**PORTAFOLIO PROJECT # 1 USING SQL**

Created By: Rodrigo Davalos  
Google Cloud BigQuery  
Link to Query:  
https://console.cloud.google.com/bigquery?sq=736774603777:701ccc3957d0479c902485fe351df7c1

-- I will start at the very basics on how to query a table from a basic and simple query to a more advanced query. The whole purpose of this project is to showcase to a potential employer that I can also do more advanced queries,

-- To see preview if the data is correct first I did a simple query:

SELECT \*

FROM my-project-001-353521.PortfolioProject.CovidDeaths

-- and then add the order by , just to make sure we have it as we do in our sheets document.

-- As a reminder Order By 1, 2, 3,refers to the sequence of columns in the select list

SELECT \*

FROM my-project-001-353521.PortfolioProject.CovidDeaths

ORDER BY 3, 4

-- Then I did the same with CovidVaccinations to make sure they match

SELECT \*

FROM my-project-001-353521.PortfolioProject.CovidVaccinations

ORDER BY 3, 4

-- Now I have our two tables and we can continue and I will start selecting data that I am going to be using ...

-- I'm using ORDER BY 1, 2 because I want to have my table organized and organize it first by location and date.

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

FROM my-project-001-353521.PortfolioProject.CovidDeaths

ORDER BY 1, 2

-- Looking at Total Cases vs Total Deaths

-- To find the Death Percentage I divided the total deaths/total cases x 100

-- This results shows the likelihood of dying if you contract covid in your country

SELECT location, date, total\_cases, total\_deaths, (total\_deaths/total\_cases)\*100 as Death\_Percentage

FROM my-project-001-353521.PortfolioProject.CovidDeaths

ORDER BY 1, 2

-- Now I want to find out the death percentage in the United States, therefore I add a WHERE location statement

-- by adding the '%States%' It will search all the words that have the 'States' word in the location column.

-- It shows in the United States how many people contracted covid and the death percentage some interesing data:

-- At the end of 2020 20,099,363 people contracted Covid and unfortunately 352,093 people died and the Death Percentage at the end of 2020 was 1.75176%

-- The highest Death Percentage in the United States happened in 2020-05-15 with a 6.259%

SELECT location, date, total\_cases, total\_deaths, (total\_deaths/total\_cases)\*100 as Death\_Percentage

FROM my-project-001-353521.PortfolioProject.CovidDeaths

WHERE LOCATION like '%States%'

ORDER BY 1, 2

-- Looking at the Total Cases vs Population

-- Shows what Percentage of the population in the United States contracted Covid

SELECT location, date, population, total\_cases,(total\_cases/population)\*100 as Percent\_Population\_Infected

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

ORDER BY 1, 2

-- Looking at Countries with Highest Infection Rate compared to population

-- I added MAX to count only the highest number of total cases and the highest percent of population in the world infected

-- I also used the GROUP BY to group my columns by location and population as well as ORDER BY Percent Population Infected desc , as I want to know which are

-- the countries with the highest population Infected with Covid

SELECT location, population, MAX(total\_cases) as Highest\_Infection\_count, MAX((total\_cases/population))\*100 as Percent\_Population\_Infected

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

GROUP BY location, population

ORDER BY Percent\_Population\_Infected desc

-- This is showing Countries with Highest Death Count Per Population

SELECT location, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

GROUP BY location

ORDER BY Total\_Death\_Count desc

-- An issue that is presenting, is that Non countries such as World, Europe, North America, European Union, etc show in our table

-- To fix this issue we will use in the WHERE the continent is not null statement

SELECT \*

FROM my-project-001-353521.PortfolioProject.CovidDeaths

WHERE continent is not null

ORDER BY 3,4

-- Now that the problem has been found, I can addthe WHERE continent is not null statement to my previous query.

-- Now the table shows only Countries

SELECT location, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is not null

GROUP BY location

ORDER BY Total\_Death\_Count desc

-- NOW, LETS BREAK THINGS DOWN BY CONTINENT

SELECT continent, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is not null

GROUP BY continent

ORDER BY Total\_Death\_Count desc

-- We have some small issues, as North America is only showing data from the United States

-- To look at the accurate data I selected Location again and in the Where statement I put continent is null which shows again the numbers from all the continents

SELECT location, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is null

GROUP BY location

ORDER BY Total\_Death\_Count desc

-- Showing continent with the highest death count per population

SELECT continent, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is not null

GROUP BY continent

ORDER BY Total\_Death\_Count desc

-- Let's calculate GLOBAL NUMBERS

-- This numbers gives the data of Total Cases, Total Deaths and Death Percentage per day across the world

SELECT date, SUM(new\_cases)as Total\_Cases, SUM(new\_deaths) as Total\_Deaths,

SUM(new\_deaths)/SUM(new\_cases)\*100 as Death\_Percentage

FROM my-project-001-353521.PortfolioProject.CovidDeaths

-- WHERE location like '%States%'

WHERE continent is not null

GROUP BY date

ORDER BY 1, 2

-- Now I removed the DATE the data will give all the total cases

-- Accross the world we have a Death Percentage of 2.11%

SELECT SUM(new\_cases)as Total\_Cases, SUM(new\_deaths) as Total\_Deaths,

SUM(new\_deaths)/SUM(new\_cases)\*100 as Death\_Percentage

FROM my-project-001-353521.PortfolioProject.CovidDeaths

-- WHERE location like '%States%'

WHERE continent is not null

-- GROUP BY date

ORDER BY 1, 2

-- I am now using the two tables using the JOIN statement to join location and date

SELECT \*

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

-- I can see that both tables where joint correclty

-- In this Next Query I'm now looking at the TOTAL POPULATION vs VACCINATIONS

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

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null

ORDER BY 2, 3

-- Now I want to add up a ROLLING COUNT

-- I am using the "partition by" to break it down by location first.

