

# 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

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

```
In [2]: # Câu 1: Đọc dữ liệu từ tập tin drinks.csv với index_col là cột đầu tiên của dữ liệu,
# và lưu vào biến drinks.
# Cho biết kiểu dữ liệu (type) của drinks
# Cho biết kích thước (shape) của drinks
# Hiển thị tên các cột (columns) của drinks
# Xem thông tin info()
# code here
```

Nhấn vào đây để xem kết quả!

```
<class 'pandas.core.frame.DataFrame'>
(193, 6)
Index(['country', 'beer_servings', 'spirit_servings', 'wine_servings',
      'total_litres_of_pure_alcohol', 'continent'],
      dtype='object')
<class 'pandas.core.frame.DataFrame'>
Int64Index: 193 entries, 0 to 192
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   country                               193 non-null    object
1   beer_servings                         193 non-null    int64
2   spirit_servings                       193 non-null    int64
3   wine_servings                        193 non-null    int64
4   total_litres_of_pure_alcohol         193 non-null    float64
5   continent                             170 non-null    object
dtypes: float64(1), int64(3), object(2)
memory usage: 10.6+ KB
None
```

```
In [3]: # Xem 5 dòng dữ liệu đầu tiên (head) của drinks
# code here
```

```
In [4]: # Xem 5 dòng dữ liệu cuối cùng (tail) của drinks
# code here
```

```
In [5]: # Đếm dữ liệu NULL
# code here
```

Nhấn vào đây để xem kết quả!

```
country          0
beer_servings     0
spirit_servings   0
wine_servings     0
total_litres_of_pure_alcohol  0
continent        23
dtype: int64
```

```
In [6]: # cho biết số lượng quốc gia ở mỗi châu Lục
# code here
```

Nhấn vào đây để xem kết quả!

```
AF    53
EU    45
AS    44
OC    16
SA    12
Name: continent, dtype: int64
```

```
In [7]: # Câu 2: Cho biết số lượng bia tiêu thụ trung bình (mean) ở mỗi châu Lục
# code here
```

Nhấn vào đây để xem kết quả!

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

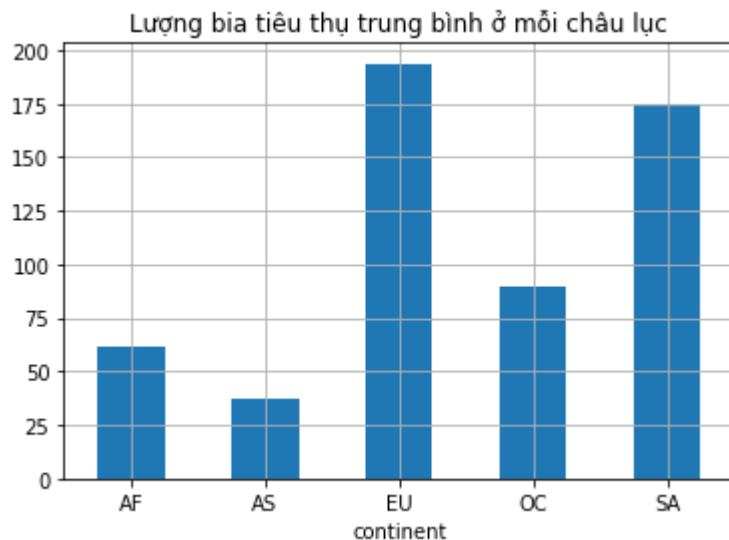
```
In [8]: # sắp giảm dần
# code here
```

Nhấn vào đây để xem kết quả!

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

```
In [9]: # Câu 2: Vẽ biểu đồ Bar với số Lượng bia tiêu thụ trung bình ở mỗi châu Lục
# HD:
# drinks.groupby('continent').beer_servings.mean().plot.bar(
#     title='Lượng bia tiêu thụ trung bình ở mỗi châu Lục', rot=0, grid=True);
```

Nhấn vào đây để xem kết quả!



```
In [10]: # Câu 3: Cho biết thông tin thống kê tổng quát (describe) số Lượng rượu vang
# được tiêu thụ ở mỗi châu Lục
# code here
```

Nhấn vào đây để xem kết quả!

	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 [11]: # Câu 4: Cho biết số Lượng bia và rượu tiêu thụ trung bình (mean) ở mỗi châu Lục
# code here
```

Nhấn vào đây để xem kết quả!

```
beer_servings  spirit_servings  wine_servings  total_litres_of_pure_alcohol
```

continent	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 [12]: *# Câu 5: Cho biết giá trị trung vị (median) của các loại bia và rượu  
# tiêu thụ ở mỗi châu lục  
# code here*

In [13]: *# Câu 6: Cho biết số Lượng rượu mạnh (spirit\_servings)  
# tiêu thụ trung bình, Lớn nhất và nhỏ nhất ở mỗi châu lục  
# Gợi ý: Dùng spirit\_servings.agg(['mean', 'min', 'max'])  
# code here*

Nhấn vào đây để xem kết quả!

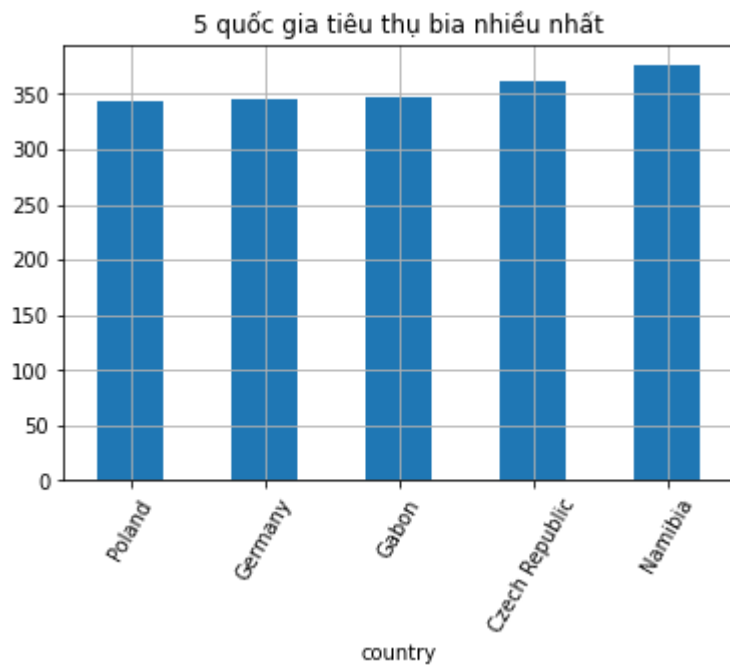
	mean	min	max
continent			
AF	19.681818	1	152
AS	78.735294	1	326
EU	138.720930	27	373
OC	66.785714	11	254
SA	114.750000	25	302

In [14]: *# Câu 7: Sắp xếp dữ liệu tăng dần (sort\_values) theo số Lượng bia tiêu thụ,  
# In 10 dòng đầu tiên  
# code here*

In [15]: *# Cho biết 5 quốc gia có Lượng tiêu thụ bia Lớn nhất  
# code here*

In [16]: *# Vẽ biểu đồ Bar với 5 quốc gia có Lượng tiêu thụ bia Lớn nhất  
# HD:  
# drinks\_des.tail().plot.bar(x='country', y='beer\_servings',  
# rot=60, legend=False, title='5 quốc gia tiêu thụ bia nhiều nhất', grid=True);*

Nhấn vào đây để xem kết quả!



In [17]: *# Cho biết 5 quốc gia có lượng tiêu thụ bia ít nhất*  
*# code here*

In [18]: *# Câu 8: Cho biết 5 quốc gia ở Châu Á có lượng tiêu thụ bia ít nhất*  
*# code here*

Nhấn vào đây để xem kết quả!

	country	beer_servings
171	Timor-Leste	1
137	Qatar	1
168	Tajikistan	2
116	Myanmar	5
167	Syria	5

In [ ]: