COVID 19 DATA EXPLORATION

Project of: Hà Nguyễn Huy Hải

Git hub: https://github.com/haiuy

Linkedin: https://www.linkedin.com/in/h%E1%BA%A3i-huy-272aa1289/

Covid 19 dataset

The dataset used in this project is from the website Our World in Data, which provides official daily counts of COVID-19 cases, deaths and vaccine utilization reported by countries, territories and areas. The dataset is updated daily and is open-access, free and available for anyone to use for all purposes. Some of the features of this dataset are:

- **Cases**: How many new cases are being confirmed each day? How many cases have been confirmed since the pandemic started? How is the number of cases changing?
- **Deaths**: How many deaths from COVID-19 have been reported? Is the number of deaths rising or falling? How does the death rate compare to other countries?
- **Vaccinations**: How many vaccine doses are being administered each day? How many doses have been administered in total? What share of the population has been vaccinated?
- **Testing**: How much testing for coronavirus do countries conduct? How do confirmed cases compare to actual infections?
- **Hospitalizations**: How many people are being hospitalized for COVID-19? How does hospital capacity compare to the number of patients?
- **Policy responses**: How do government policy responses – on travel, testing, vaccinations, face coverings, and more – vary across the world?

Data exploration by SQL server

The first information that I want to explore is the death percentage of Viet Nam. Therefore, I utilize select function to pull four columns from the table: Location, date, total_cases, and total_deaths. Then, a new column called DeathPercentage is the ratio of total deaths to total cases multiplied by 100 is added to the select function to show the result.

- Select Location, date, total_cases, total_deaths, (convert (float, total_deaths)/NULLif(Convert(Float,total_cases),0))*100 as DeathPercentage
- From CovidDeaths$
- where location like 'vietnam'
- Order by 1,2

	Location	date	total_cases	total_deaths	DeathPercentage
12...	Vietnam	2023-07-20 00:00:00.000	11621489	43206	0.37177680071...
12...	Vietnam	2023-07-21 00:00:00.000	11621522	43206	0.37177574503...
12...	Vietnam	2023-07-22 00:00:00.000	11621558	43206	0.37177459338...
12...	Vietnam	2023-07-23 00:00:00.000	11621599	43206	0.37177328180...
12...	Vietnam	2023-07-24 00:00:00.000	11621626	43206	0.37177241807...
13...	Vietnam	2023-07-25 00:00:00.000	11621681	43206	0.37177065865...
13...	Vietnam	2023-07-26 00:00:00.000	11621708	43206	0.37176979493...
13...	Vietnam	2023-07-27 00:00:00.000	11621728	43206	0.37176915515...
13...	Vietnam	2023-07-28 00:00:00.000	11621752	43206	0.37176838741...
13...	Vietnam	2023-07-29 00:00:00.000	11621728	43206	0.37176915515...
13...	Vietnam	2023-07-30 00:00:00.000	11621798	43206	0.37176691592...
13...	Vietnam	2023-07-31 00:00:00.000	11621812	43206	0.37176646808...
13...	Vietnam	2023-08-01 00:00:00.000	11621840	43206	0.37176557240...
13...	Vietnam	2023-08-02 00:00:00.000	11621883	43206	0.37176419690...
13...	Vietnam	2023-08-03 00:00:00.000	11621909	43206	0.37176336520...
13...	Vietnam	2023-08-04 00:00:00.000	11621934	43206	0.37176256550...
13...	Vietnam	2023-08-05 00:00:00.000	11621965	43206	0.37176157388...
13...	Vietnam	2023-08-06 00:00:00.000	11621978	43206	0.37176115804...
13...	Vietnam	2023-08-07 00:00:00.000	11621997	43206	0.37176055027...
13...	Vietnam	2023-08-08 00:00:00.000	11622021	43206	0.37175978257...
13...	Vietnam	2023-08-09 00:00:00.000	11622053	43206	0.37175875897...

The infected rate is calculated by dividing total cases to population

- select location, date, total_cases, population,
  (convert(float,total_cases)/population)*100 as InfectedRate

	location	date	total_cases	population	InfectedRate
25	Vietnam	2020-01-27 00:00:00.000	2	98186856	2.0369325...
26	Vietnam	2020-01-28 00:00:00.000	2	98186856	2.0369325...
27	Vietnam	2020-01-29 00:00:00.000	2	98186856	2.0369325...
28	Vietnam	2020-01-30 00:00:00.000	2	98186856	2.0369325...
29	Vietnam	2020-01-31 00:00:00.000	2	98186856	2.0369325...
30	Vietnam	2020-02-01 00:00:00.000	5	98186856	5.0923312...
31	Vietnam	2020-02-02 00:00:00.000	6	98186856	6.1107975...
32	Vietnam	2020-02-03 00:00:00.000	7	98186856	7.1292638...
33	Vietnam	2020-02-04 00:00:00.000	7	98186856	7.1292638...
34	Vietnam	2020-02-05 00:00:00.000	10	98186856	1.0184662...
35	Vietnam	2020-02-06 00:00:00.000	10	98186856	1.0184662...
36	Vietnam	2020-02-07 00:00:00.000	12	98186856	1.2221595...
37	Vietnam	2020-02-08 00:00:00.000	13	98186856	1.3240061...
38	Vietnam	2020-02-09 00:00:00.000	13	98186856	1.3240061...
39	Vietnam	2020-02-10 00:00:00.000	14	98186856	1.4258527...
40	Vietnam	2020-02-11 00:00:00.000	14	98186856	1.4258527...
41	Vietnam	2020-02-12 00:00:00.000	15	98186856	1.5276993...
42	Vietnam	2020-02-13 00:00:00.000	15	98186856	1.5276993...
43	Vietnam	2020-02-14 00:00:00.000	16	98186856	1.6295460...
44	Vietnam	2020-02-15 00:00:00.000	16	98186856	1.6295460...
45	Vietnam	2020-02-16 00:00:00.000	16	98186856	1.6295460...

