# 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 %

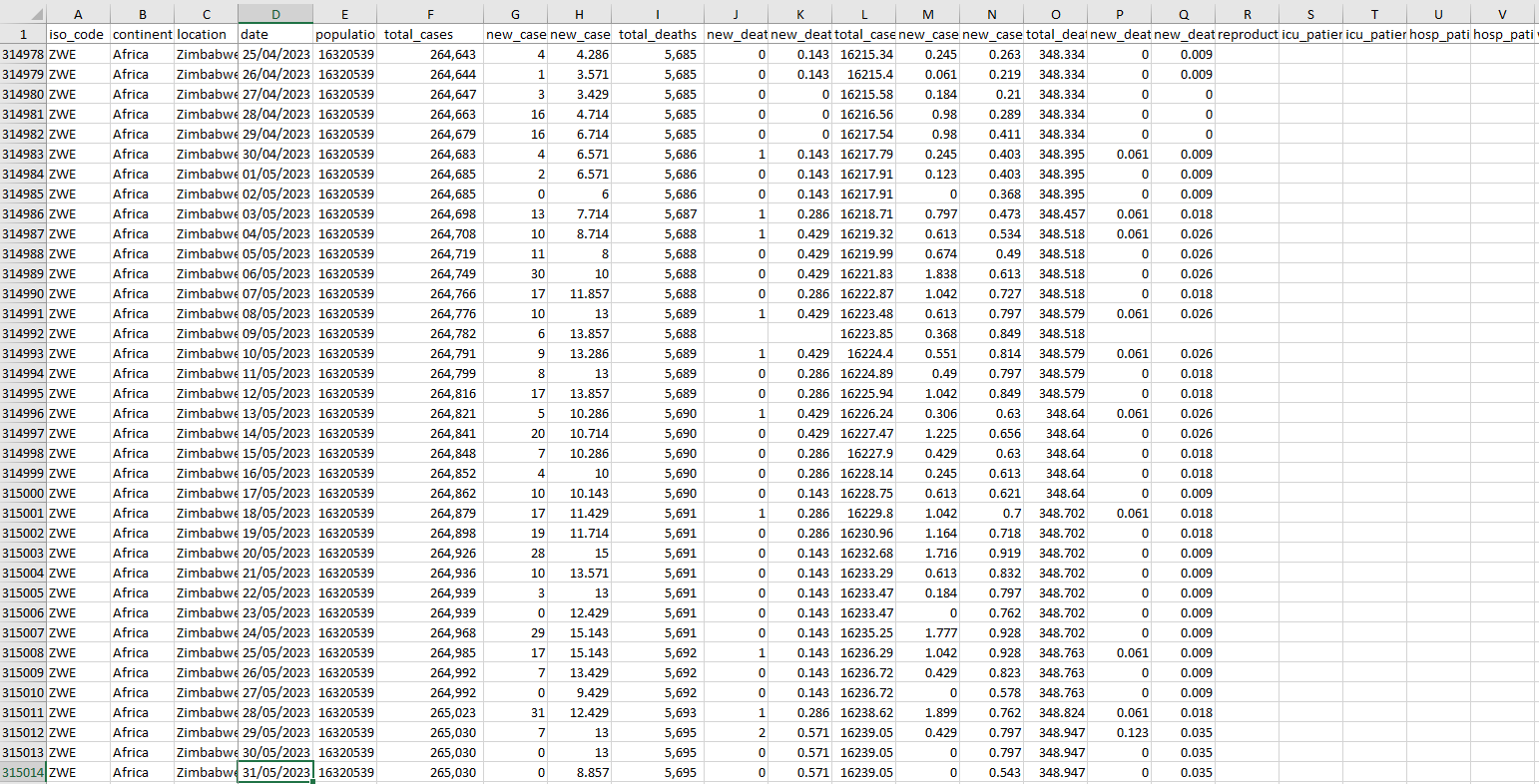
# Collecting the Data.

