Chapter 4 - Exercise 3: Lọc và sắp xếp dữ liệu UEFA_European_Championship/Euro 2012

Tạo data frame euro12 từ dữ liệu trên. In euro12: type, shape, danh sách các cột

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

```
In [2]:  # Đọc dữ liệu từ tập tin euro2012.csv vào dataframe euro12

         euro12 = pd.read_csv('data/euro2012.csv', sep=',', index_col=0)

         # Hiển thị thông tin type(euro12), shape, columns và 5 dòng đầu của euro12
         print(type(euro12))
         print(euro12.shape)
         print(euro12.columns)
         euro12.head()
```

```
<class 'pandas.core.frame.DataFrame'>
(16, 35)
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')
```

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	...	Saves made
0	Croatia	4	13	12	51.9%	16.0%	32	0	0	0	...	13
1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0	0	...	9
2	Denmark	4	10	10	50.0%	20.0%	27	1	0	0	...	10
3	England	5	11	18	50.0%	17.2%	40	0	0	0	...	22
4	France	3	22	24	37.9%	6.5%	65	1	0	0	...	6

5 rows × 35 columns

```
In [3]:  # Câu 1: Cho biết 5 đội có cột Goals cao nhất
         euro12[['Team', 'Goals']].sort_values(by='Goals', ascending=False).head(5)
```

Out[3]:

	Team	Goals
13	Spain	12
5	Germany	10
7	Italy	6
10	Portugal	6
3	England	5

```
In [4]:  euro12['Goals'].values
```

Out[4]: array([4, 4, 4, 5, 3, 10, 5, 6, 2, 2, 6, 1, 5, 12, 5, 2],
               dtype=int64)

```
In [5]:  euro12['Goals'].tolist()
```

Out[5]: [4, 4, 4, 5, 3, 10, 5, 6, 2, 2, 6, 1, 5, 12, 5, 2]

```
In [6]:  # Câu 2: Có bao nhiêu đội tham gia Euro2012?
         euro12.shape[0]
```

Out[6]: 16

```
In [7]:  # Câu 3: In thông tin của Euro2012
         euro12.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16 entries, 0 to 15
Data columns (total 35 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Team                        16 non-null     object
 1   Goals                       16 non-null     int64
 2   Shots on target             16 non-null     int64
 3   Shots off target            16 non-null     int64
 4   Shooting Accuracy           16 non-null     object
 5   % Goals-to-shots            16 non-null     object
 6   Total shots (inc. Blocked)  16 non-null     int64
 7   Hit Woodwork                16 non-null     int64
 8   Penalty goals               16 non-null     int64
 9   Penalties not scored        16 non-null     int64
 10  Headed goals                16 non-null     int64
 11  Passes                      16 non-null     int64
 12  Passes completed            16 non-null     int64
 13  Passing Accuracy            16 non-null     object
 14  Touches                     16 non-null     int64
 15  Crosses                     16 non-null     int64
 16  Dribbles                    16 non-null     int64
 17  Corners Taken               16 non-null     int64
 18  Tackles                     16 non-null     int64
 19  Clearances                  16 non-null     int64
 20  Interceptions               16 non-null     int64
 21  Clearances off line         15 non-null     float64
 22  Clean Sheets                16 non-null     int64
 23  Blocks                      16 non-null     int64
 24  Goals conceded              16 non-null     int64
 25  Saves made                  16 non-null     int64
 26  Saves-to-shots ratio        16 non-null     object
 27  Fouls Won                   16 non-null     int64
 28  Fouls Conceded              16 non-null     int64
 29  Offsides                    16 non-null     int64
 30  Yellow Cards                16 non-null     int64
 31  Red Cards                   16 non-null     int64
 32  Subs on                     16 non-null     int64
 33  Subs off                    16 non-null     int64
 34  Players Used                16 non-null     int64
dtypes: float64(1), int64(29), object(5)
memory usage: 4.5+ KB
```

In [8]:
```python
# Câu 4: Tạo 1 data frame mới từ euro12 có tên là discipline (kỷ luật)
# chỉ chứa 3 cột 'Team', 'Yellow Cards', 'Red Cards'
discipline = euro12[['Team', 'Yellow Cards', 'Red Cards']]
discipline
```

Out[8]:

	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

In [9]:
```python
# Câu 5: Sắp xếp discipline giảm dần theo 2 cột 'Red Cards', 'Yellow Cards'
discipline.sort_values(by=['Red Cards', 'Yellow Cards'],
                       ascending = [False, False])
```

Out[9]:

	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

```
In [10]:  # Câu 6a: Tính trung bình Yellow Cards
          discipline[['Yellow Cards']].mean()

Out[10]:  Yellow Cards    7.4375
          dtype: float64

In [11]:  # Câu 6b: Lọc các đội đã ghi hơn 6 bàn thắng
          euro12[euro12.Goals > 6]
```

Out[11]:

	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
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	...	10
13	Spain	12	42	33	55.9%	16.0%	100	0	1	0	...	15

2 rows × 35 columns

```
In [12]:  # Câu 6c: In ra tên các đội bị phạt thẻ vàng, sắp giảm theo thẻ vàng,
          # in 5 dòng đầu tiên
          discipline[discipline['Yellow Cards']>0].sort_values(by='Yellow Cards',
                            ascending=False)[['Team', 'Yellow Cards']].head()
```

Out[12]:

	Team	Yellow Cards
7	Italy	16
10	Portugal	12
13	Spain	11
0	Croatia	9
6	Greece	9

```
In [13]:  # Câu 6d: In ra các đội có Goals > 6 và cột Yellow Cards >0
          euro12[(euro12.Goals > 6) & (euro12['Yellow Cards'] > 0)][['Team','Goals','Yellow Cards
```

Out[13]:

	Team	Goals	Yellow Cards
5	Germany	10	4
13	Spain	12	11

```
In [14]:  # Câu 7: In các đội mà tên bắt đầu bằng 'G' (gợi ý: dùng str.startswith())
          euro12[euro12.Team.str.startswith('G')]
```

Out[14]:

	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
5	Germany	10	32	32	47.8%	15.6%	80	2	1	0	...	10
6	Greece	5	8	18	30.7%	19.2%	32	1	1	1	...	13

2 rows × 35 columns

```
In [15]:  # Câu 8: In 7 cột đầu của euro12
          euro12.iloc[: , 0:7]
```

Out[15]:

	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
15	Ukraine	2	7	26	21.2%	6.0%	38

In [16]:
```python
# Câu 9a: In tất cả các cột, trừ 3 cột cuối
euro12.iloc[: , :-3]
```

Out[16]:

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

```
In [17]:   # Câu 9b: In 3 cột cuối
           euro12.iloc[: , -3:]
```

Out[17]:

	Subs on	Subs off	Players Used
0	9	9	16
1	11	11	19
2	7	7	15
3	11	11	16
4	11	11	19
5	15	15	17
6	12	12	20
7	18	18	19
8	7	7	15
9	7	7	17
10	14	14	16
11	10	10	17
12	7	7	16
13	17	17	18
14	9	9	18
15	9	9	18

```
In [18]:   # Câu 10: In các cột Team, Goals, Shooting Accuracy, Yellow Cards, Red Cards
           euro12[['Team', 'Goals', 'Shooting Accuracy', 'Yellow Cards', 'Red Cards']]
```

Out[18]:

	Team	Goals	Shooting Accuracy	Yellow Cards	Red Cards
0	Croatia	4	51.9%	9	0
1	Czech Republic	4	41.9%	7	0
2	Denmark	4	50.0%	4	0
3	England	5	50.0%	5	0
4	France	3	37.9%	6	0
5	Germany	10	47.8%	4	0
6	Greece	5	30.7%	9	1
7	Italy	6	43.0%	16	0
8	Netherlands	2	25.0%	5	0
9	Poland	2	39.4%	7	1
10	Portugal	6	34.3%	12	0
11	Republic of Ireland	1	36.8%	6	1
12	Russia	5	22.5%	6	0
13	Spain	12	55.9%	11	0
14	Sweden	5	47.2%	7	0
15	Ukraine	2	21.2%	5	0

```
In [19]:   # Câu 11a: In các cột chỉ hiển thị 'Team','Shooting Accuracy'
           # từ 'England', 'Italy', 'Russia'
           euro12.loc[euro12.Team.isin(['England', 'Italy', 'Russia']), ['Team','Shooting Accuracy
```

Out[19]:

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

```
In [20]:   # Câu 11b: chuyển kiểu cột Shooting Accuracy sang kiểu float
           euro12['Shooting Accuracy'].dtype
```

Out[20]: dtype('O')

```
In [21]:   euro12['Shooting Accuracy'] = euro12['Shooting Accuracy'].str.replace('%', '')
           euro12['Shooting Accuracy'] = euro12['Shooting Accuracy'].astype('float')
```

```
In [22]:   # Câu 11c: In các đội có cột Shooting Accuracy >= 50.0%
           euro12[euro12['Shooting Accuracy']>=50.0][['Team','Shooting Accuracy']]
```

Out[22]:

	Team	Shooting Accuracy
0	Croatia	51.9
2	Denmark	50.0
3	England	50.0
13	Spain	55.9

```
In [ ]:
```