```
### Dataset Used:
COVID-19 Dataset (Our World in Data)


!wget https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv
```

```
--2026-01-23 13:29:07--  https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.108.133, 185.199.109.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 98391483 (94M) [text/plain]
Saving to: ‘owid-covid-data.csv’

owid-covid-data.csv 100%[===================>]  93.83M   221MB/s    in 0.4s

2026-01-23 13:29:08 (221 MB/s) - ‘owid-covid-data.csv’ saved [98391483/98391483]
```

```
!ls
```

```
owid-covid-data.csv  sample_data
```

```
import pandas as pd

covid = pd.read_csv("owid-covid-data.csv")
covid.head()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_s
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	

5 rows × 67 columns

```
covid.info()
covid['date'] = pd.to_datetime(covid['date'])
```

```
 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
```

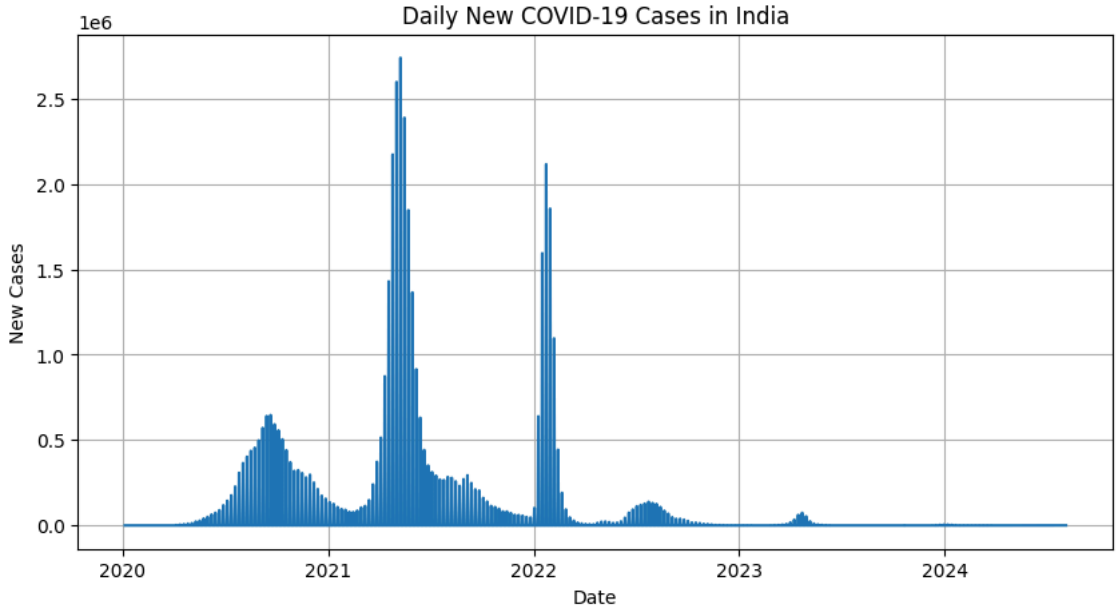
```
 41  people_vaccinated_per_hundred               81132 non-null   float64
 42  people_fully_vaccinated_per_hundred         78061 non-null   float64
 43  total_boosters_per_hundred                  53600 non-null   float64
 44  new_vaccinations_smoothed_per_million       195029 non-null  float64
 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  cardiovasc_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
```

```
india_covid = covid[covid['location'] == 'India']
india_covid.head()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_death
173549	IND	Asia	India	2020-01-05	0.0	0.0	NaN	0.0	0.0	
173550	IND	Asia	India	2020-01-06	0.0	0.0	NaN	0.0	0.0	
173551	IND	Asia	India	2020-01-07	0.0	0.0	NaN	0.0	0.0	
173552	IND	Asia	India	2020-01-08	0.0	0.0	NaN	0.0	0.0	
173553	IND	Asia	India	2020-01-09	0.0	0.0	NaN	0.0	0.0	

5 rows × 67 columns

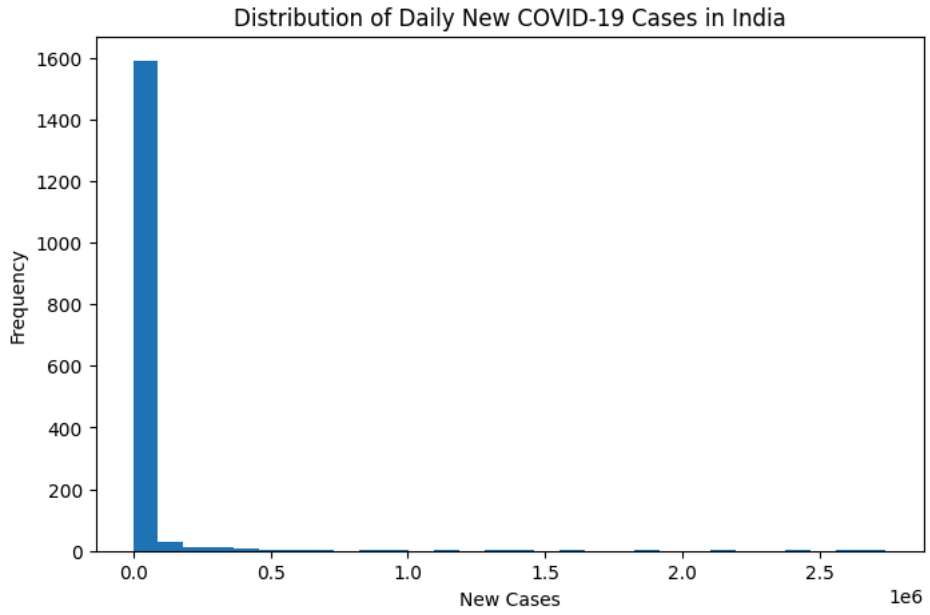
```
import matplotlib.pyplot as plt

plt.figure(figsize=(10,5))
plt.plot(india_covid['date'], india_covid['new_cases'])
plt.title("Daily New COVID-19 Cases in India")
plt.xlabel("Date")
plt.ylabel("New Cases")
plt.grid(True)
plt.show()
```



Insight:

- The graph shows multiple COVID-19 waves in India.
- The highest peak occurs during mid-2021, indicating the second wave.

```
plt.figure(figsize=(8,5))
plt.hist(india_covid['new_cases'].dropna(), bins=30)
plt.title("Distribution of Daily New COVID-19 Cases in India")
plt.xlabel("New Cases")
plt.ylabel("Frequency")
plt.show()
```

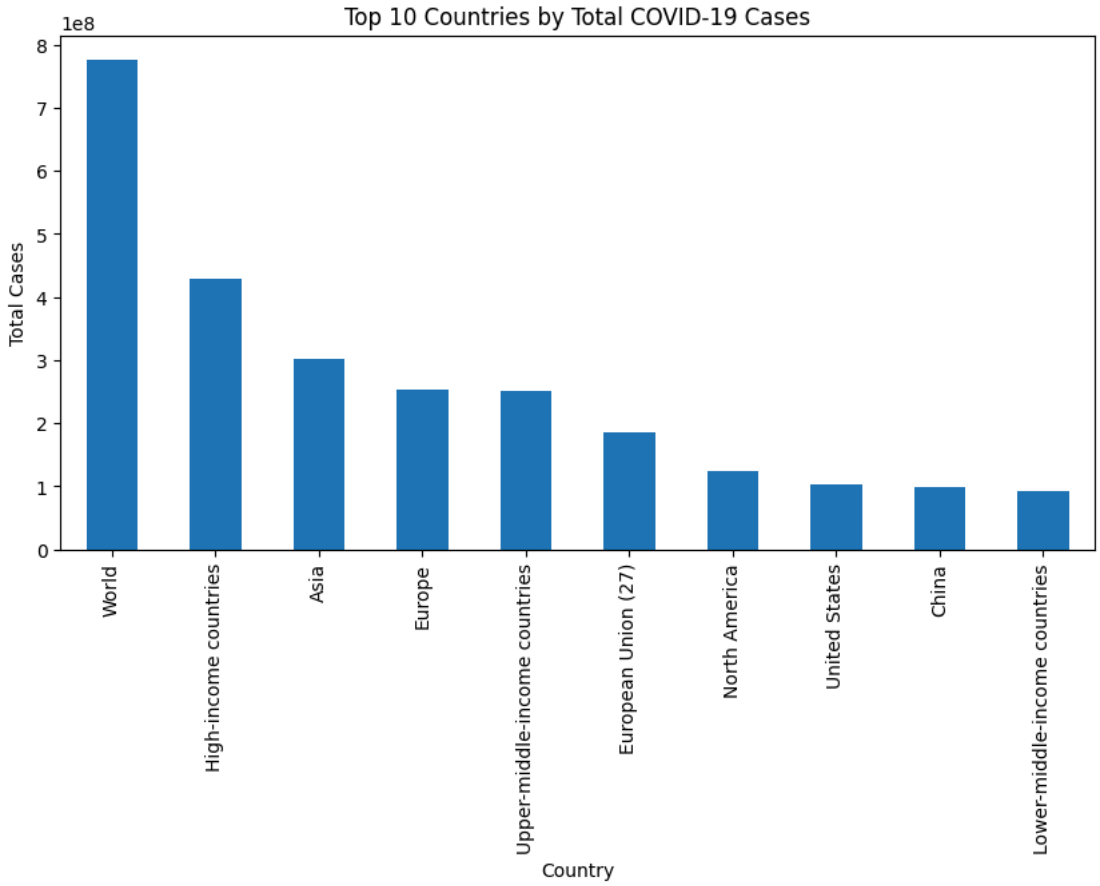


Insight:

- Most days report low case numbers.
- A few days show very high spikes, indicating outliers in the data.

```
top_countries = (
    covid.groupby('location')['total_cases']
    .max()
    .sort_values(ascending=False)
    .head(10)
)

plt.figure(figsize=(10,5))
top_countries.plot(kind='bar')
```

```
plt.title("Top 10 Countries by Total COVID-19 Cases")
plt.xlabel("Country")
plt.ylabel("Total Cases")
plt.show()
```



The bar chart shows that a small number of countries account for a large share of total COVID-19 cases.

The United States and India have the highest total cases, mainly due to large population size and widespread testing.

COVID-19 case trends show clear wave patterns over time, indicating multiple phases of outbreak and recovery.

The distribution of daily new cases is highly uneven, with a small number of days showing extreme spikes.