

[1] LOAD DATASET CovidDeaths AS csv FROM CovidDeaths.csv @ artifact file 174

On branch default

Console ▾ Timing Datasets ▾ Charts ▾

coviddeaths (232799 rows)

Views

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths
0	AFG	Asia	Afghanistan	2/24/2020	41128772	5	5		
1	AFG	Asia	Afghanistan	2/25/2020	41128772	5	0		
2	AFG	Asia	Afghanistan	2/26/2020	41128772	5	0		
3	AFG	Asia	Afghanistan	2/27/2020	41128772	5	0		
4	AFG	Asia	Afghanistan	2/28/2020	41128772	5	0		
5	AFG	Asia	Afghanistan	2/29/2020	41128772	5	0	0.7139999866485596	
6	AFG	Asia	Afghanistan	3/1/2020	41128772	5	0	0.7139999866485596	
7	AFG	Asia	Afghanistan	3/2/2020	41128772	5	0	0	
8	AFG	Asia	Afghanistan	3/3/2020	41128772	5	0	0	
9	AFG	Asia	Afghanistan	3/4/2020	41128772	5	0	0	
10	AFG	Asia	Afghanistan	3/5/2020	41128772	5	0	0	
11	AFG	Asia	Afghanistan	3/6/2020	41128772	5	0	0	
12	AFG	Asia	Afghanistan	3/7/2020	41128772	8	3	0.4289999022483826	
13	AFG	Asia	Afghanistan	3/8/2020	41128772	8	0	0.4289999022483826	
14	AFG	Asia	Afghanistan	3/9/2020	41128772	8	0	0.4289999022483826	
15	AFG	Asia	Afghanistan	3/10/2020	41128772	8	0	0.4289999022483826	
16	AFG	Asia	Afghanistan	3/11/2020	41128772	11	3	0.8569999933242798	
17	AFG	Asia	Afghanistan	3/12/2020	41128772	11	0	0.8569999933242798	
18	AFG	Asia	Afghanistan	3/13/2020	41128772	11	0	0.8569999933242798	
19	AFG	Asia	Afghanistan	3/14/2020	41128772	14	3	0.8569999933242798	

[2]

Console ▾ Timing Datasets ▾ Charts ▾

Understanding the Dataset and the attributes of the Dataset

Columns

- 1. iso_code: This likely represents the ISO code for countries or regions.
- 2. continent: The continent to which the location belongs.
- 3. location: Specific location or country.
- 4. date: Timestamp for the data recorded.
- 5. population: Total population of the location.
- 6. total_cases: Cumulative count of COVID-19 cases until the mentioned date.
- 7. new_cases: Count of new COVID-19 cases reported on the mentioned date.
- 8. new_cases_smoothed: Smoothed count of new cases, possibly averaged over a certain period to reduce daily fluctuations.
- 9. total_deaths: Cumulative count of COVID-19 related deaths until the mentioned date.
- 10. new_deaths: Count of new COVID-19 related deaths reported on the mentioned date.
- 11. new_deaths_smoothed: Smoothed count of new deaths, similar to new cases smoothed.
- 12. total_cases_per_million: Total cases standardized per million people in the population.
- 13. new_cases_per_million: New cases standardized per million people in the population.
- 14. new_cases_smoothed_per_million: Smoothed new cases standardized per million people.
- 15. total_deaths_per_million: Total deaths standardized per million people in the population.
- 16. new_deaths_per_million: New deaths standardized per million people in the population.
- 17. new_deaths_smoothed_per_million: Smoothed new deaths standardized per million people.
- 18. reproduction_rate: Estimated rate of COVID-19 transmission, indicating the average number of people to whom an infected person will transmit the virus.
- 19. icu_patients: Count of COVID-19 patients in intensive care units (ICUs).
- 20. icu_patients_per_million: ICU patients standardized per million people in the population.
- 21. hosp_patients: Count of COVID-19 patients in hospitals.
- 22. hosp_patients_per_million: Hospitalized patients standardized per million people in the population.
- 23. weekly_icu_admissions: Count of weekly admissions to ICUs due to COVID-19.
- 24. weekly_icu_admissions_per_million: Weekly ICU admissions standardized per million people.



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾

These columns seem to contain a comprehensive range of information regarding COVID-19 spread, impact on healthcare systems, and population statistic Cleaning, mapping, and integrating this data will involve tasks like handling missing values, ensuring data consistency, transforming and aggregating data as needed for analysis or modeling, and potentially merging this dataset with other relevant datasets for a more comprehensive analysis.

[3]

```
SELECT *
FROM coviddeaths
WHERE
    iso_code IS NOT NULL
    AND continent IS NOT NULL
    AND location IS NOT NULL
    AND date IS NOT NULL
    AND population IS NOT NULL
    AND total_cases IS NOT NULL
    AND new_cases IS NOT NULL
    AND new_cases_smoothed IS NOT NULL
    AND total_deaths IS NOT NULL
    AND new_deaths IS NOT NULL
    AND new_deaths_smoothed IS NOT NULL
    AND total_cases_per_million IS NOT NULL
    AND new_cases_per_million IS NOT NULL
    AND new_cases_smoothed_per_million IS NOT NULL
    AND total_deaths_per_million IS NOT NULL
    AND new_deaths_per_million IS NOT NULL
    AND new_deaths_smoothed_per_million IS NOT NULL
    AND reproduction_rate IS NOT NULL
    AND icu_patients IS NOT NULL
    AND icu_patients_per_million IS NOT NULL
    AND hosp_patients IS NOT NULL
    AND hosp_patients_per_million IS NOT NULL
    AND weekly_icu_admissions IS NOT NULL
    AND weekly_icu_admissions_per_million IS NOT NULL
    AND weekly_hosp_admissions IS NOT NULL
    AND weekly_hosp_admissions_per_million IS NOT NULL;
```

Console ▾ Timing Datasets ▾ Charts ▾

temporary_dataset (5582 rows)

Views

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths
0	BGR	Europe	Bulgaria	5/3/2020	6781955	1618	24	45.42900085449219	73
1	BGR	Europe	Bulgaria	5/10/2020	6781955	1965	44	49.57099914550781	91
2	BGR	Europe	Bulgaria	5/17/2020	6781955	2211	36	35.143001556396484	108
3	BGR	Europe	Bulgaria	5/24/2020	6781955	2427	19	30.85700035095215	130
4	BGR	Europe	Bulgaria	5/31/2020	6781955	2513	14	12.28600025177002	140
5	BGR	Europe	Bulgaria	6/7/2020	6781955	2727	16	30.570999145507812	160
6	BGR	Europe	Bulgaria	6/14/2020	6781955	3290	24	80.42900085449219	174
7	BGR	Europe	Bulgaria	6/21/2020	6781955	3905	33	87.85700225830078	199
8	BGR	Europe	Bulgaria	6/28/2020	6781955	4691	66	112.28600311279297	219
9	BGR	Europe	Bulgaria	7/5/2020	6781955	5740	63	149.85699462890625	246
10	BGR	Europe	Bulgaria	7/12/2020	6781955	7252	77	216	268
11	BGR	Europe	Bulgaria	7/19/2020	6781955	8733	95	211.5709991455078	300
12	BGR	Europe	Bulgaria	7/26/2020	6781955	10427	115	242	340
13	BGR	Europe	Bulgaria	8/2/2020	6781955	11955	119	218.28599548339844	388
14	BGR	Europe	Bulgaria	8/9/2020	6781955	13396	53	205.85699462890625	447
15	BGR	Europe	Bulgaria	8/16/2020	6781955	14365	32	138.4290008544922	498
16	BGR	Europe	Bulgaria	8/23/2020	6781955	15287	156	131.71400451660156	545
17	BGR	Europe	Bulgaria	8/30/2020	6781955	16190	26	129	613
18	BGR	Europe	Bulgaria	9/6/2020	6781955	17089	39	128.4290008544922	676
19	BGR	Europe	Bulgaria	9/13/2020	6781955	17918	27	118.42900085449219	720

[4]



