Submittes by: Muhammad Arham Adeel

Ex2 - Filtering and Sorting Data

This time we are going to pull data directly from the internet.

Step 1. Import the necessary libraries

```
In [1]:   import pandas as pd
          import numpy as np
```

Step 2. Import the dataset from this address.

```
In [2]:   url = "https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/02_Filteri
```

Step 3. Assign it to a variable called euro12.

```
In [3]:   euro12 = pd.read_csv(url)
          euro12.head()
```

Out[3]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...	S n
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	...	
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	...	
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	...	
3	England	5	11	18	50.0%	17.2%	40	0	0	0	...	
4	France	3	22	24	37.9%	6.5%	65	1	0	0	...	

5 rows × 35 columns

Step 4. Select only the Goal column.

```
In [4]:   euro12.columns
```

```
Out[4]:    Index(['Team', 'Goals', 'Shots on target', 'Shots off target',
                  'Shooting Accuracy', '% Goals-to-shots', 'Total shots (inc. Blocked)',
                  'Hit Woodwork', 'Penalty goals', 'Penalties not scored', 'Headed goals',
                  'Passes', 'Passes completed', 'Passing Accuracy', 'Touches', 'Crosses',
                  'Dribbles', 'Corners Taken', 'Tackles', 'Clearances', 'Interceptions',
                  'Clearances off line', 'Clean Sheets', 'Blocks', 'Goals conceded',
                  'Saves made', 'Saves-to-shots ratio', 'Fouls Won', 'Fouls Conceded',
                  'Offsides', 'Yellow Cards', 'Red Cards', 'Subs on', 'Subs off',
                  'Players Used'],
                 dtype='object')
```

In [5]: `euro12[["Goals"]]`

Out[5]:

	Goals
0	4
1	4
2	4
3	5
4	3
5	10
6	5
7	6
8	2
9	2
10	6
11	1
12	5
13	12
14	5
15	2

Step 5. How many team participated in the Euro2012?

In [7]: `euro12.Team.count()`

Out[7]: 16

Step 6. What is the number of columns in the dataset?

In [8]: `euro12.shape`

Out[8]: `(16, 35)`

In [9]: `print("The number of column present in this dataset is: ", euro12.shape[1])`

```
The number of column present in this dataset is:  35
```

Step 7. View only the columns Team, Yellow Cards and Red Cards and assign them to a dataframe called discipline

In [10]:
```python
discipline = euro12[['Team','Yellow Cards','Red Cards']]
discipline
```

Out[10]:

	Team	Yellow Cards	Red Cards
0	Croatia	9	0
1	Czech Republic	7	0
2	Denmark	4	0
3	England	5	0
4	France	6	0
5	Germany	4	0
6	Greece	9	1
7	Italy	16	0
8	Netherlands	5	0
9	Poland	7	1
10	Portugal	12	0
11	Republic of Ireland	6	1
12	Russia	6	0
13	Spain	11	0
14	Sweden	7	0
15	Ukraine	5	0

Step 8. Sort the teams by Red Cards, then to Yellow Cards

In [12]:
```python
discipline.sort_values(by = ['Red Cards', 'Yellow Cards'], ascending=False)
```

Out[12]:

	Team	Yellow Cards	Red Cards
6	Greece	9	1
9	Poland	7	1
11	Republic of Ireland	6	1
7	Italy	16	0
10	Portugal	12	0
13	Spain	11	0
0	Croatia	9	0
1	Czech Republic	7	0
14	Sweden	7	0
4	France	6	0
12	Russia	6	0
3	England	5	0
8	Netherlands	5	0
15	Ukraine	5	0
2	Denmark	4	0
5	Germany	4	0

Step 9. Calculate the mean Yellow Cards given per Team

In [13]:
```python
discipline.groupby("Team")['Yellow Cards'].mean()
```

Out[13]:
```
Team
Croatia               9.0
Czech Republic        7.0
Denmark               4.0
England               5.0
France                6.0
Germany               4.0
Greece                9.0
Italy                16.0
Netherlands           5.0
Poland                7.0
Portugal             12.0
Republic of Ireland   6.0
Russia                6.0
Spain                11.0
Sweden                7.0
Ukraine               5.0
Name: Yellow Cards, dtype: float64
```

Step 10. Filter teams that scored more than 6 goals

In [15]:
```python
euro12[euro12["Goals"] > 6]
```

Out[15]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	...
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	...

2 rows × 35 columns

Step 11. Select the teams that start with G

In [25]:
```python
euro12[euro12["Team"] == "G"]
```

Out[25]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...	Saves made

0 rows × 35 columns

Step 12. Select the first 7 columns

In [26]:
```python
euro12.head(7)
```

Out[26]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...	S n
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	...	
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	...	
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	...	
3	England	5	11	18	50.0%	17.2%	40	0	0	0	...	
4	France	3	22	24	37.9%	6.5%	65	1	0	0	...	
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	...	
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	...	

7 rows × 35 columns

Step 13. Select all columns except the last 3.

In [30]:
```python
euro12.iloc[ : , 0:32 ]
```

Out[30]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	..
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	..
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	..
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	..
3	England	5	11	18	50.0%	17.2%	40	0	0	0	..
4	France	3	22	24	37.9%	6.5%	65	1	0	0	..
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	..
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	..
7	Italy	6	34	45	43.0%	7.5%	110	2	0	0	..
8	Netherlands	2	12	36	25.0%	4.1%	60	2	0	0	..
9	Poland	2	15	23	39.4%	5.2%	48	0	0	0	..
10	Portugal	6	22	42	34.3%	9.3%	82	6	0	0	..
11	Republic of Ireland	1	7	12	36.8%	5.2%	28	0	0	0	..
12	Russia	5	9	31	22.5%	12.5%	59	2	0	0	..
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	..
14	Sweden	5	17	19	47.2%	13.8%	39	3	0	0	..
15	Ukraine	2	7	26	21.2%	6.0%	38	0	0	0	..

16 rows × 32 columns

Step 14. Present only the Shooting Accuracy from England, Italy and Russia

In [33]:
```
prsnt_shnt_cnt = euro12[(euro12["Team"] == "England") | (euro12["Team"] == "Italy") |
prsnt_shnt_cnt
```

Out[33]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	...	S n
3	England	5	11	18	50.0%	17.2%	40	0	0	0	...	
7	Italy	6	34	45	43.0%	7.5%	110	2	0	0	...	
12	Russia	5	9	31	22.5%	12.5%	59	2	0	0	...	

3 rows × 35 columns

In [34]: `prsnt_shnt_cnt[["Team","Shooting Accuracy"]]`

Out[34]:

	Team	Shooting Accuracy
3	England	50.0%
7	Italy	43.0%
12	Russia	22.5%