Forward School

Program Code: J620-002-4:2020

Program Name: FRONT-END SOFTWARE DEVELOPMENT

Title : Exercise 08 Filtering and Sorting Data

Name: Chuay Xiang Ze

IC Number: 021224070255

Date : 27/06/2023

Introduction : Learning how to filter and sort data using both Numpy and Pandas

Conclusion : Used to help visualize data using these functions

Ex08 - Filtering and Sorting Data

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

Step 1. Import the necessary libraries

In [4]:

```
import numpy as np
import pandas as pd
```

Step 2. Import the dataset from this address (https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/02_F

In [6]:

```
df = pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/0
df
```

Out[6]:

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

16 rows × 35 columns

Step 3. Assign it to a variable called euro12.

In [7]:

```
euro12 = df
euro12
```

Out[7]:

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

16 rows × 35 columns

Step 4. Select only the Goal column.

In [8]:

```
euro12['Goals']
```

Out[8]:

```
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 [10]:

```
len(euro12['Team'])
```

Out[10]:

```
16
```

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

In [11]:

```
len(euro12.columns)
```

Out[11]:

```
35
```

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

In [12]:

```
discipline = euro12[['Team', 'Yellow Cards', 'Red Cards']]
discipline
```

Out[12]:

	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 [16]:

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

Out[16]:

	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

Step 9. Calculate the mean Yellow Cards given per Team

In [17]:

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

Out[17]:

```
7.4375
```

Step 10. Filter teams that scored more than 6 goals

In [25]:

```
euro12[euro12['Goals'] > 6][['Team', 'Goals']]
```

Out[25]:

	Team	Goals
5	Germany	10
13	Spain	12

Step 11. Select the teams that start with G

In [27]:

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

Out[27]:

```
5    Germany
6     Greece
Name: Team, dtype: object
```

Step 12. Select the first 7 columns

In [31]:

```
euro12.columns[:7]
```

Out[31]:

```
Index(['Team', 'Goals', 'Shots on target', 'Shots off target',
       'Shooting Accuracy', '% Goals-to-shots', 'Total shots (inc. Blocke
d)'],
      dtype='object')
```

Step 13. Select all columns except the last 3.

In [34]:

```
euro12.columns[0:-3]
```

Out[34]:

```
Index(['Team', 'Goals', 'Shots on target', 'Shots off target',
       'Shooting Accuracy', '% Goals-to-shots', 'Total shots (inc. Blocke
d)',
       'Hit Woodwork', 'Penalty goals', 'Penalties not scored', 'Headed go
als',
       'Passes', 'Passes completed', 'Passing Accuracy', 'Touches', 'Cross
es',
       'Dribbles', 'Corners Taken', 'Tackles', 'Clearances', 'Interception
s',
       'Clearances off line', 'Clean Sheets', 'Blocks', 'Goals conceded',
       'Saves made', 'Saves-to-shots ratio', 'Fouls Won', 'Fouls Concede
d',
       'Offsides', 'Yellow Cards', 'Red Cards'],
      dtype='object')
```

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

In [44]:

```
euro12[(euro12['Team'] == 'England') | (euro12['Team'] == 'Italy') | (euro12['Team'] ==
```

Out[44]:

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

In [ ]:

```

```