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾

```

SUM(CASE WHEN iso_code IS NULL THEN 1 ELSE 0 END) AS missing_iso_code,
SUM(CASE WHEN continent IS NULL THEN 1 ELSE 0 END) AS missing_continent,
SUM(CASE WHEN location IS NULL THEN 1 ELSE 0 END) AS missing_location,
SUM(CASE WHEN date IS NULL THEN 1 ELSE 0 END) AS missing_date,
SUM(CASE WHEN population IS NULL THEN 1 ELSE 0 END) AS missing_population,
SUM(CASE WHEN total_cases IS NULL THEN 1 ELSE 0 END) AS missing_total_cases,
SUM(CASE WHEN new_cases IS NULL THEN 1 ELSE 0 END) AS missing_new_cases,
SUM(CASE WHEN new_cases_smoothed IS NULL THEN 1 ELSE 0 END) AS missing_new_cases_smoothed,
SUM(CASE WHEN total_deaths IS NULL THEN 1 ELSE 0 END) AS missing_total_deaths,
SUM(CASE WHEN new_deaths IS NULL THEN 1 ELSE 0 END) AS missing_new_deaths,
SUM(CASE WHEN new_deaths_smoothed IS NULL THEN 1 ELSE 0 END) AS missing_new_deaths_smoothed,
SUM(CASE WHEN total_cases_per_million IS NULL THEN 1 ELSE 0 END) AS missing_total_cases_per_million,
SUM(CASE WHEN new_cases_per_million IS NULL THEN 1 ELSE 0 END) AS missing_new_cases_per_million,
SUM(CASE WHEN new_cases_smoothed_per_million IS NULL THEN 1 ELSE 0 END) AS missing_new_cases_smoothed_per_million,
SUM(CASE WHEN total_deaths_per_million IS NULL THEN 1 ELSE 0 END) AS missing_total_deaths_per_million,
SUM(CASE WHEN new_deaths_per_million IS NULL THEN 1 ELSE 0 END) AS missing_new_deaths_per_million,
SUM(CASE WHEN new_deaths_smoothed_per_million IS NULL THEN 1 ELSE 0 END) AS missing_new_deaths_smoothed_per_million,
SUM(CASE WHEN reproduction_rate IS NULL THEN 1 ELSE 0 END) AS missing_reproduction_rate,
SUM(CASE WHEN icu_patients IS NULL THEN 1 ELSE 0 END) AS missing_icu_patients,
SUM(CASE WHEN icu_patients_per_million IS NULL THEN 1 ELSE 0 END) AS missing_icu_patients_per_million,
SUM(CASE WHEN hosp_patients IS NULL THEN 1 ELSE 0 END) AS missing_hosp_patients,
SUM(CASE WHEN hosp_patients_per_million IS NULL THEN 1 ELSE 0 END) AS missing_hosp_patients_per_million,
SUM(CASE WHEN weekly_icu_admissions IS NULL THEN 1 ELSE 0 END) AS missing_weekly_icu_admissions,
SUM(CASE WHEN weekly_icu_admissions_per_million IS NULL THEN 1 ELSE 0 END) AS missing_weekly_icu_admissions_per_million,
SUM(CASE WHEN weekly_hosp_admissions IS NULL THEN 1 ELSE 0 END) AS missing_weekly_hosp_admissions,
SUM(CASE WHEN weekly_hosp_admissions_per_million IS NULL THEN 1 ELSE 0 END) AS missing_weekly_hosp_admissions_per_million
FROM coviddeaths;

```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (1 rows)

Views >

	missing_iso_code	(long)	missing_continent	(long)	missing_location	(long)	missing_date	(long)	missing_population	(long)	missing_total_cases	(long)	missing_new_cases
0	0		13141		0		0		1003		13106		13404

[5]



```

SELECT
COALESCE(iso_code, 'N/A') AS iso_code,
COALESCE(continent, 'N/A') AS continent,
COALESCE(location, 'N/A') AS location,
COALESCE(date, 'N/A') AS date,
COALESCE(population, 'N/A') AS population,
COALESCE(total_cases, 'N/A') AS total_cases,
COALESCE(new_cases, 'N/A') AS new_cases,
COALESCE(new_cases_smoothed, 'N/A') AS new_cases_smoothed,
COALESCE(total_deaths, 'N/A') AS total_deaths,
COALESCE(new_deaths, 'N/A') AS new_deaths,
COALESCE(new_deaths_smoothed, 'N/A') AS new_deaths_smoothed,
COALESCE(total_cases_per_million, 'N/A') AS total_cases_per_million,
COALESCE(new_cases_per_million, 'N/A') AS new_cases_per_million,
COALESCE(new_cases_smoothed_per_million, 'N/A') AS new_cases_smoothed_per_million,
COALESCE(total_deaths_per_million, 'N/A') AS total_deaths_per_million,
COALESCE(new_deaths_per_million, 'N/A') AS new_deaths_per_million,
COALESCE(new_deaths_smoothed_per_million, 'N/A') AS new_deaths_smoothed_per_million,
COALESCE(reproduction_rate, 'N/A') AS reproduction_rate,
COALESCE(icu_patients, 'N/A') AS icu_patients,
COALESCE(icu_patients_per_million, 'N/A') AS icu_patients_per_million,
COALESCE(hosp_patients, 'N/A') AS hosp_patients,
COALESCE(hosp_patients_per_million, 'N/A') AS hosp_patients_per_million,
COALESCE(weekly_icu_admissions, 'N/A') AS weekly_icu_admissions,
COALESCE(weekly_icu_admissions_per_million, 'N/A') AS weekly_icu_admissions_per_million,
COALESCE(weekly_hosp_admissions, 'N/A') AS weekly_hosp_admissions,
COALESCE(weekly_hosp_admissions_per_million, 'N/A') AS weekly_hosp_admissions_per_million
FROM coviddeaths;

