Covid_country_wise

December 18, 2025

1 Country-wise Data Analysis Project

1.1 Objective

The objective of this project is to analyze a country-specific dataset to understand patterns, trends, and insights using Python and Pandas. This project focuses on basic data analysis techniques for learning purposes.

1.2 Dataset Source

The dataset used in this project is collected from Kaggle. The original dataset contains multiple files, each representing data for a different country.

1.3 Disclaimer

Only one country-wise dataset has been selected for this analysis to keep the project focused and simple. The dataset is already cleaned and complete. This project is created solely for educational and practice purposes.

[42]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

1.3.1 Dataset Loading & Basic Exploration

[7]:
```
df=pd.read_csv("C:\\Users\\Asus\\Downloads\\country_wise_latest.csv")
df
```

[7]:

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	\
0	Afghanistan	36263	1269	25198	9796	106	
1	Albania	4880	144	2745	1991	117	
2	Algeria	27973	1163	18837	7973	616	
3	Andorra	907	52	803	52	10	
4	Angola	950	41	242	667	18	
..	...	...	...	...	...	...	
182	West Bank and Gaza	10621	78	3752	6791	152	
183	Western Sahara	10	1	8	1	0	
184	Yemen	1691	483	833	375	10	

```
185              Zambia       4552     140       2815    1597         71
186            Zimbabwe       2704      36        542    2126        192

     New deaths  New recovered  Deaths / 100 Cases  Recovered / 100 Cases  \
0            10             18                3.50                  69.49
1             6             63                2.95                  56.25
2             8            749                4.16                  67.34
3             0              0                5.73                  88.53
4             1              0                4.32                  25.47
..          ...            ...                 ...                    ...
182           2              0                0.73                  35.33
183           0              0               10.00                  80.00
184           4             36               28.56                  49.26
185           1            465                3.08                  61.84
186           2             24                1.33                  20.04

     Deaths / 100 Recovered  Confirmed last week  1 week change  \
0                      5.04                35526            737
1                      5.25                 4171            709
2                      6.17                23691           4282
3                      6.48                  884             23
4                     16.94                  749            201
..                      ...                  ...            ...
182                    2.08                 8916           1705
183                   12.50                   10              0
184                   57.98                 1619             72
185                    4.97                 3326           1226
186                    6.64                 1713            991

     1 week % increase             WHO Region
0                 2.07  Eastern Mediterranean
1                17.00                 Europe
2                18.07                 Africa
3                 2.60                 Europe
4                26.84                 Africa
..                 ...                    ...
182              19.12  Eastern Mediterranean
183               0.00                 Africa
184               4.45  Eastern Mediterranean
185              36.86                 Africa
186              57.85                 Africa

[187 rows x 15 columns]
```

[8]:
```
df.head()
```

[8]:
```
  Country/Region  Confirmed  Deaths  Recovered  Active  New cases  New deaths  \
0    Afghanistan      36263    1269      25198    9796        106          10
1        Albania       4880     144       2745    1991        117           6
2        Algeria      27973    1163      18837    7973        616           8
3        Andorra        907      52        803      52         10           0
4         Angola        950      41        242     667         18           1

   New recovered  Deaths / 100 Cases  Recovered / 100 Cases  \
0             18                3.50                  69.49
1             63                2.95                  56.25
2            749                4.16                  67.34
3              0                5.73                  88.53
4              0                4.32                  25.47

   Deaths / 100 Recovered  Confirmed last week  1 week change  \
0                    5.04                35526            737
1                    5.25                 4171            709
2                    6.17                23691           4282
3                    6.48                  884             23
4                   16.94                  749            201

   1 week % increase             WHO Region
0               2.07  Eastern Mediterranean
1              17.00                 Europe
2              18.07                 Africa
3               2.60                 Europe
4              26.84                 Africa
```

1.3.2 Explanation:

The `head()` function is used to display the first few rows of the dataset to understand the structure and columns of the data.

[10]:
```
df.shape
```

[10]:
```
(187, 15)
```

1.3.3 Explanation:

The dataset contains 187 rows and 15 columns, indicating the total number of records and features.

[11]:
```
df.info
```

[11]:
```
<bound method DataFrame.info of          Country/Region  Confirmed  Deaths
Recovered  Active  New cases  \
0           Afghanistan      36263    1269      25198    9796        106
1               Albania       4880     144       2745    1991        117
2               Algeria      27973    1163      18837    7973        616
```

```
3                Andorra        907      52        803      52         10
4                 Angola        950      41        242     667         18
..                   ...        ...     ...        ...     ...        ...
182  West Bank and Gaza      10621      78       3752    6791        152
183      Western Sahara         10       1          8       1          0
184               Yemen       1691     483        833     375         10
185              Zambia       4552     140       2815    1597         71
186            Zimbabwe       2704      36        542    2126        192

     New deaths  New recovered  Deaths / 100 Cases  Recovered / 100 Cases  \
0            10             18                3.50                  69.49
1             6             63                2.95                  56.25
2             8            749                4.16                  67.34
3             0              0                5.73                  88.53
4             1              0                4.32                  25.47
..          ...            ...                 ...                    ...
182           2              0                0.73                  35.33
183           0              0               10.00                  80.00
184           4             36               28.56                  49.26
185           1            465                3.08                  61.84
186           2             24                1.33                  20.04

     Deaths / 100 Recovered  Confirmed last week  1 week change  \
0                      5.04                35526            737
1                      5.25                 4171            709
2                      6.17                23691           4282
3                      6.48                  884             23
4                     16.94                  749            201
..                      ...                  ...            ...
182                    2.08                 8916           1705
183                   12.50                   10              0
184                   57.98                 1619             72
185                    4.97                 3326           1226
186                    6.64                 1713            991

     1 week % increase             WHO Region
0                 2.07  Eastern Mediterranean
1                17.00                 Europe
2                18.07                 Africa
3                 2.60                 Europe
4                26.84                 Africa
..                 ...                    ...
182              19.12  Eastern Mediterranean
183               0.00                 Africa
184               4.45  Eastern Mediterranean
185              36.86                 Africa
186              57.85                 Africa
```

```
[187 rows x 15 columns]>
```

1.3.4 Explanation:

The `info()` function is used to understand the structure of the dataset, such as the number of columns, their data types, and the presence of missing values.

1.4 Check Missing Values

[12]:
```
df.isnull().sum()
```

[12]:
```
Country/Region            0
Confirmed                 0
Deaths                    0
Recovered                 0
Active                    0
New cases                 0
New deaths                0
New recovered             0
Deaths / 100 Cases        0
Recovered / 100 Cases     0
Deaths / 100 Recovered    0
Confirmed last week       0
1 week change             0
1 week % increase         0
WHO Region                0
dtype: int64
```

1.4.1 Explanation:

This method is used to check the number of missing (null) values in each column of the dataset. No missing values are present in the dataset.

1.5 Country-wise Summary Statistics

[13]:
```
df.describe()
```

[13]:

	Confirmed	Deaths	Recovered	Active	New cases	\
count	1.870000e+02	187.000000	1.870000e+02	1.870000e+02	187.000000	
mean	8.813094e+04	3497.518717	5.063148e+04	3.400194e+04	1222.957219	
std	3.833187e+05	14100.002482	1.901882e+05	2.133262e+05	5710.374790	
min	1.000000e+01	0.000000	0.000000e+00	0.000000e+00	0.000000	
25%	1.114000e+03	18.500000	6.265000e+02	1.415000e+02	4.000000	
50%	5.059000e+03	108.000000	2.815000e+03	1.600000e+03	49.000000	
75%	4.046050e+04	734.000000	2.260600e+04	9.149000e+03	419.500000	
max	4.290259e+06	148011.000000	1.846641e+06	2.816444e+06	56336.000000	

```
        New deaths  New recovered  Deaths / 100 Cases  Recovered / 100 Cases  \
count   187.000000     187.000000          187.000000             187.000000
mean     28.957219     933.812834            3.019519              64.820535
std     120.037173    4197.719635            3.454302              26.287694
min       0.000000       0.000000            0.000000               0.000000
25%       0.000000       0.000000            0.945000              48.770000
50%       1.000000      22.000000            2.150000              71.320000
75%       6.000000     221.000000            3.875000              86.885000
max    1076.000000   33728.000000           28.560000             100.000000

       Deaths / 100 Recovered  Confirmed last week  1 week change  \
count                  187.00         1.870000e+02     187.000000
mean                      inf         7.868248e+04    9448.459893
std                       NaN         3.382737e+05   47491.127684
min                      0.00         1.000000e+01     -47.000000
25%                      1.45         1.051500e+03      49.000000
50%                      3.62         5.020000e+03     432.000000
75%                      6.44         3.708050e+04    3172.000000
max                       inf         3.834677e+06  455582.000000

       1 week % increase
count         187.000000
mean           13.606203
std            24.509838
min            -3.840000
25%             2.775000
50%             6.890000
75%            16.855000
max           226.320000
```

1.5.1 Explanation

The 'describe()' function provides a quick statistical summary of the dataset including the mean, standard deviation etc.It is used to quickly understand the data distribution.

1.6 Confirmed Cases Analysis

[67]:
```
df1=df[['Country/Region','Confirmed']]
top_10_confirmed=df1.sort_values(by='Confirmed',ascending=False).head(10)
top_10_confirmed
```

[67]:
```
     Country/Region  Confirmed
173              US    4290259
23           Brazil    2442375
79            India    1480073
138          Russia     816680
154    South Africa     452529
```

```
111          Mexico     395489
132            Peru     389717
35            Chile     347923
177  United Kingdom     301708
81             Iran     293606
```

1.6.1 Explanation

This code sorts the dataset in descending order based on the total number of Confirmed Cases and extracts the Top 10 most affected countries.

1.7 Death Rate Analysis

[37]:
```
df2=df[['Country/Region','Deaths']]
top_10_deaths=df2.sort_values(by='Deaths', ascending=False).head(10)
top_10_deaths
```

[37]:
```
     Country/Region  Deaths
173              US  148011
23           Brazil   87618
177  United Kingdom   45844
111          Mexico   44022
85            Italy   35112
79            India   33408
61           France   30212
157           Spain   28432
132            Peru   18418
81             Iran   15912
```

1.7.1 Explanation

This code sorts the dataset in descending order based on the total number of Deaths Cases and extracts the Top 10 most deaths countries.

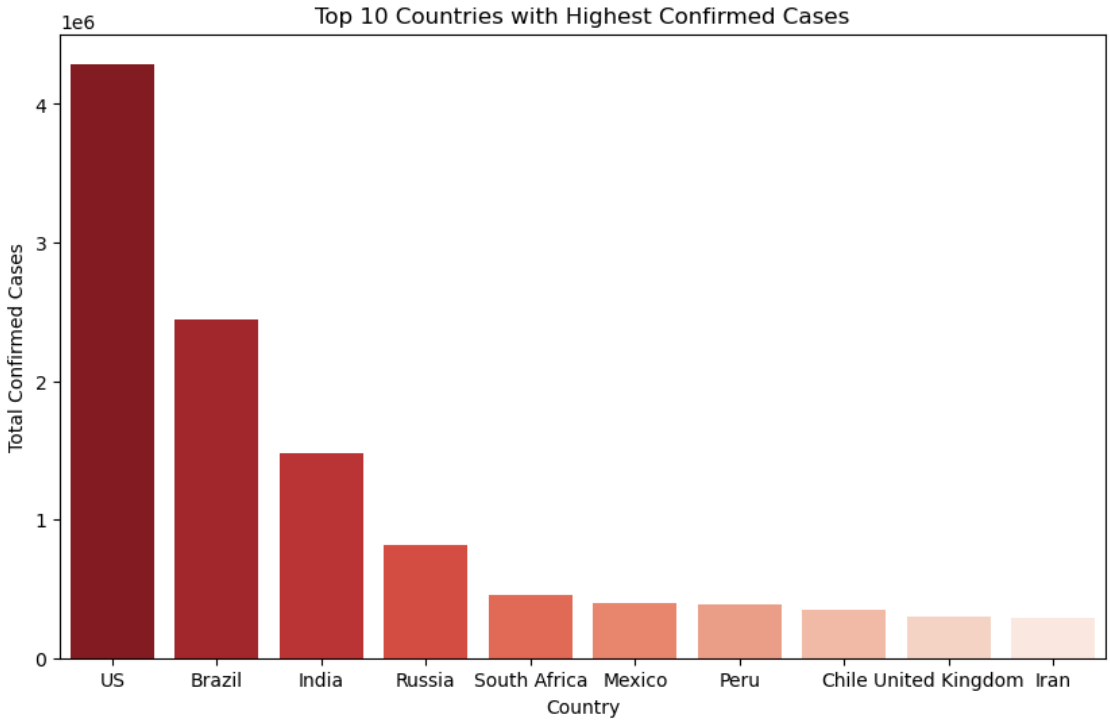
2 Data Visualization

[48]:
```
plt.figure(figsize=(10,6))
sns.barplot(y='Confirmed', x='Country/Region', data=top_10_confirmed,␣
 ↪palette='Reds_r')
plt.title('Top 10 Countries with Highest Confirmed Cases')
plt.ylabel('Total Confirmed Cases')
plt.xlabel('Country')
plt.show()
```

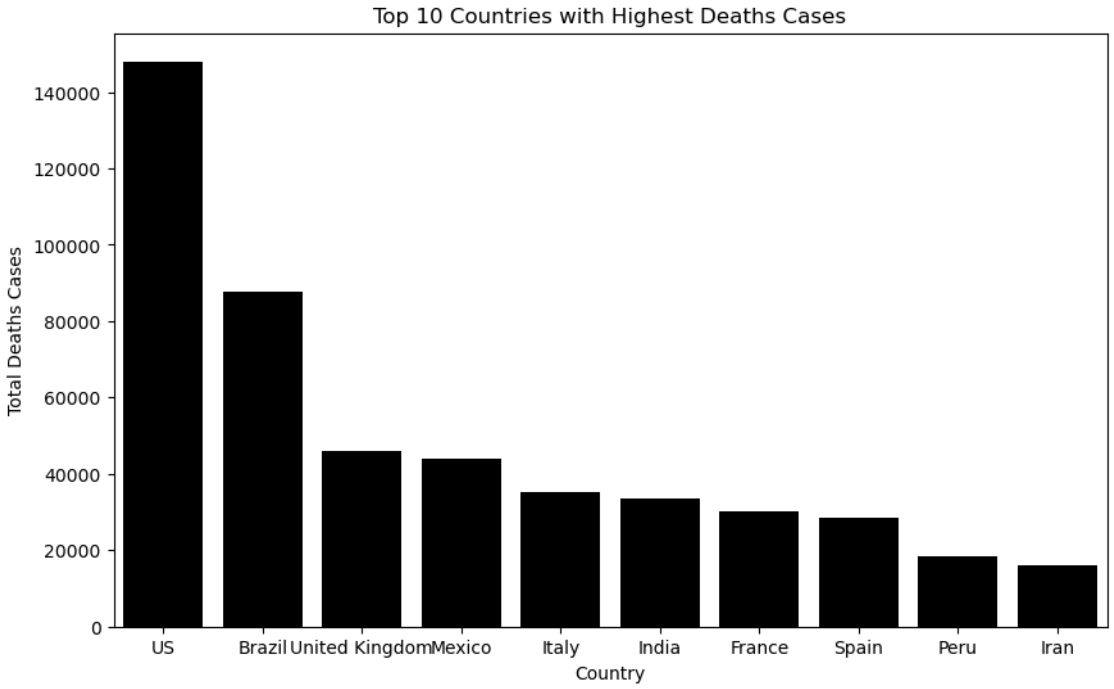
```
C:\Users\Asus\AppData\Local\Temp\ipykernel_13672\3684451670.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
```

effect.

```
  sns.barplot(y='Confirmed', x='Country/Region', data=top_10_confirmed,
palette='Reds_r')
```

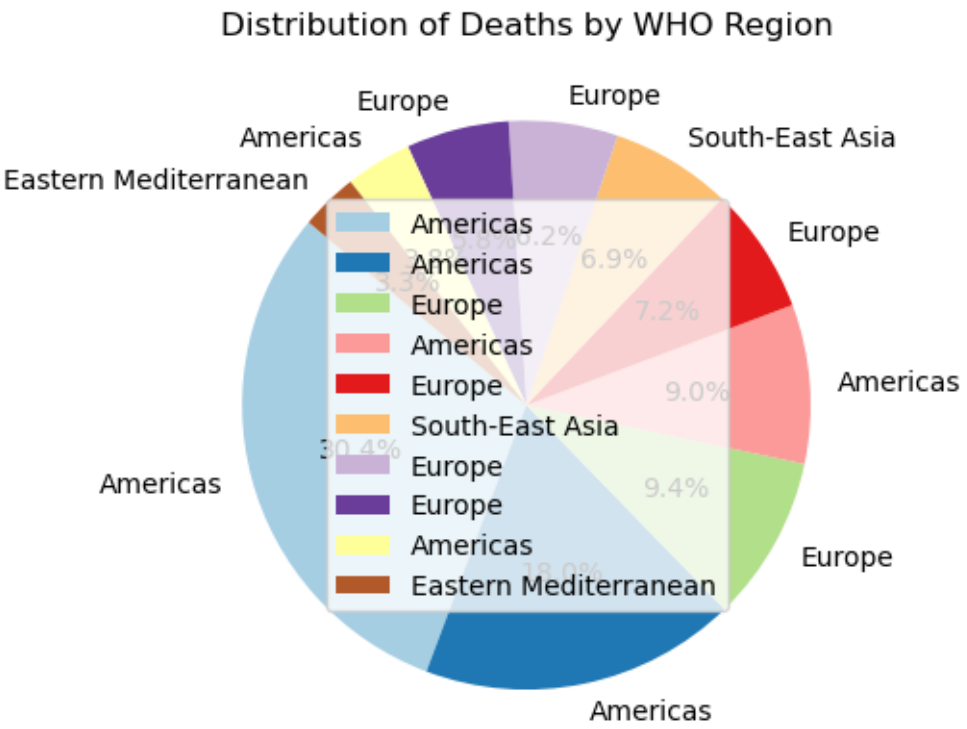


[69]:
```
plt.figure(figsize=(10,6))
sns.barplot(y='Deaths', x='Country/Region', data=top_10_deaths, color='black')
plt.title('Top 10 Countries with Highest Deaths Cases')
plt.ylabel('Total Deaths Cases')
plt.xlabel('Country')
plt.show()
```



[61]:
```
df3 = df[['WHO Region', 'Deaths']]

WHO_region_death = df3.sort_values(by='Deaths', ascending=False).head(10)

plt.figure(figsize=(8,8))
WHO_region_death.plot(
    kind='pie',
    y='Deaths',
    labels=WHO_region_death['WHO Region'],
    autopct='%1.1f%%',
    startangle=140,
    colormap='Paired'
)

plt.title('Distribution of Deaths by WHO Region')
plt.ylabel('')
plt.show()
```

```
<Figure size 800x800 with 0 Axes>
```



2.1 *Conclusion*

The analysis of the COVID-19 dataset provides a clear overview of the pandemic's impact across various countries and WHO regions. By using Pandas and NumPy, we identified significant variations in Confirmed Cases and Death Rates, highlighting global health hotspots like the US and Brazil.

[ ]: