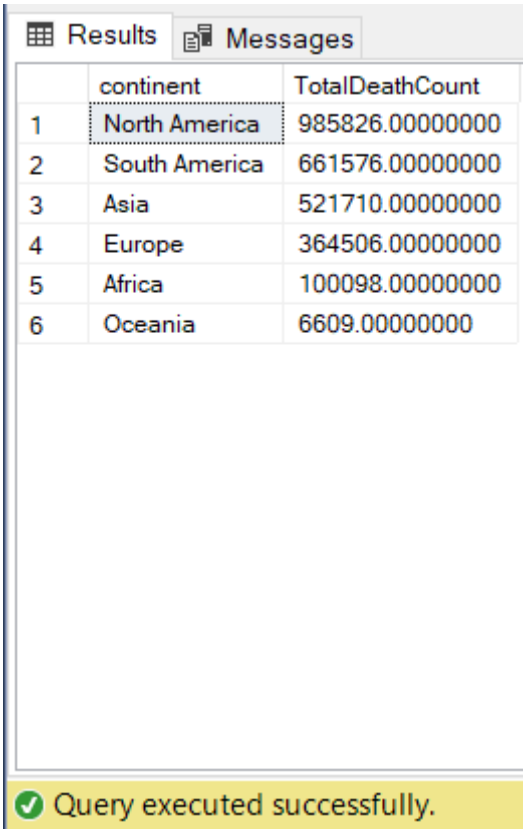
```sql
--Showing the continent with the highest death count per population

Select continent, MAX((CAST(total_deaths AS decimal(19,8)))) as
TotalDeathCount From PortfolioProject.dbo.CovidDeaths

where continent is not null

group by continent

order by TotalDeathCount desc
```



Results | Messages

	continent	TotalDeathCount
1	North America	985826.00000000
2	South America	661576.00000000
3	Asia	521710.00000000
4	Europe	364506.00000000
5	Africa	100098.00000000
6	Oceania	6609.00000000

Query executed successfully.

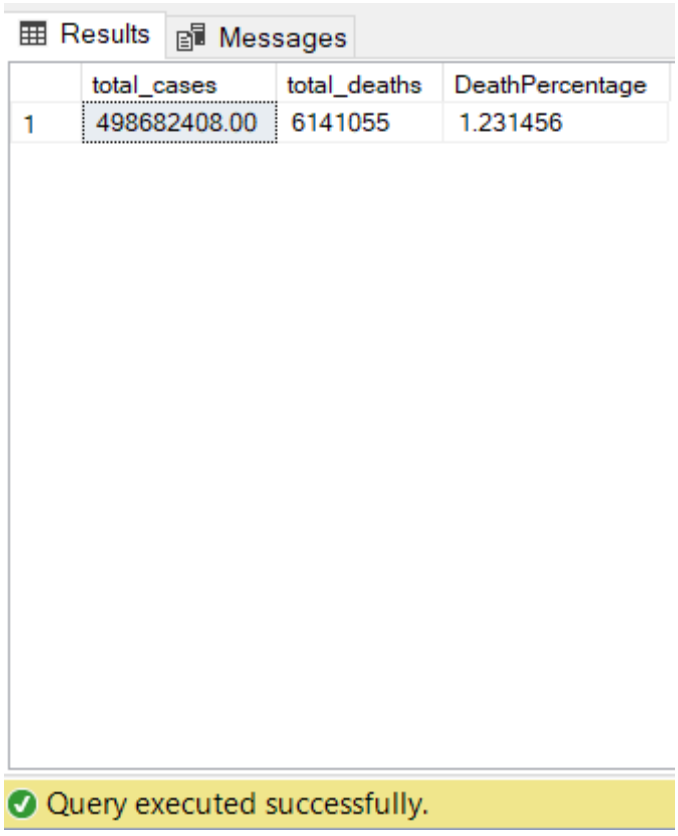
```sql
--Global Numbers

Select SUM(CAST(new_cases AS decimal(19,2))) as total_cases, SUM
(cast(new_deaths AS decimal(10,0))) as total_deaths,
SUM(cast(new_deaths AS decimal(10,0)))*100/SUM(cast(new_cases AS
decimal(10,0))) as DeathPercentage

From PortfolioProject.dbo.CovidDeaths

Where continent is not null

order by 1,2
```



```sql
--Looking at Total Population vs Vaccionations

Select dea.continent, dea.location, dea.date, dea.population,
vac.new_vaccinations
, SUM(Cast(vac.new_vaccinations AS decimal(10,2))) OVER (Partition by
dea.location order by dea.location,
dea.Date) as RollingPeopleVaccinated
from PortfolioProject.dbo.CovidDeaths as dea
join PortfolioProject.dbo.CovidVaccinations as vac
      on dea.location = vac.location
      and dea.date = vac.date

where dea.continent is not null

order by 2,3
```