```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (232799 rows)

Views >

	TEST_1	Projects ▾	Branches ▾	Notebook	Dataset ▾	Caveats ▾	Settings ▾	
0	AFG	Asia	Afghanistan	2/24/2020	41128772	5	5	N/A
1	AFG	Asia	Afghanistan	2/25/2020	41128772	5	0	N/A
2	AFG	Asia	Afghanistan	2/26/2020	41128772	5	0	N/A
3	AFG	Asia	Afghanistan	2/27/2020	41128772	5	0	N/A
4	AFG	Asia	Afghanistan	2/28/2020	41128772	5	0	N/A
5	AFG	Asia	Afghanistan	2/29/2020	41128772	5	0	0.714
6	AFG	Asia	Afghanistan	3/1/2020	41128772	5	0	0.714
7	AFG	Asia	Afghanistan	3/2/2020	41128772	5	0	0.0
8	AFG	Asia	Afghanistan	3/3/2020	41128772	5	0	0.0
9	AFG	Asia	Afghanistan	3/4/2020	41128772	5	0	0.0
10	AFG	Asia	Afghanistan	3/5/2020	41128772	5	0	0.0
11	AFG	Asia	Afghanistan	3/6/2020	41128772	5	0	0.0
12	AFG	Asia	Afghanistan	3/7/2020	41128772	8	3	0.429
13	AFG	Asia	Afghanistan	3/8/2020	41128772	8	0	0.429
14	AFG	Asia	Afghanistan	3/9/2020	41128772	8	0	0.429
15	AFG	Asia	Afghanistan	3/10/2020	41128772	8	0	0.429
16	AFG	Asia	Afghanistan	3/11/2020	41128772	11	3	0.857
17	AFG	Asia	Afghanistan	3/12/2020	41128772	11	0	0.857
18	AFG	Asia	Afghanistan	3/13/2020	41128772	11	0	0.857
19	AFG	Asia	Afghanistan	3/14/2020	41128772	14	3	0.857

[6]

```
SELECT date, location, continent
FROM coviddeaths
WHERE date >= '2020-01-01'
ORDER BY date ASC;
```



Console

Timing

Datasets ▾

Charts ▾



temporary_dataset (150275 rows)

Views

	date (string)	location (string)	continent (string)
0	3/1/2020	Afghanistan	Asia
1	3/1/2020	Africa	
2	3/1/2020	Albania	Europe
3	3/1/2020	Algeria	Africa
4	3/1/2020	Argentina	South America
5	3/1/2020	Armenia	Asia
6	3/1/2020	Asia	
7	3/1/2020	Australia	Oceania
8	3/1/2020	Austria	Europe
9	3/1/2020	Azerbaijan	Asia
10	3/1/2020	Bahrain	Asia
11	3/1/2020	Belarus	Europe
12	3/1/2020	Belgium	Europe
13	3/1/2020	Brazil	South America
14	3/1/2020	Cambodia	Asia
15	3/1/2020	Canada	North America
16	3/1/2020	Chile	South America
17	3/1/2020	China	Asia
18	3/1/2020	Croatia	Europe
19	3/1/2020	Czechia	Europe

[7]

```
SELECT *
FROM CovidDeaths
```



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (219658 rows)

Views

	iso_code	continent	location	date	population	(long)	total_cases	(int)	new_cases	(int)	new_cases_smoothed	(float)	total_deaths
0	AFG	Asia	Afghanistan	1/1/2021	41128772	52513	183	131.14300537109375	2201				
1	AFG	Asia	Afghanistan	1/1/2022	41128772	158107	23	30.285999298095703	7356				
2	AFG	Asia	Afghanistan	1/10/2021	41128772	53489	89	111.42900085449219	2277				
3	AFG	Asia	Afghanistan	1/10/2022	41128772	158394	13	30.14299964904785	7373				
4	AFG	Asia	Afghanistan	1/11/2021	41128772	53538	49	89.85700225830078	2288				
5	AFG	Asia	Afghanistan	1/11/2022	41128772	158471	77	38	7374				
6	AFG	Asia	Afghanistan	1/12/2021	41128772	53584	46	81.85700225830078	2301				
7	AFG	Asia	Afghanistan	1/12/2022	41128772	158511	40	38	7376				
8	AFG	Asia	Afghanistan	1/13/2021	41128772	53690	106	83.57099914550781	2308				
9	AFG	Asia	Afghanistan	1/13/2022	41128772	158602	91	46.7140007019043	7376				
10	AFG	Asia	Afghanistan	1/14/2021	41128772	53775	85	81.14299774169922	2314				
11	AFG	Asia	Afghanistan	1/14/2022	41128772	158639	37	48.42900085449219	7376				
12	AFG	Asia	Afghanistan	1/15/2021	41128772	53831	56	71.28600311279297	2324				
13	AFG	Asia	Afghanistan	1/15/2022	41128772	158678	39	52.7140007019043	7378				
14	AFG	Asia	Afghanistan	1/16/2021	41128772	53938	107	76.85700225830078	2336				
15	AFG	Asia	Afghanistan	1/16/2022	41128772	158717	39	48	7379				
16	AFG	Asia	Afghanistan	1/17/2021	41128772	53984	46	70.71399688720703	2339				
17	AFG	Asia	Afghanistan	1/17/2022	41128772	158826	109	61.7140007019043	7381				
18	AFG	Asia	Afghanistan	1/18/2021	41128772	54062	78	74.85700225830078	2343				
19	AFG	Asia	Afghanistan	1/18/2022	41128772	158974	148	71.85700225830078	7383				

[8]

```
--LOOKING AT THE TOTAL CASES VS TOTAL DEATHS
-- Shows Likelihood of dying if you contact covid in your country
```

```
SELECT Location, Date, total_cases, total_deaths, (total_deaths/total_cases) * 100 AS DeathPercentage
FROM CovidDeaths
WHERE Location LIKE '%states'
AND continent IS NOT NULL
ORDER BY 1, 2;
```

Console ▾ Timing Datasets ▾ Charts ▾



temporary_dataset (0 rows)

Views

Location	Date	total_cases	total_deaths	DeathPercentage
----------	------	-------------	--------------	-----------------

[9]

```
SELECT Location, Date, total_cases, total_deaths, (total_deaths/total_cases) * 100 AS DeathPercentage
FROM CovidDeaths
WHERE Location LIKE '%Afghanistan%'
AND continent IS NOT NULL
ORDER BY 1, 2;
```

Console ▾ Timing Datasets ▾ Charts ▾



temporary_dataset (986 rows)

Views

Location	Date	total_cases	total_deaths	DeathPercentage
Afghanistan	1/1/2021	52513	2201	4.191343095995277

	TEST_1	Projects ▾	Branches ▾	Notebook	Dataset ▾	Caveats ▾	Settings ▾
	2	Afghanistan		1/10/2021	53489	2277	4.2569500271083776
	3	Afghanistan		1/10/2022	158394	7373	4.654848037173125
	4	Afghanistan		1/11/2021	53538	2288	4.27360005977063
	5	Afghanistan		1/11/2022	158471	7374	4.653217307898607
	6	Afghanistan		1/12/2021	53584	2301	4.294192296207823
	7	Afghanistan		1/12/2022	158511	7376	4.6533048179621606
	8	Afghanistan		1/13/2021	53690	2308	4.298752095362265
	9	Afghanistan		1/13/2022	158602	7376	4.650634922636536
	10	Afghanistan		1/14/2021	53775	2314	4.303114830311483
	11	Afghanistan		1/14/2022	158639	7376	4.649550236700938
	12	Afghanistan		1/15/2021	53831	2324	4.317214987646524
	13	Afghanistan		1/15/2022	158678	7378	4.649667880865652
	14	Afghanistan		1/16/2021	53938	2336	4.330898438948423
	15	Afghanistan		1/16/2022	158717	7379	4.649155414983902
	16	Afghanistan		1/17/2021	53984	2339	4.332765263781861
	17	Afghanistan		1/17/2022	158826	7381	4.647224006145089
	18	Afghanistan		1/18/2021	54062	2343	4.333912914801524
	19	Afghanistan		1/18/2022	158974	7383	4.644155648093399

[10]

```
-- LOOKING AT THE TOTAL CASES VS POPULATION
-- Shows Percentage of Population infected with covid
```

```
SELECT Location, Date, population, total_cases, (total_cases/population) * 100 AS PercentPopulationInfected
FROM CovidDeaths
WHERE continent IS NOT NULL
AND Location LIKE '%Afghanistan%'
ORDER BY 1, 2;
```

Console ▾ Timing Datasets ▾ Charts ▾



temporary_dataset (986 rows)

Views

Location (string)	Date (string)	population (long)	total_cases (int)	PercentPopulationInfected (real)
0 Afghanistan	1/1/2021	41128772	52513	0.1276794746023538
1 Afghanistan	1/1/2022	41128772	158107	0.3844194521538353
2 Afghanistan	1/10/2021	41128772	53489	0.1300525092263878
3 Afghanistan	1/10/2022	41128772	158394	0.38511726049102557
4 Afghanistan	1/11/2021	41128772	53538	0.13017164723517638
5 Afghanistan	1/11/2022	41128772	158471	0.3853044773619791
6 Afghanistan	1/12/2021	41128772	53584	0.1302834910801616
7 Afghanistan	1/12/2022	41128772	158511	0.3854017328793576
8 Afghanistan	1/13/2021	41128772	53690	0.13054121820121448
9 Afghanistan	1/13/2022	41128772	158602	0.3856229891813935
10 Afghanistan	1/14/2021	41128772	53775	0.13074788617564365
11 Afghanistan	1/14/2022	41128772	158639	0.38571295053496857
12 Afghanistan	1/15/2021	41128772	53831	0.13088404389997346
13 Afghanistan	1/15/2022	41128772	158678	0.3858077746644126
14 Afghanistan	1/16/2021	41128772	53938	0.1311442024089608
15 Afghanistan	1/16/2022	41128772	158717	0.3859025987938565
16 Afghanistan	1/17/2021	41128772	53984	0.13125604625394602
17 Afghanistan	1/17/2022	41128772	158826	0.3861676200787128
18 Afghanistan	1/18/2021	41128772	54062	0.13144569451283397
19 Afghanistan	1/18/2022	41128772	158974	0.38652746549301303

[11]



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾

```
SELECT Location, population, MAX(total_cases) AS HighestInfectionCount, MAX((total_cases/population)) * 100 AS PercentPopulationInfected
FROM CovidDeaths
WHERE continent IS NOT NULL
GROUP BY Location, population
ORDER BY PercentPopulationInfected DESC;
```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (235 rows)

Views



	Location (string)	population (long)	HighestInfectionCount (int)	PercentPopulationInfected (real)
0	Cyprus	896007	602662	67.26085845311476
1	Faeroe Islands	53117	34658	65.24841387879586
2	San Marino	33690	21662	64.29801127931137
3	Gibraltar	32677	20121	61.57542002019769
4	Austria	8939617	5462911	61.10900500547171
5	Slovenia	2119843	1237624	58.38281419897605
6	Andorra	79843	46588	58.349510915171024
7	Denmark	5882259	3337464	56.73779410257181
8	Iceland	372903	206571	55.39537091415194
9	Saint Pierre and Miquelon	5885	3231	54.902293967714535
10	France	67813000	36982388	54.53583826110038
11	Portugal	10270857	5520731	53.75141529085645
12	Brunei	449002	241044	53.684393388002725
13	Liechtenstein	39355	20663	52.50412908143819
14	Latvia	1850654	954230	51.56177221674068
15	Falkland Islands	3801	1930	50.776111549592216
16	Greece	10384972	5188890	49.96537304096728
17	South Korea	51815808	25838239	49.86555261282425
18	Israel	9449000	4688662	49.62072176949942
19	Switzerland	8740471	4255891	48.69178102644583

[12]

-- SHOWING COUNTRIES WITH HIGHEST DEATH COUNT PER POPULATION

```
SELECT Location, MAX(cast(total_deaths AS int)) AS TotalDeathCount
FROM CovidDeaths
WHERE continent IS NOT NULL
GROUP BY Location
ORDER BY TotalDeathCount DESC;
```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (235 rows)

Views



	Location (string)	TotalDeathCount (int)
0	United States	1072582
1	Brazil	688332
2	India	530500
3	Russia	382644
4	Mexico	330424
5	Peru	217062
6	United Kingdom	209947
7	Italy	179436
8	Indonesia	158807
9	France	157346



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾

11	Iran	144592
12	Colombia	141850
13	Argentina	129991
14	Poland	118170
15	Ukraine	118068
16	Spain	115239
17	South Africa	102363
18	Turkey	101203
19	Romania	67211

[13]



Console ▾

Timing

Datasets ▾

Charts ▾



NOW BREAK THIS DOWN BY CONTINENT

This method kind of gives you the accurate result if you want to break it down to continent We use where continent is null because if you check in the excel sheet, the locations has the names of the continent in them, instead of the names of the countries.

[14]

```
SELECT Location, MAX(cast(total_deaths AS int)) AS TotalDeathCount
FROM CovidDeaths
WHERE continent IS NULL
GROUP BY Location
ORDER BY TotalDeathCount DESC;
```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (13 rows)

Views

	Location (string)	TotalDeathCount (int)
0	World	6599909
1	High income	2713545
2	Upper middle income	2498922
3	Europe	1965834
4	North America	1528219
5	Asia	1495280
6	Lower middle income	1339681
7	South America	1332816
8	European Union	1164346
9	Africa	257035
10	Low income	47679
11	Oceania	20710
12	International	15

[15]

```
-- Now we are going to use this for the visualization purposes
```

```
SELECT continent, MAX(cast(total_deaths AS int)) AS TotalDeathCount
FROM CovidDeaths
WHERE continent IS NOT NULL
GROUP BY continent
ORDER BY TotalDeathCount DESC;
```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (6 rows)

Views

The screenshot shows a table titled 'TEST_1' with the following data:

1	South America	688332
2	Asia	530500
3	Europe	382644
4	Africa	102363
5	Oceania	15786

[16]

```
-- GLOBAL_NUMS

SELECT Date, SUM(new_cases) AS total_cases, SUM(cast(new_deaths AS int)) AS total_deaths, SUM(cast(new_deaths AS int))/SUM(new_ca
FROM CovidDeaths
WHERE continent IS NOT NULL
AND Date IS NOT NULL
AND new_cases IS NOT NULL
AND new_deaths IS NOT NULL
GROUP BY Date
ORDER BY 1, 2;
```

Console Timing Datasets Charts

temporary_dataset (1018 rows)

Views

	Date (string)	total_cases (long)	total_deaths (long)	DeathPercentage (real)
0	1/1/2021	586578	10156	1.7313980408402634
1	1/1/2022	1228723	4455	0.3625715478590374
2	1/10/2021	601477	9554	1.5884231649755518
3	1/10/2022	3159394	6572	0.20801457494696768
4	1/11/2021	614596	10284	1.67329432667964
5	1/11/2022	3028727	9104	0.3005883329861027
6	1/12/2021	692324	16984	2.4531866582698276
7	1/12/2022	3446651	9023	0.26179035823470376
8	1/13/2021	743282	16480	2.217193474347556
9	1/13/2022	3185510	8229	0.2583259823387778
10	1/14/2021	759002	15714	2.0703502757568493
11	1/14/2022	3316566	7966	0.24018819465676244
12	1/15/2021	783872	15717	2.0050467423252774
13	1/15/2022	2505187	6363	0.2539930152918724
14	1/16/2021	656781	14029	2.1360240323639084
15	1/16/2022	2187055	4555	0.20827093968830232
16	1/17/2021	527570	9834	1.8640180449987678
17	1/17/2022	2644788	6224	0.23533077131323948
18	1/18/2021	517226	10442	2.0188466937083596
19	1/18/2022	3763316	9104	0.2419143117399655

[17]

```
SELECT SUM(new_cases) AS total_cases, SUM(cast(new_deaths AS int)) AS total_deaths, SUM(cast(new_deaths AS int))/SUM(new_cases) *
FROM CovidDeaths
WHERE continent IS NOT NULL
ORDER BY 1, 2;
```

Console Timing Datasets Charts

temporary_dataset (1 rows)

Views

	total_cases (long)	total_deaths (long)	DeathPercentage (real)
0	630711800	6561569	1.0403434659062982

vaccinations (232799 rows)

	iso_code	continent	location	date	total_tests	new_tests	total_tests_per_thousand	new_tests_per_thousand
0	AFG	Asia	Afghanistan	2/24/2020				
1	AFG	Asia	Afghanistan	2/25/2020				
2	AFG	Asia	Afghanistan	2/26/2020				
3	AFG	Asia	Afghanistan	2/27/2020				
4	AFG	Asia	Afghanistan	2/28/2020				
5	AFG	Asia	Afghanistan	2/29/2020				
6	AFG	Asia	Afghanistan	3/1/2020				
7	AFG	Asia	Afghanistan	3/2/2020				
8	AFG	Asia	Afghanistan	3/3/2020				
9	AFG	Asia	Afghanistan	3/4/2020				
10	AFG	Asia	Afghanistan	3/5/2020				
11	AFG	Asia	Afghanistan	3/6/2020				
12	AFG	Asia	Afghanistan	3/7/2020				
13	AFG	Asia	Afghanistan	3/8/2020				
14	AFG	Asia	Afghanistan	3/9/2020				
15	AFG	Asia	Afghanistan	3/10/2020				
16	AFG	Asia	Afghanistan	3/11/2020				
17	AFG	Asia	Afghanistan	3/12/2020				
18	AFG	Asia	Afghanistan	3/13/2020				
19	AFG	Asia	Afghanistan	3/14/2020				

[19]

FROM HERE ASSESSING VACCINATION DATA

DATA INTEGRATION

```
[20] LOAD DATASET vacc AS csv FROM CovidVaccinations.csv @ artifact file 218
```

vacc (232799 rows)

	iso_code	continent	location	date	total_tests	new_tests	total_tests_per_thousand	new_tests_per_thousand
0	AFG	Asia	Afghanistan	2/24/2020				
1	AFG	Asia	Afghanistan	2/25/2020				
2	AFG	Asia	Afghanistan	2/26/2020				
3	AFG	Asia	Afghanistan	2/27/2020				
4	AFG	Asia	Afghanistan	2/28/2020				
5	AFG	Asia	Afghanistan	2/29/2020				
6	AFG	Asia	Afghanistan	3/1/2020				
7	AFG	Asia	Afghanistan	3/2/2020				
8	AFG	Asia	Afghanistan	3/3/2020				
9	AFG	Asia	Afghanistan	3/4/2020				
10	AFG	Asia	Afghanistan	3/5/2020				
11	AFG	Asia	Afghanistan	3/6/2020				
12	AFG	Asia	Afghanistan	3/7/2020				
13	AFG	Asia	Afghanistan	3/8/2020				
14	AFG	Asia	Afghanistan	3/9/2020				



TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾

16	AFG	Asia	Afghanistan	3/11/2020
17	AFG	Asia	Afghanistan	3/12/2020
18	AFG	Asia	Afghanistan	3/13/2020
19	AFG	Asia	Afghanistan	3/14/2020

[21]

-- JOINING THE TWO TABLES TOGETHER FOR VIEWING

```
SELECT *
FROM CovidDeaths death
JOIN vacc vac
    ON death.location = vac.location
    AND death.date = vac.date;
```



Console ▾

Timing

Datasets ▾

Charts ▾



temporary_dataset (232799 rows)

Views

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths
0	AFG	Asia	Afghanistan	10/3/2020	41128772	39383	7	15	1466
1	AFG	Asia	Afghanistan	11/21/2020	41128772	44443	215	210.5709991455078	1661
2	AFG	Asia	Afghanistan	11/22/2021	41128772	156911	15	43	7303
3	AFG	Asia	Afghanistan	11/27/2021	41128772	157190	19	46.57099914550781	7308
4	AFG	Asia	Afghanistan	3/28/2020	41128772	106	15	11.71399974822998	2
5	AFG	Asia	Afghanistan	3/8/2020	41128772	8	0	0.4289999022483826	
6	AFG	Asia	Afghanistan	7/17/2022	41128772	183572	127	43.856998443603516	7731
7	AFG	Asia	Afghanistan	8/10/2022	41128772	188202	236	258.4289855957031	7753
8	AFG	Asia	Afghanistan	8/11/2021	41128772	151291	278	336.85699462890625	6978
9	AFG	Asia	Afghanistan	9/10/2022	41128772	195631	160	245.5709991455078	7787
10	OWID_AFR	Africa		10/1/2022	1426736614	12357818	588	1119.2860107421875	256764
11	OWID_AFR	Africa		10/10/2021	1426736614	8389905	4702	7139.85693359375	213936
12	OWID_AFR	Africa		12/13/2021	1426736614	8999010	17081	29840	224964
13	OWID_AFR	Africa		2/11/2022	1426736614	11115635	18311	18924.4296875	243533
14	OWID_AFR	Africa		3/17/2021	1426736614	4064067	10987	10360.857421875	108671
15	OWID_AFR	Africa		4/24/2021	1426736614	4497010	10152	10682.2861328125	119840
16	ALB	Europe	Albania	10/27/2020	2842318	19729	284	296.85699462890625	487
17	ALB	Europe	Albania	11/13/2020	2842318	26701	490	498.7139892578125	605
18	ALB	Europe	Albania	4/10/2021	2842318	128155	360	281.7139892578125	2310
19	ALB	Europe	Albania	4/16/2022	2842318	274462	33	38.7140007019043	3496

[22]

```
SELECT
    d.continent,
    d.location,
    d.date,
    d.population,
    SUM(v.new_vaccinations) OVER (PARTITION BY d.location ORDER BY d.date) AS RollingPeopleVaccinated
FROM
    CovidDeaths d
JOIN
    vacc v ON d.location = v.location AND d.date = v.date
WHERE
    d.continent IS NOT NULL
ORDER BY
    2, 3;
```



Console ▾

Timing

Datasets ▾

Charts ▾





TEST_1

Projects ▾

Branches ▾

Notebook

Dataset ▾

Caveats ▾

Settings ▾