Project Walkthrough

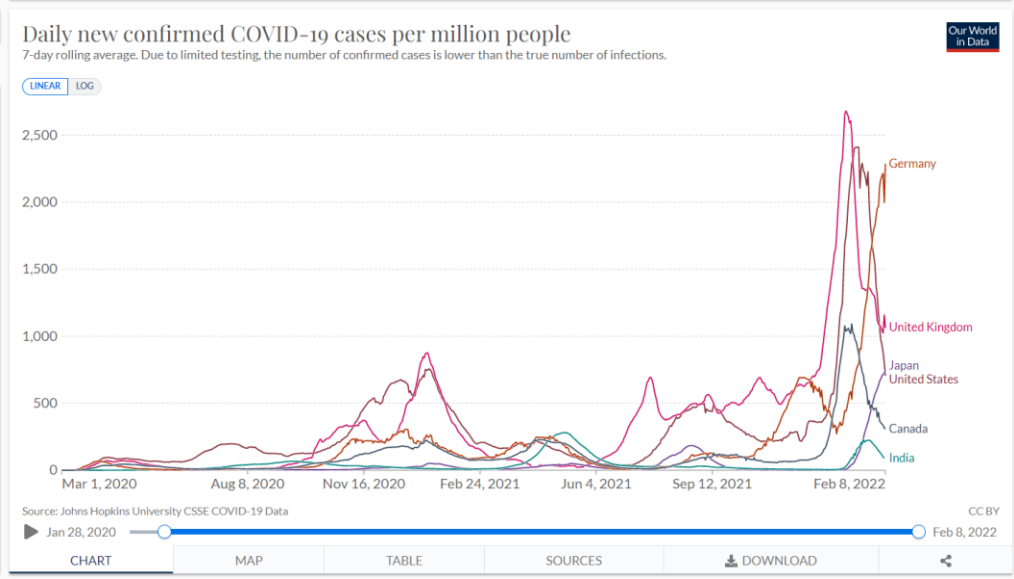
Over the past couple of years, we've been living in this pandemic that has affected the entire world. In this project we'll gather the data and dive into the statistics to find out the rate of infection for each country and number of deaths.

Tools used: Excel, SQL, & Tableau

Full SQL Script & Tableau Dashboard

Gathering and cleaning the data

1) Data is gathered from https://ourworldindata.org/coronavirus

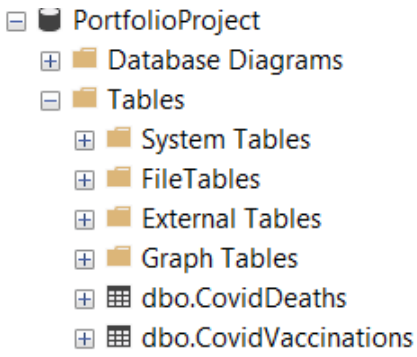


2) The downloaded datasets contain the total number of deaths and vaccinations. They are then cleaned by removing columns that are not relevant to the project.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
iso_code	continent	location	date	population	total_case	new_case	new_case	total_deat	new_deatl	new_deatl	total_case	new_case	new_case
AFG	Asia	Afghanistan	2/24/2020	39835428	5	5					0.126	0.126	
AFG	Asia	Afghanistan	2/25/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/26/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/27/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/28/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/29/2020	39835428	5	0	0.714				0.126	0	0.018
AFG	Asia	Afghanistan	3/1/2020	39835428	5	0	0.714				0.126	0	0.018
AFG	Asia	Afghanistan	3/2/2020	39835428	5	0	0				0.126	0	0
AFG	Asia	Afghanistan	3/3/2020	39835428	5	0	0				0.126	0	0
AFG	Asia	Afghanistan	3/4/2020	39835428	5	0	0				0.126	0	0

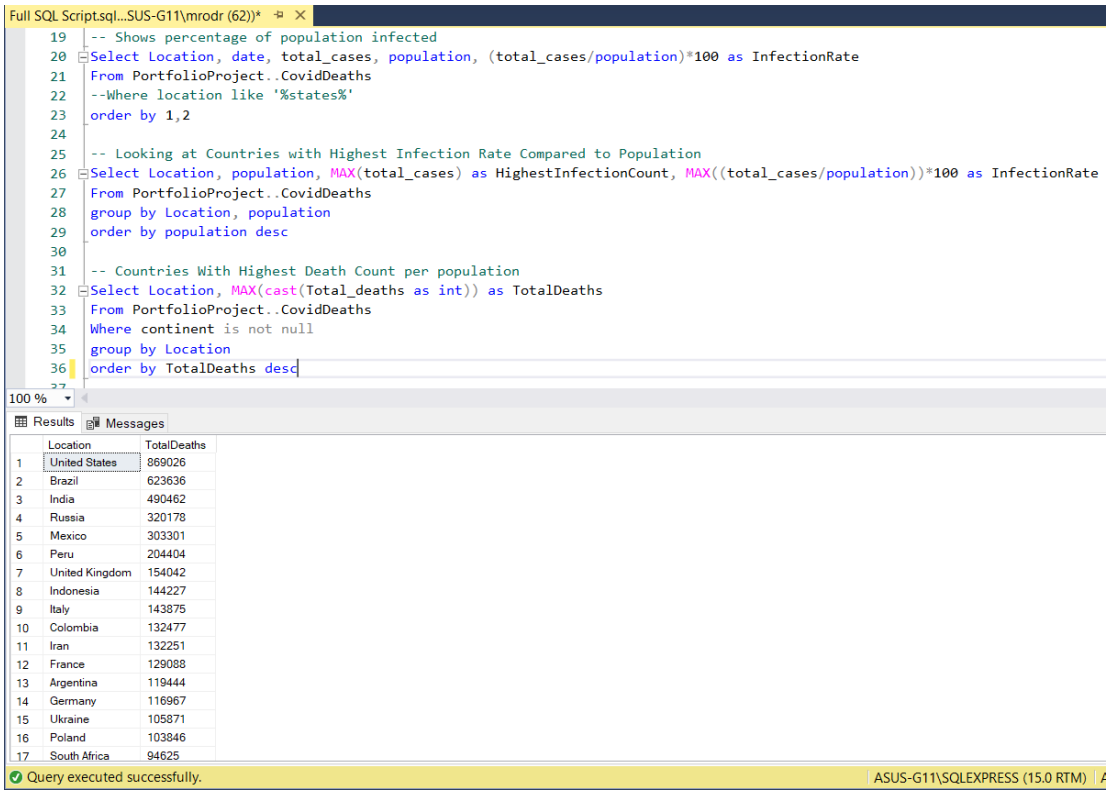
A	B	C	D	E	F	G	H	I	J	K
iso_code	continent	location	date	new_tests	total_tests	total_tests	new_tests	new_tests	new_tests	positive_ra
AFG	Asia	Afghanistan	2/24/2020							
AFG	Asia	Afghanistan	2/25/2020							
AFG	Asia	Afghanistan	2/26/2020							
AFG	Asia	Afghanistan	2/27/2020							
AFG	Asia	Afghanistan	2/28/2020							
AFG	Asia	Afghanistan	2/29/2020							
AFG	Asia	Afghanistan	3/1/2020							
AFG	Asia	Afghanistan	3/2/2020							
AFG	Asia	Afghanistan	3/3/2020							
AFG	Asia	Afghanistan	3/4/2020							

3) Data is then imported into Microsoft SQL server where it will be queried. Two tables are created in the database.

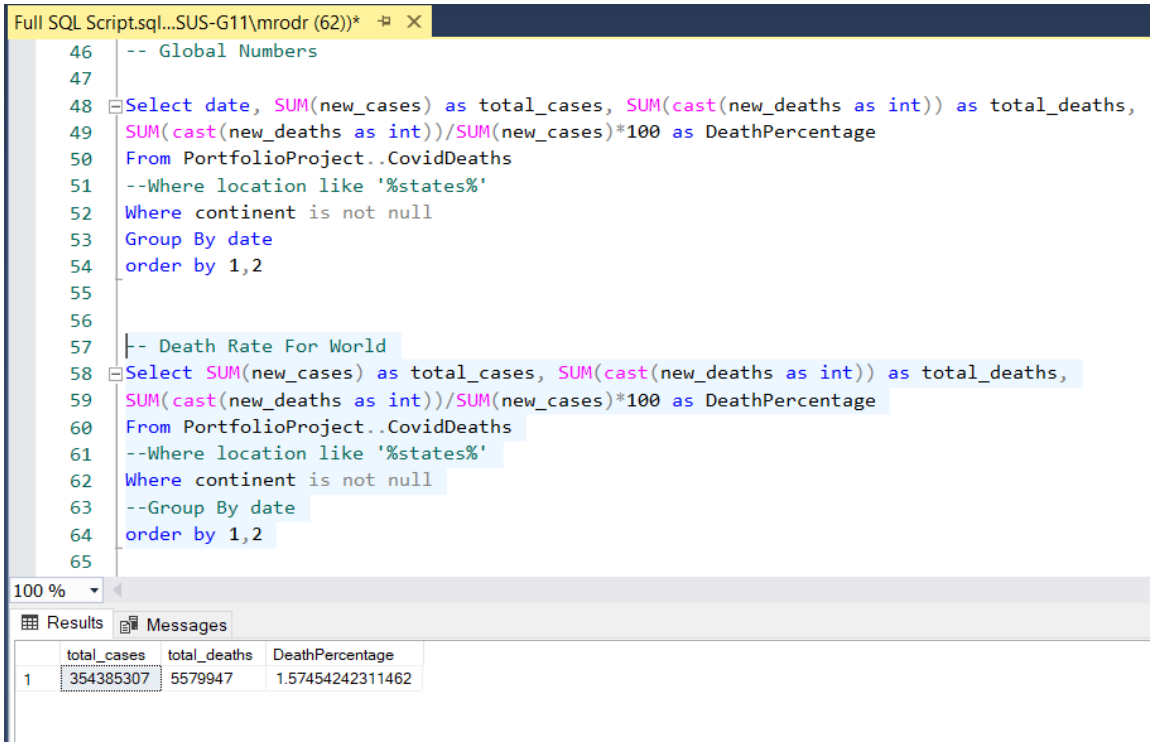


SQL Queries

1) SQL scripts are used to get different looks of the data. These scripts contain calculations to determine the percentage of the population infected and the highest death count per country.



2) The death rate for the world is calculated by creating a query that takes the total number of cases and the total number of deaths divided by the total number of cases then multiplied by 100.

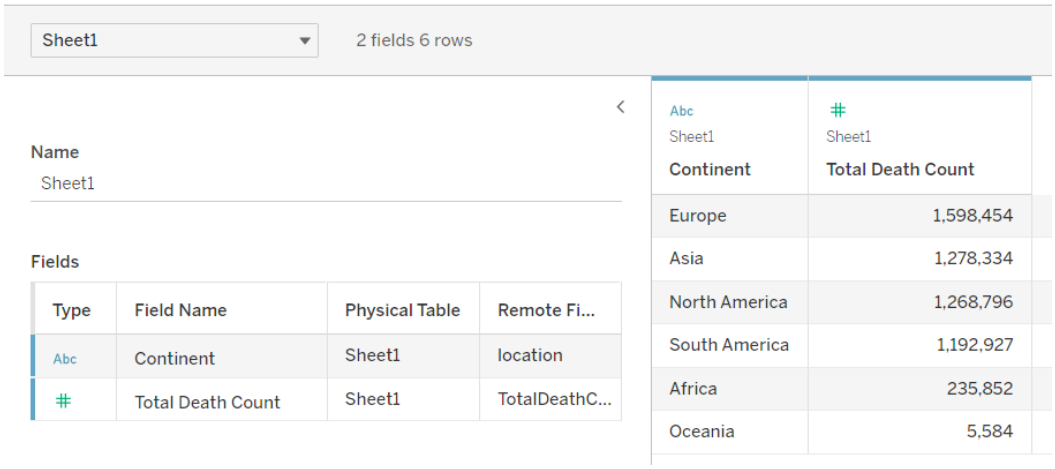


3) A view is created to store data for the visualizations that will be created in Tableau

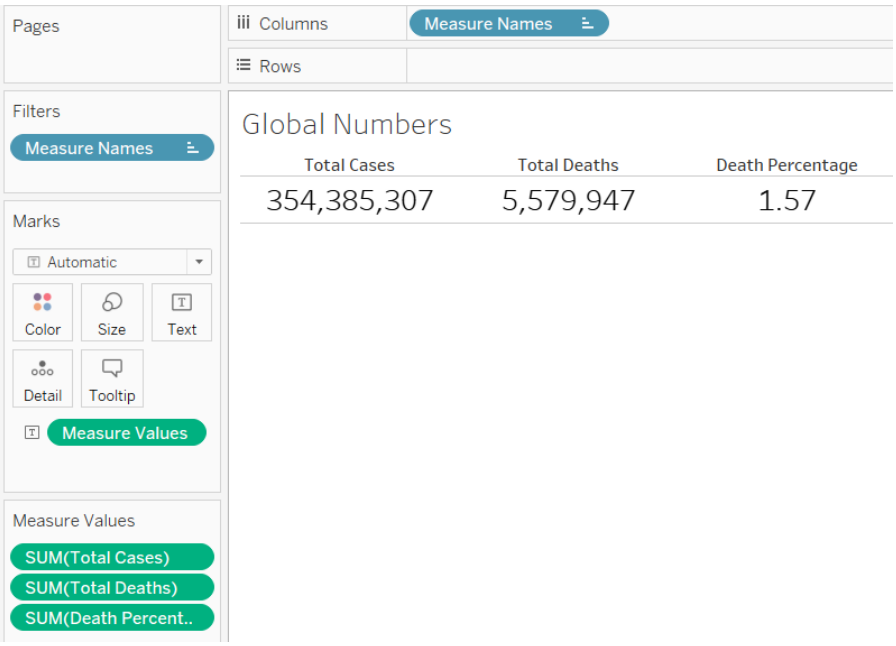
```
--Creating View to store data for visualizations
Create View DeathRateForWorldView as
Select SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths,
SUM(cast(new_deaths as int))/SUM(new_cases)*100 as DeathPercentage
From PortfolioProject..CovidDeaths
--Where location like '%states%'
Where continent is not null
Group By date
--order by 1,2

Select *
from DeathRateForWorld
```

Visuatlizations

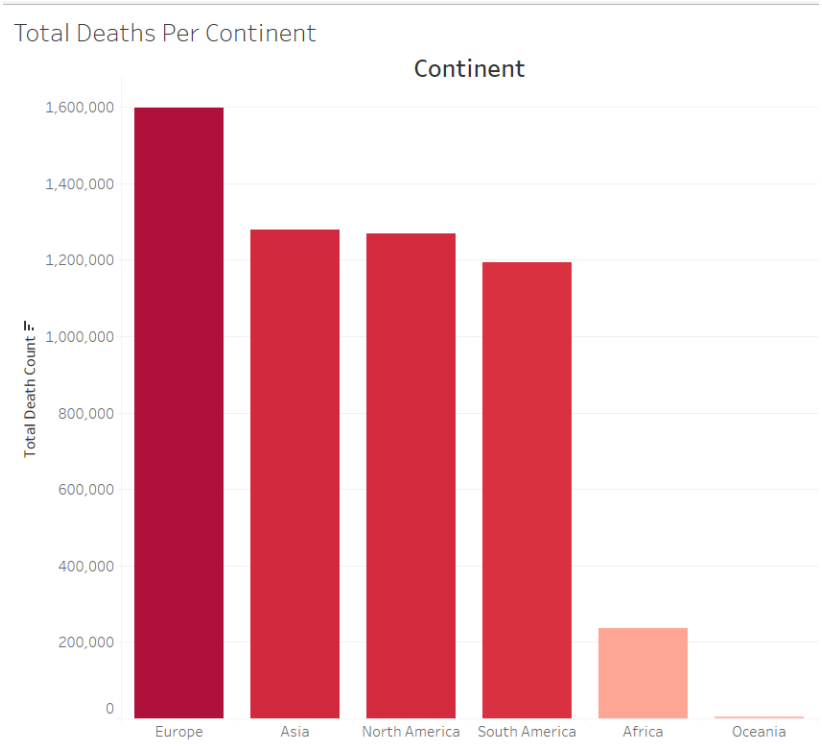
1) The data is imported into Tableau. We will be looking at the global numbers and the rate of infection delineated by country.



