

**Data Analytics Process**

**Covid Data Case Study**

**Description:**

COVID-19 had a huge impact on all of us in the last few years, and now there is data available for us to understand how it affected countries differently.

Using data exploration, I took a subset of the COVID-19 Deaths and Vaccination data to see what statistics or patterns could be found.

This project is based on the COVID-19 SQL project by [Alex the Analyst](https://www.youtube.com/@AlexTheAnalyst), but I used Ms SQL Server for my analyses and used some different queries.

**The Data**

The full COVID dataset can be found [here](https://ourworldindata.org/covid-deaths), and the subset of Deaths and Vaccination data I used for this project can be found in my [GitHub](https://github.com/manojjagdale2988/Covid-Data-Analysis-SQL), along with Ms SQL statements.

The datasets contain information collected from Jan 1, 2020 to April 30, 2021.

**Preparing the Data**

The first step was to convert both XLSX files to CSV so that they can be uploaded into Ms SQL Server. Once that was done, the files were ready to be uploaded and queried.

**The Analysis**

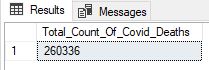
Since this is a data exploration project, there were no specific questions that needed to be answered. The goal was to get a feel for the data and see what patterns or trends — if any — could be seen, and compare stats amongst countries.

**1. Let us check how many rows are available in each table.**

Let’s take a count of the rows in each table:

SELECT COUNT(\*) as Total\_Count\_Of\_Covid\_Deaths

FROM covidportfolio..CovidDeaths



Number of rows in the covidDeath table

SELECT COUNT(\*) as Total\_Count\_Of\_Covid\_Vaccinations

FROM covidportfolio..CovidVaccinations



Number of rows in CovidVaccination Table

We can see that the Covid\_Death table has 2,60,336 and Covid\_Vaccination table has 12,445 rows.

**2. Let us take check the data in each table.**

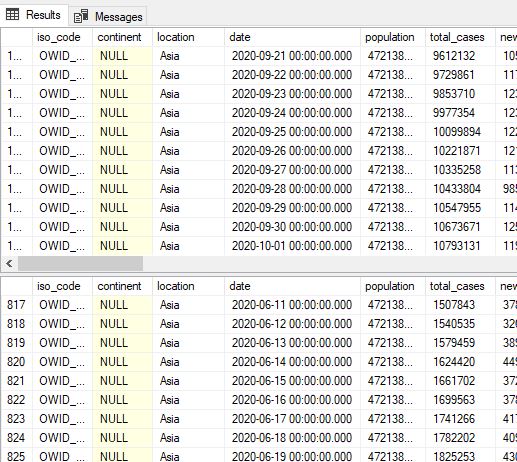
SELECT \*

FROM covidportfolio..CovidDeaths

SELECT \*

FROM covidportfolio..CovidVaccinations

As we see there is null values present in ‘continent’ column and some countries are showing the continents under the ‘Location’ Column.



So, We can handle this in SQL by using where condition – **‘WHERE continent is not NULL’**

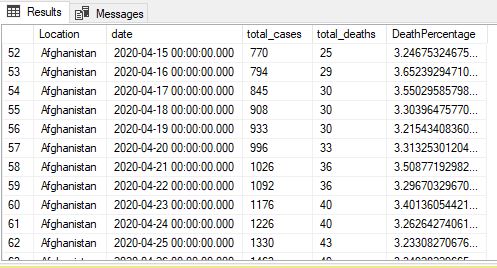
3. Let us check the Percentage of Total Deaths with Total Cases of Covid.

Select Location, date, total\_cases,total\_deaths, (total\_deaths/total\_cases)\*100 as DeathPercentage

From covidportfolio..CovidDeaths

Where continent is not null

order by 1,2



We can see that the deaths start to increase as the number of COVID cases start to increase in early 2020.

4. Now we can check the Percentage of Total Deaths with Total Cases of Covid by specific Location.

This shows the likelihood of dying if you contract COVID in the India., as measured by “DeathPercentage”.

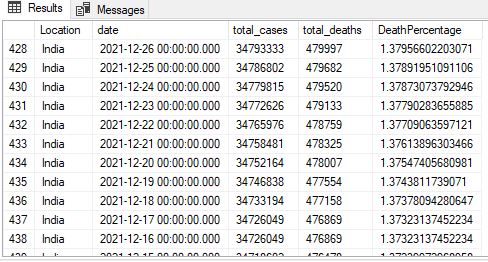
Select Location, date, total\_cases,total\_deaths, (total\_deaths/total\_cases)\*100 as DeathPercentage

From covidportfolio..CovidDeaths

Where continent is not null

And location like ‘India’

order by 2 desc



We can see the total number of cases in the millions, by 2021, which is a lot.

