# Data Analyst Portfolio Project – Covid 19 Analysis

## Business Problem

“As a data analyst working at a news company you are asked to visualize data that will help readers understand the behavior of Covid 19 virus around the world (total cases, total deaths, population percentage infected, death percentage).

You also know that there is an interest in details about the current distribution of vaccinations around the world and the percentage of people fully vaccinated.”

## Data Collection & Table Structures

The necessary data was first put into a SQL database and afterwards used a couple of statements to select different information that are going to support our analysis.

Select \*

From PortfolioProject..CovidDeaths

order by 3,4

-- Select Data that we are going to be using

Select Location, date, total\_cases, new\_cases, total\_deaths, population

From PortfolioProject..CovidDeaths

where continent is not null

order by 1,2

-- Looking at Total Cases vs Total Deaths

Select Location, date, total\_cases, total\_deaths, (total\_deaths/total\_cases)\*100 as DeathPercentage

From PortfolioProject..CovidDeaths

Where Location = 'Portugal'

order by 1,2

-- Looking at Total Cases vs Population

-- Shows what percentage of population got Covid

Select Location, date, Population, total\_cases, (total\_cases/Population)\*100 as PercentPopulationInfected

From PortfolioProject..CovidDeaths

Where Location = 'Portugal'

order by 1,2

-- 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 continent is not null

Group by Location, Population

order by PercentPopulationInfected desc

-- Showing Countries with Highest Death Count per Population

Select Location, MAX(cast(total\_deaths as int)) as TotalDeathCount

From PortfolioProject..CovidDeaths

where continent is not null

Group by Location

order by TotalDeathCount desc

--BREAKING THINGS DOWN BY CONTINENT

-- Showing continents with the highest death count per population

Select continent, MAX(cast(total\_deaths as int)) as TotalDeathCount

From PortfolioProject..CovidDeaths

where continent is not null

Group by continent

order by TotalDeathCount desc

-- GLOBAL NUMBERS

-- Total Cases, Total Deaths and Death % By Date

Select date, 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 continent is not null

Group By date

order by 1,2

-- Total Cases, Total Deaths and Death

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 continent is not null

order by 1,2

-- Looking at Total Population vs Vaccinations

Select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

SUM(CONVERT(int, vac.new\_vaccinations)) OVER(Partition by dea.location Order by dea.location, dea.date) as RollingPeopleVaccinated

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 2,3

-- Using CTE to perform Calculation on Partition By in previous query

With PopvsVac (Continent, Location, Date, Population, New\_Vaccinations, RunningTotalVaccinations, People\_Fully\_Vaccinated, VaccinatedPeoplePercentage)

as

(

Select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

SUM(CONVERT(int, vac.new\_vaccinations)) OVER(Partition by dea.location Order by dea.location, dea.date) as RunningTotalVaccinations, vac.people\_fully\_vaccinated, (vac.people\_fully\_vaccinated/dea.population)\*100 AS VaccinatedPeoplePercentage

From PortfolioProject..CovidDeaths dea

Join PortfolioProject..CovidVaccinations vac

On dea.location = vac.location

and dea.date = vac.date

where dea.continent is not null

)

Select \*

From PopvsVac

-- Using Temp Table to perform Calculation on Partition By in previous query

DROP Table if exists #PercentPopulationVaccinated

Create Table #PercentPopulationVaccinated

(

Continent nvarchar(255),

Location nvarchar(255),

Date datetime,

Population numeric,

New\_vaccinations numeric,

RunningTotalVaccinations numeric,

People\_fully\_vaccinated numeric,

VaccinatedPeoplePercentage float

)

Insert into #PercentPopulationVaccinated

Select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

SUM(CONVERT(int, vac.new\_vaccinations)) OVER(Partition by dea.location Order by dea.location, dea.date) as RunningTotalVaccinations, CONVERT(int, vac.people\_fully\_vaccinated), (vac.people\_fully\_vaccinated/dea.population)\*100 AS VaccinatedPeoplePercentage

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 2,3

Select \*

From #PercentPopulationVaccinated

-- Creating View to store data for later visualizations

DROP view if exists PercentPopulationVaccinated

Create View PercentPopulationVaccinated as

Select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

SUM(CONVERT(int, vac.new\_vaccinations)) OVER(Partition by dea.location Order by dea.location, dea.date) as RunningTotalVaccinations, vac.people\_fully\_vaccinated, (vac.people\_fully\_vaccinated/dea.population)\*100 AS VaccinatedPeoplePercentage

From PortfolioProject..CovidDeaths dea

Join PortfolioProject..CovidVaccinations vac

On dea.location = vac.location

and dea.date = vac.date

where dea.continent is not null

-- Percentage People Vaccinated by location

Select location, MAX(VaccinatedPeoplePercentage) AS VaccinatedPeoplePercentage, MAX(CONVERT(int,people\_fully\_vaccinated)) AS PeopleVaccinated, MAX(population) AS Population

FROM PercentPopulationVaccinated

WHERE VaccinatedPeoplePercentage < 100

GROUP BY location

ORDER BY Population DESC;

-- Percentage People Vaccinated Worldwide

Select SUM(PeopleVaccinated) AS TotalPeopleVaccinated, SUM(Population) AS TotalPopulation, SUM(PeopleVaccinated) / SUM(Population) AS VaccinatedPeoplePercentage

FROM(

Select MAX(CONVERT(int,people\_fully\_vaccinated)) AS PeopleVaccinated, MAX(population) AS Population

FROM PercentPopulationVaccinated

WHERE VaccinatedPeoplePercentage < 100

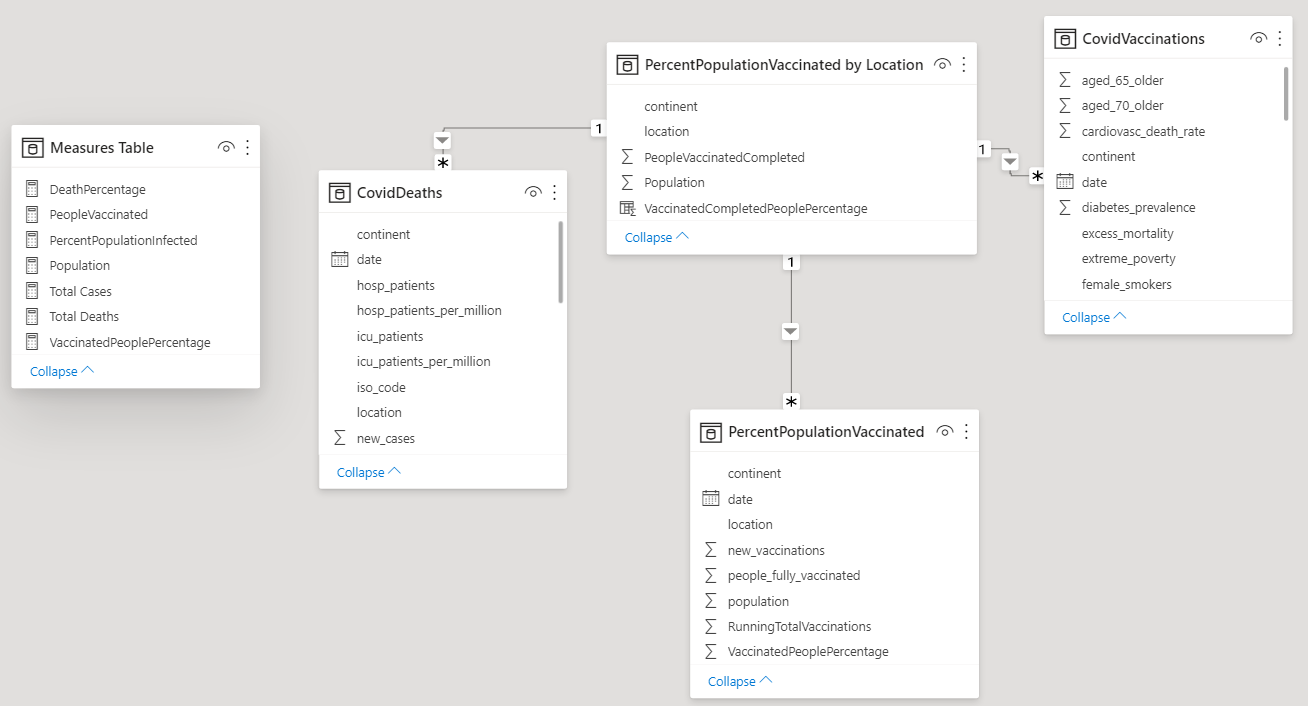
GROUP BY location)

PercentPopulationVaccinated;

## Data Model

As this is a view where dimensions and facts have been combined, the data model that is created in Power BI consists in 2 tables (*CovidDeaths* and *CovidVaccinations*) and 1 view (*PercentPopulationVaccinated*) directly from SQL. The table *PercentPopulationVaccinated by Location* was created in Power BI from *PercentPopulationVaccinated* grouping the information by Country.

Finally was created a Measures Table to specifically accommodate all the measures created in the data model.



## Calculations

The following calculations were created in the Power BI reports using DAX (Data Analysis Expressions). To lessen the extent of coding, the re-use of measures (measure branching) was emphasized:

Total Deaths = SUM(CovidDeaths[new\_deaths])

Total Cases = SUM(CovidDeaths[new\_cases])

Population = SUM('PercentPopulationVaccinated by Location'[Population])

DeathPercentage = [Total Deaths] / [Total Cases]

PeopleVaccinated = SUM('PercentPopulationVaccinated by Location'[PeopleVaccinatedCompleted])

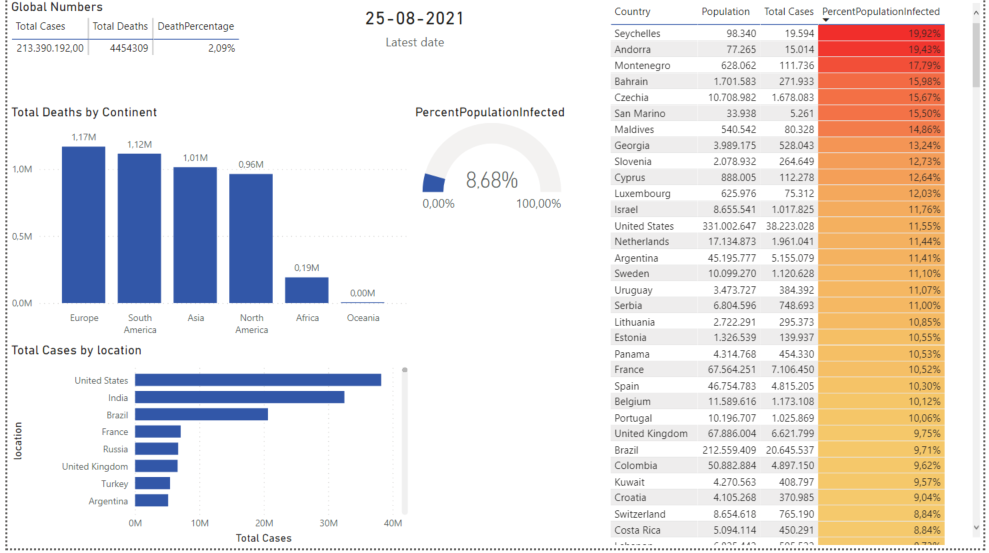
PercentPopulationInfected = [Total Cases] / MAX(CovidDeaths[population])

VaccinatedPeoplePercentage = [PeopleVaccinated] / [Population]

## Covid 19 Analysis

The finished dashboards consists of visualizations that gives an easy option for the end users to navigate the Covid 19 data regarding cases, deaths and vaccinations by location.

Click the picture to open the dashboard and try it out!

[](https://app.powerbi.com/view?r=eyJrIjoiY2Y0YjlkOGMtMDAxNC00MjJmLTg2ODEtNjA5ODIwOTg4YTQ0IiwidCI6IjBiZmE4NTAwLWIxZjItNDU2Ni1iYWYxLTZmNTkzNzA4OTNlNyIsImMiOjh9)