

# ESAME MODULO 3 - PYTHON

## LIBRERIE E IMPORTAZIONI

```
In [4]: pip install pillow
```

Requirement already satisfied: pillow in c:\users\bocci\anaconda3\lib\site-packages (10.4.0)  
Note: you may need to restart the kernel to use updated packages.

```
In [5]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
import seaborn as sns
from PIL import Image
```

```
In [6]: filepath = "C:\\Users\\bocci\\Desktop\\ESAME FINALE\\Esame.png"
Esame = Image.open(filepath)
```

```
In [7]: filepath = "C:\\Users\\bocci\\Desktop\\ESAME FINALE\\Note_traccia.png"
Note_traccia = Image.open(filepath)
```

```
In [8]: filepath = "C:\\Users\\bocci\\Desktop\\ESAME FINALE\\Traccia.png"
Traccia = Image.open(filepath)
```

```
In [9]: filepath = "C:\\Users\\bocci\\Desktop\\ESAME FINALE\\Consegna.png"
Consegna = Image.open(filepath)
```

```
In [10]: filepath = "C:\\Users\\bocci\\Desktop\\ESAME FINALE\\Ringraziamenti.png"
Ringraziamenti = Image.open(filepath)
```

## ESAME

In [12]: Esame

Out[12]:



# Esercizio: progetto di analisi dati

In [ ]:

## LEGENDA VARIABILI E DATAFRAME

PER LA PULIZIA E LA VISUALIZZAZIONE DEL DATASET MI SONO SERVITO DI UNA SERIE DI VARIABILI, AL FINE DI AVERE TUTTO PIU CHIARO. UTILI PER L'ANALISI DELLE NOTE DELLA TRACCIA E PER ESPORTARE DATASET PARZIALI O TEMPORANEI IN FORMATO CSV

- `df_covid` # Identifica l'intero dataset
- `df_continent_null` # Identifica i continenti dove all'interno della colonna `continent` il valore è null
- `df_location_puntuale` # Nuovo dataframe con solo le righe che hanno un 'continent' definito (cioè dati puntuali) nella colonna `continent`
- `aggregati_da_escludere` # Lista di valori da escludere dalla colonna `location`
- `df_paesi` # Nuovo DataFrame filtrando solo i dati a livello di paese, escludendo i valori aggregati dalla colonna `location`
- `df_aggregati` # Creo un dataframe contenente esattamente le righe escluse da `df_paesi`
- `df_covid_rimozione_null` # Nuovo dataframe aggiornato con la rimozione dei valori null di `continent`
- `percentuale_df_paesi_location` # % dataframe `df_paesi` rispetto al dataframe originario `df_covid`
- `percentuale_nulli_continent` # % null di `continent`
- `valori_nulli_continent` # totale valori null di `continent`
- `totale_righe` # numero righe `df_covid`
- `numero_righe_location` # numero righe `df_paesi`
- `percentuale_df_aggregati_location` # % righe rimosse da `location` rispetto al df originario `df_covid`
- `numero_aggregati_location` # Numero righe del `df_aggregati`
- `df_paesi_new` # Nuovo dataframe di `df_paesi` al quale sono stati esclusi oltre che gli aggregati da escludere, anche Oceania e European Union
- `df_temp` # dataframe temporaneo nel quale sono presenti le esclusioni di Oceania e European Union da `location`
- `df_covid_new` # Questo è il nuovo e definitivo dataframe `df_covid`, con esclusione dei valori null da `continent`
- `continents` # Variabile con i nomi dei continenti
- `continent_dfs` # Dizionario con lista continenti = `continent_dfs['Asia']` - `continent_dfs['Europe']` - `continent_dfs['Africa']`
- `continent_dfs['Oceania']` - `continent_dfs['df_North_America']` - `continent_dfs['df_South_America']`
- `df_asia` # Dataframe singolo per ogni continente
- `df_europe` # Dataframe singolo per ogni continente
- `df_africa` # Dataframe singolo per ogni continente
- `df_oceania` # Dataframe singolo per ogni continente
- `df_north_america` # Dataframe singolo per ogni continente
- `df_south_america` # Dataframe singolo per ogni continente
- `df_col_utili` # Qui ho creato un dataset di `df_covid_new` con le sole colonne che ritengo utili allo svolgimento delle tracce

- `df_nonnull_newcases` # Qui ho rimosso tutti i null dal `df_covid_new` z
- `df_nonzerovalues_newcases` # Qui ho rimosso tutti gli zeri dal nuovo `df_nonnull_newcases`, sarà il mio dataframe di riferimento per lo svolgimento delle tracce

#### QUESTE LE VARIABILI USATE PER LO SVOLGIMENTO DELLE TRACCE

- `df_nonnull_newcases` # Qui rimuovo i null da `new_cases`
- `df_nonzerovalues_newcases` # Qui ho rimosso tutti gli zero da `df_nonnull_newcases` (Divenuto il dataframe di riferimento per lo svolgimento delle tracce. Definitivo)
- `continent_cases` # Casi per continente
- `continent_percent` # Percentuale per continente rispetto al totale mondiale
- `df_italia_2022` # dataframe relativo ai dati italiani per l'anno 2022, utile per lo svolgimento dell'esercizio 3
- `total_world_cases` # somma totale di tutti i `new_cases`
- `icu_data` # Dataframe riguardante stati di Italia, Francia e Germania utile allo svolgimento dell'esercizio 4 relativo al numero di pazienti in terapia intensiva (Intensive Care Unit, ICU, considerando quindi la colonna `icu_patients`) da maggio 2022 (incluso) ad aprile 2023 (incluso)
- `dark_purple_palette` # Variabile per definire tonalità più scure di viola
- `hosp_data_2021` # Dataframe relativo agli ospedalizzati di Italia, Francia, Germania e Spagna, utile per lo svolgimento dell'esercizio 5
- `hosp_sum` # Somma dei totali ospedalizzati per nazione 2021 (corretta secondo il mio ragionamento)
- `hosp_sum2` # Somma dei totali ospedalizzati per nazione 2021 (non corretta secondo il mio ragionamento)
- `df_covidtemp_nonzero` # Qui ho ordinato per location e date in ordine cronologico
- `covidtempfirstcasesforlocation` # # Qui trovo il primo valore per ogni location (ordinato in ordine cronologico grazie alla variabile `df_covidtemp_nonzero`)

## NOTE

In [16]: `Note_traccia`

Out[16]:



### Note

Prima di effettuare le analisi, effettuate un'EDA del dataset per meglio comprenderne la natura (e non dimenticate ulteriori metadati e descrizioni presenti sulla pagina Github da cui avete scaricato il dataset); in particolare:

- Fate attenzione alle colonne `continent` e `location`; contengono sia dati puntuali che dati aggregati
  - Prima di iniziare l'analisi, assicuratevi di aver compreso cosa contengono
- Comprendete bene il contenuto delle colonne `new_cases` e `total_cases`
  - Hanno logiche differenti e ovviamente vanno trattate diversamente

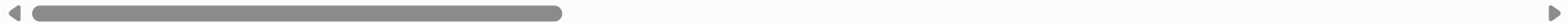
```
In [17]: df_covid = pd.read_csv("owid-covid-data.csv")
```

```
In [18]: df_covid # Verifico le dimensioni del dataset
```

Out[18]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

429435 rows × 67 columns

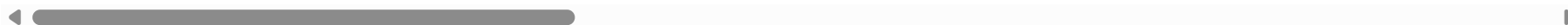


```
In [19]: # Visualizzo le prime righe del dataset
df_covid.head()
```

Out[19]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	NaN
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	NaN
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	NaN
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	NaN
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	NaN

5 rows × 67 columns



```
In [20]: # Controllo Le informazioni sulle colonne  
df_covid.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 429435 entries, 0 to 429434
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	iso_code	429435 non-null	object
1	continent	402910 non-null	object
2	location	429435 non-null	object
3	date	429435 non-null	object
4	total_cases	411804 non-null	float64
5	new_cases	410159 non-null	float64
6	new_cases_smoothed	408929 non-null	float64
7	total_deaths	411804 non-null	float64
8	new_deaths	410608 non-null	float64
9	new_deaths_smoothed	409378 non-null	float64
10	total_cases_per_million	411804 non-null	float64
11	new_cases_per_million	410159 non-null	float64
12	new_cases_smoothed_per_million	408929 non-null	float64
13	total_deaths_per_million	411804 non-null	float64
14	new_deaths_per_million	410608 non-null	float64
15	new_deaths_smoothed_per_million	409378 non-null	float64
16	reproduction_rate	184817 non-null	float64
17	icu_patients	39116 non-null	float64
18	icu_patients_per_million	39116 non-null	float64
19	hosp_patients	40656 non-null	float64
20	hosp_patients_per_million	40656 non-null	float64
21	weekly_icu_admissions	10993 non-null	float64
22	weekly_icu_admissions_per_million	10993 non-null	float64
23	weekly_hosp_admissions	24497 non-null	float64
24	weekly_hosp_admissions_per_million	24497 non-null	float64
25	total_tests	79387 non-null	float64
26	new_tests	75403 non-null	float64
27	total_tests_per_thousand	79387 non-null	float64
28	new_tests_per_thousand	75403 non-null	float64
29	new_tests_smoothed	103965 non-null	float64
30	new_tests_smoothed_per_thousand	103965 non-null	float64
31	positive_rate	95927 non-null	float64
32	tests_per_case	94348 non-null	float64
33	tests_units	106788 non-null	object
34	total_vaccinations	85417 non-null	float64
35	people_vaccinated	81132 non-null	float64



```

36 people_fully_vaccinated      78061 non-null float64
37 total_boosters               53600 non-null float64
38 new_vaccinations             70971 non-null float64
39 new_vaccinations_smoothed    195029 non-null float64
40 total_vaccinations_per_hundred 85417 non-null float64
41 people_vaccinated_per_hundred 81132 non-null float64
42 people_fully_vaccinated_per_hundred 78061 non-null float64
43 total_boosters_per_hundred    53600 non-null float64
44 new_vaccinations_smoothed_per_million 195029 non-null float64
45 new_people_vaccinated_smoothed 192177 non-null float64
46 new_people_vaccinated_smoothed_per_hundred 192177 non-null float64
47 stringency_index             196190 non-null float64
48 population_density           360492 non-null float64
49 median_age                   334663 non-null float64
50 aged_65_older                323270 non-null float64
51 aged_70_older                331315 non-null float64
52 gdp_per_capita                328292 non-null float64
53 extreme_poverty              211996 non-null float64
54 cardiovascular_death_rate     328865 non-null float64
55 diabetes_prevalence           345911 non-null float64
56 female_smokers                247165 non-null float64
57 male_smokers                  243817 non-null float64
58 handwashing_facilities        161741 non-null float64
59 hospital_beds_per_thousand    290689 non-null float64
60 life_expectancy               390299 non-null float64
61 human_development_index       319127 non-null float64
62 population                    429435 non-null int64
63 excess_mortality_cumulative_absolute 13411 non-null float64
64 excess_mortality_cumulative    13411 non-null float64
65 excess_mortality              13411 non-null float64
66 excess_mortality_cumulative_per_million 13411 non-null float64
dtypes: float64(61), int64(1), object(5)
memory usage: 219.5+ MB

```

```

In [21]: # Conto i valori nulli per colonna
df_covid.isnull().sum()

```

```
Out[21]: iso_code          0
continent        26525
location         0
date             0
total_cases      17631
...
population       0
excess_mortality_cumulative_absolute  416024
excess_mortality_cumulative          416024
excess_mortality                     416024
excess_mortality_cumulative_per_million  416024
Length: 67, dtype: int64
```

```
In [22]: # Ottengo statistiche descrittive sulle colonne numeriche
df_covid.describe()
```

```
Out[22]:
```

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	new
<b>count</b>	4.118040e+05	4.101590e+05	4.089290e+05	4.118040e+05	410608.000000	409378.000000	411804.000000	
<b>mean</b>	7.365292e+06	8.017360e+03	8.041026e+03	8.125957e+04	71.852139	72.060873	112096.199396	
<b>std</b>	4.477582e+07	2.296649e+05	8.661611e+04	4.411901e+05	1368.322990	513.636567	162240.412419	
<b>min</b>	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000	
<b>25%</b>	6.280750e+03	0.000000e+00	0.000000e+00	4.300000e+01	0.000000	0.000000	1916.100500	
<b>50%</b>	6.365300e+04	0.000000e+00	1.200000e+01	7.990000e+02	0.000000	0.000000	29145.475000	
<b>75%</b>	7.582720e+05	0.000000e+00	3.132860e+02	9.574000e+03	0.000000	3.143000	156770.190000	
<b>max</b>	7.758668e+08	4.423623e+07	6.319461e+06	7.057132e+06	103719.000000	14817.000000	763598.600000	

8 rows × 62 columns



```
In [23]: # Controllo i valori unici nella colonna 'continent'
df_covid['continent'].unique()
```

```
Out[23]: array(['Asia', nan, 'Europe', 'Africa', 'Oceania', 'North America',  
              'South America'], dtype=object)
```

```
In [24]: # Controllo i valori unici nella colonna 'location'  
df_covid['location'].unique()
```

```
Out[24]: array(['Afghanistan', 'Africa', 'Albania', 'Algeria', 'American Samoa',  
              'Andorra', 'Angola', 'Anguilla', 'Antigua and Barbuda',  
              'Argentina', 'Armenia', 'Aruba', 'Asia', 'Australia', 'Austria',  
              'Azerbaijan', 'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados',  
              'Belarus', 'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan',  
              'Bolivia', 'Bonaire Sint Eustatius and Saba',  
              'Bosnia and Herzegovina', 'Botswana', 'Brazil',  
              'British Virgin Islands', 'Brunei', 'Bulgaria', 'Burkina Faso',  
              'Burundi', 'Cambodia', 'Cameroon', 'Canada', 'Cape Verde',  
              'Cayman Islands', 'Central African Republic', 'Chad', 'Chile',  
              'China', 'Colombia', 'Comoros', 'Congo', 'Cook Islands',  
              'Costa Rica', 'Cote d'Ivoire', 'Croatia', 'Cuba', 'Curacao',  
              'Cyprus', 'Czechia', 'Democratic Republic of Congo', 'Denmark',  
              'Djibouti', 'Dominica', 'Dominican Republic', 'East Timor',  
              'Ecuador', 'Egypt', 'El Salvador', 'England', 'Equatorial Guinea',  
              'Eritrea', 'Estonia', 'Eswatini', 'Ethiopia', 'Europe',  
              'European Union (27)', 'Faroe Islands', 'Falkland Islands', 'Fiji',  
              'Finland', 'France', 'French Guiana', 'French Polynesia', 'Gabon',  
              'Gambia', 'Georgia', 'Germany', 'Ghana', 'Gibraltar', 'Greece',  
              'Greenland', 'Grenada', 'Guadeloupe', 'Guam', 'Guatemala',  
              'Guernsey', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti',  
              'High-income countries', 'Honduras', 'Hong Kong', 'Hungary',  
              'Iceland', 'India', 'Indonesia', 'Iran', 'Iraq', 'Ireland',  
              'Isle of Man', 'Israel', 'Italy', 'Jamaica', 'Japan', 'Jersey',  
              'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati', 'Kosovo', 'Kuwait',  
              'Kyrgyzstan', 'Laos', 'Latvia', 'Lebanon', 'Lesotho', 'Liberia',  
              'Libya', 'Liechtenstein', 'Lithuania', 'Low-income countries',  
              'Lower-middle-income countries', 'Luxembourg', 'Macao',  
              'Madagascar', 'Malawi', 'Malaysia', 'Maldives', 'Mali', 'Malta',  
              'Marshall Islands', 'Martinique', 'Mauritania', 'Mauritius',  
              'Mayotte', 'Mexico', 'Micronesia (country)', 'Moldova', 'Monaco',  
              'Mongolia', 'Montenegro', 'Montserrat', 'Morocco', 'Mozambique',  
              'Myanmar', 'Namibia', 'Nauru', 'Nepal', 'Netherlands',  
              'New Caledonia', 'New Zealand', 'Nicaragua', 'Niger', 'Nigeria',  
              'Niue', 'North America', 'North Korea', 'North Macedonia',  
              'Northern Cyprus', 'Northern Ireland', 'Northern Mariana Islands',  
              'Norway', 'Oceania', 'Oman', 'Pakistan', 'Palau', 'Palestine',  
              'Panama', 'Papua New Guinea', 'Paraguay', 'Peru', 'Philippines',  
              'Pitcairn', 'Poland', 'Portugal', 'Puerto Rico', 'Qatar',  
              'Reunion', 'Romania', 'Russia', 'Rwanda', 'Saint Barthelemy',
```

