

06多层索引

为什么要学习多层索引MultiIndex?

- 多层索引：在一个轴向上拥有多个索引层级，可以表达更高维度数据的形式；
- 可以更方便的进行数据筛选，如果有序则性能更好；
- `groupby`等操作的结果，如果是多KEY，结果是多层索引，需要会使用
- 一般不需要自己创建分层索引(MultiIndex有构造函数但一般不用)

演示数据：百度、阿里巴巴、爱奇艺、京东四家公司的10天股票数据

数据来自：英为财经

<https://cn.investing.com/>

本节主要内容：

- 一、Series的多层索引MultiIndex
- 二、Series有多层索引怎样筛选数据？
- 三、DataFrame的多层索引MultiIndex
- 四、DataFrame有多层索引怎样筛选数据？

0.读取数据

```
1 stocks = pd.read_excel('res/互联网公司股票.xlsx')
2 stocks.shape #查看形状
3 stocks["公司"].unique() # 去重统计
4 stocks.groupby('公司')['收盘'].mean() # 分组聚合
```

1.Series的多层索引

```
1 ser = stocks.groupby(['公司', '日期'])['收盘'].mean()
2 ser #多层索引中，空白的意思是：使用上面的值
3 ser.index
4 ser.reset_index() #重置索引
```

2.Series多层索引筛选数据

```
1 ser
2 ser.loc['BIDU']
3 # 多层索引，可以用元组的形式筛选
4 ser.loc[('BIDU', '2019-10-02')]
5 ser.loc[:, '2019-10-02']
```

3.DataFrame的多层索引

```
1 stocks.head()
2 stocks.set_index(['公司', '日期'], inplace=True)
3 stocks.index
4 stocks.sort_index(inplace=True) #排序
5 stocks
```

4.DataFrame多层索引筛选数据

【重要知识】

在选择数据时：

- 元组(key1,key2)代表筛选多层索引，其中key1是索引第一级，key2是第二级，比如key1=JD, key2=2019-10-02
- 列表[key1,key2]代表同一层的多个KEY，其中key1和key2是并列的同级索引，比如key1=JD, key2=BIDU

```
1 stocks.loc['BIDU']
2 stocks.loc(['BIDU', '2019-10-02'], :)
3 stocks.loc(['BIDU', '2019-10-02'], '开盘']
4 stocks.loc(['BIDU', 'JD'], :)
5 stocks.loc(['BIDU', 'JD'], '2019-10-03'), :)
6 stocks.loc(['BIDU', 'JD'], '2019-10-03'), '收盘']
7 stocks.loc(['BIDU', ['2019-10-02', '2019-10-03']], '收盘']
8 # slice(None)代表筛选这一索引的所有内容
9 stocks.loc[slice(None), ['2019-10-02', '2019-10-03']], :]
```