

데이터 연결하기

- 분석하기 좋은 데이터
- 데이터 연결 기초
- 데이터 연결 마무리

■ 분석하기 좋은 데이터

● 깔끔한 데이터의 조건

- 분석 목적에 맞는 데이터를 모아 표(Table) 형태로 생성
- 측정한 값은 행(row)으로 구성
- 변수는 열(column)로 구성

● 깔끔한 데이터는 데이터 연결부터

- 주식 데이터 분석 과정에서 '기업정보' 데이터 집합과 '주식 가격' 데이터 집합이 있을 때 '첨단 산업 기업의 주식 가격에 대한 데이터'를 얻으려면 공통의 Key를 찾아 연결

■ 데이터 연결 기초

● concat 메소드로 데이터 연결하기

- 데이터 준비 (csv 파일 읽어오기)

```
import pandas as pd
df1 = pd.read_csv('data/concat_1.csv')
df2 = pd.read_csv('data/concat_2.csv')
df3 = pd.read_csv('data/concat_3.csv')
```

- 데이터 연결

```
row_concat = pd.concat([df1, df2, df3])
row_concat
```

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
0	a4	b4	c4	d4
1	a5	b5	c5	d5
2	a6	b6	c6	d6
3	a7	b7	c7	d7
0	a8	b8	c8	d8
1	a9	b9	c9	d9
2	a10	b10	c10	d10
3	a11	b11	c11	d11

■ 데이터 연결 기초

● concat 메소드로 데이터 연결하기

- 4번째 행 데이터 확인 (iloc)

```
print(row_concat.iloc[3, ])
```

```
A    a3  
B    b3  
C    c3  
D    d3  
Name: 3, dtype: object
```

- 4번째 행 데이터 확인 (loc)

```
print(row_concat.loc[3, ])
```

	A	B	C	D
3	a3	b3	c3	d3
3	a7	b7	c7	d7
3	a11	b11	c11	d11

■ 데이터 연결 기초

- concat 메소드로 데이터 연결하기
 - 데이터프레임에 시리즈 연결하기

df1

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3

+

```
new_row_series = pd.Series(  
    ['n1', 'n2', 'n3', 'n4'])  
new_row_series
```

```
0    n1  
1    n2  
2    n3  
3    n4  
dtype: object
```



```
print(pd.concat([df1, new_row_series]))
```

	A	B	C	D	
0	a0	b0	c0	d0	NaN
1	a1	b1	c1	d1	NaN
2	a2	b2	c2	d2	NaN
3	a3	b3	c3	d3	NaN
0	NaN	NaN	NaN	NaN	n1
1	NaN	NaN	NaN	NaN	n2
2	NaN	NaN	NaN	NaN	n3
3	NaN	NaN	NaN	NaN	n4

■ 데이터 연결 기초

● concat 메소드로 데이터 연결하기

- 데이터프레임에 시리즈 연결하기

➔ 시리즈에는 열 이름이 없기 때문에

행으로 인식하지 않고 새로운 열로 간주

• 행이 1개인 데이터프레임으로 생성

```
new_row_df = pd.DataFrame(  
    [['n1', 'n2', 'n3', 'n4']], columns=['A', 'B', 'C', 'D'])  
new_row_df
```

	A	B	C	D
0	n1	n2	n3	n4

`pd.concat([df1, new_row_df])`

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
0	n1	n2	n3	n4

■ 데이터 연결 기초

● append 메소드로 데이터 연결하기

- 연결할 데이터프레임이 1개이면 append 메소드 사용 가능

➔ concat 메소드는 2개 이상 연결 가능

```
df1.append(new_row_df)
```

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
0	n1	n2	n3	n4

■ 데이터 연결 기초

● append 메소드로 데이터 연결하기

- append 메소드와 딕셔너리를 사용하면 간편하게 연결 가능

➔ 단, ignore_index 옵션을 True로 지정

```
df1.append(nedata_dict = {'A': 'n1', 'B': 'n2', 'C': 'n3', 'D': 'n4'})  
print(df1.append(data_dict, ignore_index=True))
```

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
4	n1	n2	n3	n4

■ 데이터 연결 기초

● 다양한 방법으로 데이터 연결하기

- ignore_index를 True 로 지정하여 인덱스 정리

```
row_concat_i = pd.concat([df1, df2, df3])  
print(row_concat_i)
```

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
0	a4	b4	c4	d4
1	a5	b5	c5	d5
2	a6	b6	c6	d6
3	a7	b7	c7	d7
0	a8	b8	c8	d8
1	a9	b9	c9	d9
2	a10	b10	c10	d10
3	a11	b11	c11	d11

```
row_concat_i = pd.concat(  
    [df1, df2, df3], ignore_index=True)  
print(row_concat_i)
```

	A	B	C	D
0	a0	b0	c0	d0
1	a1	b1	c1	d1
2	a2	b2	c2	d2
3	a3	b3	c3	d3
4	a4	b4	c4	d4
5	a5	b5	c5	d5
6	a6	b6	c6	d6
7	a7	b7	c7	d7
8	a8	b8	c8	d8
9	a9	b9	c9	d9
10	a10	b10	c10	d10
11	a11	b11	c11	d11

■ 데이터 연결 기초

● 다양한 방법으로 데이터 연결하기

- 열 방향으로 데이터 연결하기

```
col_concat = pd.concat([df1, df2, df3], axis=1)  
print(col_concat)
```

	A	B	C	D	A	B	C	D	A	B	C	D
0	a0	b0	c0	d0	a4	b4	c4	d4	a8	b8	c8	d8
1	a1	b1	c1	d1	a5	b5	c5	d5	a9	b9	c9	d9
2	a2	b2	c2	d2	a6	b6	c6	d6	a10	b10	c10	d10
3	a3	b3	c3	d3	a7	b7	c7	d7	a11	b11	c11	d11

※ 해당 열 이름을 조회하면 같은 열 이름 데이터를 모두 추출

```
print(col_concat['A'])
```

	A	A	A
0	a0	a4	a8
1	a1	a5	a9
2	a2	a6	a10
3	a3	a7	a11

■ 데이터 연결 기초

● 다양한 방법으로 데이터 연결하기

- 열 방향으로 데이터 연결하면서 인덱스 정리

```
print(pd.concat([df1, df2, df3], axis=1, ignore_index=True))
```

	0	1	2	3	4	5	6	7	8	9	10	11
0	a0	b0	c0	d0	a4	b4	c4	d4	a8	b8	c8	d8
1	a1	b1	c1	d1	a5	b5	c5	d5	a9	b9	c9	d9
2	a2	b2	c2	d2	a6	b6	c6	d6	a10	b10	c10	d10
3	a3	b3	c3	d3	a7	b7	c7	d7	a11	b11	c11	d11

- 열 이름 새로 지정

```
new_df = pd.concat([df1, df2, df3], axis=1, ignore_index=True)
new_df.columns = [
    'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L']
print(new_df)
```

	A	B	C	D	E	F	G	H	I	J	K	L
0	a0	b0	c0	d0	a4	b4	c4	d4	a8	b8	c8	d8
1	a1	b1	c1	d1	a5	b5	c5	d5	a9	b9	c9	d9
2	a2	b2	c2	d2	a6	b6	c6	d6	a10	b10	c10	d10
3	a3	b3	c3	d3	a7	b7	c7	d7	a11	b11	c11	d11

■ 데이터 연결 기초

- 다양한 방법으로 데이터 연결하기
 - 공통 열과 공통 인덱스 연결하기

```
df1.columns = ['A', 'B', 'C', 'D']
df2.columns = ['E', 'F', 'G', 'H']
df3.columns = ['A', 'C', 'F', 'H']
print(df1)
print(type(df1))
print(df2)
print(type(df2))
print(df3)
print(type(df3))
```

```
      A    B    C    D
0  a0  b0  c0  d0
1  a1  b1  c1  d1
2  a2  b2  c2  d2
3  a3  b3  c3  d3
<class 'pandas.core.frame.DataFrame'>
      E    F    G    H
0  a4  b4  c4  d4
1  a5  b5  c5  d5
2  a6  b6  c6  d6
3  a7  b7  c7  d7
<class 'pandas.core.frame.DataFrame'>
      A    C    F    H
0  a8  b8  c8  d8
1  a9  b9  c9  d9
2  a10 b10 c10 d10
3  a11 b11 c11 d11
<class 'pandas.core.frame.DataFrame'>
```