```
'Saint Helena', 'Saint Kitts and Nevis', 'Saint Lucia',  
'Saint Martin (French part)', 'Saint Pierre and Miquelon',  
'Saint Vincent and the Grenadines', 'Samoa', 'San Marino',  
'Sao Tome and Principe', 'Saudi Arabia', 'Scotland', 'Senegal',  
'Serbia', 'Seychelles', 'Sierra Leone', 'Singapore',  
'Sint Maarten (Dutch part)', 'Slovakia', 'Slovenia',  
'Solomon Islands', 'Somalia', 'South Africa', 'South America',  
'South Korea', 'South Sudan', 'Spain', 'Sri Lanka', 'Sudan',  
'Suriname', 'Sweden', 'Switzerland', 'Syria', 'Taiwan',  
'Tajikistan', 'Tanzania', 'Thailand', 'Togo', 'Tokelau', 'Tonga',  
'Trinidad and Tobago', 'Tunisia', 'Turkey', 'Turkmenistan',  
'Turks and Caicos Islands', 'Tuvalu', 'Uganda', 'Ukraine',  
'United Arab Emirates', 'United Kingdom', 'United States',  
'United States Virgin Islands', 'Upper-middle-income countries',  
'Uruguay', 'Uzbekistan', 'Vanuatu', 'Vatican', 'Venezuela',  
'Vietnam', 'Wales', 'Wallis and Futuna', 'Western Sahara', 'World',  
'Yemen', 'Zambia', 'Zimbabwe'], dtype=object)
```

In [ ]:

## VERIFICO LA PRIMA NOTA

In [26]: `df_covid.sample(50)`

Out[26]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_sm
108293	SLV	North America	El Salvador	2023-02-13	201785.0	0.0	0.000	4230.0	0.0	
66387	CPV	Africa	Cape Verde	2022-12-27	63190.0	0.0	3.714	412.0	0.0	
268715	NCL	Oceania	New Caledonia	2024-05-04	80163.0	0.0	14.143	314.0	0.0	
161036	GUY	South America	Guyana	2021-06-25	18837.0	0.0	107.000	442.0	0.0	
377021	TJK	Asia	Tajikistan	2023-10-05	17786.0	0.0	0.000	125.0	0.0	
348281	SXM	North America	Sint Maarten (Dutch part)	2022-02-04	9229.0	0.0	37.714	79.0	0.0	
384076	TKL	Oceania	Tokelau	2021-12-18	0.0	0.0	0.000	0.0	0.0	
367653	SUR	South America	Suriname	2020-02-19	0.0	0.0	0.000	0.0	0.0	
47559	BIH	Europe	Bosnia and Herzegovina	2021-11-08	259233.0	0.0	782.286	11767.0	0.0	
2857	OWID_AFR	NaN	Africa	2023-04-02	13073305.0	1783.0	254.714	258835.0	14.0	
270533	NIC	North America	Nicaragua	2020-02-18	0.0	0.0	0.000	0.0	0.0	
355570	SOM	Africa	Somalia	2023-09-20	27334.0	0.0	0.000	1361.0	0.0	
188781	JPN	Asia	Japan	2020-06-08	17141.0	0.0	41.429	916.0	0.0	

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_sm
187153	JAM	North America	Jamaica	2020-07-24	768.0	0.0	2.143	10.0	0.0	
148398	GRL	North America	Greenland	2023-07-19	11971.0	0.0	0.000	21.0	0.0	
185577	ITA	Europe	Italy	2020-11-02	679829.0	0.0	24988.714	38618.0	0.0	
403390	GBR	Europe	United Kingdom	2024-06-05	24935735.0	0.0	344.714	232112.0	0.0	
87256	CUB	North America	Cuba	2020-07-17	2420.0	0.0	7.286	87.0	0.0	
69469	CAF	Africa	Central African Republic	2022-04-05	14352.0	0.0	0.143	113.0	0.0	
101140	DOM	North America	Dominican Republic	2021-11-13	387269.0	0.0	899.286	4150.0	0.0	
217841	OWID_LIC	NaN	Low-income countries	2023-03-04	NaN	NaN	NaN	NaN	NaN	
170625	HUN	Europe	Hungary	2021-03-04	428599.0	0.0	3653.714	14974.0	0.0	
183214	IMN	Europe	Isle of Man	2023-07-15	38008.0	0.0	0.000	116.0	0.0	
131677	FRA	Europe	France	2023-08-07	38997490.0	NaN	NaN	167985.0	NaN	
70654	TCD	Africa	Chad	2020-12-02	1663.0	0.0	4.286	102.0	0.0	
126987	FRO	Europe	Faroe Islands	2024-07-04	34658.0	0.0	0.000	28.0	0.0	

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_sm
<b>24257</b>	AUT	Europe	Austria	2022-03-22	3447369.0	0.0	45586.857	18436.0	0.0	
<b>287408</b>	NOR	Europe	Norway	2024-01-08	1504834.0	0.0	129.000	5732.0	0.0	
<b>323606</b>	KNA	North America	Saint Kitts and Nevis	2022-04-11	5552.0	0.0	0.429	43.0	0.0	
<b>86553</b>	HRV	Europe	Croatia	2023-03-15	1269326.0	0.0	47.714	17987.0	0.0	
<b>4640</b>	ALB	Europe	Albania	2023-07-20	334090.0	0.0	0.000	3604.0	0.0	
<b>1654</b>	AFG	Asia	Afghanistan	2024-07-16	235214.0	0.0	0.000	7998.0	0.0	
<b>86606</b>	HRV	Europe	Croatia	2023-05-07	1272886.0	509.0	72.714	18180.0	25.0	
<b>195547</b>	KEN	Africa	Kenya	2020-08-17	29849.0	0.0	573.143	472.0	0.0	
<b>91827</b>	CYP	Europe	Cyprus	2023-11-22	662962.0	0.0	44.000	1368.0	0.0	
<b>47467</b>	BIH	Europe	Bosnia and Herzegovina	2021-08-08	206216.0	494.0	70.571	9691.0	2.0	
<b>359330</b>	KOR	Asia	South Korea	2020-04-06	10237.0	0.0	93.429	183.0	0.0	
<b>23975</b>	AUT	Europe	Austria	2021-06-13	638645.0	2011.0	287.286	12883.0	39.0	
<b>150623</b>	GLP	North America	Guadeloupe	2020-06-21	174.0	3.0	0.429	14.0	0.0	
<b>390075</b>	TUR	Asia	Turkey	2020-01-21	0.0	0.0	0.000	0.0	0.0	



	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_sm
<b>164600</b>	OWID_HIC	NaN	High-income countries	2022-12-25	NaN	NaN	NaN	NaN	NaN	
<b>417112</b>	VEN	South America	Venezuela	2021-08-31	331112.0	0.0	1083.714	3970.0	0.0	
<b>182600</b>	IMN	Europe	Isle of Man	2021-11-08	9941.0	0.0	42.429	60.0	0.0	
<b>287508</b>	NOR	Europe	Norway	2024-04-17	1507334.0	0.0	9.143	5732.0	0.0	
<b>56878</b>	BGR	Europe	Bulgaria	2024-06-14	1329381.0	0.0	3.571	38700.0	0.0	
<b>112765</b>	ERI	Africa	Eritrea	2022-06-23	9777.0	0.0	1.143	103.0	0.0	
<b>165578</b>	OWID_HIC	NaN	High-income countries	2021-01-18	49020384.0	0.0	450351.571	1035690.0	0.0	8
<b>414762</b>	VUT	Oceania	Vanuatu	2024-05-25	12019.0	0.0	0.000	14.0	0.0	
<b>153966</b>	GTM	North America	Guatemala	2020-06-16	8982.0	0.0	356.714	351.0	0.0	
<b>166992</b>	HND	North America	Honduras	2020-05-03	899.0	308.0	44.000	75.0	20.0	

50 rows × 67 columns

```
In [27]: # Contare il numero di valori nulli in 'continent'
df_covid['continent'].isnull().sum()
```

Out[27]: 26525

```
In [28]: # Creo un nuovo DataFrame con le righe dove 'continent' è NaN
df_continent_null = df_covid[df_covid['continent'].isnull()]

# Visualizzare i primi risultati
df_continent_null
```

Out[28]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
1674	OWID_AFR	NaN	Africa	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1675	OWID_AFR	NaN	Africa	2020-01-06	0.0	0.0	NaN	0.0	0.0	
1676	OWID_AFR	NaN	Africa	2020-01-07	0.0	0.0	NaN	0.0	0.0	
1677	OWID_AFR	NaN	Africa	2020-01-08	0.0	0.0	NaN	0.0	0.0	
1678	OWID_AFR	NaN	Africa	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
424408	OWID_WRL	NaN	World	2024-08-10	NaN	NaN	NaN	NaN	NaN	
424409	OWID_WRL	NaN	World	2024-08-11	NaN	NaN	NaN	NaN	NaN	
424410	OWID_WRL	NaN	World	2024-08-12	NaN	NaN	NaN	NaN	NaN	
424411	OWID_WRL	NaN	World	2024-08-13	NaN	NaN	NaN	NaN	NaN	
424412	OWID_WRL	NaN	World	2024-08-14	NaN	NaN	NaN	NaN	NaN	

26525 rows × 67 columns



In [ ]:

In [29]:

```
# Visualizzo i valori unici per identificare aggregati
print("Valori unici in 'location':", df_covid['location'].unique())
```

Valori unici in 'location': ['Afghanistan' 'Africa' 'Albania' 'Algeria' 'American Samoa' 'Andorra' 'Angola' 'Anguilla' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba' 'Asia' 'Australia' 'Austria' 'Azerbaijan' 'Bahamas' 'Bahrain' 'Bangladesh' 'Barbados' 'Belarus' 'Belgium' 'Belize' 'Benin' 'Bermuda' 'Bhutan' 'Bolivia' 'Bonaire Sint Eustatius and Saba' 'Bosnia and Herzegovina' 'Botswana' 'Brazil' 'British Virgin Islands' 'Brunei' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon' 'Canada' 'Cape Verde' 'Cayman Islands' 'Central African Republic' 'Chad' 'Chile' 'China' 'Colombia' 'Comoros' 'Congo' 'Cook Islands' 'Costa Rica' 'Cote d'Ivoire' 'Croatia' 'Cuba' 'Curacao' 'Cyprus' 'Czechia' 'Democratic Republic of Congo' 'Denmark' 'Djibouti' 'Dominica' 'Dominican Republic' 'East Timor' 'Ecuador' 'Egypt' 'El Salvador' 'England' 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia' 'Europe' 'European Union (27)' 'Faroe Islands' 'Falkland Islands' 'Fiji' 'Finland' 'France' 'French Guiana' 'French Polynesia' 'Gabon' 'Gambia' 'Georgia' 'Germany' 'Ghana' 'Gibraltar' 'Greece' 'Greenland' 'Grenada' 'Guadeloupe' 'Guam' 'Guatemala' 'Guernsey' 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti' 'High-income countries' 'Honduras' 'Hong Kong' 'Hungary' 'Iceland' 'India' 'Indonesia' 'Iran' 'Iraq' 'Ireland' 'Isle of Man' 'Israel' 'Italy' 'Jamaica' 'Japan' 'Jersey' 'Jordan' 'Kazakhstan' 'Kenya' 'Kiribati' 'Kosovo' 'Kuwait' 'Kyrgyzstan' 'Laos' 'Latvia' 'Lebanon' 'Lesotho' 'Liberia' 'Libya' 'Liechtenstein' 'Lithuania' 'Low-income countries' 'Lower-middle-income countries' 'Luxembourg' 'Macao' 'Madagascar' 'Malawi' 'Malaysia' 'Maldives' 'Mali' 'Malta' 'Marshall Islands' 'Martinique' 'Mauritania' 'Mauritius' 'Mayotte' 'Mexico' 'Micronesia (country)' 'Moldova' 'Monaco' 'Mongolia' 'Montenegro' 'Montserrat' 'Morocco' 'Mozambique' 'Myanmar' 'Namibia' 'Nauru' 'Nepal' 'Netherlands' 'New Caledonia' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria' 'Niue' 'North America' 'North Korea' 'North Macedonia' 'Northern Cyprus' 'Northern Ireland' 'Northern Mariana Islands' 'Norway' 'Oceania' 'Oman' 'Pakistan' 'Palau' 'Palestine' 'Panama' 'Papua New Guinea' 'Paraguay' 'Peru' 'Philippines' 'Pitcairn' 'Poland' 'Portugal' 'Puerto Rico' 'Qatar' 'Reunion' 'Romania' 'Russia' 'Rwanda' 'Saint Barthelemy' 'Saint Helena' 'Saint Kitts and Nevis' 'Saint Lucia' 'Saint Martin (French part)' 'Saint Pierre and Miquelon' 'Saint Vincent and the Grenadines' 'Samoa' 'San Marino' 'Sao Tome and Principe' 'Saudi Arabia' 'Scotland' 'Senegal' 'Serbia' 'Seychelles' 'Sierra Leone' 'Singapore' 'Sint Maarten (Dutch part)' 'Slovakia' 'Slovenia' 'Solomon Islands' 'Somalia' 'South Africa' 'South America' 'South Korea' 'South Sudan' 'Spain' 'Sri Lanka' 'Sudan' 'Suriname' 'Sweden' 'Switzerland' 'Syria' 'Taiwan' 'Tajikistan'

```
'Tanzania' 'Thailand' 'Togo' 'Tokelau' 'Tonga' 'Trinidad and Tobago'  
'Tunisia' 'Turkey' 'Turkmenistan' 'Turks and Caicos Islands' 'Tuvalu'  
'Uganda' 'Ukraine' 'United Arab Emirates' 'United Kingdom'  
'United States' 'United States Virgin Islands'  
'Upper-middle-income countries' 'Uruguay' 'Uzbekistan' 'Vanuatu'  
'Vatican' 'Venezuela' 'Vietnam' 'Wales' 'Wallis and Futuna'  
'Western Sahara' 'World' 'Yemen' 'Zambia' 'Zimbabwe']
```

```
In [30]: # Creo un nuovo DataFrame con solo le righe che hanno un 'continent' definito (cioè dati puntuali)  
df_location_puntuale = df_covid[df_covid['continent'].notnull()]  
  
# Visualizzare i primi risultati  
df_location_puntuale
```

Out[30]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

402910 rows × 67 columns



```
In [31]: # Definisco una lista di valori da escludere (aggregati) dalla colonna location
         aggregati_da_escludere = ['World', 'Europe', 'Asia', 'Africa', 'North America', 'South America', 'European Union']
```

```
In [32]: aggregati_da_escludere
```

```
Out[32]: ['World',  
          'Europe',  
          'Asia',  
          'Africa',  
          'North America',  
          'South America',  
          'European Union']
```

```
In [ ]:
```

```
In [33]: # Creo un nuovo DataFrame (df_paesi) filtrando solo i dati a livello di paese, escludendo i valori aggregati dalla colonna Loc  
df_paesi = df_covid[~df_covid['location'].isin(aggregati_da_escludere)].copy()
```

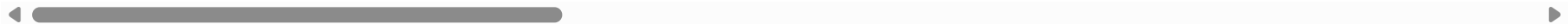
```
In [34]: # Verifico df_paesi  
print("\nNumero di righe dopo il filtro:", len(df_paesi))  
df_paesi
```

Numero di righe dopo il filtro: 419361

Out[34]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

419361 rows × 67 columns



In [35]:

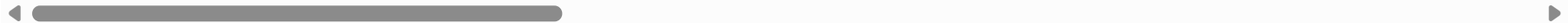
```
df_aggregati = df_covid[df_covid['location'].isin(aggregati_da_escludere)]
df_aggregati # Creo un dataframe contenente esattamente le righe escluse da df_paesi
```



Out[35]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
1674	OWID_AFR	NaN	Africa	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1675	OWID_AFR	NaN	Africa	2020-01-06	0.0	0.0	NaN	0.0	0.0	
1676	OWID_AFR	NaN	Africa	2020-01-07	0.0	0.0	NaN	0.0	0.0	
1677	OWID_AFR	NaN	Africa	2020-01-08	0.0	0.0	NaN	0.0	0.0	
1678	OWID_AFR	NaN	Africa	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
424408	OWID_WRL	NaN	World	2024-08-10	NaN	NaN	NaN	NaN	NaN	
424409	OWID_WRL	NaN	World	2024-08-11	NaN	NaN	NaN	NaN	NaN	
424410	OWID_WRL	NaN	World	2024-08-12	NaN	NaN	NaN	NaN	NaN	
424411	OWID_WRL	NaN	World	2024-08-13	NaN	NaN	NaN	NaN	NaN	
424412	OWID_WRL	NaN	World	2024-08-14	NaN	NaN	NaN	NaN	NaN	

10074 rows × 67 columns



```
In [36]: df_covid['continent'].unique() # Controllo se i valori nan in continent sono presenti in aggregati_da_escludere
```

Out[36]: array(['Asia', nan, 'Europe', 'Africa', 'Oceania', 'North America', 'South America'], dtype=object)

```
In [37]: df_continent_null['location'].unique() # Controllo se i valori nan in continent sono presenti in aggregati_da_escludere
```

```
Out[37]: array(['Africa', 'Asia', 'Europe', 'European Union (27)',  
              'High-income countries', 'Low-income countries',  
              'Lower-middle-income countries', 'North America', 'Oceania',  
              'South America', 'Upper-middle-income countries', 'World'],  
          dtype=object)
```

```
In [38]: aggregati_da_escludere
```

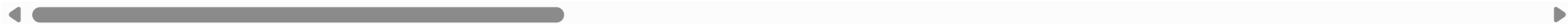
```
Out[38]: ['World',  
          'Europe',  
          'Asia',  
          'Africa',  
          'North America',  
          'South America',  
          'European Union']
```

```
In [39]: df_covid_rimozione_null = df_covid.dropna(subset=['continent'])  
df_covid_rimozione_null # Elimino le righe con continent null creando una nuova variabile (nuovo df)
```

Out[39]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

402910 rows × 67 columns



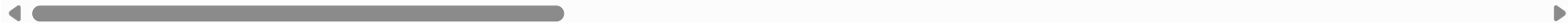
In [40]:

```
df_continent_null
```

Out[40]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
1674	OWID_AFR	NaN	Africa	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1675	OWID_AFR	NaN	Africa	2020-01-06	0.0	0.0	NaN	0.0	0.0	
1676	OWID_AFR	NaN	Africa	2020-01-07	0.0	0.0	NaN	0.0	0.0	
1677	OWID_AFR	NaN	Africa	2020-01-08	0.0	0.0	NaN	0.0	0.0	
1678	OWID_AFR	NaN	Africa	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
424408	OWID_WRL	NaN	World	2024-08-10	NaN	NaN	NaN	NaN	NaN	
424409	OWID_WRL	NaN	World	2024-08-11	NaN	NaN	NaN	NaN	NaN	
424410	OWID_WRL	NaN	World	2024-08-12	NaN	NaN	NaN	NaN	NaN	
424411	OWID_WRL	NaN	World	2024-08-13	NaN	NaN	NaN	NaN	NaN	
424412	OWID_WRL	NaN	World	2024-08-14	NaN	NaN	NaN	NaN	NaN	

26525 rows × 67 columns



```
In [41]: #esporto in csv il dataframe con i valori null della colonna continent
df_continent_null.to_csv("valori_null_continent.csv", index=False)
```

```
In [42]: df_paesi.to_csv("valori_aggregati_location.csv", index=False)
```

```
In [43]: #esporto in csv il dataframe con i valori al netto dei null della colonna continent (null rimossi)
df_covid_rimozione_null.to_csv("valori_continent_rimozione_null.csv", index=False)
```

```
In [44]: #esporto in csv il dataframe con i valori rimossi dalla colonna location (aggregati rimossi)
df_aggregati.to_csv("valori_aggregati_location_rimossi.csv", index=False)
```

```
In [45]: # Esporto in csv il df_paesi
df_paesi.to_csv("df_paesi.csv" , index=False)
```

```
In [46]: print("\nNumero di righe dopo il filtro:", len(df_paesi)) # MI sono reso conto che è ancora presente European Union

# Per cercare una location nel dataframe filtrato (se pensi sia presente)
if 'European Union' in df_paesi['location'].values:
    print(df_paesi.loc[df_paesi['location'] == 'European Union'])
else:
    print("'European Union' non trovata - è stata esclusa dal filtro")
```

Numero di righe dopo il filtro: 419361

'European Union' non trovata - è stata esclusa dal filtro

```
In [47]: # Crea un DataFrame temporaneo con solo le righe da escludere
df_temp = df_paesi[df_paesi['location'].str.contains('Oceania|Europ', na=False, case=False)]

# Visualizzo questo DataFrame temporaneo
print("Righe che verranno escluse:")
print(df_temp)

# Ora lo escludo dal DataFrame principale
df_paesi_new = df_paesi[~df_paesi['location'].str.contains('Oceania|European Union', na=False, regex=True)]

# 4. Verifico
print("\nDataFrame dopo l'esclusione:")
print(f"Righe totali rimaste: {len(df_paesi_new)}")
```

Righe che verranno escluse:

	iso_code	continent	location	date	total_cases	\
120253	OWID_EUN	NaN	European Union (27)	2020-12-04	NaN	
120254	OWID_EUN	NaN	European Union (27)	2020-12-05	NaN	
120255	OWID_EUN	NaN	European Union (27)	2020-12-06	NaN	
120256	OWID_EUN	NaN	European Union (27)	2020-12-07	NaN	
120257	OWID_EUN	NaN	European Union (27)	2020-12-08	NaN	
...	...	...	...	...	...	
289294	OWID_OCE	NaN	Oceania	2024-08-07	NaN	
289295	OWID_OCE	NaN	Oceania	2024-08-08	NaN	
289296	OWID_OCE	NaN	Oceania	2024-08-09	NaN	
289297	OWID_OCE	NaN	Oceania	2024-08-10	NaN	
289298	OWID_OCE	NaN	Oceania	2024-08-11	NaN	

	new_cases	new_cases_smoothed	total_deaths	new_deaths	\
120253	NaN	NaN	NaN	NaN	
120254	NaN	NaN	NaN	NaN	
120255	NaN	NaN	NaN	NaN	
120256	NaN	NaN	NaN	NaN	
120257	NaN	NaN	NaN	NaN	
...	...	...	...	...	
289294	NaN	NaN	NaN	NaN	
289295	NaN	NaN	NaN	NaN	
289296	NaN	NaN	NaN	NaN	
289297	NaN	NaN	NaN	NaN	
289298	NaN	NaN	NaN	NaN	

	new_deaths_smoothed	...	male_smokers	handwashing_facilities	\
120253	NaN	...	NaN	NaN	
120254	NaN	...	NaN	NaN	
120255	NaN	...	NaN	NaN	
120256	NaN	...	NaN	NaN	
120257	NaN	...	NaN	NaN	
...	...	...	...	...	
289294	NaN	...	NaN	NaN	
289295	NaN	...	NaN	NaN	
289296	NaN	...	NaN	NaN	
289297	NaN	...	NaN	NaN	
289298	NaN	...	NaN	NaN	

hospital_beds_per_thousand	life_expectancy	human_development_index	\
----------------------------	-----------------	-------------------------	---

120253	NaN	NaN	NaN
120254	NaN	NaN	NaN
120255	NaN	NaN	NaN
120256	NaN	NaN	NaN
120257	NaN	NaN	NaN
...	...	...	...
289294	NaN	NaN	NaN
289295	NaN	NaN	NaN
289296	NaN	NaN	NaN
289297	NaN	NaN	NaN
289298	NaN	NaN	NaN

	population	excess_mortality_cumulative_absolute	\
120253	450146793	NaN	
120254	450146793	NaN	
120255	450146793	NaN	
120256	450146793	NaN	
120257	450146793	NaN	
...	...	...	
289294	45038860	NaN	
289295	45038860	NaN	
289296	45038860	NaN	
289297	45038860	NaN	
289298	45038860	NaN	

	excess_mortality_cumulative	excess_mortality	\
120253	NaN	NaN	
120254	NaN	NaN	
120255	NaN	NaN	
120256	NaN	NaN	
120257	NaN	NaN	
...	...	...	
289294	NaN	NaN	
289295	NaN	NaN	
289296	NaN	NaN	
289297	NaN	NaN	
289298	NaN	NaN	

	excess_mortality_cumulative_per_million
120253	NaN
120254	NaN

120255	NaN
120256	NaN
120257	NaN
...	...
289294	NaN
289295	NaN
289296	NaN
289297	NaN
289298	NaN

[4705 rows x 67 columns]

DataFrame dopo l'esclusione:  
Righe totali rimaste: 414656

```
In [48]: # Esporto in csv il df_paesi
df_paesi_new.to_csv("df_paesi_new.csv" , index=False)
```

```
In [49]: df_temp
```



Out[49]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smo
120253	OWID_EUN	NaN	European Union (27)	2020-12-04	NaN	NaN	NaN	NaN	NaN	
120254	OWID_EUN	NaN	European Union (27)	2020-12-05	NaN	NaN	NaN	NaN	NaN	
120255	OWID_EUN	NaN	European Union (27)	2020-12-06	NaN	NaN	NaN	NaN	NaN	
120256	OWID_EUN	NaN	European Union (27)	2020-12-07	NaN	NaN	NaN	NaN	NaN	
120257	OWID_EUN	NaN	European Union (27)	2020-12-08	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	
289294	OWID_OCE	NaN	Oceania	2024-08-07	NaN	NaN	NaN	NaN	NaN	
289295	OWID_OCE	NaN	Oceania	2024-08-08	NaN	NaN	NaN	NaN	NaN	
289296	OWID_OCE	NaN	Oceania	2024-08-09	NaN	NaN	NaN	NaN	NaN	
289297	OWID_OCE	NaN	Oceania	2024-08-10	NaN	NaN	NaN	NaN	NaN	
289298	OWID_OCE	NaN	Oceania	2024-08-11	NaN	NaN	NaN	NaN	NaN	

4705 rows × 67 columns



```
In [50]: print("\nNumero di righe dopo il filtro:", len(df_paesi))

# Per cercare una location nel dataframe filtrato
if 'Oceania' in df_paesi['location'].values:
    print(df_paesi.loc[df_paesi['location'] == 'Oceania'])
else:
    print("'Oceania' non trovata - è stata esclusa dal filtro")
```

Numero di righe dopo il filtro: 419361

	iso_code	continent	location	date	total_cases	new_cases	\
287618	OWID_OCE	NaN	Oceania	2020-01-05	0.0	0.0	
287619	OWID_OCE	NaN	Oceania	2020-01-06	0.0	0.0	
287620	OWID_OCE	NaN	Oceania	2020-01-07	0.0	0.0	
287621	OWID_OCE	NaN	Oceania	2020-01-08	0.0	0.0	
287622	OWID_OCE	NaN	Oceania	2020-01-09	0.0	0.0	
...	...	...	...	...	...	...	
289294	OWID_OCE	NaN	Oceania	2024-08-07	NaN	NaN	
289295	OWID_OCE	NaN	Oceania	2024-08-08	NaN	NaN	
289296	OWID_OCE	NaN	Oceania	2024-08-09	NaN	NaN	
289297	OWID_OCE	NaN	Oceania	2024-08-10	NaN	NaN	
289298	OWID_OCE	NaN	Oceania	2024-08-11	NaN	NaN	
	new_cases_smoothed		total_deaths	new_deaths	new_deaths_smoothed		\
287618	NaN		0.0	0.0	NaN		
287619	NaN		0.0	0.0	NaN		
287620	NaN		0.0	0.0	NaN		
287621	NaN		0.0	0.0	NaN		
287622	NaN		0.0	0.0	NaN		
...	...		...	...	...		
289294	NaN		NaN	NaN	NaN		
289295	NaN		NaN	NaN	NaN		
289296	NaN		NaN	NaN	NaN		
289297	NaN		NaN	NaN	NaN		
289298	NaN		NaN	NaN	NaN		
	...	male_smokers	handwashing_facilities		hospital_beds_per_thousand		\
287618	...	NaN	NaN		NaN		
287619	...	NaN	NaN		NaN		
287620	...	NaN	NaN		NaN		
287621	...	NaN	NaN		NaN		
287622	...	NaN	NaN		NaN		
...	...	...	...		...		
289294	...	NaN	NaN		NaN		
289295	...	NaN	NaN		NaN		
289296	...	NaN	NaN		NaN		
289297	...	NaN	NaN		NaN		
289298	...	NaN	NaN		NaN		

life\_expectancy human\_development\_index population \

287618	NaN	NaN	45038860
287619	NaN	NaN	45038860
287620	NaN	NaN	45038860
287621	NaN	NaN	45038860
287622	NaN	NaN	45038860
...	...	...	...
289294	NaN	NaN	45038860
289295	NaN	NaN	45038860
289296	NaN	NaN	45038860
289297	NaN	NaN	45038860
289298	NaN	NaN	45038860

	excess_mortality_cumulative_absolute	excess_mortality_cumulative	\
287618	NaN	NaN	
287619	NaN	NaN	
287620	NaN	NaN	
287621	NaN	NaN	
287622	NaN	NaN	
...	...	...	
289294	NaN	NaN	
289295	NaN	NaN	
289296	NaN	NaN	
289297	NaN	NaN	
289298	NaN	NaN	

	excess_mortality	excess_mortality_cumulative_per_million
287618	NaN	NaN
287619	NaN	NaN
287620	NaN	NaN
287621	NaN	NaN
287622	NaN	NaN
...	...	...
289294	NaN	NaN
289295	NaN	NaN
289296	NaN	NaN
289297	NaN	NaN
289298	NaN	NaN

[1681 rows x 67 columns]

```
In [51]: # Così ho ripristinato i null rimossi dal df_covid con la riga 194 prima di creare la variabile df_covid_rimozione_null
# df_covid = pd.concat([df_covid, df_continent_null], ignore_index=True)
```

```
In [52]: df_paesi
```

Out[52]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

419361 rows × 67 columns



```
In [53]: df_covid_rimozione_null
```

Out[53]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
--	----------	-----------	----------	------	-------------	-----------	--------------------	--------------	------------	-----------------

0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

402910 rows × 67 columns



```
In [54]: df_covid
```

Out[54]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

429435 rows × 67 columns



```
In [55]: totale_righe = len(df_covid)
         totale_righe # totale righe del df_covid (df originario)
```

Out[55]: 429435

```
In [56]: valori_nulli_continent = df_covid['continent'].isnull().sum()  
valori_nulli_continent # Totale dei valori null di continent
```

```
Out[56]: 26525
```

```
In [57]: percentuale_nulli_continent = (valori_nulli_continent / totale_righe) * 100  
percentuale_nulli_continent
```

```
Out[57]: 6.176720574708629
```

```
In [58]: numero_righe_location = len(df_paesi)  
numero_righe_location # Numero righe del df_paesi al netto delle righe rimosse (numero_aggregati_location)
```

```
Out[58]: 419361
```

```
In [59]: numero_righe_location_new = len(df_paesi_new)  
numero_righe_location_new # Numero righe del df_paesi_new (dopo la rimozione di Oceania e European Union da Location)  
# al netto delle righe rimosse (numero_aggregati_location + Oceania e European Union)
```

```
Out[59]: 414656
```

```
In [60]: percentuale_df_paesi_location = (numero_righe_location / totale_righe) * 100  
percentuale_df_paesi_location
```

```
Out[60]: 97.65412693422753
```

```
In [61]: percentuale_df_paesi_location = (numero_righe_location_new / totale_righe) * 100  
percentuale_df_paesi_location
```

```
Out[61]: 96.55850128657421
```

```
In [62]: numero_aggregati_location = len(df_aggregati)  
numero_aggregati_location # Numero di righe del df_aggregati (location) rimosse dal df_covid (df originario)
```

```
Out[62]: 10074
```

```
In [63]: percentuale_df_aggregati_location = (numero_aggregati_location / totale_righe) * 100  
percentuale_df_aggregati_location
```



Out[63]: 2.3458730657724685

```
In [64]: # 1. Calcola il numero di righe ORIGINALI (prima di qualsiasi filtro)
totale_righe_originali = len(df_paesi) # Assicurati di aver salvato il df prima del filtro

# 2. Calcola le righe da rimuovere (Oceania/Europe)
righe_da_rimuovere = len(df_temp)

# 3. Calcola la PERCENTUALE delle righe rimosse
percentuale_rimosse = (righe_da_rimuovere / totale_righe_originali) * 100

print(f"Righe totali originali: {totale_righe_originali}")
print(f"Righe rimosse: {righe_da_rimuovere}")
print(f"Percentuale rimossa: {percentuale_rimosse:.2f}%")
```

Righe totali originali: 419361

Righe rimosse: 4705

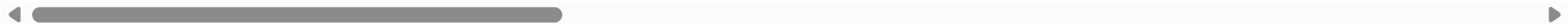
Percentuale rimossa: 1.12%

