# Defining the Question (The Problem Statement).

Covid-19 had a huge impact in the world. In this exploratory project, we attempt to find out the countries that were the most affected by it. Further, the study tries to find out the following:

* Which country (having more than a million residents) had the highest percentage of citizens infected by covid?
* Highest death by country and continent.
* Countries with the highest death rate
* Countries with the highest vaccination %

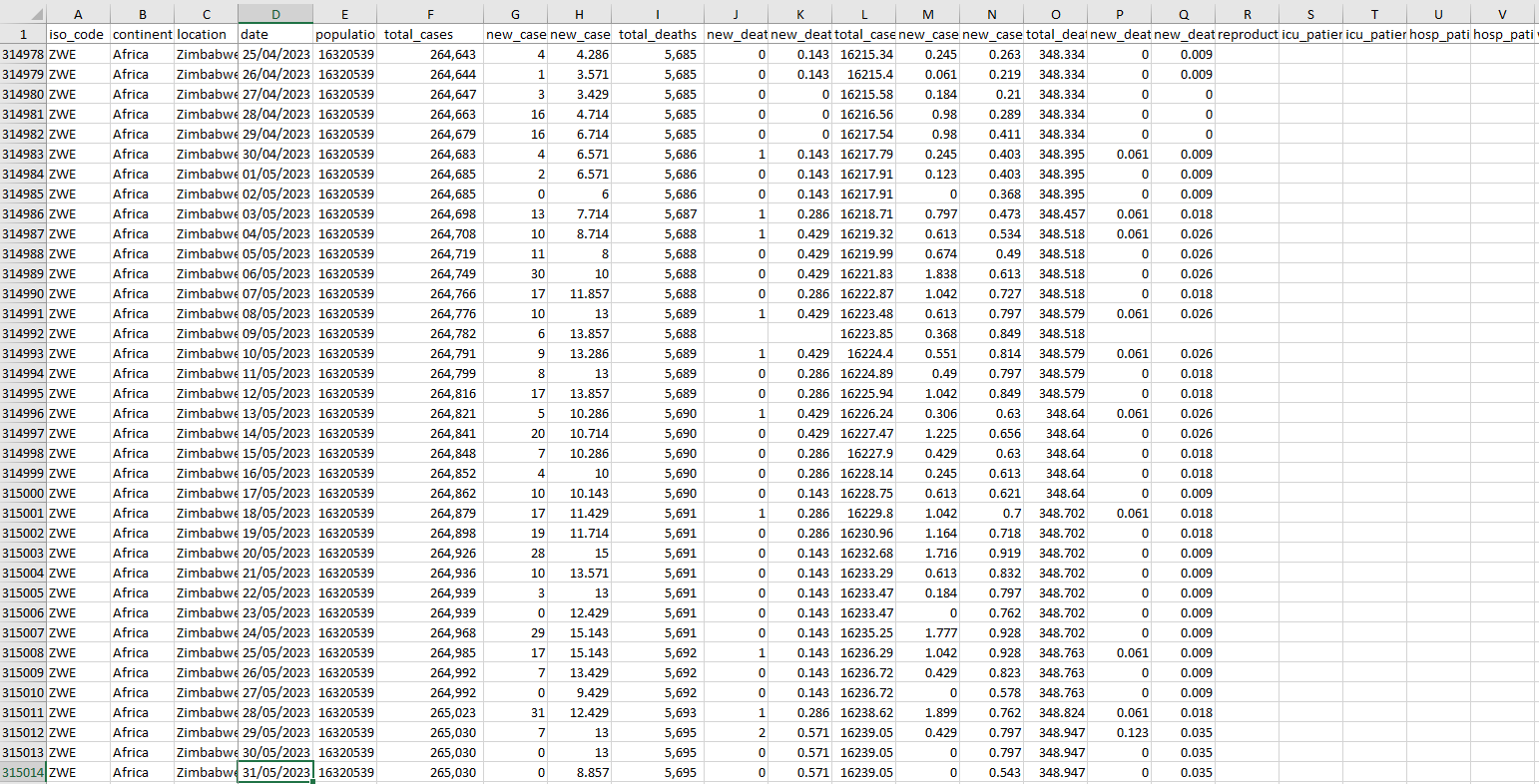
# Collecting the Data.

