Case Study: Covid-19 Exploration

Luis Prieto Uzcategui

2022-10-24

1. ASK

1.1 What is the problem you are trying to solve?

Strategies aimed at exploring Covid-19 deaths by location, continent, population, reproductive rate, and positive rate.

1.2 How can your insights drive business decisions?

Using historical data from February 2020 to October 2022 to understand Covid-19 data.

1.3 Key tasks

- Show the likelihood of dying if you contract covid in your country
- Show what percentage of the population got covid
- Looking at Countries with Highest Infection Rate compared to Population
- Showing Countries with Highest Death Count per Population
- What are the total cases per country or location
- Shows the reproduction rate risk in each location by date, and population
- Show the average reproduction rate by location and population
- Showing continents with the highest death count per population
- Showing the total cases vs Total Death as a percentage
- Looking at Total Population vs Vaccinations
- Show the positive rate in each location in relation to total cases and total deaths

2. PREPARE

2.1 Where is your data located?

The data collected is from Our World in Data https://ourworldindata.org/covid-deaths

2.2 How is the data organized?

The data is considered structured data because is organized in a certain format, like rows and columns.

2.3 Are their issues with bias or credibility in this data? Does your data ROCCC?

Data has been downloaded from Our World in Data. This data meet ROCCC because is reliable, original (based on the author, so we believe with precaution), comprehensive, current, and cited.

2.4 How are you addressing licensing, privacy, security, and accessibility?

The data has been made available by Our World in Data under this link https://ourworldindata.org/covid-deaths

2.5 How did you verify the data's integrity?

This is public data that you can use the explore how Covid-19 many data fields. We are going to assume the data is credible.

3. PROCESS

3.1 Choose your tools

SQL Server for processing, cleaning, and analysis. Data visualization for Tableau.

3.2 Check the data for errors

Many data rows have NULL values because of a lack of data, and data that keep updating, so I used the WHERE statement that is not null to reduce any bad data that could damage the dataset.

I also transform some data types to another data type because I was doing exercises, but it wouldn't let me until I change the data to int or float.

I also check Null values when I transferred my data to a spreadsheet in Excel for data visualization. I clicked Ctrl + H to replace Null cells with 0 values.

4. ANALYZE

4.1 How should you organize your data to perform analysis on it?

```
-- Looking at Total Cases vs Total Deaths
-- Show the likelyhood of dying if you contract covid in your country
Select location,date,total_cases,total_deaths, (total_deaths/total_cases)*100 AS
DeathPercentage
From PortfolioProject.dbo.CovidDeaths
--Where location LIKE '%United States%' and continent is not null
--Where location LIKE '%Venezuela%' and  continent is not null
Where continent is not null and date is not null and total_cases is not null and
total_deaths is not null
Order by location,date
```

Observation: The death percentage reflects that the higher percentage the higher the chances of dying in that particular country and day.

```
-- Looking Total Cases vs Population
-- Shows what percentage of population got covid

Select location,date,population,total_cases, (total_cases/population)*100 AS
PercentagePopulation
From PortfolioProject.dbo.CovidDeaths
```

```
--Where location LIKE '%Venezuela%' and continent is not null and continent is not null
and date is not null and population is not null and total_cases is not null
--Where location LIKE '%states%' and continent is not null and date is not null and
population is not null and total_cases is not null
Where continent is not null and date is not null and population is not null and
total_cases is not null
Order by location,date
```

Observation: The graph reflects what percentage of the population got covid. The higher percentage of the population the higher the chances of people getting infected.

```
-- Looking at Countries with Highest Infection Rate compared to Population
-- Graph 3 for data visualization
Select location,population,Max(total_cases) AS HighestInfectionCount,
Max((total_cases/population)*100) AS PercentPopulation_Infected
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and population is not null  and total_cases is not null
Group by Location, population
Order by PercentPopulation_Infected desc
```

Observation: Looking at Countries with the Highest Infection Rate compared to Population.

```
-- Graph 4 for data visualization
Select Location, Population,date, MAX(total_cases) as HighestInfectionCount,
Max((total_cases/population))*100 as PercentPopulationInfected
From PortfolioProject..CovidDeaths
Where population is not null and date is not null and total_cases is not null
Group by Location, Population, date
order by PercentPopulationInfected desc
```

Observation: Looking at Countries with percent of the population infected. Cyprus accounted for 66.24 in 2022, which reflects the highest percent population infected by the country by far.

```
-- Showing Countries with Highest Death Count per Population

Select location, Max(cast(Total_deaths as int)) AS TotalDeathCount
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and total_deaths is not null
Group by Location, population
Order by TotalDeathCount desc
```