5. Now we can same calculate same as above for other country say China

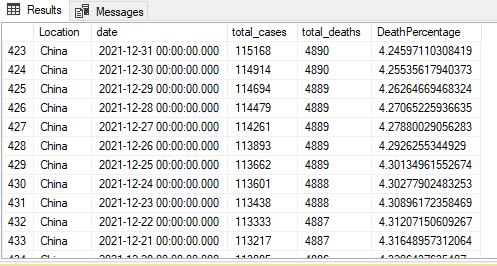
Select Location, date, total\_cases,total\_deaths, (total\_deaths/total\_cases)\*100 as DeathPercentage

From covidportfolio..CovidDeaths

Where continent is not null

And location like ‘China’

order by 2 desc



Not as many cases as in the India but still a lot, by 2021

6. Let’s Check percentage of population infected with Covid for the India.

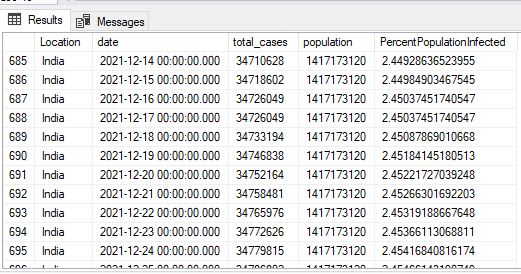
select Location, date, total\_cases, population, (total\_cases/population)\*100 as PercentPopulationInfected

from covidportfolio..CovidDeaths

Where location like 'India'

and continent is not null

order by 1,2



Showing Percentage of Population Infected in the India.

7. Now Check percentage of population infected with Covid for the China.

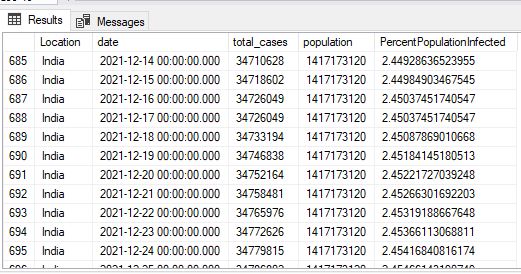
select Location, date, total\_cases, population, (total\_cases/population)\*100 as PercentPopulationInfected

from covidportfolio..CovidDeaths

Where location like 'China'

and continent is not null

order by 1,2



Showing Percentage of Population Infected in the China nearly same as the India.

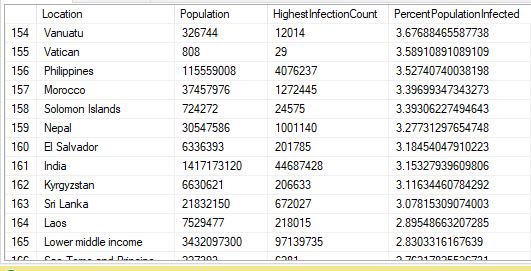
8. Let’s Countries with Highest Infection Rate compared to Population

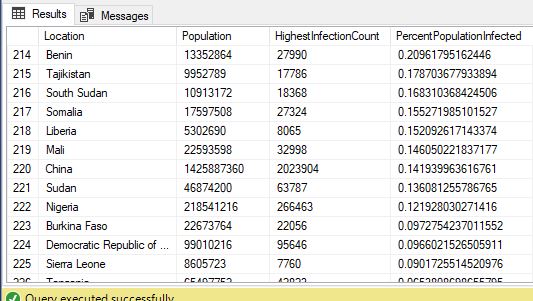
Select Location, Population, MAX(total\_cases) as HighestInfectionCount, Max((total\_cases/population))\*100 as PercentPopulationInfected

From covidportfolio..CovidDeaths

Group by Location, Population

order by PercentPopulationInfected desc





Showing Highest Infection Count in the India is 44 million and on 161 rank, Whereas in the China it’s showing only 2 million and 220 rank. It’s showing a contradiction.

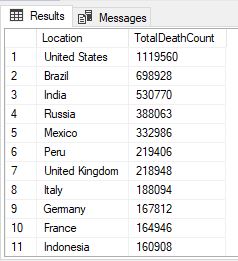
9. Let us show Countries with Highest Death Count per Population

Select Location, MAX( cast (total\_deaths as int)) as TotalDeathCount

From covidportfolio..CovidDeaths

Group by Location

order by TotalDeathCount desc



Showing highest Death Count/ Populations of Countries.

10. Showing continents with the highest death count per population

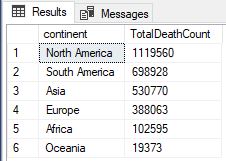
Select continent, MAX(cast(Total\_deaths as int)) as TotalDeathCount

From covidportfolio..CovidDeaths

Where continent is not null

Group by continent

order by TotalDeathCount desc



Showing highest Death Count/ Populations of Continents.

11. Now Check Global Numbers to each day.

Here is some problem occurred, total deaths and new\_deaths column has char datatype. So we can’t directly find the sum. So, we have to use cast function to change columns char datatype to Integer.

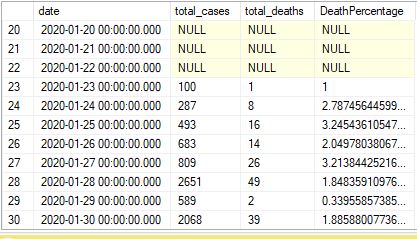
Select date, SUM(new\_cases) as total\_cases, SUM(cast(new\_deaths as int)) as total\_deaths, SUM(cast(new\_deaths as int))/SUM(New\_Cases)\*100 as DeathPercentage

From covidportfolio..CovidDeaths

where continent is not null

group by date

order by 1,2



Here is showing early days Global Numbers.

12. Now Check Global Number.

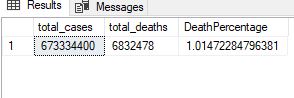
Here is some problem occurred, total deaths and new\_deaths column has char datatype. So we can’t directly find the sum. So, we have to use cast function to change columns char datatype to Integer.

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 DeathPercentage

From covidportfolio..CovidDeaths

where continent is not null

order by 1,2



This shows that, globally, the proportion of new COVID cases resulting in deaths is 2 percent.

13. Let us check the rolling number of people vaccinated per country over each day.

SELECT dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations,

SUM(cast(vac.new\_vaccinations as int)) OVER (PARTITION BY dea.location ORDER BY dea.location, dea.date) as RollingPeopleVaccinated

FROM covidportfolio..CovidDeaths AS dea

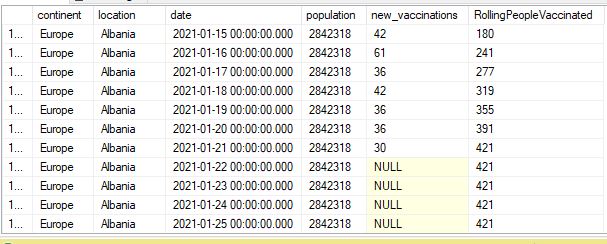
JOIN covidportfolio..CovidVaccinations AS vac

ON dea.location = vac.location

AND dea.date = vac.date

WHERE dea.continent is not NULL

ORDER BY 2,3;



So if there were no vaccinations that day (i.e. new\_vaccinations is NULL) then the rolling number stays the same, until the next day that occurs when there are new vaccinations.

14. Let us check Percentage of Population that has received at least one Covid Vaccine

With PopvsVac (Continent, Location, Date, Population, New\_Vaccinations, RollingPeopleVaccinated)

as

(

Select dea.continent, dea.location, dea.date, dea.population, vac.new\_vaccinations

, SUM(CONVERT(int,vac.new\_vaccinations)) OVER (Partition by dea.Location Order by dea.location, dea.Date) as RollingPeopleVaccinated

From covidportfolio..CovidDeaths dea

Join covidportfolio..CovidVaccinations vac

On dea.location = vac.location

and dea.date = vac.date

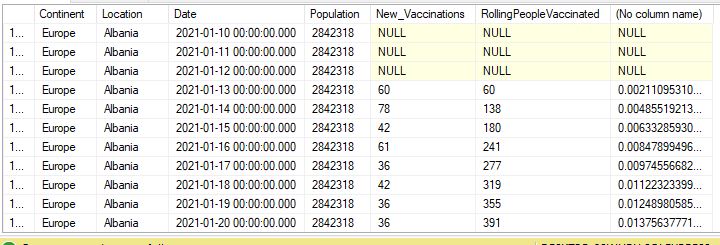
where dea.continent is not null

--order by 2,3

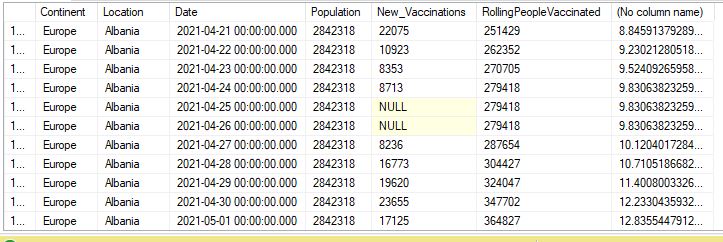
)

Select \*, (RollingPeopleVaccinated/Population)\*100

From PopvsVac



If we scroll further down the table on Albania, we can see that about 12% of the population is vaccinated by the end of April 2021:



**Conclusion**

1. Percentage of Population Infected in the China nearly same as the India 2.45%, by the end of 2021.
2. The country with the highest infection rate was Andorra, with 17% of the population being infected, though their population is quite small compared to other countries.
3. The U.S. has the highest death count per population out of all the countries.
4. North America has the highest death count per population, out of all the continents.