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

1. In thông tin cột Goals
2. Có bao nhiêu đội tham gia Euro2012?
3. In thông tin của Euro2012
4. Tạo 1 data frame mới từ euro12 có tên là discipline chỉ chứa 3 cột 'Team', 'Yellow Cards', 'Red Cards'
5. Sắp xếp discipline giảm dần theo 2 cột 'Red Cards', 'Yellow Cards'
6. Lọc các đội đã ghi hơn 6 bàn thắng
7. In các đội mà tên bắt đầu bằng 'G' dùng str.startswith('G')
8. In 7 cột đầu của euro12
9. In tất cả các cột trừ 3 cột cuối
10. In các cột Team, Goals, Shooting Accuracy, Yellow Cards, Red Cards
11. In các cột Chỉ hiển thị 'Team','Shooting Accuracy' từ 'England', 'Italy', 'Russia'

In [1]:
```
%config IPCompleter.greedy = True
import numpy as np
import pandas as pd
```

Lấy dữ liệu từ đường dẫn

Euro 12 (https://raw.githubusercontent.com/jokecamp/FootballData/master/UEFA_European_Championship/E

Tạo data frame euro12 tử dữ liệu trên. In euro2: type, shape, danh sách các cột

In [2]:
```
euro12 = pd.read_csv('euro2012.csv', sep=',')
print(type(euro12))
print(euro12.shape)
print(euro12.columns)
euro12
```

```
<class 'pandas.core.frame.DataFrame'>
(16, 36)
Index(['Unnamed: 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', 'Headed goal
s',
       '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]:

	Unnamed: 0	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals
0	0	Croatia	4	13	12	51.9%	16.0%	32	0	0
1	1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0
2	2	Denmark	4	10	10	50.0%	20.0%	27	1	0
3	3	England	5	11	18	50.0%	17.2%	40	0	0
4	4	France	3	22	24	37.9%	6.5%	65	1	0
5	5	Germany	10	32	32	47.8%	15.6%	80	2	1
6	6	Greece	5	8	18	30.7%	19.2%	32	1	1
7	7	Italy	6	34	45	43.0%	7.5%	110	2	0
8	8	Netherlands	2	12	36	25.0%	4.1%	60	2	0
9	9	Poland	2	15	23	39.4%	5.2%	48	0	0
10	10	Portugal	6	22	42	34.3%	9.3%	82	6	0
11	11	Republic of Ireland	1	7	12	36.8%	5.2%	28	0	0
12	12	Russia	5	9	31	22.5%	12.5%	59	2	0
13	13	Spain	12	42	33	55.9%	16.0%	100	0	1
14	14	Sweden	5	17	19	47.2%	13.8%	39	3	0
15	15	Ukraine	2	7	26	21.2%	6.0%	38	0	0

16 rows × 36 columns

In [3]:
```
# Câu 1: In thông tin cột Goals
euro12['Goals']
```

Out[3]:
```
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
```

In [4]:
```
# Câu 2: Có bao nhiêu đội tham gia Euro2012?
euro12.shape[0]
```

Out[4]:
```
16
```

In [5]:
```
# Câu 3: In thông tin của Euro2012
euro12.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15
Data columns (total 36 columns):
Unnamed: 0                    16 non-null int64
Team                          16 non-null object
Goals                         16 non-null int64
Shots on target               16 non-null int64
Shots off target              16 non-null int64
Shooting Accuracy             16 non-null object
% Goals-to-shots              16 non-null object
Total shots (inc. Blocked)    16 non-null int64
Hit Woodwork                  16 non-null int64
Penalty goals                 16 non-null int64
Penalties not scored          16 non-null int64
Headed goals                  16 non-null int64
Passes                        16 non-null int64
Passes completed              16 non-null int64
Passing Accuracy              16 non-null object
Touches                       16 non-null int64
Crosses                       16 non-null int64
Dribbles                      16 non-null int64
Corners Taken                 16 non-null int64
Tackles                       16 non-null int64
Clearances                    16 non-null int64
Interceptions                 16 non-null int64
Clearances off line           15 non-null float64
Clean Sheets                  16 non-null int64
Blocks                        16 non-null int64
Goals conceded                16 non-null int64
Saves made                    16 non-null int64
Saves-to-shots ratio          16 non-null object
Fouls Won                     16 non-null int64
Fouls Conceded                16 non-null int64
Offsides                      16 non-null int64
Yellow Cards                  16 non-null int64
Red Cards                     16 non-null int64
Subs on                       16 non-null int64
Subs off                      16 non-null int64
Players Used                  16 non-null int64
dtypes: float64(1), int64(30), object(5)
memory usage: 4.6+ KB
```

In [6]:
```
# Câu 4: Tạo 1 data frame mới từ euro12 có tên là discipline chỉ chứa 3 cột 'Team'
discipline = euro12[['Team', 'Yellow Cards', 'Red Cards']]
discipline
```

Out[6]:

	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 [7]:
```
# Câu 5: Sắp xếp discipline giảm dần theo 2 cột 'Red Cards', 'Yellow Cards'
discipline.sort_values(['Red Cards', 'Yellow Cards'], ascending = False)
```

Out[7]:

	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 [8]:
```
# Câu 6: Tính trung bình Yellow Cards
round(discipline['Yellow Cards'].mean())
```

Out[8]: 7

In [9]:
```
# Câu 6: Lọc các đội đã ghi hơn 6 bàn thắng
euro12[euro12.Goals > 6]
```

Out[9]:

	Unnamed: 0	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	..
5	5	Germany	10	32	32	47.8%	15.6%	80	2	1	..
13	13	Spain	12	42	33	55.9%	16.0%	100	0	1	..

2 rows × 36 columns

Step 11. Select the teams that start with G

In [10]:
```
# Câu 7: In các đội mà tên bắt đầu bằng 'G' dùng str.startswith('G')
euro12[euro12.Team.str.startswith('G')]
```

Out[10]:

	Unnamed: 0	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals	...
5	5	Germany	10	32	32	47.8%	15.6%	80	2	1	...
6	6	Greece	5	8	18	30.7%	19.2%	32	1	1	...

2 rows × 36 columns

In [11]:
```
# Câu 8: In 7 cột đầu của euro12
euro12.iloc[: , 0:7]
```

Out[11]:

	Unnamed: 0	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots
0	0	Croatia	4	13	12	51.9%	16.0%
1	1	Czech Republic	4	13	18	41.9%	12.9%
2	2	Denmark	4	10	10	50.0%	20.0%
3	3	England	5	11	18	50.0%	17.2%
4	4	France	3	22	24	37.9%	6.5%
5	5	Germany	10	32	32	47.8%	15.6%
6	6	Greece	5	8	18	30.7%	19.2%
7	7	Italy	6	34	45	43.0%	7.5%
8	8	Netherlands	2	12	36	25.0%	4.1%
9	9	Poland	2	15	23	39.4%	5.2%
10	10	Portugal	6	22	42	34.3%	9.3%
11	11	Republic of Ireland	1	7	12	36.8%	5.2%
12	12	Russia	5	9	31	22.5%	12.5%
13	13	Spain	12	42	33	55.9%	16.0%
14	14	Sweden	5	17	19	47.2%	13.8%
15	15	Ukraine	2	7	26	21.2%	6.0%

In [12]:
```
# Câu 9: In tất cả các cột trừ 3 cột cuối
euro12.iloc[: , :-3]
```

Out[12]:

	Unnamed: 0	Team	Goals	Shots on target	Shots off target	Shooting Accuracy	% Goals-to-shots	Total shots (inc. Blocked)	Hit Woodwork	Penalty goals
0	0	Croatia	4	13	12	51.9%	16.0%	32	0	0
1	1	Czech Republic	4	13	18	41.9%	12.9%	39	0	0
2	2	Denmark	4	10	10	50.0%	20.0%	27	1	0
3	3	England	5	11	18	50.0%	17.2%	40	0	0
4	4	France	3	22	24	37.9%	6.5%	65	1	0
5	5	Germany	10	32	32	47.8%	15.6%	80	2	1
6	6	Greece	5	8	18	30.7%	19.2%	32	1	1
7	7	Italy	6	34	45	43.0%	7.5%	110	2	0
8	8	Netherlands	2	12	36	25.0%	4.1%	60	2	0
9	9	Poland	2	15	23	39.4%	5.2%	48	0	0
10	10	Portugal	6	22	42	34.3%	9.3%	82	6	0
11	11	Republic of Ireland	1	7	12	36.8%	5.2%	28	0	0
12	12	Russia	5	9	31	22.5%	12.5%	59	2	0
13	13	Spain	12	42	33	55.9%	16.0%	100	0	1
14	14	Sweden	5	17	19	47.2%	13.8%	39	3	0
15	15	Ukraine	2	7	26	21.2%	6.0%	38	0	0

16 rows × 33 columns

In [13]:
```
# Câu 10: In các cột Team, Goals, Shooting Accuracy, Yellow Cards, Red Cards
euro12[['Team', 'Goals', 'Shooting Accuracy', 'Yellow Cards', 'Red Cards']]
```

Out[13]:

	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 [14]:
```
# Câu 11: In các cột Chỉ hiển thị 'Team','Shooting Accuracy' từ 'England', 'Italy'
euro12.loc[euro12.Team.isin(['England', 'Italy', 'Russia']), ['Team','Shooting Acc
```

Out[14]:

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

In [ ]:
```

```