■ 데이터 연결 기초

- 다양한 방법으로 데이터 연결하기
 - 공통 열과 공통 인덱스 연결하기

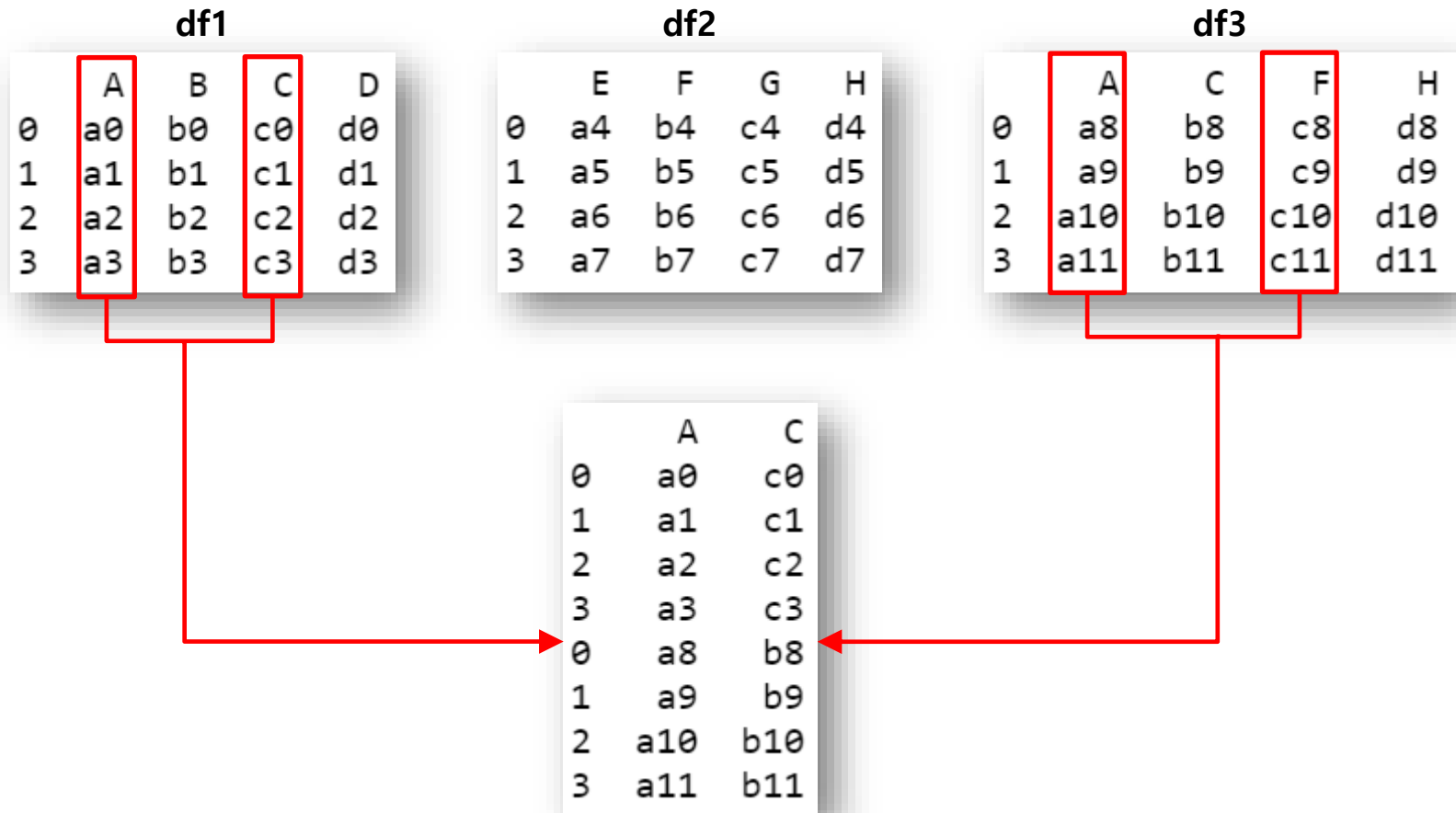
```
row_concat = pd.concat([df1, df2, df3])  
print(row_concat)
```

	A	B	C	D	E	F	G	H
0	a0	b0	c0	d0	NaN	NaN	NaN	NaN
1	a1	b1	c1	d1	NaN	NaN	NaN	NaN
2	a2	b2	c2	d2	NaN	NaN	NaN	NaN
3	a3	b3	c3	d3	NaN	NaN	NaN	NaN
0	NaN	NaN	NaN	NaN	a4	b4	c4	d4
1	NaN	NaN	NaN	NaN	a5	b5	c5	d5
2	NaN	NaN	NaN	NaN	a6	b6	c6	d6
3	NaN	NaN	NaN	NaN	a7	b7	c7	d7
0	a8	NaN	b8	NaN	NaN	c8	NaN	d8
1	a9	NaN	b9	NaN	NaN	c9	NaN	d9
2	a10	NaN	b10	NaN	NaN	c10	NaN	d10
3	a11	NaN	b11	NaN	NaN	c11	NaN	d11

데이터프레임에 해당 열 이름이 없으면
누락값으로 처리 (NaN)

■ 데이터 연결 기초

- 다양한 방법으로 데이터 연결하기
 - 공통 열과 공통 인덱스만 연결하기 (join)

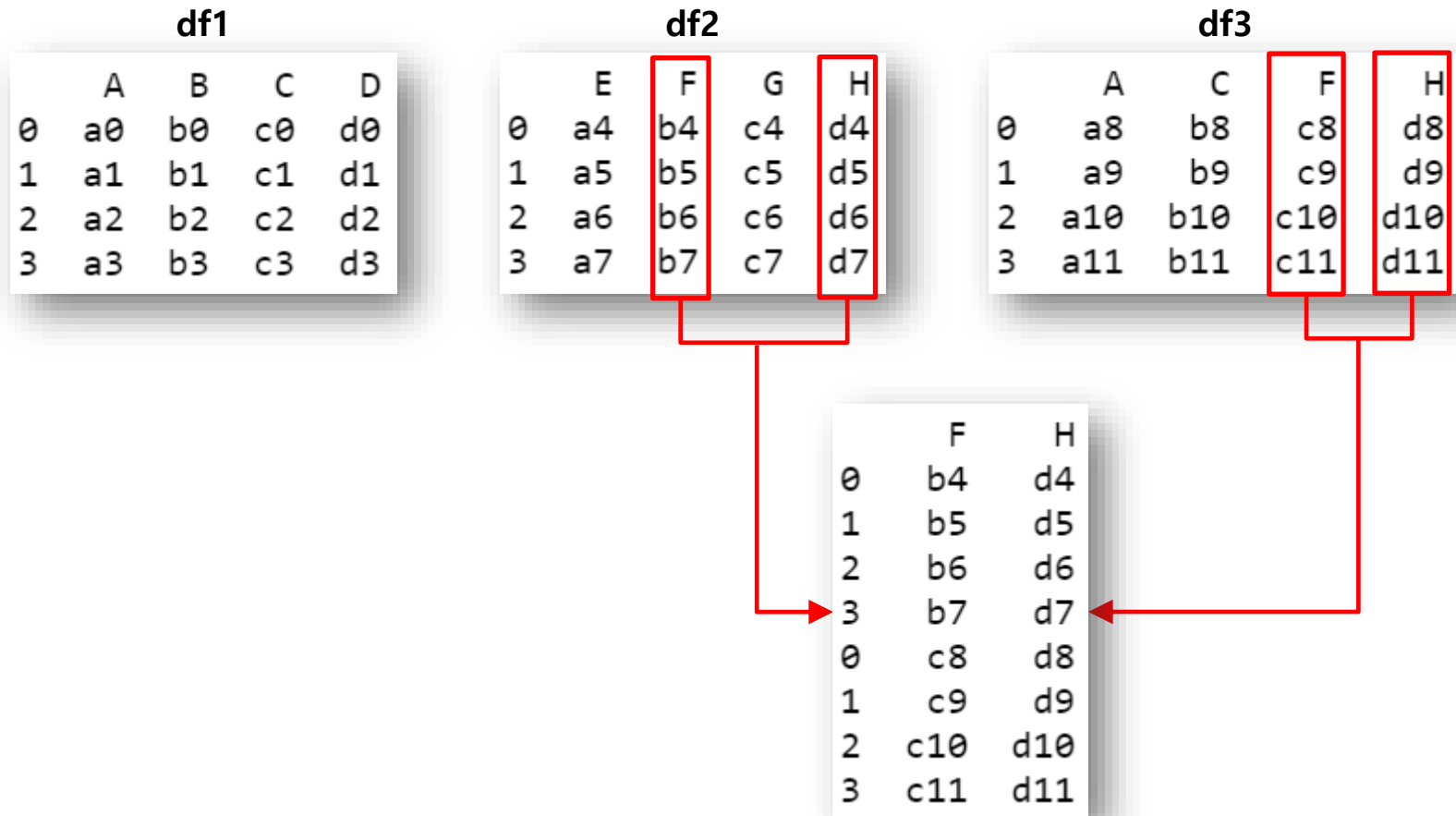


```
print(pd.concat([df1, df3], join='inner'))
```

■ 데이터 연결 기초

● 다양한 방법으로 데이터 연결하기

- 공통 열과 공통 인덱스만 연결하기 (join)



```
print(pd.concat([df2, df3], join='inner'))
```

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 날씨 정보 데이터

```
person = pd.read_csv('data/survey_person.csv')
site = pd.read_csv('data/survey_site.csv')
visited = pd.read_csv('data/survey_visited.csv')
survey = pd.read_csv('data/survey_survey.csv')
```

person (관측자)

	ident	personal	family
0	dye	William	Dye
1	pb	Frank	Pabodie
2	lake	Anderson	Lake
3	roe	Valentina	Roerich
4	danforth	Frank	Danforth

site (장소)

	name	lat	long
0	DR-1	-49.85	-128.57
1	DR-3	-47.15	-126.72
2	MSK-4	-48.87	-123.40

visited (날짜)

	ident	site	dated
0	619	DR-1	1927-02-08
1	622	DR-1	1927-02-10
2	734	DR-3	1939-01-07
3	735	DR-3	1930-01-12
4	751	DR-3	1930-02-26
5	752	DR-3	NaN
6	837	MSK-4	1932-01-14
7	844	DR-1	1932-03-22

survey (날씨정보)

	taken	person	quant	reading
0	619	dye	rad	9.82
1	619	dye	sal	0.13
2	622	dye	rad	7.80
3	622	dye	sal	0.09
4	734	pb	rad	8.41
5	734	lake	sal	0.05
6	734	pb	temp	-21.50
7	735	pb	rad	7.22
8	735	NaN	sal	0.06
9	735	NaN	temp	-26.00
10	751	pb	rad	4.35
11	751	pb	temp	-18.50
12	751	lake	sal	0.10
13	752	lake	rad	2.19
14	752	lake	sal	0.09
15	752	lake	temp	-16.00
16	752	roe	sal	41.60
17	837	lake	rad	1.46
18	837	lake	sal	0.21
19	837	roe	sal	22.50
20	844	roe	rad	11.25

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 연결하려는 열 이름이 다른 경우 left_on, right_on 옵션 사용

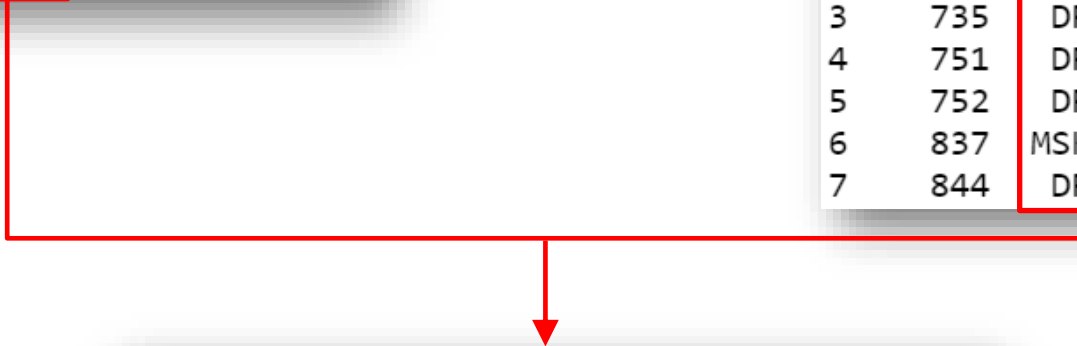
```
sv_merge = site.merge(visited, left_on='name', right_on='site')  
print(sv_merge)
```

site (장소)

	name	lat	long
0	DR-1	-49.85	-128.57
1	DR-3	-47.15	-126.72
2	MSK-4	-48.87	-123.40

visited (날짜)

	ident	site	dated
0	619	DR-1	1927-02-08
1	622	DR-1	1927-02-10
2	734	DR-3	1939-01-07
3	735	DR-3	1930-01-12
4	751	DR-3	1930-02-26
5	752	DR-3	NaN
6	837	MSK-4	1932-01-14
7	844	DR-1	1932-03-22



	name	lat	long	ident	site	dated
0	DR-1	-49.85	-128.57	619	DR-1	1927-02-08
1	DR-1	-49.85	-128.57	622	DR-1	1927-02-10
2	DR-1	-49.85	-128.57	844	DR-1	1932-03-22
3	DR-3	-47.15	-126.72	734	DR-3	1939-01-07
4	DR-3	-47.15	-126.72	735	DR-3	1930-01-12
5	DR-3	-47.15	-126.72	751	DR-3	1930-02-26
6	DR-3	-47.15	-126.72	752	DR-3	NaN
7	MSK-4	-48.87	-123.40	837	MSK-4	1932-01-14

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 연결하려는 열 이름이 다른 경우 left_on, right_on 옵션 사용

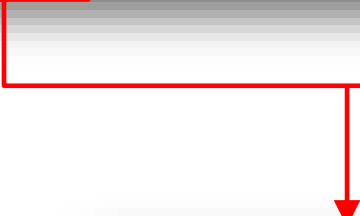
```
ps_merge = person.merge(survey, left_on='ident', right_on='person')  
print(ps_merge)
```

person (관측자)

	ident	personal	family
0	dyer	William	Dyer
1	pb	Frank	Pabodie
2	lake	Anderson	Lake
3	roe	Valentina	Roerich
4	danforth	Frank	Danforth

survey (날씨정보)

	taken	person	quant	reading
0	619	dyer	rad	9.82
1	619	dyer	sal	0.13
2	622	dyer	rad	7.80
3	622	dyer	sal	0.09
4	734	pb	rad	8.41
5	734	lake	sal	0.05
6	734	pb	temp	-21.50
7	735	pb	rad	7.22
8	735	NaN	sal	0.06
9	735	NaN	temp	-26.00
10	751	pb	rad	4.35
11	751	pb	temp	-18.50
12	751	lake	sal	0.10
13	752	lake	rad	2.19
14	752	lake	sal	0.09
15	752	lake	temp	-16.00
16	752	roe	sal	41.60
17	837	lake	rad	1.46
18	837	lake	sal	0.21
19	837	roe	sal	22.50
20	844	roe	rad	11.25



	ident	personal	family	taken	person	quant	reading
0	dyer	William	Dyer	619	dyer	rad	9.82
1	dyer	William	Dyer	619	dyer	sal	0.13
2	dyer	William	Dyer	622	dyer	rad	7.80
3	dyer	William	Dyer	622	dyer	sal	0.09
4	pb	Frank	Pabodie	734	pb	rad	8.41
5	pb	Frank	Pabodie	734	pb	temp	-21.50
6	pb	Frank	Pabodie	735	pb	rad	7.22
7	pb	Frank	Pabodie	751	pb	rad	4.35
8	pb	Frank	Pabodie	751	pb	temp	-18.50
9	lake	Anderson	Lake	734	lake	sal	0.05
10	lake	Anderson	Lake	751	lake	sal	0.10
11	lake	Anderson	Lake	752	lake	rad	2.19
12	lake	Anderson	Lake	752	lake	sal	0.09
13	lake	Anderson	Lake	752	lake	temp	-16.00
14	lake	Anderson	Lake	837	lake	rad	1.46
15	lake	Anderson	Lake	837	lake	sal	0.21
16	roe	Valentina	Roerich	752	roe	sal	41.60
17	roe	Valentina	Roerich	837	roe	sal	22.50
18	roe	Valentina	Roerich	844	roe	rad	11.25

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 연결하려는 열 이름이 다른 경우 left_on, right_on 옵션 사용

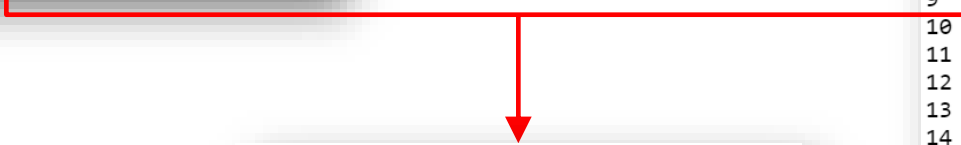
```
vs_merge = visited.merge(survey, left_on='ident', right_on='taken')  
print(vs_merge)
```

visited (날짜)

	ident	site	dated
0	619	DR-1	1927-02-08
1	622	DR-1	1927-02-10
2	734	DR-3	1939-01-07
3	735	DR-3	1930-01-12
4	751	DR-3	1930-02-26
5	752	DR-3	NaN
6	837	MSK-4	1932-01-14
7	844	DR-1	1932-03-22

survey (날씨정보)

	taken	person	quant	reading
0	619	dye	rad	9.82
1	619	dye	sal	0.13
2	622	dye	rad	7.80
3	622	dye	sal	0.09
4	734	pb	rad	8.41
5	734	lake	sal	0.05
6	734	pb	temp	-21.50
7	735	pb	rad	7.22
8	735	NaN	sal	0.06
9	735	NaN	temp	-26.00
10	751	pb	rad	4.35
11	751	pb	temp	-18.50
12	751	lake	sal	0.10
13	752	lake	rad	2.19
14	752	lake	sal	0.09
15	752	lake	temp	-16.00
16	752	roe	sal	41.60
17	837	lake	rad	1.46
18	837	lake	sal	0.21
19	837	roe	sal	22.50
20	844	roe	rad	11.25



