# 第四阶段 Python数据分析

# Python 数据分析库3

## 网校相关视频

## 明确本次课知识点，明确重点难点

### 【知识点目标】

* Series对象基本操作
* DataFrame对象基本操作

### 【重点】

* Series对象基本操作
* DataFrame对象基本操作

### 【难点】

* Series对象基本操作
* DataFrame对象基本操作
* 数据规整化

## 复习巩固作业讲解

## 本次课程任务讲解

### 【知识点1】Series对象类型

* 1. 时间序列的操作
     1. 创建时间序列（TimeSeries）:索引由时间戳构成
     2. 截取时间段数据：针对较长的时间序列，可以使用年月进行切片
     3. 滞后或超前操作：shift(n) #n阶滞后/超前，滞后或超前由n的符号决定，正数为滞后，负数为超前

### 【知识点2】DataFrame对象类型

1. DataFrame对象简介

DataFrame极大地简化了数据分析过程中一些烦琐操作，它是一个表格型的数据结构, 每一列代表一个变量，而每一行则是一条记录。简单地说，DataFrame是共享同一个index 的Series的集合。

1. DataFrame对象的创建

DataFrame对象的创建方法与Series对象类似，只不过可以同时接受多条一维的列表, 每个列表都会成为单独的一列。在创建DataFrame对象之前，要先创建一个索引。索引是一个DataFrame对象必须有的元素，起到标识的作用。

* 1. 手动创建：pd.DataFrame()
  2. 从文件中读取：pd.read\_csv(‘path\filename’,sep=’’)

1. 查看DataFrame对象
   1. 基本操作：DataFrame名称
   2. 查看行：head(n),tail(n)
   3. 查看列：columns
   4. 查看索引：index
   5. 查看值：values
   6. 查看DataFrame的信息：info()
   7. 查看各变量的描述：descibie()
2. DataFrame对象的索引与切片
   1. 对行进行切片：df[start:end]
   2. 对列进行切片：df.loc[:,startcolumn,endcolumn]
   3. 条件切片：df.loc[条件表达式]
   4. 标签索引与切片：行或列标签，闭区间，loc
   5. 位置索引与切片：行或列索引，左闭右开区间，iloc
3. DataFrame对象的常用操作
   1. 转置：DataFrame.T
   2. 排序与排名：
      1. 排序：DataFrame.sort\_values(by ,ascending,axis=0) #值的排序

DataFrame.sort\_index(axis=0,ascending)#索引的排序

排名：df.rank(axis=0,ascending) #排名

* 1. 增加行
  2. 增加列
  3. 横向合并
  4. 纵向合并

concat

* 1. 删除
     1. drop(indexrange)#按索引进行删除
     2. drop(columnsrange,axis=1)#按列进行删除
  2. 替换：找出要替换的位置，对该位置上的值进行赋值

## 自主学习作业讲解

无

## 课程总结

#### Serires基本操作

#### DataFrame对象基本操作

#### 数据规整化

## 下次自主学习任务布置

作业：1、生成如下图所示的DataFrame，输出转置之后的第三行数据



1. 在上题的基础上加上“g,John,19”一行，将年龄设置为索引，删除第三行后输出结果