

pandas 전처리

1. 제공된 sungjuk.csv 파일을 “c:\python_work\ 폴더에 복사한 후, 전처리 작업을 진행하시오.

	name	kor	eng	math
0	윤수진	90.0	80.0	NaN
1	김은숙	90.0	NaN	NaN
2	강수길	80.0	70.0	NaN
3	문영숙	NaN	90.0	NaN
4	김경은	100.0	80.0	NaN
5	정유진	70.0	NaN	NaN
6	윤경희	90.0	85.0	NaN
7	이은비	60.0	90.0	NaN
8	장은아	NaN	NaN	NaN
9	남희진	NaN	60.0	NaN
10	김수빈	70.0	0.0	NaN

2. 결측치 값이 있는지 조회하시오.

```
name      3
kor       8
eng       6
math     18
dtype: int64
```

3. ‘kor’, ‘eng’, ‘math’ 세 과목 모두 결측치 값을 삭제하시오.

	name	kor	eng	math
0	윤수진	90.0	80.0	NaN
1	김은숙	90.0	NaN	NaN
2	강수길	80.0	70.0	NaN
3	문영숙	NaN	90.0	NaN
4	김경은	100.0	80.0	NaN
5	정유진	70.0	NaN	NaN
6	윤경희	90.0	85.0	NaN
7	이은비	60.0	90.0	NaN
8	장은아	NaN	NaN	NaN
9	남희진	NaN	60.0	NaN
10	김수빈	70.0	0.0	NaN
11	박은진	85.0	85.0	NaN
12	김성경	100.0	85.0	NaN
13	이수정	NaN	90.0	NaN
14	박현아	NaN	80.0	NaN

4. 행 방향으로 2개 이상의 NULL 값이 있는 데이터를 모두 삭제하시오.

	name	kor	eng	math
0	윤수진	90.0	80.0	NaN
2	강수길	80.0	70.0	NaN
4	김경은	100.0	80.0	NaN
6	윤경희	90.0	85.0	NaN
7	이은비	60.0	90.0	NaN
10	김수빈	70.0	0.0	NaN
11	박은진	85.0	85.0	NaN
12	김성경	100.0	85.0	NaN

5. 'math' 과목에 모두 0 값으로 채워넣으시오.

	name	kor	eng	math
0	윤수진	90.0	80.0	0.0
2	강수길	80.0	70.0	0.0
4	김경은	100.0	80.0	0.0
6	윤경희	90.0	85.0	0.0
7	이은비	60.0	90.0	0.0
10	김수빈	70.0	0.0	0.0
11	박은진	85.0	85.0	0.0
12	김성경	100.0	85.0	0.0

6. 'kor', 'eng', 'math' 세과목 모두 정수 처리하시오.

	name	kor	eng	math
0	윤수진	90	80	0
2	강수길	80	70	0
4	김경은	100	80	0
6	윤경희	90	85	0
7	이은비	60	90	0
10	김수빈	70	0	0
11	박은진	85	85	0
12	김성경	100	85	0

7. 인덱스 번호를 순서대로 재설정하시오.

	name	kor	eng	math
0	윤수진	90	80	0
1	강수길	80	70	0
2	김경은	100	80	0
3	윤경희	90	85	0
4	이은비	60	90	0
5	김수빈	70	0	0
6	박은진	85	85	0
7	김성경	100	85	0

8. 8번 인덱스 위치에 아래 데이터를 삽입하시오.

'이은비',60,90,0]

	name	kor	eng	math
0	윤수진	90	80	0
1	강수길	80	70	0
2	김경은	100	80	0
3	윤경희	90	85	0
4	이은비	60	90	0
5	김수빈	70	0	0
6	박은진	85	85	0
7	김성경	100	85	0
8	이은비	60	90	0

9. 중복 데이터가 있는지 조회하시오.

	name	kor	eng	math
4	이은비	60	90	0
8	이은비	60	90	0

10. 중복된 데이터 중 첫 번째 조회되는 중복 데이터는 유지하고 그 이후 중복되는 데이터를 삭제하시오.

	name	kor	eng	math
0	윤수진	90	80	0
1	강수길	80	70	0
2	김경은	100	80	0
3	윤경희	90	85	0
4	이은비	60	90	0
5	김수빈	70	0	0
6	박은진	85	85	0
7	김성경	100	85	0

11. 데이터가 없는 ‘math’ 열을 삭제해주세요.

	name	kor	eng
0	윤수진	90	80
1	강수길	80	70
2	김경은	100	80
3	윤경희	90	85
4	이은비	60	90
5	김수빈	70	0
6	박은진	85	85
7	김성경	100	85

12. 두 과목 중 한 과목이라도 0점인 과목은 삭제해 주세요.

	name	kor	eng
0	윤수진	90	80
1	강수길	80	70
2	김경은	100	80
3	윤경희	90	85
4	이은비	60	90
6	박은진	85	85
7	김성경	100	85

13. 'kor' 과목이 60점 이하인 경우 65점으로 데이터를 상향 조정해 주세요.

	name	kor	eng
0	윤수진	90	80
1	강수길	80	70
2	김경은	100	80
3	윤경희	90	85
4	이은비	65	90
6	박은진	85	85
7	김성경	100	85

14. 순위를 부여하시오. 단, 순위는 kor+eng 점수에 대해 점수가 높은 순으로 순위를 부여하시오.
Hint : df['열'].rank(method='dense', ascending=False)

	name	kor	eng	순위
0	윤수진	90	80	4
1	강수길	80	70	6
2	김경은	100	80	2
3	윤경희	90	85	3
4	이은비	65	90	5
6	박은진	85	85	4
7	김성경	100	85	1

15. 성적이 우수한 성적 순위 기준으로 정렬하시오.

	name	kor	eng	순위
7	김성경	100	85	1
2	김경은	100	80	2
3	윤경희	90	85	3
0	윤수진	90	80	4
6	박은진	85	85	4
4	이은비	65	90	5
1	강수길	80	70	6