

Covid-19 Infections and Deaths Data Analytics project

Data analytics using SQL and Tableau

Executive Summary

The aim of this analysis is to focus on Covid-19 pandemic as one of the most fatal diseases of the 21st century to gain insights into how it has affected countries and continents in different manner from January 2020 to October 2022.

This report covers three (3) different phases in my analysis to effectively draw relevant conclusions. These phases include Data collection and preparation, Data wrangling, Data visualization, and Analysis summary.

Methodology

Before performing the analysis, the data was collected through a public domain, then wrangled to make sure it's cleaned, reliable and error-free. After that, I explored the different variables, proceeded to data visualization to better capture trends and insights and finally draw conclusion.

Tools: Excel, SQL, Tableau

Data collection and preparation

The data to explore and analyze was made available through <http://ourworldindata.org/covid-deaths> which is a Public Domain collecting and updating daily data related to Covid-19 in all of the countries in the world. I proceeded to the collection and the storage of data by making sure they meet the requirements in terms of integrity, reliability, credibility and security.

The data collected is "owid-covid-data". However for the purposes of this project, I split it into two (2) data sets using excel:

CovidDeaths

CovidVaccinations

Data Wrangling

Let's take a quick look at our data sets: CovidDeaths and CovidVaccinations:

```
select*  
from [Portfolio Project]..CovidDeaths  
  
select*
```

```
from [Portfolio Project]..CovidVaccinations
```

Some inconsistencies have been found into some rows from “iso_code” column which represents acronyms used for all of the countries in these data sets. For example, United States is represented by “USA”. However, it turns out that some countries iso_code are preceded by “Owid_” like for example “Owid_KOS” for Kosovo. To make our data consistent, let’s delete these “Owid_” from our data sets.

```
-- deleting row values containing "OWID_" from CovidDeaths table
```

```
Select iso_code, location,  
REPLACE(iso_code, 'OWID_', '')  
from [Portfolio Project]..CovidDeaths  
Group by iso_code, location
```

```
Alter Table [Portfolio Project]..CovidDeaths  
Add IsoCodeEdited Nvarchar(255)
```

```
Update [Portfolio Project]..CovidDeaths  
Set IsoCodeEdited = REPLACE(iso_code, 'OWID_', '')
```

```
-- deleting row values containing "OWID_" from CovidVaccinations table
```

```
Select iso_code, location,  
REPLACE(iso_code, 'OWID_', '')  
from [Portfolio Project]..CovidVaccinations  
Group by iso_code, location
```

```
Alter Table [Portfolio Project]..CovidVaccinations  
Add IsoCodeEdited Nvarchar(255)
```

```
Update [Portfolio Project]..CovidVaccinations  
Set IsoCodeEdited = REPLACE(iso_code, 'OWID_', '')
```

Now, we will delete “iso_code” column as this has been replaced with “IsoCodeEdited” which contains consistent data.

```
-- Deleting Unused Columns
```

```
Alter Table [Portfolio Project]..CovidDeaths  
DROP COLUMN iso_code
```

```
Alter Table [Portfolio Project]..CovidVaccinations  
DROP COLUMN iso_code
```

```
-- Checking new columns
```

```
select*  
from [Portfolio Project]..CovidDeaths
```

```
select*  
from [Portfolio Project]..CovidVaccinations
```

Data Exploration And Analysis

Let’s focus our analysis on Brazil as one of the most infected countries in the world

--Showing death percentage by day in Brazil

```
select location, date, total_cases, total_deaths, (total_deaths/total_cases)*100 as
deaths_percentage
from [Portfolio Project]..CovidDeaths
where location = 'Brazil'
order by deaths_percentage desc
```

This query shows the daily death percentage since the first infection in Brazil, and this can also be viewed as likelihood of dying if an individual contracts covid-19 in Brazil.

The lowest death percentage was 0.31% which represents the first death from Covid-19 in the country on March 17th, 2020 while the highest death percentage was 7% on May 12th, 2020.

On October 20th, 2022, the death percentage decreased significantly to 1.97%

--Showing what percentage of population got Covid

```
select location, date, total_cases, population, (total_cases/population)*100 as
percent_population_infected
from [Portfolio Project]..CovidDeaths
where location = 'Brazil'
order by date desc
```

This query shows that the highest percentage of population infected in Brazil is 16.22% and occurred on October 20th, 2022. As the number of cases increases as does the percentage of population infected.

In this section of our analysis, we will compare infections across countries and continent and conduct a comparative analysis.

--Showing countries with highest infection rate compared to population

```
select location, population, MAX(total_cases) as HighestInfectionCount,
MAX((total_cases/population))*100 as percent_population_infected
from [Portfolio Project]..CovidDeaths
group by location, population
order by percent_population_infected DESC
```

Here we notice that Cyprus has the highest infection rate: 66.24%. However, it is important to note that the total population of this country is only 896,007 for a total of 593,542 people infected. This may be due to a rapid spread of the disease and/or weak policy responses.

North Korea has the lowest infection rate: 3.85031381405194E-06% which is significantly low as the country has a total population of 25,971,909 for only one (1) individual infected.

--Showing countries with highest death count per population

```
select location, MAX(cast(total_deaths as int)) as total_death_count
from [Portfolio Project]..CovidDeaths
```

```
where continent is not null
group by location
order by total_death_count DESC
```

This query indicates that the United States have the highest death count: 1,067,105 which is significantly high compared to Nauru which records only one (1) death.

--Showing total death count by continent

```
select continent, SUM(cast(new_deaths as int)) as total_death_count
from [Portfolio Project]..CovidDeaths
where continent is not null
group by continent
order by total_death_count DESC
```

Europe has the highest death count with 1,955,528 which is significantly high compared to Oceania recording 20,005 deaths.

--Showing global numbers in terms of total deaths and death percentage in the world

```
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 death_percentage
from [Portfolio Project]..CovidDeaths
where continent is not null
```

As of October 20th, 2022 there were a total of 625,153,703 people infected, 6,536,912 deaths for a death percentage of 1.05%.

-- Total people vaccinated in the world by continent, country and date using Common Table Expressions (CTE)

```
With PopvsVac (Continent, date, Location, Population, New_Vaccinations,
TotalVaccinated)
as
(
select dea.continent, dea.date, dea.location, dea.population, vac.new_vaccinations
, sum(convert(int,vac.new_vaccinations)) OVER (Partition by dea.location) as
TotalVaccinated
from [Portfolio Project]..CovidDeaths dea
join [Portfolio Project]..CovidVaccinations vac
    On dea.location = vac.location
    and dea.date = vac.date
where dea.continent is not null
)
select*, (TotalVaccinated/population)*100 as PercentPopulationVaccinated
from PopvsVac
```

In this section both data sets (CovidDeaths and CovidVaccinations) have been merged to facilitate my analysis related to the percentage of population vaccinated by continent, country and date.

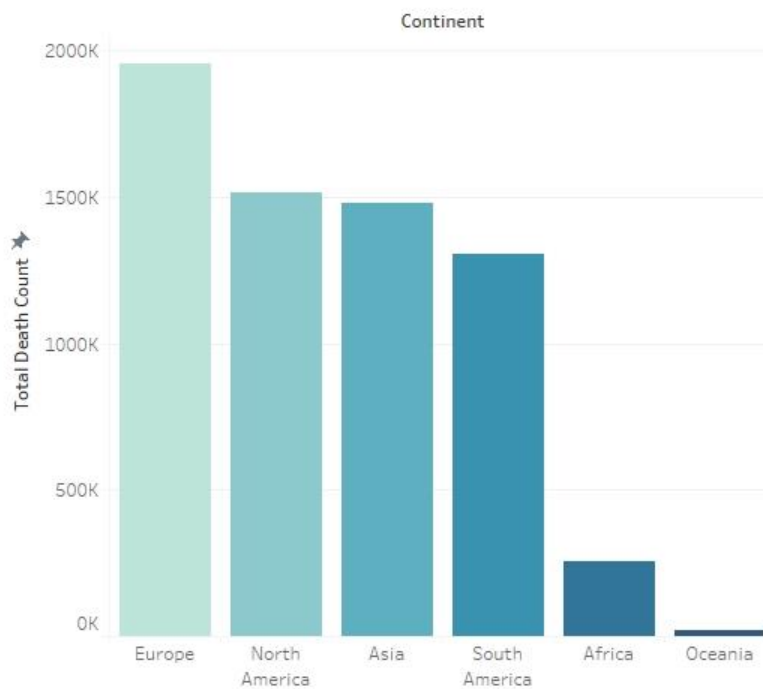
In Canada for example 239% population have been vaccinated taking into account the number of doses injected per person.

Data Visualization

Here I will go through a series of visualizations which come from previous data sets created in my analysis.

For this graph, I plot total death count against continent.

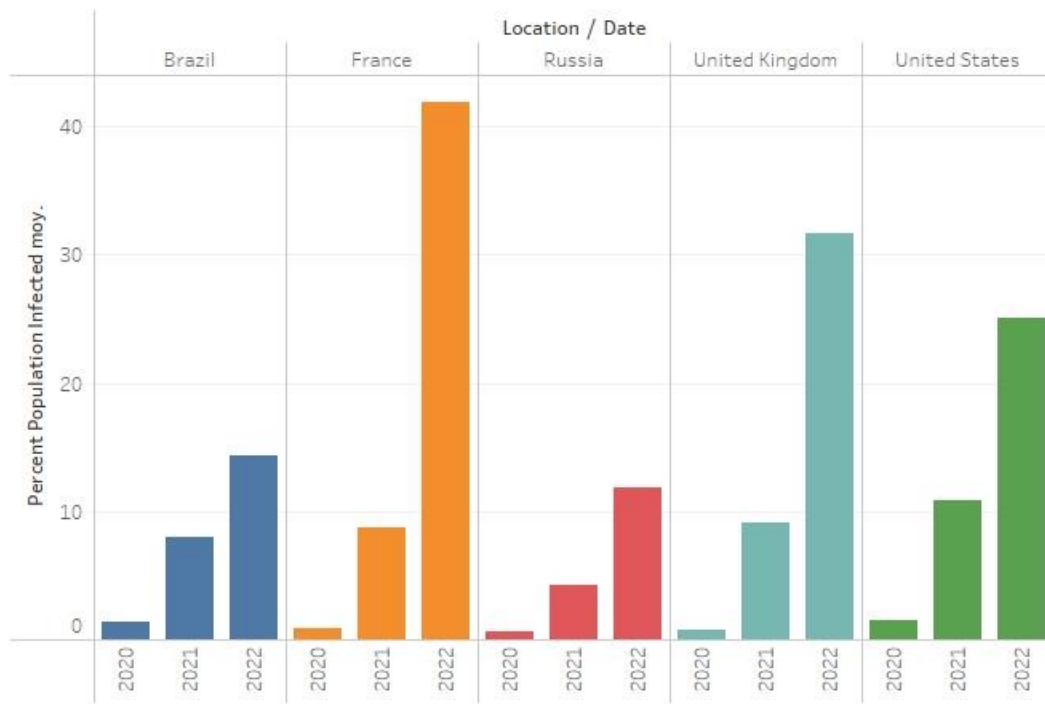
Total Death Count by Continent



We notice that Europe has the highest death count, followed by North America while Oceania has the lowest number of death.

For this graph, I plot percent population infected against location and date

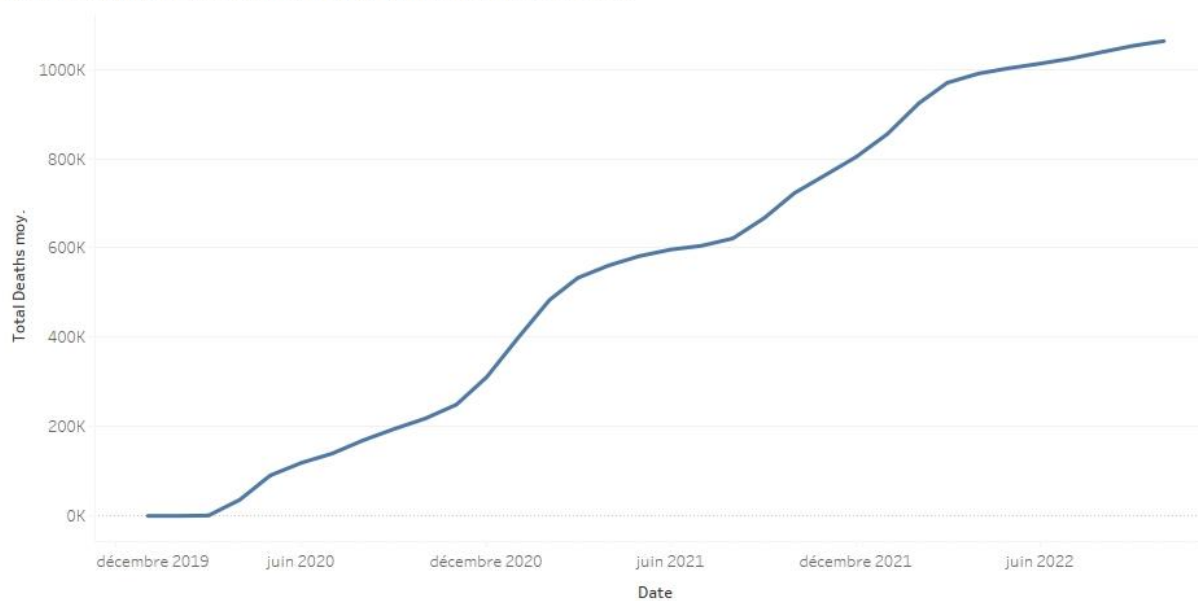
Percent Population Infected in Brazil, France, Russia, UK and USA in 2020, 2021 and 2022



This graph shows the evolution of Covid-19 since the outbreak until now in five (5) specific countries including Brazil, France, Russia, United Kingdom and United States. We can see that France and United Kingdom has experienced an important rate of population infected in 2022. France has the highest rate with 41.90% population infected.

Here, I plot total deaths against date in the United States

Total Deaths in the United States from 2020 to 2022



This graph shows the evolution of Covid-19 in the United States since the first death on January 3rd, 2020. As we can see the curve is rising and as of October 2022, total deaths was 1,067,105

Tableau Dashboard

To visualize an interactive dashboard containing more of these visuals, please see

<https://public.tableau.com/app/profile/andre3721/viz/WorldCovidInfectionsDashboard/Tableaudebord1>

Analysis summary

- As of October 20th, 2022 in the world, there were a total of 625,153,703 people infected with Covid-19, and 6,536,912 deaths
- The United States have the highest death count: 1,067,105
- Europe has the highest death count compared to other continents: 1,955,528
- Cyprus has the highest infection rate: 66.24%.