SQL Data exploration Project in Snowflake

Covid data from *ourworldindata.org* was used to create tables in snowflake.

Link to Dataset: [https://ourworldindata.org/covid-deaths](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbHdwV0dRZU5ZM2RrTjZkN01pc1RjcDBWWWwxUXxBQ3Jtc0tsVjFTcS1GT3YydURMTnVpbGoxVGFneVBidVF0LXNlQUhUNTFoOXVRMVltQnFNMDRpbF9VS3lGSDV6bDZ1Wkt6eEF3LXl5NkhZcXBVNmJhX3N3ckoxMjdIMXRGVHAtWWxsQ3NzT3NQWm1oU2JDbWk2WQ&q=https%3A%2F%2Fourworldindata.org%2Fcovid-deaths&v=qfyynHBFOsM)

***Skills used:*** Joins, CTE's, Temp Tables, Windows Functions, Aggregate Functions, Creating Views, Converting Data Types

SELECT \* from covid\_deaths limit 100;

SELECT \* from covid\_vacc limit 100;

**--Select data that we are going to use**

Select "LOCATION", "DATE","TOTAL\_CASES","NEW\_CASES","TOTAL\_DEATHS","POPULATION" FROM COVID\_DEATHS ORDER BY 1,2;

**--Total cases vs total deaths**

Select "LOCATION", "DATE","TOTAL\_CASES","TOTAL\_DEATHS",("TOTAL\_DEATHS"/"TOTAL\_CASES")\*100 as "Death Percentage"

FROM COVID\_DEATHS ORDER BY 1,2;

**--Total cases vs total deaths filter by country**

Select "LOCATION", "DATE","TOTAL\_CASES","TOTAL\_DEATHS",("TOTAL\_DEATHS"/"TOTAL\_CASES")\*100 as "Death Percentage"

FROM COVID\_DEATHS Where "LOCATION" like 'India';

**--Total cases vs Population (Shows what percetage of population got covid)**

Select "LOCATION", "DATE","TOTAL\_CASES","POPULATION" ,("TOTAL\_CASES"/"POPULATION")\*100 as "TCases v Population"

FROM COVID\_DEATHS Where "LOCATION" like 'India';

**--Total Deaths vs Population (Shows what percentage of population died with Covid)**

Select "LOCATION", "DATE","TOTAL\_DEATHS","POPULATION" ,("TOTAL\_DEATHS"/"POPULATION")\*100 as "Death vs Population" FROM COVID\_DEATHS Where "LOCATION" like 'India';

**--Find countries with highest infection rate**

Select "LOCATION", "POPULATION", MAX("TOTAL\_CASES")as "Highest Infection Count", MAX("TOTAL\_CASES"/"POPULATION")\*100 as "TCases v Population"

FROM COVID\_DEATHS

Group BY "LOCATION", "POPULATION"

ORDER BY "TCases v Population" DESC;

**--Showing countries with highest death count**

SELECT "LOCATION", max("TOTAL\_DEATHS") from COVID\_DEATHS

where "CONTINENT" is not NULL

GROUP BY "LOCATION"

ORDER BY max("TOTAL\_DEATHS") DESC;

**--Breaking down with continents with highest death count**

SELECT "CONTINENT", max("TOTAL\_DEATHS") from COVID\_DEATHS

where "CONTINENT" is not NULL

GROUP BY "CONTINENT"

ORDER BY max("TOTAL\_DEATHS") DESC;

**-- Getting Global death percentage by date**

SELECT "DATE", SUM("NEW\_CASES")as "Total Cases",SUM("TOTAL\_DEATHS") as "Total Deaths",(SUM("NEW\_DEATHS")/SUM("TOTAL\_DEATHS"))\*100 as "Death Percentage" FROM COVID\_DEATHS

WHERE "CONTINENT" is not null

GROUP BY "DATE";

**--Joining of 2 tables Covid daths and Covid vaccinations**

SELECT \* from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date;

**--Getting data for Total vaccinations vs Populations**

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date

where covid\_deaths.continent is not null

order by 2,3;

**--Getting data for Total vaccinations vs Populations for specific location**

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date

where covid\_deaths.LOCATION like 'Canada'

order by 2,3;

**--Rolling total of new vaccinations by location and date**

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS,

Sum(covid\_vacc.new\_vaccinations) OVER (Partition by covid\_deaths.LOCATION ORDER BY covid\_deaths.LOCATION,covid\_deaths.DATE ) as "Rolling total"

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date

where covid\_deaths.continent is not null

order by 2,3;

**--Rolling total of new vaccinations by location and date**

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS as "N\_Vaccs",

Sum(covid\_vacc.new\_vaccinations) OVER (Partition by covid\_deaths.LOCATION ORDER BY covid\_deaths.LOCATION,covid\_deaths.DATE ) as "Rolling total"

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date

where covid\_deaths.LOCATION like 'Canada'

order by 2,3;

**--Common Table Expression (CTE)**

**-- Goal is the get the percentage of the people getting vaccinated. As we are joining 2 tables we need to create a CTE first and then the calculation**

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

as

(SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS,

Sum(covid\_vacc.new\_vaccinations) OVER (Partition by covid\_deaths.LOCATION ORDER BY covid\_deaths.LOCATION,covid\_deaths.DATE ) as "Rolling\_total"

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date

where covid\_deaths.LOCATION like 'Canada'

)

Select \* ,(Rolling\_Total/Population)\*100 as "Vacc%" from PopvsVac;

**--Creating Temp table to get percentage of people vaccinated**

DROP Table if exists EDP\_GTMA.GBS\_SCHEMA\_EXTERNAL.PercentPeopleVaccinated;

CREATE TABLE EDP\_GTMA.GBS\_SCHEMA\_EXTERNAL.PercentPeopleVaccinated(

Continent nvarchar(255),

Location nvarchar(255),

Date datetime,

Population numeric,

New\_vaccinations numeric,

Rolling\_Total numeric

)

INSERT INTO PercentPeopleVaccinated

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS,

Sum(covid\_vacc.new\_vaccinations) OVER (Partition by covid\_deaths.LOCATION ORDER BY covid\_deaths.LOCATION,covid\_deaths.DATE ) as "Rolling\_total"

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date;

--where covid\_deaths.LOCATION like 'Canada'

Select \* ,(Rolling\_Total/Population)\*100 as "Vacc%" from PercentPeopleVaccinated;

**--Creating a view**

Create VIEW EDP\_GTMA.GBS\_SCHEMA\_EXTERNAL.PercentageofVaccPopulation as

SELECT covid\_deaths.CONTINENT,covid\_deaths.LOCATION,covid\_deaths.DATE,covid\_deaths.POPULATION,covid\_vacc.NEW\_VACCINATIONS,

Sum(covid\_vacc.new\_vaccinations) OVER (Partition by covid\_deaths.LOCATION ORDER BY covid\_deaths.LOCATION,covid\_deaths.DATE ) as "Rolling\_total"

from covid\_deaths

join covid\_vacc

on covid\_deaths.location = covid\_vacc. location

AND

covid\_deaths.date = covid\_vacc.date;