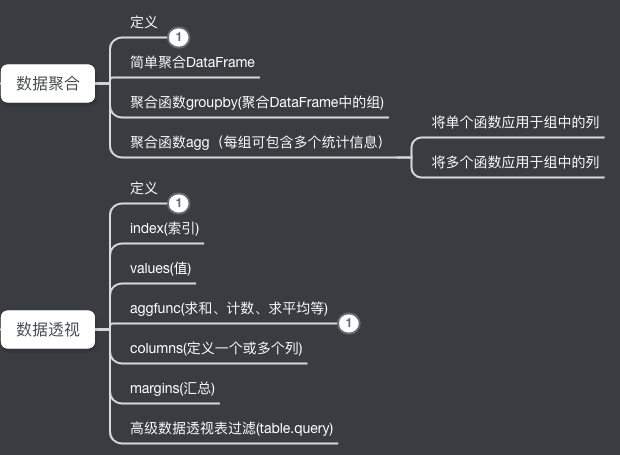
基础|Pandas常用知识点汇总(五)

Pandas 是一套用于 Python 的快速、高效的数据分析工具。本文将详细讲解数据聚合与数据透视，目录如下：



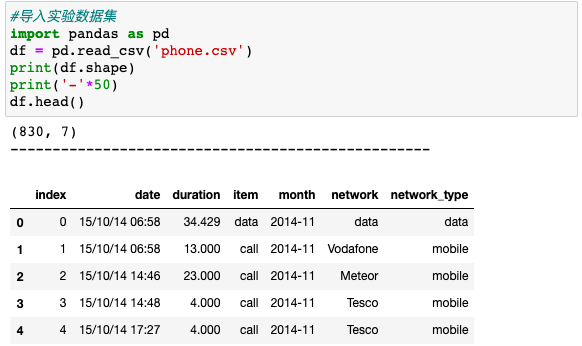
本文数据聚合实验数据集为**phone.csv**，数据透视实验数据集为**sale.xlsx**，后台回复「**datasets**」即可自动获取。