Data was collected from <https://ourworldindata.org/covid-deaths>

**Time period of the study:** 3rd Jan 2020 to 31st May 2023

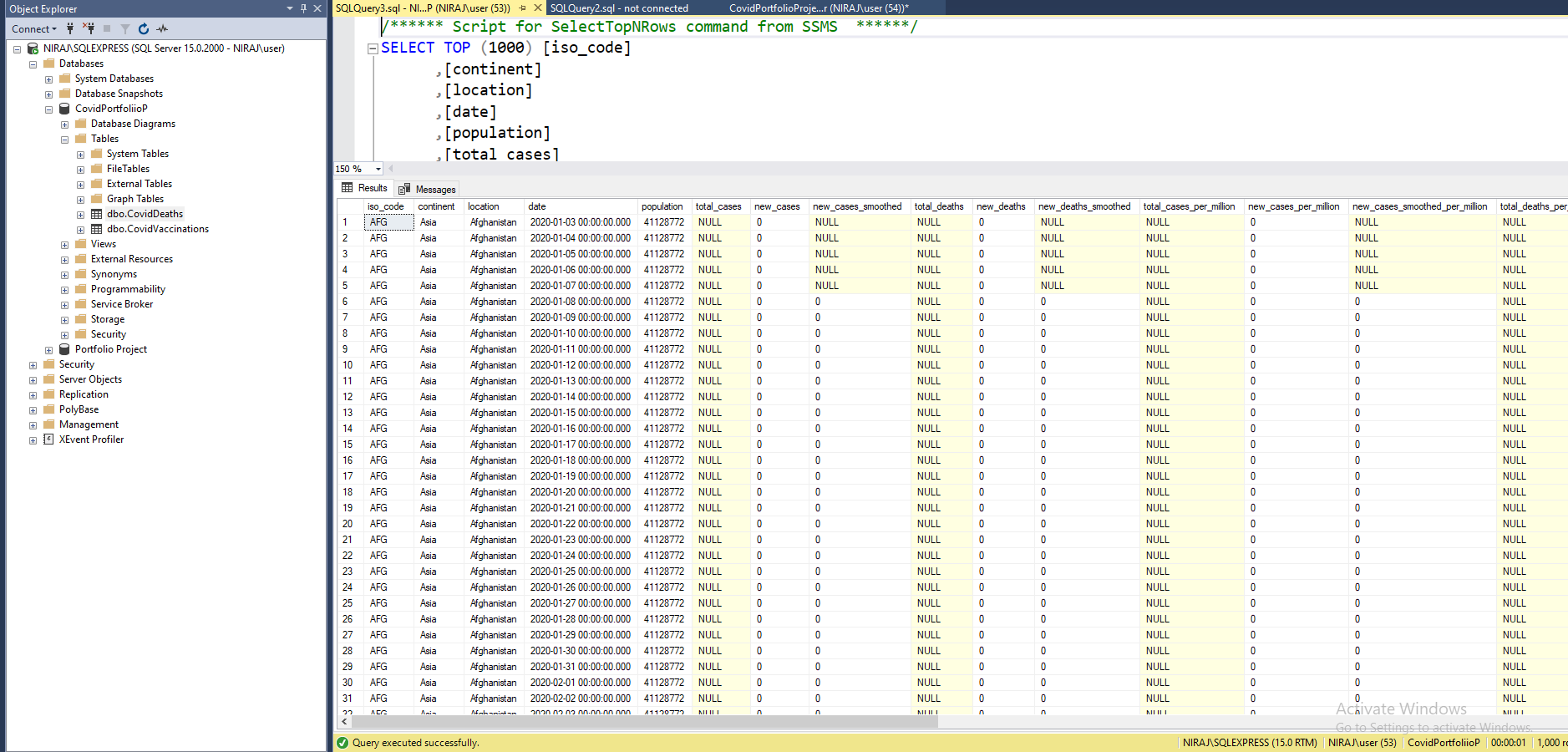
**Background**: The above website collects the data related to covid-19 deaths and vaccinations over the mentioned period.

