

olympics-2020

June 25, 2024

****Olympic Games Analytics Project Using Python and Pandas****

1 Olympics medal count 2021: Final table of gold, silver, bronze medals in overall standings

```
[185]: from IPython import display
display.Image("/content/drive/MyDrive/Data_Engineering/Tokyo Olympics_
↳2020(2021)/
↳vector-set-pictograms-summer-olympic-games-saint-petersburg-russia-june-sport-icons-logos-t
↳jpeg", width = 1000, height = 400)
```

[185]:



[185]:

```
[ ]: #pip install pywedge
!pip install pywedge
```

```
[187]: import pandas as pd
import pywedge as pw
import seaborn as sns
```

2 Which country has won the most Olympic medals?

```
[188]: # Import library python

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[189]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
[190]: # Import Excel files

dfom = pd.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo Olympics_
↳2020(2021)/Olympic Medal Count 2020.xlsx',
                    sheet_name='Olympic Medal Count 2020')

dfom.set_index('Team/NOC', inplace=True)
```

```
[190]:
```

```
[191]: dfom
```

```
[191]:
```

	Rank	Gold Medal	Silver Medal	Bronze Medal	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	
...	
Ghana	86.0	0.0	0.0	1.0	
Grenada	86.0	0.0	0.0	1.0	
Kuwait	86.0	0.0	0.0	1.0	
Republic of Moldova	86.0	0.0	0.0	1.0	
Syrian Arab Republic	86.0	0.0	0.0	1.0	

	Total	Rank by Total
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
...
Ghana	1.0	77.0
Grenada	1.0	77.0
Kuwait	1.0	77.0
Republic of Moldova	1.0	77.0
Syrian Arab Republic	1.0	77.0

[93 rows x 6 columns]

```
[192]: ## conversion: Columns to string
## https://www.codeforests.com/2020/08/16/pandas-format-column-headers/
dfom.columns = list(map(str, dfom.columns)) # or

dfom.columns = dfom.columns.astype("str") # or

#Index.str method
dfom.columns = dfom.columns.str.upper() # or

#Index.map method
dfom.columns = dfom.columns.map(str.upper) # or

#Python built-in map method
dfom.columns = map(str.upper, dfom.columns)
```

```
[193]: df_check = dfom.head()
df_check
```

```
[193]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0

Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0

```
[194]: # Information about DataFrame including the index dtype and column dtypes,
      ↪non-null values and memory usage
df_check.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 5 entries, United States of America to ROC
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   RANK                   5 non-null     float64
1   GOLD MEDAL             5 non-null     float64
2   SILVER MEDAL            5 non-null     float64
3   BRONZE MEDAL            5 non-null     float64
4   TOTAL                   5 non-null     float64
5   RANK BY TOTAL           5 non-null     float64
dtypes: float64(6)
memory usage: 280.0+ bytes
```

3 Descriptive Statistics

```
[195]: dfom[["GOLD MEDAL", "SILVER MEDAL", "BRONZE MEDAL", "TOTAL"]].describe()
```

```
[195]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	TOTAL
count	93.000000	93.000000	93.000000	93.000000
mean	3.655914	3.634409	4.322581	11.612903
std	7.022471	6.626339	6.210372	19.091332
min	0.000000	0.000000	0.000000	1.000000
25%	0.000000	0.000000	1.000000	2.000000
50%	1.000000	1.000000	2.000000	4.000000
75%	3.000000	4.000000	5.000000	11.000000
max	39.000000	41.000000	33.000000	113.000000

4 Select rows in a DataFrame based on column values.

```
[196]: #
dfom[(dfom['GOLD MEDAL'] > 3.655914) & (dfom['SILVER MEDAL'] > 3.634409) &
      ↪(dfom['BRONZE MEDAL'] > 4.322581)]
```

```
[196]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					

United States of America	1.0	39.0	41.0	33.0
People's Republic of China	2.0	38.0	32.0	18.0
Japan	3.0	27.0	14.0	17.0
Great Britain	4.0	22.0	21.0	22.0
ROC	5.0	20.0	28.0	23.0
Australia	6.0	17.0	7.0	22.0
Netherlands	7.0	10.0	12.0	14.0
France	8.0	10.0	12.0	11.0
Germany	9.0	10.0	11.0	16.0
Italy	10.0	10.0	10.0	20.0
Canada	11.0	7.0	6.0	11.0
Brazil	12.0	7.0	6.0	8.0
New Zealand	13.0	7.0	6.0	7.0
Hungary	15.0	6.0	7.0	7.0
Republic of Korea	16.0	6.0	4.0	10.0
Poland	17.0	4.0	5.0	5.0

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
Australia	46.0	6.0
Netherlands	36.0	9.0
France	33.0	10.0
Germany	37.0	8.0
Italy	40.0	7.0
Canada	24.0	11.0
Brazil	21.0	12.0
New Zealand	20.0	13.0
Hungary	20.0	13.0
Republic of Korea	20.0	13.0
Poland	14.0	19.0

[196]:

5 Select the rows by conditions if the values of the rows correspond to the average value of the different medals

[197]:

```
above_mean = dfom[(dfom['GOLD MEDAL'] > 3.655914) & (dfom['SILVER MEDAL'] > 3.
↪634409) & (dfom['BRONZE MEDAL'] > 4.322581)]
above_mean
```

```
[197]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	
Australia	6.0	17.0	7.0	22.0	
Netherlands	7.0	10.0	12.0	14.0	
France	8.0	10.0	12.0	11.0	
Germany	9.0	10.0	11.0	16.0	
Italy	10.0	10.0	10.0	20.0	
Canada	11.0	7.0	6.0	11.0	
Brazil	12.0	7.0	6.0	8.0	
New Zealand	13.0	7.0	6.0	7.0	
Hungary	15.0	6.0	7.0	7.0	
Republic of Korea	16.0	6.0	4.0	10.0	
Poland	17.0	4.0	5.0	5.0	

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
Australia	46.0	6.0
Netherlands	36.0	9.0
France	33.0	10.0
Germany	37.0	8.0
Italy	40.0	7.0
Canada	24.0	11.0
Brazil	21.0	12.0
New Zealand	20.0	13.0
Hungary	20.0	13.0
Republic of Korea	20.0	13.0
Poland	14.0	19.0

```
[198]: above_Mean = above_mean.drop(['RANK', 'TOTAL', 'RANK BY TOTAL'], axis = 1)
```

```
[199]: above_Mean
```

```
[199]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
United States of America	39.0	41.0	33.0
People's Republic of China	38.0	32.0	18.0
Japan	27.0	14.0	17.0

Great Britain	22.0	21.0	22.0
ROC	20.0	28.0	23.0
Australia	17.0	7.0	22.0
Netherlands	10.0	12.0	14.0
France	10.0	12.0	11.0
Germany	10.0	11.0	16.0
Italy	10.0	10.0	20.0
Canada	7.0	6.0	11.0
Brazil	7.0	6.0	8.0
New Zealand	7.0	6.0	7.0
Hungary	6.0	7.0	7.0
Republic of Korea	6.0	4.0	10.0
Poland	4.0	5.0	5.0

[200]: dfom

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	
...	
Ghana	86.0	0.0	0.0	1.0	
Grenada	86.0	0.0	0.0	1.0	
Kuwait	86.0	0.0	0.0	1.0	
Republic of Moldova	86.0	0.0	0.0	1.0	
Syrian Arab Republic	86.0	0.0	0.0	1.0	
	TOTAL	RANK BY TOTAL			
Team/NOC					
United States of America	113.0	1.0			
People's Republic of China	88.0	2.0			
Japan	58.0	5.0			
Great Britain	65.0	4.0			
ROC	71.0	3.0			
...			
Ghana	1.0	77.0			
Grenada	1.0	77.0			
Kuwait	1.0	77.0			
Republic of Moldova	1.0	77.0			
Syrian Arab Republic	1.0	77.0			

[93 rows x 6 columns]

```
[201]: # "GOLD MEDAL", "SILVER MEDAL", "BRONZE MEDAL", "TOTAL"
dfom[(dfom['GOLD MEDAL'] < 3.655914) & (dfom['SILVER MEDAL'] < 3.634409) &
↪(dfom['BRONZE MEDAL'] < 4.322581)]
```

```
[201]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	TOTAL	\
Team/NOC						
Croatia	26.0	3.0	3.0	2.0	8.0	
Islamic Republic of Iran	27.0	3.0	2.0	2.0	7.0	
Belgium	29.0	3.0	1.0	3.0	7.0	
Bulgaria	30.0	3.0	1.0	2.0	6.0	
Slovenia	31.0	3.0	1.0	1.0	5.0	
Uzbekistan	32.0	3.0	0.0	2.0	5.0	
Greece	36.0	2.0	1.0	1.0	4.0	
Uganda	36.0	2.0	1.0	1.0	4.0	
Ecuador	38.0	2.0	1.0	0.0	3.0	
Ireland	39.0	2.0	0.0	2.0	4.0	
Israel	39.0	2.0	0.0	2.0	4.0	
Qatar	41.0	2.0	0.0	1.0	3.0	
Bahamas	42.0	2.0	0.0	0.0	2.0	
Kosovo	42.0	2.0	0.0	0.0	2.0	
Belarus	45.0	1.0	3.0	3.0	7.0	
Romania	46.0	1.0	3.0	0.0	4.0	
Venezuela	46.0	1.0	3.0	0.0	4.0	
India	48.0	1.0	2.0	4.0	7.0	
Hong Kong, China	49.0	1.0	2.0	3.0	6.0	
Philippines	50.0	1.0	2.0	1.0	4.0	
Slovakia	50.0	1.0	2.0	1.0	4.0	
South Africa	52.0	1.0	2.0	0.0	3.0	
Egypt	54.0	1.0	1.0	4.0	6.0	
Indonesia	55.0	1.0	1.0	3.0	5.0	
Ethiopia	56.0	1.0	1.0	2.0	4.0	
Portugal	56.0	1.0	1.0	2.0	4.0	
Tunisia	58.0	1.0	1.0	0.0	2.0	
Estonia	59.0	1.0	0.0	1.0	2.0	
Fiji	59.0	1.0	0.0	1.0	2.0	
Latvia	59.0	1.0	0.0	1.0	2.0	
Thailand	59.0	1.0	0.0	1.0	2.0	
Bermuda	63.0	1.0	0.0	0.0	1.0	
Morocco	63.0	1.0	0.0	0.0	1.0	
Puerto Rico	63.0	1.0	0.0	0.0	1.0	
Azerbaijan	67.0	0.0	3.0	4.0	7.0	
Dominican Republic	68.0	0.0	3.0	2.0	5.0	
Armenia	69.0	0.0	2.0	2.0	4.0	
Kyrgyzstan	70.0	0.0	2.0	1.0	3.0	
Mongolia	71.0	0.0	1.0	3.0	4.0	
Argentina	72.0	0.0	1.0	2.0	3.0	
San Marino	72.0	0.0	1.0	2.0	3.0	

Jordan	74.0	0.0	1.0	1.0	2.0
Malaysia	74.0	0.0	1.0	1.0	2.0
Nigeria	74.0	0.0	1.0	1.0	2.0
Bahrain	77.0	0.0	1.0	0.0	1.0
Saudi Arabia	77.0	0.0	1.0	0.0	1.0
Lithuania	77.0	0.0	1.0	0.0	1.0
North Macedonia	77.0	0.0	1.0	0.0	1.0
Namibia	77.0	0.0	1.0	0.0	1.0
Turkmenistan	77.0	0.0	1.0	0.0	1.0
Mexico	84.0	0.0	0.0	4.0	4.0
Finland	85.0	0.0	0.0	2.0	2.0
Botswana	86.0	0.0	0.0	1.0	1.0
Burkina Faso	86.0	0.0	0.0	1.0	1.0
Côte d'Ivoire	86.0	0.0	0.0	1.0	1.0
Ghana	86.0	0.0	0.0	1.0	1.0
Grenada	86.0	0.0	0.0	1.0	1.0
Kuwait	86.0	0.0	0.0	1.0	1.0
Republic of Moldova	86.0	0.0	0.0	1.0	1.0
Syrian Arab Republic	86.0	0.0	0.0	1.0	1.0

RANK BY TOTAL

Team/NOC	
Croatia	29.0
Islamic Republic of Iran	33.0
Belgium	33.0
Bulgaria	39.0
Slovenia	42.0
Uzbekistan	42.0
Greece	47.0
Uganda	47.0
Ecuador	60.0
Ireland	47.0
Israel	47.0
Qatar	60.0
Bahamas	66.0
Kosovo	66.0
Belarus	33.0
Romania	47.0
Venezuela	47.0
India	33.0
Hong Kong, China	39.0
Philippines	47.0
Slovakia	47.0
South Africa	60.0
Egypt	39.0
Indonesia	42.0
Ethiopia	47.0

Portugal	47.0
Tunisia	66.0
Estonia	66.0
Fiji	66.0
Latvia	66.0
Thailand	66.0
Bermuda	77.0
Morocco	77.0
Puerto Rico	77.0
Azerbaijan	33.0
Dominican Republic	42.0
Armenia	47.0
Kyrgyzstan	60.0
Mongolia	47.0
Argentina	60.0
San Marino	60.0
Jordan	66.0
Malaysia	66.0
Nigeria	66.0
Bahrain	77.0
Saudi Arabia	77.0
Lithuania	77.0
North Macedonia	77.0
Namibia	77.0
Turkmenistan	77.0
Mexico	47.0
Finland	66.0
Botswana	77.0
Burkina Faso	77.0
Côte d'Ivoire	77.0
Ghana	77.0
Grenada	77.0
Kuwait	77.0
Republic of Moldova	77.0
Syrian Arab Republic	77.0

```
[202]: below_mean = dfom[(dfom['GOLD MEDAL'] < 3.655914) & (dfom['SILVER MEDAL'] < 3.
↪634409) & (dfom['BRONZE MEDAL'] < 4.322581)]
```

```
[203]: below_Mean = below_mean.drop(['RANK', 'TOTAL', 'RANK BY TOTAL'], axis = 1)
```

```
[204]: below_Mean
```

```
[204]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
Croatia	3.0	3.0	2.0
Islamic Republic of Iran	3.0	2.0	2.0

Belgium	3.0	1.0	3.0
Bulgaria	3.0	1.0	2.0
Slovenia	3.0	1.0	1.0
Uzbekistan	3.0	0.0	2.0
Greece	2.0	1.0	1.0
Uganda	2.0	1.0	1.0
Ecuador	2.0	1.0	0.0
Ireland	2.0	0.0	2.0
Israel	2.0	0.0	2.0
Qatar	2.0	0.0	1.0
Bahamas	2.0	0.0	0.0
Kosovo	2.0	0.0	0.0
Belarus	1.0	3.0	3.0
Romania	1.0	3.0	0.0
Venezuela	1.0	3.0	0.0
India	1.0	2.0	4.0
Hong Kong, China	1.0	2.0	3.0
Philippines	1.0	2.0	1.0
Slovakia	1.0	2.0	1.0
South Africa	1.0	2.0	0.0
Egypt	1.0	1.0	4.0
Indonesia	1.0	1.0	3.0
Ethiopia	1.0	1.0	2.0
Portugal	1.0	1.0	2.0
Tunisia	1.0	1.0	0.0
Estonia	1.0	0.0	1.0
Fiji	1.0	0.0	1.0
Latvia	1.0	0.0	1.0
Thailand	1.0	0.0	1.0
Bermuda	1.0	0.0	0.0
Morocco	1.0	0.0	0.0
Puerto Rico	1.0	0.0	0.0
Azerbaijan	0.0	3.0	4.0
Dominican Republic	0.0	3.0	2.0
Armenia	0.0	2.0	2.0
Kyrgyzstan	0.0	2.0	1.0
Mongolia	0.0	1.0	3.0
Argentina	0.0	1.0	2.0
San Marino	0.0	1.0	2.0
Jordan	0.0	1.0	1.0
Malaysia	0.0	1.0	1.0
Nigeria	0.0	1.0	1.0
Bahrain	0.0	1.0	0.0
Saudi Arabia	0.0	1.0	0.0
Lithuania	0.0	1.0	0.0
North Macedonia	0.0	1.0	0.0
Namibia	0.0	1.0	0.0

Turkmenistan	0.0	1.0	0.0
Mexico	0.0	0.0	4.0
Finland	0.0	0.0	2.0
Botswana	0.0	0.0	1.0
Burkina Faso	0.0	0.0	1.0
Côte d'Ivoire	0.0	0.0	1.0
Ghana	0.0	0.0	1.0
Grenada	0.0	0.0	1.0
Kuwait	0.0	0.0	1.0
Republic of Moldova	0.0	0.0	1.0
Syrian Arab Republic	0.0	0.0	1.0

```
[205]: mean = dfom[(dfom['GOLD MEDAL'] == 3.655914) & (dfom['SILVER MEDAL'] == 3.
↳634409) & (dfom['BRONZE MEDAL'] == 4.322581)]
```

```
[206]: mean # No country has the average number of medals compared to the 3 types of
↳medals
```

```
[206]: Empty DataFrame
Columns: [RANK, GOLD MEDAL, SILVER MEDAL, BRONZE MEDAL, TOTAL, RANK BY TOTAL]
Index: []
```

```
[207]: atleast_one_mean = dfom[(dfom['GOLD MEDAL'] > 3.655914) | (dfom['SILVER MEDAL']
↳> 3.634409) | (dfom['BRONZE MEDAL'] > 4.322581)]
```

```
[208]: atleast_one_mean
```

```
[208]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	
Australia	6.0	17.0	7.0	22.0	
Netherlands	7.0	10.0	12.0	14.0	
France	8.0	10.0	12.0	11.0	
Germany	9.0	10.0	11.0	16.0	
Italy	10.0	10.0	10.0	20.0	
Canada	11.0	7.0	6.0	11.0	
Brazil	12.0	7.0	6.0	8.0	
New Zealand	13.0	7.0	6.0	7.0	
Cuba	14.0	7.0	3.0	5.0	
Hungary	15.0	6.0	7.0	7.0	
Republic of Korea	16.0	6.0	4.0	10.0	
Poland	17.0	4.0	5.0	5.0	
Czech Republic	18.0	4.0	4.0	3.0	

Kenya	19.0	4.0	4.0	2.0
Norway	20.0	4.0	2.0	2.0
Jamaica	21.0	4.0	1.0	4.0
Spain	22.0	3.0	8.0	6.0
Sweden	23.0	3.0	6.0	0.0
Switzerland	24.0	3.0	4.0	6.0
Denmark	25.0	3.0	4.0	4.0
Serbia	28.0	3.0	1.0	5.0
Georgia	33.0	2.0	5.0	1.0
Chinese Taipei	34.0	2.0	4.0	6.0
Turkey	35.0	2.0	2.0	9.0
Ukraine	44.0	1.0	6.0	12.0
Austria	53.0	1.0	1.0	5.0
Colombia	66.0	0.0	4.0	1.0
Kazakhstan	83.0	0.0	0.0	8.0

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
Australia	46.0	6.0
Netherlands	36.0	9.0
France	33.0	10.0
Germany	37.0	8.0
Italy	40.0	7.0
Canada	24.0	11.0
Brazil	21.0	12.0
New Zealand	20.0	13.0
Cuba	15.0	18.0
Hungary	20.0	13.0
Republic of Korea	20.0	13.0
Poland	14.0	19.0
Czech Republic	11.0	23.0
Kenya	10.0	25.0
Norway	8.0	29.0
Jamaica	9.0	26.0
Spain	17.0	17.0
Sweden	9.0	26.0
Switzerland	13.0	20.0
Denmark	11.0	23.0
Serbia	9.0	26.0
Georgia	8.0	29.0
Chinese Taipei	12.0	22.0
Turkey	13.0	20.0

Ukraine	19.0	16.0
Austria	7.0	33.0
Colombia	5.0	42.0
Kazakhstan	8.0	29.0

```
[209]: atleast_one_Mean = atleast_one_mean.drop(['RANK', 'TOTAL', 'RANK BY TOTAL'],
↪axis = 1)
```

```
[210]: atleast_one_Mean
```

```
[210]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
United States of America	39.0	41.0	33.0
People's Republic of China	38.0	32.0	18.0
Japan	27.0	14.0	17.0
Great Britain	22.0	21.0	22.0
ROC	20.0	28.0	23.0
Australia	17.0	7.0	22.0
Netherlands	10.0	12.0	14.0
France	10.0	12.0	11.0
Germany	10.0	11.0	16.0
Italy	10.0	10.0	20.0
Canada	7.0	6.0	11.0
Brazil	7.0	6.0	8.0
New Zealand	7.0	6.0	7.0
Cuba	7.0	3.0	5.0
Hungary	6.0	7.0	7.0
Republic of Korea	6.0	4.0	10.0
Poland	4.0	5.0	5.0
Czech Republic	4.0	4.0	3.0
Kenya	4.0	4.0	2.0
Norway	4.0	2.0	2.0
Jamaica	4.0	1.0	4.0
Spain	3.0	8.0	6.0
Sweden	3.0	6.0	0.0
Switzerland	3.0	4.0	6.0
Denmark	3.0	4.0	4.0
Serbia	3.0	1.0	5.0
Georgia	2.0	5.0	1.0
Chinese Taipei	2.0	4.0	6.0
Turkey	2.0	2.0	9.0
Ukraine	1.0	6.0	12.0
Austria	1.0	1.0	5.0
Colombia	0.0	4.0	1.0
Kazakhstan	0.0	0.0	8.0

```
[211]: atleast_two_above_mean = dfom[(dfom['GOLD MEDAL'] > 3.655914) & (dfom['SILVER_
MEDAL'] > 3.634409)]
```

```
[212]: atleast_two_above_mean
```

```
[212]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL \
Team/NOC				
United States of America	1.0	39.0	41.0	33.0
People's Republic of China	2.0	38.0	32.0	18.0
Japan	3.0	27.0	14.0	17.0
Great Britain	4.0	22.0	21.0	22.0
ROC	5.0	20.0	28.0	23.0
Australia	6.0	17.0	7.0	22.0
Netherlands	7.0	10.0	12.0	14.0
France	8.0	10.0	12.0	11.0
Germany	9.0	10.0	11.0	16.0
Italy	10.0	10.0	10.0	20.0
Canada	11.0	7.0	6.0	11.0
Brazil	12.0	7.0	6.0	8.0
New Zealand	13.0	7.0	6.0	7.0
Hungary	15.0	6.0	7.0	7.0
Republic of Korea	16.0	6.0	4.0	10.0
Poland	17.0	4.0	5.0	5.0
Czech Republic	18.0	4.0	4.0	3.0
Kenya	19.0	4.0	4.0	2.0

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
Australia	46.0	6.0
Netherlands	36.0	9.0
France	33.0	10.0
Germany	37.0	8.0
Italy	40.0	7.0
Canada	24.0	11.0
Brazil	21.0	12.0
New Zealand	20.0	13.0
Hungary	20.0	13.0
Republic of Korea	20.0	13.0
Poland	14.0	19.0
Czech Republic	11.0	23.0
Kenya	10.0	25.0

```
[213]: atleast_two_above_Mean = atleast_two_above_mean .drop(['RANK', 'TOTAL', 'RANK_
↳BY TOTAL'], axis = 1)
atleast_two_above_Mean
```

```
[213]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
United States of America	39.0	41.0	33.0
People's Republic of China	38.0	32.0	18.0
Japan	27.0	14.0	17.0
Great Britain	22.0	21.0	22.0
ROC	20.0	28.0	23.0
Australia	17.0	7.0	22.0
Netherlands	10.0	12.0	14.0
France	10.0	12.0	11.0
Germany	10.0	11.0	16.0
Italy	10.0	10.0	20.0
Canada	7.0	6.0	11.0
Brazil	7.0	6.0	8.0
New Zealand	7.0	6.0	7.0
Hungary	6.0	7.0	7.0
Republic of Korea	6.0	4.0	10.0
Poland	4.0	5.0	5.0
Czech Republic	4.0	4.0	3.0
Kenya	4.0	4.0	2.0

```
[214]: atleast_two_below_mean = dfom[(dfom['GOLD MEDAL'] < 3.655914) & (dfom['SILVER_
↳MEDAL'] < 3.634409)]
```

```
[215]: atleast_two_below_mean
```

```
[215]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	TOTAL	\
Team/NOC						
Croatia	26.0	3.0	3.0	2.0	8.0	
Islamic Republic of Iran	27.0	3.0	2.0	2.0	7.0	
Serbia	28.0	3.0	1.0	5.0	9.0	
Belgium	29.0	3.0	1.0	3.0	7.0	
Bulgaria	30.0	3.0	1.0	2.0	6.0	
...	
Ghana	86.0	0.0	0.0	1.0	1.0	
Grenada	86.0	0.0	0.0	1.0	1.0	
Kuwait	86.0	0.0	0.0	1.0	1.0	
Republic of Moldova	86.0	0.0	0.0	1.0	1.0	
Syrian Arab Republic	86.0	0.0	0.0	1.0	1.0	

```
RANK BY TOTAL

Team/NOC
Croatia
```


Islamic Republic of Iran	33.0
Serbia	26.0
Belgium	33.0
Bulgaria	39.0
...	...
Ghana	77.0
Grenada	77.0
Kuwait	77.0
Republic of Moldova	77.0
Syrian Arab Republic	77.0

[64 rows x 6 columns]

```
[216]: atleast_two_below_Mean = atleast_two_below_mean .drop(['RANK', 'TOTAL', 'RANK_↓
↳BY TOTAL'], axis = 1)
atleast_two_below_Mean
```

```
[216]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
Croatia	3.0	3.0	2.0
Islamic Republic of Iran	3.0	2.0	2.0
Serbia	3.0	1.0	5.0
Belgium	3.0	1.0	3.0
Bulgaria	3.0	1.0	2.0
...
Ghana	0.0	0.0	1.0
Grenada	0.0	0.0	1.0
Kuwait	0.0	0.0	1.0
Republic of Moldova	0.0	0.0	1.0
Syrian Arab Republic	0.0	0.0	1.0

[64 rows x 3 columns]

```
[217]: atleast_two_below_Mean.head(20)
```

```
[217]:
```

	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL
Team/NOC			
Croatia	3.0	3.0	2.0
Islamic Republic of Iran	3.0	2.0	2.0
Serbia	3.0	1.0	5.0
Belgium	3.0	1.0	3.0
Bulgaria	3.0	1.0	2.0
Slovenia	3.0	1.0	1.0
Uzbekistan	3.0	0.0	2.0
Turkey	2.0	2.0	9.0
Greece	2.0	1.0	1.0
Uganda	2.0	1.0	1.0

Ecuador	2.0	1.0	0.0
Ireland	2.0	0.0	2.0
Israel	2.0	0.0	2.0
Qatar	2.0	0.0	1.0
Bahamas	2.0	0.0	0.0
Kosovo	2.0	0.0	0.0
Belarus	1.0	3.0	3.0
Romania	1.0	3.0	0.0
Venezuela	1.0	3.0	0.0
India	1.0	2.0	4.0

6 Top 10 countries by total number of medals

```
[218]: dfom.head(10)
```

```
[218]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	\
Team/NOC					
United States of America	1.0	39.0	41.0	33.0	
People's Republic of China	2.0	38.0	32.0	18.0	
Japan	3.0	27.0	14.0	17.0	
Great Britain	4.0	22.0	21.0	22.0	
ROC	5.0	20.0	28.0	23.0	
Australia	6.0	17.0	7.0	22.0	
Netherlands	7.0	10.0	12.0	14.0	
France	8.0	10.0	12.0	11.0	
Germany	9.0	10.0	11.0	16.0	
Italy	10.0	10.0	10.0	20.0	

	TOTAL	RANK BY TOTAL
Team/NOC		
United States of America	113.0	1.0
People's Republic of China	88.0	2.0
Japan	58.0	5.0
Great Britain	65.0	4.0
ROC	71.0	3.0
Australia	46.0	6.0
Netherlands	36.0	9.0
France	33.0	10.0
Germany	37.0	8.0
Italy	40.0	7.0

```
[219]: dfomtop10_by_value_of_medals = dfom.head(10).drop(["RANK", "GOLD MEDAL", "SILVER_
↳MEDAL", "BRONZE MEDAL", "RANK BY TOTAL"], axis = 1)
```

```
[220]: dfomtop10_by_value_of_medals # Ranking by value (not total quantity) of medals_
↳won by the 10 First Nations.
```

```
[220]:
```

	TOTAL
Team/NOC	
United States of America	113.0
People's Republic of China	88.0
Japan	58.0
Great Britain	65.0
ROC	71.0
Australia	46.0
Netherlands	36.0
France	33.0
Germany	37.0
Italy	40.0

```
[221]: # Ranking by total quantity (not by value) of medals won by each of the 10
        ↪First Nations.
        # df.sort_values(['a', 'b'], ascending=[True, False])

dfomtop10_by_quantity_of_medals = dfomtop10_by_value_of_medals .
        ↪sort_values(["TOTAL"], ascending=[False])
dfomtop10_by_quantity_of_medals
```

```
[221]:
```

	TOTAL
Team/NOC	
United States of America	113.0
People's Republic of China	88.0
ROC	71.0
Great Britain	65.0
Japan	58.0
Australia	46.0
Italy	40.0
Germany	37.0
Netherlands	36.0
France	33.0

7 Visualization of the Top 10 countries by total number of medals

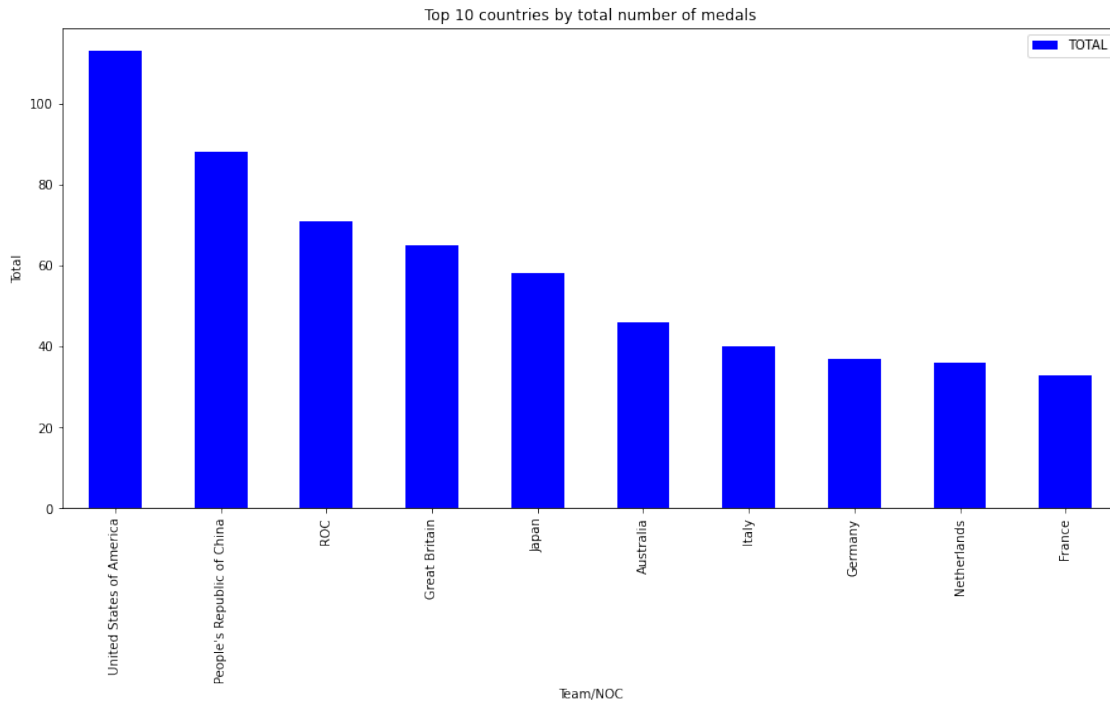
```
[222]: dfomtop10_by_quantity_of_medals.plot(kind = 'bar',color = 'b',figsize = (15,7))

plt.title("Top 10 countries by total number of medals")

plt.xlabel("Team/NOC")

plt.ylabel("Total")

plt.show()
```



[222]:

8 Visualization of the Top 10 countries by Value of medals

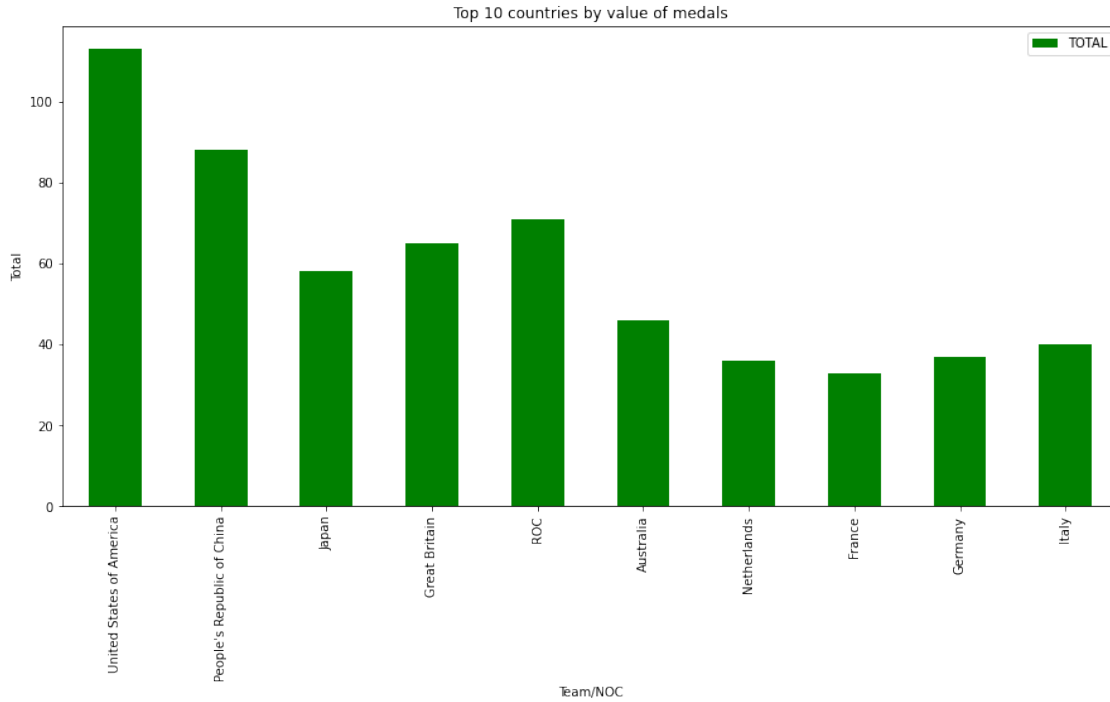
```
[223]: dfomtop10_by_value_of_medals.plot(kind = 'bar',color = 'g',figsize = (15,7))

plt.title("Top 10 countries by value of medals")

plt.xlabel("Team/NOC")

plt.ylabel("Total")

plt.show()
```



9 The last 10 countries by total number of medals

```
[224]: dfom.tail(10)
```

```
[224]:
```

	RANK	GOLD MEDAL	SILVER MEDAL	BRONZE MEDAL	TOTAL \
Team/NOC					
Mexico	84.0	0.0	0.0	4.0	4.0
Finland	85.0	0.0	0.0	2.0	2.0
Botswana	86.0	0.0	0.0	1.0	1.0
Burkina Faso	86.0	0.0	0.0	1.0	1.0
Côte d'Ivoire	86.0	0.0	0.0	1.0	1.0
Ghana	86.0	0.0	0.0	1.0	1.0
Grenada	86.0	0.0	0.0	1.0	1.0
Kuwait	86.0	0.0	0.0	1.0	1.0
Republic of Moldova	86.0	0.0	0.0	1.0	1.0
Syrian Arab Republic	86.0	0.0	0.0	1.0	1.0

```
RANK BY TOTAL
```

Team/NOC	
Mexico	47.0
Finland	66.0
Botswana	77.0
Burkina Faso	77.0

Côte d'Ivoire	77.0
Ghana	77.0
Grenada	77.0
Kuwait	77.0
Republic of Moldova	77.0
Syrian Arab Republic	77.0

```
[225]: dfomlast10_by_value_of_medals = dfom.tail(10).drop(["RANK", "GOLD",
↳MEDAL", "SILVER MEDAL", "BRONZE MEDAL", "RANK BY TOTAL"], axis = 1)
dfomlast10_by_value_of_medals
```

```
[225]:
```

	TOTAL
Team/NOC	
Mexico	4.0
Finland	2.0
Botswana	1.0
Burkina Faso	1.0
Côte d'Ivoire	1.0
Ghana	1.0
Grenada	1.0
Kuwait	1.0
Republic of Moldova	1.0
Syrian Arab Republic	1.0

```
[226]: dfomlast10_by_quantity_of_medals = dfomlast10_by_value_of_medals.
↳sort_values(["TOTAL"], ascending=[False])
dfomlast10_by_quantity_of_medals
```

```
[226]:
```

	TOTAL
Team/NOC	
Mexico	4.0
Finland	2.0
Botswana	1.0
Burkina Faso	1.0
Côte d'Ivoire	1.0
Ghana	1.0
Grenada	1.0
Kuwait	1.0
Republic of Moldova	1.0
Syrian Arab Republic	1.0

10 Visualization of the last 10 countries by total number of medals = Visualization of the last 10 countries by total value of medals

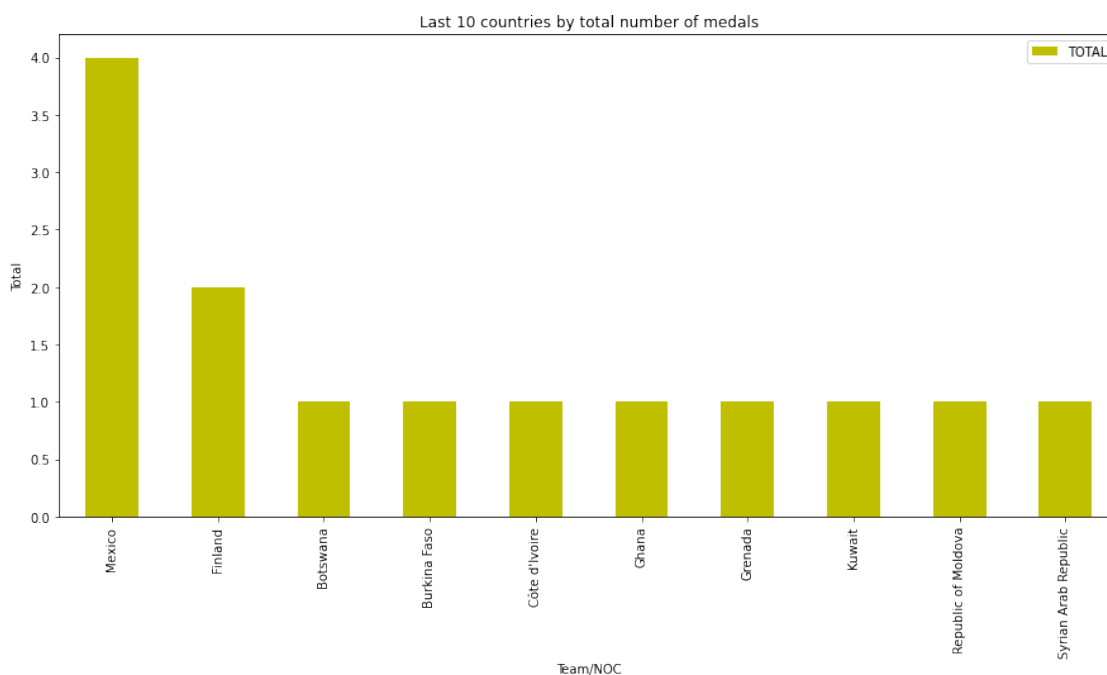
```
[227]: dfomlast10_by_quantity_of_medals .plot(kind = 'bar',color ='y',figsize = (15,7))

plt.title("Last 10 countries by total number of medals")

plt.xlabel("Team/NOC")

plt.ylabel("Total")

plt.show()
```



11 Analyze the data in relation to the different continents that were present at the Tokyo 2020 Olympic Games

```
[228]: import numpy as np
import pandas as pdsa
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[229]: dfosa = pdsa.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo Olympics_
↳2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='South American countries')
```

```
[230]: dfosa
```

```
[230]:
```

	COUNTRIES	ATHLETES	GOLD	SILVER	BRONZE
0	Argentina	174.0	0.0	1.0	2.0
1	Aruba	3.0	0.0	0.0	0.0
2	Bolivia	5.0	0.0	0.0	0.0
3	Brazil	301.0	7.0	6.0	8.0
4	Chile	57.0	0.0	0.0	0.0
5	Colombia	70.0	0.0	4.0	1.0
6	Ecuador	48.0	2.0	1.0	0.0
7	Guyana	7.0	0.0	0.0	0.0
8	Paraguay	8.0	0.0	0.0	0.0
9	Peru	35.0	0.0	0.0	0.0
10	Suriname	3.0	0.0	0.0	0.0
11	Uruguay	11.0	0.0	0.0	0.0
12	Venezuela	44.0	1.0	3.0	0.0

```
[231]: dfosa.set_index('COUNTRIES', inplace=True)
```

```
[232]: dfosa
```

```
[232]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Argentina	174.0	0.0	1.0	2.0
Aruba	3.0	0.0	0.0	0.0
Bolivia	5.0	0.0	0.0	0.0
Brazil	301.0	7.0	6.0	8.0
Chile	57.0	0.0	0.0	0.0
Colombia	70.0	0.0	4.0	1.0
Ecuador	48.0	2.0	1.0	0.0
Guyana	7.0	0.0	0.0	0.0
Paraguay	8.0	0.0	0.0	0.0
Peru	35.0	0.0	0.0	0.0
Suriname	3.0	0.0	0.0	0.0
Uruguay	11.0	0.0	0.0	0.0
Venezuela	44.0	1.0	3.0	0.0

```
[233]: dfosa.columns =list(map( str,dfosa.columns))
```

```
[234]: dfosa
```

```
[234]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				

Argentina	174.0	0.0	1.0	2.0
Aruba	3.0	0.0	0.0	0.0
Bolivia	5.0	0.0	0.0	0.0
Brazil	301.0	7.0	6.0	8.0
Chile	57.0	0.0	0.0	0.0
Colombia	70.0	0.0	4.0	1.0
Ecuador	48.0	2.0	1.0	0.0
Guyana	7.0	0.0	0.0	0.0
Paraguay	8.0	0.0	0.0	0.0
Peru	35.0	0.0	0.0	0.0
Suriname	3.0	0.0	0.0	0.0
Uruguay	11.0	0.0	0.0	0.0
Venezuela	44.0	1.0	3.0	0.0

```
[235]: dfosa.loc['Total_South America'] = pdsa.Series([dfosa['ATHLETES'].
↳sum(),dfosa['GOLD'].sum(),dfosa['SILVER'].sum(),dfosa['BRONZE'].sum()],
↳index = ["ATHLETES","GOLD","SILVER","BRONZE"])
```

```
[236]: dfosa
```

```
[236]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Argentina	174.0	0.0	1.0	2.0
Aruba	3.0	0.0	0.0	0.0
Bolivia	5.0	0.0	0.0	0.0
Brazil	301.0	7.0	6.0	8.0
Chile	57.0	0.0	0.0	0.0
Colombia	70.0	0.0	4.0	1.0
Ecuador	48.0	2.0	1.0	0.0
Guyana	7.0	0.0	0.0	0.0
Paraguay	8.0	0.0	0.0	0.0
Peru	35.0	0.0	0.0	0.0
Suriname	3.0	0.0	0.0	0.0
Uruguay	11.0	0.0	0.0	0.0
Venezuela	44.0	1.0	3.0	0.0
Total_South America	766.0	10.0	15.0	11.0

```
[237]: import numpy as np
import pandas as pdaf
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[238]: dfoaf = pdaf.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo Olympics_
↳2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='African countries')
```

[239]: dfoaf

```
[239]:
```

	COUNTRIES	ATHLETES	GOLD	SILVER	BRONZE
0	Algeria	44.0	0.0	0.0	0.0
1	Angola	20.0	0.0	0.0	0.0
2	Benin	7.0	0.0	0.0	0.0
3	Botswana	13.0	0.0	0.0	1.0
4	Burkina Faso	7.0	0.0	0.0	1.0
5	Burundi	6.0	0.0	0.0	0.0
6	Cameroon	12.0	0.0	0.0	0.0
7	Cape Verde	6.0	0.0	0.0	0.0
8	Central African Republic	2.0	0.0	0.0	0.0
9	Chad	3.0	0.0	0.0	0.0
10	Comoros	3.0	0.0	0.0	0.0
11	Democratic Republic of the Congo	7.0	0.0	0.0	0.0
12	Djibouti	4.0	0.0	0.0	0.0
13	Egypt	133.0	1.0	1.0	4.0
14	Equatorial Guinea	3.0	0.0	0.0	0.0
15	Eritrea	13.0	0.0	0.0	0.0
16	Eswatini	4.0	0.0	0.0	0.0
17	Ethiopia	38.0	1.0	1.0	2.0
18	Gabon	5.0	0.0	0.0	0.0
19	Gambia	4.0	0.0	0.0	0.0
20	Ghana	14.0	0.0	0.0	1.0
21	Guinea	5.0	0.0	0.0	0.0
22	Guinea-Bissau	4.0	0.0	0.0	0.0
23	Ivory Coast	28.0	0.0	0.0	1.0
24	Kenya	85.0	4.0	4.0	2.0
25	Lesotho	2.0	0.0	0.0	0.0
26	Liberia	3.0	0.0	0.0	0.0
27	Libya	4.0	0.0	0.0	0.0
28	Madagascar	6.0	0.0	0.0	0.0
29	Malawi	5.0	0.0	0.0	0.0
30	Mali	4.0	0.0	0.0	0.0
31	Mauritania	2.0	0.0	0.0	0.0
32	Mauritius	8.0	0.0	0.0	0.0
33	Morocco	50.0	1.0	0.0	0.0
34	Mozambique	10.0	0.0	0.0	0.0
35	Namibia	11.0	0.0	1.0	0.0
36	Niger	7.0	0.0	0.0	0.0
37	Nigeria	60.0	0.0	1.0	1.0
38	Republic of the Congo	3.0	0.0	0.0	0.0
39	Rwanda	6.0	0.0	0.0	0.0
40	Sao Tome and Principe	3.0	0.0	0.0	0.0
41	Senegal	9.0	0.0	0.0	0.0
42	Seychelles	5.0	0.0	0.0	0.0
43	Sierra Leone	4.0	0.0	0.0	0.0

44	Somalia	2.0	0.0	0.0	0.0
45	South Africa	177.0	1.0	2.0	0.0
46	South Sudan	2.0	0.0	0.0	0.0
47	Sudan	5.0	0.0	0.0	0.0
48	Tanzania	3.0	0.0	0.0	0.0
49	Togo	4.0	0.0	0.0	0.0
50	Tunisia	63.0	1.0	1.0	0.0
51	Uganda	25.0	2.0	1.0	1.0
52	Zambia	26.0	0.0	0.0	0.0
53	Zimbabwe	5.0	0.0	0.0	0.0

```
[240]: dfoaf.set_index('COUNTRIES', inplace=True)
```

```
[241]: dfoaf
```

```
[241]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Algeria	44.0	0.0	0.0	0.0
Angola	20.0	0.0	0.0	0.0
Benin	7.0	0.0	0.0	0.0
Botswana	13.0	0.0	0.0	1.0
Burkina Faso	7.0	0.0	0.0	1.0
Burundi	6.0	0.0	0.0	0.0
Cameroon	12.0	0.0	0.0	0.0
Cape Verde	6.0	0.0	0.0	0.0
Central African Republic	2.0	0.0	0.0	0.0
Chad	3.0	0.0	0.0	0.0
Comoros	3.0	0.0	0.0	0.0
Democratic Republic of the Congo	7.0	0.0	0.0	0.0
Djibouti	4.0	0.0	0.0	0.0
Egypt	133.0	1.0	1.0	4.0
Equatorial Guinea	3.0	0.0	0.0	0.0
Eritrea	13.0	0.0	0.0	0.0
Eswatini	4.0	0.0	0.0	0.0
Ethiopia	38.0	1.0	1.0	2.0
Gabon	5.0	0.0	0.0	0.0
Gambia	4.0	0.0	0.0	0.0
Ghana	14.0	0.0	0.0	1.0
Guinea	5.0	0.0	0.0	0.0
Guinea-Bissau	4.0	0.0	0.0	0.0
Ivory Coast	28.0	0.0	0.0	1.0
Kenya	85.0	4.0	4.0	2.0
Lesotho	2.0	0.0	0.0	0.0
Liberia	3.0	0.0	0.0	0.0
Libya	4.0	0.0	0.0	0.0
Madagascar	6.0	0.0	0.0	0.0
Malawi	5.0	0.0	0.0	0.0

Mali	4.0	0.0	0.0	0.0
Mauritania	2.0	0.0	0.0	0.0
Mauritius	8.0	0.0	0.0	0.0
Morocco	50.0	1.0	0.0	0.0
Mozambique	10.0	0.0	0.0	0.0
Namibia	11.0	0.0	1.0	0.0
Niger	7.0	0.0	0.0	0.0
Nigeria	60.0	0.0	1.0	1.0
Republic of the Congo	3.0	0.0	0.0	0.0
Rwanda	6.0	0.0	0.0	0.0
Sao Tome and Principe	3.0	0.0	0.0	0.0
Senegal	9.0	0.0	0.0	0.0
Seychelles	5.0	0.0	0.0	0.0
Sierra Leone	4.0	0.0	0.0	0.0
Somalia	2.0	0.0	0.0	0.0
South Africa	177.0	1.0	2.0	0.0
South Sudan	2.0	0.0	0.0	0.0
Sudan	5.0	0.0	0.0	0.0
Tanzania	3.0	0.0	0.0	0.0
Togo	4.0	0.0	0.0	0.0
Tunisia	63.0	1.0	1.0	0.0
Uganda	25.0	2.0	1.0	1.0
Zambia	26.0	0.0	0.0	0.0
Zimbabwe	5.0	0.0	0.0	0.0

```
[242]: dfoaf.loc['Total_Africa'] = pdaf.Series([dfoaf['ATHLETES'].sum(),dfoaf['GOLD'].
↪sum(),dfoaf['SILVER'].sum(),dfoaf['BRONZE'].sum()], index =
↪["ATHLETES", "GOLD", "SILVER", "BRONZE"])
```

```
[243]: dfoaf
```

```
[243]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Algeria	44.0	0.0	0.0	0.0
Angola	20.0	0.0	0.0	0.0
Benin	7.0	0.0	0.0	0.0
Botswana	13.0	0.0	0.0	1.0
Burkina Faso	7.0	0.0	0.0	1.0
Burundi	6.0	0.0	0.0	0.0
Cameroon	12.0	0.0	0.0	0.0
Cape Verde	6.0	0.0	0.0	0.0
Central African Republic	2.0	0.0	0.0	0.0
Chad	3.0	0.0	0.0	0.0
Comoros	3.0	0.0	0.0	0.0
Democratic Republic of the Congo	7.0	0.0	0.0	0.0
Djibouti	4.0	0.0	0.0	0.0
Egypt	133.0	1.0	1.0	4.0

Equatorial Guinea	3.0	0.0	0.0	0.0
Eritrea	13.0	0.0	0.0	0.0
Eswatini	4.0	0.0	0.0	0.0
Ethiopia	38.0	1.0	1.0	2.0
Gabon	5.0	0.0	0.0	0.0
Gambia	4.0	0.0	0.0	0.0
Ghana	14.0	0.0	0.0	1.0
Guinea	5.0	0.0	0.0	0.0
Guinea-Bissau	4.0	0.0	0.0	0.0
Ivory Coast	28.0	0.0	0.0	1.0
Kenya	85.0	4.0	4.0	2.0
Lesotho	2.0	0.0	0.0	0.0
Liberia	3.0	0.0	0.0	0.0
Libya	4.0	0.0	0.0	0.0
Madagascar	6.0	0.0	0.0	0.0
Malawi	5.0	0.0	0.0	0.0
Mali	4.0	0.0	0.0	0.0
Mauritania	2.0	0.0	0.0	0.0
Mauritius	8.0	0.0	0.0	0.0
Morocco	50.0	1.0	0.0	0.0
Mozambique	10.0	0.0	0.0	0.0
Namibia	11.0	0.0	1.0	0.0
Niger	7.0	0.0	0.0	0.0
Nigeria	60.0	0.0	1.0	1.0
Republic of the Congo	3.0	0.0	0.0	0.0
Rwanda	6.0	0.0	0.0	0.0
Sao Tome and Principe	3.0	0.0	0.0	0.0
Senegal	9.0	0.0	0.0	0.0
Seychelles	5.0	0.0	0.0	0.0
Sierra Leone	4.0	0.0	0.0	0.0
Somalia	2.0	0.0	0.0	0.0
South Africa	177.0	1.0	2.0	0.0
South Sudan	2.0	0.0	0.0	0.0
Sudan	5.0	0.0	0.0	0.0
Tanzania	3.0	0.0	0.0	0.0
Togo	4.0	0.0	0.0	0.0
Tunisia	63.0	1.0	1.0	0.0
Uganda	25.0	2.0	1.0	1.0
Zambia	26.0	0.0	0.0	0.0
Zimbabwe	5.0	0.0	0.0	0.0
Total_Africa	984.0	11.0	12.0	14.0

```
[244]: import numpy as np
import pandas as pdoc
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
import openpyxl
```

```
[245]: dfooc = pdoc.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo Olympics_
↳2020(2021)/Olympic Medal Count 2020.xlsx',
        sheet_name='Oceanian countries')
```

```
[246]: dfooc.set_index('COUNTRIES', inplace=True)
```

```
[247]: dfooc.loc['Total_Oceania'] = pdoc.Series([dfooc['ATHLETES'].sum(),dfooc['GOLD'].
↳sum(),dfooc['SILVER'].sum(),dfooc['BRONZE'].sum()], index =_
↳["ATHLETES", "GOLD", "SILVER", "BRONZE"])
```

```
[248]: dfooc
```

```
[248]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
American Samoa	6.0	0.0	0.0	0.0
Australia	478.0	17.0	7.0	22.0
Cook Islands	6.0	0.0	0.0	0.0
Federal States of Micronesia	3.0	0.0	0.0	0.0
Fiji	30.0	1.0	0.0	1.0
Guam	5.0	0.0	0.0	0.0
Kiribati	3.0	0.0	0.0	0.0
Marshall Islands	2.0	0.0	0.0	0.0
Nauru	2.0	0.0	0.0	0.0
New Zealand	223.0	7.0	6.0	7.0
Palau	3.0	0.0	0.0	0.0
Papua New Guinea	8.0	0.0	0.0	0.0
Samoa	8.0	0.0	0.0	0.0
Solomon Islands	3.0	0.0	0.0	0.0
Tonga	6.0	0.0	0.0	0.0
Tuvalu	2.0	0.0	0.0	0.0
Vanuatu	3.0	0.0	0.0	0.0
Total_Oceania	791.0	25.0	13.0	30.0

```
[249]: dfooc.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 18 entries, American Samoa to Total_Oceania
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ATHLETES    18 non-null    float64
1   GOLD        18 non-null    float64
2   SILVER      18 non-null    float64
3   BRONZE      18 non-null    float64
dtypes: float64(4)
```

memory usage: 720.0+ bytes

```
[250]: import numpy as np
import pandas as pdasian
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[251]: dfoasian = pdasian.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo_
↳Olympics 2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='Asian countries')
```

```
[252]: dfoasian
```

```
[252]:
```

	COUNTRIES	ATHLETES	GOLD	SILVER	BRONZE
0	Afghanistan	5.0	0.0	0.0	0.0
1	Armenia	17.0	0.0	2.0	2.0
2	Azerbaijan	44.0	0.0	3.0	4.0
3	Bahrain	32.0	0.0	1.0	0.0
4	Bangladesh	6.0	0.0	0.0	0.0
5	Bhutan	4.0	0.0	0.0	0.0
6	Brunei	2.0	0.0	0.0	0.0
7	Cambodia	3.0	0.0	0.0	0.0
8	China	406.0	38.0	32.0	18.0
9	Cyprus	15.0	0.0	0.0	0.0
10	East Timor	3.0	0.0	0.0	0.0
11	Hong Kong	42.0	1.0	2.0	3.0
12	India	127.0	1.0	2.0	4.0
13	Indonesia	28.0	1.0	1.0	3.0
14	Iran	66.0	3.0	2.0	2.0
15	Iraq	4.0	0.0	0.0	0.0
16	Israel	90.0	2.0	0.0	2.0
17	Japan	522.0	27.0	14.0	17.0
18	Jordan	14.0	0.0	1.0	1.0
19	Kazakhstan	93.0	0.0	0.0	8.0
20	Kuwait	11.0	0.0	0.0	1.0
21	Kyrgyzstan	16.0	0.0	2.0	1.0
22	Laos	4.0	0.0	0.0	0.0
23	Lebanon	6.0	0.0	0.0	0.0
24	Malaysia	30.0	0.0	1.0	1.0
25	Maldives	4.0	0.0	0.0	0.0
26	Mongolia	43.0	0.0	1.0	3.0
27	Myanmar	3.0	0.0	0.0	0.0
28	Nepal	5.0	0.0	0.0	0.0
29	Oman	5.0	0.0	0.0	0.0
30	Pakistan	10.0	0.0	0.0	0.0

31	Palestine	5.0	0.0	0.0	0.0
32	Philippines	19.0	1.0	2.0	1.0
33	Qatar	16.0	2.0	0.0	1.0
34	Saudi Arabia	29.0	0.0	1.0	0.0
35	Singapore	23.0	0.0	0.0	0.0
36	South Korea	236.0	6.0	4.0	10.0
37	Sri Lanka	9.0	0.0	0.0	0.0
38	Syria	6.0	0.0	0.0	1.0
39	Taiwan	68.0	2.0	4.0	6.0
40	Tajikistan	11.0	0.0	0.0	0.0
41	Thailand	42.0	1.0	0.0	1.0
42	Turkey	108.0	2.0	2.0	9.0
43	Turkmenistan	9.0	0.0	1.0	0.0
44	United Arab Emirates	5.0	0.0	0.0	0.0
45	Uzbekistan	65.0	3.0	0.0	2.0
46	Vietnam	18.0	0.0	0.0	0.0
47	Yemen	5.0	0.0	0.0	0.0

```
[253]: dfoasian.set_index('COUNTRIES', inplace=True)
```

```
[254]: dfoasian
```

```
[254]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Afghanistan	5.0	0.0	0.0	0.0
Armenia	17.0	0.0	2.0	2.0
Azerbaijan	44.0	0.0	3.0	4.0
Bahrain	32.0	0.0	1.0	0.0
Bangladesh	6.0	0.0	0.0	0.0
Bhutan	4.0	0.0	0.0	0.0
Brunei	2.0	0.0	0.0	0.0
Cambodia	3.0	0.0	0.0	0.0
China	406.0	38.0	32.0	18.0
Cyprus	15.0	0.0	0.0	0.0
East Timor	3.0	0.0	0.0	0.0
Hong Kong	42.0	1.0	2.0	3.0
India	127.0	1.0	2.0	4.0
Indonesia	28.0	1.0	1.0	3.0
Iran	66.0	3.0	2.0	2.0
Iraq	4.0	0.0	0.0	0.0
Israel	90.0	2.0	0.0	2.0
Japan	522.0	27.0	14.0	17.0
Jordan	14.0	0.0	1.0	1.0
Kazakhstan	93.0	0.0	0.0	8.0
Kuwait	11.0	0.0	0.0	1.0
Kyrgyzstan	16.0	0.0	2.0	1.0
Laos	4.0	0.0	0.0	0.0

Lebanon	6.0	0.0	0.0	0.0
Malaysia	30.0	0.0	1.0	1.0
Maldives	4.0	0.0	0.0	0.0
Mongolia	43.0	0.0	1.0	3.0
Myanmar	3.0	0.0	0.0	0.0
Nepal	5.0	0.0	0.0	0.0
Oman	5.0	0.0	0.0	0.0
Pakistan	10.0	0.0	0.0	0.0
Palestine	5.0	0.0	0.0	0.0
Philippines	19.0	1.0	2.0	1.0
Qatar	16.0	2.0	0.0	1.0
Saudi Arabia	29.0	0.0	1.0	0.0
Singapore	23.0	0.0	0.0	0.0
South Korea	236.0	6.0	4.0	10.0
Sri Lanka	9.0	0.0	0.0	0.0
Syria	6.0	0.0	0.0	1.0
Taiwan	68.0	2.0	4.0	6.0
Tajikistan	11.0	0.0	0.0	0.0
Thailand	42.0	1.0	0.0	1.0
Turkey	108.0	2.0	2.0	9.0
Turkmenistan	9.0	0.0	1.0	0.0
United Arab Emirates	5.0	0.0	0.0	0.0
Uzbekistan	65.0	3.0	0.0	2.0
Vietnam	18.0	0.0	0.0	0.0
Yemen	5.0	0.0	0.0	0.0

```
[255]: dfoasian.loc['Total_Asia'] = pdasian.Series([dfoasian['ATHLETES'].
↪sum(),dfoasian['GOLD'].sum(),dfoasian['SILVER'].sum(),dfoasian['BRONZE'].
↪sum()], index = ["ATHLETES","GOLD","SILVER","BRONZE"])
```

```
[256]: dfoasian
```

```
[256]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Afghanistan	5.0	0.0	0.0	0.0
Armenia	17.0	0.0	2.0	2.0
Azerbaijan	44.0	0.0	3.0	4.0
Bahrain	32.0	0.0	1.0	0.0
Bangladesh	6.0	0.0	0.0	0.0
Bhutan	4.0	0.0	0.0	0.0
Brunei	2.0	0.0	0.0	0.0
Cambodia	3.0	0.0	0.0	0.0
China	406.0	38.0	32.0	18.0
Cyprus	15.0	0.0	0.0	0.0
East Timor	3.0	0.0	0.0	0.0
Hong Kong	42.0	1.0	2.0	3.0
India	127.0	1.0	2.0	4.0

Indonesia	28.0	1.0	1.0	3.0
Iran	66.0	3.0	2.0	2.0
Iraq	4.0	0.0	0.0	0.0
Israel	90.0	2.0	0.0	2.0
Japan	522.0	27.0	14.0	17.0
Jordan	14.0	0.0	1.0	1.0
Kazakhstan	93.0	0.0	0.0	8.0
Kuwait	11.0	0.0	0.0	1.0
Kyrgyzstan	16.0	0.0	2.0	1.0
Laos	4.0	0.0	0.0	0.0
Lebanon	6.0	0.0	0.0	0.0
Malaysia	30.0	0.0	1.0	1.0
Maldives	4.0	0.0	0.0	0.0
Mongolia	43.0	0.0	1.0	3.0
Myanmar	3.0	0.0	0.0	0.0
Nepal	5.0	0.0	0.0	0.0
Oman	5.0	0.0	0.0	0.0
Pakistan	10.0	0.0	0.0	0.0
Palestine	5.0	0.0	0.0	0.0
Philippines	19.0	1.0	2.0	1.0
Qatar	16.0	2.0	0.0	1.0
Saudi Arabia	29.0	0.0	1.0	0.0
Singapore	23.0	0.0	0.0	0.0
South Korea	236.0	6.0	4.0	10.0
Sri Lanka	9.0	0.0	0.0	0.0
Syria	6.0	0.0	0.0	1.0
Taiwan	68.0	2.0	4.0	6.0
Tajikistan	11.0	0.0	0.0	0.0
Thailand	42.0	1.0	0.0	1.0
Turkey	108.0	2.0	2.0	9.0
Turkmenistan	9.0	0.0	1.0	0.0
United Arab Emirates	5.0	0.0	0.0	0.0
Uzbekistan	65.0	3.0	0.0	2.0
Vietnam	18.0	0.0	0.0	0.0
Yemen	5.0	0.0	0.0	0.0
Total_Asia	2334.0	90.0	78.0	101.0

```
[257]: import numpy as np
import pandas as pdeuro
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[258]: dfoeuro = pdeuro.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo_
↳Olympics 2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='European countries')
```

[259] : dfoeuro

```
[259] :
      COUNTRIES  ATHLETES  GOLD  SILVER  BRONZE
0      Albania      9.0    0.0    0.0    0.0
1      Andorra      2.0    0.0    0.0    0.0
2      Austria     60.0    1.0    1.0    5.0
3      Belarus    101.0    1.0    3.0    3.0
4      Belgium    121.0    3.0    1.0    3.0
5  Bosnia and Herzegovina    7.0    0.0    0.0    0.0
6      Bulgaria     42.0    3.0    1.0    2.0
7      Croatia     59.0    3.0    3.0    2.0
8      Czech Republic    115.0    4.0    4.0    3.0
9      Denmark    107.0    3.0    4.0    4.0
10     Estonia     33.0    1.0    0.0    1.0
11     Finland     29.0    0.0    0.0    2.0
12     France     398.0   10.0   12.0   11.0
13     Georgia     35.0    2.0    5.0    1.0
14     Germany    425.0   10.0   11.0   16.0
15     Great Britain    376.0   22.0   21.0   22.0
16     Greece      83.0    2.0    1.0    1.0
17     Hungary    166.0    6.0    7.0    7.0
18     Iceland      4.0    0.0    0.0    0.0
19     Ireland    116.0    2.0    0.0    2.0
20     Italy     372.0   10.0   10.0   20.0
21     Kosovo      11.0    2.0    0.0    0.0
22     Latvia      33.0    1.0    0.0    1.0
23     Liechtenstein     5.0    0.0    0.0    0.0
24     Lithuania     41.0    0.0    1.0    0.0
25     Luxembourg     12.0    0.0    0.0    0.0
26     Malta        6.0    0.0    0.0    0.0
27     Moldova     20.0    0.0    0.0    1.0
28     Monaco       6.0    0.0    0.0    0.0
29     Montenegro     34.0    0.0    0.0    0.0
30     Netherlands    278.0   10.0   12.0   14.0
31     North Macedonia     8.0    0.0    1.0    0.0
32     Norway      75.0    4.0    2.0    2.0
33     Poland     210.0    4.0    5.0    5.0
34     Portugal     92.0    1.0    1.0    2.0
35     Romania    101.0    1.0    3.0    0.0
36     Russia     328.0   17.0   24.0   22.0
37     San Marino      5.0    0.0    1.0    2.0
38     Serbia      86.0    3.0    1.0    5.0
39     Slovakia     41.0    1.0    2.0    1.0
40     Slovenia     53.0    3.0    1.0    1.0
41     Spain     321.0    3.0    8.0    6.0
42     Sweden     134.0    3.0    6.0    0.0
43     Switzerland    107.0    3.0    4.0    6.0
```

44 Ukraine 155.0 1.0 6.0 12.0

```
[260]: dfoeuro.set_index('COUNTRIES', inplace=True)
```

```
[261]: dfoeuro
```

```
[261]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Albania	9.0	0.0	0.0	0.0
Andorra	2.0	0.0	0.0	0.0
Austria	60.0	1.0	1.0	5.0
Belarus	101.0	1.0	3.0	3.0
Belgium	121.0	3.0	1.0	3.0
Bosnia and Herzegovina	7.0	0.0	0.0	0.0
Bulgaria	42.0	3.0	1.0	2.0
Croatia	59.0	3.0	3.0	2.0
Czech Republic	115.0	4.0	4.0	3.0
Denmark	107.0	3.0	4.0	4.0
Estonia	33.0	1.0	0.0	1.0
Finland	29.0	0.0	0.0	2.0
France	398.0	10.0	12.0	11.0
Georgia	35.0	2.0	5.0	1.0
Germany	425.0	10.0	11.0	16.0
Great Britain	376.0	22.0	21.0	22.0
Greece	83.0	2.0	1.0	1.0
Hungary	166.0	6.0	7.0	7.0
Iceland	4.0	0.0	0.0	0.0
Ireland	116.0	2.0	0.0	2.0
Italy	372.0	10.0	10.0	20.0
Kosovo	11.0	2.0	0.0	0.0
Latvia	33.0	1.0	0.0	1.0
Liechtenstein	5.0	0.0	0.0	0.0
Lithuania	41.0	0.0	1.0	0.0
Luxembourg	12.0	0.0	0.0	0.0
Malta	6.0	0.0	0.0	0.0
Moldova	20.0	0.0	0.0	1.0
Monaco	6.0	0.0	0.0	0.0
Montenegro	34.0	0.0	0.0	0.0
Netherlands	278.0	10.0	12.0	14.0
North Macedonia	8.0	0.0	1.0	0.0
Norway	75.0	4.0	2.0	2.0
Poland	210.0	4.0	5.0	5.0
Portugal	92.0	1.0	1.0	2.0
Romania	101.0	1.0	3.0	0.0
Russia	328.0	17.0	24.0	22.0
San Marino	5.0	0.0	1.0	2.0
Serbia	86.0	3.0	1.0	5.0

Slovakia	41.0	1.0	2.0	1.0
Slovenia	53.0	3.0	1.0	1.0
Spain	321.0	3.0	8.0	6.0
Sweden	134.0	3.0	6.0	0.0
Switzerland	107.0	3.0	4.0	6.0
Ukraine	155.0	1.0	6.0	12.0

```
[262]: dfoeuro.loc['Total_Europe'] = pdeuro.Series([dfoeuro['ATHLETES'].
↪sum(),dfoeuro['GOLD'].sum(),dfoeuro['SILVER'].sum(),dfoeuro['BRONZE'].
↪sum()], index = ["ATHLETES","GOLD","SILVER","BRONZE"])
```

```
[263]: dfoeuro
```

```
[263]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Albania	9.0	0.0	0.0	0.0
Andorra	2.0	0.0	0.0	0.0
Austria	60.0	1.0	1.0	5.0
Belarus	101.0	1.0	3.0	3.0
Belgium	121.0	3.0	1.0	3.0
Bosnia and Herzegovina	7.0	0.0	0.0	0.0
Bulgaria	42.0	3.0	1.0	2.0
Croatia	59.0	3.0	3.0	2.0
Czech Republic	115.0	4.0	4.0	3.0
Denmark	107.0	3.0	4.0	4.0
Estonia	33.0	1.0	0.0	1.0
Finland	29.0	0.0	0.0	2.0
France	398.0	10.0	12.0	11.0
Georgia	35.0	2.0	5.0	1.0
Germany	425.0	10.0	11.0	16.0
Great Britain	376.0	22.0	21.0	22.0
Greece	83.0	2.0	1.0	1.0
Hungary	166.0	6.0	7.0	7.0
Iceland	4.0	0.0	0.0	0.0
Ireland	116.0	2.0	0.0	2.0
Italy	372.0	10.0	10.0	20.0
Kosovo	11.0	2.0	0.0	0.0
Latvia	33.0	1.0	0.0	1.0
Liechtenstein	5.0	0.0	0.0	0.0
Lithuania	41.0	0.0	1.0	0.0
Luxembourg	12.0	0.0	0.0	0.0
Malta	6.0	0.0	0.0	0.0
Moldova	20.0	0.0	0.0	1.0
Monaco	6.0	0.0	0.0	0.0
Montenegro	34.0	0.0	0.0	0.0
Netherlands	278.0	10.0	12.0	14.0
North Macedonia	8.0	0.0	1.0	0.0

Norway	75.0	4.0	2.0	2.0
Poland	210.0	4.0	5.0	5.0
Portugal	92.0	1.0	1.0	2.0
Romania	101.0	1.0	3.0	0.0
Russia	328.0	17.0	24.0	22.0
San Marino	5.0	0.0	1.0	2.0
Serbia	86.0	3.0	1.0	5.0
Slovakia	41.0	1.0	2.0	1.0
Slovenia	53.0	3.0	1.0	1.0
Spain	321.0	3.0	8.0	6.0
Sweden	134.0	3.0	6.0	0.0
Switzerland	107.0	3.0	4.0	6.0
Ukraine	155.0	1.0	6.0	12.0
Total_Europe	4822.0	140.0	162.0	185.0

```
[264]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import openpyxl
```

```
[265]: dfonordAme = pdnordAme.read_excel('/content/drive/MyDrive/Data_Engineering/
↳Tokyo Olympics 2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='North American countries')
```

```
[266]: dfonordAme
```

```
[266]:
```

	COUNTRIES	ATHLETES	GOLD	SILVER	BRONZE
0	Antigua and Barbuda	6.0	0.0	0.0	0.0
1	Bahamas	16.0	2.0	0.0	0.0
2	Barbados	8.0	0.0	0.0	0.0
3	Belize	3.0	0.0	0.0	0.0
4	Bermuda	2.0	1.0	0.0	0.0
5	Bolivia	5.0	0.0	0.0	0.0
6	British Virgin Islands	3.0	0.0	0.0	0.0
7	Canada	371.0	7.0	6.0	11.0
8	Cayman Islands	5.0	0.0	0.0	0.0
9	Costa Rica	14.0	0.0	0.0	0.0
10	Cuba	70.0	7.0	3.0	5.0
11	Dominica	2.0	0.0	0.0	0.0
12	Dominican Republic	62.0	0.0	3.0	2.0
13	El Salvador	5.0	0.0	0.0	0.0
14	Grenada	6.0	0.0	0.0	1.0
15	Guatemala	24.0	0.0	0.0	0.0
16	Haiti	6.0	0.0	0.0	0.0
17	Honduras	22.0	0.0	0.0	0.0

18	Jamaica	50.0	4.0	1.0	4.0
19	Mexico	164.0	0.0	0.0	4.0
20	Nicaragua	8.0	0.0	0.0	0.0
21	Panama	10.0	0.0	0.0	0.0
22	Puerto Rico	37.0	1.0	0.0	0.0
23	Saint Kitts and Nevis	2.0	0.0	0.0	0.0
24	Saint Lucia	5.0	0.0	0.0	0.0
25	Saint Vincent and the Grenadines	3.0	0.0	0.0	0.0
26	Trinidad and Tobago	22.0	0.0	0.0	0.0
27	United States	613.0	39.0	41.0	33.0

```
[267]: dfonordAme.set_index('COUNTRIES', inplace=True)
```

```
[268]: dfonordAme
```

```
[268]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Antigua and Barbuda	6.0	0.0	0.0	0.0
Bahamas	16.0	2.0	0.0	0.0
Barbados	8.0	0.0	0.0	0.0
Belize	3.0	0.0	0.0	0.0
Bermuda	2.0	1.0	0.0	0.0
Bolivia	5.0	0.0	0.0	0.0
British Virgin Islands	3.0	0.0	0.0	0.0
Canada	371.0	7.0	6.0	11.0
Cayman Islands	5.0	0.0	0.0	0.0
Costa Rica	14.0	0.0	0.0	0.0
Cuba	70.0	7.0	3.0	5.0
Dominica	2.0	0.0	0.0	0.0
Dominican Republic	62.0	0.0	3.0	2.0
El Salvador	5.0	0.0	0.0	0.0
Grenada	6.0	0.0	0.0	1.0
Guatemala	24.0	0.0	0.0	0.0
Haiti	6.0	0.0	0.0	0.0
Honduras	22.0	0.0	0.0	0.0
Jamaica	50.0	4.0	1.0	4.0
Mexico	164.0	0.0	0.0	4.0
Nicaragua	8.0	0.0	0.0	0.0
Panama	10.0	0.0	0.0	0.0
Puerto Rico	37.0	1.0	0.0	0.0
Saint Kitts and Nevis	2.0	0.0	0.0	0.0
Saint Lucia	5.0	0.0	0.0	0.0
Saint Vincent and the Grenadines	3.0	0.0	0.0	0.0
Trinidad and Tobago	22.0	0.0	0.0	0.0
United States	613.0	39.0	41.0	33.0

```
[269]: dfonordAme.loc['Total_North America'] = pdnordAme.
        ↪Series([dfonordAme['ATHLETES'].sum(),dfonordAme['GOLD'].
        ↪sum(),dfonordAme['SILVER'].sum(),dfonordAme['BRONZE'].sum()], index =_
        ↪["ATHLETES", "GOLD", "SILVER", "BRONZE"])
```

```
[270]: dfonordAme
```

```
[270]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Antigua and Barbuda	6.0	0.0	0.0	0.0
Bahamas	16.0	2.0	0.0	0.0
Barbados	8.0	0.0	0.0	0.0
Belize	3.0	0.0	0.0	0.0
Bermuda	2.0	1.0	0.0	0.0
Bolivia	5.0	0.0	0.0	0.0
British Virgin Islands	3.0	0.0	0.0	0.0
Canada	371.0	7.0	6.0	11.0
Cayman Islands	5.0	0.0	0.0	0.0
Costa Rica	14.0	0.0	0.0	0.0
Cuba	70.0	7.0	3.0	5.0
Dominica	2.0	0.0	0.0	0.0
Dominican Republic	62.0	0.0	3.0	2.0
El Salvador	5.0	0.0	0.0	0.0
Grenada	6.0	0.0	0.0	1.0
Guatemala	24.0	0.0	0.0	0.0
Haiti	6.0	0.0	0.0	0.0
Honduras	22.0	0.0	0.0	0.0
Jamaica	50.0	4.0	1.0	4.0
Mexico	164.0	0.0	0.0	4.0
Nicaragua	8.0	0.0	0.0	0.0
Panama	10.0	0.0	0.0	0.0
Puerto Rico	37.0	1.0	0.0	0.0
Saint Kitts and Nevis	2.0	0.0	0.0	0.0
Saint Lucia	5.0	0.0	0.0	0.0
Saint Vincent and the Grenadines	3.0	0.0	0.0	0.0
Trinidad and Tobago	22.0	0.0	0.0	0.0
United States	613.0	39.0	41.0	33.0
Total_North America	1544.0	61.0	54.0	60.0

```
[271]: dfonordAme.iloc[28, 0:4]
```

```
[271]: ATHLETES    1544.0
        GOLD       61.0
        SILVER     54.0
        BRONZE     60.0
        Name: Total_North America, dtype: float64
```



```
[272]: dfonordAme.iloc[[28]]
```

```
[272]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_North America	1544.0	61.0	54.0	60.0

```
[273]: dfonordAme1 = dfonordAme.iloc[[28]]
```

```
[274]: dfonordAme1
```

```
[274]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_North America	1544.0	61.0	54.0	60.0

```
[275]: dfoeuro1 = dfoeuro.iloc[[45]]
```

```
[276]: dfoeuro1
```

```
[276]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_Europe	4822.0	140.0	162.0	185.0

```
[277]: dfoasian1 = dfoasian.iloc[[48]]
```

```
[278]: dfoasian1
```

```
[278]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_Asia	2334.0	90.0	78.0	101.0

```
[279]: dfooc1 = dfooc.iloc[[17]]
```

```
[280]: dfooc1
```

```
[280]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_Oceania	791.0	25.0	13.0	30.0

```
[281]: dfoaf1 = dfoaf.iloc[[54]]
```

```
[282]: dfoaf1
```

```
[282]:
```

	ATHLETES	GOLD	SILVER	BRONZE
COUNTRIES				
Total_Africa	984.0	11.0	12.0	14.0

```
[283]: import pandas as pdcon
import matplotlib.pyplot as plt
%matplotlib inline
```

```
[284]: df_continents = pdcon.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo_
↳Olympics 2020(2021)/Olympic Medal Count 2020.xlsx',
sheet_name='Continents Medals')
```

```
[285]: df_continents
```

```
[285]:
```

	CONTINENTS	ATHLETES	GOLD	SILVER	BRONZE
0	South America	766.0	10.0	15.0	11.0
1	Africa	984.0	11.0	12.0	14.0
2	Oceania	791.0	25.0	13.0	30.0
3	Asia	2334.0	90.0	78.0	101.0
4	Europe	4822.0	140.0	162.0	185.0
5	North America	1544.0	61.0	54.0	60.0

```
[286]: df_continents.set_index('CONTINENTS', inplace=True)
```

```
[287]: df_continents
```

```
[287]:
```

	ATHLETES	GOLD	SILVER	BRONZE
CONTINENTS				
South America	766.0	10.0	15.0	11.0
Africa	984.0	11.0	12.0	14.0
Oceania	791.0	25.0	13.0	30.0
Asia	2334.0	90.0	78.0	101.0
Europe	4822.0	140.0	162.0	185.0
North America	1544.0	61.0	54.0	60.0

```
[288]: df_continents['TOTAL_MEDAL_BY_CONTINENT'] = df_continents.iloc[:, 1:].sum(axis_
↳= 1)
df_continents
```

```
[288]:
```

	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT
CONTINENTS					
South America	766.0	10.0	15.0	11.0	36.0
Africa	984.0	11.0	12.0	14.0	37.0
Oceania	791.0	25.0	13.0	30.0	68.0
Asia	2334.0	90.0	78.0	101.0	269.0
Europe	4822.0	140.0	162.0	185.0	487.0
North America	1544.0	61.0	54.0	60.0	175.0

```
[289]: Continents = df_continents.reset_index()
Continents
```

```
[289]:
```

	CONTINENTS	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT
0	South America	766.0	10.0	15.0	11.0	36.0
1	Africa	984.0	11.0	12.0	14.0	37.0
2	Oceania	791.0	25.0	13.0	30.0	68.0
3	Asia	2334.0	90.0	78.0	101.0	269.0
4	Europe	4822.0	140.0	162.0	185.0	487.0
5	North America	1544.0	61.0	54.0	60.0	175.0

```
[290]: CONT = pw.Pywedge_Charts(Continents, c=None, y="CONTINENTS")
# For Visualization
chart = CONT.make_charts()
```

/usr/local/lib/python3.7/dist-packages/pywedge/pywedge.py:27: FutureWarning:

In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only

HTML(value='<h2>Pywedge Make_Charts </h2>')

Tab(children=(Output(), Output(), Output(), Output(), Output(), Output(),
Output(), Output()), _titles={'0': '...

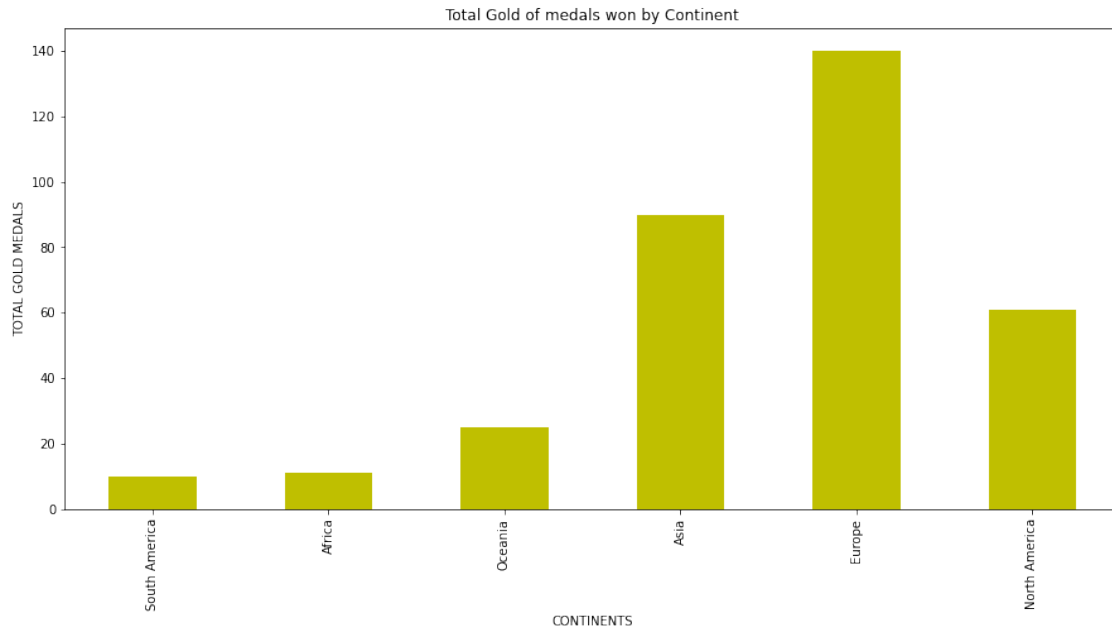
12 Visualization of the Total Gold of medals won by Continent

```
[291]: df_continents["GOLD"].plot(kind = 'bar',color = 'y',figsize = (15,7))
plt.title("Total Gold of medals won by Continent")

plt.xlabel("CONTINENTS")

plt.ylabel("TOTAL GOLD MEDALS")

plt.show()
```



13 Percentage of Total gold medals won by Continent

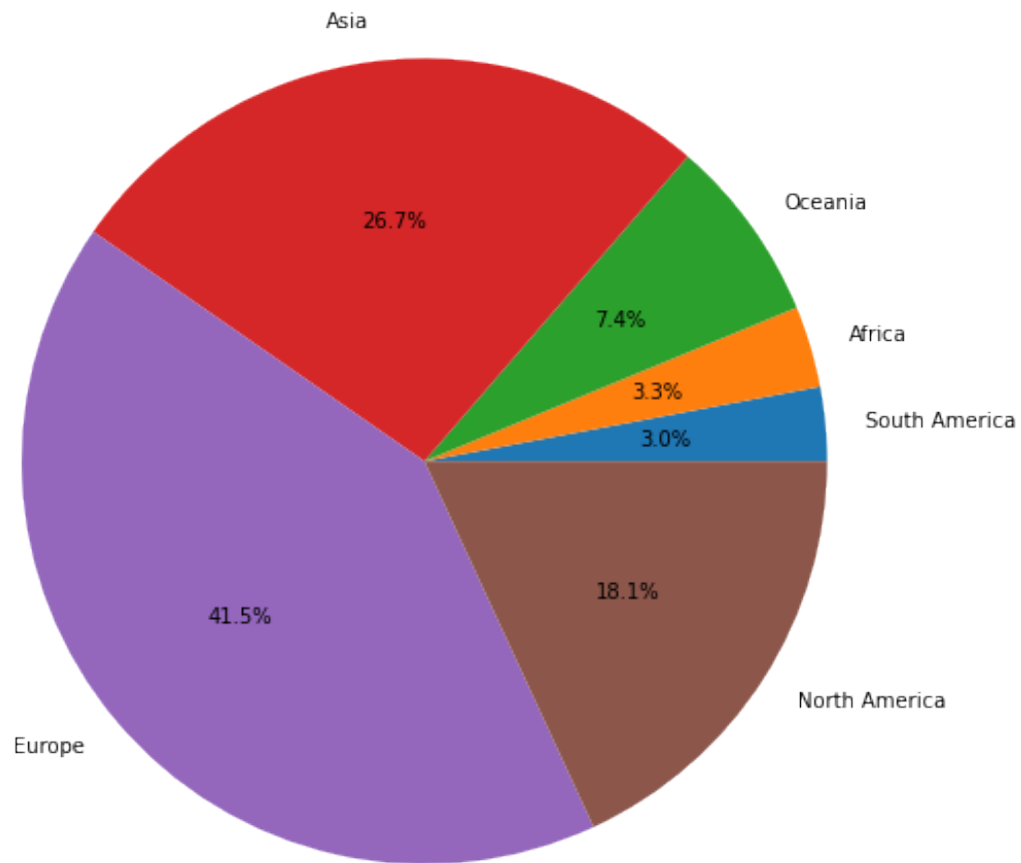
```
[292]: df_continents['GOLD'].plot( kind = 'pie', figsize = (18,9),autopct = '%1.1f%%')
plt.title("Percentage of Total gold medals won by Continent")

plt.xlabel("")

plt.ylabel("")

plt.show()
```

Percentage of Total gold medals won by Continent



14 Percentage of Total Silver medals won by Continent

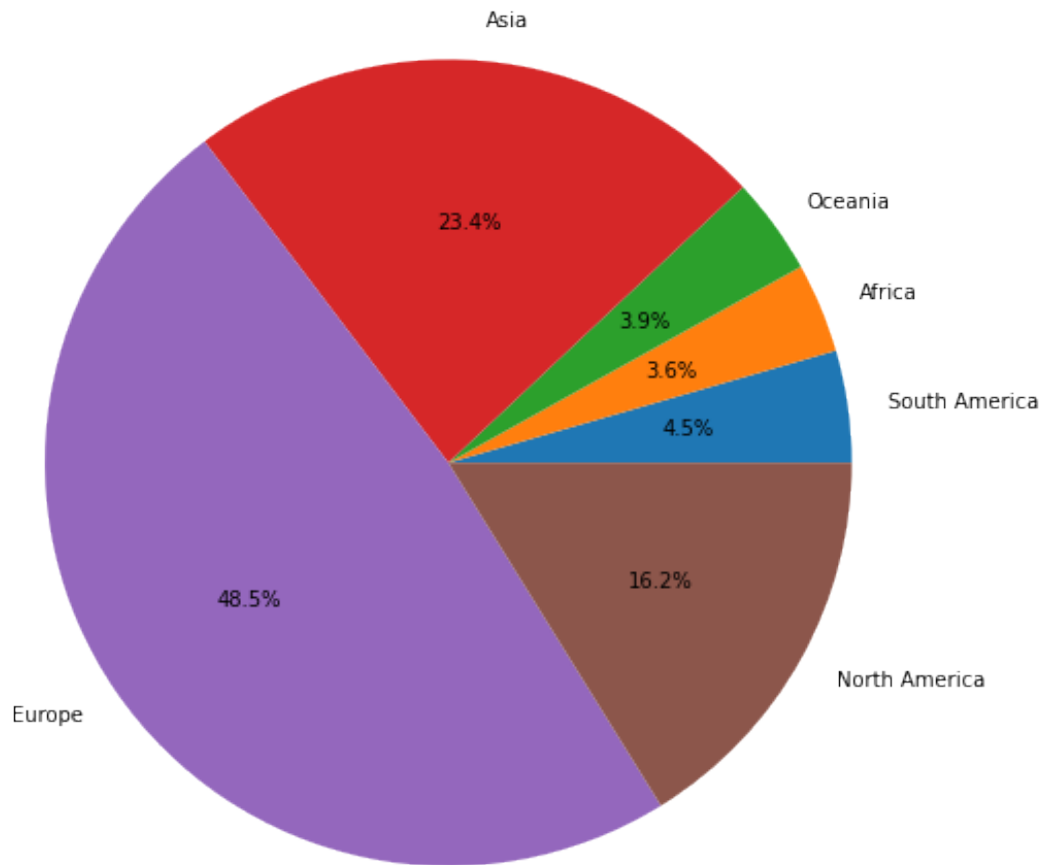
```
[293]: df_continents['SILVER'].plot( kind = 'pie', figsize = (18,9),autopct = '%1.1f%%')
plt.title("Percentage of Total Silver medals won by Continent")

plt.xlabel("")

plt.ylabel("")

plt.show()
```

Percentage of Total Silver medals won by Continent



15 Percentage of Total Bronze medals won by Continent

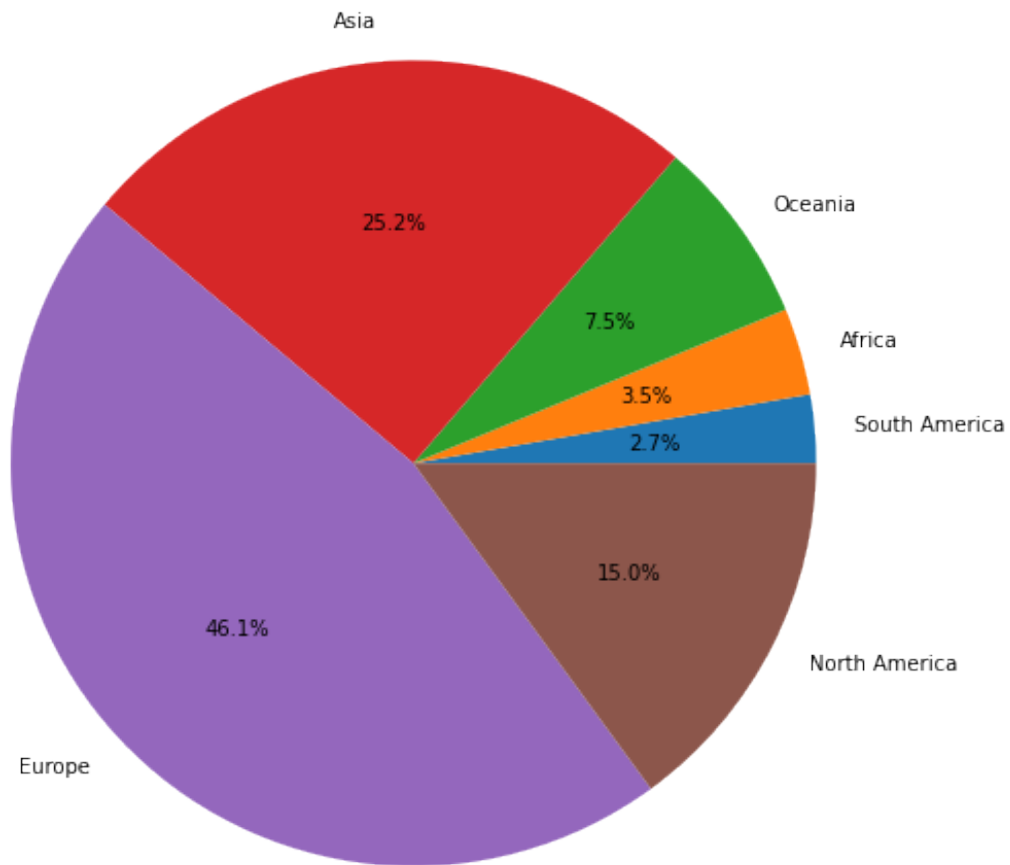
```
[294]: df_continents['BRONZE'].plot( kind = 'pie', figsize = (18,9),autopct = '%1.1f%%')
plt.title("Percentage of Total Bronze medals won by Continent")

plt.xlabel("")

plt.ylabel("")

plt.show()
```

Percentage of Total Bronze medals won by Continent



16 PERCENTAGE OF TOTAL ATHLETES PER CONTINENT

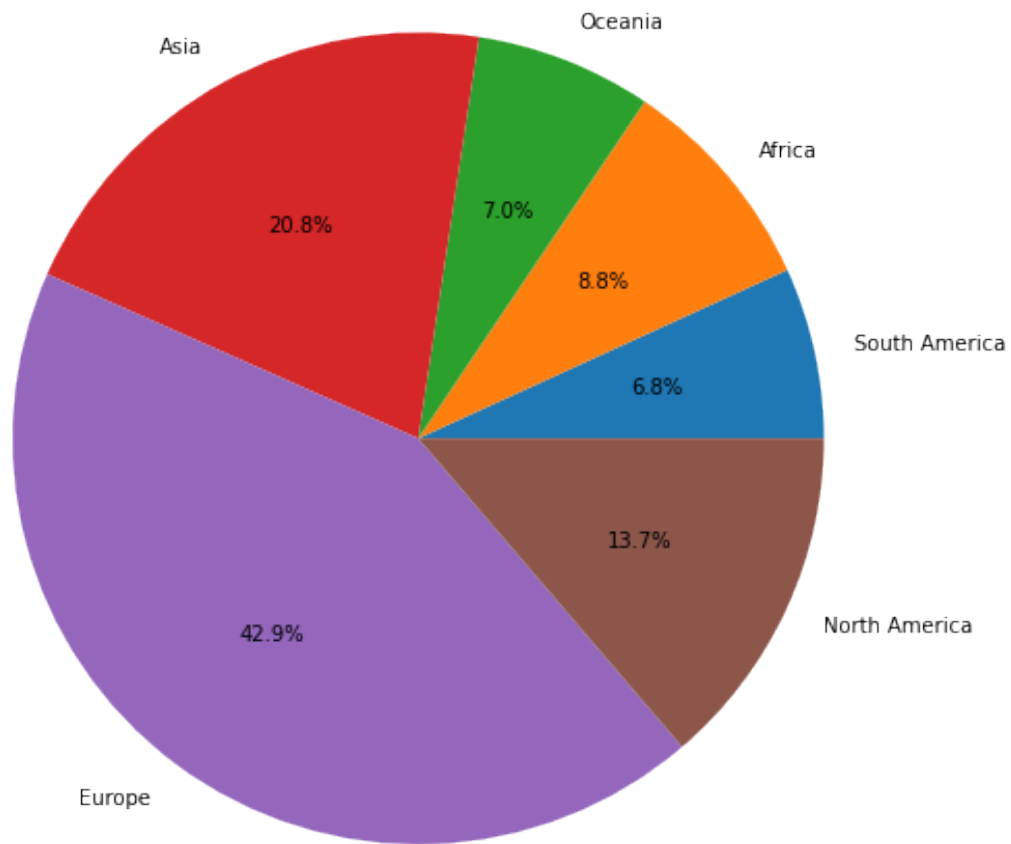
```
[295]: df_continents['ATHLETES'].plot( kind = 'pie', figsize = (18,9),autopct = '%1.
↪1f%%')
plt.title("PERCENTAGE OF TOTAL ATHLETES PER CONTINENT")

plt.xlabel("")

plt.ylabel("")

plt.show()
```

PERCENTAGE OF TOTAL ATHLETES PER CONTINENT

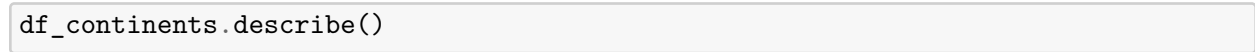


[295]:



17 Descriptive Statistics

[296]: `df_continents.describe()`



[296]:

	ATHLETES	GOLD	SILVER	BRONZE \
count	6.000000	6.000000	6.000000	6.000000
mean	1873.500000	56.166667	55.666667	66.833333
std	1562.296739	51.650428	58.633324	66.996766
min	766.000000	10.000000	12.000000	11.000000
25%	839.250000	14.500000	13.500000	18.000000
50%	1264.000000	43.000000	34.500000	45.000000

75%	2136.500000	82.750000	72.000000	90.750000
max	4822.000000	140.000000	162.000000	185.000000

TOTAL_MEDAL_BY_CONTINENT	
count	6.000000
mean	178.666667
std	176.586145
min	36.000000
25%	44.750000
50%	121.500000
75%	245.500000
max	487.000000

18 Statistical and graphical interpretation

19 On average, a continent sent 1873 athletes and won 56.16 gold medals, 55.66 silver medals and 66.83 bronze medals. ATHLETES, GOLDS, SILVER, BRONZE have high standard deviation

20 TOTAL ATHLETES BY CONTINENT

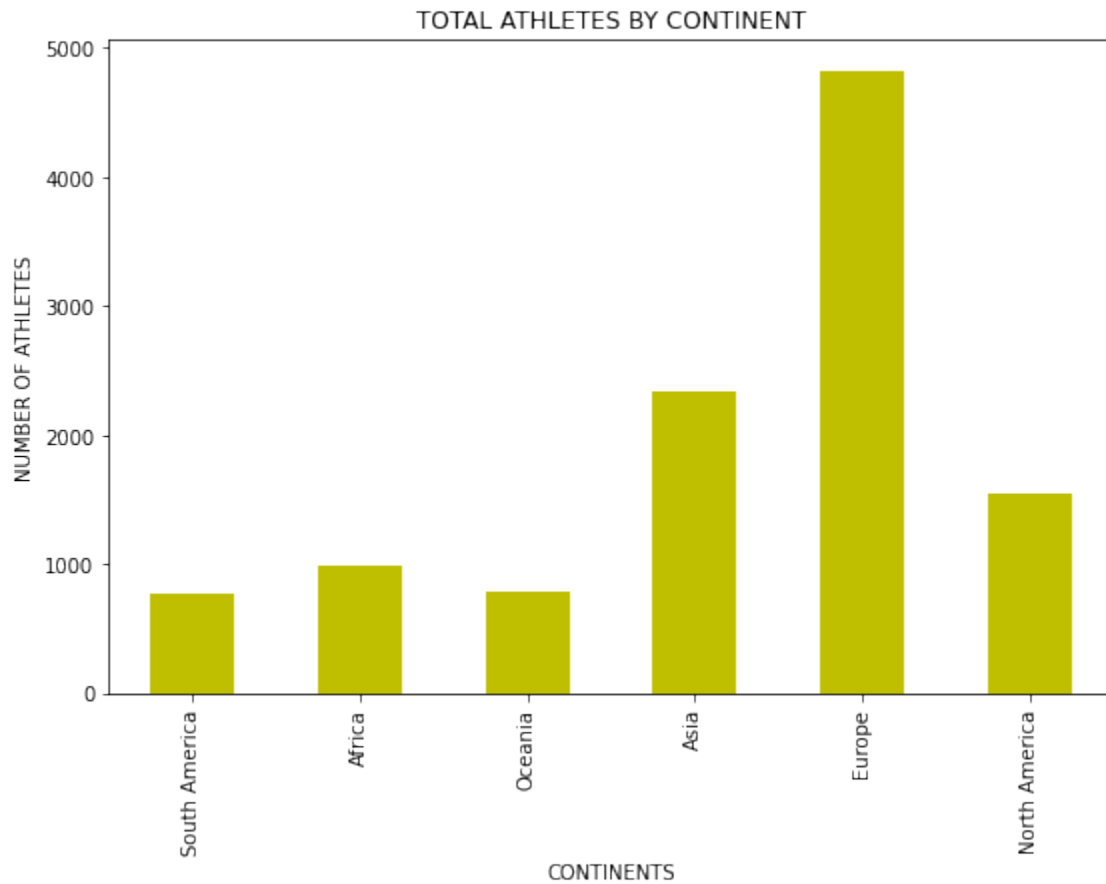
```
[297]: df_continents["ATHLETES"].plot(kind = 'bar',color = 'y',figsize = (9,6))

plt.title("TOTAL ATHLETES BY CONTINENT")

plt.xlabel("CONTINENTS")

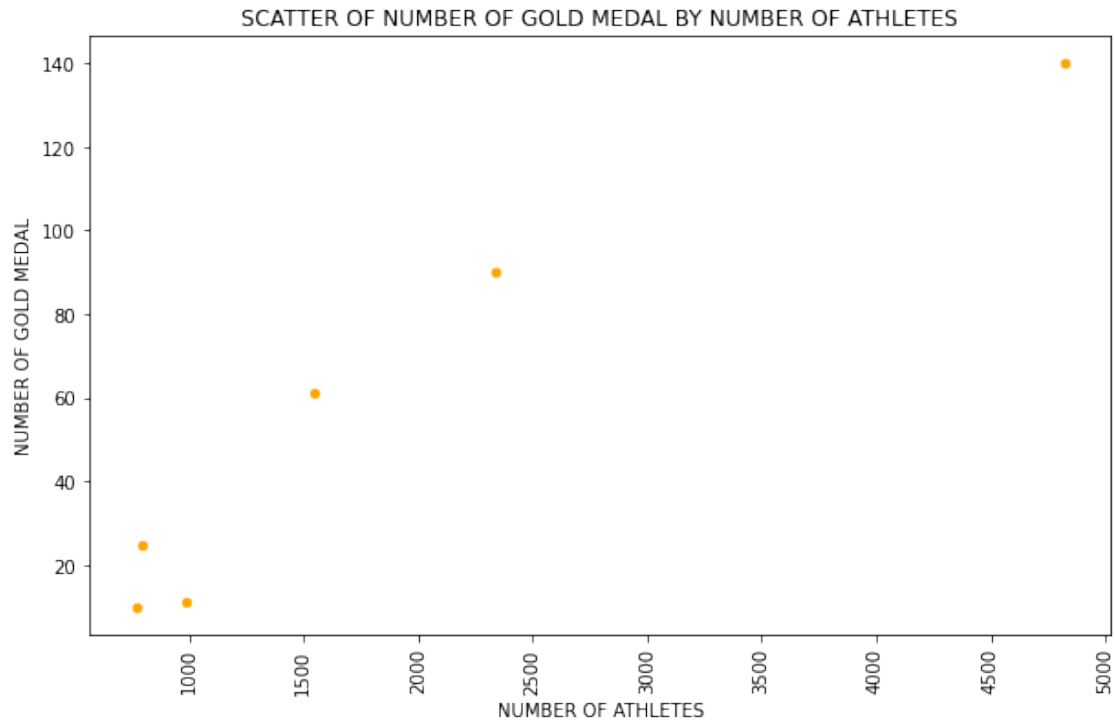
plt.ylabel("NUMBER OF ATHLETES")

plt.show()
```



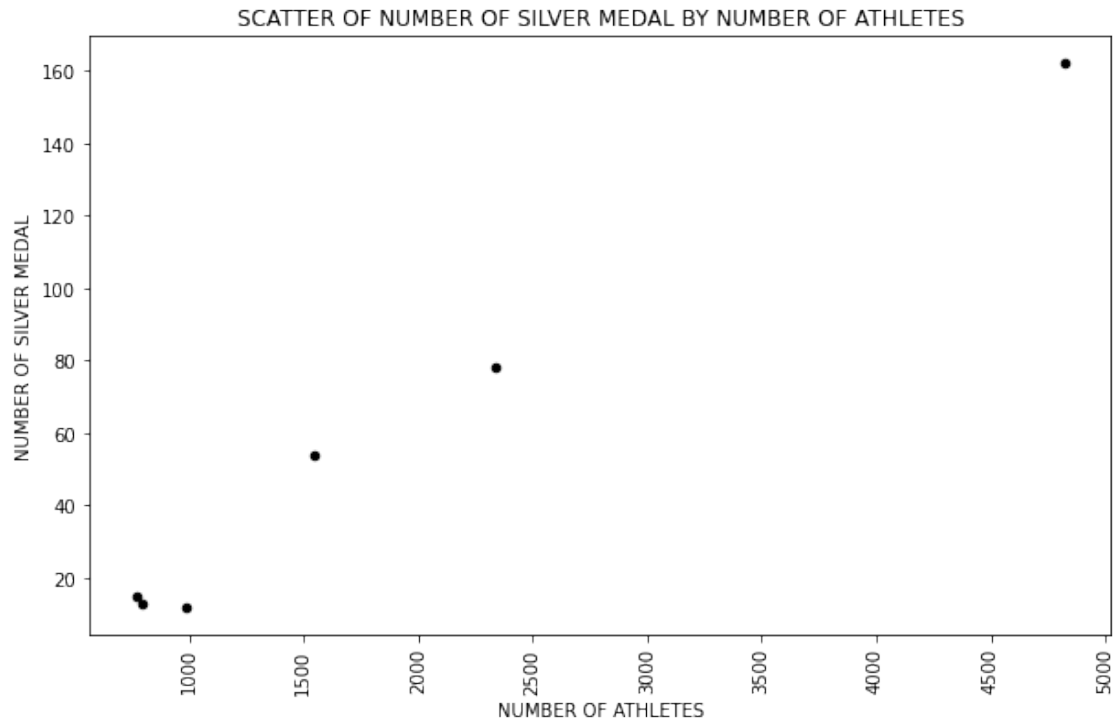
21 SCATTER OF NUMBER OF GOLD MEDAL BY NUMBER OF ATHLETES

```
[298]: df_continents.plot(kind = "scatter", x = "ATHLETES", y = "GOLD",figsize =(10,6), color = "orange")
plt.xlabel('NUMBER OF ATHLETES')
plt.ylabel('NUMBER OF GOLD MEDAL')
plt.xticks(rotation = 90)
plt.title('SCATTER OF NUMBER OF GOLD MEDAL BY NUMBER OF ATHLETES')
plt.show()
```



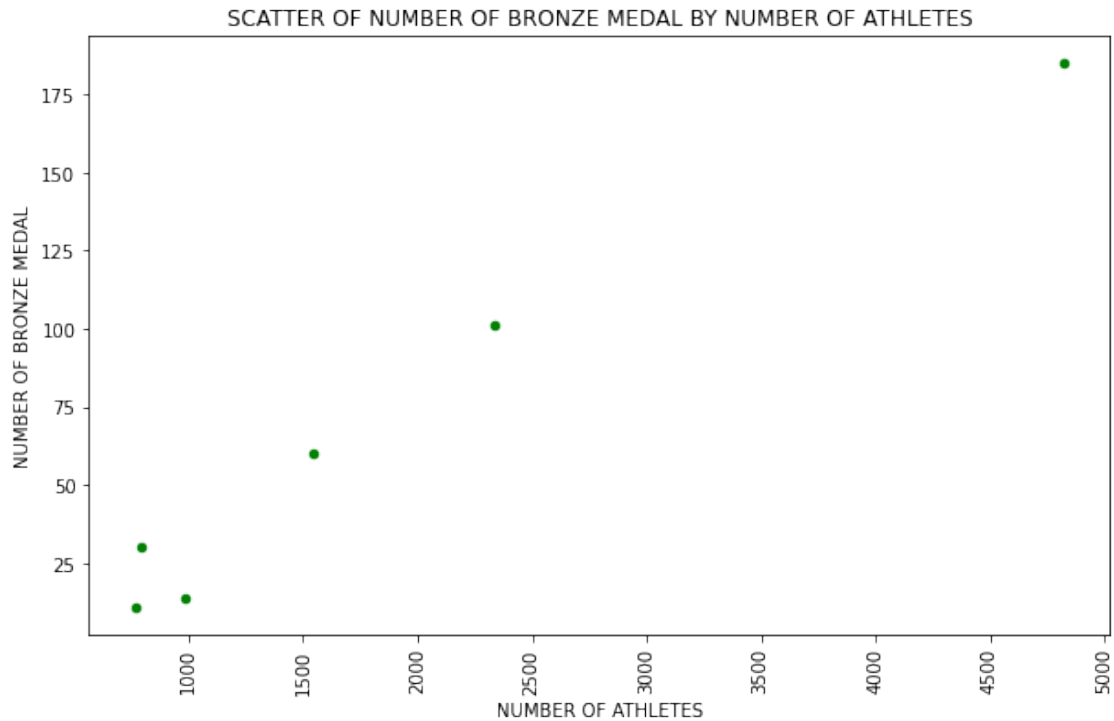
22 SCATTER OF NUMBER OF SILVER MEDAL BY NUMBER OF ATHLETES

```
[299]: df_continents.plot(kind = "scatter", x = "ATHLETES", y = "SILVER",figsize =(10,6), color = "k")
plt.xlabel('NUMBER OF ATHLETES')
plt.ylabel('NUMBER OF SILVER MEDAL')
plt.xticks(rotation = 90)
plt.title('SCATTER OF NUMBER OF SILVER MEDAL BY NUMBER OF ATHLETES')
plt.show()
```



23 SCATTER OF NUMBER OF BRONZE MEDAL BY NUMBER OF ATHLETES

```
[300]: df_continents.plot(kind = "scatter", x = "ATHLETES", y = "BRONZE",figsize =(
    ↳(10,6), color = "green")
plt.xlabel('NUMBER OF ATHLETES')
plt.ylabel('NUMBER OF BRONZE MEDAL')
plt.xticks(rotation = 90)
plt.title('SCATTER OF NUMBER OF BRONZE MEDAL BY NUMBER OF ATHLETES')
plt.show()
```



24 Find the correlation between the number of athletes and the total number of gold,silver,and bronze medals by continent.

```
[301]: corr_Athlete_Gold = df_continents["GOLD"].corr(df_continents["ATHLETES"])
```

```
[302]: corr_Athlete_Gold
```

```
[302]: 0.957238150372163
```

```
[303]: corr_Athlete_Gold1 = df_continents["ATHLETES"].corr(df_continents["GOLD"])
```

```
[304]: corr_Athlete_Gold1
```

```
[304]: 0.957238150372163
```

```
[305]: corr_Athlete_Silver = df_continents["SILVER"].corr(df_continents["ATHLETES"])
```

```
[306]: corr_Athlete_Silver
```

```
[306]: 0.9921634293063021
```

```
[307]: corr_Athlete_bronze = df_continents["BRONZE"].corr(df_continents["ATHLETES"])
```

```
[308]: corr_Athlete_bronze
```

```
[308]: 0.9842519574746943
```

```
[309]: correlation_df_continents = df_continents.corr()
```

```
[310]: correlation_df_continents
```

```
[310]:
```

	ATHLETES	GOLD	SILVER	BRONZE	\
ATHLETES	1.000000	0.957238	0.992163	0.984252	
GOLD	0.957238	1.000000	0.980132	0.990701	
SILVER	0.992163	0.980132	1.000000	0.992233	
BRONZE	0.984252	0.990701	0.992233	1.000000	
TOTAL_MEDAL_BY_CONTINENT	0.982848	0.993807	0.995174	0.998633	

	TOTAL_MEDAL_BY_CONTINENT
ATHLETES	0.982848
GOLD	0.993807
SILVER	0.995174
BRONZE	0.998633
TOTAL_MEDAL_BY_CONTINENT	1.000000

```
[311]: correlation_df_continents = correlation_df_continents.  
↳drop(["GOLD", "SILVER", "BRONZE"], axis = 1)
```

```
[312]: correlation_df_continents
```

```
[312]:
```

	ATHLETES	TOTAL_MEDAL_BY_CONTINENT
ATHLETES	1.000000	0.982848
GOLD	0.957238	0.993807
SILVER	0.992163	0.995174
BRONZE	0.984252	0.998633
TOTAL_MEDAL_BY_CONTINENT	0.982848	1.000000

25 Conclusion: Because we have OR (0.957238) SILVER (0.992163) BRONZEA (0.984252), it means a strong upward (positive) linear relationship between the number of medals and the number of athletes

```
[313]: df_continents
```

```
[313]:
```

	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT
CONTINENTS					
South America	766.0	10.0	15.0	11.0	36.0
Africa	984.0	11.0	12.0	14.0	37.0

Oceania	791.0	25.0	13.0	30.0	68.0
Asia	2334.0	90.0	78.0	101.0	269.0
Europe	4822.0	140.0	162.0	185.0	487.0
North America	1544.0	61.0	54.0	60.0	175.0

26 Create new column consisting of row sums across specific columns.

```
[314]: df_continents['TOTAL MEDAL BY CONTINENT'] = df_continents.iloc[:,0:4].
      ↪sum(axis=1)
```

```
[315]: df_continents2 = df_continents.iloc[:,0:4].sum(axis=1)
```

```
[316]: df_continents2
```

```
[316]: CONTINENTS
South America    802.0
Africa          1021.0
Oceania          859.0
Asia            2603.0
Europe          5309.0
North America   1719.0
dtype: float64
```

```
[317]: df_continents['TOTAL MEDAL BY CONTINENT']
```

```
[317]: CONTINENTS
South America    802.0
Africa          1021.0
Oceania          859.0
Asia            2603.0
Europe          5309.0
North America   1719.0
Name: TOTAL MEDAL BY CONTINENT, dtype: float64
```

```
[318]: df_continents
```

```
[318]:
```

	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT \
CONTINENTS					
South America	766.0	10.0	15.0	11.0	36.0
Africa	984.0	11.0	12.0	14.0	37.0
Oceania	791.0	25.0	13.0	30.0	68.0
Asia	2334.0	90.0	78.0	101.0	269.0
Europe	4822.0	140.0	162.0	185.0	487.0
North America	1544.0	61.0	54.0	60.0	175.0

TOTAL MEDAL BY CONTINENT	
CONTINENTS	
South America	802.0
Africa	1021.0
Oceania	859.0
Asia	2603.0
Europe	5309.0
North America	1719.0

```
[319]: df_continents
```

```
[319]:
```

	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT \
CONTINENTS					
South America	766.0	10.0	15.0	11.0	36.0
Africa	984.0	11.0	12.0	14.0	37.0
Oceania	791.0	25.0	13.0	30.0	68.0
Asia	2334.0	90.0	78.0	101.0	269.0
Europe	4822.0	140.0	162.0	185.0	487.0
North America	1544.0	61.0	54.0	60.0	175.0

TOTAL MEDAL BY CONTINENT	
CONTINENTS	
South America	802.0
Africa	1021.0
Oceania	859.0
Asia	2603.0
Europe	5309.0
North America	1719.0

27 Sort DataFrame based on column names or row

```
[320]: df_continents.sort_values(by='CONTINENTS', inplace=True)
```

```
[321]: df_continents
```

```
[321]:
```

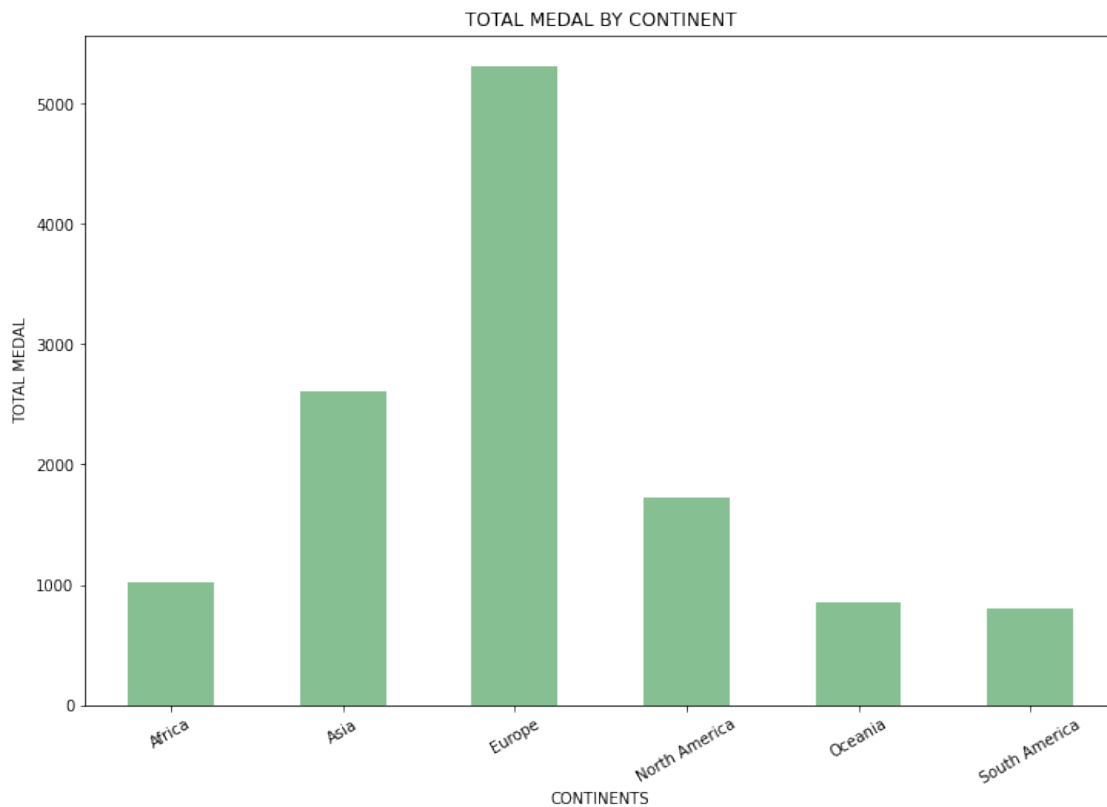
	ATHLETES	GOLD	SILVER	BRONZE	TOTAL_MEDAL_BY_CONTINENT \
CONTINENTS					
Africa	984.0	11.0	12.0	14.0	37.0
Asia	2334.0	90.0	78.0	101.0	269.0
Europe	4822.0	140.0	162.0	185.0	487.0
North America	1544.0	61.0	54.0	60.0	175.0
Oceania	791.0	25.0	13.0	30.0	68.0
South America	766.0	10.0	15.0	11.0	36.0

TOTAL MEDAL BY CONTINENT	
CONTINENTS	

Africa	1021.0
Asia	2603.0
Europe	5309.0
North America	1719.0
Oceania	859.0
South America	802.0

28 TOTAL MEDAL BY CONTINENT

```
[322]: df_continents['TOTAL MEDAL BY CONTINENT'].plot(kind = 'bar', figsize=(12,8),  
color='#86bf91')  
plt.title("TOTAL MEDAL BY CONTINENT")  
  
plt.xlabel("CONTINENTS")  
plt.xticks(rotation=30, horizontalalignment="center")  
plt.ylabel("TOTAL MEDAL")  
  
plt.show()
```



29 Which continent did better or worse at the Tokyo Olympics? Why? Is there an economic impact on the performance of athletes by continent?

30 To answer this question, I collected information from wikipedia.

Sources: [https://en.wikipedia.org/wiki/List_of_continents_by_GDP_\(nominal\)#cite_note-IMF01-3](https://en.wikipedia.org/wiki/List_of_continents_by_GDP_(nominal)#cite_note-IMF01-3)

```
[323]: #Import data from the web to Excel,  
  
#Import Excel File into Python using Pandas,  
  
#I combined data in Pandas by merging, joining and concating,  
  
#I selected rows from my DataFrames based on column values,  
  
#I have selected columns from my DataFrames by index,
```

```
[324]: import pandas as pdGDP  
import matplotlib.pyplot as plt  
%matplotlib inline
```

```
[325]: dfGDP = pdGDP.read_excel('/content/drive/MyDrive/Data_Engineering/Tokyo_OLympics 2020(2021)/Olympic Medal Count 2020.xlsx',  
                             sheet_name='GDP (nominal) by continents')
```

```
[326]: dfGDP
```

```
[326]:
```

	Rank	Continent	GDP (US\$billion)	Share (%)
0	NaN	World	93864.0	100.0
1	1.0	Asia	36383.0	38.8
2	2.0	North America	26298.0	28.0
3	3.0	Europe	23614.0	25.2
4	4.0	South America	3019.0	3.2
5	5.0	Africa	2601.0	2.8
6	6.0	Oceania	1894.0	2.0
7	NaN	Unaccounted	55.0	0.1

```
[327]: dfGDP.info()  
  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 8 entries, 0 to 7  
Data columns (total 4 columns):  
#   Column                Non-Null Count  Dtype  
---  -
```

```

0    Rank          6 non-null    float64
1    Continent      8 non-null    object
2    GDP (US$billion) 8 non-null    float64
3    Share (%)      8 non-null    float64
dtypes: float64(3), object(1)
memory usage: 384.0+ bytes

```

```
[328]: dfGDP
```

```
[328]:
```

	Rank	Continent	GDP (US\$billion)	Share (%)
0	NaN	World	93864.0	100.0
1	1.0	Asia	36383.0	38.8
2	2.0	North America	26298.0	28.0
3	3.0	Europe	23614.0	25.2
4	4.0	South America	3019.0	3.2
5	5.0	Africa	2601.0	2.8
6	6.0	Oceania	1894.0	2.0
7	NaN	Unaccounted	55.0	0.1

```
[329]: dfGDP.rename(columns={'Continent': 'CONTINENTS'},
                    inplace=True, errors='raise')
```

```
[330]: dfGDP
```

```
[330]:
```

	Rank	CONTINENTS	GDP (US\$billion)	Share (%)
0	NaN	World	93864.0	100.0
1	1.0	Asia	36383.0	38.8
2	2.0	North America	26298.0	28.0
3	3.0	Europe	23614.0	25.2
4	4.0	South America	3019.0	3.2
5	5.0	Africa	2601.0	2.8
6	6.0	Oceania	1894.0	2.0
7	NaN	Unaccounted	55.0	0.1

```
[331]: dfGDP.set_index('CONTINENTS', inplace=True)
```

```
[332]: dfGDP
```

```
[332]:
```

	Rank	GDP (US\$billion)	Share (%)
CONTINENTS			
World	NaN	93864.0	100.0
Asia	1.0	36383.0	38.8
North America	2.0	26298.0	28.0
Europe	3.0	23614.0	25.2
South America	4.0	3019.0	3.2
Africa	5.0	2601.0	2.8
Oceania	6.0	1894.0	2.0

Unaccounted	NaN	55.0	0.1
-------------	-----	------	-----

```
[333]: dfGDP.drop(['Rank', 'Share (%)'], inplace=True, axis=1)
```

```
[334]: dfGDP
```

```
[334]:
```

	GDP (US\$billion)
CONTINENTS	
World	93864.0
Asia	36383.0
North America	26298.0
Europe	23614.0
South America	3019.0
Africa	2601.0
Oceania	1894.0
Unaccounted	55.0

```
[335]: # Delete rows with index name
dfGDP2 = dfGDP.drop(['World' , 'Unaccounted'])
```

```
[336]: dfGDP2.sort_values(by='CONTINENTS')
```

```
[336]:
```

	GDP (US\$billion)
CONTINENTS	
Africa	2601.0
Asia	36383.0
Europe	23614.0
North America	26298.0
Oceania	1894.0
South America	3019.0

```
[337]: dfconticoncate = pd.concat([dfGDP2, df_continents], axis=1, join='inner')
```

```
[338]: dfconticoncate
```

```
[338]:
```

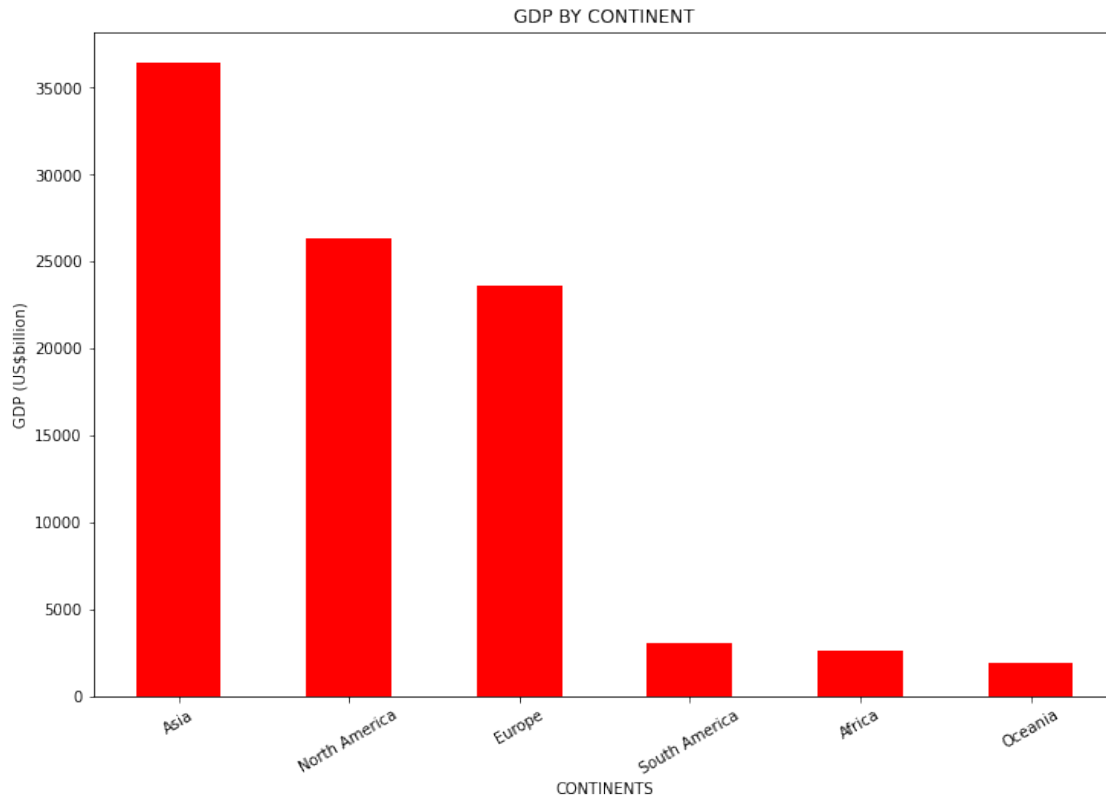
	GDP (US\$billion)	ATHLETES	GOLD	SILVER	BRONZE	\
CONTINENTS						
Asia	36383.0	2334.0	90.0	78.0	101.0	
North America	26298.0	1544.0	61.0	54.0	60.0	
Europe	23614.0	4822.0	140.0	162.0	185.0	
South America	3019.0	766.0	10.0	15.0	11.0	
Africa	2601.0	984.0	11.0	12.0	14.0	
Oceania	1894.0	791.0	25.0	13.0	30.0	
	TOTAL_MEDAL_BY_CONTINENT	TOTAL MEDAL BY CONTINENT				
CONTINENTS						
Asia		269.0			2603.0	

North America	175.0	1719.0
Europe	487.0	5309.0
South America	36.0	802.0
Africa	37.0	1021.0
Oceania	68.0	859.0

```
[339]: dfGDP2['GDP (US$billion)'].plot(kind = 'bar', figsize=(12,8), color='red')
plt.title("GDP BY CONTINENT")

plt.xlabel("CONTINENTS")
plt.xticks(rotation=30, horizontalalignment="center")
plt.ylabel("GDP (US$billion)")

plt.show()
```



```
[340]: dfGDP
```

```
[340]:          GDP (US$billion)
CONTINENTS
World          93864.0
Asia           36383.0
```

North America	26298.0
Europe	23614.0
South America	3019.0
Africa	2601.0
Oceania	1894.0
Unaccounted	55.0

31 Find the correlation between the GDP (billions of US dollars) by continent and the total number of medal by continent.

```
[341]: corr_GDP_Total_medals = dfconticoncate["GDP (US$billion)"].
      ↪corr(dfconticoncate["TOTAL MEDAL BY CONTINENT"])
```

```
[342]: corr_GDP_Total_medals
```

```
[342]: 0.6101211549068609
```

32 Conclusion: corr_GDP_Total_medals = 0.61: A moderate positive relationship.

33 Find the correlation between the GDP (billions of US dollars) by continent and the total number of athlete by continent.

```
[343]: corr_GDP_Total_Athletes = dfconticoncate["GDP (US$billion)"].
      ↪corr(dfconticoncate["ATHLETES"])
```

```
[344]: corr_GDP_Total_Athletes
```

```
[344]: 0.5972888225383218
```

34 Conclusion: corr_GDP_Total_Athletes = 0.59: A moderate positive relationship.

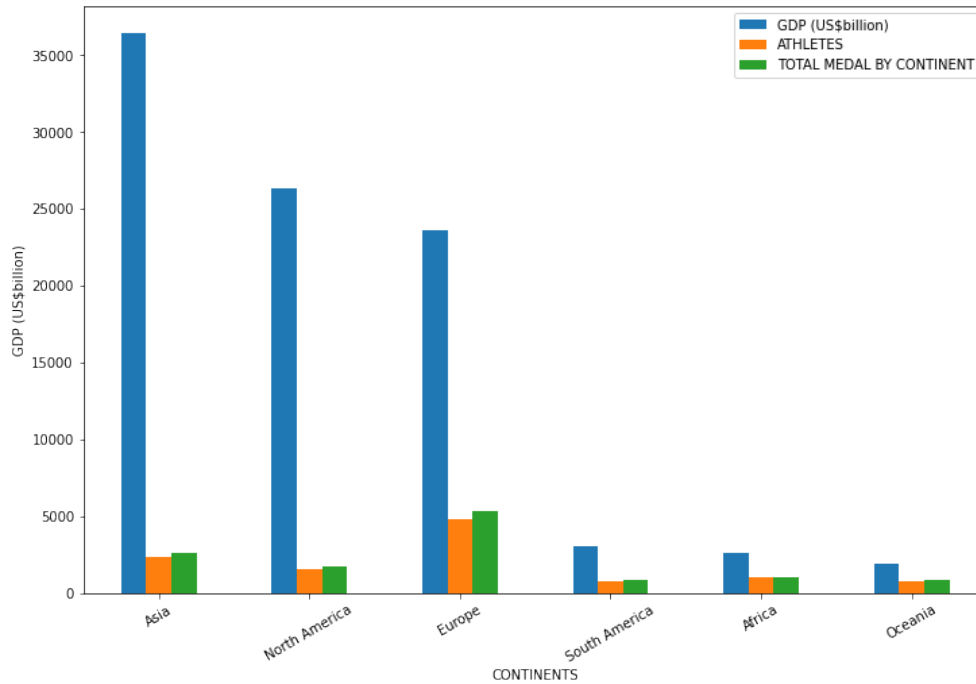
35 REPRESENTATION OF GDP (in billions of US dollars) BY CONTINENT, TOTAL ATHLETES BY CONTINENT AND TOTAL MEDALS BY CONTINENT

```
[345]: dfconticoncate[['GDP (US$billion)', 'ATHLETES', 'TOTAL MEDAL BY CONTINENT']].
      ↪plot(kind = 'bar', figsize=(12,8))
plt.title("REPRESENTATION OF GDP (in billions of US dollars) BY CONTINENT,
      ↪TOTAL ATHLETES BY CONTINENT AND TOTAL MEDALS BY CONTINENT")
```

```
plt.xlabel("CONTINENTS")
plt.xticks(rotation=30, horizontalalignment="center")
plt.ylabel("GDP (US$billion)")

plt.show()
```

REPRESENTATION OF GDP (in billions of US dollars) BY CONTINENT, TOTAL ATHLETES BY CONTINENT AND TOTAL MEDALS BY CONTINENT



36 Conclusion: No strong impact of the economy on the performance of athletes by continent