```python
#1 Importing necessary Python Libraries:
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

#2 Reading the dataset:
df=pd.read_csv(r"C:\Users\ADITI DUNGYAN\Downloads\covid_data.csv")
df
```

```
               Country                Other names ISO 3166-1 alpha-3
CODE  \
0          Afghanistan                Afghanistan
AFG
1              Albania                    Albania
ALB
2              Algeria                    Algeria
DZA
3              Andorra                    Andorra
AND
4               Angola                     Angola
AGO
..                 ...                        ...
...
220  Wallis and Futuna  Wallis and Futuna Islands
WLF
221     Western Sahara             Western Sahara
ESHÂ
222              Yemen                      Yemen
YEM
223             Zambia                     Zambia
ZMB
224           Zimbabwe                   Zimbabwe
ZWE

     Population Continent  Total Cases  Total Deaths  Tot Cases//1M
pop  \
0      40462186      Asia       177827          7671
4395
1       2872296    Europe       273870          3492
95349
2      45236699    Africa       265691          6874
5873
3         77481    Europe        40024           153
516565
4      34654212    Africa        99194          1900
2862
..          ...       ...          ...           ...                ..
.
220       10894   Oceania          454             7
41674
```

```
221      623031    Africa           10              1
16
222    30975258      Asia        11806           2143
381
223    19284482    Africa       317076           3967
16442
224    15241601    Africa       246525           5446
16174

     Tot Deaths/1M pop  Death percentage
0                  190          4.313743
1                 1216          1.275058
2                  152          2.587216
3                 1975          0.382271
4                   55          1.915438
..                 ...               ...
220                643          1.541850
221                  2         10.000000
222                 69         18.151787
223                206          1.251120
224                357          2.209107

[225 rows x 10 columns]
```

```
#3 Display the first 10 rows of the dataset:
df.head(10)
```

```
               Country          Other names ISO 3166-1 alpha-3 CODE  \
0          Afghanistan          Afghanistan                      AFG
1              Albania              Albania                      ALB
2              Algeria              Algeria                      DZA
3              Andorra              Andorra                      AND
4               Angola               Angola                      AGO
5             Anguilla             Anguilla                      AIA
6  Antigua and Barbuda  Antigua and Barbuda                      ATG
7            Argentina            Argentina                      ARG
8              Armenia              Armenia                      ARM
9                Aruba                Aruba                      ABW

   Population                          Continent  Total Cases  Total
Deaths  \
0    40462186                               Asia       177827
7671
1     2872296                             Europe       273870
3492
2    45236699                             Africa       265691
6874
3       77481                             Europe        40024
153
4    34654212                             Africa        99194
```

```
1900
5       15237  Latin America and the Caribbean          2700
9
6       99348  Latin America and the Caribbean          7493
135
7    45921761  Latin America and the Caribbean       9041124
128065
8     2972939                                Asia       422574
8617
9      107560  Latin America and the Caribbean         34051
212

   Tot Cases//1M pop  Tot Deaths/1M pop  Death percentage
0               4395                190          4.313743
1              95349               1216          1.275058
2               5873                152          2.587216
3             516565               1975          0.382271
4               2862                 55          1.915438
5             177200                591          0.333333
6              75422               1359          1.801682
7             196881               2789          1.416472
8             142140               2898          2.039169
9             316577               1971          0.622596
```

```
#4 Display the last 5 rows of the dataset:
df.tail(5)
```

```
               Country                Other names ISO 3166-1 alpha-3
CODE  \
220  Wallis and Futuna  Wallis and Futuna Islands
WLF
221     Western Sahara             Western Sahara
ESHÂ
222              Yemen                      Yemen
YEM
223             Zambia                     Zambia
ZMB
224           Zimbabwe                   Zimbabwe
ZWE

     Population Continent  Total Cases  Total Deaths  Tot Cases//1M
pop  \
220       10894   Oceania          454             7
41674
221      623031    Africa           10             1
16
222    30975258      Asia        11806          2143
381
223    19284482    Africa       317076          3967
16442
```

```
224    15241601    Africa       246525          5446
16174

     Tot Deaths/1M pop  Death percentage
220                643          1.541850
221                  2         10.000000
222                 69         18.151787
223                206          1.251120
224                357          2.209107
```

```
#5 total number of rows and columns of the dataset:
df.shape
```

```
(225, 10)
```

```
#6 Concise information about the dataset:
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 225 entries, 0 to 224
Data columns (total 10 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Country                  225 non-null    object
 1   Other names              224 non-null    object
 2   ISO 3166-1 alpha-3 CODE  225 non-null    object
 3   Population               225 non-null    int64
 4   Continent                225 non-null    object
 5   Total Cases              225 non-null    int64
 6   Total Deaths             225 non-null    int64
 7   Tot Cases//1M pop        225 non-null    int64
 8   Tot Deaths/1M pop        225 non-null    int64
 9   Death percentage         225 non-null    float64
dtypes: float64(1), int64(5), object(4)
memory usage: 17.7+ KB
```

```
#7 Statistical summary of the numerical columns:
df.describe()
```

```
         Population   Total Cases  Total Deaths  Tot Cases//1M pop  \
count  2.250000e+02  2.250000e+02  2.250000e+02         225.000000
mean   3.507321e+07  2.184781e+06  2.744813e+04      136900.373333
std    1.392418e+08  7.275938e+06  9.689177e+04      145060.340289
min    8.050000e+02  1.000000e+00  0.000000e+00           9.000000
25%    5.665570e+05  2.407100e+04  1.890000e+02       11384.000000
50%    5.827911e+06  1.639360e+05  1.965000e+03       88987.000000
75%    2.190585e+07  1.092547e+06  1.366000e+04      223335.000000
max    1.439324e+09  8.183905e+07  1.008222e+06      696044.000000

       Tot Deaths/1M pop  Death percentage
count         225.000000        225.000000
```

```
mean         1096.715556          1.444125
std          1195.715543          1.741728
min             0.000000          0.000000
25%           123.000000          0.511291
50%           708.000000          1.036905
75%          1795.000000          1.977017
max          6286.000000         18.151787
```

```
#8 Missing values in each column.
df.isnull().sum()
```

```
Country                     0
Other names                 1
ISO 3166-1 alpha-3 CODE     0
Population                  0
Continent                   0
Total Cases                 0
Total Deaths                0
Tot Cases//1M pop           0
Tot Deaths/1M pop           0
Death percentage            0
dtype: int64
```

```
#9 Row which contains null values (Hint axis=1):
nan_df1=df[df.isnull().any(axis=1)]
nan_df1
```

```
        Country Other names ISO 3166-1 alpha-3 CODE  Population
Continent  \
135  Montenegro         NaN                     MNE      628205
Europe

     Total Cases  Total Deaths  Tot Cases//1M pop  Tot Deaths/1M
pop  \
135       233326          2705             371417                4306


     Death percentage
135          1.159322
```

```
#10 Filling null value with most suitable value:
df1=df.fillna(df.mean(numeric_only=True)).fillna(df.mode().iloc[0])
df1
```

```
               Country                Other names ISO 3166-1 alpha-3
CODE  \
0          Afghanistan                Afghanistan
AFG
1              Albania                    Albania
ALB
2              Algeria                    Algeria
```

```
DZA
3              Andorra                    Andorra
AND
4               Angola                     Angola
AGO
..                 ...                        ...
...
220  Wallis and Futuna  Wallis and Futuna Islands
WLF
221     Western Sahara             Western Sahara
ESHÂ
222              Yemen                      Yemen
YEM
223             Zambia                     Zambia
ZMB
224           Zimbabwe                   Zimbabwe
ZWE

     Population Continent  Total Cases  Total Deaths  Tot Cases//1M
pop  \
0      40462186      Asia       177827          7671
4395
1       2872296    Europe       273870          3492
95349
2      45236699    Africa       265691          6874
5873
3         77481    Europe        40024           153
516565
4      34654212    Africa        99194          1900
2862
..          ...       ...          ...           ...                  ..
.
220       10894   Oceania          454             7
41674
221      623031    Africa           10             1
16
222    30975258      Asia        11806          2143
381
223    19284482    Africa       317076          3967
16442
224    15241601    Africa       246525          5446
16174

     Tot Deaths/1M pop  Death percentage
0                  190          4.313743
1                 1216          1.275058
2                  152          2.587216
3                 1975          0.382271
4                   55          1.915438
```

```
..                  ...                 ...
220                 643            1.541850
221                   2           10.000000
222                  69           18.151787
223                 206            1.251120
224                 357            2.209107

[225 rows x 10 columns]
```

```
#11 checking whether null values are replaced or not:
df1.info()
```

