COVID-19 Portfolio Project- Data exploration and Data Manipulation using SQL

Let's have an overview of the Covid-19 Data:

Select all the data

MySQL> select * from CovidDeaths;

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	new_cases_per_million	new_cases_smoothed_per_million	total_deaths_per_million
2	AFG	Asia	Afghanistan	2020-02-24	38928341	1	1	0	0	0	0	0.026	0.026	0	0
3	AFG	Asia	Afghanistan	2020-02-25	38928341	1	0	0	0	0	0	0.026	0	0	0
4	AFG	Asia	Afghanistan	2020-02-26	38928341	1	0	0	0	0	0	0.026	0	0	0
5	AFG	Asia	Afghanistan	2020-02-27	38928341	1	0	0	0	0	0	0.026	0	0	0
6	AFG	Asia	Afghanistan	2020-02-28	38928341	1	0	0	0	0	0	0.026	0	0	0
7	AFG	Asia	Afghanistan	2020-02-29	38928341	1	0	0.143	0	0	0	0.026	0	0.004	0
8	AFG	Asia	Afghanistan	2020-03-01	38928341	1	0	0.143	0	0	0	0.026	0	0.004	0
9	AFG	Asia	Afghanistan	2020-03-02	38928341	1	0	0	0	0	0	0.026	0	0	0
10	AFG	Asia	Afghanistan	2020-03-03	38928341	2	1	0.143	0	0	0	0.051	0.026	0.004	0
11	AFG	Asia	Afghanistan	2020-03-04	38928341	4	2	0.429	0	0	0	0.103	0.051	0.011	0

Total cases vs Total Deaths

MySQL> select location, date, total_cases, total_deaths, (total_deaths/total_cases) * 100 as DeathRatio from CovidDeaths where location like '%states%' order by DeathRatio desc;

location	date	total_cases	total_deaths	DeathRatio
United States	2020-03-02	55	6	10.9091
United States	2020-03-04	107	11	10.2804
United States	2020-03-03	74	7	9.4595
United States	2020-03-05	184	12	6.5217
United States	2020-05-15	1453214	90966	6.2596
United States	2020-05-14	1428467	89298	6.2513
United States	2020-05-13	1401649	87502	6.2428
United States	2020-05-16	1477373	92164	6.2384
United States	2020-05-08	1295019	80688	6.2306
United States	2020-05-07	1268180	78925	6.2235

Likelihood of contracting COVID-19

MySQL> Select Location, date, total_cases, total_deaths, (total_deaths/total_cases)*100 as DeathRatio from CovidDeaths Where location like '%states%' and continent != '' order by date desc, DeathRatio desc;

From the image below,

#1:

The initial peak covid data shows that the total cases skyrocketed from 55 to 107.

#2:

The total cases peaked on November 8, 2020 with over 10M+ cases and a DeathRatio of 2.37%.

#3:

The DeathRatio also plummeted to 1.7% in the later stage.

This shows that you have 1.7% chance of contracting COVID-19

#1

Location	date	total_cases	total_deaths	DeathRatio
United States	2020-03-02	55	6	10.9091
United States	2020-03-04	107	11	10.2804

#2

Location	date	total_cases	total_deaths	DeathRatio
United States	2020-11-08	10087380	239540	2.3747

#3

Location	date	total_cases	total_deaths	DeathRatio
United States	2021-04-30	32346971	576232	1.7814

Shows the percentage of population got covid

MySQL> select location, date, population, total_cases, (total_cases/population)
* 100 as DeathRatio from CovidDeaths where location like '%states%' order by DeathRatio desc;

location	date	population	total_cases	DeathRatio
United States	2021-04-30	331002647	32346971	9.7724
United States	2021-04-29	331002647	32289049	9.7549
United States	2021-04-28	331002647	32230850	9.7373
United States	2021-04-27	331002647	32175725	9.7207
United States	2021-04-26	331002647	32124869	9.7053
United States	2021-04-25	331002647	32077178	9.6909
United States	2021-04-24	331002647	32045113	9.6812
United States	2021-04-23	331002647	31991750	9.6651
United States	2021-04-22	331002647	31929351	9.6463
United States	2021-04-21	331002647	31862094	9.6259

Looking at countries with highest infection rate vs population

MySQL> select location, population, Max(total_cases) as HighestInfectionCount, Max((total_cases/population)) * 100 as PercentPopulationInfected from CovidDeaths group by location, population order by PercentPopulationInfected desc;

location	population	HighestInfectionCount	PercentPopulationInfected
Andorra	77265	13232	17.1300
Montenegro	628062	97389	15.5100
Czechia	10708982	1630758	15.2300
San Marino	33938	5066	14.9300
Slovenia	2078932	240292	11.5600
Luxembourg	625976	67205	10.7400
Bahrain	1701583	176934	10.4000
Serbia	6804596	689557	10.1300
United States	331002647	32346971	9.7700
Israel	8655541	838481	9.6900
Sweden	10099270	973604	9.6400
Estonia	1326539	122019	9.2000
Lithuania	2722291	247269	9.0800
Netherlands	17134873	1522973	8.8900
Belgium	11589616	990229	8.5400
Panama	4314768	364576	8.4500
France	68147687	5677835	8.3300
Portugal	10196707	836493	8.2000
Croatia	4105268	332183	8.0900

Showing the countries with highest death count per population

MySQL> select location, max(cast(total_deaths as UNSIGNED)) as TotalDeathCount from CovidDeaths where continent != '' group by location order by TotalDeathCount desc;

Breaking things down by continent

showing deaths count by continent

MySQL> select date, sum(new_cases) as TotalCases, sum(new_deaths) as TotalDeaths, sum(new_deaths) / sum(new_cases) * 100 as DeathPercent from CovidDeaths where continent != '' group by date order by 1,2; — globally

MySQL> select continent, max(cast(total_deaths as UNSIGNED)) as TotalDeathCount from CovidDeaths where continent != '' group by continent order by TotalDeathCount desc;— continent

MySQL> select sum(new_cases) as TotalCases, sum(new_deaths) as TotalDeaths, sum(new_deaths) / sum(new_cases) * 100 as DeathPercent from CovidDeaths where continent!= '' order by 1,2;— across the world

Looking for total pop vs Vacc

MySQL> With PopvsVacc (continent, location, date, population, new_vaccinations, RVacc) as (select d.continent, d.location, d.date, d.population, v.new_vaccinations, sum(cast(v.new_vaccinations as unsigned)) over (partition by d.location order by d.location, d.date) as RVacc from CovidDeaths as d JOIN CovidVaccinations as v on d.location = v.location and d.date = v.date where d.continent != '') select *, (RVacc/population) * 100 from PopvsVacc limit 1000;

Creating a temp table

MySQL> Drop temporary table if exists PercentPopVaccinated; create temporary table PercentPopVaccinated (continent varchar(50), location varchar(100), date DATE, population bigint, new_vaccinations bigint, RVacc bigint); insert into PercentPopVaccinated select d.continent, d.location, d.date, d.population, v.new_vaccinations, sum(cast(v.new_vaccinations as unsigned)) over (partition by d.location order by d.location, d.date) as RVacc from CovidDeaths as d JOIN CovidVaccinations as v on d.location = v.location and d.date = v.date; select *, RVacc/population * 100 as PercentVaccinated from PercentPopVaccinated limit 1000;

Creating a view

MySQL> create view PercentPopVaccinated as select d.continent, d.location, d.date, d.population, v.new_vaccinations, sum(cast(v.new_vaccinations as unsigned)) over (partition by d.location order by d.location, d.date) as RVacc from CovidDeaths as d JOIN CovidVaccinations as v on d.location =
v.location and d.date = v.date where d.continent != '' limit 1000;