



AniPython

分组

groupby()

groupby()

1. 拆分成多个组

	BY	

groupby()

1. 拆分成多个组

	BY	

	BY	

groupby()

1. 拆分成多个组
2. 分别对每个组应用函数

	BY	

	BY	

groupby()

1. 拆分成多个组
2. 分别对每个组应用函数
3. 合并

	BY	

	BY	

groupby()

1. 拆分成多个组
2. 分别对每个组应用函数
3. 合并

	BY	


```
df = pd.DataFrame({  
    '货号': ['A', 'B', 'B', 'A', 'B'],  
    '平台': ['淘宝', '淘宝', '京东',  
            '京东', '淘宝'],  
    '销量': [1, 2, 3, 4, 5]  
})
```

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

```
df.groupby('货号')
```


	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

df.groupby('货号')

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

df.groupby('货号')

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

```
df.groupby('平台')
```


	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
4	B	淘宝	5

df.groupby('平台')

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

df.groupby('平台')

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
4	B	淘宝	5

	货号	平台	销量
2	B	京东	3
3	A	京东	4

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

```
df.groupby(['货号','平台'])
```


	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1

df.groupby(['货号','平台'])

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1

	货号	平台	销量
3	A	京东	4

df.groupby(['货号','平台'])

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1

	货号	平台	销量
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
4	B	淘宝	5

df.groupby(['货号','平台'])

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

df.groupby(['货号','平台'])

	货号	平台	销量
0	A	淘宝	1

	货号	平台	销量
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
4	B	淘宝	5

	货号	平台	销量
2	B	京东	3

	货号	平台	销量
0	A	淘宝	1
1	B	淘宝	2
2	B	京东	3
3	A	京东	4
4	B	淘宝	5

df.groupby('货号')

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

```
df.groupby('货号')
[('A', df_A),
 ('B', df_B)]
<DataFrameGroupBy>
```

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

```
df.groupby('货号')['销量']
[('A', s_A),
 ('B', s_B)]
<SeriesGroupBy>
```


	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

分组之后通常
可以做三件事情

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

1: 聚合(aggregate)

```
df.groupby('货号').sum()
```

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

	销量
货号	
A	5
B	10

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

2: 转换(transform)

```
df.groupby('货号')['销量']
    .transform('sum')
```

0	5
1	10
2	10
3	5
4	10
Name: 销量, dtype: int64	

	货号	平台	销量
0	A	淘宝	1
3	A	京东	4

3: 过滤(filter)

```
df.groupby('货号').filter(
    lambda x: x['销量'].sum() > 6)
```

	货号	平台	销量
1	B	淘宝	2
2	B	京东	3
4	B	淘宝	5

	平台	销量
1	淘宝	2
2	京东	5
4	淘宝	10

Function	Description
mean()	Compute mean of groups
sum()	Compute sum of group values
size()	Compute group sizes
count()	Compute count of group
std()	Standard deviation of groups
var()	Compute variance of groups
sem()	Standard error of the mean of groups
describe()	Generates descriptive statistics
first()	Compute first of group values
last()	Compute last of group values
nth()	Take nth value, or a subset if n is a list
min()	Compute min of group values
max()	Compute max of group values

	货号	平台	销量
0	A	淘宝	1
4	B	淘宝	5

	货号	平台	销量
0	A	淘宝	1
2	B	京东	3
4	B	淘宝	5

groupby() 分组

df.groupby('货号')