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

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

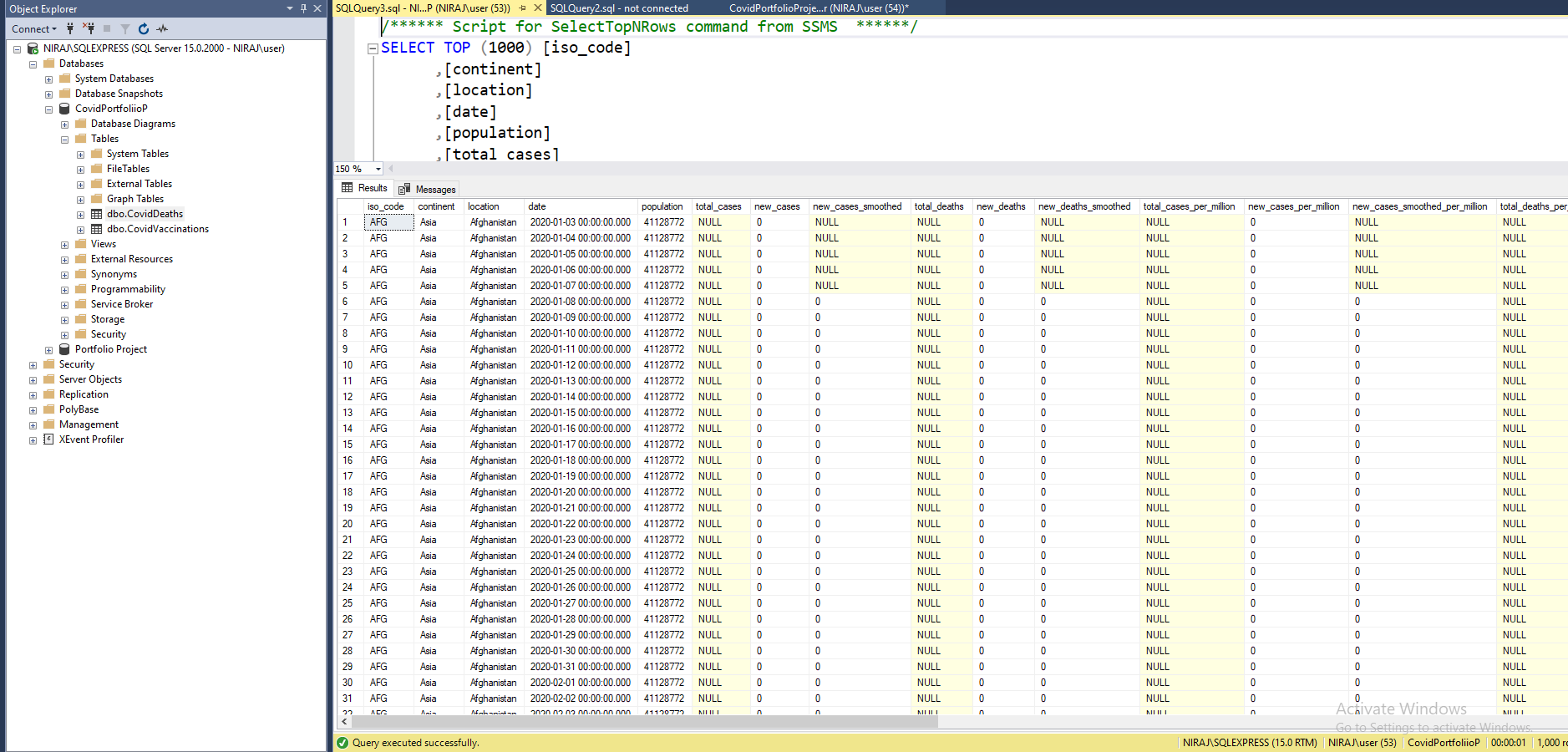
**Background**: The above website collects the data related to covid-19 deaths and vaccinations over the mentioned period.

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.

Each of the 11 Excel files had at least 40 tables. Only 10-15 tables were required for our study.

Macros were used initially to convert the tables into Excel recognizable tables and nomenclature was given to each table as “Table**N**”, where “N” stands for 1,2,3…., N.

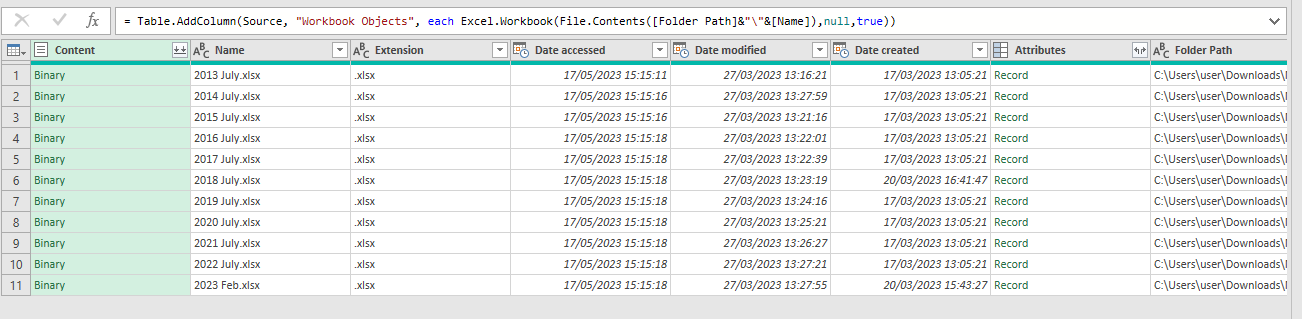
This nomenclature was required to distinguish the tables required for the study from other tables.

