Exercise - Filtering and Sorting Data-Euro Team Dataset

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

Step 1. Import the necessary libraries

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

Step 2. Import the dataset from this address.

In [2]:
```
pd.read_csv("https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/0
```

Out[2]:

	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

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

Step 3. Assign it to a variable called euro12.

In [5]:
```
euro12=pd.read_csv("https://raw.githubusercontent.com/guipsamora/pandas_exercises/r
euro12.head()
```

Out[5]:

	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 rows × 35 columns

Step 4. Select only the Goal column.

In [7]:
```
euro12.Goals
```

Out[7]:
```
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
Name: Goals, dtype: int64
```

Step 5. How many team participated in the Euro2012?

In [14]:
```
euro12.Team.count()
```

Out[14]:
```
16
```

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

In [15]:
```
euro12.shape[1]
```

Out[15]:
```
35
```

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

In [23]:
```
discipline=euro12.loc[:,['Team','Yellow Cards','Red Cards']]
discipline
```

Out[23]:

	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 [44]:
```
print(discipline.sort_values(['Red Cards', 'Yellow Cards'],ascending=False))
```

```
                  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 [43]:
```
discipline['Yellow Cards'].mean()
```

Out[43]:
```
7.4375
```

Step 10. Filter teams that scored more than 6 goals

In [61]:
```
euro12[['Team','Goals']][(euro12.Goals>6)]
```

Out[61]:

	Team	Goals
5	Germany	10
13	Spain	12

Step 11. Select the teams that start with G

In [62]:
```
euro12[euro12.Team.str.startswith('G')]
```

Out[62]:

	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	...
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	...

2 rows × 35 columns

Step 12. Select the first 7 columns

In [66]:
```
euro12.iloc[:,:7]
```

Out[66]:

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)
0	Croatia	4	13	12	51.9%	16.0%	32
1	Czech Republic	4	13	18	41.9%	12.9%	39
2	Denmark	4	10	10	50.0%	20.0%	27
3	England	5	11	18	50.0%	17.2%	40
4	France	3	22	24	37.9%	6.5%	65
5	Germany	10	32	32	47.8%	15.6%	80
6	Greece	5	8	18	30.7%	19.2%	32
7	Italy	6	34	45	43.0%	7.5%	110
8	Netherlands	2	12	36	25.0%	4.1%	60
9	Poland	2	15	23	39.4%	5.2%	48
10	Portugal	6	22	42	34.3%	9.3%	82
11	Republic of Ireland	1	7	12	36.8%	5.2%	28
12	Russia	5	9	31	22.5%	12.5%	59
13	Spain	12	42	33	55.9%	16.0%	100
14	Sweden	5	17	19	47.2%	13.8%	39

	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)
15	Ukraine	2	7	26	21.2%	6.0%	38

Step 13. Select all columns except the last 3.

In [69]:
```
euro12.head()
```

Out[69]:

	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 rows × 35 columns

In [68]:
```
euro12.iloc[:,:-3]
```

Out[68]:

	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

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

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

In [81]:
```
Shooting_Accuracy=euro12.set_index("Team")
Shooting_Accuracy
```

Out[81]:

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

Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	Penalties not scored	H
Russia	5	9	31	22.5%	12.5%	59	2	0	0	
Spain	12	42	33	55.9%	16.0%	100	0	1	0	
Sweden	5	17	19	47.2%	13.8%	39	3	0	0	
Ukraine	2	7	26	21.2%	6.0%	38	0	0	0	

In [82]:
```
Shooting_Accuracy.loc[['England', 'Italy', 'Russia']]['Shooting Accuracy']
```

Out[82]:
```
Team
England    50.0%
Italy      43.0%
Russia     22.5%
Name: Shooting Accuracy, dtype: object
```

In [83]:
```
# .loc is another way to slice, using the labels of the columns and indexes

euro12.loc[euro12.Team.isin(['England', 'Italy', 'Russia']), ['Team','Shooting Acc
```

Out[83]:

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

In [ ]: