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 percetage 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 TABLEIF 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.