	ident	site	dated	taken	person	quant	reading
0	619	DR-1	1927-02-08	619	dye	rad	9.82
1	619	DR-1	1927-02-08	619	dye	sal	0.13
2	622	DR-1	1927-02-10	622	dye	rad	7.80
3	622	DR-1	1927-02-10	622	dye	sal	0.09
4	734	DR-3	1939-01-07	734	pb	rad	8.41
5	734	DR-3	1939-01-07	734	lake	sal	0.05
6	734	DR-3	1939-01-07	734	pb	temp	-21.50
7	735	DR-3	1930-01-12	735	pb	rad	7.22
8	735	DR-3	1930-01-12	735	NaN	sal	0.06
9	735	DR-3	1930-01-12	735	NaN	temp	-26.00
10	751	DR-3	1930-02-26	751	pb	rad	4.35
11	751	DR-3	1930-02-26	751	pb	temp	-18.50
12	751	DR-3	1930-02-26	751	lake	sal	0.10
13	752	DR-3	NaN	752	lake	rad	2.19
14	752	DR-3	NaN	752	lake	sal	0.09
15	752	DR-3	NaN	752	lake	temp	-16.00
16	752	DR-3	NaN	752	roe	sal	41.60
17	837	MSK-4	1932-01-14	837	lake	rad	1.46
18	837	MSK-4	1932-01-14	837	lake	sal	0.21
19	837	MSK-4	1932-01-14	837	roe	sal	22.50
20	844	DR-1	1932-03-22	844	roe	rad	11.25

■ 데이터 연결 마무리

- 데이터 연결 전용 메소드 merge 사용하기
 - 4개의 데이터 모두 연결

person (관측자)

	ident	personal	family
0	dyer	William	Dyer
1	pb	Frank	Pabodie
2	lake	Anderson	Lake
3	roe	Valentina	Roerich
4	danforth	Frank	Danforth

visited (날짜)

	ident	site	dated
0	619	DR-1	1927-02-08
1	622	DR-1	1927-02-10
2	734	DR-3	1939-01-07
3	735	DR-3	1930-01-12
4	751	DR-3	1930-02-26
5	752	DR-3	NaN
6	837	MSK-4	1932-01-14
7	844	DR-1	1932-03-22

site (장소)

	name	lat	long
0	DR-1	-49.85	-128.57
1	DR-3	-47.15	-126.72
2	MSK-4	-48.87	-123.40

survey (날씨정보)

	taken	person	quant	reading
0	619	dyer	rad	9.82
1	619	dyer	sal	0.13
2	622	dyer	rad	7.80
3	622	dyer	sal	0.09
4	734	pb	rad	8.41
5	734	lake	sal	0.05
6	734	pb	temp	-21.50
7	735	pb	rad	7.22
8	735	NaN	sal	0.06
9	735	NaN	temp	-26.00
10	751	pb	rad	4.35
11	751	pb	temp	-18.50
12	751	lake	sal	0.10
13	752	lake	rad	2.19
14	752	lake	sal	0.09
15	752	lake	temp	-16.00
16	752	roe	sal	41.60
17	837	lake	rad	1.46
18	837	lake	sal	0.21
19	837	roe	sal	22.50
20	844	roe	rad	11.25

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 4개의 데이터 모두 연결 (survey + **person**)

```
sp = survey.merge(person, left_on='person', right_on='ident')  
print(sp)
```

	taken	person	quant	reading	ident	personal	family
0	619	dyer	rad	9.82	dyer	William	Dyer
1	619	dyer	sal	0.13	dyer	William	Dyer
2	622	dyer	rad	7.80	dyer	William	Dyer
3	622	dyer	sal	0.09	dyer	William	Dyer
4	734	pb	rad	8.41	pb	Frank	Pabodie
5	734	pb	temp	-21.50	pb	Frank	Pabodie
6	735	pb	rad	7.22	pb	Frank	Pabodie
7	751	pb	rad	4.35	pb	Frank	Pabodie
8	751	pb	temp	-18.50	pb	Frank	Pabodie
9	734	lake	sal	0.05	lake	Anderson	Lake
10	751	lake	sal	0.10	lake	Anderson	Lake
11	752	lake	rad	2.19	lake	Anderson	Lake
12	752	lake	sal	0.09	lake	Anderson	Lake
13	752	lake	temp	-16.00	lake	Anderson	Lake
14	837	lake	rad	1.46	lake	Anderson	Lake
15	837	lake	sal	0.21	lake	Anderson	Lake
16	752	roe	sal	41.60	roe	Valentina	Roerich
17	837	roe	sal	22.50	roe	Valentina	Roerich
18	844	roe	rad	11.25	roe	Valentina	Roerich

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 4개의 데이터 모두 연결 (survey + person + visit)

```
vs_merge = visited.merge(survey, left_on='ident', right_on='taken')  
print(vs_merge)
```

	taken	person	quant	reading	ident_x	personal	family	ident_y	site	dated
0	619	dye	rad	9.82	dye	William	Dye	619	DR-1	1927-02-08
1	619	dye	sal	0.13	dye	William	Dye	619	DR-1	1927-02-08
2	622	dye	rad	7.80	dye	William	Dye	622	DR-1	1927-02-10
3	622	dye	sal	0.09	dye	William	Dye	622	DR-1	1927-02-10
4	734	pb	rad	8.41	pb	Frank	Pabodie	734	DR-3	1939-01-07
5	734	pb	temp	-21.50	pb	Frank	Pabodie	734	DR-3	1939-01-07
6	734	lake	sal	0.05	lake	Anderson	Lake	734	DR-3	1939-01-07
7	735	pb	rad	7.22	pb	Frank	Pabodie	735	DR-3	1930-01-12
8	751	pb	rad	4.35	pb	Frank	Pabodie	751	DR-3	1930-02-26
9	751	pb	temp	-18.50	pb	Frank	Pabodie	751	DR-3	1930-02-26
10	751	lake	sal	0.10	lake	Anderson	Lake	751	DR-3	1930-02-26
11	752	lake	rad	2.19	lake	Anderson	Lake	752	DR-3	NaN
12	752	lake	sal	0.09	lake	Anderson	Lake	752	DR-3	NaN
13	752	lake	temp	-16.00	lake	Anderson	Lake	752	DR-3	NaN
14	752	roe	sal	41.60	roe	Valentina	Roerich	752	DR-3	NaN
15	837	lake	rad	1.46	lake	Anderson	Lake	837	MSK-4	1932-01-14
16	837	lake	sal	0.21	lake	Anderson	Lake	837	MSK-4	1932-01-14
17	837	roe	sal	22.50	roe	Valentina	Roerich	837	MSK-4	1932-01-14
18	844	roe	rad	11.25	roe	Valentina	Roerich	844	DR-1	1932-03-22

■ 데이터 연결 마무리

● 데이터 연결 전용 메소드 merge 사용하기

- 4개의 데이터 모두 연결 (survey + person + visit + site)

```
spvs = spv.merge(site, left_on='site', right_on='name')  
print(spvs)
```

	taken	person	quant	reading	ident_x	personal	family	ident_y	site	dated	name	lat	long
0	619	dye	rad	9.82	dye	William	Dye	619	DR-1	1927-02-08	DR-1	-49.85	-128.57
1	619	dye	sal	0.13	dye	William	Dye	619	DR-1	1927-02-08	DR-1	-49.85	-128.57
2	622	dye	rad	7.80	dye	William	Dye	622	DR-1	1927-02-10	DR-1	-49.85	-128.57
3	622	dye	sal	0.09	dye	William	Dye	622	DR-1	1927-02-10	DR-1	-49.85	-128.57
4	734	pb	rad	8.41	pb	Frank	Pabodie	734	DR-3	1939-01-07	DR-1	-49.85	-128.57
5	734	pb	temp	-21.50	pb	Frank	Pabodie	734	DR-3	1939-01-07	DR-3	-47.15	-126.72
6	734	lake	sal	0.05	lake	Anderson	Lake	734	DR-3	1939-01-07	DR-3	-47.15	-126.72
7	735	pb	rad	7.22	pb	Frank	Pabodie	735	DR-3	1930-01-12	DR-3	-47.15	-126.72
8	751	pb	rad	4.35	pb	Frank	Pabodie	751	DR-3	1930-02-26	DR-3	-47.15	-126.72
9	751	pb	temp	-18.50	pb	Frank	Pabodie	751	DR-3	1930-02-26	DR-3	-47.15	-126.72
10	751	lake	sal	0.10	lake	Anderson	Lake	751	DR-3	1930-02-26	DR-3	-47.15	-126.72
11	752	lake	rad	2.19	lake	Anderson	Lake	752	DR-3	NaN	DR-3	-47.15	-126.72
12	752	lake	sal	0.09	lake	Anderson	Lake	752	DR-3	NaN	DR-3	-47.15	-126.72
13	752	lake	temp	-16.00	lake	Anderson	Lake	752	DR-3	NaN	DR-3	-47.15	-126.72
14	752	roe	sal	41.60	roe	Valentina	Roerich	752	DR-3	NaN	DR-3	-47.15	-126.72
15	837	lake	rad	1.46	lake	Anderson	Lake	837	MSK-4	1932-01-14	DR-3	-47.15	-126.72
16	837	lake	sal	0.21	lake	Anderson	Lake	837	MSK-4	1932-01-14	MSK-4	-48.87	-123.40
17	837	roe	sal	22.50	roe	Valentina	Roerich	837	MSK-4	1932-01-14	MSK-4	-48.87	-123.40
18	844	roe	rad	11.25	roe	Valentina	Roerich	844	DR-1	1932-03-22	MSK-4	-48.87	-123.40