Following is a snapshot of the data collected in excel files:



There are more than 300k rows in the excel file. The excel file is further divided into two files - CovidDeaths and CovidVaccinations and loaded in Microsoft SQL Server Management Studio.

CovidDeaths in SQL:



CovidVaccinations in SQL:

A screenshot of a computer

Description automatically generated

# Cleaning the Data.

# Analyzing the Data using SQL.

1. Total Cases vs Total Deaths

A screenshot of a computer

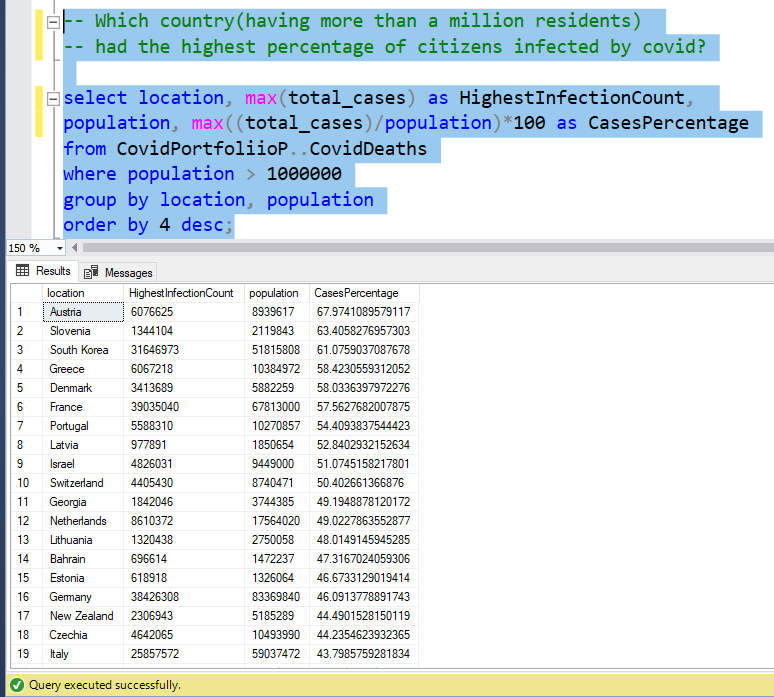
Description automatically generated with medium confidence

1. Total Cases vs Population of a Particular Country

A screenshot of a computer

Description automatically generated

1. Country having highest covid infection?



1. Countries with highest death count?

A screenshot of a computer

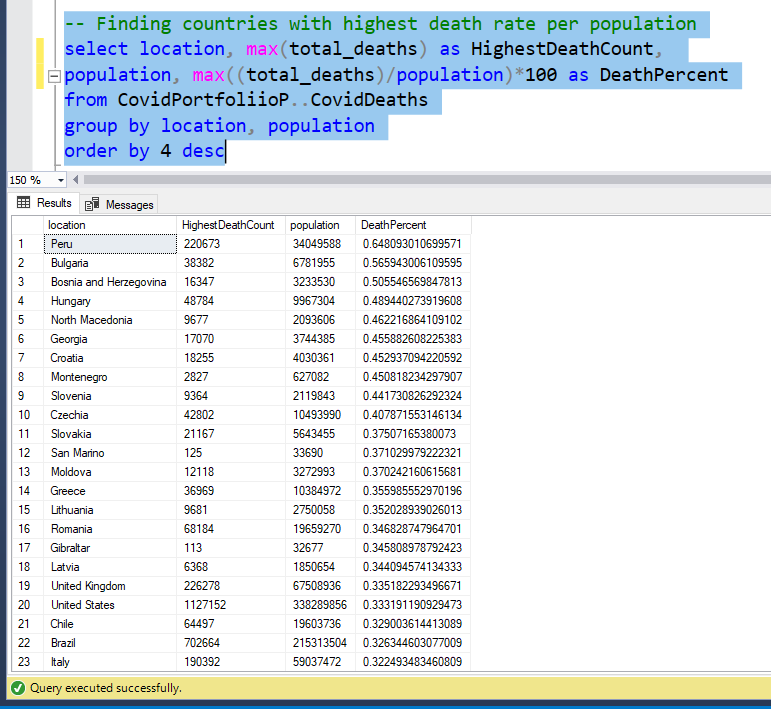
Description automatically generated

1. Highest Death count By Continent

A screenshot of a computer

Description automatically generated

1. Countries with highest death rate per population?

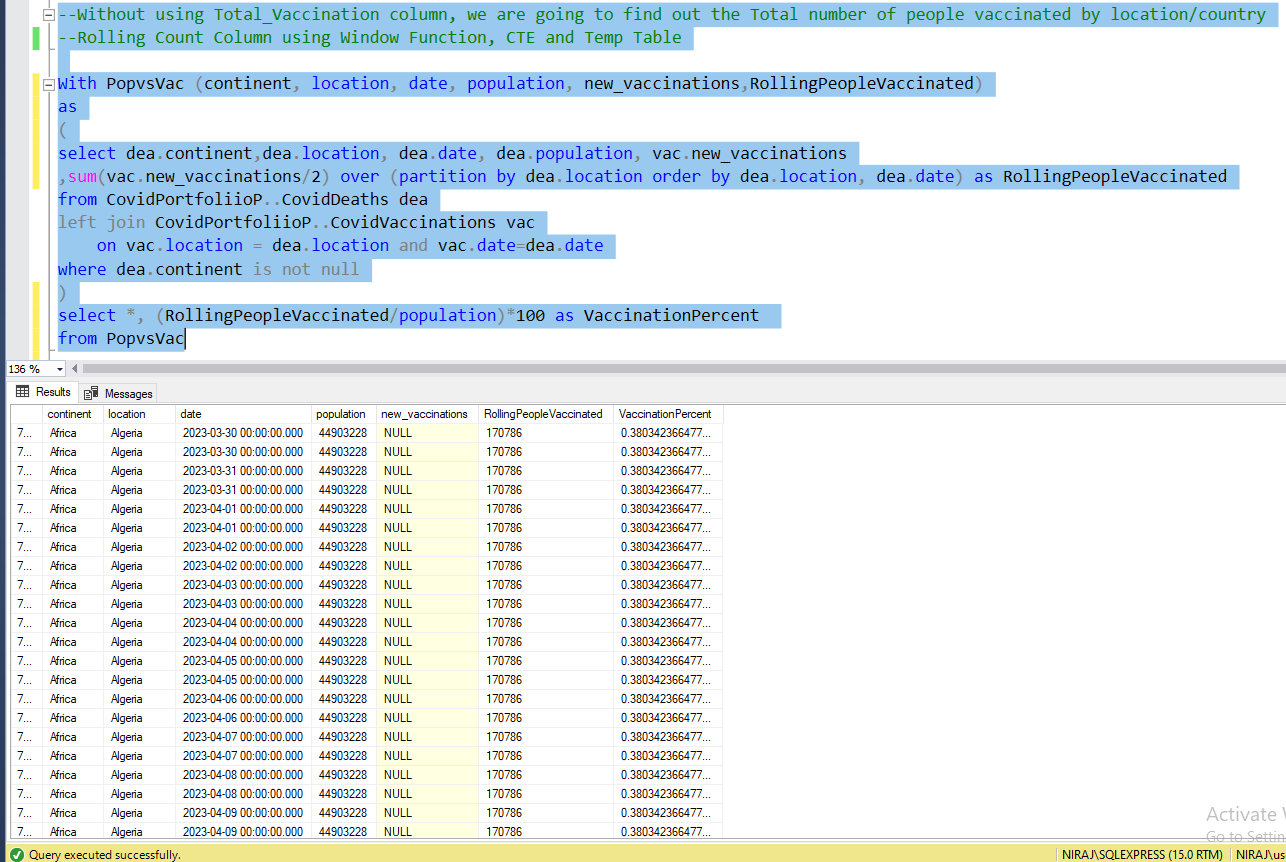


1. Joining deaths and vaccination tables on locations and dates

A screenshot of a computer

Description automatically generated with medium confidence

1. Total number of people vaccinated by location/country (Using Window function, CTE, and Temp Table)



