

# 03\_데이터 연결하기

## • 데이터 연결 기초

- > Concat 메서드로 데이터 연결하기
  - 데이터 준비(csv로 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')
```

Out :

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

df2

	A	B	C	D
0	a4	b4	c4	d4
1	a5	b5	c5	d5
2	a6	b6	c6	d6
3	a7	b7	c7	d7

df3

	A	B	C	D
0	a8	b8	c8	d8
1	a9	b9	c9	d9
2	a10	b10	c10	d10
3	a11	b11	c11	d11

## • 데이터 연결 기초

### > Concat 메서드로 데이터 연결하기

- 데이터 연결

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

Out :

	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

04

### > Concat 메서드로 데이터 연결하기

- iloc로 4번째 행 확인

```
row_concat.iloc[3]
```

**Out :**

A	a3
B	b3
C	c3
D	d3

Name: 3, dtype: object

- loc로 3 인덱스 확인

```
row_concat.loc[3]
```

**Out :**

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

## • 데이터 연결 기초

05

> Concat 메서드를 사용하여 열 방향으로 연결하기

- axis = 1

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

```
col_concat
```

Out :

	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

```
col_concat['A']
```

Out :

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

## • 데이터 연결 기초

> ignore\_index = True

- 데이터프레임을 연결할 때, 인덱스를 초기화시키는 옵션

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

Out :

	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

> Concat 메서드를 사용하여 열 방향으로 연결하기

- axis = 1

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

Out :

	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

## • 데이터 연결 기초

> Concat 메서드를 사용하여 열 방향으로 연결하기

- `df.columns = []` 로 컬럼명 변경

```
col_concat.columns = ['A','B','C','D','E','F','G','H','I','J','K','L']
```

```
col_concat
```

Out :

	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 인덱스로 연결하기

```
df1.columns = ['A', 'B', 'C', 'D']
```

```
df2.columns = ['E', 'F', 'G', 'H']
```

```
df3.columns = ['A', 'C', 'F', 'H']
```

Out :

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

df2

	E	F	G	H
0	a4	b4	c4	d4
1	a5	b5	c5	d5
2	a6	b6	c6	d6
3	a7	b7	c7	d7

df3

	A	C	F	H
0	a8	b8	c8	d8
1	a9	b9	c9	d9
2	a10	b10	c10	d10
3	a11	b11	c11	d11

## • 데이터 연결 기초

### > 다양한 방법으로 데이터 연결하기

- 공통 열과 인덱스로 연결하기

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

Out :

	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

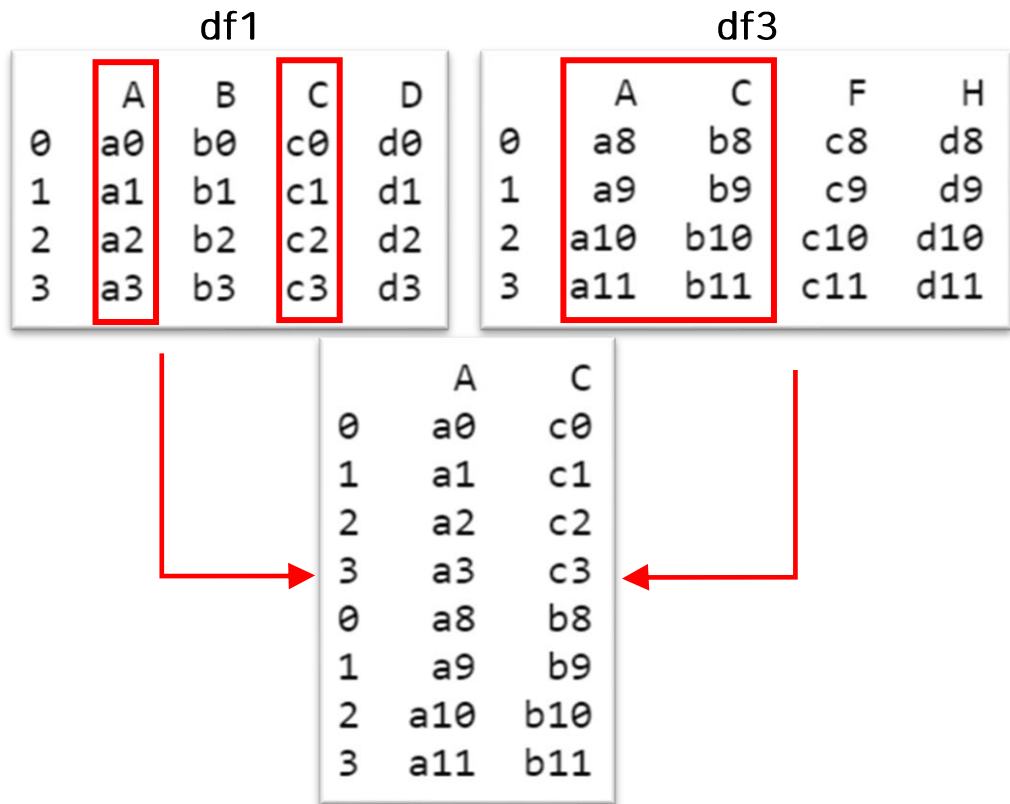
## • 데이터 연결 기초

- > 다양한 방법으로 데이터 연결하기
  - 공통 열과 인덱스로 연결하기(join)

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

Out :

	A	C
0	a0	c0
1	a1	c1
2	a2	c2
3	a3	c3
0	a8	b8
1	a9	b9
2	a10	b10
3	a11	b11



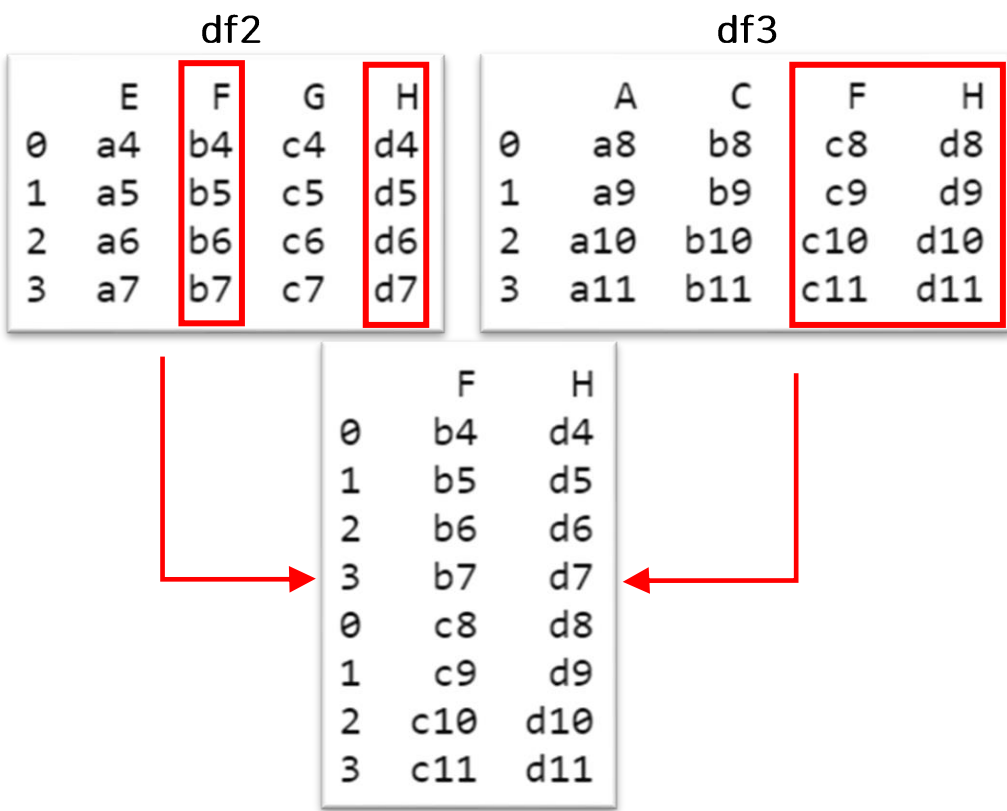
- 데이터 연결 기초

- > 다양한 방법으로 데이터 연결하기
  - 공통 열과 인덱스로 연결하기(join)

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

Out :

	F	H
0	b4	d4
1	b5	d5
2	b6	d6
3	b7	d7
0	c8	d8
1	c9	d9
2	c10	d10
3	c11	d11



## • 데이터 연결 기초

- > merge 메서드를 사용한 데이터 연결하기
  - 서로 같은 키를 가지고 있는 경우

```
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')
```

- 데이터 연결 기초

- > merge 메서드를 사용한 데이터 연결하기

- 서로 같은 키를 가지고 있는 경우

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

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	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 기초

- > merge 메서드를 사용한 데이터 연결하기
  - left\_on, right\_on을 사용하여 공통 키 선택

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

Out :

	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 기초

- > merge 메서드를 사용한 데이터 연결하기
  - left\_on, right\_on을 사용하여 공통 키 선택

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

Out :

	name	lat	long		ident	site	dated
0	DR-1	-49.85	-128.57	0	619	DR-1	1927-02-08
1	DR-3	-47.15	-126.72	1	622	DR-1	1927-02-10
2	MSK-4	-48.87	-123.40	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



## • 데이터 연결 기초

017

- > merge 메서드를 사용한 데이터 연결하기
  - left\_on, right\_on을 사용하여 공통 키 선택

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

Out :

	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 기초

- > merge 메서드를 사용한 데이터 연결하기
  - left\_on, right\_on을 사용하여 공통 키 선택

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

Out :

	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 연결하기

## • 데이터 연결 기초

### > merge 메서드를 사용한 데이터 연결하기

- 서로 같은 키를 가지고 있는 경우

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

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	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 기초

020

### > merge 메서드를 사용한 데이터 연결하기

- 4개의 데이터 모두 연결하기(survey + person)

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

sp

Out :

	taken	person	quant	reading	ident	personal	family
0	619	dye	rad	9.82	dye	William	Dye
1	619	dye	sal	0.13	dye	William	Dye
2	622	dye	rad	7.80	dye	William	Dye
3	622	dye	sal	0.09	dye	William	Dye
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 기초

021

### > merge 메서드를 사용한 데이터 연결하기

- 4개의 데이터 모두 연결하기(survey + person + visited)

