■ 데이터 그룹 만들기

● 기본 데이터

```
df = pd.read_csv('data/member.txt')
df.head(8)
```

	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

```
df.tail(8)
```

	name	class	gender	age
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

■ 데이터 그룹 만들기

● 성별(gender) 그룹

```
group_gender = df.groupby('gender')
group_gender.groups  # 그룹의 정보 확인
```

```
{'female': Int64Index([1, 2, 6, 8, 9], dtype='int64'),
 'male': Int64Index([0, 3, 4, 5, 7, 10, 11, 12, 13, 14, 15], dtype='int64')}
```

● 그룹 길이 확인

```
len(group_gender)
```

```
2
```

● 그룹 크기(현황) 확인

```
group_gender.size()
```

```
gender
female     5
male      11
dtype: int64
```

● 그룹 결과 데이터프레임 생성

```
gender_cnt = pd.DataFrame({'count' : group_gender.size()}).reset_index()
gender_cnt
```

	gender	count
0	female	5
1	male	11

■ 데이터 그룹 만들기

● 세부 내용 확인

```
for name, group in group_gender:
    print(name, ':', str(len(group)))
    print(group)
    print()
```

female : 5

	name	class	gender	age
1	Liam	E-Business	female	19
2	Noah	Mathematics	female	20
6	Oliver	E-Business	female	21
8	Elijah	Computer	female	22
9	Aiden	Accounting	female	22

male : 11

	name	class	gender	age
0	John	Mathematics	male	18
3	Logan	Computer	male	20
4	Lucas	E-Business	male	21
5	Mason	E-Business	male	21
7	Ethan	Computer	male	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

■ 데이터 그룹 만들기

● 학과(class) 그룹

```
group_class = df.groupby('class')
group_class.groups
```

```
{'Accounting': Int64Index([9, 12], dtype='int64'),
 'Computer': Int64Index([3, 7, 8, 10, 14], dtype='int64'),
 'E-Business': Int64Index([1, 4, 5, 6], dtype='int64'),
 'Finance': Int64Index([11, 13, 15], dtype='int64'),
 'Mathematics': Int64Index([0, 2], dtype='int64')}
```

● 그룹 길이 확인

```
len(group_class)
```

```
5
```

● 그룹 크기(현황) 확인

```
group_class.size()
```

```
class
Accounting     2
Computer       5
E-Business     4
Finance        3
Mathematics    2
dtype: int64
```

■ 데이터 그룹 만들기

● 세부 내용 확인

```
for key, group in group_class:
    print(key, ':', str(len(group)))
```

```
Accounting : 2
Computer : 5
E-Business : 4
Finance : 3
Mathematics : 2
```

● 그룹 결과 데이터프레임 생성

```
class_cnt = pd.DataFrame({'count' : group_class.size()}).reset_index()
class_cnt
```

	class	count
0	Accounting	2
1	Computer	5
2	E-Business	4
3	Finance	3
4	Mathematics	2

■ 데이터 그룹 만들기

● 나이(age) 그룹

```
group_age = df.groupby('age')
group_age.groups
```

```
{18: Int64Index([0], dtype='int64'),
 19: Int64Index([1], dtype='int64'),
 20: Int64Index([2, 3], dtype='int64'),
 21: Int64Index([4, 5, 6], dtype='int64'),
 22: Int64Index([7, 8, 9, 10], dtype='int64'),
 23: Int64Index([11, 12, 13, 14, 15], dtype='int64')}
```

● 세부 내용 확인

```
for key, group in group_age:
    print(key, ':', str(len(group)))
```

```
18 : 1
19 : 1
20 : 2
21 : 3
22 : 4
23 : 5
```

■ 데이터 그룹 만들기

● 2개 이상의 컬럼을 이용하여 그룹화

```
group = df.groupby(['class', 'gender'])
group.size()
```

```
class        gender
Accounting   female    1
             male      1
Computer     female    1
             male      4
E-Business   female    2
             male      2
Finance      male      3
Mathematics  female    1
             male      1
dtype: int64
```

● 데이터프레임 생성

```
pd.DataFrame({'count' : group.size()}).reset_index()
```

	class	gender	count
0	Accounting	female	1
1	Accounting	male	1
2	Computer	female	1
3	Computer	male	4
4	E-Business	female	2
5	E-Business	male	2
6	Finance	male	3
7	Mathematics	female	1
8	Mathematics	male	1