dataframe groupby

pandas_data3 ➔

	name	gender	year	mid	final	att	proj
0	kim2	F	1	12	36	5	75
1	kim3	M	2	17	20	5	96
2	kim6	M	2	28	20	4	83
3	kim7	M	2	20	44	6	82
4	kim8	F	1	20	24	8	88
5	kim9	M	3	23	45	10	92
6	kim10	M	3	16	47	5	83
7	lee5	M	3	21	35	5	92
8	lee9	M	3	20	36	5	87
9	lee10	F	4	28	41	9	87
10	park1	M	4	30	25	7	86
11	park2	M	4	20	31	7	76
12	park6	F	1	24	37	7	91
13	kim1	F	2	19	18	8	88
14	park10	M	3	23	20	6	80

groupby and select a column

```
d1=pd.read_table('pandas_data3.txt', index_col=0)
d1=d1.head(10)

for key, gp in d1.groupby('gender'):
    print(key)
    gp
    gp['mid']
```

	name	gender	year	mid	final	att	proj
0	kim2	F	1	12	36	5	75
4	kim8	F	1	20	24	8	88
9	lee10	F	4	28	41	9	87

	name	gender	year	mid	final	att	proj
1	kim3	M	2	17	20	5	96
2	kim6	M	2	28	20	4	83
3	kim7	M	2	20	44	6	82
5	kim9	M	3	23	45	10	92
6	kim10	M	3	16	47	5	83
7	lee5	M	3	21	35	5	92
8	lee9	M	3	20	36	5	87

```
for key, col in d1.groupby('gender')['mid']:
    print(key)
    col
```

```
F

11]: name
     kim2     12
     kim8     20
     lee10    28
     Name: mid, dtype: int64

M

11]: name
     kim3     17
     kim6     28
     kim7     20
     kim9     23
     kim10    16
     lee5     21
     lee9     20
     Name: mid, dtype: int64
```

groupby and operation function

```
d1=pd.read_table('pandas_data3.txt', index_col=0)
d1=d1.head(10)

for key, col in d1.groupby('gender')['mid']:
    print(key)
    col
```

```
d1.groupby('gender')['mid'].sum()
```

```
          F

Out[17]:  name
          kim2     12
          kim8     20
          lee10    28
          Name: mid, dtype: int64

          M

Out[17]:  name
          kim3     17
          kim6     28
          kim7     20
          kim9     23
          kim10    16
          lee5     21
          lee9     20
          Name: mid, dtype: int64

Out[17]:  gender
          F      60
          M     145
          Name: mid, dtype: int64
```

groupby with two columns

```
d1=pd.read_table('pandas_data3.txt', index_col=0)
d1=d1.head(10)

for key, gp in d1.groupby(['gender', 'year']):
    print('# key:', key)
    gp
```

```
# key: ('F', 1)
```

[8]:

name	gender	year	mid	final	att	proj
kim2	F	1	12	36	5	75
kim8	F	1	20	24	8	88

```
# key: ('F', 4)
```

[8]:

name	gender	year	mid	final	att	proj
lee10	F	4	28	41	9	87

```
# key: ('M', 2)
```

[8]:

name	gender	year	mid	final	att	proj
kim3	M	2	17	20	5	96
kim6	M	2	28	20	4	83
kim7	M	2	20	44	6	82

```
# key: ('M', 3)
```

[8]:

name	gender	year	mid	final	att	proj
kim9	M	3	23	45	10	92
kim10	M	3	16	47	5	83
lee5	M	3	21	35	5	92
lee9	M	3	20	36	5	87

groupby and operation function

```
d1=pd.read_table('pandas_data3.txt', index_col=0)
d1=d1.head(10)

for key, col in d1.groupby(['gender','year'])['mid']:
    print(key)
    col
```

```
('F', 1)

[13]: name
kim2    12
kim8    20
Name: mid, dtype: int64

('F', 4)

[13]: name
lee10    28
Name: mid, dtype: int64

('M', 2)

[13]: name
kim3    17
kim6    28
kim7    20
Name: mid, dtype: int64

('M', 3)

[13]: name
kim9     23
kim10    16
lee5     21
lee9     20
Name: mid, dtype: int64
```

```
d1.groupby(['gender','year'])['mid'].sum()
```

```
gender  year
F       1       32
        4       28
M       2       65
        3       80
Name: mid, dtype: int64
```