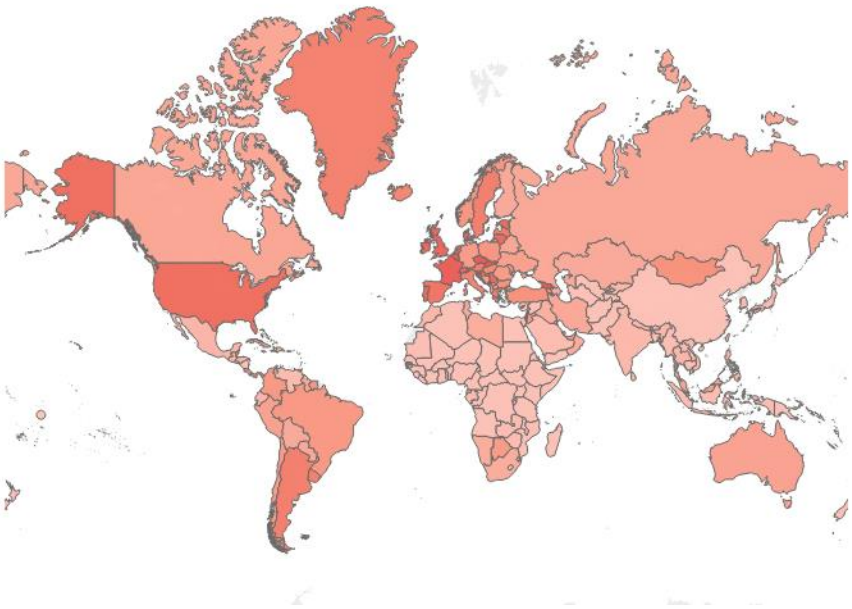
2) The first sheet contains a simple table describing the global numbers for the total number of cases, total, deaths, and the death percentage.



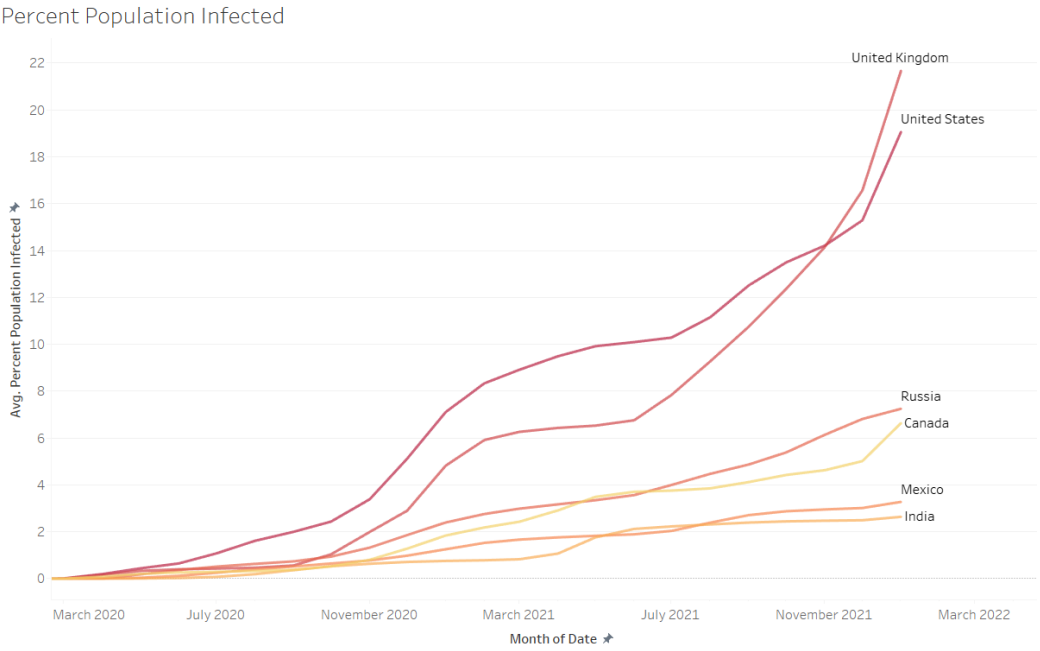
3) The 2nd sheet is a bar chart that is easy to digest. It shows the number of deaths per continent with Europe being the highest at 1,598,454 deaths.



4) The third sheet is an interactive map of the world where you can hover over each country and see the rate of infection. The darker countries have the highest infection rate for its population.



5) The last sheet contains the average percent of the population infected per month since the beginning of the pandemic. This is also a comparison between 6 countries. These countries are Canada, India, Mexico, Russia, the United Kingdom, and the United States. We can see that the US and the UK have the highest numbers.



Dashboard

The final piece of the project is an interactive dashboard containing all the previous sheets. You can see the full visualization Here