```
spv = sp.merge(visited, left_on = 'taken', right_on='ident')
```

spv

Out :

	taken	person	quant	reading	ident_x	personal	family	ident_y	site	dated
0	619	dyer	rad	9.82	dyer	William	Dyer	619	DR-1	1927-02-08
1	619	dyer	sal	0.13	dyer	William	Dyer	619	DR-1	1927-02-08
2	622	dyer	rad	7.80	dyer	William	Dyer	622	DR-1	1927-02-10
3	622	dyer	sal	0.09	dyer	William	Dyer	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 기초

### > merge 메서드를 사용한 데이터 연결하기

- 4개의 데이터 모두 연결하기(survey + person + visited + site)

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

spvs

Out :

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

## • 데이터 연결 기초

> merge 메서드를 사용한 데이터 연결하기

- 중복 컬럼 제거(ident\_x, ident\_y, name)

```
spvs.drop(['ident_x', 'ident_y', 'name'], axis = 1)
```

Out :

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

- 데이터베이스 불러오기

- > 데이터베이스 불러오기

```
import cx_Oracle as oci
import pandas as pd

dsn = oci.makedsn('localhost', '1521', 'xe')
conn = oci.connect('scott', 'tiger', dsn)

emp = pd.read_sql('select * from emp', conn)
dept = pd.read_sql('select * from dept', conn)

conn.close()
```



• 데이터베이스 불러오기

> 데이터베이스 불러오기

emp

Out :

	EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
0	7369	SMITH	CLERK	7902.0	1980-12-17	800.0	NaN	20
1	7499	ALLEN	SALESMAN	7698.0	1981-02-20	1600.0	300.0	30
2	7521	WARD	SALESMAN	7698.0	1981-02-22	1250.0	500.0	30
3	7566	JONES	MANAGER	7839.0	1981-04-02	2975.0	NaN	20
4	7654	MARTIN	SALESMAN	7698.0	1981-09-28	1250.0	1400.0	30
5	7698	BLAKE	MANAGER	7839.0	1981-05-01	2850.0	NaN	30
6	7782	CLARK	MANAGER	7839.0	1981-06-09	2450.0	NaN	10
7	7839	KING	PRESIDENT	NaN	1981-07-11	5000.0	NaN	10
8	7844	TURNER	SALESMAN	7698.0	1981-09-08	1500.0	0.0	30
9	7900	JAMES	CLERK	7698.0	1981-12-03	950.0	NaN	30
10	7902	FORD	ANALYST	7566.0	1981-12-03	3000.0	NaN	20
11	7934	MILLER	CLERK	7782.0	1982-01-23	1300.0	NaN	10

dept

Out :

	DEPTNO	DNAME	LOC
0	10	ACCOUNTING	NEW YORK
1	20	RESEARCH	DALLAS
2	30	SALES	CHICAGO
3	40	OPERATIONS	BOSTON

- 데이터베이스 불러오기

026

- > 데이터베이스(SQLAlchemy) 불러오기

```
# 설치
```

```
pip install sqlalchemy==1.4.16
```

```
import sqlalchemy as sql
```

```
engine = sql.create_engine('oracle://scott:tiger@localhost:1521/xe')
```

```
conn = engine.connect()
```

```
book_list = pd.read_sql('select * from book_list', conn)
```

```
book_store = pd.read_sql('select * from book_store', conn)
```

```
conn.close()
```

- 데이터베이스 불러오기

- > 데이터베이스(SQLAlchemy) 불러오기

```
# 데이터 연결하기
```

```
book = book_list.merge(book_store, left_on = 'publisher', right_on = 'bscode')
```

- > 저장하기

```
# engine을 사용하여 저장
```

```
book.to_sql('book', engine, index=False, if_exists = 'replace')
```

fail	기존 테이블이 있으면 실패(기본값)
replace	기존 테이블이 있으면 교체
append	기존 테이블이 있으면 추가