

## ▼ HW02 (제출기한-4월5일 자정)

Pandas 이해도 평가를 위한 문제입니다.

- 반드시 수업시간 배운 내용만을 사용하여 코드를 작성합니다.
- 한 셀에 여러줄의 코드를 사용해도 무방합니다.

제출시 다음 사항에 유의하기 바랍니다.

- 텍스트 셀에 설명되어 있는 각 문항을 잘 읽은 뒤, '답안을 작성하시오'라고 적힌 코드 셀에 적절한 코드를 작성합니다.
- '지우지 마시오'라고 적힌 코드 셀은 절대로 지우면 안 됩니다.
- 작성 후 파일명에 학번 을 자신의 학번으로 고친 후 ULMS 해당과제 제출란에 업로드합니다.

## ▼ 가정

다음은 100명 학생에 관한 정보를 담고 있는 변수들이다.

- sname : 학생 이름
- noclass : 반 번호
- gender : 성별
- height : 키
- weight : 몸무게

아래 코드는 수정없이 사용하면 된다.

```
1 ### 제출 시 수정하지 마시오.
2 import numpy as np
3 import pandas as pd
4 np.random.seed(0)
5
6 sname=['s'+format(i, '03d') for i in range(100)]
7 noclass=np.random.randint(1, 6, 100)
8 gender=['female']*50 + ['male']*50
9 height=np.random.randint(160, 190, 100)
10 weight=np.random.randint(50, 80, 100)
```

## ▼ 문제1

sname, noclass, gender, height, weight를 column으로 사용하는 DataFrame 변수 student 를 생성하시오. (index는 자동으로 생성되는 정수)

```
1 ### 답안을 작성하시오.
2 student=pd.DataFrame({'sname':sname , 'noclass': noclass,'gender':gender, 'height':height, 'weig
3
```

```
1 ### 지우지 마시오.  
2 student.head() ### 첫 5행을 출력한 결과이다.
```

## ▼ 문제2

1반부터 5반까지 담임교사의 성이 각각 Kim, Lee, Park, Choi, Yun이라고 할 때 반번호(1반~5반)와 담임교사의 이름(tname)을 column으로 가지는 DataFrame 변수 teacher 를 생성하시오. (index는 자동으로 생성되는 정수값이다)

```
1 ### 답안을 작성하시오.  
2 tname = ['kim', 'Lee', 'Park', 'Choi', 'Yun']  
3 noclase = [1,2,3,4,5]  
4 teacher = pd.DataFrame({'noclass': noclase, 'tname': tname})  
5
```

```
1 ### 지우지 마시오.  
2 teacher
```

## ▼ 문제3

두 DataFrame 변수 student 와 teacher 를 합쳐서 info 라는 이름의 DataFrame 변수를 생성하시오.

- noclase column이 key값이 된다.

```
1 ### 답안을 작성하시오.  
2  
3 info=pd.merge(student, teacher)
```

```
1 ### 지우지 마시오.  
2 info.tail() ### 마지막 5행을 보여준다.
```

#### ▼ 문제4

$ratio = height - weight$  일 때 새로운 column인 `ratio`를 `info`에 추가하시오.

```
1 ### 답안을 작성하시오.  
2  
3 info['ratio']=info['height']-info['weight']  
4
```

```
1 ### 지우지 마시오.  
2 info.head()
```

#### ▼ 문제5

`info`에서 `sname`변수가 `index`로 되도록 변경한 `info_sname` 변수를 생성하시오.

```
1 ### 답안을 작성하시오.  
2  
3 info_sname=info.set_index('sname')  
4
```

```
1 ### 지우지 마시오.  
2 info_sname.head()
```

## ▼ 문제6

반별로 height, weight, ratio의 평균값을 구하여 아래 결과와 같이 dataframe 변수 info\_avg에 저장하시오.

- info와 info\_sname 중 어떤 변수를 사용하여도 좋다.

```
1 ### 답안을 작성하시오.  
2 info_avg=info.groupby('noclass').mean()  
3  
4
```

```
1 ### 지우지 마시오.  
2 info_avg
```

## ▼ 문제7

pivot table을 사용하여 반별, 성별 weight 평균값을 구한 결과를 info\_pivot 변수에 저장하시오.

```
1 ### 답안을 작성하시오.  
2 info_pivot = info.pivot_table('weight', index='noclass', columns='gender')  
3
```

```
1 ### 지우지 마시오.  
2 info_pivot
```

## ▼ 문제8

각 반에서 남학생과 여학생 수를 가리키는 DataFrame 변수 `ct` 를 생성하시오.

(단, 이 때 `noclass`와 `gender` 순서로 multi-index하며 column이름은 `count`로 한다)

```
1 ### 답안을 작성하시오.  
2 ct=info.groupby(['noclass','gender']).count()  
3 ct['count']=ct['ratio']  
4 ct=pd.DataFrame(ct['count'])  
5
```

```
1 ### 지우지 마시오.  
2 ct
```

## ▼ 문제9

각 학생의 키에서 남학생 혹은 여학생 전체 평균키를 뺀 값을 `height_dev` 제목의 column으로 `info` 변수에 추가하시오.

```
1 ### 답안을 작성하시오.
2 info['height_dev']=np.where(info['gender'],
3                               info['height']-info.groupby(['gender']).mean()['height']['female']
4                               ,info['height']-info.groupby(['gender']).mean()['height']['male'])
```

```
1 ### 지우지 마시오.
2 info
```

## ▼ 문제10

*height\_dev* 가 0보다 큰 여학생 수 및 남학생 수를 각각 *ct\_tall\_female* 및 *ct\_tall\_male* 변수에 저장하시오.

```
1 ### 답안을 작성하시오.
2 ### loc : 행 또는 열의 데이터를 조회(loc는 소괄호()가 아닌 대괄호[] 사용)
3
4 ct_tall_female=info.loc[info['height_dev']>0,:].groupby(['gender']).count()['ratio']['female']
5 ct_tall_male=info.loc[info['height_dev']>0,:].groupby(['gender']).count()['ratio']['male']
```

```
1 ### 지우지 마시오.
2 ct_tall_female, ct_tall_male
```

(25, 25)

