**Capstone Project #1 Covid 19 Data**

**Scenario**

You are a junior data analyst working in the marketing analyst team at the World Health Organization (WHO). You have been at your job for six months, and your boss feels you are ready for more responsibility. He has asked you to lead a project detailing analyzing global pandemic data — this will involve everything from defining the business task all the way through presenting your data-driven recommendations. You will choose the topic, ask the right questions, identify a fresh dataset and ensure its integrity, conduct analysis, create compelling data visualizations, and prepare a presentation.

***Question #3:*** *How did it look in Japan compared to US for percent deaths.*

The data is provided by Johns Hopkins University and shared on the following link: <https://ourworldindata.org/covid-deaths>

The data is organized by total\_cases, new\_cases, total\_deaths, new\_deaths, total vaccinations, people vaccinated, population, etc. IT’S A LOT OF DATA!!

The authenticity of this data is based on the accuracy/willingness of governments reporting their information.

**Data Cleaning**

I took the “population” data from the “covid deaths” sheet and pasted it next to the total\_deaths chart. Easier to reference and would’ve been difficult to use a bunch of joins. Then I deleted all of the data related to covid cases that populated in the “covid deaths” spreadsheet. I saved a copy of this spreadsheet and called it “Covid vaccinations”. Then I deleted all of the Covid Vaccination info and called it “Covid Deaths”. Review this part in the video and my spreadsheets before posting online.

***SQL Server doesn’t read csv files when importing excel documents. Only reads xlsx!! Didn’t know that when I was trying to import data initially.***

431,716 rows of data combined between the two sets.

Notes:

--SELECT \*

--FROM Case\_Study\_#1.dbo.Covid\_Deaths$

--WHERE location = 'Japan'

--SELECT \*

--FROM Case\_Study\_#1.dbo.Covid\_Deaths$

--WHERE location = 'United States'

I used the above query to look at Japan and the United States early responses to Covid out of curiosity. Will expand upon this later.

To look at the death percentages between Japan and the United States:

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

--FROM Case\_Study\_#1.dbo.Covid\_Deaths$

--WHERE location = 'United States'

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

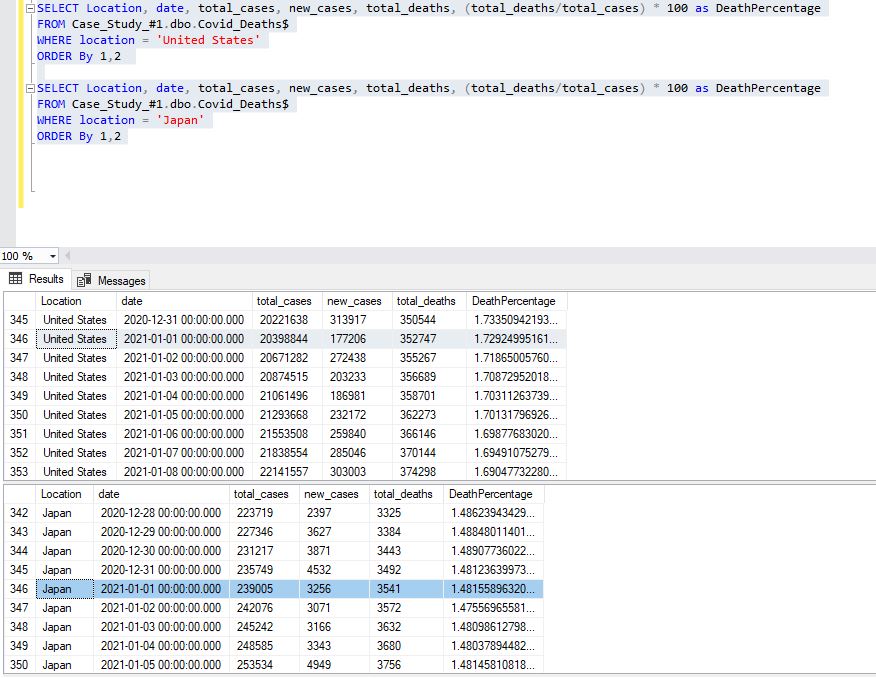
--FROM Case\_Study\_#1.dbo.Covid\_Deaths$

--WHERE location = 'Japan'

Looking at the above Query, if you look at year end. United States had a 1.7% death rate for Covid while Japan had a 1.4%.

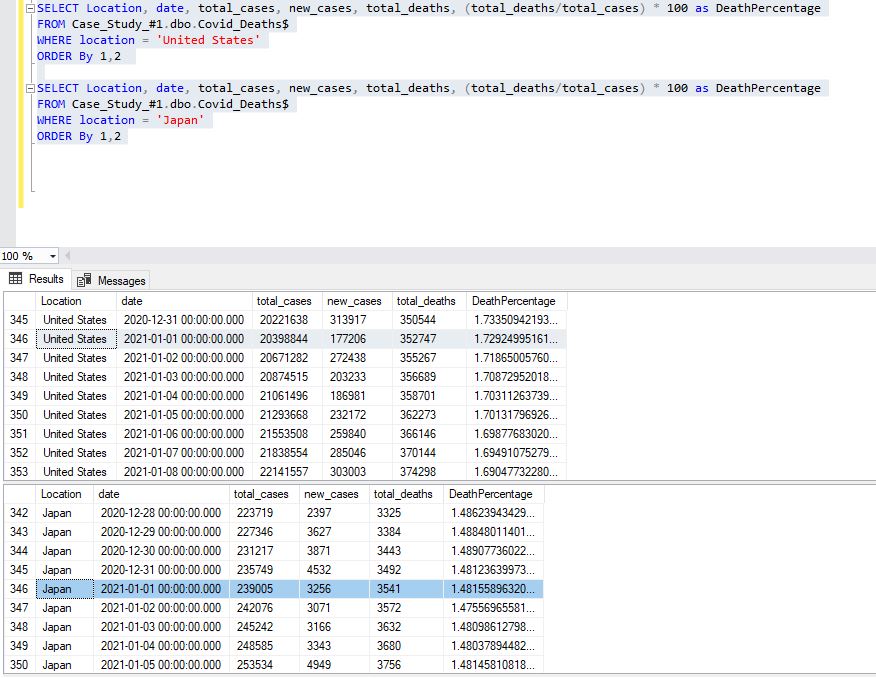
By this point, the United States in total lost 352,747 people

Japan lost 3541 people.



***I could totally look at the death rate for vaccinated and fully vaccinated people too. Once I’m more familiarized with Joins.***

By the holiday season, Japan had a .19% of its population testing positive for covid. While the United States had 6% on the same date (New Years Day)



Code written below

SELECT Location, date, total\_cases, new\_cases, population, total\_deaths, (total\_deaths/total\_cases) \* 100 as DeathPercentage, (total\_cases/population) \* 100 as Population\_Had\_Covid

FROM Case\_Study\_#1.dbo.Covid\_Deaths$

WHERE location = 'United States'

ORDER By 1,2

SELECT Location, date, total\_cases, new\_cases, population, total\_deaths, (total\_deaths/total\_cases) \* 100 as DeathPercentage, (total\_cases/population) \* 100 as Population\_Had\_Covid

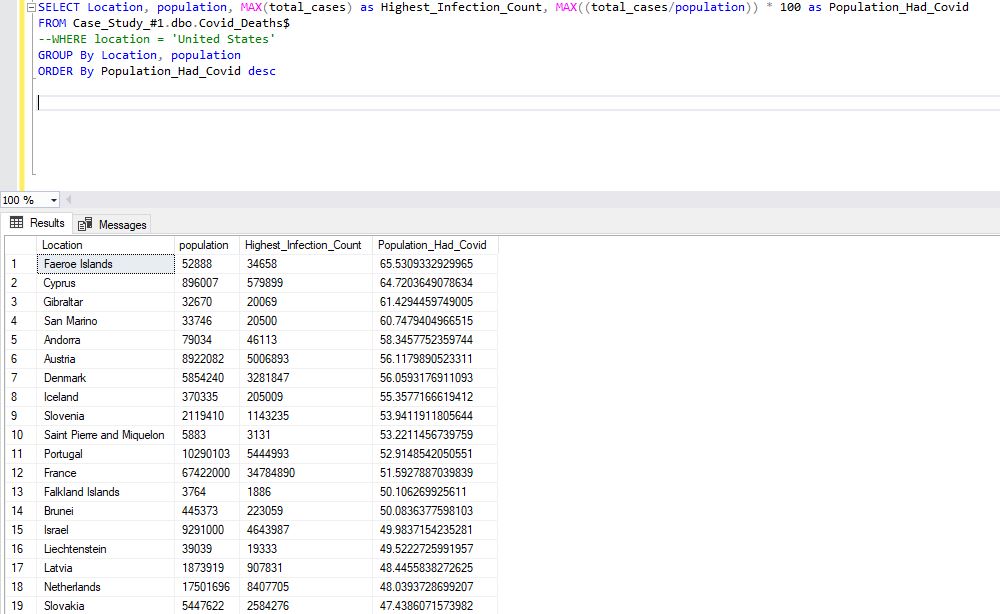
FROM Case\_Study\_#1.dbo.Covid\_Deaths$

WHERE location = 'Japan'

ORDER By 1,2

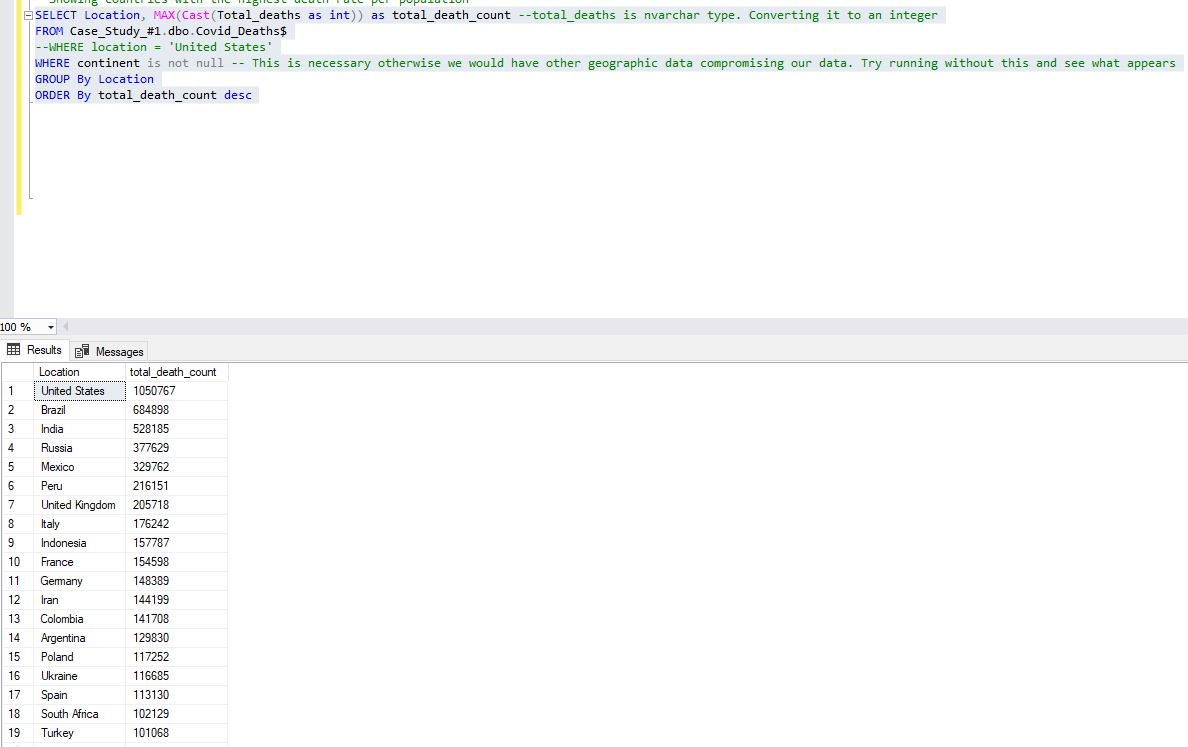
***Could use this as a visualization example!!***

**Next we will look at countries with the highest infection rate compared to the population.**

****

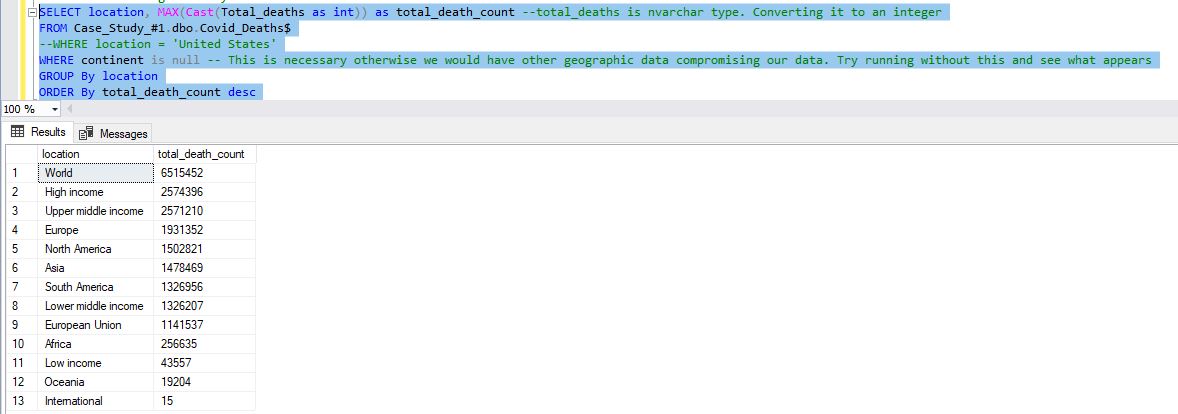
The United States ranks at 57 (28%) of the population which has tested positive. While Japan ranked at 90 with 16.2% of the population having tested positive. As of currently (9/14) the US has had a registered 95 million covid cases. While Japan has had 20 million covid cases.

**Next I will be showing the countries with the highest death rate per population.**

**At the time of which I downloaded the data 9/13/2022, AS FAR AS REGISTERED DEATHS GO. The United States ranked #1.** 



Meanwhile Japan was #26 for deaths from Covid.



**Global Covid Death Rate**

SELECT SUM(new\_cases) as Global\_Cases, SUM(CAST(new\_deaths as int)) as Global\_Deaths, SUM(Cast(new\_deaths as int))/SUM(new\_cases) \* 100 as Death\_Percentage-- total\_deaths, (total\_deaths/total\_cases) \* 100 as DeathPercentage

FROM Case\_Study\_#1.dbo.Covid\_Deaths$

WHERE continent is not null

--GROUP By date

ORDER By 1,2

We had 607,945,608 total cases. 6477416 deaths at a 1% death rate.

For the United States:

SELECT SUM(new\_cases) as Global\_Cases, SUM(CAST(new\_deaths as int)) as Global\_Deaths, SUM(Cast(new\_deaths as int))/SUM(new\_cases) \* 100 as Death\_Percentage

FROM Case\_Study\_#1.dbo.Covid\_Deaths$

WHERE location = 'United States'

ORDER By 1,2

For the United States we had 95,320,070 total cases. 1,051,300 deaths at a 1.10% death rate which is higher than the global average.

For Japan:

SELECT SUM(new\_cases) as Global\_Cases, SUM(CAST(new\_deaths as int)) as Global\_Deaths, SUM(Cast(new\_deaths as int))/SUM(new\_cases) \* 100 as Death\_Percentage

FROM Case\_Study\_#1.dbo.Covid\_Deaths$

WHERE location = 'Japan'

ORDER By 1,2

For Japan they had 20209161 total cases. 42803 deaths at a .21% death rate which is significantly lower than the global average.