```
In [65]: df_covid_new = df_covid_rimozione_null
df_covid_new
```

Out[65]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

402910 rows × 67 columns



In [66]: df\_covid\_new.info()

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 402910 entries, 0 to 429434
```

```
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	iso_code	402910 non-null	object
1	continent	402910 non-null	object
2	location	402910 non-null	object
3	date	402910 non-null	object
4	total_cases	391716 non-null	float64
5	new_cases	390071 non-null	float64
6	new_cases_smoothed	388901 non-null	float64
7	total_deaths	391716 non-null	float64
8	new_deaths	390520 non-null	float64
9	new_deaths_smoothed	389350 non-null	float64
10	total_cases_per_million	391716 non-null	float64
11	new_cases_per_million	390071 non-null	float64
12	new_cases_smoothed_per_million	388901 non-null	float64
13	total_deaths_per_million	391716 non-null	float64
14	new_deaths_per_million	390520 non-null	float64
15	new_deaths_smoothed_per_million	389350 non-null	float64
16	reproduction_rate	183741 non-null	float64
17	icu_patients	39116 non-null	float64
18	icu_patients_per_million	39116 non-null	float64
19	hosp_patients	40656 non-null	float64
20	hosp_patients_per_million	40656 non-null	float64
21	weekly_icu_admissions	10993 non-null	float64
22	weekly_icu_admissions_per_million	10993 non-null	float64
23	weekly_hosp_admissions	24497 non-null	float64
24	weekly_hosp_admissions_per_million	24497 non-null	float64
25	total_tests	79387 non-null	float64
26	new_tests	75403 non-null	float64
27	total_tests_per_thousand	79387 non-null	float64
28	new_tests_per_thousand	75403 non-null	float64
29	new_tests_smoothed	103965 non-null	float64
30	new_tests_smoothed_per_thousand	103965 non-null	float64
31	positive_rate	95927 non-null	float64
32	tests_per_case	94348 non-null	float64
33	tests_units	106788 non-null	object
34	total_vaccinations	70150 non-null	float64
35	people_vaccinated	65865 non-null	float64

```

36 people_fully_vaccinated      62980 non-null float64
37 total_boosters               39539 non-null float64
38 new_vaccinations             55747 non-null float64
39 new_vaccinations_smoothed    179805 non-null float64
40 total_vaccinations_per_hundred 70150 non-null float64
41 people_vaccinated_per_hundred 65865 non-null float64
42 people_fully_vaccinated_per_hundred 62980 non-null float64
43 total_boosters_per_hundred    39539 non-null float64
44 new_vaccinations_smoothed_per_million 179805 non-null float64
45 new_people_vaccinated_smoothed 176953 non-null float64
46 new_people_vaccinated_smoothed_per_hundred 176953 non-null float64
47 stringency_index             196190 non-null float64
48 population_density           358808 non-null float64
49 median_age                   332979 non-null float64
50 aged_65_older                321586 non-null float64
51 aged_70_older                329631 non-null float64
52 gdp_per_capita                326608 non-null float64
53 extreme_poverty              210312 non-null float64
54 cardiovascular_death_rate     327181 non-null float64
55 diabetes_prevalence           344227 non-null float64
56 female_smokers                245481 non-null float64
57 male_smokers                  242133 non-null float64
58 handwashing_facilities        160057 non-null float64
59 hospital_beds_per_thousand    289005 non-null float64
60 life_expectancy               388615 non-null float64
61 human_development_index       317443 non-null float64
62 population                    402910 non-null int64
63 excess_mortality_cumulative_absolute 13411 non-null float64
64 excess_mortality_cumulative    13411 non-null float64
65 excess_mortality              13411 non-null float64
66 excess_mortality_cumulative_per_million 13411 non-null float64
dtypes: float64(61), int64(1), object(5)
memory usage: 209.0+ MB

```

```

In [67]: # Verifico la presenza di null in seguito alla pulizia (della colonna continent) e alla creazione del dataframe finale df_covi
df_covid_new.isnull().sum()

```

```
Out[67]: iso_code          0
continent          0
location          0
date              0
total_cases      11194
...
population        0
excess_mortality_cumulative_absolute  389499
excess_mortality_cumulative          389499
excess_mortality                    389499
excess_mortality_cumulative_per_million  389499
Length: 67, dtype: int64
```

```
In [68]: # Ora genero un dizionario per memorizzare i dataframe continent separati
continents = ['Asia', 'Europe', 'Africa', 'Oceania', 'North America', 'South America'] # Qui creo una variabile continenti
continent_dfs = {} #Qui creo il dizionario
for continent in continents:
    continent_dfs[continent] = df_covid_new[df_covid_new['continent'] == continent].copy() # Qui filtro il DataFrame per il c
# Ora posso accedere a ciascun DataFrame con:
# continent_dfs['Asia'] - continent_dfs['Europe'] - continent_dfs['Africa'] - continent_dfs['Oceania'] - continent_dfs['df_Nor
```

```
In [69]: df_covid_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 402910 entries, 0 to 429434
```

```
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	iso_code	402910 non-null	object
1	continent	402910 non-null	object
2	location	402910 non-null	object
3	date	402910 non-null	object
4	total_cases	391716 non-null	float64
5	new_cases	390071 non-null	float64
6	new_cases_smoothed	388901 non-null	float64
7	total_deaths	391716 non-null	float64
8	new_deaths	390520 non-null	float64
9	new_deaths_smoothed	389350 non-null	float64
10	total_cases_per_million	391716 non-null	float64
11	new_cases_per_million	390071 non-null	float64
12	new_cases_smoothed_per_million	388901 non-null	float64
13	total_deaths_per_million	391716 non-null	float64
14	new_deaths_per_million	390520 non-null	float64
15	new_deaths_smoothed_per_million	389350 non-null	float64
16	reproduction_rate	183741 non-null	float64
17	icu_patients	39116 non-null	float64
18	icu_patients_per_million	39116 non-null	float64
19	hosp_patients	40656 non-null	float64
20	hosp_patients_per_million	40656 non-null	float64
21	weekly_icu_admissions	10993 non-null	float64
22	weekly_icu_admissions_per_million	10993 non-null	float64
23	weekly_hosp_admissions	24497 non-null	float64
24	weekly_hosp_admissions_per_million	24497 non-null	float64
25	total_tests	79387 non-null	float64
26	new_tests	75403 non-null	float64
27	total_tests_per_thousand	79387 non-null	float64
28	new_tests_per_thousand	75403 non-null	float64
29	new_tests_smoothed	103965 non-null	float64
30	new_tests_smoothed_per_thousand	103965 non-null	float64
31	positive_rate	95927 non-null	float64
32	tests_per_case	94348 non-null	float64
33	tests_units	106788 non-null	object
34	total_vaccinations	70150 non-null	float64
35	people_vaccinated	65865 non-null	float64

```

36 people_fully_vaccinated      62980 non-null    float64
37 total_boosters               39539 non-null    float64
38 new_vaccinations             55747 non-null    float64
39 new_vaccinations_smoothed    179805 non-null   float64
40 total_vaccinations_per_hundred 70150 non-null    float64
41 people_vaccinated_per_hundred 65865 non-null    float64
42 people_fully_vaccinated_per_hundred 62980 non-null    float64
43 total_boosters_per_hundred    39539 non-null    float64
44 new_vaccinations_smoothed_per_million 179805 non-null   float64
45 new_people_vaccinated_smoothed 176953 non-null   float64
46 new_people_vaccinated_smoothed_per_hundred 176953 non-null   float64
47 stringency_index             196190 non-null   float64
48 population_density           358808 non-null   float64
49 median_age                   332979 non-null   float64
50 aged_65_older                321586 non-null   float64
51 aged_70_older                329631 non-null   float64
52 gdp_per_capita                326608 non-null   float64
53 extreme_poverty              210312 non-null   float64
54 cardiovascular_death_rate     327181 non-null   float64
55 diabetes_prevalence           344227 non-null   float64
56 female_smokers                245481 non-null   float64
57 male_smokers                  242133 non-null   float64
58 handwashing_facilities        160057 non-null   float64
59 hospital_beds_per_thousand    289005 non-null   float64
60 life_expectancy               388615 non-null   float64
61 human_development_index       317443 non-null   float64
62 population                    402910 non-null   int64
63 excess_mortality_cumulative_absolute 13411 non-null    float64
64 excess_mortality_cumulative    13411 non-null    float64
65 excess_mortality              13411 non-null    float64
66 excess_mortality_cumulative_per_million 13411 non-null    float64
dtypes: float64(61), int64(1), object(5)
memory usage: 209.0+ MB

```

```

In [70]: # Ora genero un dataframe per ogni continente, creando variabili separate per ciascun continente
df_asia = df_covid_new[df_covid_new['continent'] == 'Asia'].copy()
df_europe = df_covid_new[df_covid_new['continent'] == 'Europe'].copy()
df_africa = df_covid_new[df_covid_new['continent'] == 'Africa'].copy()
df_oceania = df_covid_new[df_covid_new['continent'] == 'Oceania'].copy()

```

```
df_north_america = df_covid_new[df_covid_new['continent'] == 'North America'].copy()
df_south_america = df_covid_new[df_covid_new['continent'] == 'South America'].copy()
```

In [71]: df\_south\_america

Out[71]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
15066	ARG	South America	Argentina	2020-01-01	NaN	NaN	NaN	NaN	NaN	NaN
15067	ARG	South America	Argentina	2020-01-02	NaN	NaN	NaN	NaN	NaN	NaN
15068	ARG	South America	Argentina	2020-01-03	NaN	NaN	NaN	NaN	NaN	NaN
15069	ARG	South America	Argentina	2020-01-04	NaN	NaN	NaN	NaN	NaN	NaN
15070	ARG	South America	Argentina	2020-01-05	0.0	0.0	NaN	0.0	0.0	NaN
...	...	...	...	...	...	...	...	...	...	...
418177	VEN	South America	Venezuela	2024-07-31	552695.0	0.0	0.0	5856.0	0.0	0.0
418178	VEN	South America	Venezuela	2024-08-01	552695.0	0.0	0.0	5856.0	0.0	0.0
418179	VEN	South America	Venezuela	2024-08-02	552695.0	0.0	0.0	5856.0	0.0	0.0
418180	VEN	South America	Venezuela	2024-08-03	552695.0	0.0	0.0	5856.0	0.0	0.0
418181	VEN	South America	Venezuela	2024-08-04	552695.0	0.0	0.0	5856.0	0.0	0.0

23440 rows × 67 columns





```
In [72]: df_north_america
```

Out[72]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
11718	AIA	North America	Anguilla	2020-01-05	0.0	0.0	NaN	0.0	0.0	NaN
11719	AIA	North America	Anguilla	2020-01-06	0.0	0.0	NaN	0.0	0.0	NaN
11720	AIA	North America	Anguilla	2020-01-07	0.0	0.0	NaN	0.0	0.0	NaN
11721	AIA	North America	Anguilla	2020-01-08	0.0	0.0	NaN	0.0	0.0	NaN
11722	AIA	North America	Anguilla	2020-01-09	0.0	0.0	NaN	0.0	0.0	NaN
...	...	...	...	...	...	...	...	...	...	...
406794	VIR	North America	United States Virgin Islands	2024-07-31	25389.0	0.0	0.0	132.0	0.0	0.0
406795	VIR	North America	United States Virgin Islands	2024-08-01	25389.0	0.0	0.0	132.0	0.0	0.0
406796	VIR	North America	United States Virgin Islands	2024-08-02	25389.0	0.0	0.0	132.0	0.0	0.0
406797	VIR	North America	United States Virgin Islands	2024-08-03	25389.0	0.0	0.0	132.0	0.0	0.0

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
406798	VIR	North America	United States Virgin Islands	2024-08-04	25389.0	0.0	0.0	132.0	0.0	0

68638 rows × 67 columns

```
In [73]: df_oceania
```

Out[73]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smooth
6696	ASM	Oceania	American Samoa	2020-01-05	0.0	0.0	NaN	0.0	0.0	N
6697	ASM	Oceania	American Samoa	2020-01-06	0.0	0.0	NaN	0.0	0.0	N
6698	ASM	Oceania	American Samoa	2020-01-07	0.0	0.0	NaN	0.0	0.0	N
6699	ASM	Oceania	American Samoa	2020-01-08	0.0	0.0	NaN	0.0	0.0	N
6700	ASM	Oceania	American Samoa	2020-01-09	0.0	0.0	NaN	0.0	0.0	N
...	...	...	...	...	...	...	...	...	...	
422723	WLF	Oceania	Wallis and Futuna	2024-07-31	3760.0	0.0	0.0	9.0	0.0	
422724	WLF	Oceania	Wallis and Futuna	2024-08-01	3760.0	0.0	0.0	9.0	0.0	
422725	WLF	Oceania	Wallis and Futuna	2024-08-02	3760.0	0.0	0.0	9.0	0.0	
422726	WLF	Oceania	Wallis and Futuna	2024-08-03	3760.0	0.0	0.0	9.0	0.0	
422727	WLF	Oceania	Wallis and Futuna	2024-08-04	3760.0	0.0	0.0	9.0	0.0	

40183 rows × 67 columns



```
In [74]: df_africa
```

Out[74]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
--	----------	-----------	----------	------	-------------	-----------	--------------------	--------------	------------	---------------------

5022	DZA	Africa	Algeria	2020-01-05	0.0	0.0	NaN	0.0	0.0	NaN
5023	DZA	Africa	Algeria	2020-01-06	0.0	0.0	NaN	0.0	0.0	NaN
5024	DZA	Africa	Algeria	2020-01-07	0.0	0.0	NaN	0.0	0.0	NaN
5025	DZA	Africa	Algeria	2020-01-08	0.0	0.0	NaN	0.0	0.0	NaN
5026	DZA	Africa	Algeria	2020-01-09	0.0	0.0	NaN	0.0	0.0	NaN
...	...	...	...	...	...	...	...	...	...	...
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	0.0
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	0.0
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	0.0
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	0.0
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	0.0

95419 rows × 67 columns

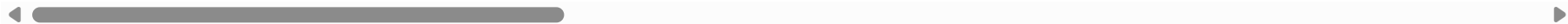


```
In [75]: df_europe
```

Out[75]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	
3348	ALB	Europe	Albania	2020-01-05	0.0	0.0	NaN	0.0	0.0		
3349	ALB	Europe	Albania	2020-01-06	0.0	0.0	NaN	0.0	0.0		
3350	ALB	Europe	Albania	2020-01-07	0.0	0.0	NaN	0.0	0.0		
3351	ALB	Europe	Albania	2020-01-08	0.0	0.0	NaN	0.0	0.0		
3352	ALB	Europe	Albania	2020-01-09	0.0	0.0	NaN	0.0	0.0		
...	...	...	...	...	...	...	...	...	...		
421049	OWID_WLS	Europe	Wales	2023-07-08	NaN	NaN	NaN	NaN	NaN		
421050	OWID_WLS	Europe	Wales	2023-07-09	NaN	NaN	NaN	NaN	NaN		
421051	OWID_WLS	Europe	Wales	2023-07-10	NaN	NaN	NaN	NaN	NaN		
421052	OWID_WLS	Europe	Wales	2023-07-11	NaN	NaN	NaN	NaN	NaN		
421053	OWID_WLS	Europe	Wales	2023-07-12	NaN	NaN	NaN	NaN	NaN		

91031 rows × 67 columns



In [76]:

```
df_asia
```

Out[76]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
426082	YEM	Asia	Yemen	2024-07-31	11945.0	0.0	0.0	2159.0	0.0	
426083	YEM	Asia	Yemen	2024-08-01	11945.0	0.0	0.0	2159.0	0.0	
426084	YEM	Asia	Yemen	2024-08-02	11945.0	0.0	0.0	2159.0	0.0	
426085	YEM	Asia	Yemen	2024-08-03	11945.0	0.0	0.0	2159.0	0.0	
426086	YEM	Asia	Yemen	2024-08-04	11945.0	0.0	0.0	2159.0	0.0	

84199 rows × 67 columns



```
In [77]: # Lista dei dataframe continentali
continental_dfs = {
    'Asia': df_asia,
    'Europa': df_europe,
    'Africa': df_africa,
```

```

'Oceania': df_oceania,
'Nord America': df_north_america,
'Sud America': df_south_america
}

# Genero una stampa ordinata delle dimensioni
print("\nDIMENSIONI DATAFRAME CONTINENTALI")
print("-" * 40)
for name, df in continental_dfs.items():
    print(f"{name:<15}: {len(df):>6} righe | {len(df.columns):>2} colonne")
print("-" * 40)
total_rows = sum(len(df) for df in continental_dfs.values())
print(f"TOTALE{' ':>8}: {total_rows:>6} righe")

```

#### DIMENSIONI DATAFRAME CONTINENTALI

```

-----
Asia          : 84199 righe | 67 colonne
Europa        : 91031 righe | 67 colonne
Africa        : 95419 righe | 67 colonne
Oceania       : 40183 righe | 67 colonne
Nord America  : 68638 righe | 67 colonne
Sud America   : 23440 righe | 67 colonne

```

```

-----
TOTALE        : 402910 righe

```

VALUTAZIONI: DOPO AVER GENERATO UNA SERIE DI VARIABILI TALI DA PERMETTERMI DI INDIVIDUARE I VALORI NULLI NELLA COLONNA CONTINENT E I VALORI AGGREGATI NELLA COLONNA LOCATION, HO VALUTATO DI VOLER RIMUOVERE DAL DATASET I VALORI NULL DALLA COLONNA CONTINENT (QUINDI NON LA RIMOZIONE DEI VALORI AGGREGATI DALLA COLONNA LOCATION), IN QUANTO, IL RESULT SET FINALE CON IL NUOVO DATAFRAME df\_paesi e successivamente df\_paesi\_new ANDAVA A RIMUOVERE SINGOLARMENTE E PASSO PASSO TUTTI I VALORI RIDONDANTI NELLA COLONNA LOCATION MA CONTINUAVANO A PERSISTERE ALCUNI VALORI NULL NELLA COLONNA CONTINENT; PERTANTO HO OPTATO PER LA DECISIONE PIU' DRASTICA MA ANCHE PIU' COMPLETA E DEFINITIVA' (MANO A MANO CHE PROSEGUIVO CON LA SCREMATURA, GENERAVO FILE CSV CON I QUALI HO APPROFONDITO L'ANALISI VISIVA TRAMITE EXCEL, IN MODO PIU' APPROFONDITO); IN CONCLUSIONE HO RIMOSSO I VALORI NULL DA CONTINENT (OVVIAMENTE NON HO RIMOSSO NULLA IN MODO DEFINITIVO FACENDO UN IMPLACE, MA HO GENERATO UN DATA FRAME PIU PULITO CON IL QUALE AFFRONTARE LE TRACCE DELL'ESERCIZIO).



## VERIFICO LA SECONDA NOTA

```
In [80]: casi_per_continente = df_covid.groupby('continent')['new_cases'].sum().sum()  
print(f'I casi per continente sono:{casi_per_continente}') # Con questo groupby vado a calcolare il numero di new_cases sul da
```

I casi per continente sono:775935057.0

```
In [81]: casi_per_continente = df_covid.groupby('location')['new_cases'].sum().sum()  
print(f'I casi per continente sono:{casi_per_continente}')
```

I casi per continente sono:3288392333.0

```
In [82]: casi_per_continente = df_covid_new.groupby('continent')['new_cases'].sum().sum()  
print(f'I casi per continente sono:{casi_per_continente}')
```

I casi per continente sono:775935057.0

```
In [83]: casi_per_continente = df_covid_new.groupby('location')['new_cases'].sum().sum()  
print(f'I casi per continente sono:{casi_per_continente}')
```

I casi per continente sono:775935057.0

```
In [84]: df_covid_new["date"] = pd.to_datetime(df_covid_new["date"])
```

C:\Users\bocci\AppData\Local\Temp\ipykernel\_29880\1539548751.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_covid_new["date"] = pd.to_datetime(df_covid_new["date"])
```

```
In [85]: df_covid_new["date"].dtype
```

Out[85]: dtype('<M8[ns]')

```
In [348]: filtro_italia_2022 = (df_covid_new["location"] == "Italy") & (df_covid_new["date"].dt.year == 2022)  
df_covid_new.loc[filtro_italia_2022]
```

Out[348...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186002</b>	ITA	Europe	Italy	2022-01-01	5622431.0	0.0	36797.000	136530.0	0.0	140.81
<b>186003</b>	ITA	Europe	Italy	2022-01-02	6267035.0	644604.0	92086.286	137513.0	983.0	140.42
<b>186004</b>	ITA	Europe	Italy	2022-01-03	6267035.0	0.0	92086.286	137513.0	0.0	140.42
<b>186005</b>	ITA	Europe	Italy	2022-01-04	6267035.0	0.0	92086.286	137513.0	0.0	140.42
<b>186006</b>	ITA	Europe	Italy	2022-01-05	6267035.0	0.0	92086.286	137513.0	0.0	140.42
...	...	...	...	...	...	...	...	...	...	...
<b>186362</b>	ITA	Europe	Italy	2022-12-27	25060503.0	0.0	18918.143	184168.0	0.0	114.00
<b>186363</b>	ITA	Europe	Italy	2022-12-28	25060503.0	0.0	18918.143	184168.0	0.0	114.00
<b>186364</b>	ITA	Europe	Italy	2022-12-29	25060503.0	0.0	18918.143	184168.0	0.0	114.00
<b>186365</b>	ITA	Europe	Italy	2022-12-30	25060503.0	0.0	18918.143	184168.0	0.0	114.00
<b>186366</b>	ITA	Europe	Italy	2022-12-31	25060503.0	0.0	18918.143	184168.0	0.0	114.00

365 rows × 67 columns



```
In [87]: filtro_italia_2022 = (df_covid_new["location"] == "Italy") & (df_covid_new["date"].dt.year == 2022)
maxmin= df_covid_new.loc[filtro_italia_2022].head(80)
pd.set_option('display.max_rows', 100) # Numero massimo di righe da visualizzare
```

```
pd.set_option('display.max_columns', 50) # Numero massimo di colonne da visualizzare  
maxmin
```

Out[87]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
186002	ITA	Europe	Italy	2022-01-01	5622431.0	0.0	36797.000	136530.0	0.0	140.81
186003	ITA	Europe	Italy	2022-01-02	6267035.0	644604.0	92086.286	137513.0	983.0	140.42
186004	ITA	Europe	Italy	2022-01-03	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186005	ITA	Europe	Italy	2022-01-04	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186006	ITA	Europe	Italy	2022-01-05	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186007	ITA	Europe	Italy	2022-01-06	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186008	ITA	Europe	Italy	2022-01-07	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186009	ITA	Europe	Italy	2022-01-08	6267035.0	0.0	92086.286	137513.0	0.0	140.42
186010	ITA	Europe	Italy	2022-01-09	7281297.0	1014262.0	144894.571	138881.0	1368.0	195.42
186011	ITA	Europe	Italy	2022-01-10	7281297.0	0.0	144894.571	138881.0	0.0	195.42
186012	ITA	Europe	Italy	2022-01-11	7281297.0	0.0	144894.571	138881.0	0.0	195.42
186013	ITA	Europe	Italy	2022-01-12	7281297.0	0.0	144894.571	138881.0	0.0	195.42
186014	ITA	Europe	Italy	2022-01-13	7281297.0	0.0	144894.571	138881.0	0.0	195.42
186015	ITA	Europe	Italy	2022-01-14	7281297.0	0.0	144894.571	138881.0	0.0	195.42

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186016</b>	ITA	Europe	Italy	2022-01-15	7281297.0	0.0	144894.571	138881.0	0.0	195.46
<b>186017</b>	ITA	Europe	Italy	2022-01-16	8549450.0	1268153.0	181164.714	140856.0	1975.0	282.14
<b>186018</b>	ITA	Europe	Italy	2022-01-17	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186019</b>	ITA	Europe	Italy	2022-01-18	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186020</b>	ITA	Europe	Italy	2022-01-19	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186021</b>	ITA	Europe	Italy	2022-01-20	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186022</b>	ITA	Europe	Italy	2022-01-21	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186023</b>	ITA	Europe	Italy	2022-01-22	8549450.0	0.0	181164.714	140856.0	0.0	282.14
<b>186024</b>	ITA	Europe	Italy	2022-01-23	9781191.0	1231741.0	175963.000	143296.0	2440.0	348.57
<b>186025</b>	ITA	Europe	Italy	2022-01-24	9781191.0	0.0	175963.000	143296.0	0.0	348.57
<b>186026</b>	ITA	Europe	Italy	2022-01-25	9781191.0	0.0	175963.000	143296.0	0.0	348.57
<b>186027</b>	ITA	Europe	Italy	2022-01-26	9781191.0	0.0	175963.000	143296.0	0.0	348.57
<b>186028</b>	ITA	Europe	Italy	2022-01-27	9781191.0	0.0	175963.000	143296.0	0.0	348.57
<b>186029</b>	ITA	Europe	Italy	2022-01-28	9781191.0	0.0	175963.000	143296.0	0.0	348.57

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186030</b>	ITA	Europe	Italy	2022-01-29	9781191.0	0.0	175963.000	143296.0	0.0	348.57
<b>186031</b>	ITA	Europe	Italy	2022-01-30	10821375.0	1040184.0	148597.714	145914.0	2618.0	374.00
<b>186032</b>	ITA	Europe	Italy	2022-01-31	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186033</b>	ITA	Europe	Italy	2022-02-01	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186034</b>	ITA	Europe	Italy	2022-02-02	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186035</b>	ITA	Europe	Italy	2022-02-03	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186036</b>	ITA	Europe	Italy	2022-02-04	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186037</b>	ITA	Europe	Italy	2022-02-05	10821375.0	0.0	148597.714	145914.0	0.0	374.00
<b>186038</b>	ITA	Europe	Italy	2022-02-06	11542793.0	721418.0	103059.714	148542.0	2628.0	375.42
<b>186039</b>	ITA	Europe	Italy	2022-02-07	11542793.0	0.0	103059.714	148542.0	0.0	375.42
<b>186040</b>	ITA	Europe	Italy	2022-02-08	11542793.0	0.0	103059.714	148542.0	0.0	375.42
<b>186041</b>	ITA	Europe	Italy	2022-02-09	11542793.0	0.0	103059.714	148542.0	0.0	375.42
<b>186042</b>	ITA	Europe	Italy	2022-02-10	11542793.0	0.0	103059.714	148542.0	0.0	375.42
<b>186043</b>	ITA	Europe	Italy	2022-02-11	11542793.0	0.0	103059.714	148542.0	0.0	375.42

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
186044	ITA	Europe	Italy	2022-02-12	11542793.0	0.0	103059.714	148542.0	0.0	375.46
186045	ITA	Europe	Italy	2022-02-13	12053330.0	510537.0	72933.857	150824.0	2282.0	326.00
186046	ITA	Europe	Italy	2022-02-14	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186047	ITA	Europe	Italy	2022-02-15	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186048	ITA	Europe	Italy	2022-02-16	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186049	ITA	Europe	Italy	2022-02-17	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186050	ITA	Europe	Italy	2022-02-18	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186051	ITA	Europe	Italy	2022-02-19	12053330.0	0.0	72933.857	150824.0	0.0	326.00
186052	ITA	Europe	Italy	2022-02-20	12427773.0	374443.0	53491.857	152848.0	2024.0	289.14
186053	ITA	Europe	Italy	2022-02-21	12427773.0	0.0	53491.857	152848.0	0.0	289.14
186054	ITA	Europe	Italy	2022-02-22	12427773.0	0.0	53491.857	152848.0	0.0	289.14
186055	ITA	Europe	Italy	2022-02-23	12427773.0	0.0	53491.857	152848.0	0.0	289.14
186056	ITA	Europe	Italy	2022-02-24	12427773.0	0.0	53491.857	152848.0	0.0	289.14
186057	ITA	Europe	Italy	2022-02-25	12427773.0	0.0	53491.857	152848.0	0.0	289.14

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
186058	ITA	Europe	Italy	2022-02-26	12427773.0	0.0	53491.857	152848.0	0.0	289.14
186059	ITA	Europe	Italy	2022-02-27	12732680.0	304907.0	43558.143	154416.0	1568.0	224.00
186060	ITA	Europe	Italy	2022-02-28	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186061	ITA	Europe	Italy	2022-03-01	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186062	ITA	Europe	Italy	2022-03-02	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186063	ITA	Europe	Italy	2022-03-03	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186064	ITA	Europe	Italy	2022-03-04	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186065	ITA	Europe	Italy	2022-03-05	12732680.0	0.0	43558.143	154416.0	0.0	224.00
186066	ITA	Europe	Italy	2022-03-06	12990223.0	257543.0	36791.857	155782.0	1366.0	195.14
186067	ITA	Europe	Italy	2022-03-07	12990223.0	0.0	36791.857	155782.0	0.0	195.14
186068	ITA	Europe	Italy	2022-03-08	12990223.0	0.0	36791.857	155782.0	0.0	195.14
186069	ITA	Europe	Italy	2022-03-09	12990223.0	0.0	36791.857	155782.0	0.0	195.14
186070	ITA	Europe	Italy	2022-03-10	12990223.0	0.0	36791.857	155782.0	0.0	195.14
186071	ITA	Europe	Italy	2022-03-11	12990223.0	0.0	36791.857	155782.0	0.0	195.14



	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186072</b>	ITA	Europe	Italy	2022-03-12	12990223.0	0.0	36791.857	155782.0	0.0	195.14
<b>186073</b>	ITA	Europe	Italy	2022-03-13	13323128.0	332905.0	47557.857	156782.0	1000.0	142.85
<b>186074</b>	ITA	Europe	Italy	2022-03-14	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186075</b>	ITA	Europe	Italy	2022-03-15	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186076</b>	ITA	Europe	Italy	2022-03-16	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186077</b>	ITA	Europe	Italy	2022-03-17	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186078</b>	ITA	Europe	Italy	2022-03-18	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186079</b>	ITA	Europe	Italy	2022-03-19	13323128.0	0.0	47557.857	156782.0	0.0	142.85
<b>186080</b>	ITA	Europe	Italy	2022-03-20	13800179.0	477051.0	68150.143	157692.0	910.0	130.00
<b>186081</b>	ITA	Europe	Italy	2022-03-21	13800179.0	0.0	68150.143	157692.0	0.0	130.00

80 rows × 67 columns

VALUTAZIONI: AVENDO ANALIZZATO LE DUE COLONNE TOTAL\_CASES E NEW\_CASES, OPTO PER UTILIZZARE LA COLONNA NEW\_CASES PER I MIEI CALCOLI

In [89]: `df_covid_new.info()`

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 402910 entries, 0 to 429434
```

```
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	iso_code	402910 non-null	object
1	continent	402910 non-null	object
2	location	402910 non-null	object
3	date	402910 non-null	datetime64[ns]
4	total_cases	391716 non-null	float64
5	new_cases	390071 non-null	float64
6	new_cases_smoothed	388901 non-null	float64
7	total_deaths	391716 non-null	float64
8	new_deaths	390520 non-null	float64
9	new_deaths_smoothed	389350 non-null	float64
10	total_cases_per_million	391716 non-null	float64
11	new_cases_per_million	390071 non-null	float64
12	new_cases_smoothed_per_million	388901 non-null	float64
13	total_deaths_per_million	391716 non-null	float64
14	new_deaths_per_million	390520 non-null	float64
15	new_deaths_smoothed_per_million	389350 non-null	float64
16	reproduction_rate	183741 non-null	float64
17	icu_patients	39116 non-null	float64
18	icu_patients_per_million	39116 non-null	float64
19	hosp_patients	40656 non-null	float64
20	hosp_patients_per_million	40656 non-null	float64
21	weekly_icu_admissions	10993 non-null	float64
22	weekly_icu_admissions_per_million	10993 non-null	float64
23	weekly_hosp_admissions	24497 non-null	float64
24	weekly_hosp_admissions_per_million	24497 non-null	float64
25	total_tests	79387 non-null	float64
26	new_tests	75403 non-null	float64
27	total_tests_per_thousand	79387 non-null	float64
28	new_tests_per_thousand	75403 non-null	float64
29	new_tests_smoothed	103965 non-null	float64
30	new_tests_smoothed_per_thousand	103965 non-null	float64
31	positive_rate	95927 non-null	float64
32	tests_per_case	94348 non-null	float64
33	tests_units	106788 non-null	object
34	total_vaccinations	70150 non-null	float64
35	people_vaccinated	65865 non-null	float64

