

酒类消费数据 - 数据分组



加载数据

```
1 import pandas as pd
2 drinks = pd.read_csv('data/drinks.csv')
3 drinks.head()
```

```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>country</th>
4   <th>beer_servings</th>
5   <th>spirit_servings</th>
6   <th>wine_servings</th>
7   <th>total_litres_of_pure_alcohol</th>
8   <th>continent</th>
9 </tr>
```

```
1 <tr>
2   <th>0</th>
3   <td>Afghanistan</td>
4   <td>0</td>
5   <td>0</td>
6   <td>0</td>
7   <td>0.0</td>
8   <td>AS</td>
9 </tr>
10 <tr>
11   <th>1</th>
12   <td>Albania</td>
13   <td>89</td>
14   <td>132</td>
15   <td>54</td>
16   <td>4.9</td>
17   <td>EU</td>
18 </tr>
19 <tr>
20   <th>2</th>
21   <td>Algeria</td>
22   <td>25</td>
23   <td>0</td>
24   <td>14</td>
25   <td>0.7</td>
26   <td>AF</td>
27 </tr>
28 <tr>
29   <th>3</th>
30   <td>Andorra</td>
31   <td>245</td>
32   <td>138</td>
33   <td>312</td>
34   <td>12.4</td>
```

```
35     <td>EU</td>
36 </tr>
37 <tr>
38     <th>4</th>
39     <td>Angola</td>
40     <td>217</td>
41     <td>57</td>
42     <td>45</td>
43     <td>5.9</td>
44     <td>AF</td>
45 </tr>
```

```
1 drinks.info()
```

```
1 <class 'pandas.core.frame.DataFrame'>
2 RangeIndex: 193 entries, 0 to 192
3 Data columns (total 6 columns):
4 country                                193 non-null object
5 beer_servings                          193 non-null int64
6 spirit_servings                         193 non-null int64
7 wine_servings                          193 non-null int64
8 total_litres_of_pure_alcohol           193 non-null float64
9 continent                              170 non-null object
10 dtypes: float64(1), int64(3), object(2)
11 memory usage: 9.1+ KB
```

查询：哪个大陆(continent)平均消耗的啤酒(beer)更多？

```
1 drinks.groupby('continent').beer_servings.mean().sort_values  
  (ascending=False).rename('Beer servings mean')
```

```
1 continent  
2 EU      193.777778  
3 SA      175.083333  
4 OC       89.687500  
5 AF       61.471698  
6 AS       37.045455  
7 Name: Beer servings mean, dtype: float64
```

输出每个大陆(continent)的红酒消耗 (wine_servings)的描述性统计值

```
1 drinks.groupby('continent').wine_servings.describe()
```

```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>count</th>
4   <th>mean</th>
5   <th>std</th>
6   <th>min</th>
7   <th>25%</th>
8   <th>50%</th>
9   <th>75%</th>
10  <th>max</th>
11 </tr>
12 <tr>
13   <th>continent</th>
14   <th></th>
15   <th></th>
16   <th></th>
17   <th></th>
18   <th></th>
19   <th></th>
20   <th></th>
21   <th></th>
22 </tr>
```

```
1 <tr>
2   <th>AF</th>
3   <td>53.0</td>
4   <td>16.264151</td>
5   <td>38.846419</td>
6   <td>0.0</td>
7   <td>1.0</td>
8   <td>2.0</td>
9   <td>13.00</td>
10  <td>233.0</td>
11 </tr>
12 <tr>
13   <th>AS</th>
14   <td>44.0</td>
15   <td>9.068182</td>
16   <td>21.667034</td>
17   <td>0.0</td>
18   <td>0.0</td>
19   <td>1.0</td>
20   <td>8.00</td>
21   <td>123.0</td>
22 </tr>
23 <tr>
24   <th>EU</th>
25   <td>45.0</td>
26   <td>142.222222</td>
27   <td>97.421738</td>
28   <td>0.0</td>
29   <td>59.0</td>
30   <td>128.0</td>
31   <td>195.00</td>
32   <td>370.0</td>
33 </tr>
34 <tr>
```

```
35 <th>OC</th>
36 <td>16.0</td>
37 <td>35.625000</td>
38 <td>64.555790</td>
39 <td>0.0</td>
40 <td>1.0</td>
41 <td>8.5</td>
42 <td>23.25</td>
43 <td>212.0</td>
44 </tr>
45 <tr>
46 <th>SA</th>
47 <td>12.0</td>
48 <td>62.416667</td>
49 <td>88.620189</td>
50 <td>1.0</td>
51 <td>3.0</td>
52 <td>12.0</td>
53 <td>98.50</td>
54 <td>221.0</td>
55 </tr>
```

打印出每个大陆每种酒类别的消耗平均值

```
1 drinks.groupby('continent').mean()
```



```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>beer_servings</th>
4   <th>spirit_servings</th>
5   <th>wine_servings</th>
6   <th>total_litres_of_pure_alcohol</th>
7 </tr>
8 <tr>
9   <th>continent</th>
10  <th></th>
11  <th></th>
12  <th></th>
13  <th></th>
14 </tr>
```

```
1 <tr>
2   <th>AF</th>
3   <td>61.471698</td>
4   <td>16.339623</td>
5   <td>16.264151</td>
6   <td>3.007547</td>
7 </tr>
8 <tr>
9   <th>AS</th>
10  <td>37.045455</td>
11  <td>60.840909</td>
12  <td>9.068182</td>
13  <td>2.170455</td>
14 </tr>
15 <tr>
16  <th>EU</th>
17  <td>193.777778</td>
18  <td>132.555556</td>
19  <td>142.222222</td>
20  <td>8.617778</td>
21 </tr>
22 <tr>
23  <th>OC</th>
24  <td>89.687500</td>
25  <td>58.437500</td>
26  <td>35.625000</td>
27  <td>3.381250</td>
28 </tr>
29 <tr>
30  <th>SA</th>
31  <td>175.083333</td>
32  <td>114.750000</td>
33  <td>62.416667</td>
34  <td>6.308333</td>
```

打印出每个大陆每种酒类别的消耗中位数

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

```
1 <tr style="text-align: right;">
2   <th></th>
3   <th>beer_servings</th>
4   <th>spirit_servings</th>
5   <th>wine_servings</th>
6   <th>total_litres_of_pure_alcohol</th>
7 </tr>
8 <tr>
9   <th>continent</th>
10  <th></th>
11  <th></th>
12  <th></th>
13  <th></th>
14 </tr>
```

```
1 <tr>
2   <th>AF</th>
3   <td>32.0</td>
4   <td>3.0</td>
5   <td>2.0</td>
6   <td>2.30</td>
7 </tr>
8 <tr>
9   <th>AS</th>
10  <td>17.5</td>
11  <td>16.0</td>
12  <td>1.0</td>
13  <td>1.20</td>
14 </tr>
15 <tr>
16  <th>EU</th>
17  <td>219.0</td>
18  <td>122.0</td>
19  <td>128.0</td>
20  <td>10.00</td>
21 </tr>
22 <tr>
23  <th>OC</th>
24  <td>52.5</td>
25  <td>37.0</td>
26  <td>8.5</td>
27  <td>1.75</td>
28 </tr>
29 <tr>
30  <th>SA</th>
31  <td>162.5</td>
32  <td>108.5</td>
33  <td>12.0</td>
34  <td>6.85</td>
```

打印出每个大陆对spirit饮品消耗的平均值，最大值和最小值

```
1 drinks.groupby('continent').spirit_servings.agg(['mean',  
    'max', 'min'])
```

```
1 <tr style="text-align: right;">  
2   <th></th>  
3   <th>mean</th>  
4   <th>max</th>  
5   <th>min</th>  
6 </tr>  
7 <tr>  
8   <th>continent</th>  
9   <th></th>  
10  <th></th>  
11  <th></th>  
12 </tr>
```

```
1 <tr>
2   <th>AF</th>
3   <td>16.339623</td>
4   <td>152</td>
5   <td>0</td>
6 </tr>
7 <tr>
8   <th>AS</th>
9   <td>60.840909</td>
10  <td>326</td>
11  <td>0</td>
12 </tr>
13 <tr>
14   <th>EU</th>
15   <td>132.555556</td>
16   <td>373</td>
17   <td>0</td>
18 </tr>
19 <tr>
20   <th>OC</th>
21   <td>58.437500</td>
22   <td>254</td>
23   <td>0</td>
24 </tr>
25 <tr>
26   <th>SA</th>
27   <td>114.750000</td>
28   <td>302</td>
29   <td>25</td>
30 </tr>
```

查询：中国的啤酒消费在亚洲和全球的排名

```
1 china_beer = drinks[drinks.country ==  
  'China'].beer_servings.iloc[0]  
2 print ('中国啤酒消费量: ', china_beer)  
3 print ('中国啤酒消费在亚洲排名: ', (drinks[drinks.continent ==  
  'AS'].beer_servings > china_beer).sum()+1, '/',  
  drinks[drinks.continent == 'AS'].beer_servings.count())  
4 print ('中国啤酒消费在全球排名: ', (drinks['beer_servings'] >  
  china_beer).sum()+1, '/', drinks['beer_servings'].count())
```

```
1 中国啤酒消费量: 79  
2 中国啤酒消费在亚洲排名: 6 / 44  
3 中国啤酒消费在全球排名: 91 / 193
```