Results | Messages

	continent	location	date	population	new_vaccinations	RollingPeopleVaccinated
14...	Europe	Spain	2021-01-07 00:00:00.000	46745211.0	NULL	56505.00
14...	Europe	Spain	2021-01-08 00:00:00.000	46745211.0	70653.0	127158.00
14...	Europe	Spain	2021-01-09 00:00:00.000	46745211.0	NULL	127158.00
14...	Europe	Spain	2021-01-10 00:00:00.000	46745211.0	NULL	127158.00
14...	Europe	Spain	2021-01-11 00:00:00.000	46745211.0	NULL	127158.00
14...	Europe	Spain	2021-01-12 00:00:00.000	46745211.0	82031.0	209189.00
14...	Europe	Spain	2021-01-13 00:00:00.000	46745211.0	93516.0	302705.00
14...	Europe	Spain	2021-01-14 00:00:00.000	46745211.0	94548.0	397253.00
14...	Europe	Spain	2021-01-15 00:00:00.000	46745211.0	92764.0	490017.00
14...	Europe	Spain	2021-01-16 00:00:00.000	46745211.0	NULL	490017.00
14...	Europe	Spain	2021-01-17 00:00:00.000	46745211.0	NULL	490017.00
14...	Europe	Spain	2021-01-18 00:00:00.000	46745211.0	NULL	490017.00
14...	Europe	Spain	2021-01-19 00:00:00.000	46745211.0	68155.0	558172.00
14...	Europe	Spain	2021-01-20 00:00:00.000	46745211.0	59840.0	618012.00
14...	Europe	Spain	2021-01-21 00:00:00.000	46745211.0	77364.0	695376.00
14...	Europe	Spain	2021-01-22 00:00:00.000	46745211.0	62524.0	757900.00
14...	Europe	Spain	2021-01-23 00:00:00.000	46745211.0	NULL	757900.00

Query executed successfully.

```sql
--Use CTE

With PopvsVac (Continent, Location, Date, Population, New_Vaccination,
RollingPeopleVaccinated)
as
(
Select dea.continent, dea.location, dea.date, dea.population,
vac.new_vaccinations
, SUM(Cast(vac.new_vaccinations AS decimal(10,0) ))

OVER (Partition by dea.location order by dea.location, dea.Date) as
RollingPeopleVaccinated
from PortfolioProject.dbo.CovidDeaths as dea
join PortfolioProject.dbo.CovidVaccinations as vac
      on dea.location = vac.location
      and dea.date = vac.date
where dea.continent is not null
)



Select *, (RollingPeopleVaccinated/Population)*100
From PopvsVac
```

Results | Messages

	Continent	Location	Date	Population	New_Vaccination	RollingPeopleVaccinated	(No column name)
14...	Europe	Spain	2022-01-20 00:00:00.000	4674521...	347976.0	60181016	128.742600
14...	Europe	Spain	2022-01-21 00:00:00.000	4674521...	NULL	60181016	128.742600
14...	Europe	Spain	2022-01-22 00:00:00.000	4674521...	NULL	60181016	128.742600
14...	Europe	Spain	2022-01-23 00:00:00.000	4674521...	NULL	60181016	128.742600
14...	Europe	Spain	2022-01-24 00:00:00.000	4674521...	225682.0	60406698	129.225400
14...	Europe	Spain	2022-01-25 00:00:00.000	4674521...	285076.0	60691774	129.835200
14...	Europe	Spain	2022-01-26 00:00:00.000	4674521...	238167.0	60929941	130.344700
14...	Europe	Spain	2022-01-27 00:00:00.000	4674521...	209621.0	61139562	130.793200
14...	Europe	Spain	2022-01-28 00:00:00.000	4674521...	NULL	61139562	130.793200
14...	Europe	Spain	2022-01-29 00:00:00.000	4674521...	NULL	61139562	130.793200
14...	Europe	Spain	2022-01-30 00:00:00.000	4674521...	NULL	61139562	130.793200
14...	Europe	Spain	2022-01-31 00:00:00.000	4674521...	137550.0	61277112	131.087400
14...	Europe	Spain	2022-02-01 00:00:00.000	4674521...	176208.0	61453320	131.464400
14...	Europe	Spain	2022-02-02 00:00:00.000	4674521...	143754.0	61597074	131.771900
14...	Europe	Spain	2022-02-03 00:00:00.000	4674521...	122514.0	61719588	132.034000
14...	Europe	Spain	2022-02-04 00:00:00.000	4674521...	NULL	61719588	132.034000
14...	Europe	Spain	2022-02-05 00:00:00.000	4674521...	NULL	61719588	132.034000

