

▼ Chapter 5 - Exercise 1: GroupBy

- ▼ drinks.csv là tập tin cung cấp dữ liệu về tình hình tiêu thụ rượu bia ở các quốc gia theo từng châu lục

country: tên quốc gia tiêu thụ

beer_servings: số lượng bia tiêu thụ

spirit_servings: số lượng rượu mạnh ($\geq 40^0$)

wine_servings: số lượng rượu vang

total_litres_of_pure_alcohol: tổng số lít rượu nguyên chất

continent: châu lục

```
import pandas as pd
```

Câu 1: Đọc dữ liệu từ tập tin drinks.csv với index_col là cột đầu tiên của dữ liệu, và 1

```
drinks = pd.read_csv('drinks.csv', index_col=0)
```

Cho biết kiểu dữ liệu (type) của drink

```
print(type(drinks))
```

Cho biết ra kích thước (shape) của drink

```
print(drinks.shape)
```

Hiển thị tên các cột (columns) của drink

```
print(drinks.columns)
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
(193, 6)
```

```
Index(['country', 'beer_servings', 'spirit_servings', 'wine_servings',  
      'total_litres_of_pure_alcohol', 'continent'],  
      dtype='object')
```

Xem 5 dòng dữ liệu đầu tiên (head) của drink

```
drinks.head()
```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_a:
0	Afghanistan	0	0	0	
1	Albania	89	132	54	
2	Algeria	25	0	14	
3	Andorra	245	138	312	
4	Angola	217	57	45	

Xem 5 dòng dữ liệu cuối cùng (tail) của drink

```
drinks.tail()
```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_i
188	Venezuela	333	100	3	
189	Vietnam	111	2	1	
190	Yemen	6	0	0	
191	Zambia	32	19	4	
192	Zimbabwe	64	18	4	

Câu 2: Cho biết số lượng bia tiêu thụ trung bình ở mỗi châu lục
drinks.groupby('continent').beer_servings.mean()

```
continent
AF      61.471698
AS      37.045455
EU     193.777778
OC      89.687500
SA     175.083333
Name: beer_servings, dtype: float64
```

Câu 3: Cho biết thông tin thống kê tổng quát (describe) số lượng rượu vang được tiêu thụ
drinks.groupby('continent').wine_servings.describe()

	count	mean	std	min	25%	50%	75%	max
continent								
AF	53.0	16.264151	38.846419	0.0	1.0	2.0	13.00	233.0
AS	44.0	9.068182	21.667034	0.0	0.0	1.0	8.00	123.0
EU	45.0	142.222222	97.421738	0.0	59.0	128.0	195.00	370.0
OC	16.0	35.625000	64.555790	0.0	1.0	8.5	23.25	212.0
SA	12.0	62.416667	88.620189	1.0	3.0	12.0	98.50	221.0

Câu 4: Cho biết số lượng các loại bia và rượu tiêu thụ trung bình (mean) ở mỗi châu lục
drinks.groupby('continent').mean()

	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcoh
continent				
AF	61.471698	16.339623	16.264151	3.0075
AS	37.045455	60.840909	9.068182	2.1704
EU	193.777778	132.555556	142.222222	8.6177
OC	89.687500	58.437500	35.625000	3.3812
SA	175.083333	114.750000	62.416667	6.3083

Câu 5: Cho biết giá trị trung vị (median) của các loại bia và rượu tiêu thụ ở mỗi châu lục

```
drinks.groupby('continent').median()
```

	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
continent				
AF	32.0	3.0	2.0	2.0
AS	17.5	16.0	1.0	1.0
EU	219.0	122.0	128.0	10.0
OC	52.5	37.0	8.5	1.0
SA	162.5	108.5	12.0	6.0

```
# Câu 6: Cho biết số lượng rượu mạnh (spirit_servings) tiêu thụ trung bình, lớn nhất và nhỏ nhất
# Gợi ý: Dùng spirit_servings.agg(['mean', 'min', 'max'])
drinks.groupby('continent').spirit_servings.agg(['mean', 'min', 'max'])
```

	mean	min	max
continent			
AF	16.339623	0	152
AS	60.840909	0	326
EU	132.555556	0	373
OC	58.437500	0	254
SA	114.750000	25	302

```
# Câu 7: Sắp xếp dữ liệu tăng dần (sort_values) theo số lượng bia tiêu thụ
drinks_des = drinks.sort_values(by='beer_servings', ascending=True)
drinks_des
```

```

country beer_servings spirit_servings wine_servings total_litres_of_pure_
# Cho biết 5 quốc gia có lượng tiêu thụ bia lớn nhất
drinks_des.tail()

```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_a
135	Poland	343	215	56	
65	Germany	346	117	175	
62	Gabon	347	98	59	
45	Czech Republic	361	170	134	
117	Namibia	376	3	1	
62	Gabon	347	98	59	

```

# Cho biết 5 quốc gia có lượng tiêu thụ bia nhỏ nhất
drinks_des.head()

```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_a
0	Afghanistan	0	0	0	
40	Cook Islands	0	254	74	
79	Iran	0	0	0	
90	Kuwait	0	0	0	
97	Libya	0	0	0	