■ 행(row) / 열(column) 추가하기

● 기본 데이터

```
product_list = [
    {'name':'MOUSE', 'price':100, 'company':'A'},
    {'name':'SSD', 'price':200, 'company':'B'},
    {'name':'CPU', 'price':300, 'company':'C'}
]
df = pd.DataFrame(product_list, columns=['name', 'price', 'company'])
df
```

	name	price	company
0	MOUSE	100	A
1	SSD	200	B
2	CPU	300	C

● 단일 값

```
df['expires'] = 5
df
```

	name	price	company	expires
0	MOUSE	100	A	5
1	SSD	200	B	5
2	CPU	300	C	5

■ 열(column) 추가하기

● Series

```
series = pd.Series([10, 2, 2], index=[0, 1, 2])
df['count'] = series
df
```

	name	price	company	expires	count
0	MOUSE	100	A	5	10
1	SSD	200	B	5	2
2	CPU	300	C	5	2

● 연산 결과

```
df['total'] = df['price'] * df['count']
df
```

	name	price	company	count	expires	total
0	MOUSE	100	A	10	5	1000
1	SSD	200	B	2	5	400
2	CPU	300	C	2	5	600

■ 행(row) / 열(column) 추가하기

● List

```
discount = []

for row in df['total']:
    if row >= 1000:
        discount.append('10%')
    elif row >= 500:
        discount.append('5%')
    else:
        discount.append('0%')

df['discount'] = discount
df
```

	name	price	company	count	expires	total	discount
0	MOUSE	100	A	10	5	1000	10%
1	SSD	200	B	2	5	400	0%
2	CPU	300	C	2	5	600	5%

■ 사용자 함수 (apply)

● 무선 지원 여부 추가

```
def is_wireless(row):
    if(row == 'MOUSE'):
        return True
    else:
        return False

df['wireless'] = df.name.apply(is_wireless)
df
```

	name	price	company	count	expires	total	discount	wireless
0	MOUSE	100	A	10	5	1000	10%	True
1	SSD	200	B	2	5	400	0%	False
2	CPU	300	C	2	5	600	5%	False

■ 사용자 함수 (apply)

● 생산일자 추가

```
create_date = ['2004-08-09', '2009-11-19', '2012-03-25']
df['create_date'] = create_date
df
```

	name	price	company	count	expires	total	discount	wireless	create_date
0	MOUSE	100	A	10	5	1000	10%	True	2004-08-09
1	SSD	200	B	2	5	400	0%	False	2009-11-19
2	CPU	300	C	2	5	600	5%	False	2012-03-25

```
def extract_year(row):
    return row.split('-')[0]

df['year'] = df['create_date'].apply(extract_year)
df
```

	name	price	company	count	expires	total	discount	wireless	create_date	year
0	MOUSE	100	A	10	5	1000	10%	True	2004-08-09	2004
1	SSD	200	B	2	5	400	0%	False	2009-11-19	2009
2	CPU	300	C	2	5	600	5%	False	2012-03-25	2012

■ 열(column) 수정하기

● 단일 값

```
df['expires'] = 7
df
```

	name	price	company	expires	count	total	discount	wireless	create_date	year
0	MOUSE	100	A	7	10	1000	10%	True	2004-08-09	2004
1	SSD	200	B	7	2	400	0%	False	2009-11-19	2009
2	CPU	300	C	7	2	600	5%	False	2012-03-25	2012

● Numpy

```
import numpy as np
```

```
df['price'] = np.where(df['year'] < '2010', df['price'] * 0.95, df['price'])
```

```
df
```

	name	price	company	expires	count	total	discount	wireless	create_date	year
0	MOUSE	95.0	A	7	10	1000	10%	True	2004-08-09	2004
1	SSD	190.0	B	7	2	400	0%	False	2009-11-19	2009
2	CPU	300.0	C	7	2	600	5%	False	2012-03-25	2012

