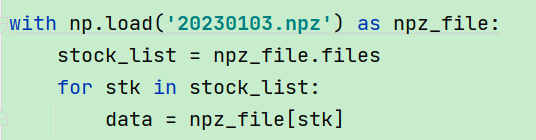
**一、数据：**从链接：https://pan.baidu.com/s/1op4ywi6ke4JhgPNWVFJT3A 提取码：ec18 下载20230103.npz文件，文件里存储的是20230103当天所有股票的逐笔成交数据。可以用下面代码打开：



其中data是一个股票的逐笔成交数据，为np.array格式，共12列：【交易所代码，自然日，时间，成交编号，成交代码，委托代码，BS标志，成交价格，成交数量，叫卖序号，叫买序号，空列】。

**交易所代码**：int类型的股票代码，比如1表示股票‘000001’；

**自然日**：指交易日；

**时间列**：102505000表示10:25:05.000，最后三位是毫秒数；

**成交编号，成交代码，委托代码：**这三列本题用不上；

**BS标志**：列数值为1，-1，nan.；

**成交价格**：本题用不上；

**成交数量**：指每笔成交的股票数量；

**叫卖序号**：卖单编号；

**叫买序号**：买单编号；

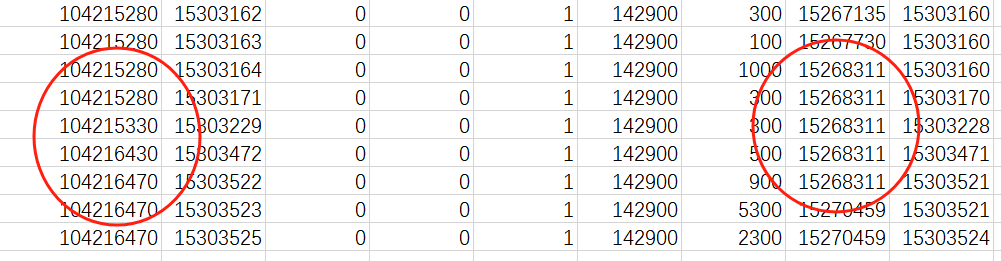
最后一列为空列，全是nan

**二、题目：**对某个股票逐笔成交数据的array，按下面步骤完成计算：

1，去掉BS标志列为nan的行；

2，用时间列去掉九点三十分之前的所有数据；

3，统计当天每个卖单成交的起始和结束时间，计算出成交的时间间隔，精确到秒；



如上图所示，统计出订单号15268311的所有成交记录，起始时间是104215280，即10:42:15.280, 结束时间是104216470（10:42:16.470）所以成交时长为1秒（忽略毫秒数）。有两个特殊情况需要处理：1）如果某卖单成交起始结束时间跨过了中午休盘时间则需要去掉午休时间，从11:30:00到13:00:00, 共5400秒，2）如果某卖单只有一行成交记录，则成交时间为0秒；

4. 将所有卖单的成交时长排序并统计出0.9分位数的值，将大于0.9分位数的卖单标记为长卖单；

5，统计计算出所有长卖单的当天总成交量，并除以当天总成交量（去掉九点三十分之前的所有数据）得到长卖单成交量占比因子；

6，对npz文件中所有股票计算长卖单成交量占比因子；

**三、要求：**计算中除了时间处理部分可以使用其他库如pandas等之外，整个统计计算过程只能使用numpy完成。请在收到题目后尽快完成并提交，答题效率也是考核因素之一。

**四、交付：**将相关交付物与你的最新中文简历一起打成压缩包，以【金讯扬行笔试答题：你的姓名】格式命名后，发送至邮箱caichao@castaliafund.com

1. 相应计算代码；

2. 计算出的所有股票的长卖单成交量占比因子结（csv格式）；