Observation: The graph shows countries with the highest death count per population. The United States hit the highest death count by 1065076.

```
-- What is the total cases per country or location
--Graph 6
Select location, Sum(total_cases) as Total_cases_percountry
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and total_cases is not null
Group by location, new_cases
Order by 2 desc
```

Observation: What are the total cases per country or location.

```sql
-- Change the nvarchar to float because this will give me an error that would give error
Msg 245
ALTER TABLE PortfolioProject.dbo.CovidDeaths
ALTER COLUMN reproduction_rate FLOAT

--Shows the reproduction rate risk in each location by date, population
-- Use is not null and reduced 210,932 to 165,734
-- Graph 7 used for data visulization
Select location, date, population, total_deaths, reproduction_rate,
Case
	When reproduction_rate >= 1 Then 'spreading exponientally'
	When reproduction_rate <= 1 then 'outbreak is subsiding'
	Else 'Null'
End AS reproductionrate_indicator
From PortfolioProject.dbo.CovidDeaths
Where continent is not null  and reproduction_rate is not null and date is not null and
population is not null
--Group by location,date
Order by 1, 2
```

Observation: Show the reproduction rate risk in each location by date, and population. If the value was one or higher Covid-19 infection was spreading exponentially and if the value was less than 1 the Covid-19 infection was decreasing.

```sql
-- Show the average reproduction rate by location and population
-- Graph 8 for data visualization
Select location, population, avg(reproduction_rate) as  Avg_reprductionrate
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and reproduction_rate is not null
Group by location, population
Order by 2 desc
```

Observation: Show the average reproduction rate risk in each location, and population. If the value was one or higher Covid-19 infection was spreading exponentially and if the value was less than 1 the Covid-19 infection was decreasing.

```sql
-- Let's Break Things Down By Continent

-- Showing contintents with the highest death count per population

Select location, Max(cast(Total_deaths as int)) AS TotalDeathCount
From PortfolioProject.dbo.CovidDeaths
Where continent is null
Group by location
Order by TotalDeathCount desc
```

Observation: Show continents with the highest death count per population.

```sql
--Use for data visualization graph 2
Select location, SUM(cast(new_deaths as int)) as TotalDeathCount
From PortfolioProject..CovidDeaths
```

```
--Where location like '%states%'
Where continent is null
and location not in ('World', 'European Union', 'International')
Group by location
order by TotalDeathCount desc
```

Observation: Show continents the total deaths per location.

```
-- Global Number

Select date,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 DeathPercent--
(total_cases/population)*100 AS PercentagePopulation
From PortfolioProject.dbo.CovidDeaths
--Where location LIKE '%states%'
Where continent is not null and new_cases is not null and new_deaths is not null
Group by date
Order by 1, 2
```

Observation: Show the total number of cases, total deaths, and death percentage by date.

```
-- Showing the total cases vs Total Death as percentage
-- Graph 1 use for data visulization
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 DeathPercent--
(total_cases/population)*100 AS PercentagePopulation
From PortfolioProject.dbo.CovidDeaths
--Where location LIKE '%states%'
 Where continent is not null and new_cases is not null and new_deaths is not null
--Group by date
Order by 1, 2
```

Observation: Show the total number of cases, total deaths, and death percentage.

```
-- Looking at Total Population vs Vaccinations

Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations,
Sum (Cast(vac.new_vaccinations AS bigint)) OVER (Partition by dea.location Order by
dea.location,
dea.Date) AS RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
From PortfolioProject.dbo.CovidDeaths AS Dea
JOIN PortfolioProject.dbo.CovidVaccinations Vac
       ON dea.location = vac.location
       and Dea.date = Vac.date
Where dea.continent is not null and dea.date is not null and dea.population is not null
and vac.new_vaccinations is not null
Order by 2,3

Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations,
SUM(CONVERT(bigint,vac.new_vaccinations)) OVER (Partition by dea.Location Order by
dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
From PortfolioProject..CovidDeaths dea
```

```
Join PortfolioProject..CovidVaccinations vac
	On dea.location = vac.location
	and dea.date = vac.date
Where dea.continent is not null and dea.date is not null and dea.population is not null
and vac.new_vaccinations is not null
order by 2,3
```

Observation: Show the vaccinations per country. The number of vaccinations was added per day in each country using Partition By.

```
--
-- Show the positive rate in each location  in relation with total_Cases and total_deaths
Select dea.location, dea.date, dea.total_cases,dea.total_deaths,vac.positive_rate
From PortfolioProject..CovidDeaths dea
Join PortfolioProject..CovidVaccinations vac
	On dea.location = vac.location
	and dea.date = vac.date
where dea.continent is not null and dea.date is not null and dea.total_cases is not null
and dea.total_deaths is not null and vac.positive_rate is not null
Group by dea.location, dea.date, dea.total_cases, dea.total_deaths, vac.positive_rate
Order by 2
```

Observation: Show the positive rate in each location per day.

```
-- USE CTE

With PopvsVac ( Continent, Location, Date, Population, New_Vaccinations,
RollingPeopleVaccinated)
As
(
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations,
SUM(CONVERT(bigint,vac.new_vaccinations)) OVER (Partition by dea.Location Order by
dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
From PortfolioProject..CovidDeaths dea
Join PortfolioProject..CovidVaccinations vac
	On dea.location = vac.location
	and dea.date = vac.date
Where dea.continent is not null and dea.date is not null and dea.population is not null
and vac.new_vaccinations is not null
--order by 2,3
)
Select *, (RollingPeopleVaccinated/Population)*100
From PopvsVac



-- Using Temp Table to perform Calculation on Partition By in previous query

DROP Table if exists #PercentPopulationVaccinated
Create Table #PercentPopulationVaccinated
(
Continent nvarchar (255),
Location nvarchar (255),
Date datetime,
Population numeric,
```

```
New_vaccinations numeric,
RollingPeopleVaccinated numeric
)

Insert Into #PercentPopulationVaccinated
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations,
SUM(CONVERT(bigint,vac.new_vaccinations)) OVER (Partition by dea.Location Order by
dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
From PortfolioProject..CovidDeaths dea
Join PortfolioProject..CovidVaccinations vac
       On dea.location = vac.location
       and dea.date = vac.date
Where dea.continent is not null and dea.date is not null and dea.population is not null
and vac.new_vaccinations is not null

Select *, (RollingPeopleVaccinated/Population)*100
From #PercentPopulationVaccinated


-- Creating View to store data for later visualizations

Create View PercentPopulationVaccinated As
Select dea.continent, dea.location, dea.date, dea.population, vac.new_vaccinations,
SUM(CONVERT(bigint,vac.new_vaccinations)) OVER (Partition by dea.Location Order by
dea.location, dea.Date) as RollingPeopleVaccinated
--, (RollingPeopleVaccinated/population)*100
From PortfolioProject..CovidDeaths dea
Join PortfolioProject..CovidVaccinations vac
       On dea.location = vac.location
       and dea.date = vac.date
Where dea.continent is not null and dea.date is not null and dea.population is not null
and vac.new_vaccinations is not null
```

5. SHARE

5.1 Global numbers

```
-- Showing the total cases vs Total Death as percentage
-- Graph 1 use for data visulization
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 DeathPercent--
(total_cases/population)*100 AS PercentagePopulation
From PortfolioProject.dbo.CovidDeaths
--Where location LIKE '%states%'
 Where continent is not null and new_cases is not null and new_deaths is not null
--Group by date
Order by 1, 2
```

Observation: The total cases from 2020 to 2022 reflects 621,452,254. The total number of deaths from 2020 to 2022 reflects 6,520,512 and representing 1.05% death percent.

5.2 Total Deaths Per Continent

```
--Use for data visualization graph 2
Select location, SUM(cast(new_deaths as int)) as TotalDeathCount
```

```
From PortfolioProject..CovidDeaths
--Where location like '%states%'
Where continent is null
and location not in ('World', 'European Union', 'International')
Group by location
order by TotalDeathCount desc
```

Observation: Europe's total deaths accounted for the most deaths per continent and Oceania accounted for the least deaths per continent.

5.3 Percent Population Infected Per Country

```
-- Looking at Countries with Highest Infection Rate compared to Population
-- Graph 3 for data visualization
Select location,population,Max(total_cases) AS HighestInfectionCount,
Max((total_cases/population)*100) AS PercentPopulation_Infected
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and population is not null  and total_cases is not null
Group by Location, population
Order by PercentPopulation_Infected desc
```

Observation: The map showed which country was the most country infected per population. The lighter the color the least that country was infected and the darker the color the more infected was that country.

5.4 Percent of the Population Infected

```
-- Graph 4 for data visualization
Select Location, Population,date, MAX(total_cases) as HighestInfectionCount,
Max((total_cases/population))*100 as PercentPopulationInfected
From PortfolioProject..CovidDeaths
Where population is not null and date is not null and total_cases is not null
Group by Location, Population, date
order by PercentPopulationInfected desc
```

Observation: This graph reflected how the population is getting infected on average by Covid-19, which increases from 2020 to 2022 in France, the United States, and Japan. The forecasting also shows an alarming number.

5.5 Total Cases per Country 2020-2022

```
-- What is the total cases per country or location
--Graph 6
Select location, Sum(total_cases) as Total_cases_percountry
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and total_cases is not null
Group by location, new_cases
Order by 2 desc
```

Observation: This graph reflected the total number of cases per country, countries that start from blue reflect the least country with the most cases, and countries with red reflected the most cases of Covid-19 cases such as the United States.

