**Data Analyst Portfolio Project | SQL Data Exploration | Project 1/4**

1. Descargamos el dataset

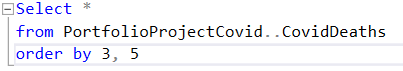
<https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-data.csv>

1. Como es un CSV (separados por comas) vamos a pasarlo a tablas/columnas. Datos -> Obtener datos -> de texto/csv
2. Si nos damos cuenta, la tabla es muuuy grande. Por eso, vamos a achicarla para este ejercicio. Primero vamos a cortar “population” y ponerla al lado de “date” (fecha). Despues, vamos a borrar desde “new\_tests” hacia la derecha.
3. Teniendo los dos datasets, vamos a guardarlos como libro de Excel no como csv.
4. Los importamos a SQL. **Aclaración:** Office debe ser 32 bits sino no anda.
5. Hacemos nuestra **Primera query**. Vamos a *new query*.
6. 

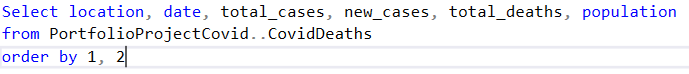
O puede ser



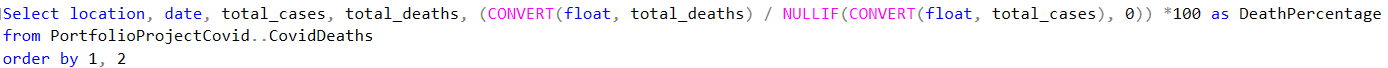
1. Puede que acá esté ordenando por país y población



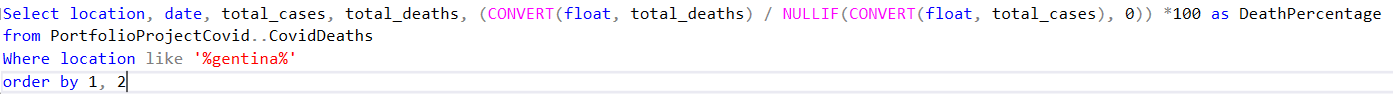
1. Seleccionamos los datos que queremos usar

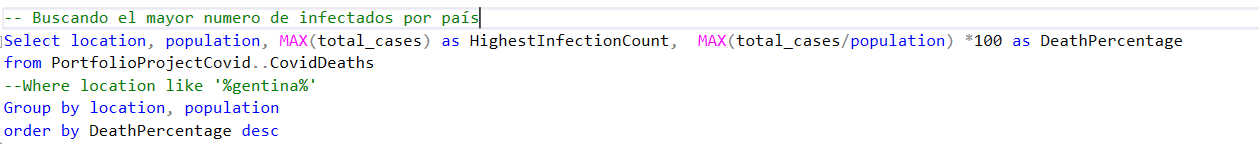
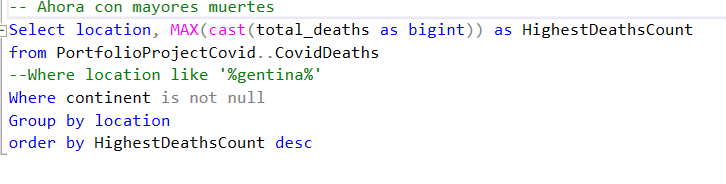


1. Vamos a calcular el porcentaje de muertos por total de casos. Hicimos una conversión a float y también para que el divisor no sea 0.

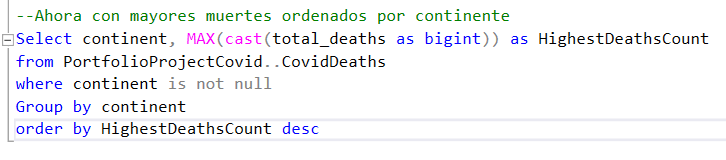


1. Hacemos el porcentaje pero usamos el like para buscar a Argentina

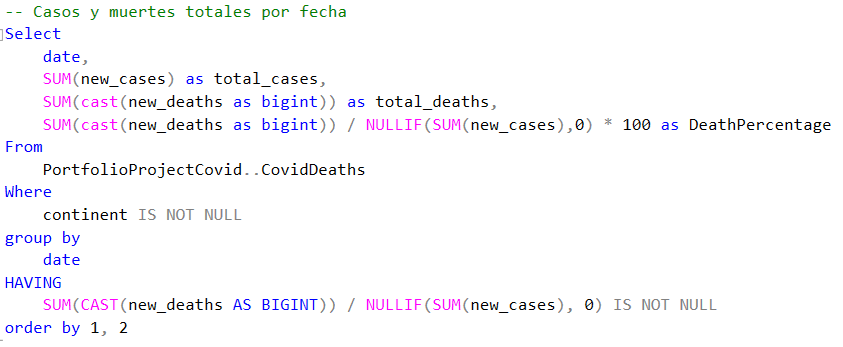


1. 
2. 

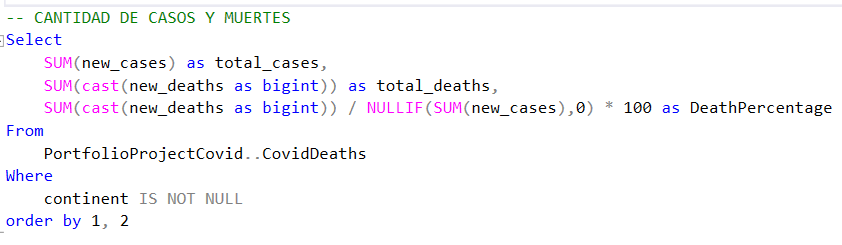
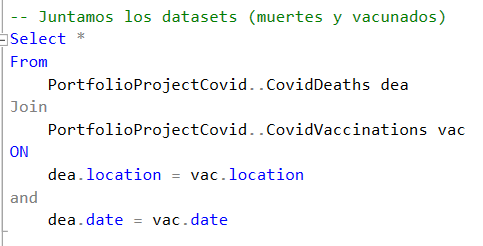
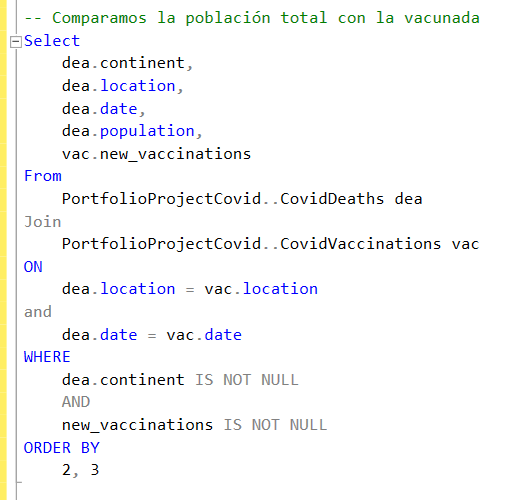
Ponemos continent is not null porque nos aparece WORLD, ASIA y no queremos continentes, sino países.

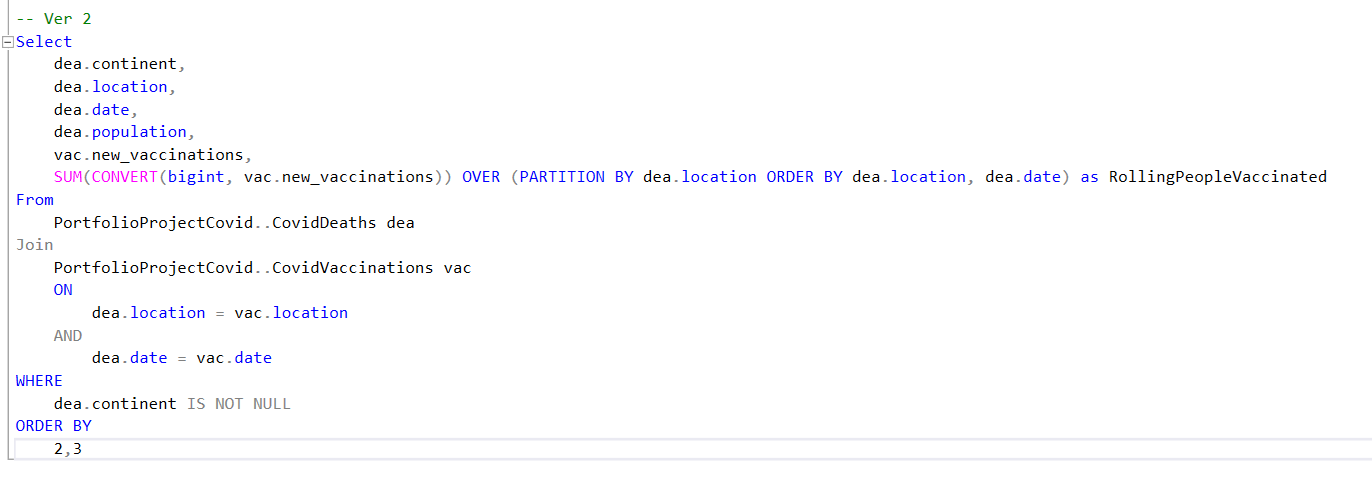
1. 

**Pero no es EFICIENTE.** Ya que North America es solamente EEUU. Y no hay otros países que no tienen catalogado un continente.

1. 

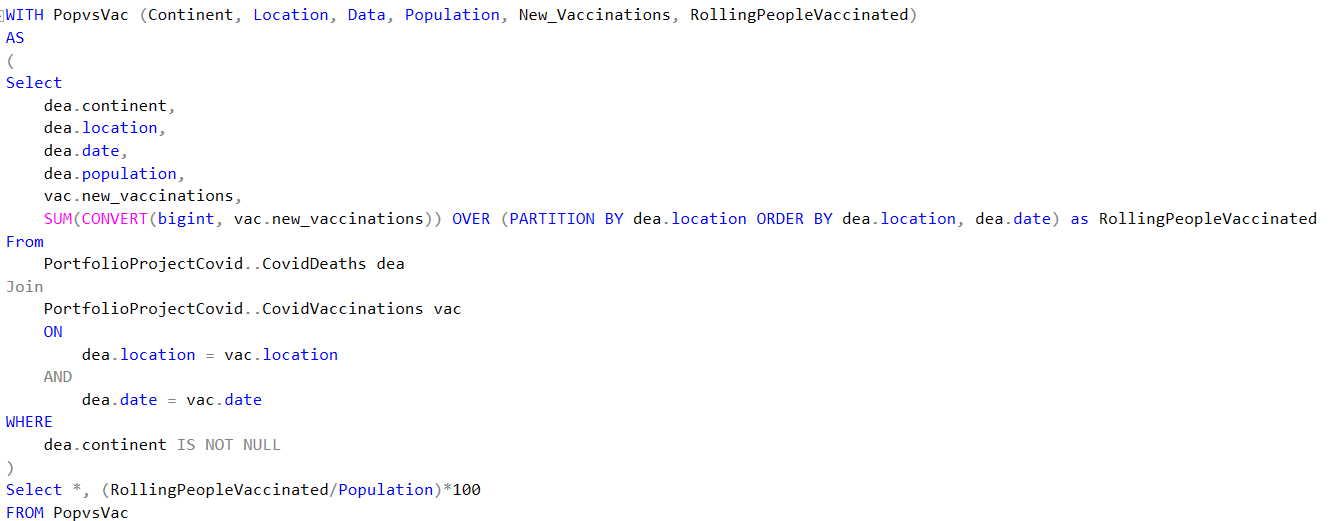
Usamos el **Having** para filtrar los días que no hubieron casos nuevos ni muertes

1. 
2. 
3. 



Lo que hacemos con **OVER PARTITION** Es hacer un conteo hasta que cambie de location. Y el order es importante para que haga la suma ordenadamente.

1. 



1. 