```

36 people_fully_vaccinated      62980 non-null float64
37 total_boosters               39539 non-null float64
38 new_vaccinations             55747 non-null float64
39 new_vaccinations_smoothed    179805 non-null float64
40 total_vaccinations_per_hundred 70150 non-null float64
41 people_vaccinated_per_hundred 65865 non-null float64
42 people_fully_vaccinated_per_hundred 62980 non-null float64
43 total_boosters_per_hundred    39539 non-null float64
44 new_vaccinations_smoothed_per_million 179805 non-null float64
45 new_people_vaccinated_smoothed 176953 non-null float64
46 new_people_vaccinated_smoothed_per_hundred 176953 non-null float64
47 stringency_index             196190 non-null float64
48 population_density           358808 non-null float64
49 median_age                   332979 non-null float64
50 aged_65_older                321586 non-null float64
51 aged_70_older                329631 non-null float64
52 gdp_per_capita                326608 non-null float64
53 extreme_poverty              210312 non-null float64
54 cardiovascular_death_rate    327181 non-null float64
55 diabetes_prevalence          344227 non-null float64
56 female_smokers                245481 non-null float64
57 male_smokers                  242133 non-null float64
58 handwashing_facilities       160057 non-null float64
59 hospital_beds_per_thousand   289005 non-null float64
60 life_expectancy              388615 non-null float64
61 human_development_index      317443 non-null float64
62 population                   402910 non-null int64
63 excess_mortality_cumulative_absolute 13411 non-null float64
64 excess_mortality_cumulative   13411 non-null float64
65 excess_mortality             13411 non-null float64
66 excess_mortality_cumulative_per_million 13411 non-null float64
dtypes: datetime64[ns](1), float64(61), int64(1), object(4)
memory usage: 209.0+ MB

```

## TRACCIA

In [91]: Traccia

Out[91]:



## Analisi Diffusione COVID-19

Il committente richiede di avere un report su casi e vaccinazioni in diverse aree del mondo; a tal fine, richiede di utilizzare il dataset, curato da Our World in Data, all'indirizzo <https://github.com/owid/covid-19-data/tree/master/public/data> alla voce "📂 Download our complete COVID-19 dataset" scaricare il dataset nel formato che si preferisce.

1. Si richiede di verificare le dimensioni del dataset e i relativi metadati
2. Si chiede di trovare, per ogni continente:
  - a. il numero di casi fin dall'inizio della pandemia
  - b. la percentuale rispetto al totale mondiale del numero di casi
3. Selezionare i dati relativi all'Italia nel 2022 e, poiché i nuovi casi vengono registrati settimanalmente, filtrare via i giorni che non hanno misurazioni; quindi mostrare con dei grafici adeguati:
  - a. l'evoluzione dei casi totali dall'inizio alla fine dell'anno
  - b. il numero di nuovi casi rispetto alla data
4. Riguardo le nazioni di Italia, Germania e Francia:
  - a. mostrare in un *boxplot* la differenza tra queste nazioni riguardo il numero di pazienti in terapia intensiva (Intensive Care Unit, ICU, considerare quindi la colonna `icu_patients`) da maggio 2022 (incluso) ad aprile 2023 (incluso)
  - b. scrivere un breve commento (una o due righe) riguardo che conclusioni possiamo trarre osservando il grafico risultante
5. Riguardo le nazioni di Italia, Germania, Francia e Spagna in tutto il 2021:
  - a. mostrare, in maniera grafica oppure numerica, la somma dei pazienti ospitalizzati per ognuna (colonna `hosp_patients`)
  - b. se ci sono dati nulli, con un breve commento scrivere se può essere possibile gestirli tramite sostituzione o meno

```
In [92]: # Inizio con la verifica delle dimensioni del dataset
df_covid_new
```

Out[92]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

402910 rows × 67 columns



In [93]:

```
# veifico i metadati
df_covid_new.describe()
```

Out[93]:

	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_ca
count	402910	3.917160e+05	3.900710e+05	3.889010e+05	3.917160e+05	390520.000000	389350.000000	
mean	2022-04-18 09:33:33.108634368	1.827082e+06	1.989215e+03	1.995095e+03	2.044839e+04	18.080989	18.133540	
min	2020-01-01 00:00:00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.000000	
25%	2021-02-28 00:00:00	5.516000e+03	0.000000e+00	0.000000e+00	3.700000e+01	0.000000	0.000000	
50%	2022-04-17 00:00:00	5.093700e+04	0.000000e+00	8.714000e+00	6.580000e+02	0.000000	0.000000	
75%	2023-06-07 00:00:00	5.610120e+05	0.000000e+00	2.040000e+02	7.378000e+03	0.000000	2.000000	
max	2024-08-14 00:00:00	1.034368e+08	4.047548e+07	5.782211e+06	1.193165e+06	47687.000000	6812.429000	
std	NaN	7.857184e+06	8.585827e+04	3.244760e+04	8.262329e+04	314.885156	118.004772	

8 rows × 63 columns



In [94]:

```
# verifico le colonne del dataset
df_covid_new.columns
```

```
Out[94]: Index(['iso_code', 'continent', 'location', 'date', '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', 'new_deaths_per_million',
               'new_deaths_smoothed_per_million', 'reproduction_rate', 'icu_patients',
               'icu_patients_per_million', 'hosp_patients',
               'hosp_patients_per_million', 'weekly_icu_admissions',
               'weekly_icu_admissions_per_million', 'weekly_hosp_admissions',
               'weekly_hosp_admissions_per_million', 'total_tests', 'new_tests',
               'total_tests_per_thousand', 'new_tests_per_thousand',
               'new_tests_smoothed', 'new_tests_smoothed_per_thousand',
               'positive_rate', 'tests_per_case', 'tests_units', 'total_vaccinations',
               'people_vaccinated', 'people_fully_vaccinated', 'total_boosters',
               'new_vaccinations', 'new_vaccinations_smoothed',
               'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred',
               'people_fully_vaccinated_per_hundred', 'total_boosters_per_hundred',
               'new_vaccinations_smoothed_per_million',
               'new_people_vaccinated_smoothed',
               'new_people_vaccinated_smoothed_per_hundred', 'stringency_index',
               'population_density', 'median_age', 'aged_65_old', 'aged_70_old',
               'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate',
               'diabetes_prevalence', 'female_smokers', 'male_smokers',
               'handwashing_facilities', 'hospital_beds_per_thousand',
               'life_expectancy', 'human_development_index', 'population',
               'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative',
               'excess_mortality', 'excess_mortality_cumulative_per_million'],
              dtype='object')
```

```
In [95]: print(len(df_covid_new.columns))
```

67

```
In [96]: df_col_utili= df_covid_new[["date", "continent", "location", "total_cases", "new_cases", "icu_patients", "hosp_patients"]]
df_col_utili
```

Out[96]:

	date	continent	location	total_cases	new_cases	icu_patients	hosp_patients
0	2020-01-05	Asia	Afghanistan	0.0	0.0	NaN	NaN
1	2020-01-06	Asia	Afghanistan	0.0	0.0	NaN	NaN
2	2020-01-07	Asia	Afghanistan	0.0	0.0	NaN	NaN
3	2020-01-08	Asia	Afghanistan	0.0	0.0	NaN	NaN
4	2020-01-09	Asia	Afghanistan	0.0	0.0	NaN	NaN
...	...	...	...	...	...	...	...
429430	2024-07-31	Africa	Zimbabwe	266386.0	0.0	NaN	NaN
429431	2024-08-01	Africa	Zimbabwe	266386.0	0.0	NaN	NaN
429432	2024-08-02	Africa	Zimbabwe	266386.0	0.0	NaN	NaN
429433	2024-08-03	Africa	Zimbabwe	266386.0	0.0	NaN	NaN
429434	2024-08-04	Africa	Zimbabwe	266386.0	0.0	NaN	NaN

402910 rows × 7 columns

In [97]: `print(len(df_col_utili.columns))`

7

In [98]: `df_nonull_newcases = df_covid_new.dropna(subset=['new_cases'])  
df_nonull_newcases`



Out[98]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
0	AFG	Asia	Afghanistan	2020-01-05	0.0	0.0	NaN	0.0	0.0	
1	AFG	Asia	Afghanistan	2020-01-06	0.0	0.0	NaN	0.0	0.0	
2	AFG	Asia	Afghanistan	2020-01-07	0.0	0.0	NaN	0.0	0.0	
3	AFG	Asia	Afghanistan	2020-01-08	0.0	0.0	NaN	0.0	0.0	
4	AFG	Asia	Afghanistan	2020-01-09	0.0	0.0	NaN	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	
429430	ZWE	Africa	Zimbabwe	2024-07-31	266386.0	0.0	0.0	5740.0	0.0	
429431	ZWE	Africa	Zimbabwe	2024-08-01	266386.0	0.0	0.0	5740.0	0.0	
429432	ZWE	Africa	Zimbabwe	2024-08-02	266386.0	0.0	0.0	5740.0	0.0	
429433	ZWE	Africa	Zimbabwe	2024-08-03	266386.0	0.0	0.0	5740.0	0.0	
429434	ZWE	Africa	Zimbabwe	2024-08-04	266386.0	0.0	0.0	5740.0	0.0	

390071 rows × 67 columns



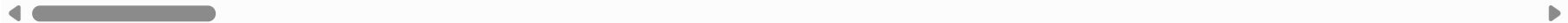
In [99]:

```
df_nonzerovalues_newcases = df_nonull_newcases[df_nonull_newcases['new_cases'] != 0]
df_nonzerovalues_newcases
```

Out[99]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
56	AFG	Asia	Afghanistan	2020-03-01	1.0	1.0	0.143	0.0	0.0	
70	AFG	Asia	Afghanistan	2020-03-15	7.0	6.0	0.857	0.0	0.0	
77	AFG	Asia	Afghanistan	2020-03-22	24.0	17.0	2.429	0.0	0.0	
84	AFG	Asia	Afghanistan	2020-03-29	91.0	67.0	9.571	2.0	2.0	
91	AFG	Asia	Afghanistan	2020-04-05	274.0	183.0	26.143	5.0	3.0	
...	...	...	...	...	...	...	...	...	...	
429385	ZWE	Africa	Zimbabwe	2024-06-16	266374.0	9.0	1.286	5740.0	0.0	
429392	ZWE	Africa	Zimbabwe	2024-06-23	266378.0	4.0	0.571	5740.0	0.0	
429399	ZWE	Africa	Zimbabwe	2024-06-30	266384.0	6.0	0.857	5740.0	0.0	
429406	ZWE	Africa	Zimbabwe	2024-07-07	266385.0	1.0	0.143	5740.0	0.0	
429420	ZWE	Africa	Zimbabwe	2024-07-21	266386.0	1.0	0.143	5740.0	0.0	

38929 rows × 67 columns



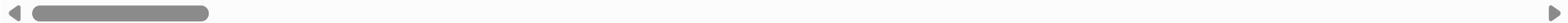
In [100...

```
df_nonzerovalues_newcases.describe()
```

Out[100...

	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cas
count	38929	3.892900e+04	3.892900e+04	3.892700e+04	3.892900e+04	38919.000000	38917.000000	
mean	2022-02-05 08:26:48.343394304	2.012611e+06	1.993206e+04	2.847583e+03	2.348808e+04	179.165986	25.596469	
min	2020-01-05 00:00:00	1.000000e+00	1.000000e+00	1.430000e-01	0.000000e+00	0.000000	0.000000	
25%	2021-02-21 00:00:00	1.150500e+04	4.300000e+01	6.143000e+00	1.230000e+02	0.000000	0.000000	
50%	2022-01-16 00:00:00	1.073230e+05	3.950000e+02	5.642900e+01	1.388000e+03	3.000000	0.429000	
75%	2022-12-11 00:00:00	8.588480e+05	3.997000e+03	5.710000e+02	1.070500e+04	41.000000	5.857000	
max	2024-08-04 00:00:00	1.034368e+08	4.047548e+07	5.782211e+06	1.127152e+06	47687.000000	6812.429000	
std	NaN	7.816414e+06	2.711243e+05	3.873303e+04	8.068520e+04	980.954972	140.139903	

8 rows × 63 columns



In [238...

```
# Qui produco il numero di casi per continente e la percentuale rispetto al totale mondiale
continent_cases = df_nonzerovalues_newcases.groupby('continent')['new_cases'].sum()
total_world_cases = continent_cases.sum()
continent_percent = (continent_cases / total_world_cases) * 100
continent_percent = continent_percent.round(2)

print("Numero di casi per continente:")
print(continent_cases)
print("\nPercentuale rispetto al totale mondiale:")
print(continent_percent)
print(total_world_cases)
```

Numero di casi per continente:

continent

Africa 13146831.0

Asia 301564180.0

Europe 252916868.0

North America 124492698.0

Oceania 15003468.0

South America 68811012.0

Name: new\_cases, dtype: float64

Percentuale rispetto al totale mondiale:

continent

Africa 1.69

Asia 38.86

Europe 32.60

North America 16.04

Oceania 1.93

South America 8.87

Name: new\_cases, dtype: float64

775935057.0

In [102...

```
df_italia_2022 = df_nonzerovalues_newcases [(df_nonzerovalues_newcases['location'] == 'Italy') & (df_nonzerovalues_newcases['d
df_italia_2022 # Qui seleziono i dati relativi all'Italia nel 2022
```

Out[102...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186003</b>	ITA	Europe	Italy	2022-01-02	6267035.0	644604.0	92086.286	137513.0	983.0	140.42
<b>186010</b>	ITA	Europe	Italy	2022-01-09	7281297.0	1014262.0	144894.571	138881.0	1368.0	195.42
<b>186017</b>	ITA	Europe	Italy	2022-01-16	8549450.0	1268153.0	181164.714	140856.0	1975.0	282.14
<b>186024</b>	ITA	Europe	Italy	2022-01-23	9781191.0	1231741.0	175963.000	143296.0	2440.0	348.57
<b>186031</b>	ITA	Europe	Italy	2022-01-30	10821375.0	1040184.0	148597.714	145914.0	2618.0	374.00
<b>186038</b>	ITA	Europe	Italy	2022-02-06	11542793.0	721418.0	103059.714	148542.0	2628.0	375.42
<b>186045</b>	ITA	Europe	Italy	2022-02-13	12053330.0	510537.0	72933.857	150824.0	2282.0	326.00
<b>186052</b>	ITA	Europe	Italy	2022-02-20	12427773.0	374443.0	53491.857	152848.0	2024.0	289.14
<b>186059</b>	ITA	Europe	Italy	2022-02-27	12732680.0	304907.0	43558.143	154416.0	1568.0	224.00
<b>186066</b>	ITA	Europe	Italy	2022-03-06	12990223.0	257543.0	36791.857	155782.0	1366.0	195.14
<b>186073</b>	ITA	Europe	Italy	2022-03-13	13323128.0	332905.0	47557.857	156782.0	1000.0	142.85
<b>186080</b>	ITA	Europe	Italy	2022-03-20	13800179.0	477051.0	68150.143	157692.0	910.0	130.00
<b>186087</b>	ITA	Europe	Italy	2022-03-27	14304111.0	503932.0	71990.286	158700.0	1008.0	144.00
<b>186094</b>	ITA	Europe	Italy	2022-04-03	14790806.0	486695.0	69527.857	159666.0	966.0	138.00

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186101</b>	ITA	Europe	Italy	2022-04-10	15238128.0	447322.0	63903.143	160658.0	992.0	141.77
<b>186108</b>	ITA	Europe	Italy	2022-04-17	15659835.0	421707.0	60243.857	161602.0	944.0	134.85
<b>186115</b>	ITA	Europe	Italy	2022-04-24	16079209.0	419374.0	59910.571	162609.0	1007.0	143.85
<b>186122</b>	ITA	Europe	Italy	2022-05-01	16463200.0	383991.0	54855.857	163507.0	898.0	128.28
<b>186129</b>	ITA	Europe	Italy	2022-05-08	16767773.0	304573.0	43510.429	164417.0	910.0	130.00
<b>186136</b>	ITA	Europe	Italy	2022-05-15	17030147.0	262374.0	37482.000	165182.0	765.0	109.28
<b>186143</b>	ITA	Europe	Italy	2022-05-22	17229263.0	199116.0	28445.143	165918.0	736.0	105.14
<b>186150</b>	ITA	Europe	Italy	2022-05-29	17373741.0	144478.0	20639.714	166542.0	624.0	89.14
<b>186157</b>	ITA	Europe	Italy	2022-06-05	17490451.0	116710.0	16672.857	166922.0	380.0	54.28
<b>186164</b>	ITA	Europe	Italy	2022-06-12	17634065.0	143614.0	20516.286	167365.0	443.0	63.28
<b>186171</b>	ITA	Europe	Italy	2022-06-19	17844905.0	210840.0	30120.000	167703.0	338.0	48.28
<b>186178</b>	ITA	Europe	Italy	2022-06-26	18184917.0	340012.0	48573.143	168058.0	355.0	50.71
<b>186185</b>	ITA	Europe	Italy	2022-07-03	18695954.0	511037.0	73005.286	168488.0	430.0	61.43
<b>186192</b>	ITA	Europe	Italy	2022-07-10	19357938.0	661984.0	94569.143	169062.0	574.0	82.00

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186199</b>	ITA	Europe	Italy	2022-07-17	20076863.0	718925.0	102703.571	169846.0	784.0	112.00
<b>186206</b>	ITA	Europe	Italy	2022-07-24	20608190.0	531327.0	75903.857	170798.0	952.0	136.00
<b>186213</b>	ITA	Europe	Italy	2022-07-31	21002773.0	394583.0	56369.000	172003.0	1205.0	172.14
<b>186220</b>	ITA	Europe	Italy	2022-08-07	21286771.0	283998.0	40571.143	173062.0	1059.0	151.28
<b>186227</b>	ITA	Europe	Italy	2022-08-14	21480076.0	193305.0	27615.000	173982.0	920.0	131.42
<b>186234</b>	ITA	Europe	Italy	2022-08-21	21630998.0	150922.0	21560.286	174659.0	677.0	96.77
<b>186241</b>	ITA	Europe	Italy	2022-08-28	21788862.0	157864.0	22552.000	175306.0	647.0	92.42
<b>186248</b>	ITA	Europe	Italy	2022-09-04	21925073.0	136211.0	19458.714	175802.0	496.0	70.89
<b>186255</b>	ITA	Europe	Italy	2022-09-11	22035717.0	110644.0	15806.286	176175.0	373.0	53.28
<b>186262</b>	ITA	Europe	Italy	2022-09-18	22148935.0	113218.0	16174.000	176546.0	371.0	53.00
<b>186269</b>	ITA	Europe	Italy	2022-09-25	22284812.0	135877.0	19411.000	176867.0	321.0	45.89
<b>186276</b>	ITA	Europe	Italy	2022-10-02	22500346.0	215534.0	30790.571	177130.0	263.0	37.55
<b>186283</b>	ITA	Europe	Italy	2022-10-09	22781293.0	280947.0	40135.286	177478.0	348.0	49.77
<b>186290</b>	ITA	Europe	Italy	2022-10-16	23069745.0	288452.0	41207.429	177956.0	478.0	68.28

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>186297</b>	ITA	Europe	Italy	2022-10-23	23322522.0	252777.0	36111.000	178542.0	586.0	83.7
<b>186304</b>	ITA	Europe	Italy	2022-10-30	23531023.0	208501.0	29785.857	179101.0	559.0	79.8
<b>186311</b>	ITA	Europe	Italy	2022-11-06	23697012.0	165989.0	23712.714	179655.0	554.0	79.1
<b>186318</b>	ITA	Europe	Italy	2022-11-13	23878193.0	181181.0	25883.000	180139.0	484.0	69.1
<b>186325</b>	ITA	Europe	Italy	2022-11-20	24099206.0	221013.0	31573.286	180679.0	540.0	77.1
<b>186332</b>	ITA	Europe	Italy	2022-11-27	24327664.0	228458.0	32636.857	181271.0	592.0	84.5
<b>186339</b>	ITA	Europe	Italy	2022-12-04	24555456.0	227792.0	32541.714	181944.0	673.0	96.1
<b>186346</b>	ITA	Europe	Italy	2022-12-11	24753889.0	198433.0	28347.571	182619.0	675.0	96.4
<b>186353</b>	ITA	Europe	Italy	2022-12-18	24928076.0	174187.0	24883.857	183370.0	751.0	107.2
<b>186360</b>	ITA	Europe	Italy	2022-12-25	25060503.0	132427.0	18918.143	184168.0	798.0	114.0

52 rows × 67 columns

Qui salvo i colori per servirmene in seguito e stabilire varianti di viola più scure da utilizzare nei grafici sottostanti

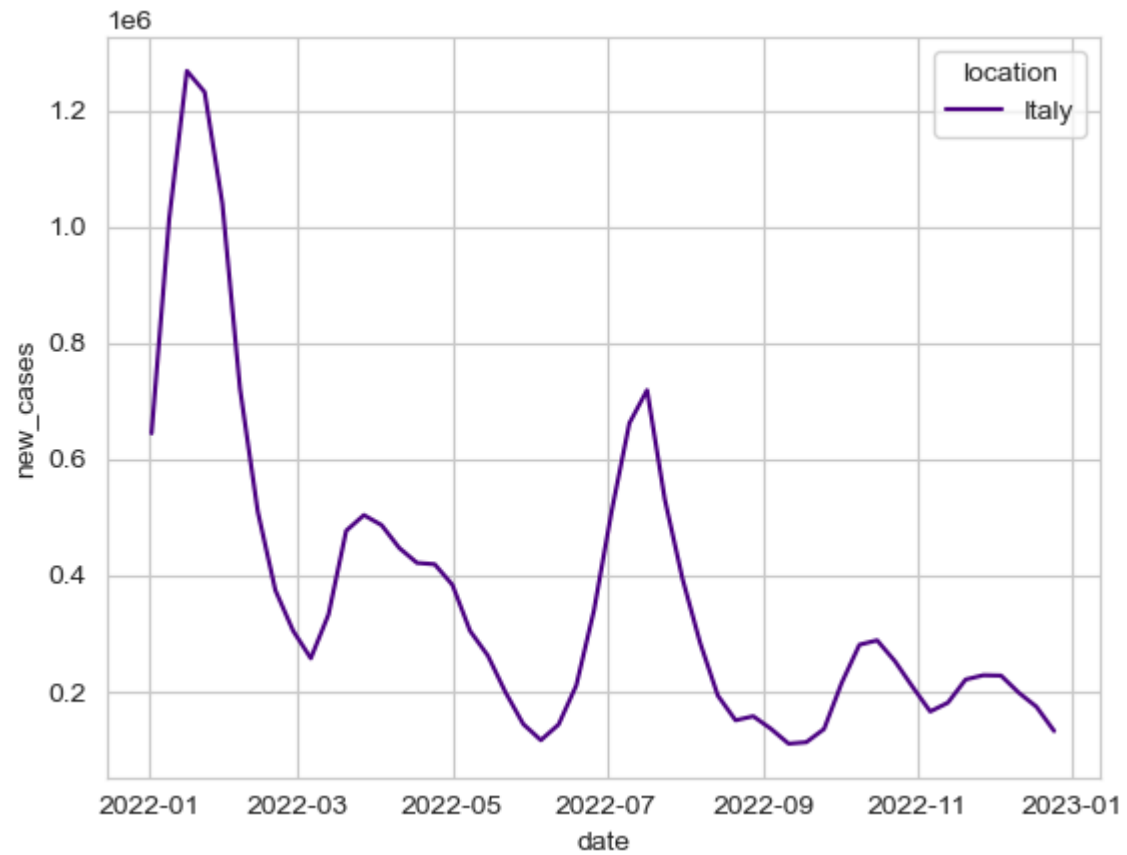
(["#4B0082", "#6A0DAD", "#800080", "#551A8B", "#301934"])

In [275...

```
# Crea il grafico con una linea viola scura
sns.lineplot(data = df_italia_2022, x = 'date', y = 'new_cases', hue = 'location', color = "#4B0082")
```

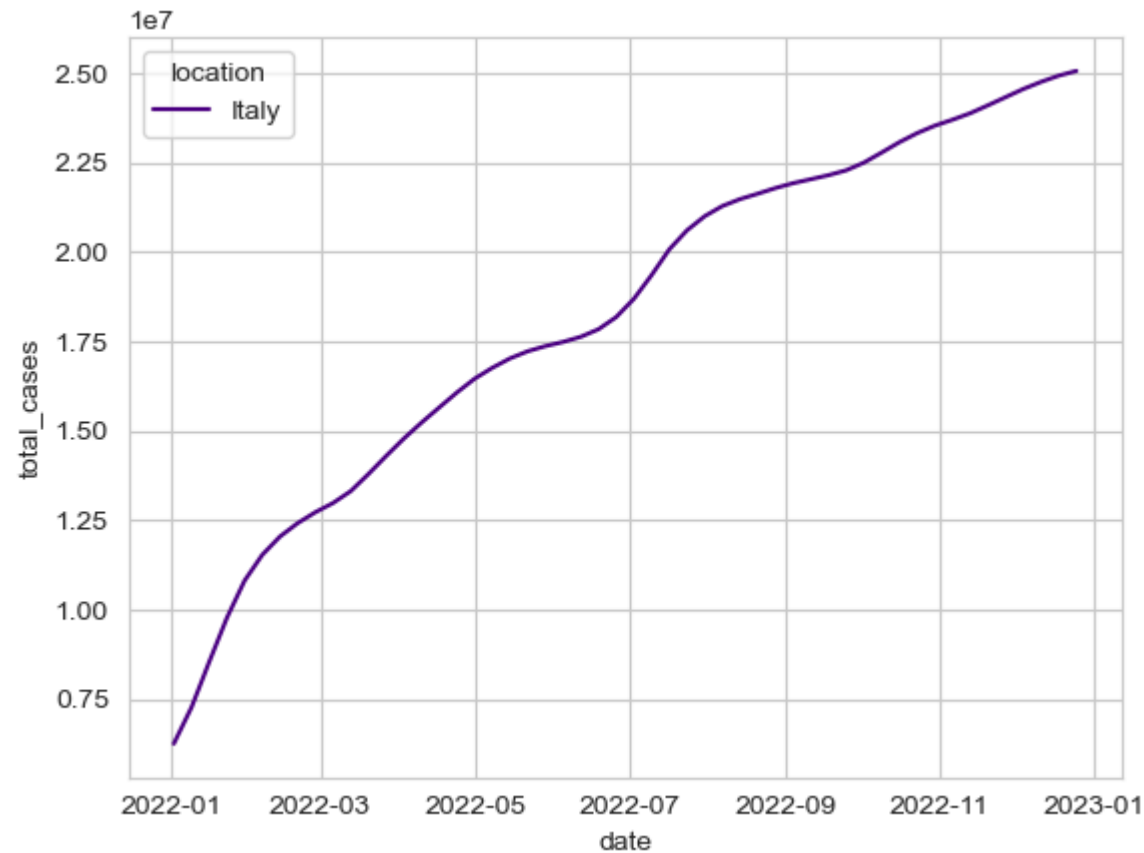


Out[275... <Axes: xlabel='date', ylabel='new\_cases'>



```
In [279... sns.lineplot(data = df_italia_2022, x = 'date', y = 'total_cases', hue = 'location', color = "#800080")
```

Out[279... <Axes: xlabel='date', ylabel='total\_cases'>

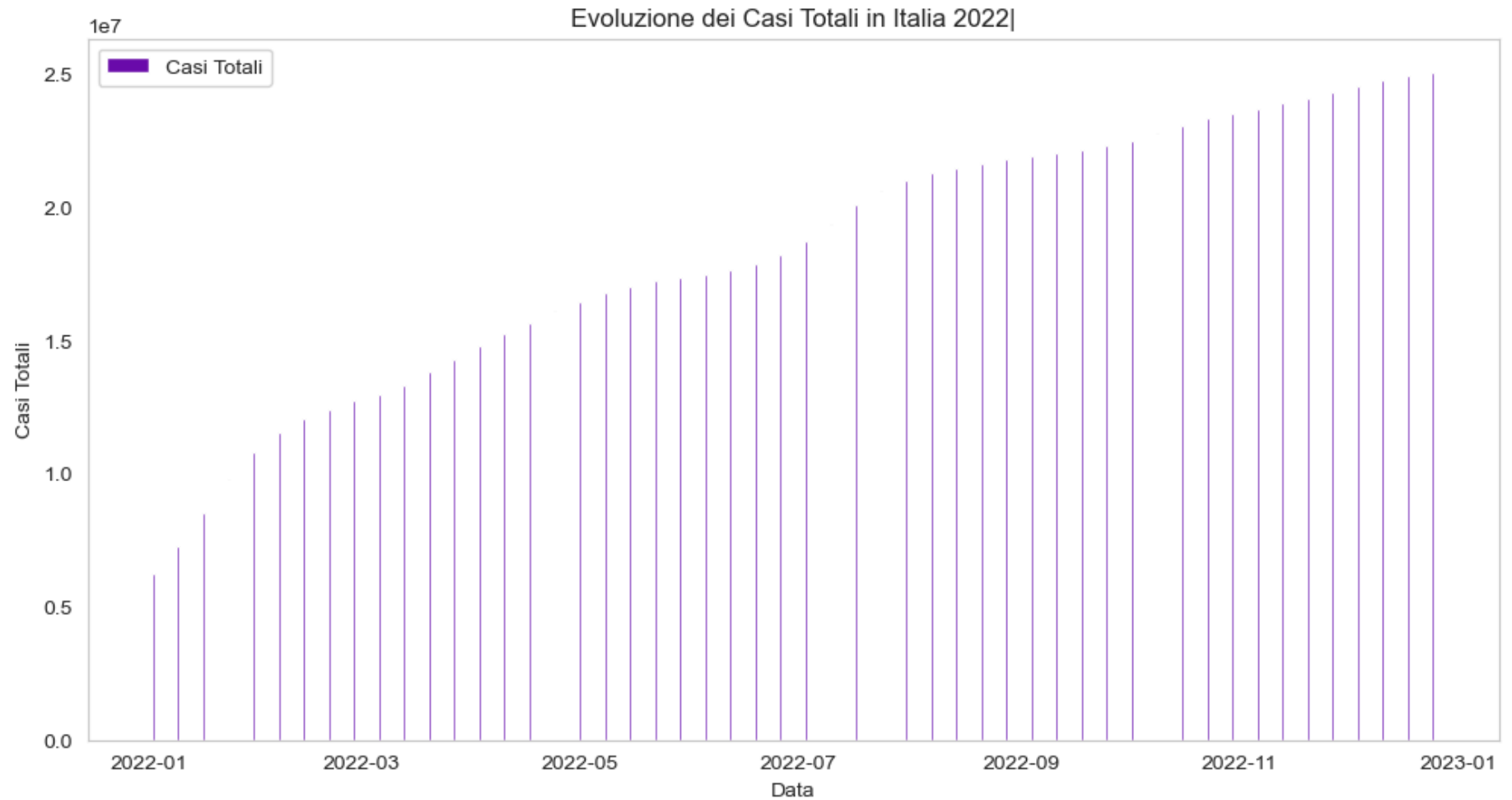


```
In [295... # dati relativi all'Italia nel 2022
df_italia_2022 = italy_data.dropna(subset=["new_cases"])

italy_data["date"] = pd.to_datetime(italy_data["date"])

#Grafico dell'evoluzione dei casi totali in Italia nel 2022
plt.figure(figsize=(12, 6))
plt.bar(italy_data["date"], italy_data["total_cases"], label="Casi Totali", color= "#6A0DAD")
plt.xlabel("Data")
plt.ylabel("Casi Totali")
plt.title("Evoluzione dei Casi Totali in Italia 2022|")
plt.legend()
```

```
plt.grid()  
plt.show()
```



```
In [106... icu_data = df_nonzerovalues_newcases[(df_nonzerovalues_newcases['location'].isin(['Italy', 'Germany', 'France'])) &  
(df_nonzerovalues_newcases['date'] >= '2022-05-01') & (df_nonzerovalues_newcases['date'] <= '2023-04-30')]  
icu_data
```

Out[106...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
131214	FRA	Europe	France	2022-05-01	27742592.0	376510.0	53787.143	146744.0	867.0	123.89
131221	FRA	Europe	France	2022-05-08	28007520.0	264928.0	37846.857	147493.0	749.0	107.00
131228	FRA	Europe	France	2022-05-15	28224235.0	216715.0	30959.286	148107.0	614.0	87.77
131235	FRA	Europe	France	2022-05-22	28386474.0	162239.0	23177.000	148573.0	466.0	66.55
131242	FRA	Europe	France	2022-05-29	28499829.0	113355.0	16193.571	148873.0	300.0	42.89
...	...	...	...	...	...	...	...	...	...	...
186458	ITA	Europe	Italy	2023-04-02	25701147.0	20999.0	2999.857	189155.0	181.0	25.89
186465	ITA	Europe	Italy	2023-04-09	25721054.0	19907.0	2843.857	189303.0	148.0	21.14
186472	ITA	Europe	Italy	2023-04-16	25745800.0	24746.0	3535.143	189440.0	137.0	19.55
186479	ITA	Europe	Italy	2023-04-23	25772459.0	26659.0	3808.429	189627.0	187.0	26.77
186486	ITA	Europe	Italy	2023-04-30	25795509.0	23050.0	3292.857	189786.0	159.0	22.77

159 rows × 67 columns

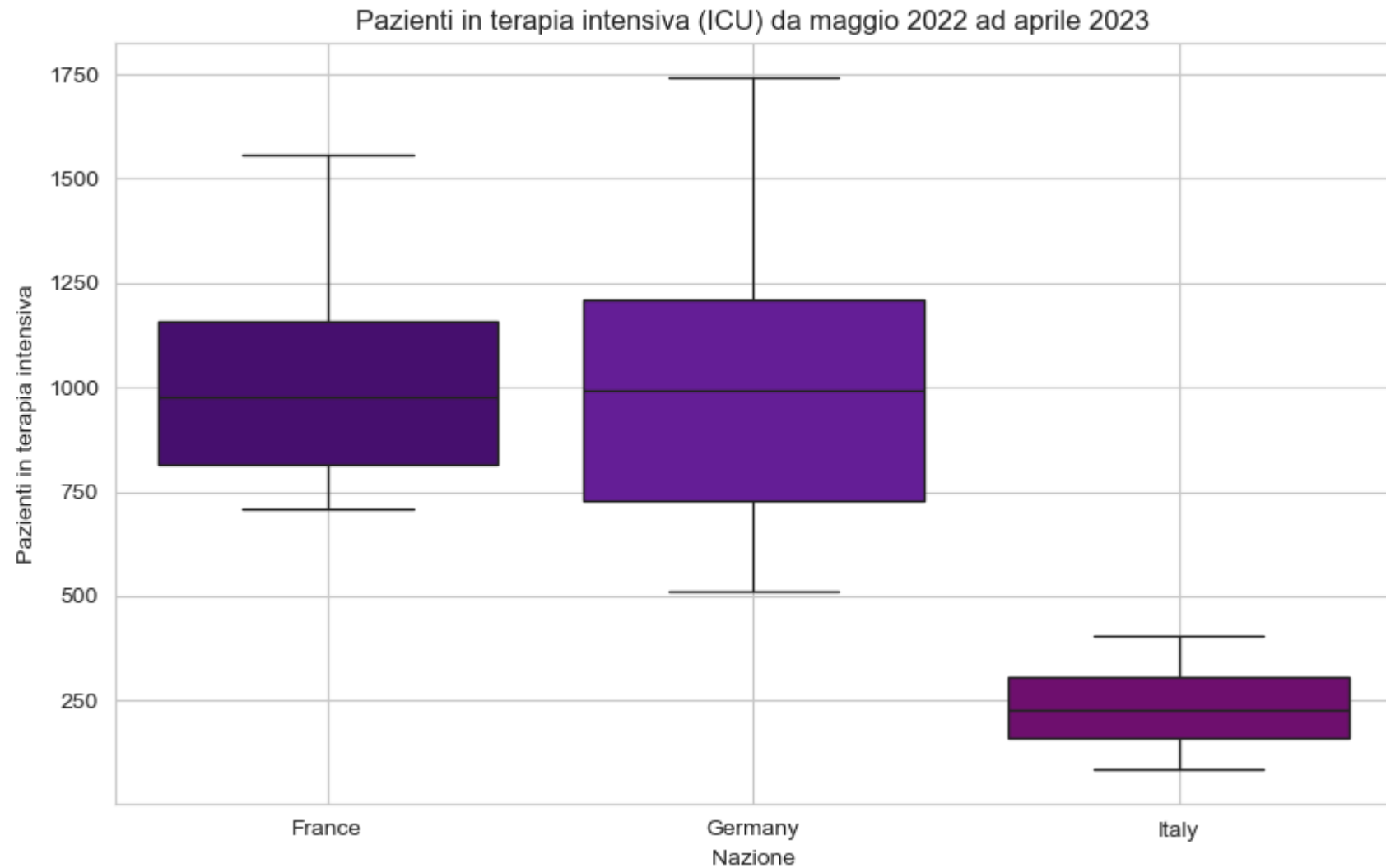
◀

▶

```
# Qui ho definito una variabile di colori viola più scuri
dark_purple_palette = ["#4B0082", "#6A0DAD", "#800080"]
```

```
In [301... # Creare un boxplot per mostrare la differenza nel numero di pazienti in terapia intensiva
plt.figure(figsize=(10, 6))
sns.boxplot(x='location', y='icu_patients', data=icu_data, palette=dark_purple_palette, hue = "location")
plt.title('Pazienti in terapia intensiva (ICU) da maggio 2022 ad aprile 2023')
plt.xlabel('Nazione')
plt.ylabel('Pazienti in terapia intensiva')
plt.grid(True)
plt.show()

print("Commento: Il boxplot mostra la distribuzione dei pazienti in terapia intensiva per Italia, Germania e Francia. \
Possiamo osservare che l'Italia ha un valore più basso di ICU avendo provveduto per prima alla chiusura del paese - lockdown")
```



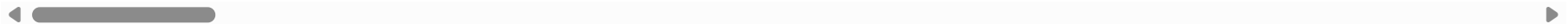
Commento: Il boxplot mostra la distribuzione dei pazienti in terapia intensiva per Italia, Germania e Francia. Possiamo osservare che la mediana dei pazienti in terapia intensiva varia tra le nazioni, con alcune nazioni che hanno una maggiore variabilità rispetto ad altre.

In [314... `df_nonzerovalues_newcases`

Out[314...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoc
56	AFG	Asia	Afghanistan	2020-03-01	1.0	1.0	0.143	0.0	0.0	
70	AFG	Asia	Afghanistan	2020-03-15	7.0	6.0	0.857	0.0	0.0	
77	AFG	Asia	Afghanistan	2020-03-22	24.0	17.0	2.429	0.0	0.0	
84	AFG	Asia	Afghanistan	2020-03-29	91.0	67.0	9.571	2.0	2.0	
91	AFG	Asia	Afghanistan	2020-04-05	274.0	183.0	26.143	5.0	3.0	
...	...	...	...	...	...	...	...	...	...	
429385	ZWE	Africa	Zimbabwe	2024-06-16	266374.0	9.0	1.286	5740.0	0.0	
429392	ZWE	Africa	Zimbabwe	2024-06-23	266378.0	4.0	0.571	5740.0	0.0	
429399	ZWE	Africa	Zimbabwe	2024-06-30	266384.0	6.0	0.857	5740.0	0.0	
429406	ZWE	Africa	Zimbabwe	2024-07-07	266385.0	1.0	0.143	5740.0	0.0	
429420	ZWE	Africa	Zimbabwe	2024-07-21	266386.0	1.0	0.143	5740.0	0.0	

38929 rows × 67 columns



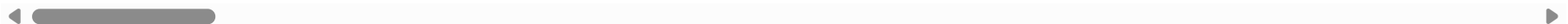
In [108...

```
# Filtra i dati per Italia, Germania, Francia e Spagna nel 2021
hosp_data_2021 = df_nonzerovalues_newcases.loc[(df_nonzerovalues_newcases['location'].isin(['Italy', 'Germany', 'France', 'Spain']) && (df_nonzerovalues_newcases['date'].dt.year == 2021))
hosp_data_2021
```

Out[108...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>130731</b>	FRA	Europe	France	2021-01-03	2427206.0	88948.0	12706.857	66449.0	2915.0	416.42
<b>130738</b>	FRA	Europe	France	2021-01-10	2549469.0	122263.0	17466.143	69878.0	3429.0	489.85
<b>130745</b>	FRA	Europe	France	2021-01-17	2668311.0	118842.0	16977.429	73332.0	3454.0	493.42
<b>130752</b>	FRA	Europe	France	2021-01-24	2798460.0	130149.0	18592.714	76746.0	3414.0	487.71
<b>130759</b>	FRA	Europe	France	2021-01-31	2931084.0	132624.0	18946.286	79905.0	3159.0	451.28
...	...	...	...	...	...	...	...	...	...	...
<b>363279</b>	ESP	Europe	Spain	2021-11-28	5171584.0	59198.0	8456.857	89540.0	194.0	27.71
<b>363286</b>	ESP	Europe	Spain	2021-12-05	5261477.0	89893.0	12841.857	89835.0	295.0	42.14
<b>363293</b>	ESP	Europe	Spain	2021-12-12	5377493.0	116016.0	16573.714	90190.0	355.0	50.71
<b>363300</b>	ESP	Europe	Spain	2021-12-19	5609092.0	231599.0	33085.571	90691.0	501.0	71.51
<b>363307</b>	ESP	Europe	Spain	2021-12-26	6100138.0	491046.0	70149.429	91275.0	584.0	83.42

208 rows × 67 columns



In [318...

```
# Calcola la somma dei pazienti ospitalizzati per ognuna delle nazioni
hosp_sum = hosp_data_2021.groupby('location')['hosp_patients'].sum()
print("Somma dei pazienti ospitalizzati nel 2021:")
print(hosp_sum)
```



Somma dei pazienti ospitalizzati nel 2021:

location

France 854767.0

Germany 0.0

Italy 626339.0

Spain 357643.0

Name: hosp\_patients, dtype: float64

In [110...

hosp\_data\_2021

Out[110...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>130731</b>	FRA	Europe	France	2021-01-03	2427206.0	88948.0	12706.857	66449.0	2915.0	416.42
<b>130738</b>	FRA	Europe	France	2021-01-10	2549469.0	122263.0	17466.143	69878.0	3429.0	489.85
<b>130745</b>	FRA	Europe	France	2021-01-17	2668311.0	118842.0	16977.429	73332.0	3454.0	493.42
<b>130752</b>	FRA	Europe	France	2021-01-24	2798460.0	130149.0	18592.714	76746.0	3414.0	487.77
<b>130759</b>	FRA	Europe	France	2021-01-31	2931084.0	132624.0	18946.286	79905.0	3159.0	451.28
...	...	...	...	...	...	...	...	...	...	...
<b>363279</b>	ESP	Europe	Spain	2021-11-28	5171584.0	59198.0	8456.857	89540.0	194.0	27.77
<b>363286</b>	ESP	Europe	Spain	2021-12-05	5261477.0	89893.0	12841.857	89835.0	295.0	42.14
<b>363293</b>	ESP	Europe	Spain	2021-12-12	5377493.0	116016.0	16573.714	90190.0	355.0	50.77
<b>363300</b>	ESP	Europe	Spain	2021-12-19	5609092.0	231599.0	33085.571	90691.0	501.0	71.55
<b>363307</b>	ESP	Europe	Spain	2021-12-26	6100138.0	491046.0	70149.429	91275.0	584.0	83.42

208 rows × 67 columns



In [311...

```
# Verifico la presenza di null e stampo un commento in merito
null_values = hosp_data_2021['hosp_patients'].isnull().sum()
if null_values > 0:
    print(f"\nCi sono {null_values} valori nulli nella colonna 'hosp_patients'.")
    print("Possiamo considerare la sostituzione con l'interpolazione, con la media o la mediana dei valori disponibili. Io ho
```

```
else:  
    print("\n Non ci sono valori nulli nella colonna 'hosp_patients'.")
```

Ci sono 52 valori nulli nella colonna 'hosp\_patients'.

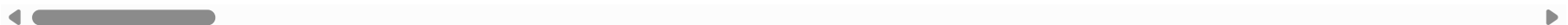
Possiamo considerare la sostituzione con l'interpolazione, con la media o la mediana dei valori disponibili. Io ho deciso di non rimuoverli in questo caso, in quanto i valori null appartengono interamente alla Germania, la quale non ha valori di pazienti ospedalizzati. Mantenendo la presenza dei null si nota subito l'assenza di valori per la Germania

```
In [307... # Seleziono i dati per Italia, Germania, Francia e Spagna nel 2021 - hosp_data2_2021  
# sulla base del df_covid_new al quale non ho rimosso alcun dato, al contrario di df_nonzerovalues_newcases  
hosp_data2_2021 = df_covid_new.loc[(df_covid_new['location'].isin(['Italy', 'Germany', 'France', 'Spain'])) &  
(df_covid_new['date'].dt.year == 2021)]  
hosp_data2_2021
```

Out[307...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
<b>130729</b>	FRA	Europe	France	2021-01-01	2338258.0	0.0	11984.000	63534.0	0.0	445.57
<b>130730</b>	FRA	Europe	France	2021-01-02	2338258.0	0.0	11984.000	63534.0	0.0	445.57
<b>130731</b>	FRA	Europe	France	2021-01-03	2427206.0	88948.0	12706.857	66449.0	2915.0	416.42
<b>130732</b>	FRA	Europe	France	2021-01-04	2427206.0	0.0	12706.857	66449.0	0.0	416.42
<b>130733</b>	FRA	Europe	France	2021-01-05	2427206.0	0.0	12706.857	66449.0	0.0	416.42
...	...	...	...	...	...	...	...	...	...	...
<b>363308</b>	ESP	Europe	Spain	2021-12-27	6100138.0	0.0	70149.429	91275.0	0.0	83.42
<b>363309</b>	ESP	Europe	Spain	2021-12-28	6100138.0	0.0	70149.429	91275.0	0.0	83.42
<b>363310</b>	ESP	Europe	Spain	2021-12-29	6100138.0	0.0	70149.429	91275.0	0.0	83.42
<b>363311</b>	ESP	Europe	Spain	2021-12-30	6100138.0	0.0	70149.429	91275.0	0.0	83.42
<b>363312</b>	ESP	Europe	Spain	2021-12-31	6100138.0	0.0	70149.429	91275.0	0.0	83.42

1460 rows × 67 columns



In [324...

```
# Calcola la somma dei pazienti ospitalizzati per ognuna delle nazioni per il 2021 - con il data2 (quello senza rimozione degli
hosp_sum2 = hosp_data2_2021.groupby('location')['hosp_patients'].sum()
print("Somma dei pazienti ospitalizzati nel 2021:")
```

```
print(hosp_sum)
# Ritengo che l'output sia errato, invece quello che ritengo essere corretto è l'output con l'utilizzo del hosp_data_2021 (tre
```

Somma dei pazienti ospitalizzati nel 2021:

location

France 6008717.0

Germany 0.0

Italy 4419950.0

Spain 2411706.0

Name: hosp\_patients, dtype: float64

```
In [337... # Qui ordino per location e date in ordine cronologico
df_covidtemp_nonzero = df_covidtemp_nonzero.sort_values(by=["location", "date"])

# Qui trovo il primo valore per ogni location
covidtempfirstcasesforlocation = df_covidtemp_nonzero.groupby("location").first()

# Qui visualizzo le colonne rilevanti
print(covidtempfirstcasesforlocation[["date", "new_cases"]])
covidtempfirstcasesforlocation.sample(5)
```

	date	new_cases
location		
Afghanistan	2020-03-01	1.0
Albania	2020-03-15	33.0
Algeria	2020-03-01	1.0
American Samoa	2021-09-19	1.0
Andorra	2020-03-08	1.0
...	...	...
Vietnam	2020-01-26	2.0
Wallis and Futuna	2020-10-25	1.0
Yemen	2020-04-12	1.0
Zambia	2020-03-22	2.0
Zimbabwe	2020-03-22	2.0

[232 rows x 2 columns]

Out[337...

	iso_code	continent	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	tota
<b>location</b>										
<b>Cook Islands</b>	COK	Oceania	2022-02-20	4.0	4.0	0.571	0.0	0.0	0.0	
<b>Mayotte</b>	MYT	Africa	2020-03-15	5.0	5.0	0.714	0.0	0.0	0.0	
<b>Cuba</b>	CUB	North America	2020-03-15	4.0	4.0	0.571	0.0	0.0	0.0	
<b>Barbados</b>	BRB	North America	2020-03-22	5.0	5.0	0.714	0.0	0.0	0.0	
<b>Trinidad and Tobago</b>	TTO	North America	2020-03-15	1.0	1.0	0.143	0.0	0.0	0.0	

5 rows × 66 columns



In [116... `df_nonzerovalues_newcases.to_csv("df_nonzerovalues_newcases.csv")`

In [117... `df_albania = df_nonzerovalues_newcases[df_nonzerovalues_newcases["location"] == "Albania"]`  
`df_albania.sample(10)`

Out[117...

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
4048	ALB	Europe	Albania	2021-12-05	201045.0	2313.0	330.429	3108.0	31.0	4.429
4510	ALB	Europe	Albania	2023-03-12	333651.0	80.0	11.429	3600.0	1.0	0.143
3936	ALB	Europe	Albania	2021-08-15	135140.0	1549.0	221.286	2461.0	3.0	0.429
4468	ALB	Europe	Albania	2023-01-29	333219.0	50.0	7.143	3596.0	0.0	0.000
4111	ALB	Europe	Albania	2022-02-06	263172.0	7431.0	1061.571	3371.0	37.0	5.286
3873	ALB	Europe	Albania	2021-06-13	132437.0	65.0	9.286	2453.0	2.0	0.286
4524	ALB	Europe	Albania	2023-03-26	333799.0	83.0	11.857	3601.0	0.0	0.000
3971	ALB	Europe	Albania	2021-09-19	161324.0	6031.0	861.571	2569.0	34.0	4.857
4433	ALB	Europe	Albania	2022-12-25	332670.0	132.0	18.857	3596.0	2.0	0.286
4209	ALB	Europe	Albania	2022-05-15	275485.0	219.0	31.286	3497.0	1.0	0.143

10 rows × 67 columns



# CONSEGNA

In [119...

Consegna

Out[119...

## Consegna

- Per consegnare, effettuate direttamente l'upload del file `.ipynb` su Github, senza passare per file `.zip` o simili
  - *Caricati direttamente in questo modo, i Notebook possono essere direttamente consultati dal browser; possono quindi essere utilizzati, ad esempio, per costruire un portfolio*

## RINGRAZIAMENTI

In [333...

Ringraziamenti



Out[333...



Un sentito ringraziamento ai miei colleghi di corso, con cui ho avuto il piacere di confrontarmi, ragionare e scambiare idee. Le nostre discussioni mi hanno arricchito e offerto spunti davvero interessanti. Ho apprezzato molto questo percorso condiviso. Un grazie anche ad Epicode e ai professori per aver reso stimolante questo modulo con Python.

**Davide Buccino**

In [ ]: