



Chapter 5 - exercise 1: GroupBy

Cung cấp tập tin 'drinks.csv'

Hãy thực hiện những yêu cầu sau:

1. Tạo data frame drink có dữ liệu được lấy ra từ tập tin drinks.csv với index_col là cột đầu tiên của dữ liệu. In type, shape, column, head, tail
2. Thống kê trung bình beer_servings nhóm theo 'continent'
3. Thống kê tổng quát wine_servings nhóm theo 'continent'
4. Thống kê trung bình cho tất cả các cột nhóm theo 'continent'
5. Thống kê median cho tất cả các cột theo 'continent'
6. Thống kê trung bình, min, max của spirit_servings nhóm theo 'continent' dùng spirit_servings.agg(['mean', 'min', 'max'])
7. Sắp xếp dữ liệu tăng dần theo cột beer_servings. In 5 country có beer_servings lớn nhất, nhỏ nhất

In [43]: `import pandas as pd`

In [44]: `# Câu 1: Tạo data frame drink có dữ liệu được lấy ra từ tập tin drinks.csv với index_col=0
In type, shape, head, tail
drinks = pd.read_csv('drinks.csv', index_col=0)
print(type(drinks))
print(drinks.shape)
print(drinks.columns)
drinks.head()`

```
<class 'pandas.core.frame.DataFrame'>  
(193, 6)  
Index(['country', 'beer_servings', 'spirit_servings', 'wine_servings',  
      'total_litres_of_pure_alcohol', 'continent'],  
      dtype='object')
```

Out[44]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	AS
1	Albania	89	132	54	4.9	EU
2	Algeria	25	0	14	0.7	AF
3	Andorra	245	138	312	12.4	EU
4	Angola	217	57	45	5.9	AF

In [45]: `drinks.tail()`

Out[45]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
188	Venezuela	333	100	3	7.7	SA
189	Vietnam	111	2	1	2.0	AS
190	Yemen	6	0	0	0.1	AS
191	Zambia	32	19	4	2.5	AF
192	Zimbabwe	64	18	4	4.7	AF

In [46]: *# Câu 2: Thống kê trung bình beer_servings nhóm theo 'continent'*
`drinks.groupby('continent').beer_servings.mean()`

Out[46]: continent
 AF 61.471698
 AS 37.045455
 EU 193.777778
 OC 89.687500
 SA 175.083333
 Name: beer_servings, dtype: float64

In [47]: *# Câu 3: Thống kê tổng quát wine_servings nhóm theo 'continent'*
`drinks.groupby('continent').wine_servings.describe()`

Out[47]:

	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

In [48]: *# Câu 4: Thống kê trung bình cho tất cả các cột theo 'continent'*
`drinks.groupby('continent').mean()`

Out[48]:

	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
continent				
AF	61.471698	16.339623	16.264151	3.007547
AS	37.045455	60.840909	9.068182	2.170455
EU	193.777778	132.555556	142.222222	8.617778
OC	89.687500	58.437500	35.625000	3.381250
SA	175.083333	114.750000	62.416667	6.308333



In [49]: *# Câu 5: Thống kê median cho tất cả các cột theo 'continent'*
drinks.groupby('continent').median()

Out[49]:

	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
continent				
AF	32.0	3.0	2.0	2.30
AS	17.5	16.0	1.0	1.20
EU	219.0	122.0	128.0	10.00
OC	52.5	37.0	8.5	1.75
SA	162.5	108.5	12.0	6.85

In [50]: *# Câu 6: Thống kê trung bình, min, max của spirit_servings nhóm theo 'continent' d*
drinks.groupby('continent').spirit_servings.agg(['mean', 'min', 'max'])

Out[50]:

	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

In [51]: *# Câu 7: Sắp xếp dữ liệu tăng dần theo cột beer_servings. In 5 country có beer_ser*
drinks_des = drinks.sort_values(by='beer_servings', ascending=False)

In [52]: drinks_des.head()

Out[52]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
117	Namibia	376	3	1	6.8	AF
45	Czech Republic	361	170	134	11.8	EU
62	Gabon	347	98	59	8.9	AF
65	Germany	346	117	175	11.3	EU
98	Lithuania	343	244	56	12.9	EU



In [53]: `drinks_des.tail()`

Out[53]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	container
107	Mauritania	0	0	0	0.0	A
158	Somalia	0	0	0	0.0	A
111	Monaco	0	0	0	0.0	E
128	Pakistan	0	0	0	0.0	A
0	Afghanistan	0	0	0	0.0	A



In []:

