Dataset Link: https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases

Importing Necessary Libraries

In [1]:

```
#Write Your Code Here
import pandas as pd
import numpy as np
import seaborn as sns
```

Import and read dataset

In [2]:

```
import pandas as pd
data = pd.read_csv("Covid_19_cases4 (2).csv")
print(data)
print(data.head(10))
```

```
        dateRep day month year cases deaths countriesAndTerritories
0 31-05-2021 31 5 2021 366 5 Austria
```

```
1 30-05-2021 30 5 2021 570 6 Austria
2 29-05-2021 29 5 2021 538 11 Austria
3 28-05-2021 28 5 2021 639 4 Austria
4 27-05-2021 27 5 2021 405 19 Austria
... ... ... ... ... ... ... ...
2725 06-03-2021 6 3 2021 3455 17 Sweden
2726 05-03-2021 5 3 2021 4069 12 Sweden
2727 04-03-2021 4 3 2021 4884 14 Sweden
2728 03-03-2021 3 3 2021 4876 19 Sweden
2729 02-03-2021 2 3 2021 6191 19 Sweden

[2730 rows x 7 columns]
        dateRep day month year cases deaths countriesAndTerritories
0 31-05-2021 31 5 2021 366 5 Austria
1 30-05-2021 30 5 2021 570 6 Austria
2 29-05-2021 29 5 2021 538 11 Austria
3 28-05-2021 28 5 2021 639 4 Austria
4 27-05-2021 27 5 2021 405 19 Austria
5 26-05-2021 26 5 2021 287 8 Austria
6 25-05-2021 25 5 2021 342 3 Austria
7 24-05-2021 24 5 2021 520 3 Austria
8 23-05-2021 23 5 2021 626 8 Austria
9 22-05-2021 22 5 2021 671 12 Austria
```

Click Here for the Hint

Data Cleaning

In [3]:

aa eanng a. Missing Value

```
import pandas as pd
data = pd.read_csv("Covid_19_cases4 (2).csv")
print(data.isnull().sum())
```

```
dateRep 0
day 0
month 0
year 0
cases 0
deaths 0
countriesAndTerritories 0
dtype: int64
```

Click Here for the Hint

b. Duplicate data

```
import pandas as pd
data = pd.read_csv("Covid_19_cases4 (2).csv")
print(data.duplicated().sum())
```

0

Click Here for the Hint

c. drop unecessary columns

```
import pandas as pd
data = pd.read_csv("Covid_19_cases4 (2).csv")
data.drop(["day", "month"], axis = 1, inplace = True)
data["dateRep"] = pd.to_datetime(data["dateRep"])
data.set_index("dateRep", inplace = True)
print(data.head())
```

dateRep	year	cases	deaths	countriesAndTerritories
2021-05-31	2021	366	5	Austria
2021-05-30	2021	570	6	Austria
2021-05-29	2021	538	11	Austria
2021-05-28	2021	639	4	Austria
2021-05-27	2021	405	19	Austria

Data Analysis

1. Count the total number of cases and deaths in the dataset.

```
import pandas as pd
data = pd.read_csv("Covid_19_cases4 (2).csv")
total_cases = data["cases"].sum()
total_deaths = data["deaths"].sum()
print("Total Number Of Cases:", total_cases)
print("Total Number Of Deaths: ", total_deaths)
```

Total Number Of Cases: 9994560
Total Number Of Deaths: 178247

Click Here for the Hint

1. Calculate the percentage of cases and deaths by country.

```
import pandas as pd
covid_data = pd.read_csv("Covid_19_cases4 (2).csv")
grouped = covid_data.groupby("countriesAndTerritories")
total_cases = grouped["cases"].sum()
total_deaths = grouped["deaths"].sum()
case_percent = (total_cases / total_cases.sum()) * 100
death_percent = (total_deaths / total_deaths.sum()) * 100
covid_data["case_percent"] = covid_data["countriesAndTerritories"].map(case_percent) covid_data["death_percent"] =
covid_data["countriesAndTerritories"].map(death_percent) percentages = covid_data[["countriesAndTerritories", "case_percent",
"death_percent"]].drop_duplicates() print(percentages)
```

countriesAndTerritories case_percent death_percent 0 Austria 1.845164 1.079962 91 Belgium 2.882758 1.512508 182 Bulgaria 1.713292 4.191375 273 Croatia 1.132296 1.395816 364 Cyprus 0.377205 0.072371 455 Czechia 4.214503 5.407665 546 Denmark 0.692257 0.086958 637 Estonia 0.629502 0.366907 728 Finland 0.347789 0.099300 819 France 20.219079 12.890540 910 Germany 12.347297 10.287410 1001 Greece 2.103154 3.113657 1092 Hungary 3.718153 8.232958 1183 Iceland 0.005273 0.000561 1274 Ireland 0.420799 0.348954 1365 Italy 12.914405 15.903213 1456 Latvia 0.469375 0.421886 1547 Liechtenstein 0.004372 0.002244 1638 Lithuania 0.770819 0.573362 1729 Luxembourg 0.144719 0.098739 1820 Malta 0.075901 0.058346 1911 Netherlands 5.582867 1.152895 2002 Norway 0.540244 0.090324 2093 Poland 11.655981 16.813186 2184 Portugal 0.441200 0.396080 2275 Romania 2.757400 5.568677 2366 Slovakia 1.785721 2.889249 2457 Slovenia 0.635846 0.326513 2548 Spain 5.530238 5.803183 2639 Sweden 4.042389 0.815161

Click Here for the Hint

1. Find the country with the highest number of cases and deaths.

```
import pandas as pd
df = pd.read_csv("Covid_19_cases4 (2).csv")
country_totals = df.groupby("countriesAndTerritories")["cases", "deaths"].sum()
max_cases_country = country_totals["cases"].idxmax()
max_deaths_country = country_totals["deaths"].idxmax()
max_cases = country_totals["cases"].max()
max_deaths = country_totals["deaths"].max()
print(f"The Country With The Highest Number Of COVID-19 cases is {max_cases_country} With {max_cases} cases") print(f"The Country With The Highest Number Of COVID-19 cases is {max_deaths_country} With {max_deaths} cases")
```

```
The Country With The Highest Number Of COVID-19 cases is France With 2020808 cases
The Country With The Highest Number Of COVID-19 cases is Poland With 29969 cases
C:\Users\ELCOT-Lenovo\AppData\Local\Temp\ipykernel_4056\2632207238.py:3: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
  country_totals = df.groupby("countriesAndTerritories")["cases", "deaths"].sum()
```

Click Here for the Hint

Data Visualization

1. Find top five countries in terms of cases, store them in a new dataframe and Visualize them

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

covid_data = pd.read_csv("Covid_19_cases4 (2).csv")
covid_cases = covid_data[["countriesAndTerritories", "cases", "deaths"]]

print(covid_cases.head(10))

country_totals = covid_cases.groupby("countriesAndTerritories").sum()
top_5_counteries = country_totals.sort_values(by = "cases", ascending = False).head(5) top_5_counteries =

top_5_counteries.reset_index()

print(top_5_counteries)

sns.set(style = "whitegrid")
plt.figure(figsize = (10,6))
ax = sns.barplot(x = "countriesAndTerritories", y = "cases", data = top_5_counteries) ax.set(xlabel = "Country", ylabel = "Number Of Cases", title = "To
Countries With The Highest Number Of COVID-19 plt.show()
```

```
  countriesAndTerritories cases deaths
0 Austria 366 5
1 Austria 570 6
2 Austria 538 11
3 Austria 639 4
4 Austria 405 19
5 Austria 287 8
6 Austria 342 3
7 Austria 520 3
8 Austria 626 8
9 Austria 671 12
  countriesAndTerritories cases deaths
0 France 2020808 22977
1 Italy 1290738 28347
2 Germany 1234058 18337
3 Poland 1164964 29969
4 Netherlands 557983 2055
```

Click Here for the Hint

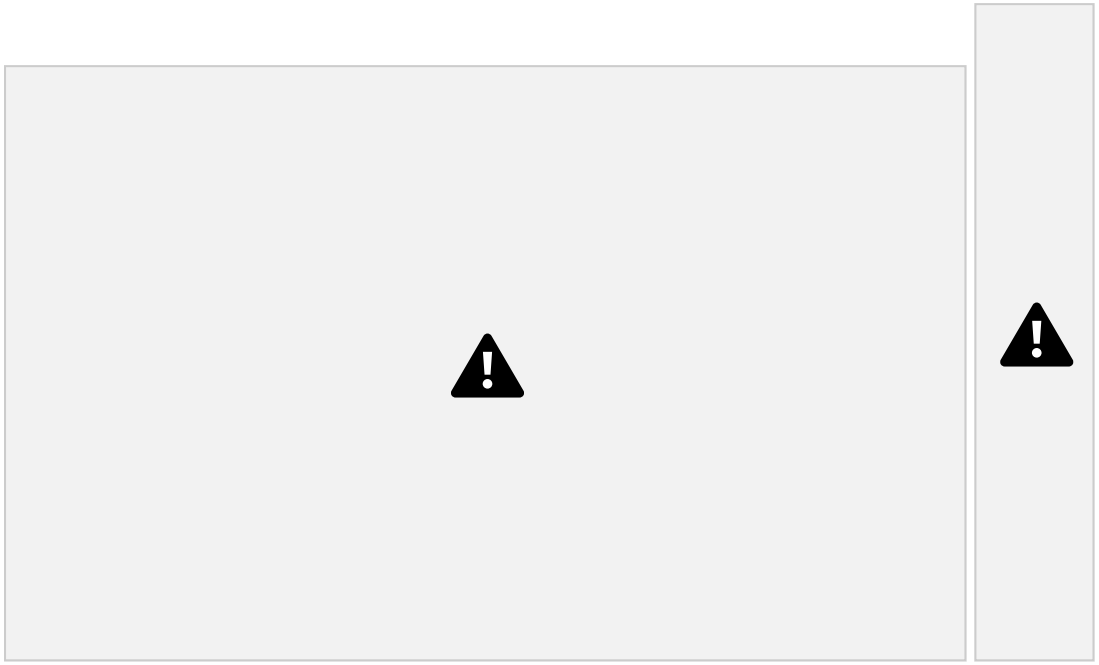
1. Find top five countries in terms of deaths, store them in a new dataframe and Visualize them

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

covid_data = pd.read_csv("Covid_19_cases4 (2).csv")
covid_cases = covid_data[["countriesAndTerritories", "cases", "deaths"]]

print(covid_cases.head(10))

country_totals = covid_cases.groupby("countriesAndTerritories").sum()
top_5_counteries = country_totals.sort_values(by = "deaths", ascending = False).head(5) top_5_counteries =

top_5_counteries.reset_index()

print(top_5_counteries)

sns.set(style = "whitegrid")
plt.figure(figsize = (10,6))
ax = sns.barplot(x = "countriesAndTerritories", y = "deaths", data = top_5_counteries) ax.set(xlabel = "Country", ylabel = "Number Of Deaths", title = "
Countries With The Highest Number Of COVID-19plt.show()
```

```
  countriesAndTerritories cases deaths
0 Austria 366 5
1 Austria 570 6
2 Austria 538 11
3 Austria 639 4
```

4 Austria 405 19
5 Austria 287 8
6 Austria 342 3
7 Austria 520 3
8 Austria 626 8
9 Austria 671 12

countriesAndTerritories cases deaths
0 Poland 1164964 29969
1 Italy 1290738 28347
2 France 2020808 22977
3 Germany 1234058 18337
4 Hungary 371613 14675



Click Here for the Hint

Loading
[MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js