**Data Analyst Portfolio Project with C-19 Dataset**

**[Source :** [**https://www.youtube.com/watch?v=qfyynHBFOsM**](https://www.youtube.com/watch?v=qfyynHBFOsM)**]**

1. **SQL Data Exploration Project 1 of 4 (Data Exploration)**

**Steps**

1. Download the dataset from <https://ourworldindata.org/covid-deaths> (the dataset is downloaded on June 17-2022)
2. Do some excel formatting
   1. Two excel tables “CovidDeaths” and “CovidVaccination” are prepared and saved in downloads Folder. The file format is .xlsx not .csv
3. Upload the two excel workbooks to SQL server and rename them as “CovidDeaths” and “CovidVaccination”
4. Verif which table is which using SQL statements

Select \* from PortfolioProject .. CovidDeaths

Order by 3,4

1. Actual exploration of the data with SQL starts here
   1. **Select Data that We are going to be using  (the fields I am going to work with)**

SQL :

Select location,date,total\_cases,new\_cases, total\_deaths,population

from PortfolioProject..CovidDeaths

order by 1,2

1. **Looking at Total Cases vs Total Death (Fatality Rate) – Shows likelihood of dying covid if you contract covid in ethiopia**

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

from PortfolioProject..CovidDeaths

where location like '%Ethiopia%'

order by 1,2

1. Looking at total cases vs Population – shows what percentage of population got covid

Select location,date,total\_cases, population, (total\_cases/population)\*100 as PercentPopulationInfected

from PortfolioProject..CovidDeaths

where location like '%Ethiopia%'

order by 1,2

1. For future visualization let’s go from ethiopia to the world

Select location,date,total\_cases, population, (total\_cases/population)\*100 as PercentPopulationInfected

from PortfolioProject..CovidDeaths

– – where location like '%Ethiopia%' (comment out)

order by 1,2

1. Looking at countries with highest infection rate compared to population

Select location,population, max(total\_cases) as HighestInfectionCount,max((total\_cases/population))\*100 as PercentagePopulationInfected

from PortfolioProject..CovidDeaths

--where location like '%Ethiopia%'

Group by location, population

Order by PercentagePopulationInfected desc

The output is right at June 26 the highest the infection got is 487,430 : 

1. Showing Countries with highest Death Count per Population

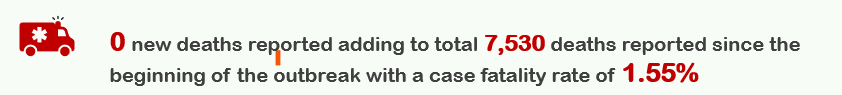
Select location, max(cast (total\_deaths as bigint)) as TotalDeathCount

from PortfolioProject..CovidDeaths

--where location like '%Ethiopia%'

Group by location

Order by TotalDeathCount desc



The above is the proof at June 26 total death in ethiopia reported to be 7,530

1. **Global numbers (stopped at minute 50)**

Select sum(new\_cases) as totalcases, sum(cast (new\_deaths as bigint)) as totaldeaths, sum(cast (new\_deaths as bigint))/sum(new\_cases)\*100 as deathpercentage

from PortfolioProject..CovidDeaths

where continent is not null

--group by date

Order by 1,2

1. **Looking at Total Population vs. Vaccination**

Select Dea.continent,Dea.location,Dea.population,Dea.date,Vac.new\_vaccinations,sum (convert (bigint, vac.new\_vaccinations))

over (partition by Dea.location order by Dea.location, Dea.date) as PeopleVaccinated

from PortfolioProject..CovidDeaths as Dea

Join PortfolioProject..CovidVaccination as Vac

on Dea.location = Vac.location

and Dea.date = Vac.date

where Dea.continent is not null

Order by 2,3

!! Verify with the Ethiopia data - released by the UN !!

1. **Use CTE (Common Table Expression) to calculate the percentage of People vaccinated**

With PopvsVac (continent, location, population, date, new\_vaccination, PeopleVaccinated)

 as

 (

 Select Dea.continent,Dea.location,Dea.population,Dea.date,Vac.new\_vaccinations,sum (convert (bigint, vac.new\_vaccinations))

over (partition by Dea.location order by Dea.location, Dea.date) as PeopleVaccinated

from PortfolioProject..CovidDeaths as Dea

Join PortfolioProject..CovidVaccination as Vac

on Dea.location = Vac.location

and Dea.date = Vac.date

where Dea.location like '%eth%'

--where Dea.continent is not null

--Order by 2,3

)

Select \*, (PeopleVaccinated/population)\*100

from PopvsVac

**Verified, currently 19.55 % of the population is vaccinated in Ethiopia (googled)**

1. **Temp table to get percentage of population Vaccinated**

Drop table if exists #PercentagePopVaccinated

Create table #PercentagePopVaccinated

(

continent nvarchar(255),

location nvarchar(255),

population numeric,

Date datetime,

new\_vaccinations numeric,

PeopleVaccinated numeric

)

Insert into #PercentagePopVaccinated

Select Dea.continent,Dea.location,Dea.population, Dea.date,Vac.new\_vaccinations,sum (convert (bigint, vac.new\_vaccinations))

over (partition by Dea.location order by Dea.location, Dea.date) as PeopleVaccinated

from PortfolioProject..CovidDeaths as Dea

Join PortfolioProject..CovidVaccination as Vac

on Dea.location = Vac.location

and Dea.date = Vac.date

--where Dea.location like '%eth%'

where Dea.continent is not null

--Order by 2,3

Select \*, (PeopleVaccinated/population)\*100

from #PercentagePopVaccinated

1. **Creating view to store data for future visualization**

Create view PercentagePopulationVaccinated as

Select Dea.continent,Dea.location,Dea.population, Dea.date,Vac.new\_vaccinations,sum (convert (bigint, vac.new\_vaccinations))

over (partition by Dea.location order by Dea.location, Dea.date) as PeopleVaccinated

from PortfolioProject..CovidDeaths as Dea

Join PortfolioProject..CovidVaccination as Vac

on Dea.location = Vac.location

and Dea.date = Vac.date

--where Dea.location like '%eth%'

where Dea.continent is not null

--Order by 2,3