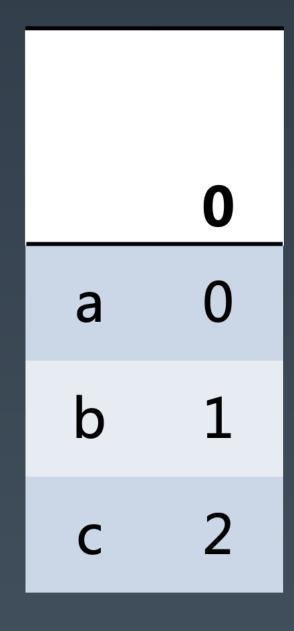
## Pandas 的基本数据结构

Series 数据结构详解
DataFrame 数据结构详解



## Series 数据结构

两个基本属性: index 和 value





# DataFrame 数据结构

	index	0
0	a	0
1	b	1
2	C	2



### 数据的获取和预处理

数据的导入和新建

数据预处理: 缺失、重复、异常值的处理方法

索引设置

高阶函数 map reduce filter



### 数据的导入

```
Pandas 支持大量格式的导入,使用的是 read_*() 的形式如:
import pandas as pd
pd.read_excel(r'1.xlsx')
pd.read_csv(r'c:\file.csv',sep=' ', nrows=10, encoding='utf-8')
pd.read_sql(sql,conn)
```



### 数据预处理

#### 缺失值处理:

- 使用平均值填充
- 前向填充
- 缺失值删除

#### 重复值处理:

• 删除

### 数据操作实战

数据的排序操作 数据行列选择与查找操作 数据行列互换操作 数据透视表

# 数据的导出

导出为 .csv 格式 导出为 excel 格式

# 使用 jieba 进行自动分词

jieba 分词与提取关键词

jieba 调整词典

jieba 词性标注



### 使用 SnowNLP 进行语义情感标注

SnowNLP 进行语义情感标注

SnowNLP 情感词典的定制



#