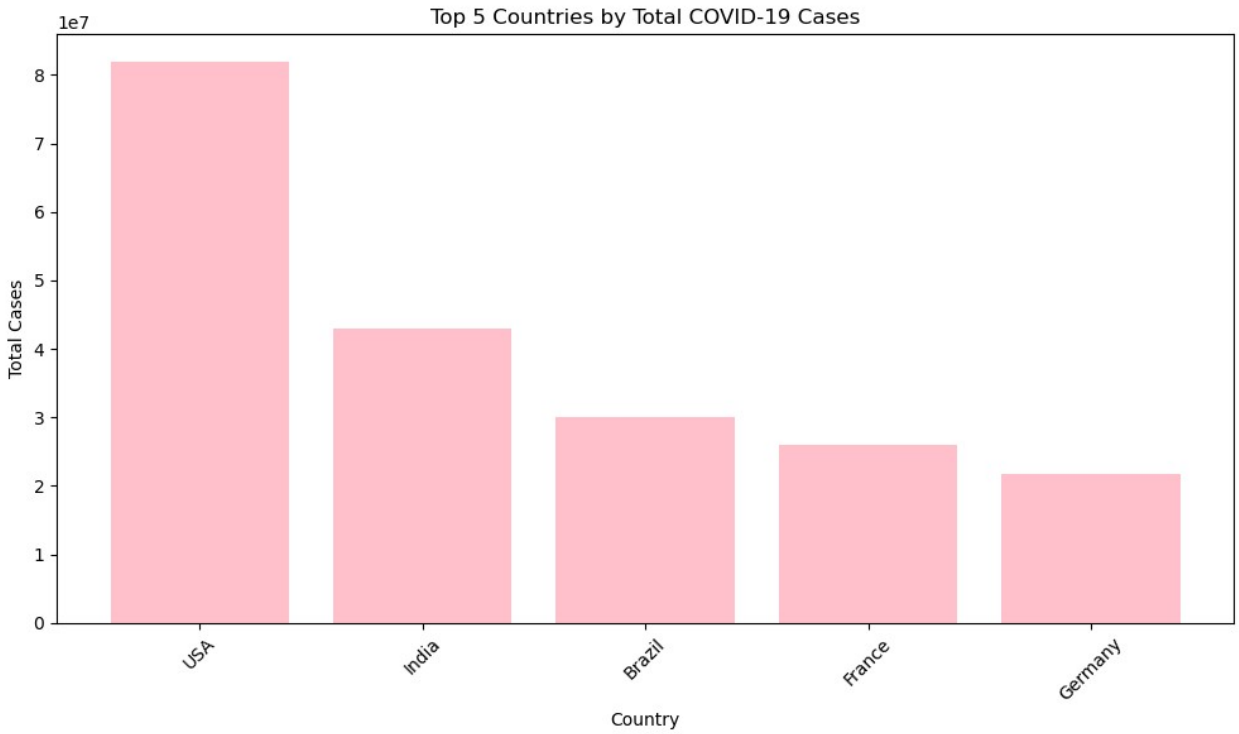
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 225 entries, 0 to 224
Data columns (total 10 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Country                  225 non-null    object 
 1   Other names              225 non-null    object 
 2   ISO 3166-1 alpha-3 CODE  225 non-null    object 
 3   Population               225 non-null    int64  
 4   Continent                225 non-null    object 
 5   Total Cases              225 non-null    int64  
 6   Total Deaths             225 non-null    int64  
 7   Tot Cases//1M pop        225 non-null    int64  
 8   Tot Deaths/1M pop        225 non-null    int64  
 9   Death percentage         225 non-null    float64
dtypes: float64(1), int64(5), object(4)
memory usage: 17.7+ KB
```

```
#12 Top 5 countries which have the greatest number of covid cases:
top_5_cases = df.nlargest(5,'Total Cases')[['Country', 'Total Cases']]
top_5_cases
```

```
     Country  Total Cases
214      USA     81839052
92     India     43029044
26    Brazil     29999816
70    France     25997852
76   Germany     21646375
```

```
#13. Plotting the graph for top 5 countries which have a greater
number of covid cases:
plt.figure(figsize=(10,6))
plt.bar(top_5_cases['Country'], top_5_cases['Total Cases'],
color='pink')
plt.title('Top 5 Countries by Total COVID-19 Cases')
plt.xlabel('Country')
plt.ylabel('Total Cases')
```

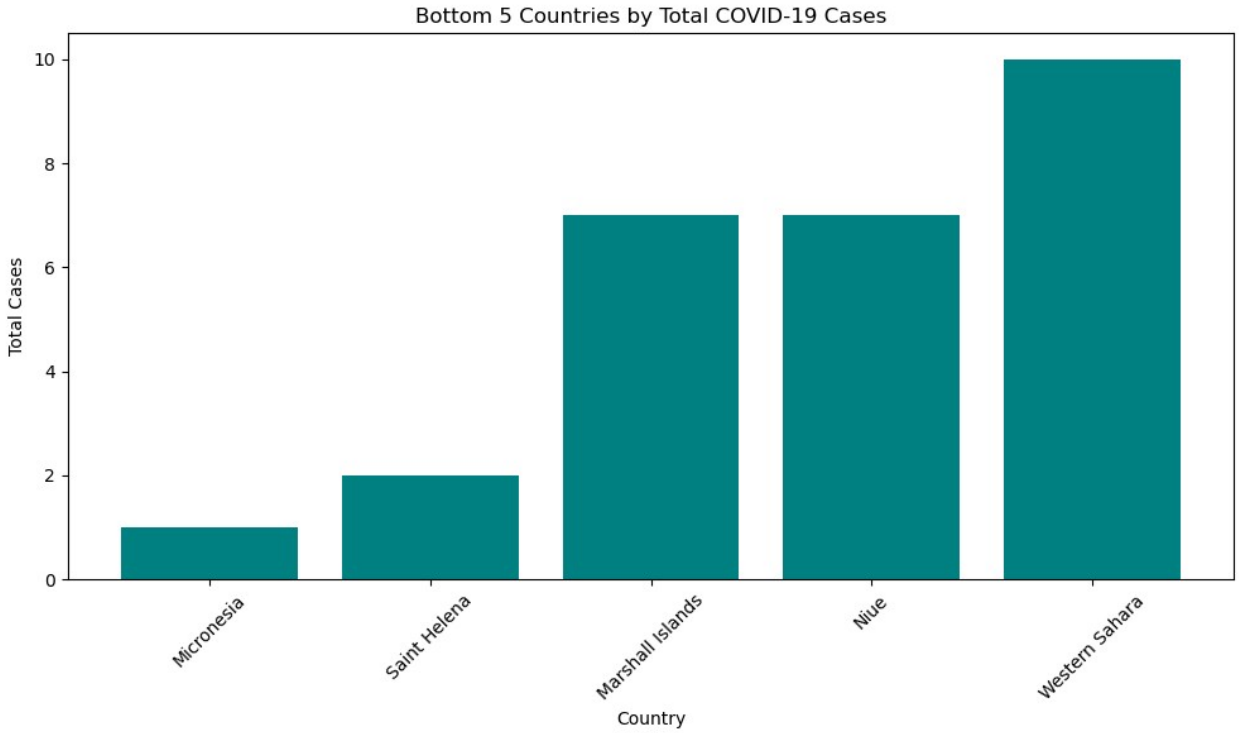
```
plt.xticks(rotation=45)
plt.tight_layout()
```




```
#14 least 5 countries which have least number of covid cases:
least_5_cases = df.nsmallest(5, 'Total Cases')[['Country', 'Total
Cases']]
least_5_cases
```

	Country	Total Cases
131	Micronesia	1
168	Saint Helena	2
125	Marshall Islands	7
148	Niue	7
221	Western Sahara	10

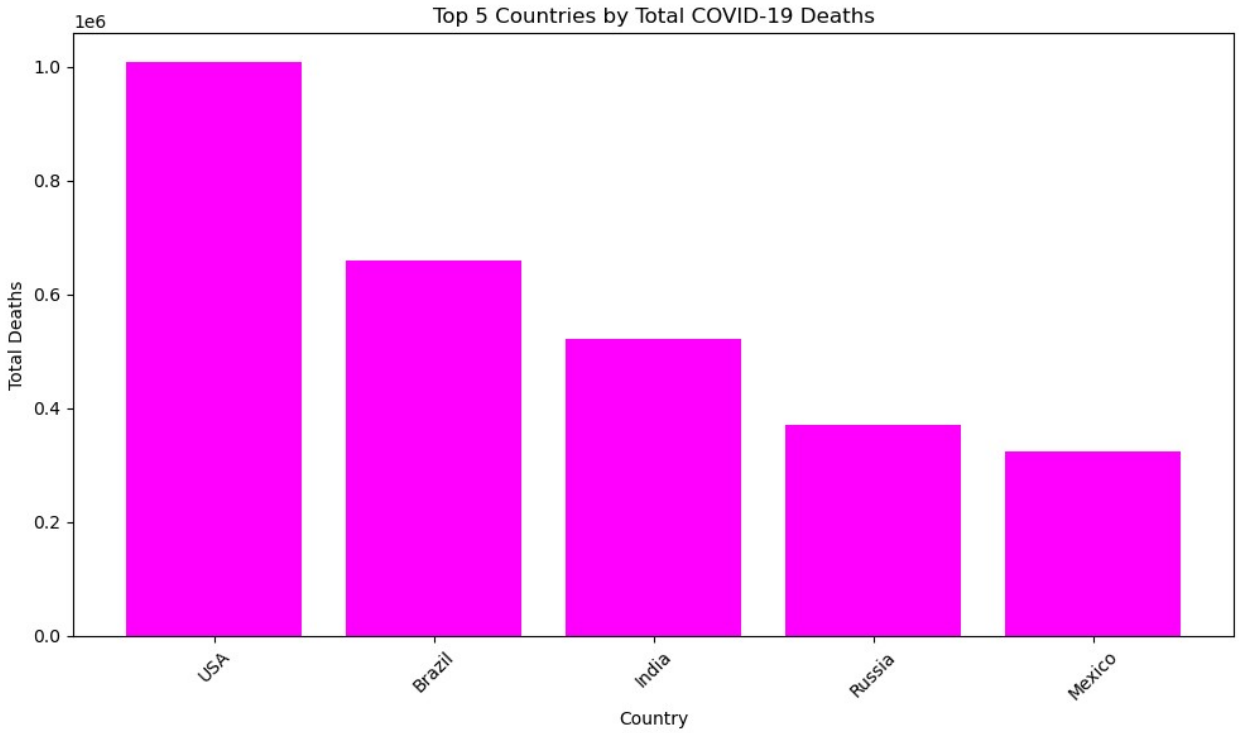
```
#15 Plotting the graph for the least 5 countries which have least
number of covid cases:
plt.figure(figsize=(10,6))
plt.bar(least_5_cases['Country'], least_5_cases['Total Cases'],
color='teal')
plt.title('Bottom 5 Countries by Total COVID-19 Cases')
plt.xlabel('Country')
plt.ylabel('Total Cases')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```




```
#16. Top 5 countries which have a greater number of covid Deaths:
top_5_deaths = df.nlargest(5, 'Total Deaths')[['Country', 'Total
Deaths']]
top_5_deaths
```

	Country	Total Deaths
214	USA	1008222
26	Brazil	660269
92	India	521388
165	Russia	369708
130	Mexico	323212

```
#17 Plotting the graph for top 5 countries which have a greater number
of covid Deaths.
plt.figure(figsize=(10,6))
plt.bar(top_5_deaths['Country'], top_5_deaths['Total Deaths'],
color='magenta')
plt.title('Top 5 Countries by Total COVID-19 Deaths')
plt.xlabel('Country')
plt.ylabel('Total Deaths')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

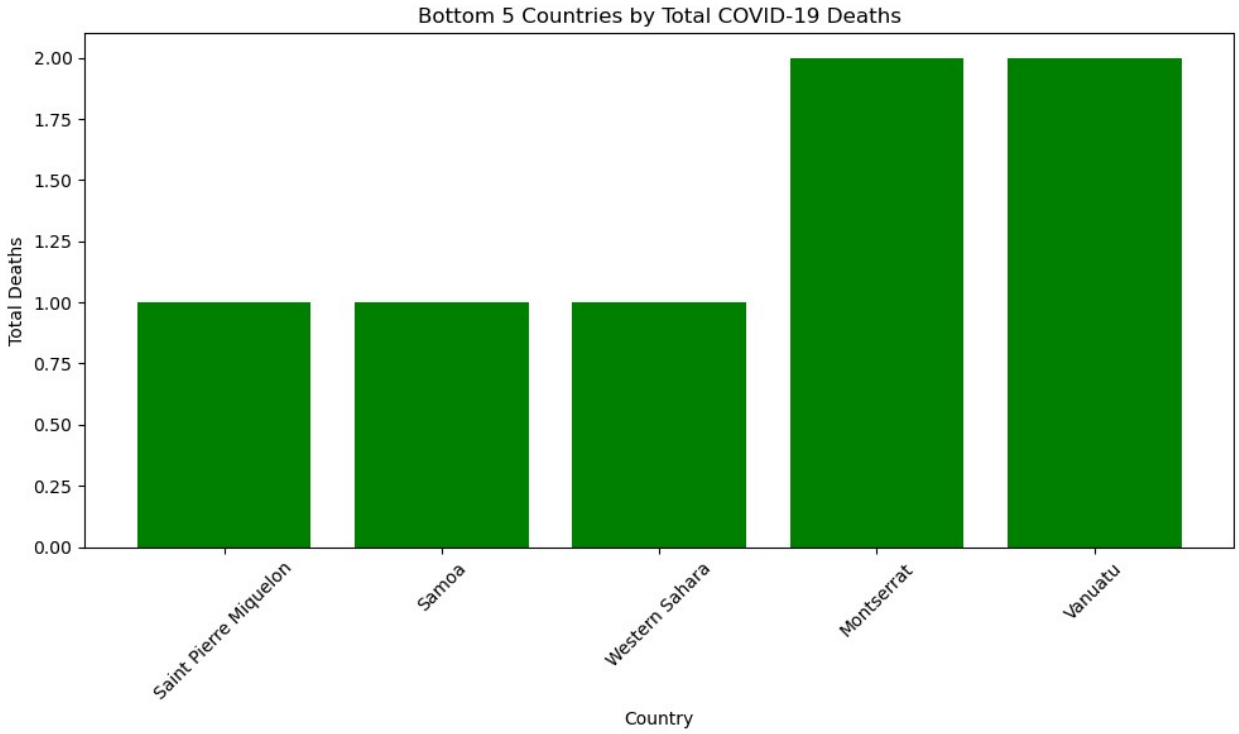



```
#18 Top 5 countries which have least number of covid Deaths:
least_5_deaths = df.nsmallest(5, 'Total Deaths')[['Country', 'Total
Deaths']]
least_5_deaths
```

	Country	Total Deaths
46	Cook Islands	0
67	Falkland Islands	0
118	Macao	0
125	Marshall Islands	0
131	Micronesia	0

```
#19 Plotting the graph for top 5 countries which have least number of
covid Death:
df_nonzero_deaths = df[df['Total Deaths'] > 0]
least_5_deaths = df_nonzero_deaths.nsmallest(5, 'Total Deaths')
[['Country', 'Total Deaths']]
least_5_deaths
plt.figure(figsize=(10,6))
plt.bar(least_5_deaths['Country'], least_5_deaths['Total Deaths'],
color='green')
plt.title('Bottom 5 Countries by Total COVID-19 Deaths')
plt.xlabel('Country')
plt.ylabel('Total Deaths')
plt.xticks(rotation=45)
plt.tight_layout()
```

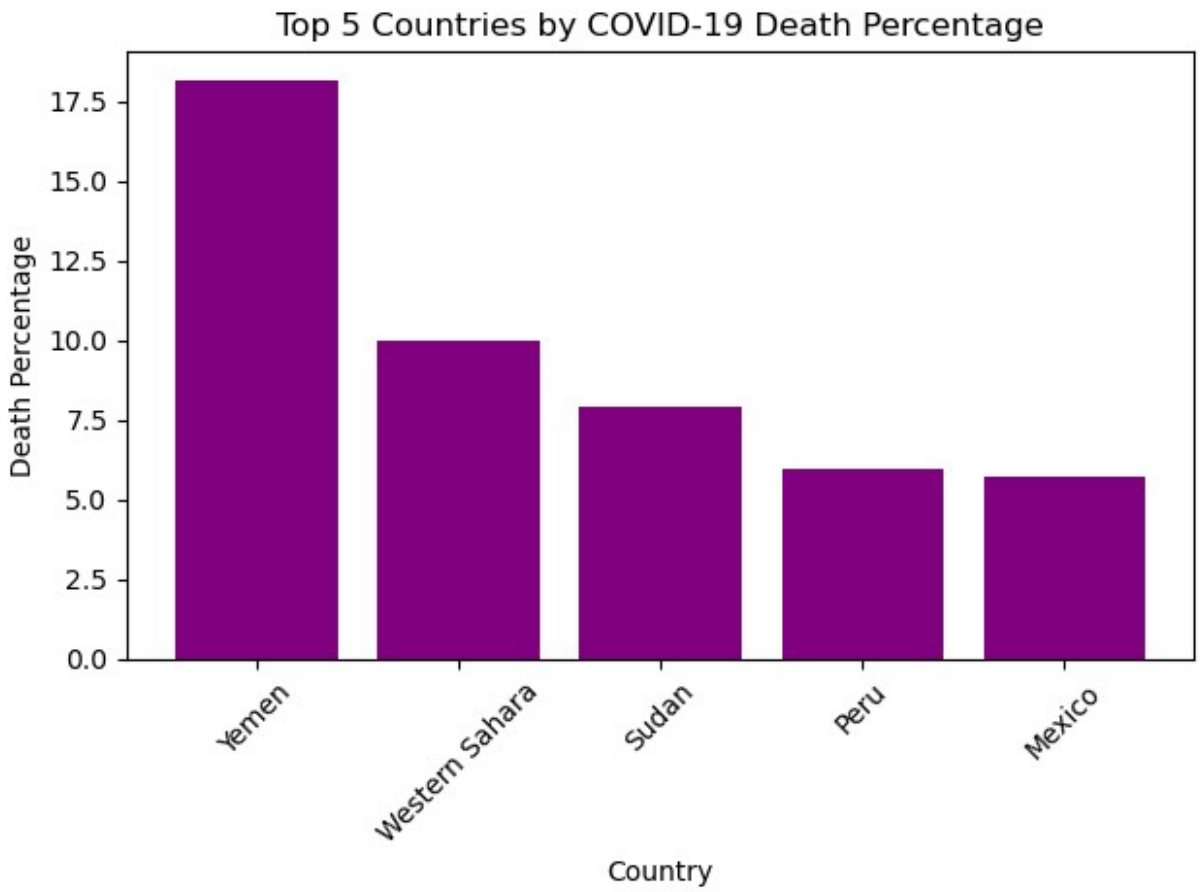
```
plt.show()
```




```
#20 Top 5 countries which have a greater number of covid Death percentage:
top_5_death_pct = df.nlargest(5, 'Death percentage')[['Country', 'Death percentage']]
top_5_death_pct
```

	Country	Death percentage
222	Yemen	18.151787
221	Western Sahara	10.000000
193	Sudan	7.920265
158	Peru	5.983499
130	Mexico	5.705041

```
#21 Plotting graph for top 5 countries which have a greater number of covid Death percentaplt.figure(figsize=(10,6))
plt.bar(top_5_death_pct['Country'], top_5_death_pct['Death percentage'], color='purple')
plt.title('Top 5 Countries by COVID-19 Death Percentage')
plt.xlabel('Country')
plt.ylabel('Death Percentage')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

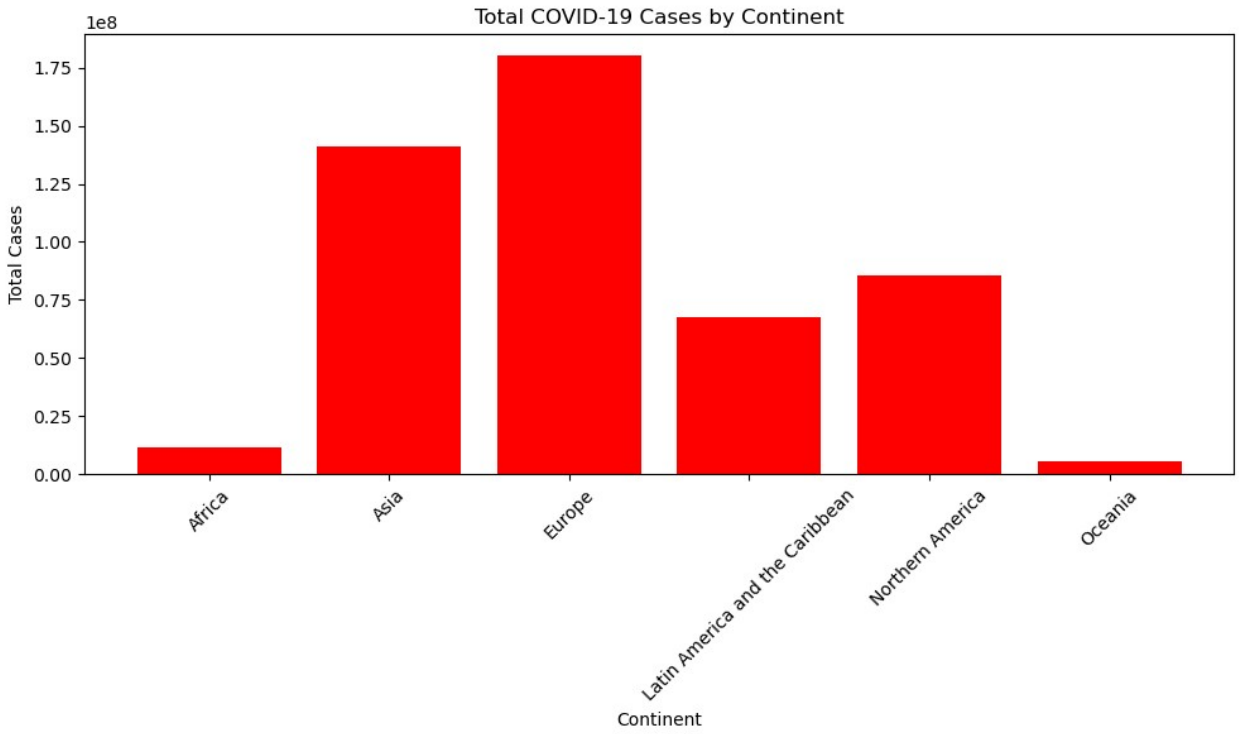



```
#22 Find the continent wise cases (hint: - use group by).
continent_cases = df.groupby('Continent')['Total
Cases'].sum().reset_index()
continent_cases
```

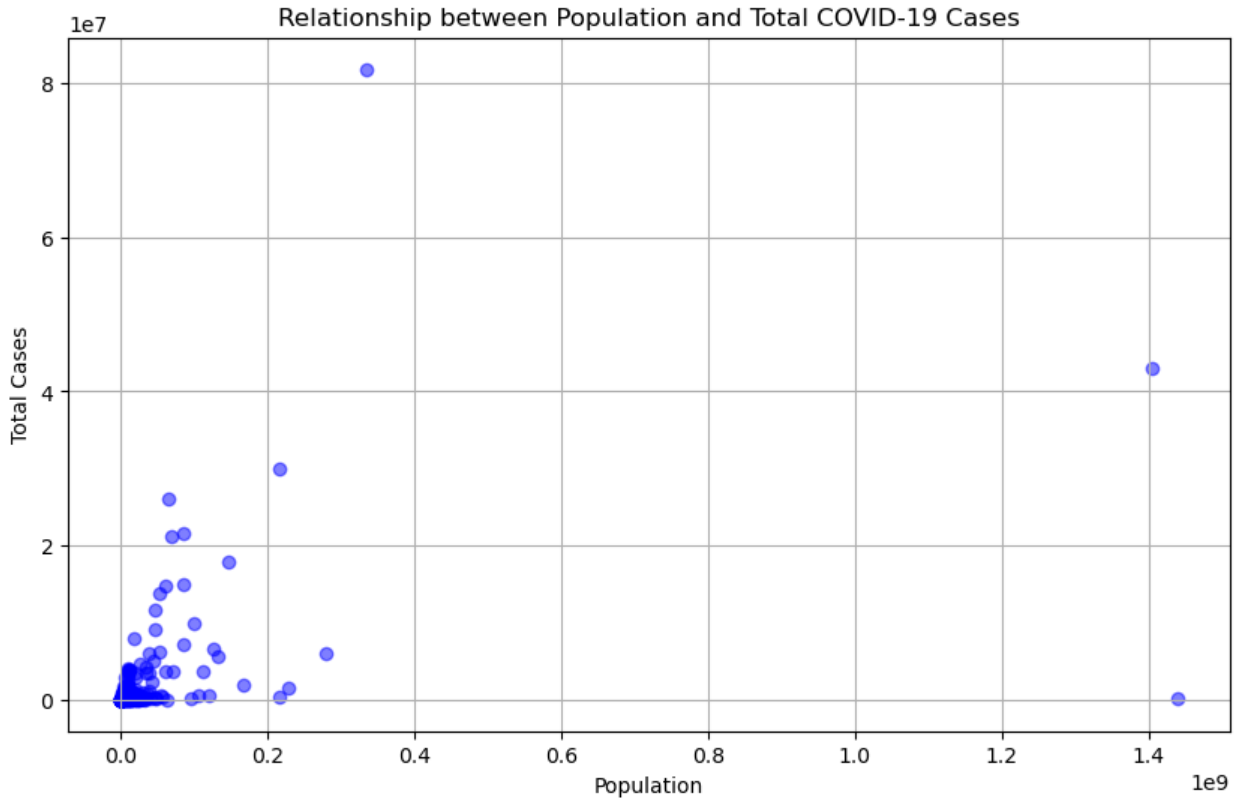
	Continent	Total Cases
0	Africa	11764207
1	Asia	140957179
2	Europe	180332483
3	Latin America and the Caribbean	67509231
4	Northern America	85364770
5	Oceania	5647957

```
#23 Plotting the graph for continent wise case:
plt.figure(figsize=(10,6))
plt.bar(continent_cases['Continent'], continent_cases['Total Cases'],
color='red')
plt.title('Total COVID-19 Cases by Continent')
plt.xlabel('Continent')
plt.ylabel('Total Cases')
plt.xticks(rotation=45)
```

```
plt.tight_layout()
plt.show()
```

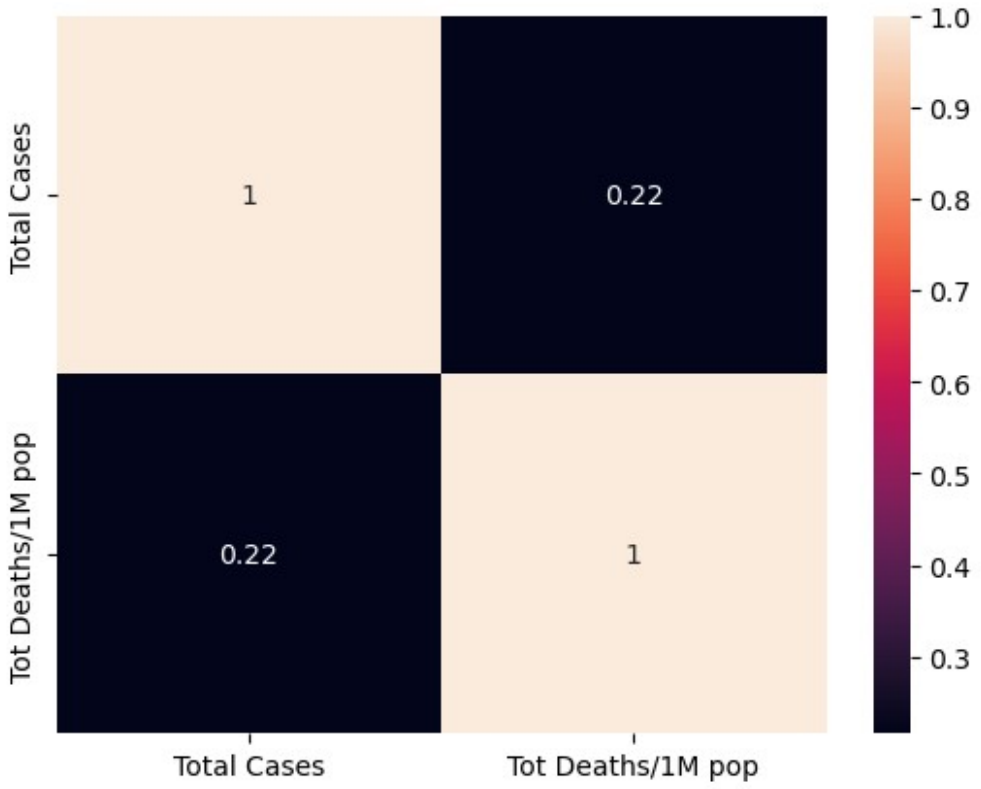



```
#24 Finding relationship between Population and total cases. Comment
on outliers.
plt.figure(figsize=(10,6))
plt.scatter(df['Population'], df['Total Cases'], color='blue',
alpha=0.5)
plt.title('Relationship between Population and Total COVID-19 Cases')
plt.xlabel('Population')
plt.ylabel('Total Cases')
plt.grid(True)
```




```
#25Find the co-relation between Total Cases and Total Deaths/1m pop
and
#What does this  strong positive correlation indicate about the
relationship between Total Cases and  Total Deaths/1m pop across
different countries? (hint: - heatmap)
numeric_df = df1[['Total Cases', 'Tot\xa0Deaths/1M pop']]
sns.heatmap(numeric_df.corr(), annot=True)
```

```
<Axes: >
```




```
# The graph shows 22% relation between Total Cases and Total Deaths
for one million population.
```

OVERVIEW OF THE DATASET:

- The dataset contains 255 rows and 10 columns.
- There is only one null value present.
- The null value is then imputed with value that is most likely to be interpreted by the dataset.

KEY OBSERVATION:

- USA was the leading country for the most cases and death rate .
- The country with least number of cases and deaths was Micronesia but there are many more countries with least number of deaths

INSIGHTS FROM VISUALIZATION:

- Due to unavailability of Recovery and Fatality rate analysis cant be conducted on such parameters and no insights or conclusions can be made.

CONCLUSION:

- Countries with more population were affected the most by the virus.
- Countries that had less resources faced more problems.
- Due to high population , the virus spread faster.

- Countries with less devolopment didn't have enough facilities to control the spread .
- Both highly populated and less developed countries showed the same trend in cases and death rate.
- Recovery and fatality rate couldn't be analyzed due to lack of data.