

Python For Data Analysis

Review Test

Chapter_10-Part.2

이석희 장은조

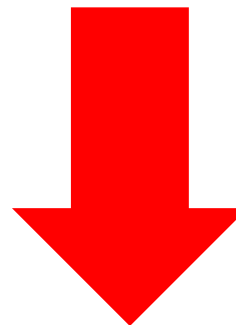
보너스 문제) . 다음 데이터프레임을 생성하세요

In [12]:

```
df
```

Out [12]:

	key1	key2	data1	data2
0	a	one	0.826763	0.736411
1	a	two	-1.563122	0.793367
2	a	one	-0.375564	-2.607418
3	b	two	0.036945	1.126356
4	b	two	-2.743886	0.196364
5	b	one	0.440431	0.244225
6	aa	one	0.337922	1.962843
7	aa	two	0.168380	0.471902
8	aa	two	0.179708	1.070554



```
df = pd.DataFrame({'key1': ['a','a','a','b','b','b','aa','aa','aa'],  
                  'key2': ['one','two','one','two','two','one','one','two','two'],  
                  'data1': np.random.randn(9),  
                  'data2': np.random.randn(9)})
```

1-1) . df 의 data2 값을 key1와 key2 두개의 색인으로 묶고 다음과 같은 계층적 색인을 가지는 Series 를 출력하세요

Out [14] :

key1	a	aa	b
key2			
one	-0.935503	1.962843	0.244225
two	0.793367	0.771228	0.661360

1-2) . df 데이터프레임에 data3~data5 변수를 추가하고 각각의 값은 표준정규분포에서 난수 9개씩 추출하세요

In [21]:

df

Out [21]:

	key1	key2	data1	data2	data3	data4	data5
0	a	one	0.826763	0.736411	-0.394829	0.645587	1.883280
1	a	two	-1.563122	0.793367	-0.987467	-1.367048	-0.366103
2	a	one	-0.375564	-2.607418	-0.320879	0.740404	0.423720
3	b	two	0.036945	1.126356	-0.397231	0.525121	0.195863
4	b	two	-2.743886	0.196364	-0.167586	1.799061	0.315768
5	b	one	0.440431	0.244225	0.248910	0.354121	0.528650
6	aa	one	0.337922	1.962843	-0.806798	-0.804123	0.467688
7	aa	two	0.168380	0.471902	0.919700	1.039658	1.171889
8	aa	two	0.179708	1.070554	0.161665	0.639892	-0.465629

1-3) . df의 data1~data5 각 컬럼을 data1, data2, data4를 Yes / data3, data5를 No로 그룹목록을 사전으로 만든 후, 그룹별로 컬럼의 값을 모두 더해보자.
(hint : axis = 1)

Out [28] :

	No	Yes
0	1.488451	2.208760
1	-1.353570	-2.136802
2	0.102841	-2.242578
3	-0.201368	1.688422
4	0.148182	-0.748461
5	0.777560	1.038777
6	-0.339111	1.496642
7	2.091588	1.679939
8	-0.303963	1.890154

1. AB_NYC_2019.csv 를 읽어 와서,
오른쪽에서 8번째까지의 열만 남긴 후,
'reviews_per_month', 'minimum_nights', 'price' 중
하나의 열이라도 결측치가 있으면 그 행을 제거하시오.
(hint: dropna)

2. 1번의 데이터 프레임에

'reviews_per_month', 'minimum_nights', 'price' 을 모두 곱한 값인
'minimum_monthly_sales' 열을 추가하시오.

3-1. 'minimum_monthly_sales' 열을 기준으로 상위 10개를 남기는 함수를 만드시오.

3-2. 'room_type' 그룹별로 'minimum_monthly_sales' 상위 10개를 남기시오. (hint: groupby 활용)

4. room_type 별로 minimum_monthly_sales 피벗 테이블을 만드시오.

minimum_monthly_sales	
room_type	
Entire home/apt	991.299183
Private room	321.137804
Shared room	145.838676