

* 작성한 코드

https://github.com/dayo2n/2021-winterMGC/blob/main/활동/10회차/WISET_Python_Pandas_HW.ipynb

1. drinks.csv 데이터를 이용하여 다음을 분석하시오.

* 각 문제별로 소스 및 나온 값 혹은 그래프, 검정결과 등은 캡처를 해서 제출해 주시면 됩니다.

** 분석에 앞서 anaconda prompt 창에서 아래 패키지를 설치해야 합니다.

---> pip install seaborn

---> pip install scipy

(1) beer_servings', 'wine_servings' 두 변수간의 피어슨 상관계수를 계산

```
# (1) beer_servings, wine_servings 두 변수간 피어슨 상관계수를 계산
corr = drinks[['beer_servings', 'wine_servings']].corr(method = 'pearson')
print(corr)
```

	beer_servings	wine_servings
beer_servings	1.000000	0.527172
wine_servings	0.527172	1.000000

(2) corr 행렬 히트맵을 통한 시각화하기

```
# (2) corr 행렬 히트맵을 통한 시각화하기
import seaborn as sns
```

```
cols_view = ['beer', 'spirit', 'wine', 'alcohol']
sns.set(font_scale=1.5)
```

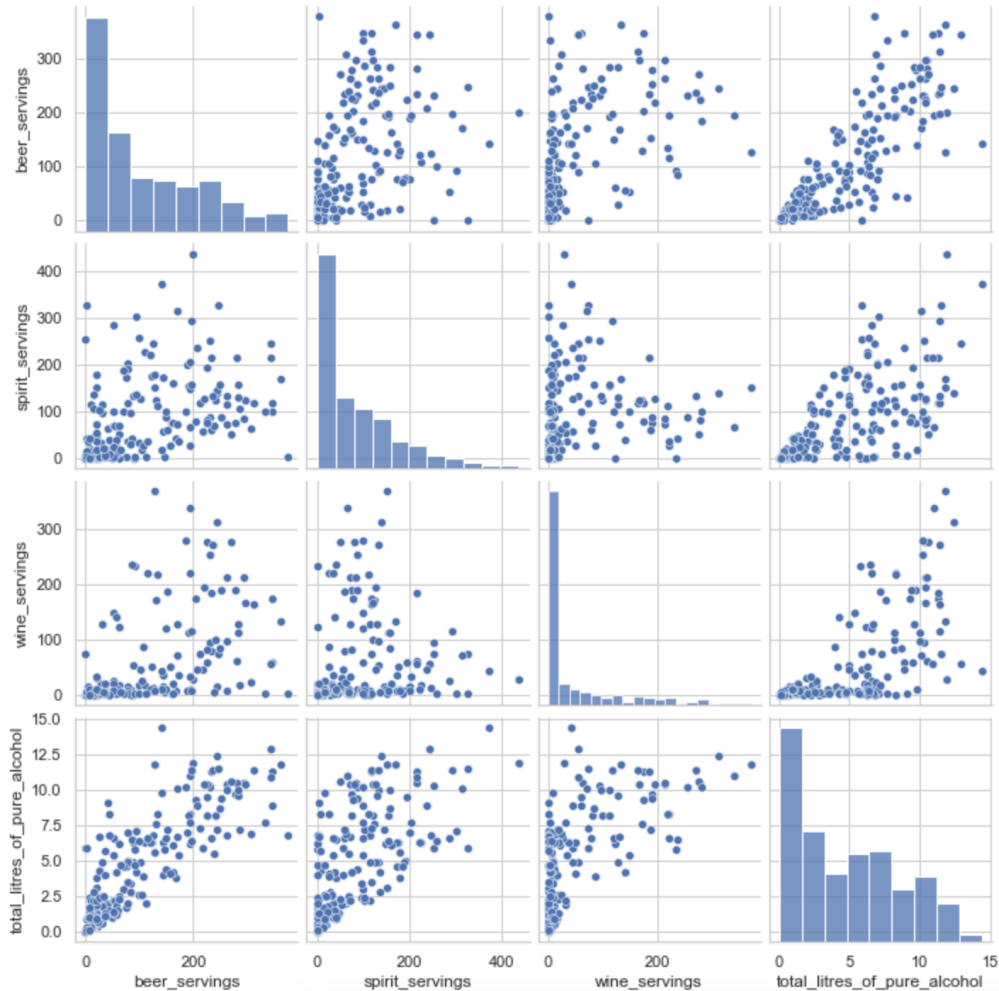
```
hm = sns.heatmap(corr.values, cbar=True, annot=True, fmt='.2f', square=True,
                  annot_kws={'size': 15}, yticklabels=cols_view, xticklabels=cols_view)

plt.tight_layout()
plt.show()
```



