



Python数据分析——第8周

【声明】 本视频和幻灯片为炼数成金网络课程的教学资料，所有资料只能在课程内使用，不得在课程以外范围散播，违者将可能被追究法律和经济责任。

课程详情访问炼数成金培训网站

<http://edu.dataguru.cn>

关注炼数成金企业微信

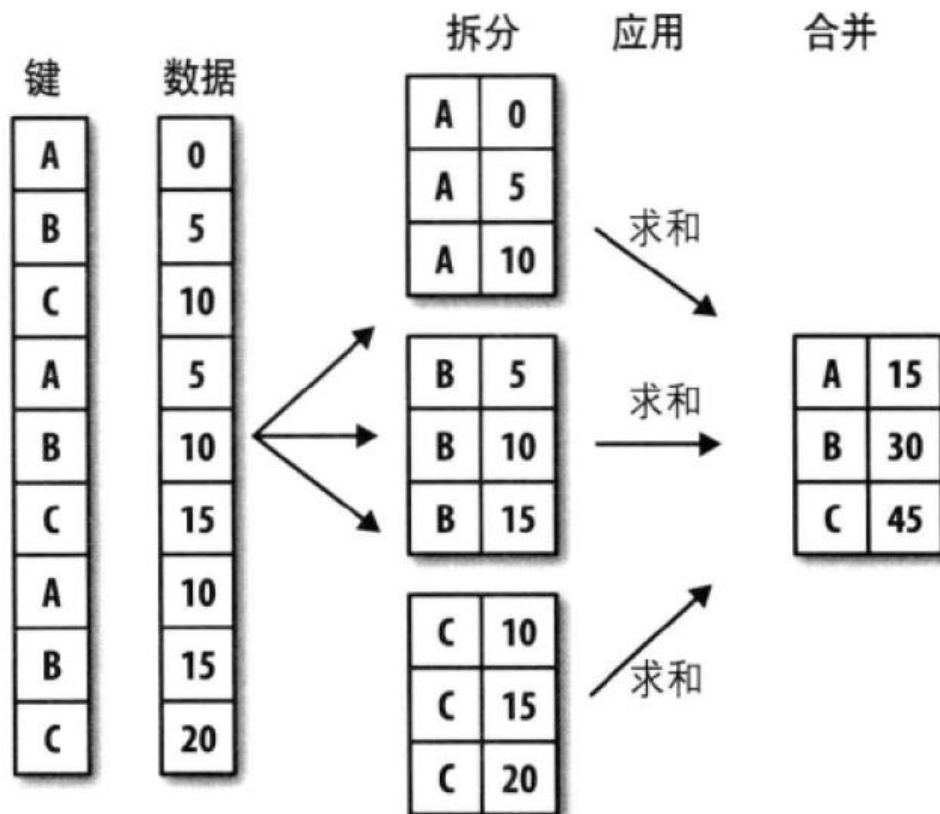


- 提供全面的数据价值资讯，涵盖商业智能与数据分析、大数据、企业信息化、数字化技术等，各种高性价比课程信息，赶紧掏出您的手机关注吧！



- ◆ 根据一个或多个键拆分pandas对象
- ◆ 计算分组摘要统计
- ◆ 对DataFrame的列应用各种各样的函数
- ◆ 应用组内转换或其他运算
- ◆ 计算透视表或交叉表
- ◆ 执行分位数分析以及其他分组分析

◆ Split-apply-combine(拆分-应用-合并)



◆ 分组键

- 列表或数组，其长度与待分组的轴一样
- 表示DataFrame某个列名的值
- 字典或Series，给出待分组轴上的值与分组名之间的对应关系
- 函数，用于处理轴索引或索引中的各个标签

- ◆ 对分组进行迭代
- ◆ 选取一个或一组列
- ◆ 通过字典或Series进行分组
- ◆ 通过函数分组
- ◆ 根据索引级别分组

◆ 聚合——任何能够从数组产生标量值的数据转换过程

函数名	说明
count	分组中非NA值的数量
sum	非NA值的和
mean	非NA值的平均值
median	非NA值的算术中位数
std、var	无偏（分母为n - 1）标准差和方差
min、max	非NA值的最小值和最大值
prod	非NA值的积
first、last	第一个和最后一个非NA值

◆ Transform

- 产生一个标量值
- 产生一个大小相同的数组

◆ Apply——一般性的“拆分-应用-合并”

- 用特定于分组的值填充缺失值
- 随机采样和排列分组加权平均数和相关系数

- ◆ 透视表——根据一个或多个键对数据进行聚合，并根据行和列上的分组键将数据分配到各个矩形区域中
- ◆ 交叉表——一种用于计算分组频率的特殊透视表

参数名	说明
values	待聚合的列的名称。默认聚合所有数值列
rows	用于分组的列名或其他分组键，出现在结果透视表的行
cols	用于分组的列名或其他分组键，出现在结果透视表的列
aggfunc	聚合函数或函数列表，默认为'mean'。可以是任何对groupby有效的函数
fill_value	用于替换结果表中的缺失值
margins	添加行/列小计和总计，默认为False

示例：2012联邦选举委员会数据分析

- ◆ 数据：赞助者姓名、职业、雇主、地址、出资额等
- ◆ 根据职业和雇主统计赞助信息
- ◆ 对出资额分组
- ◆ 根据州统计赞助信息

- ◆ Dataguru (炼数成金) 是专业数据分析网站，提供教育，媒体，内容，社区，出版，数据分析业务等服务。我们的课程采用新兴的互联网教育形式，独创地发展了逆向收费式网络培训课程模式。既继承传统教育重学习氛围，重竞争压力的特点，同时又发挥互联网的威力打破时空限制，把天南地北志同道合的朋友组织在一起交流学习，使到原先孤立的学习个体组合成有组织的探索力量。并且把原先动辄成千上万的学习成本，直线下降至百元范围，造福大众。我们的目标是：低成本传播高价值知识，构架中国第一的网上知识流转阵地。
- ◆ 关于逆向收费式网络的详情，请看我们的培训网站 <http://edu.dataguru.cn>



Thanks

FAQ时间