All 11 Excel files were loaded in a new Excel file using **Power Query**.

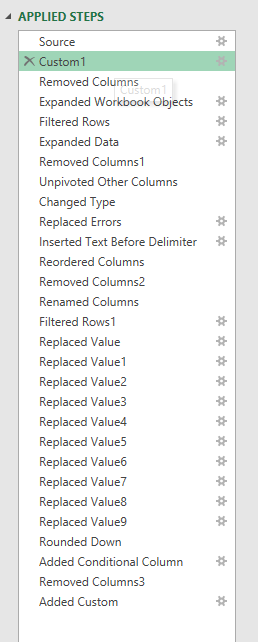
Only required tables were selected for the study using:





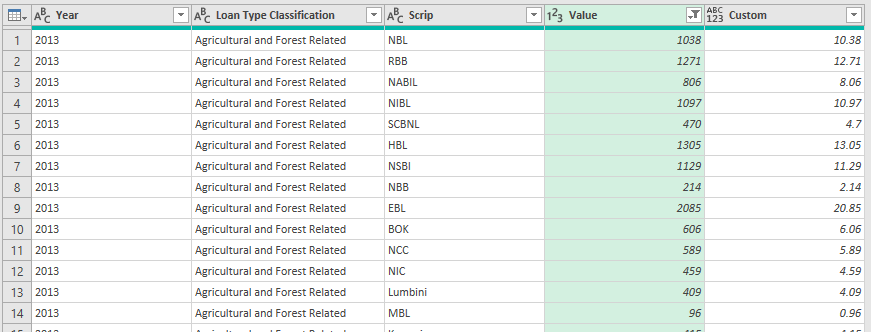


The following steps were used to clean the data:



Multiple “Replaced Value” was used to filter out the non-listed, merged, or acquired banks.

Finally, the data were transformed into five columns and 33,908 rows. It was then loaded to the Excel sheet and added to the data model.



Following relationships were created among the three tables based on a primary fields “Scrip” and “Loan Type Classification”.

A screenshot of a computer

Description automatically generated with medium confidence

# Analyzing the Data using SQL.

1. Total Cases vs Total Deaths

A screenshot of a computer

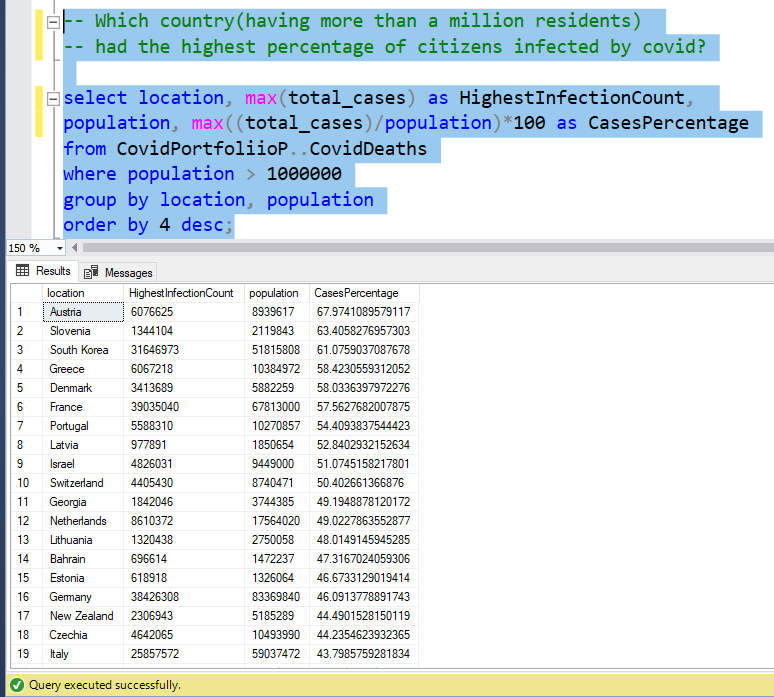
Description automatically generated with medium confidence

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

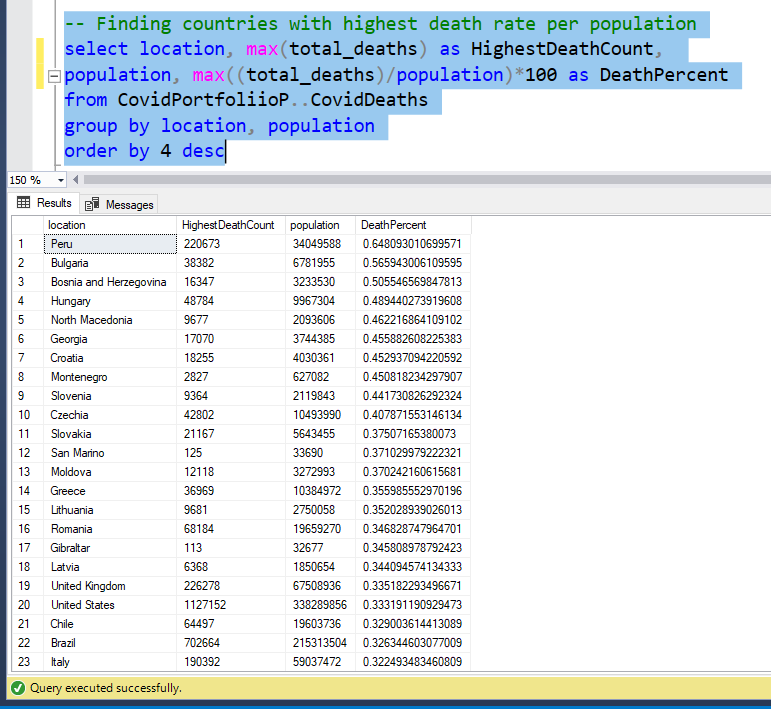
Description automatically generated

1. Highest Death count By Continent

A screenshot of a computer

Description automatically generated

1. Countries with highest death rate per population?

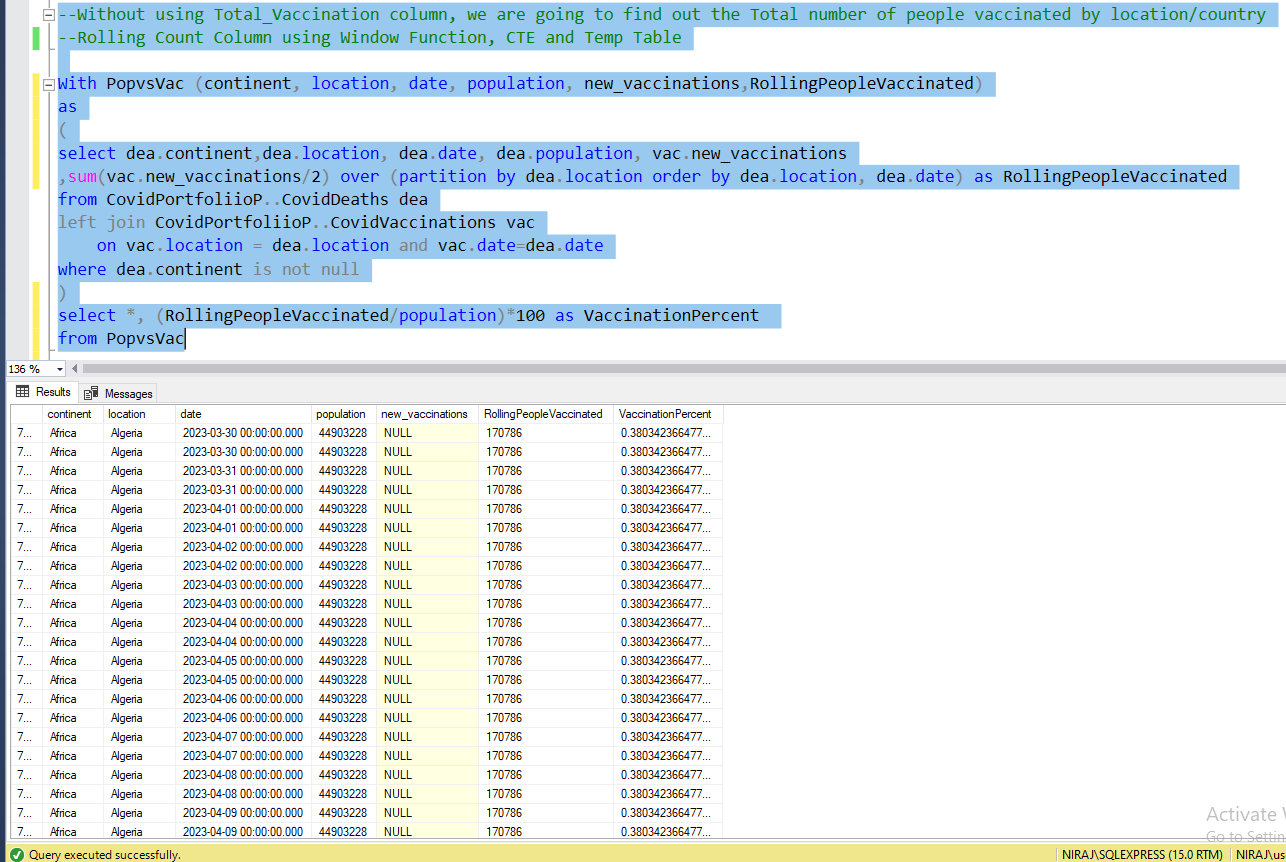


1. Joining deaths and vaccination tables on locations and dates

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.