5.6 Average Reproduction rate per Country

```
--Shows the reproduction rate risk in each location by date, population
```

```
-- Use is not null and reduced 210,932 to 165,734
-- Graph 7 used for data visulization
Select location, date, population, total_deaths, reproduction_rate,
Case
	When reproduction_rate >= 1 Then 'spreading exponientally'
	When reproduction_rate <= 1 then 'outbreak is subsiding'
	Else 'Null'
End AS reproductionrate_indicator
From PortfolioProject.dbo.CovidDeaths
Where continent is not null  and reproduction_rate is not null and date is not null and
population is not null
--Group by location,date
Order by 1, 2
```

Observation: This graph reflected the reproduction rate per country a number of 1 or more reflects that people are infecting more and a value of 1 or less reflects that the infection is decreasing. From 2020 each country was at its peak, but over the two years have been decreasing.

5.7 Reproduction Rate Indicator per day

```
-- Show the average reproduction rate by location and population
-- Graph 8 for data visualization
Select location, population, avg(reproduction_rate) as  Avg_reprductionrate
From PortfolioProject.dbo.CovidDeaths
Where continent is not null and reproduction_rate is not null
Group by location, population
Order by 2 desc
```

Observation: This graph reflected the reproduction rate per day a number of 1 or more reflects that people are spreading exponentially and a value of 1 or less reflects that the infection outbreak is subsiding. In 2020 the spreading exponentially was over outbreak subsided, but by 2022 the outbreak is decreasing running over the spreading exponentially.

6. ACT

6.1 Conclusions and findings

- The death percentage shows that North Korea accounted for the highest death percentage in 2022 at 600%, and the Cook Islands with the smallest death percentage of 0.0156519017060573% in 2022. This data reflects the death percentage of each country, so still affects the present of each country as this data was collected until October 2022. Countries must take into account that the death of Covid-19 is still relevant to the health of each country.

- The population percentage that got Covid showed Cyprus 66.24 by October 2022. The United States population percentage that got Covid increased from 2.96% in 2020 to 28.76%. The data reflect how contagious Covid-19 is still relevant to people in the US and how people are more likely to get Covid-19 in 2022 than in 2020. The United States must take into account this result as in 2022 there are already Covid-19 vaccinations appearing and yet people are more like to get infected.

- Countries with the highest infection include the United States, India, France, Brazil, Germany, and many others. The United States accounted for 96940217 infections making the highest of all countries and North Korea accounted for only one infection, but this show not enough data, so next with more reliable data is Macao accounted 793 for infection.

- The Country with the highest death count was the United States with 1065076 and Cook Islands with one death count.

- The total cases per country showed the United States leading the table with 39,160,628,619, India with 22,608,644,817, and Brazil with 15,928,442,709. This data reflects how big countries have the most cases of Covid-19 accounted for.

- The reproduction rate from January 2020 to December 2020, but from there it kept decreasing and increasing until it reached its lowest in February 2022, but is still increasing at a decreasing rate. Covid-19 reproduction rate have decreased exponentially by 2022, but still, show signs of increase if precautions are not enforced even if it is not a threat anymore.

- China's average reproduction rate daily in January 2020 hit 3.53%, which commonly 1 or more reflects that Covid-19 was a dangerous infection rate in China, but by October 2022 have been 0.74%, which reflects the reproduction rate is decreasing compared to 2020. This data showed how many countries reach their peaked back in 2020, but in 2022 the Covid-19 reproduction rate by day is decreasing, but still relevant.

- The continent with the highest total deaths in Europe and Oceania with the less total deaths.

- Global numbers of total cases,621,452,254; total deaths, 6,520,512; and death percent, 1.05.

- The vaccinations per country showed how vaccinations are per day in each country. The United States accounted for Dec 2020 4747 to October 2022 627815993 vaccinations in total. This data can vary by each country and the United States shows how vaccinations have increased since they came out of the market in the US.

- The positive rate for each country varies by day. If the positive rate is at least 1% the Covid-19 infection is under control and higher means less ability to control, it.

7. REFERENCE PAGE

https://ourworldindata.org/covid-deaths. This was dataset I took from it.

https://www.reuters.com/article/us-health-coronavirus-transmission-expla/explainer-tracking-the-coronavirus-reproduction-rate-as-lockdowns-ease-idUSKBN22N2ZS

https://publichealth.jhu.edu/2020/covid-19-testing-understanding-the-percent-positive

8. CONTACT PAGE

www.linkedin.com/in/luis-prieto-uzcategui