/\* Write SQL queries to extract data for exploration which late used for Tableau Project\*/

Use JOINS to combine two table to see how each country was managing vaccination drive in order reduce Covid infections then lastly create temp table to apply query on about percent of population vaccinated

WITH PopvsVac

as

(

SELECT dea.continent,dea.location,dea.date,dea.population,vac.new\_vacc,SUM(vac.new\_vacc) OVER (PARTITION BY dea.location order by dea.location,dea.date) as rolling\_vac\_count

FROM `massive-anagram-341913.covid19\_start.deaths` AS dea

JOIN

     `massive-anagram-341913.covid19\_start.vaccination` AS vac

    ON dea.location = vac.location

    and dea.date = vac.date

WHERE dea.continent is not null

--order by 2,3

)

SELECT \*,(rolling\_vac\_count/population)\*100

FROM PopvsVac;

Calculated global death percentage from the people who contracted Covid using two variables of total cases and total deaths

Select SUM(new\_cases) as total\_cases, SUM(new\_deaths) as total\_deaths, SUM(new\_deaths)/SUM(New\_Cases)\*100 as DeathPercentage

FROM `massive-anagram-341913.covid19\_start.deaths`

--Where location like '%South Africa%'

where continent is not null

--Group By date

order by 1,2;

Created a View to store the about results for later view in my visualization phase

Create View `massive-anagram-341913.covid19\_start.DeathPercentage`

AS

Select SUM(new\_cases) as total\_cases, SUM(new\_deaths) as total\_deaths, SUM(new\_deaths)/SUM(New\_Cases)\*100 as DeathPercentage

FROM `massive-anagram-341913.covid19\_start.deaths`

--Where location like '%South Africa%'

where continent is not null

--Group By date

order by 1,2;

Calculating the total death count in each continent to see which region was affected the most.

Select location, SUM(cast(new\_deaths) as TotalDeathCount

FROM `massive-anagram-341913.covid19\_start.deaths`

--Where location like '%states%'

Where continent is null

and location not in ('World', 'European Union', 'International','High income','Upper middle income','Low income')

Group by location

order by TotalDeathCount desc;

Looking for countries that had high percentage of it population infected by Covid

Select Location, Population,date, MAX(total\_cases) as HighestInfectionCount, Max((total\_cases/population))\*100 as PercentPopulationInfected

FROM `massive-anagram-341913.covid19\_start.deaths`

--Where location like '%South Africa%'

Group by Location, Population, date

order by PercentPopulationInfected desc;

How was the above percentage of people infected increasing over time period

Select Location,date, Population, MAX(total\_cases) as HighestInfectionCount, Max((total\_cases/population))\*100 as PercentPopulationInfected

From `massive-anagram-341913.covid19\_start.deaths`

--Where location like '%South Africa%'

Group by Location, Population

order by PercentPopulationInfected desc;