# Visualizing the Data.

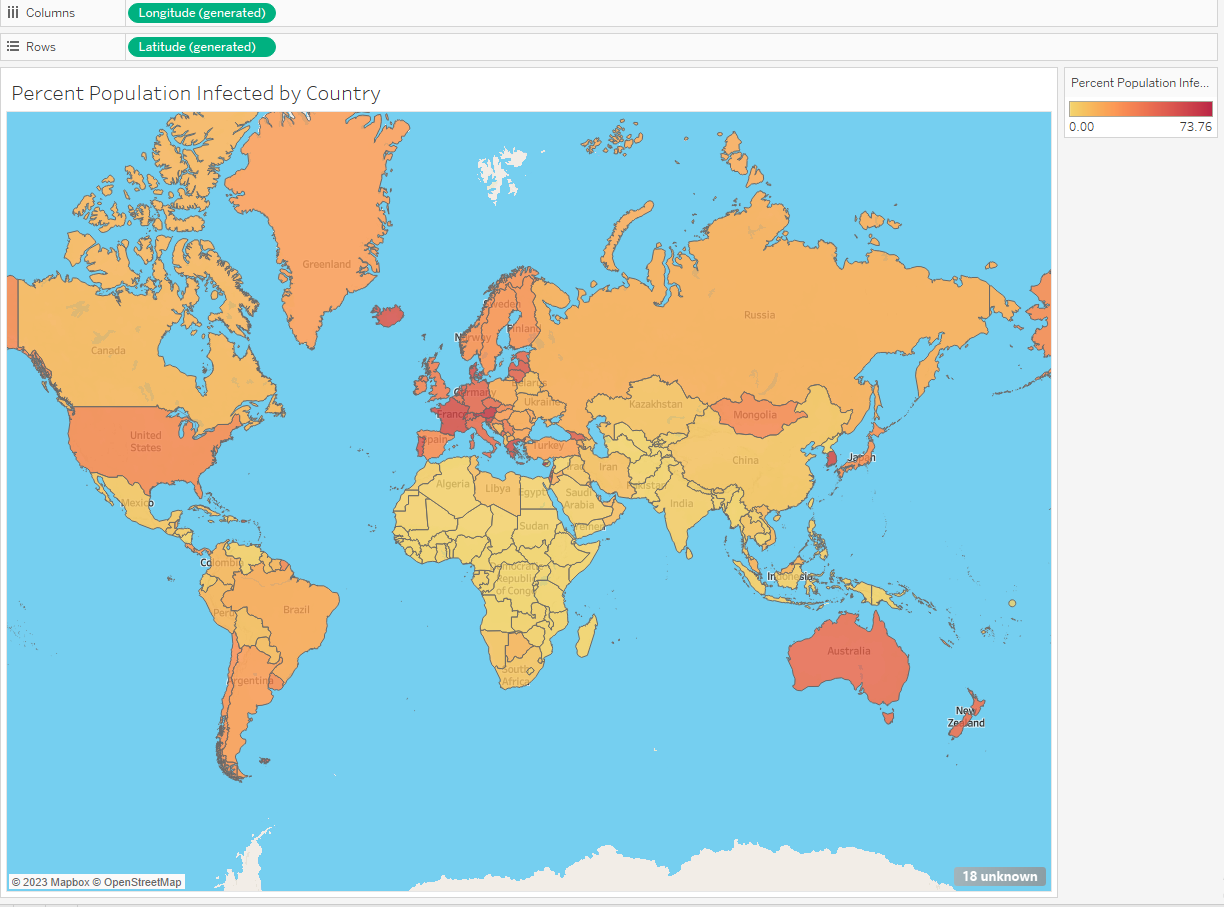
Tableau was used to visualize the data collected from SQL.

Most covid deaths were reported in North America 1.12 million, followed by South America 0.7 million, and Assia 0.53 million. Least deaths were reported in Oceania.

A screenshot of a graph

Description automatically generated

Similarly, we can also see percentage of people affected by covid by country. Australia, USA, and European countries were highly affected whereas the countries in Africa were affected the least.



Dashboard:

A screenshot of a graph

Description automatically generated

# Future References:

This project can be used by organizations like WHO and UN for efficiently allocation resources related to health and vaccination.