Next, I calculate the infected rate of each country in order to find out which nation has the highest ratio of infected. In this code, I divide total cases by population multiply by 100 to get the infected rate and select the maximum values with descending order to create a global ranking table. The result shows that Cyprus has the highest Infected Rate.

	location	population	HighestInfectionCount	InfectedRate
1	Cyprus	896007	999	73.7554505712567
2	San Marino	33690	9900	72.3716236271891
3	Brunei	449002	9828	69.1264181451308
4	Austria	8939617	989928	68.0262588430802
5	South Korea	51815808	99839	65.419296366082
6	Faeroe Islands	53117	998	65.2484138787959
7	Slovenia	2119843	997973	63.4324334396462
8	Gibraltar	32677	9969	62.8882700370291
9	Martinique	367512	99753	62.6793138727443
10	Andorra	79843	9972	60.1367684080007
11	Jersey	110796	9995	59.9218383335138
12	Luxembourg	647601	9664	59.152935217827
13	Saint Pierre and Miquelon	5885	995	58.2158028887001
14	Denmark	5882259	9938	58.0574741778626
15	France	67813000	9850650	57.5073953371772
16	Iceland	372903	9568	56.0051809719954
17	Guernsey	63329	996	55.7817113802523
18	Liechtenstein	39355	9965	54.5648583407445
19	Portugal	10270857	998388	54.5344268740184
20	Latvia	1850654	975826	52.7287110394488
21	Greece	10384972	97551	51.5422381495107

The highest death count belongs to the United States with 1127152 cases recorded

	location	HighestDeathCount
1	United States	1127152
2	Brazil	704659
3	India	531925
4	Russia	399938
5	Mexico	334336
6	United Kingdom	228707
7	Peru	221364
8	Italy	191167
9	Germany	174979
10	France	167985
11	Indonesia	161916
12	Iran	146321
13	Colombia	142961
14	Argentina	130472
15	Spain	121852
16	China	121628
17	Poland	119636
18	Ukraine	109904
19	South Africa	102595
20	Turkey	101419
21	Japan	74694

Looking from a continental scale, high income nations have a higher number of death case than those of upper and middle income. Europe, Asia, and the Americas have also experienced high death counts, ranging from 1.6 million to 2.1 million.

	location	HighestDeathCount
1	World	6955484
2	High income	2899122
3	Upper middle income	2665630
4	Europe	2075405
5	Asia	1632877
6	North America	1602820
7	South America	1356082
8	Lower middle income	1338801
9	European Union	1243032
10	Africa	259004
11	Low income	47979
12	Oceania	29291

After gathering information about the damage of Covid pandemic, I then turn to explore more on the global attempts to fight back the disease by looking further into the vaccination number. It is clearly seen that the UK just started their first vaccination campaign with 165844 doses in 01-11-2021.

	continent	location	date	population	new_vaccinations	CumulativeVaccination
30...	Europe	United Ki...	2021-01-07 00:00:00.000	67508936	NULL	NULL
30...	Europe	United Ki...	2021-01-08 00:00:00.000	67508936	NULL	NULL
30...	Europe	United Ki...	2021-01-09 00:00:00.000	67508936	NULL	NULL
30...	Europe	United Ki...	2021-01-10 00:00:00.000	67508936	NULL	NULL
30...	Europe	United Ki...	2021-01-11 00:00:00.000	67508936	165844	165844
30...	Europe	United Ki...	2021-01-12 00:00:00.000	67508936	223726	389570
30...	Europe	United Ki...	2021-01-13 00:00:00.000	67508936	288688	678258
30...	Europe	United Ki...	2021-01-14 00:00:00.000	67508936	321951	1000209
30...	Europe	United Ki...	2021-01-15 00:00:00.000	67508936	345525	1345734
30...	Europe	United Ki...	2021-01-16 00:00:00.000	67508936	298751	1644485
30...	Europe	United Ki...	2021-01-17 00:00:00.000	67508936	192346	1836831
30...	Europe	United Ki...	2021-01-18 00:00:00.000	67508936	208641	2045472
30...	Europe	United Ki...	2021-01-19 00:00:00.000	67508936	346922	2392394
30...	Europe	United Ki...	2021-01-20 00:00:00.000	67508936	366919	2759313
30...	Europe	United Ki...	2021-01-21 00:00:00.000	67508936	412615	3171928
30...	Europe	United Ki...	2021-01-22 00:00:00.000	67508936	480069	3651997
30...	Europe	United Ki...	2021-01-23 00:00:00.000	67508936	493013	4145010

Finally I use temp table or CTE function in order to save those results and create view to store data for later visualization.