# 思路总结

## 读数据

把数据从log文件中读到python的list，读取的时候只取解锁状态下的数据。

## **判断解锁状态**

因为每个log文件的数据的Timeus是以递增的形式存在的，所以采取标志位的方法来判断是否为解锁状态，当MKF1的ARD为0是ard\_status为False，直到ARD=0，ard\_status才变为True，直到ARD=0再变为False，如此循环。

## **处理数据**

把数据分为三类，一类是FMT变量包数据，保存的是其他变量包的变量格式；第二类是PRM和MSG，这类数据的变量里面含有字符串，不起作用，采取忽略的形式处理；第三类是剩下的其他变量包，这类变量包的格式是变量包名+多个变量数据。

## **转换数据类型**

把list转成pandas的DataFrame，这样的类型有利于索引及后续的分组聚合。

## **分组聚合**

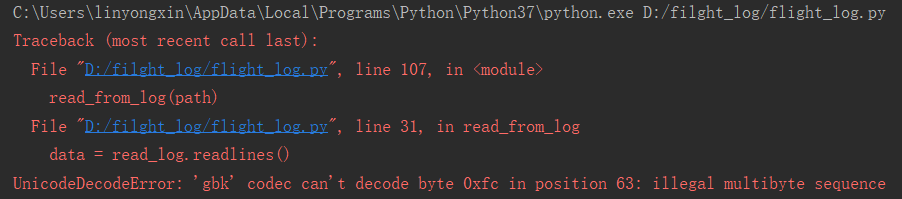
分组聚合采用GroupBy技术，Pandas自带的一种专门用来对数据分组聚合的工具。

## **导出到Excel**

把分组聚合好的数据导出到Excel的sheet里面，这里采取的是每个变量包存到一个sheet里面。

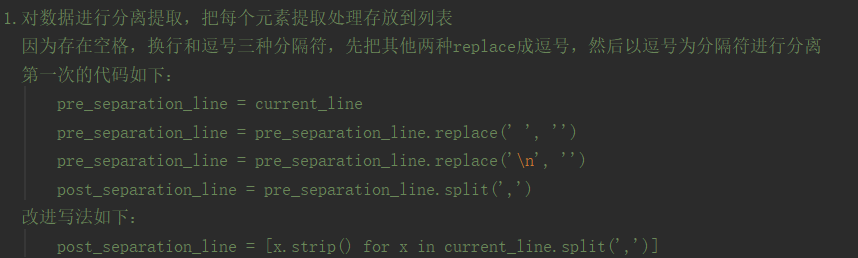
# 错误总结

## 读数据环节



原因分析：当前文件夹里面有一个Excel文件，里面的数据读取的是有错误。如果文件夹全部是log文件是不会报错的。

## **处理数据环节**



注意：split是字符串的方法，list不能用这个进行分离

## **分组聚合环节**

1. **用什么方法进行分组聚合？Numpy？List？For循环？**

**解决方法：**这一环节的主要问题在于如何把DataFrame进行按某一列进行分组，想了很多方法，后来看书才发现GroupBy这个技术。

1. **因为第一列是变量包名称，如何删除DataFrame的第一列？**

**解决方法：**del（obj[0]）可以进行删除处理，DataFrame中索引第一列是obj[val]，索引第一行是obj.ix[val]（这个很重要，跟二维数组不一样，卡在这里很久），对于DataFrame的索引方法不熟悉会导致删错数据而导致出错。

## **导出到Excel环节**

1. **运行成功，但找不到Excel文件？**

**解决方法：**缺少out\_writer.save()，没有这个保存指令，文件是不存在的！！！

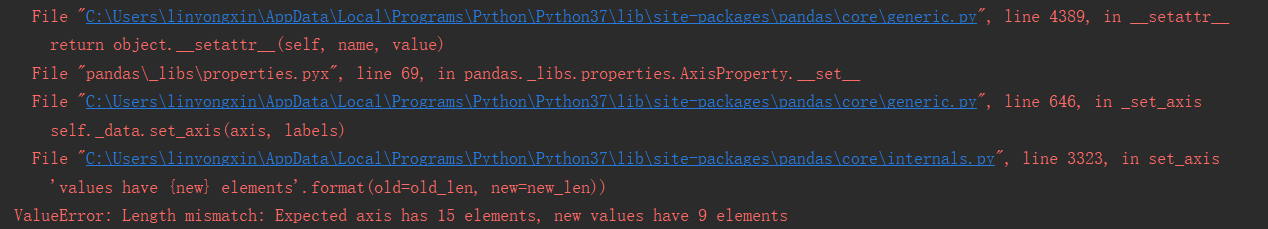
1. **GroupBy分组后无法直接使用.to\_excel这个方法，如何导出？**

**解决方法：**在PyCharm中要显示GroupBy后的内容需要对GroupBy后的DataFrame进行for循环一个一个打印，因此可以用for循环把每一个group打印到一个sheet里面。（这里研究了好久，问题出在对GroupBy不熟悉。）

1. **导出到Excel中的数据开始没有列索引名称，也就是没有变量名称，怎么解决？**

**解决方法：**GroupBy后的数据中的每一个group是DataFrame的数据结构，因此可以使用DataFrame的修改列索引名称的方法来修改

1. **DataFrame的修改列索引名称时报错，提示如下：**



**解决方法：**明显是修改后的列索引无法满足所有列元素的个数，后来发现原理是把list数据转换成DataFrame后DataFrame把原先没有数值的元素变为NAN，也当成元素来看待。比如，原先只有9列，但是DataFrame是整个list进行转换的，不是单个变量包，它的对齐模式以最大个数的那个为准，所有就会变成15列，而且这15列的后6列全部是NA。这种情况下修改索引就出现个数不匹配，列索引名称只有9个，但有15列。后来查书本发现这个问题，采取了删去无效值的方法来解决,用**.dropna(axis=1)**即可解决。

1. **GroupBy导出的过程中使用mean、var、std进行数值计算时报错？**

**解决方法：**问题出在GroupBy之后的DataFrame里面的数据类型是str，而这些数值计算方法只能是数值型，所以需要进行转换，把字符串类型转换成float，用法如下：DataFrame（groups，dtype=np.float）

# 整体总结

## 方法总结

本次的小项目让我快速掌握用python进行数据分析的一些常用工具，我采取的方法是先快速看书，然后开始实践，实践过程中遇到不懂的再看书或者谷歌。推荐一本好书《利用python进行数据分析》，这本书专门介绍用python进行数据分析，去除很多python在其他方面的运用，因为这个项目的目的是进行数据分析，所以可以用这本书做参考。我编写程序的时候大部分工具都是采用这本书介绍的。第一遍看书快速过一遍，把一些基础的以及标题浏览一遍，简单的工具也跟着步骤实验一边，做到大概知道每个章节每个小节讲什么内容，然后开始针对本项目写程序，最后是碰到问题先找书，找不出来再谷歌，然后再参考书本进行修改。

## 2.学习体会

因为python是一门高级语言，所以我们不能把它用成C语言这样的底层语言，也就是说要善于运用python的库或者说工具包，很多底层的逻辑不需要自己去写，不然的话就没有用python的必要了，鉴于这一点，我碰到问题解决不了第一反应是考虑有没有实现这样功能的库可以给我用，而不是去考虑如何通过逻辑去实现这个功能。换句话说，我个人认为python比的是谁知道的库多，谁用的库多，谁开发起来就高效。这也是我觉得先看书再去开始写程序的原因，书上会介绍一些好用的库，有了这些库开发起来就事半功倍了。比如，存数据的时候我们都只想着用list，但如果知道python有个pandas的模块，里面有DataFrame的数据结构，那对数据的处理就会容易很多，包括后面的运用GroupBy进行分组聚合。

**附录：pycharm查看性能**