Query executed successfully. DESKTOP-J

```sql
--TEMP TABLE

DROP Table if exists #PercentPopulationVaccinated
Create Table #PercentPopulationVaccinated
(
Continent nvarchar(255),
Location nvarchar(255),
Date datetime,
Population numeric,
New_vaccinations numeric,
RollingPeopleVaccinated numeric
)


 Insert into #PercentPopulationVaccinated

Select dea.continent, dea.location, dea.date, dea.population,
vac.new_vaccinations
, SUM(Cast(vac.new_vaccinations AS decimal(10,2) )) OVER (Partition by
dea.location order by dea.location, dea.Date) as
RollingPeopleVaccinated

from PortfolioProject.dbo.CovidDeaths as dea
join PortfolioProject.dbo.CovidVaccinations as vac
      on dea.location = vac.location
      and dea.date = vac.date
where dea.continent is not null
order by 2,3


Select *, (RollingPeopleVaccinated/Population)*100
From #PercentPopulationVaccinated
```

Results | Messages

	Continent	Location	Date	Population	New_vaccinations	RollingPeopleVaccinated	(No column name)
34...	Europe	Spain	2021-03-13 00:00:00.000	46745211	NULL	3863782	8.2656210493947712
34...	Europe	Spain	2021-03-14 00:00:00.000	46745211	NULL	3863782	8.2656210493947712
34...	Europe	Spain	2021-03-15 00:00:00.000	46745211	NULL	3863782	8.2656210493947712
34...	Europe	Spain	2021-03-16 00:00:00.000	46745211	97323	3961105	8.4738199170819873
34...	Europe	Spain	2021-03-17 00:00:00.000	46745211	114867	4075972	8.7195499021279420
34...	Europe	Spain	2021-03-18 00:00:00.000	46745211	136278	4212250	9.0110835097096898
34...	Europe	Spain	2021-03-19 00:00:00.000	46745211	NULL	4212250	9.0110835097096898
34...	Europe	Spain	2021-03-20 00:00:00.000	46745211	NULL	4212250	9.0110835097096898
34...	Europe	Spain	2021-03-21 00:00:00.000	46745211	NULL	4212250	9.0110835097096898
34...	Europe	Spain	2021-03-22 00:00:00.000	46745211	NULL	4212250	9.0110835097096898
34...	Europe	Spain	2021-03-23 00:00:00.000	46745211	87288	4299538	9.1978149376628121
34...	Europe	Spain	2021-03-24 00:00:00.000	46745211	210897	4510435	9.6489777316439966
34...	Europe	Spain	2021-03-25 00:00:00.000	46745211	219643	4730078	10.1188504636335...
34...	Europe	Spain	2021-03-26 00:00:00.000	46745211	227635	4957713	10.6058201341737...
34...	Europe	Spain	2021-03-27 00:00:00.000	46745211	NULL	4957713	10.6058201341737...
34...	Europe	Spain	2021-03-28 00:00:00.000	46745211	NULL	4957713	10.6058201341737...
34...	Europe	Spain	2021-03-29 00:00:00.000	46745211	NULL	4957713	10.6058201341737...

Query executed successfully. DESKTOP-J871ETT

```sql
--Creating view to store data for a later visualization

Create View PercentPopulationVaccinated1 as

Select dea.continent, dea.location, dea.date, dea.population,
vac.new_vaccinations
, SUM(Cast(vac.new_vaccinations AS decimal(10,0) )) OVER (Partition by
dea.location order by dea.location, dea.Date) as
RollingPeopleVaccinated
from PortfolioProject.dbo.CovidDeaths as dea
join PortfolioProject.dbo.CovidVaccinations as vac
      on dea.location = vac.location
      and dea.date = vac.date
where dea.continent is not null



Select *
from PercentPopulationVaccinated1
```

Results | Messages

	continent	location	date	population	new_vaccinations	RollingPeopleVaccinated
14...	Europe	Spain	2022-03-06 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-07 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-08 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-09 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-10 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-11 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-12 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-13 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-14 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-15 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-16 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-17 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-18 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-19 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-20 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-21 00:00:00.000	46745211.0	NULL	63202201
14...	Europe	Spain	2022-03-22 00:00:00.000	46745211.0	NULL	63202201

Query executed successfully.