(3) scatter plot을 출력하기

```
# (3) scatter plot 출력하기
sns.set(style = 'whitegrid', context = 'notebook')
sns.pairplot(drinks[['beer_servings', 'spirit_servings',
                    'wine_servings', 'total_litres_of_pure_alcohol']], height=2.5)
plt.show()
```



(4) 대륙별로 spirit_servings의 평균, 최소, 최대, 총합을 계산하기

```
# (4) 대륙별로 spirit_servings의 평균, 최소, 최대, 총합을 계산하기
result = drinks.groupby('continent').spirit_servings.agg(['mean', 'min', 'max', 'sum'])
result.head(10)
```

	mean	min	max	sum
continent				
AF	16.339623	0	152	866
AS	60.840909	0	326	2677
EU	132.555556	0	373	5965
OC	58.437500	0	254	935
OT	165.739130	68	438	3812
SA	114.750000	25	302	1377

(5) 아프리카 대륙과 유럽 대륙간의 맥주 소비량 차이를 검정

```
# (5) 아프리카 대륙과 유럽 대륙간의 맥주 소비량 차이를 검정
beer_continent = drinks.groupby('continent').beer_servings.mean()
beer_continent
```

```
continent
AF    61.471698
AS    37.045455
EU   193.777778
OC    89.687500
OT   145.434783
SA   175.083333
Name: beer_servings, dtype: float64
```

대륙 별로 그룹핑하여 맥주 소비량에 대해 평균을 도출했을 때, 문제에서 요구되는 아프리카 대륙의 맥주 소비량 평균은 약 61, 유럽 대륙의 맥주 소비량 평균은 약 194이다. 평균으로만 차이를 비교해보면 유럽 대륙의 맥주 소비량은 아프리카의 3배인 것을 알 수 있다.

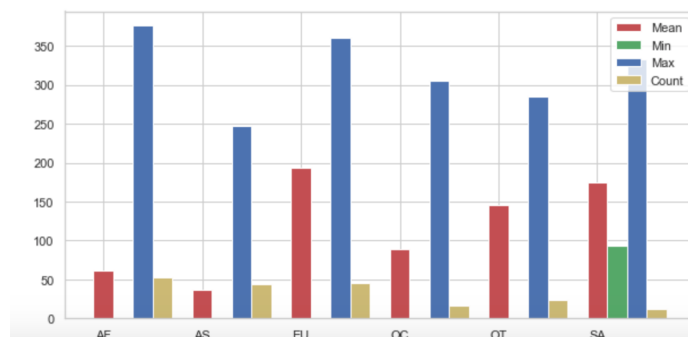
아래에서는 각 대륙의 맥주 소비량에 대해 평균, 최소, 최대, 국가의 수를 시각화한다.

```
result = drinks.groupby('continent').beer_servings.agg(['mean', 'min', 'max', 'count'])
number_groups = len(result.index)
means = result['mean'].tolist()
mins = result['min'].tolist()
maxs = result['max'].tolist()
counts = result['count'].tolist()
index = np.arange(number_groups)

plt.figure(figsize = (10,5))
bar_width = 0.2

rects_1 = plt.bar(index, means, bar_width, color = 'r', label = 'Mean')
rects_2 = plt.bar(index+bar_width, mins, bar_width, color = 'g', label = 'Min')
rects_3 = plt.bar(index+bar_width * 2, maxs, bar_width, color='b', label = 'Max')
rects_4 = plt.bar(index+bar_width * 3, counts, bar_width, color = 'y', label = 'Count')

plt.xticks(index, result.index.tolist())
plt.legend()
plt.show()
```



시각화를 통해 나타난 그래프를 보면, 유럽 대륙의 맥주 소비량 평균이 아프리카 대륙의 것보다 훨씬 높은 것을 알 수 있지만 아프리카 대륙 내의 맥주 최대 소비량은 유럽의 것보다 높다.

```
beer_continent2 = drinks.groupby('continent').beer_servings.count()
beer_continent2
```

```
continent
AF    53
AS    44
EU    45
OC    16
OT    23
SA    12
Name: beer_servings, dtype: int64
```

아프리카 대륙 내의 국가 수는 53, 유럽은 45이므로 평균에 영향을 크게 미친 것은 각국의 맥주 소비량의 총 합이다. 앞에서 살펴본 결과로 보았을 때, 아프리카는 모든 대륙과 비교했을 때 맥주 최대 소비량이 가장 높은 대륙이지만 대륙 내의 모든 국가의 맥주 소비량이 특히 유럽에 비해 전반적으로 높지는 않은 것을 알 수 있다.

(6) 대륙별 술 소비량 대비 알콜 비율 구하기

```
# (6) 대륙별 술 소비량 대비 알콜 비율 구하기
continent_mean = drinks.groupby('continent')['total_litres_of_pure_alcohol'].mean()
continent_mean
```

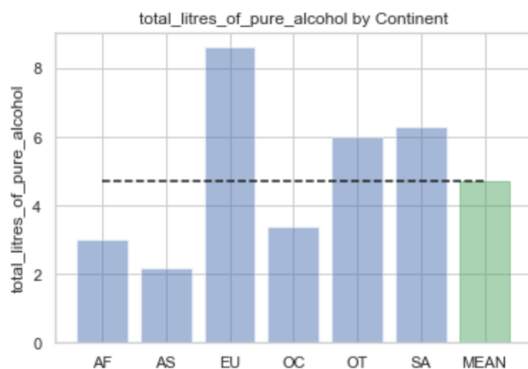
```
continent
AF      3.007547
AS      2.170455
EU      8.617778
OC      3.381250
OT      5.995652
SA      6.308333
Name: total_litres_of_pure_alcohol, dtype: float64
```

다음은 대륙별 술 소비량 대비 알코올 비율에 대한 평균이다. 이를 시각화하면 아래와 같다.

```
bar_list = plt.bar(x_pos, alcohol, align='center', alpha=0.5)
bar_list[len(continent)-1].set_color('g')
plt.plot([0., 6], [total_mean, total_mean], "k--")
plt.xticks(x_pos, continent)

plt.ylabel('total_litres_of_pure_alcohol')
plt.title('total_litres_of_pure_alcohol by Continent')

plt.show()
```



그 평균 비율이 가장 높은 대륙은 유럽이고, 그 다음으로는 남아메리카로 이 두 대륙은 평균보다 높은 평균 비율을 차지한다. 이 두 대륙 외에도 OT(other : SA ...) 부분도 그러하다. 그 다음으로는 오세아니아, 아프리카, 아시아 순으로 평균 이하의 평균 비율을 보인다.

(7) 평균보다 알코올을 적게 마시는 대륙 중에서, spirit을 가장 많이 마시는 국가를 구하기

```
total_mean = drinks.total_litres_of_pure_alcohol.mean()
total_mean
```

```
4.71098445595855
```

```
continent_sum['mean'] = drinks.groupby('continent')['total_litres_of_pure_alcohol'].mean()
continent_sum
```

	sum	mean
continent		
AF	866	3.007547
AS	2677	2.170455
EU	5965	8.617778
OC	935	3.381250
OT	3812	5.995652
SA	1377	6.308333

```
continent_sum[continent_mean < total_mean]
```

	sum	mean
continent		
AF	866	3.007547
AS	2677	2.170455
OC	935	3.381250

(6)의 데이터만으로 분석한 내용을 파이썬을 이용해 쉽게 분석하고 도출할 수 있다. total_litres_of_pure_alcohol 평균을 먼저 구하고, 각 대륙의 해당 데이터의 합계와 평균 도출한다. 그 값 중 대륙간 평균보다 낮은 평균비율을 차지하는 대륙만 뽑아낸다. 이제 대륙이 AF, AS 또는 OC인 국가를 가려내본다.

```
a = drinks['continent'].str.contains('AS|AF|OC')
print(a)
```

```
0      True
1     False
2      True
3     False
4      True
...
188    False
189     True
190     True
191     True
192     True
Name: continent, Length: 193, dtype: bool
```

```
b = drinks[a]
print(b)
```

	country	beer_servings	spirit_servings	wine_servings	\
0	Afghanistan	0	0	0	
2	Algeria	25	0	14	
4	Angola	217	57	45	
8	Australia	261	72	212	
12	Bahrain	42	63	7	
..	
187	Vanuatu	21	18	11	
189	Vietnam	111	2	1	
190	Yemen	6	0	0	
191	Zambia	32	19	4	
192	Zimbabwe	64	18	4	

	total_litres_of_pure_alcohol	continent	total_servings	alcohol_rate
0	0.0	AS	0	0.000000
2	0.7	AF	39	0.017949
4	5.9	AF	319	0.018495
8	10.4	OC	545	0.019083
12	2.0	AS	112	0.017857
..
187	0.9	OC	50	0.018000
189	2.0	AS	114	0.017544
190	0.1	AS	6	0.016667
191	2.5	AF	55	0.045455
192	4.7	AF	86	0.054651

```
[113 rows x 8 columns]
```

해당하는 국가 리스트를 주욱 뽑아낸 후 이 중 spirit_servings의 최댓값을 가리키는 국가를 선별해보면 아래와 같이 Russian Federation인 것을 알 수 있다.

```
spirit_continent_max = b[b['spirit_servings'] == b['spirit_servings'].max()]
spirit_continent_max
```

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent	total_servings	alcohol_rate
141	Russian Federation	247	326	73	11.5	AS	646	0.017802