■ 행(row) / 열(column) 삭제하기

● drop

– index 삭제

```
df.drop([0])
```

drop 함수의 기본 속성 – index (axis=0)

	name	price	company	expires	count	total	discount	wireless	create_date	year
1	SSD	190.0	B	7	2	400	0%	False	2009-11-19	2009
2	CPU	300.0	C	7	2	600	5%	False	2012-03-25	2012

```
df.drop([1])
```

drop 함수의 기본 속성 – index (axis=0)

	name	price	company	expires	count	total	discount	wireless	create_date	year
0	MOUSE	95.0	A	7	10	1000	10%	True	2004-08-09	2004
2	CPU	300.0	C	7	2	600	5%	False	2012-03-25	2012

```
df.drop([0, 2], axis=0)
```

	name	price	company	expires	count	total	discount	wireless	create_date	year
1	SSD	190.0	B	7	2	400	0%	False	2009-11-19	2009

■ 행(row) / 열(column) 삭제하기

● drop

– column 삭제

```
df.drop(['year'])   기본 속성이 axis=0 이므로 오류
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-133-3c29a5b1e49a> in <module>()
----> 1 df.drop(['year'])
```

```
df.drop(['year', 'discount'], axis=1)
```

	name	price	company	expires	count	total	wireless	create_date
0	MOUSE	95.0	A	7	10	1000	True	2004-08-09
1	SSD	190.0	B	7	2	400	False	2009-11-19

– 삭제 결과 저장

```
df = df.drop(['year', 'discount'], axis=1)
```

```
df.drop(['year', 'discount'], axis=1, inplace=True)
```

■ 행(row) / 열(column) 삭제하기

● 조건에 해당하는 데이터 조회 (데이터 삭제)

```
df[df['wireless'] == False]
```

	name	price	company	expires	count	total	wireless	create_date
1	SSD	190.0	B	7	2	400	False	2009-11-19
2	CPU	300.0	C	7	2	600	False	2012-03-25

```
df[df['total'] >= 500]
```

	name	price	company	expires	count	total	wireless	create_date
0	MOUSE	95.0	A	7	10	1000	True	2004-08-09
2	CPU	300.0	C	7	2	600	False	2012-03-25

■ 행(row) / 열(column) 삭제하기

● 행 데이터 삭제 후 index 정리

```
df[df['total'] >= 500].reset_index()
```

	index	name	price	company	expi
0	0	MOUSE	95.0	A	
1	2	CPU	300.0	C	

```
df[df['total'] >= 500].reset_index(drop=True)
```

	name	price	company	expires	count	total
0	MOUSE	95.0	A	7	10	1000
1	CPU	300.0	C	7	2	600

● 인덱스 직접 지정

```
df2 = df[df['total'] >= 500]
df2.index = ['1번', '2번']
df2
```

	name	price	company	expires	count	total	wireless	create_date
1번	MOUSE	95.0	A	7	10	1000	True	2004-08-09
2번	CPU	300.0	C	7	2	600	False	2012-03-25

■ 행(row) / 열(column) 삭제하기

● 중복 데이터 삭제

```
member = pd.read_csv('data/member.txt')
member
```

	name	class	gender	age
0	John	Mathematics	male	18
1	Liam	E-Business	female	19
2	Noah	Mathematics	female	20
3	Logan	Computer	male	20
4	Lucas	E-Business	male	21
5	Mason	E-Business	male	21
6	Oliver	E-Business	female	21
7	Ethan	Computer	male	22
8	Elijah	Computer	female	22
9	Aiden	Accounting	female	22
10	James	Computer	male	22
11	Sophia	Finance	male	23
12	Emma	Accounting	male	23
13	Emily	Finance	male	23
14	Grace	Computer	male	23
15	Hailey	Finance	male	23

```
member.drop_duplicates(['class'])
```

	name	class	gender	age
0	John	Mathematics	male	18
1	Liam	E-Business	female	19
3	Logan	Computer	male	20
9	Aiden	Accounting	female	22
11	Sophia	Finance	male	23

```
member.drop_duplicates(['class'], keep='last')
```

	name	class	gender	age
2	Noah	Mathematics	female	20
6	Oliver	E-Business	female	21
12	Emma	Accounting	male	23
14	Grace	Computer	male	23
15	Hailey	Finance	male	23

```
member.drop_duplicates(['class'], keep=False)
```

	name	class	gender	age