

```
-- This is a data set that was downloaded by https://ourworldindata.org/covid-cases &
https://ourworldindata.org/covid-deaths.
```

```
-- Data Exploration will begin with a query to see if the dataset is uploaded.
```

```
SELECT *
FROM `project-000-392922.CDD.covid_data_death`
```

```
-- Since we have our data uploaded we can answer a few questions using the dataset.
```

```
## REFERENCE QUERY
```

```
SELECT location, date, total_cases, new_cases, total_deaths, population
FROM `project-000-392922.CDD.covid_data_death`
ORDER BY 1,2
```

```
////////////////////////////////////
////////////////////////////////////
//
```

```
-- 1. Total Cases vs Total Deaths
```

```
-- What is the percent of people who died of Covid in the United States ?
```

```
SELECT location, date, total_cases, total_deaths, (total_deaths/total_cases) * 100 as
Death_Percentage
FROM `project-000-392922.CDD.covid_data_death`
Where location like '%State%' AND location is not null
ORDER BY 1,2 DESC
```

```
## Comment: DESC is placed to see the current death percentage of people who died from
Covid.
```

```
////////////////////////////////////
////////////////////////////////////
//
```

```
-- 2. Total Cases vs Population
```

-- What is the Percent on the poulation of people who got Covid in the United States?

```
SELECT location, date, total_cases, population, (total_cases/population) * 100 as
Percent_of_Population_with_Covid
FROM `project-000-392922.CDD.covid_data_death`
Where location like '%State%' AND location is not null
ORDER BY 1,2 DESC
```

```
////////////////////////////////////
////////////////////////////////////
//
```

-- 3. Countries Infection/Death Rate

-- 3A. Which Country has the Highest Infection rate in comparrison to population ?.

```
SELECT location, population, MAX(total_cases) AS Highest_Infection_Count,
MAX(total_cases)/(population) * 100 as Percent_of_Population_Infected
FROM `project-000-392922.CDD.covid_data_death`
Where location is not null
GROUP BY location, population
ORDER BY Percent_of_Population_Infected DESC
```

Comments: In order to find the Highest Infection rate I aggrigated the function where we can have the MAX total of cases representing the Infection count. To Keep the Table organized before a Query, its best to use the GROUP BY on location and population. To find Which location has the Highest Infection rate in comparison to population. DESC is used in the ORDER BY clause to see the the highest percent of population infected within the location.

--3B. Which countries has the highest death count?

```
SELECT location, MAX(total_deaths) AS Total_Death_Count
FROM `project-000-392922.CDD.covid_data_death`
Where continent is not null AND location is not null
GROUP BY location
```

```
ORDER BY Total_Death_Count DESC
```

#Comment: In the Data set there were some nulls that affected the grouping of location and continent. With the WHERE clause it's best to have continent and location is not null. With all other script it's best to have that Where clause to insure data integrity on accurate results from functions.

--3C. Which Continent has the Highest Death Count?

```
SELECT continent, MAX(total_deaths) AS Total_Death_Count
FROM `project-000-392922.CDD.covid_data_death`
Where continent is not null AND location is not null
GROUP BY continent
ORDER BY Total_Death_Count DESC
```

```
////////////////////////////////////
////////////////////////////////////
//
```

-- 4.Total Population vs Vaccinations.

-- What is the Global percent of people who are vaccinated?.

--PART 1/2 A TEMP TABLE will be created because the system cannot use a column that is just created to have it used in a next column. So when creating a TEMP TABLE a PARTITION is needed for location and date. Once SQL reaches a new location we want the system to reset the count. Once our Aggregated function is applied it will not continuously run where our numbers are ruined. An INSERT INTO function will be applied to connect the TEMPORARY TABLE with two different datasets. JOIN function will be used to execute this query.

```
CREATE OR REPLACE TEMP TABLE POPVSVAC
(
continent string,
location string,
date datetime,
```

```

population FLOAT64,
new_vaccinations FLOAT64,
People_Vaccinated FLOAT64);
INSERT INTO POPVSVAC
SELECT A.continent, A.location, A.date, A.population, B.new_vaccinations,
Sum(B.new_vaccinations) OVER (PARTITION BY A.location ORDER BY A.location,A.date) as
People_Vaccinated
FROM `project-000-392922.CDD.covid_data_cases` B
JOIN `project-000-392922.CDD.covid_data_death` A
on B.location=A.location AND B.date = A.date
where A.continent is not null

```

-- PART 2/2 After creating a TEMP TABLE we can now find the percent of population who are vaccinated.

```

SELECT *, (People_Vaccinated)/(population)*100 AS Global_Percent
FROM `project-000-392922._scriptcd5139e2214133ebf0955165d99a9611ef042ecb.POPVSVAC`
Where continent is not null AND location is not null
ORDER BY continent

```

```

////////////////////////////////////
////////////////////////////////////
//

```

This query will result in percentage of people who are vaccinated in different location of the continent.
A rolling number will appear in the query that adds its total Global Percentage of people who are vaccinated.

-- Creating View to store Data for later Visualizations

```

CREATE VIEW IF NOT EXISTS
`project-000-392922._scriptcd5139e2214133ebf0955165d99a9611ef042ecb.POPVSVAC` As
SELECT A.continent, A.location, A.date, A.population, B.new_vaccinations,
Sum(B.new_vaccinations) OVER (PARTITION BY A.location ORDER BY A.location,A.date) as
People_Vaccinated
FROM `project-000-392922.CDD.covid_data_cases` B
JOIN `project-000-392922.CDD.covid_data_death` A

```

```
on B.location=A.location AND B.date = A.date  
where A.continent is not null
```