-- Also I added the "ORDER BY" after the partition so It can also order by location and date when counting new vaccinations and rolling the count by location.

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

SUM(CAST(vac.new\_vaccinations as int64)) OVER (partition by dea.location ORDER BY dea.location, dea.date) as Rolling\_People\_Vaccinated

--, (Rolling\_People\_Vaccinated/population)\*100 <-- To use that I need to create a CTE or a TEMP TABLE

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null

ORDER BY 2, 3

-- USING CTE

-- When using a CTE I use the "WITH" statement

-- By using the CTE now I can use the formula and add it to the JOINT table

WITH PopvsVac --(continent, location, date, population, new\_vaccinations, Rolling\_People\_Vaccinated)

as

(

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

SUM(CAST(vac.new\_vaccinations as int64)) OVER (partition by dea.location ORDER BY dea.location, dea.date) as Rolling\_People\_Vaccinated

--, (Rolling\_People\_Vaccinated/population)\*100

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null

-- ORDER BY 2, 3

)

SELECT \*, (Rolling\_People\_Vaccinated/population)\*100 as People\_Vaccinated\_Percentage

FROM PopvsVac

--The Second way I'm Using a TEMP TABLE with same results:  
-- The difference between this TEMP TABLE and the CTE TEMP TABLE, is that This TEMP

-- TABLE is available during the database session, and the CTE TEMP TABLE or SUB TABLE -- is only available during for that one Query

CREATE TEMP TABLE PopvsVac as

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

SUM(CAST(vac.new\_vaccinations as int64)) OVER (partition by dea.location ORDER BY dea.location, dea.date) as Rolling\_People\_Vaccinated

--, (Rolling\_People\_Vaccinated/population)\*100

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null;

-- ORDER BY 2, 3)

SELECT \*, (Rolling\_People\_Vaccinated/population)\*100 as People\_Vaccinated\_Percentage

FROM PopvsVac

CREATE TABLE `my-project-001-353521.PortfolioProject.People\_Vaccinated\_Percentage`(

Continent varchar(255),

Location varchar(255),

Date datetime,

Population numeric,

New\_Vaccinations numeric,

Rolling\_People\_Vaccinated numeric

);

INSERT INTO `my-project-001-353521.PortfolioProject.People\_Vaccinated\_Percentage`

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

SUM(CAST(vac.new\_vaccinations as int64)) OVER (partition by dea.location ORDER BY dea.location, dea.date) as Rolling\_People\_Vaccinated

--, (Rolling\_People\_Vaccinated/population)\*100

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null;

-- ORDER BY 2, 3)

SELECT \*, (Rolling\_People\_Vaccinated/population)\*100 as People\_Vaccinated\_Percentage

FROM `my-project-001-353521.PortfolioProject.People\_Vaccinated\_Percentage`

-- CREATING VIEW TO STORE DATA FOR LATER VISUALIZATIONS

-- VIEW #1: PERCENTAGE OF VACCINATED PEOPLE

CREATE VIEW `my-project-001-353521.PortfolioProject.PeopleVaccinatedPercentage` as

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

SUM(CAST(vac.new\_vaccinations as int64)) OVER (partition by dea.location ORDER BY dea.location, dea.date) as Rolling\_People\_Vaccinated

--, (Rolling\_People\_Vaccinated/population)\*100

FROM `my-project-001-353521.PortfolioProject.CovidDeaths` as dea

JOIN `my-project-001-353521.PortfolioProject.CovidVaccinations` as vac

ON dea.location = vac.location AND dea.date = vac.date

WHERE dea.continent is not null;

-- ORDER BY 2, 3)

-- VIEW #2: TOTAL DEATHS BY COUNTRY

CREATE VIEW `my-project-001-353521.PortfolioProject.TotalDeathByCountry` as

SELECT location, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is not null

GROUP BY location

ORDER BY Total\_Death\_Count desc

SELECT \*

FROM `my-project-001-353521.PortfolioProject.TotalDeathByCountry`

-- VIEW #3: TOTAL DEATHS BY CONTINENT

CREATE VIEW `my-project-001-353521.PortfolioProject.TotalDeathByContinent` as

SELECT continent, MAX(total\_deaths) as Total\_Death\_Count

FROM my-project-001-353521.PortfolioProject.CovidDeaths

--Where location like '%States%'

WHERE continent is not null

GROUP BY continent

ORDER BY Total\_Death\_Count desc

SELECT \*

FROM `my-project-001-353521.PortfolioProject.TotalDeathByContinent`