①数据聚合

**一.定义**

通过不同变量对大型数据帧进行分组，以及对每个组应用汇总函数的任务。这时需在pandas中使用DataFrame对象的groupby函数和 agg函数完成。

**二.导入实验数据集**



**三.数据集字段释义**

phone.csv数据集包含830个记录共7个字段，以下对各字段进行简单解释。

**index**：数据集索引；

**date**：输入的日期和时间；

**duration**：每个呼叫的持续时间（以秒为单位），每个数据条目的数据量（以 MB 为单位），以及每个 sms 条目发送的文本数（通常为 1）；

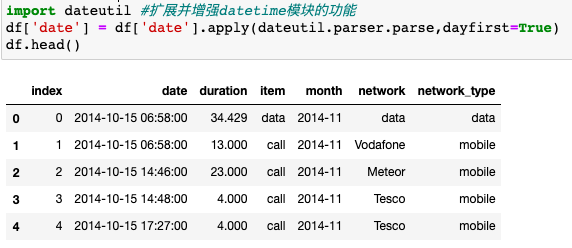
**item**：事件发生的描述——可以是呼叫，短信或者数据之一；

**month**：每个条目所属的结算月份——形式为 “YYYY-MM”；

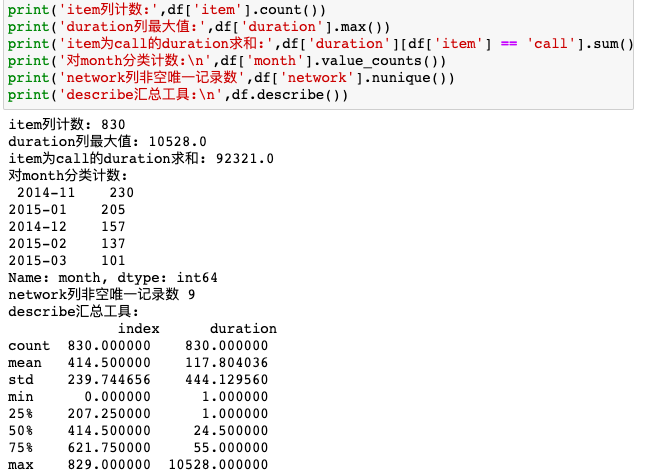
**network**：为每个条目调用 / 发送的移动网络；

**network\_type**：被呼叫的号码是移动，国际，语音邮件，固定电话还是其他特殊号码。

预处理date字段。



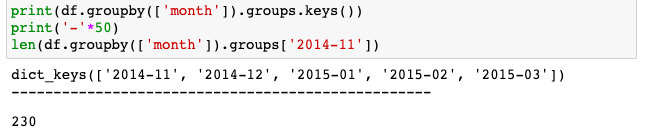
**四.简单聚合DataFrame**



**五.groupby**

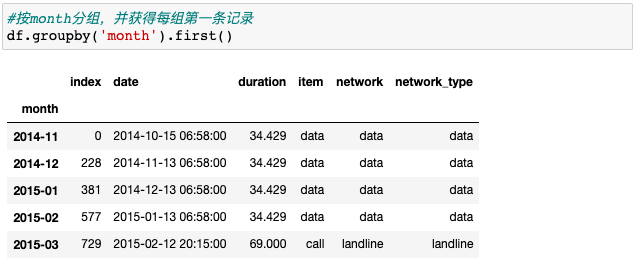
groupby函数将返回groupby对象，但实际上描述了如何拆分原始数据集的行。groupby对象groups变量是一个字典，其键(keys)是计算的唯一组，对应的值是属于每个组的轴标签。

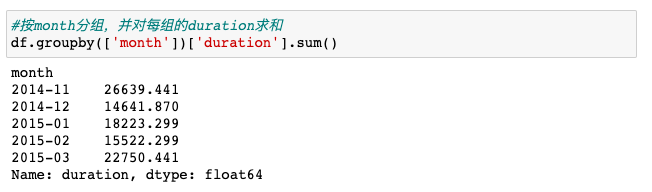
**1.聚合DataFrame中的组**

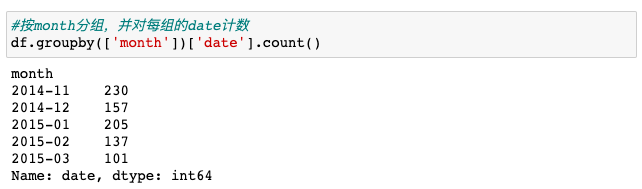


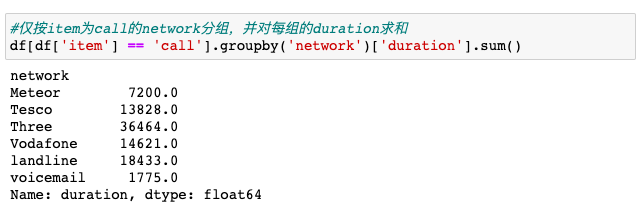
**2.具体示例**

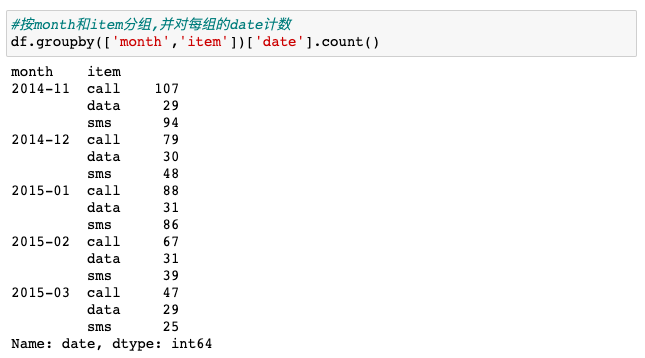
max()，min()，mean()，first()，last()等函数，可以快速应用于groupby对象，以获取每个组的汇总统计信息。groupby是一个非常有用的函数，可以从每个摘要中查看不同的变量。

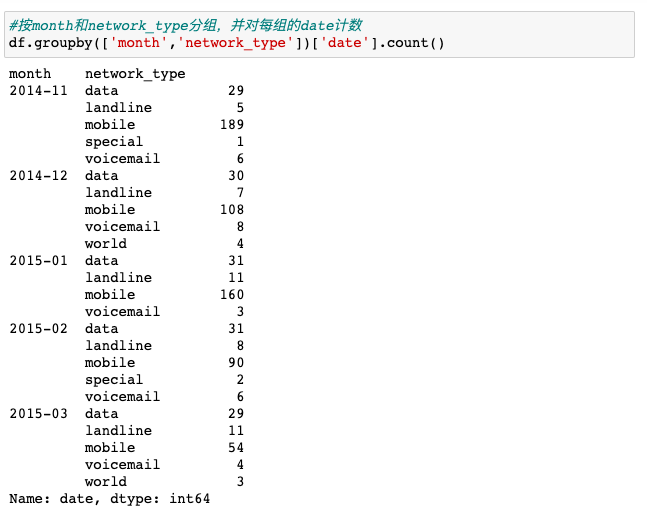












**3.注意要点**

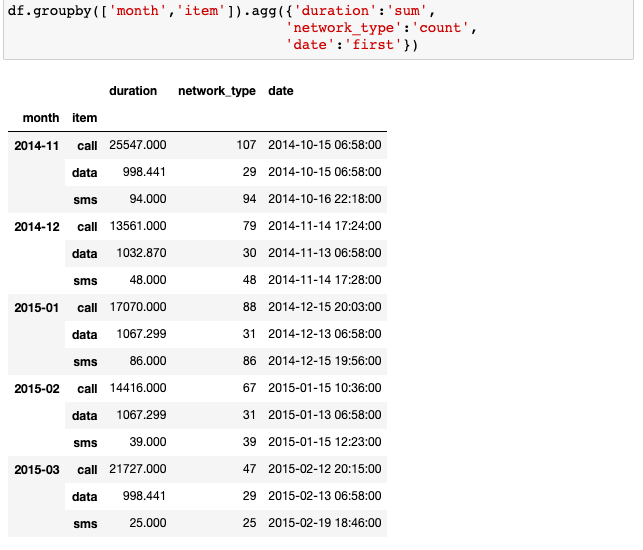




**六.agg**

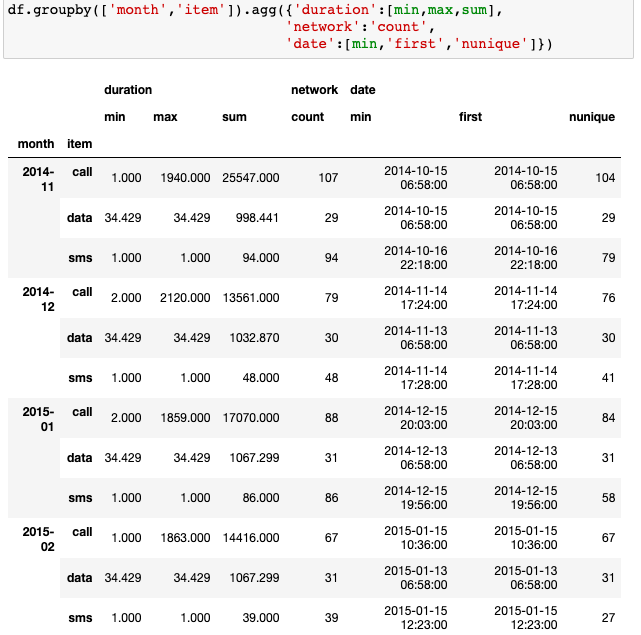
agg函数提供的聚合功能允许在一次计算中按组计算多个统计数据。语法很简单，聚合指令以字典或列表的形式提供。

**1.将单个函数应用于组中的列**





**2.将多个函数应用于组中的列**

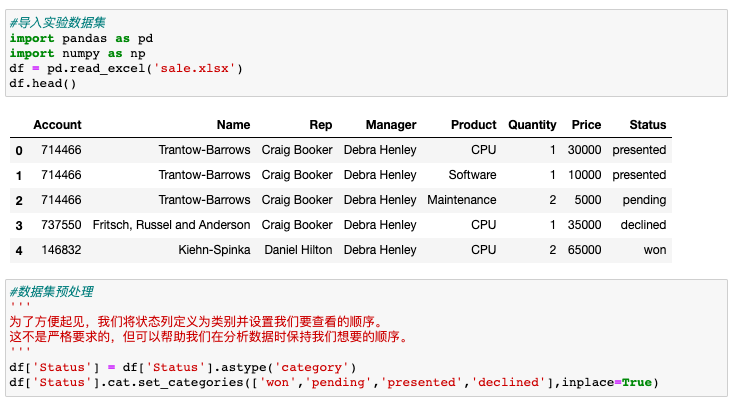


②  
数据透视

**一.定义**

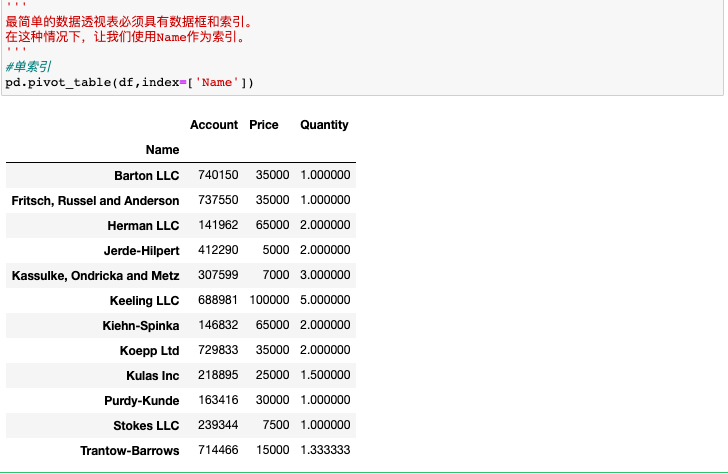
大多数人可能都有使用Excel中的数据透视表的经验。pandas提供了一个类似的功能，称为pivot\_table。

**二.导入实验数据集**

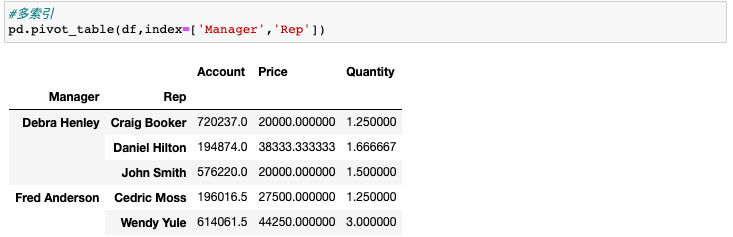


**三.数据透视**

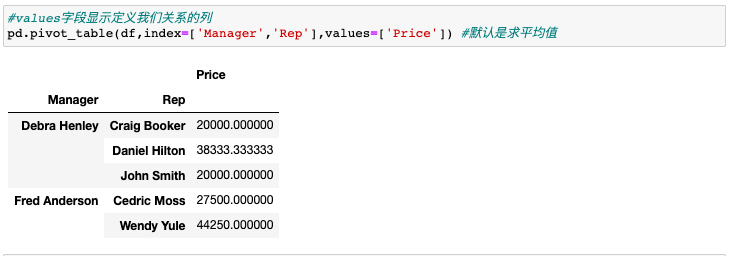
**1.单索引**



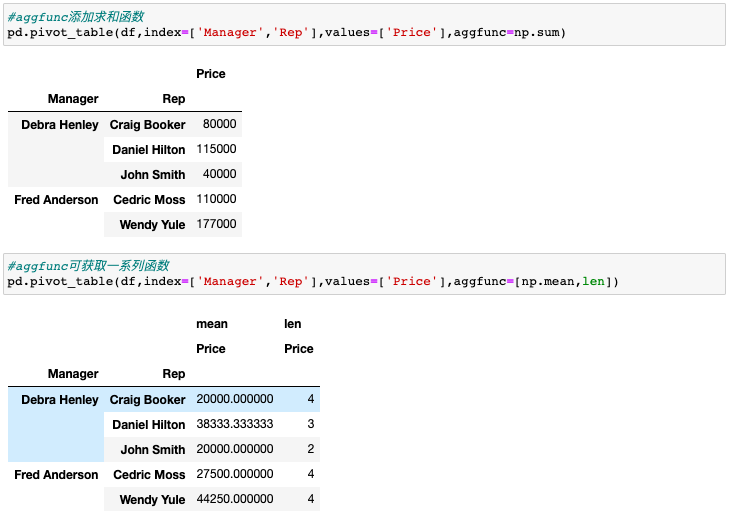
**2.多索引**

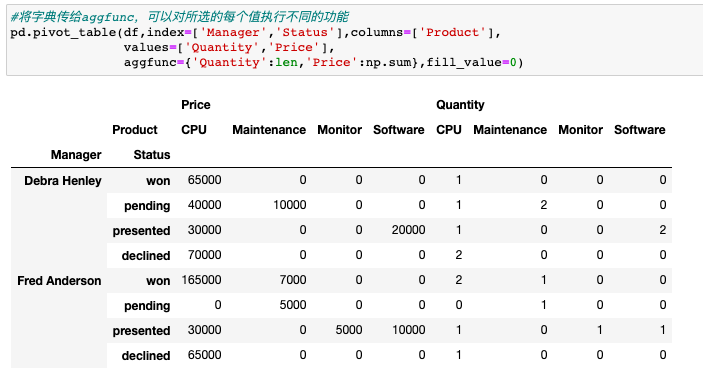


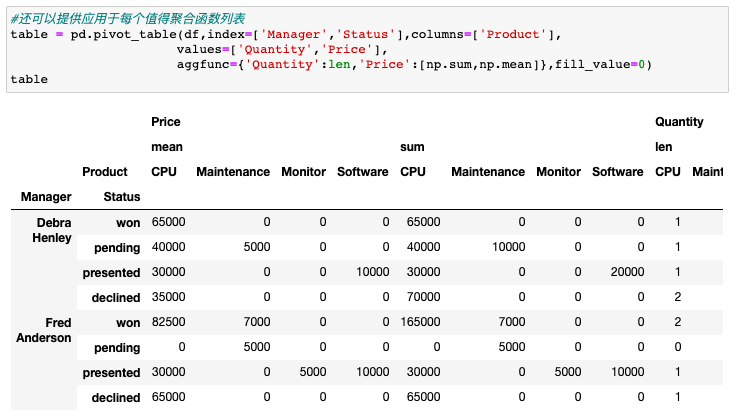
**3.values字段显示定义我们关系的列**



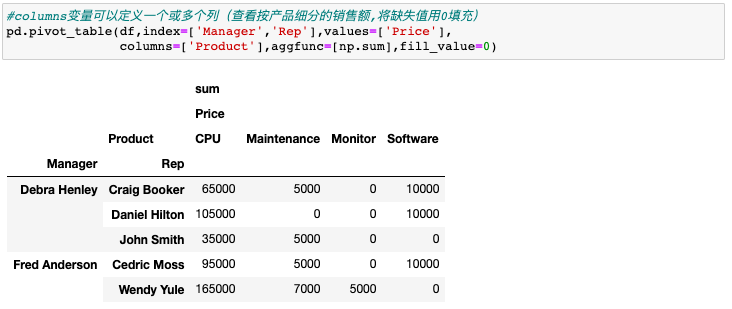
**4.aggfunc可进行数据运算**







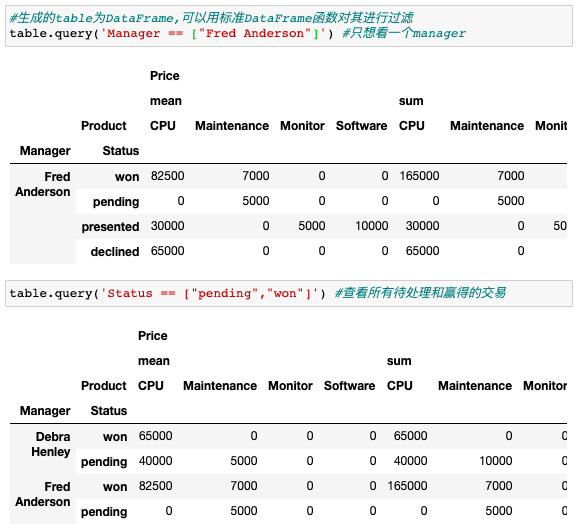
**5.columns变量可以定义一个或多个列**



**6.margins = True可以查看汇总数据**



**7.高级数据透视表过滤**



**结语**OK，第五篇完！至此，pandas基础系列结束。欢迎关注菜J学Python，我们坚持认真写Python基础，幽默写Python实战。你可在公众号后台免费领取相关学习资料或学习交流。

[睡地摊or租房？爬取某大型房产网站24685个房源信息并分析，助你选择](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456405069&idx=1&sn=f74df8b3f4fc9383d3da2da7211dc612&chksm=8d5ad85dba2d514b7b128b5b361b8d496f1bac5498950ffc073cf714151b3917e948854bbe04&scene=21#wechat_redirect)  
[摆地摊or打工，爬取某大型招聘网站3万条招聘信息并分析，作何选择?](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404450&idx=1&sn=4cee4c349aaadfaba69a9ec3919b42e3&chksm=8d5addf2ba2d54e43b6306f68524c9c602a9a9e77041071b792d0bd4ebe96a0061d3c8d7616e&scene=21#wechat_redirect)[实战|手把手教你用Python爬虫(附详细源码)](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456403634&idx=1&sn=41e3e71b16be2a0a25c1eaff063e16af&chksm=8d5aa2a2ba2d2bb4dc8e13bac62d8fef84910535d3c88c1d346de4f30f0995687cef810a3f74&scene=21#wechat_redirect)  
[实战|利用Python获取财务金融数据的几种实用方法](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404246&idx=1&sn=1e3b647baf6b2c208033cc2b5376f9fa&chksm=8d5add06ba2d5410810b39467900035efcd9acf240998ceebbe1456b44631069d72b8bc12b1e&scene=21#wechat_redirect)[实战|Python爬虫并用Flask框架搭建可视化网站](http://mp.weixin.qq.com/s?__biz=MzAwMjYxNjcwMw==&mid=2456404005&idx=1&sn=381e740a8169ff5ac31a5df587383cb0&chksm=8d5adc35ba2d5523121f17d7ddb7bc06ae262548f0b8ca9e6020928fdc78e5bfdee9a40f5edc&scene=21#wechat_redirect)