

pandas中的数据分组与透视表

```
In [1]: import pandas as pd
```

```
In [2]: import numpy as np
```

数据分组----groupby

```
In [3]: df = pd.read_excel('data.xlsx')  
df
```

```
Out[3]:
```

	日期	名称	类别	单价	数量	金额
0	2018-07-01	商品A	服装	20	2	40
1	2018-07-02	商品B	服装	200	3	600
2	2018-07-03	商品C	食品	1200	4	4800
3	2018-07-04	商品A	服装	22	5	110
4	2018-07-05	商品B	服装	220	6	1320
5	2018-07-06	商品C	食品	1000	7	7000
6	2018-07-07	商品A	服装	30	3	90
7	2018-07-08	商品A	服装	800	1	800
8	2018-07-09	商品C	食品	1300	4	5200
9	2018-07-10	商品B	服装	230	3	690
10	2018-07-11	商品A	服装	28	1	28

```
In [6]: grouped = df.groupby('类别')  
grouped
```

```
Out[6]: <pandas.core.groupby.groupby.DataFrameGroupBy object at 0x116c14f60>
```

```
In [5]: grouped.sum()
```

Out[5]:

	单价	数量	金额
类别			
服装	1550	24	3678
食品	3500	15	17000

```
In [7]: grouped[['数量', '金额']].sum()
```

Out[7]:

	数量	金额
类别		
服装	24	3678
食品	15	17000

```
In [9]: grouped[['单价']].max()
```

Out[9]:

	单价
类别	
服装	800
食品	1300

```
In [10]: grouped = df.groupby(['类别', '名称'])
```

```
In [11]: grouped[['单价']].max()
```

Out[11]:

		单价
类别	名称	
服装	商品A	800
	商品B	230
食品	商品C	1300

```
In [12]: grouped[['单价']].mean()
```

Out[12]:

		单价
类别	名称	
服装	商品A	180.000000
	商品B	216.666667
食品	商品C	1166.666667

```
In [14]: for name, data in df.groupby('类别'):
          print(name)
          print(data)
```

服装

	日期	名称	类别	单价	数量	金额
0	2018-07-01	商品A	服装	20	2	40
1	2018-07-02	商品B	服装	200	3	600
3	2018-07-04	商品A	服装	22	5	110
4	2018-07-05	商品B	服装	220	6	1320
6	2018-07-07	商品A	服装	30	3	90
7	2018-07-08	商品A	服装	800	1	800
9	2018-07-10	商品B	服装	230	3	690
10	2018-07-11	商品A	服装	28	1	28

食品

	日期	名称	类别	单价	数量	金额
2	2018-07-03	商品C	食品	1200	4	4800
5	2018-07-06	商品C	食品	1000	7	7000
8	2018-07-09	商品C	食品	1300	4	5200

```
In [15]: for name, data in df.groupby('类别'):
          print(name)
          print(data['名称'].unique())
```

服装

['商品A' '商品B']

食品

['商品C']

透视表----pivot_table

```
In [16]: df
```

```
Out[16]:
```

	日期	名称	类别	单价	数量	金额
0	2018-07-01	商品A	服装	20	2	40
1	2018-07-02	商品B	服装	200	3	600
2	2018-07-03	商品C	食品	1200	4	4800
3	2018-07-04	商品A	服装	22	5	110
4	2018-07-05	商品B	服装	220	6	1320
5	2018-07-06	商品C	食品	1000	7	7000
6	2018-07-07	商品A	服装	30	3	90
7	2018-07-08	商品A	服装	800	1	800
8	2018-07-09	商品C	食品	1300	4	5200
9	2018-07-10	商品B	服装	230	3	690
10	2018-07-11	商品A	服装	28	1	28

```
In [17]: df.pivot_table(index=['类别', '名称'], values=['单价', '数量'])
```

```
Out[17]:
```

		单价	数量
类别	名称		
服装	商品A	180.000000	2.4
	商品B	216.666667	4.0
食品	商品C	1166.666667	5.0

```
In [19]: df.pivot_table(index=['类别', '名称'], values=['数量', '金额'], aggfun  
c=np.sum)      # 默认是np.mean 平均
```

```
Out[19]:
```

		数量	金额
类别	名称		
服装	商品A	12	1068
	商品B	12	2610
食品	商品C	15	17000