MorningStar

目录

[00.URLSet](#_Toc940201419_WPSOffice_Level1) [2](#_Toc940201419_WPSOffice_Level1)

[01.Download 下载源数据](#_Toc517637441_WPSOffice_Level1) [2](#_Toc517637441_WPSOffice_Level1)

[02.QC数据清洗](#_Toc476216890_WPSOffice_Level1) [3](#_Toc476216890_WPSOffice_Level1)

[03.Calculate 基础运算](#_Toc105717861_WPSOffice_Level1) [4](#_Toc105717861_WPSOffice_Level1)

[04.ReCalculate 数据补全](#_Toc831113758_WPSOffice_Level1) [4](#_Toc831113758_WPSOffice_Level1)

[05.Sum 数据总结](#_Toc1295290618_WPSOffice_Level1) [5](#_Toc1295290618_WPSOffice_Level1)

前提：开始以下步骤前，先把最新的Master\_symbol\_v1.7\_2020.7.xlsx

置于以下代码文件的相同目录下，和安装所需python库。

Setting:

Python 3.8

Pandas 1.2.3

Request 2.25.1

Openpyxl 3.0.7

Numpy 1.19.5

Running time:

Total: 1530min

- TSX 30min

- China 100min

- US 400min

Data Clean and Manipulate :1000min

# 00.URLSet

打开文件：00 URLSet.py

用途: 设置股票的URL，并保存为json格式的文件中，为后续下载作准备

操作流程：

把所需要的Master\_symbol\_v1.7\_2020.7.xlsx文件放到同级目录下，设置你所要跑的国家名称即可，无需输入或修改参数，直接运行程序。获取每支股票Ownership页面的网址，网址以字典形式保存于json文件之中，一般只需要在项目，开始时运行一次，生成的文件将用于后续工作，(有部分股票网址为空，可继续更新)。

程序输出：

./util/Owner\_URLs\_"xx".json

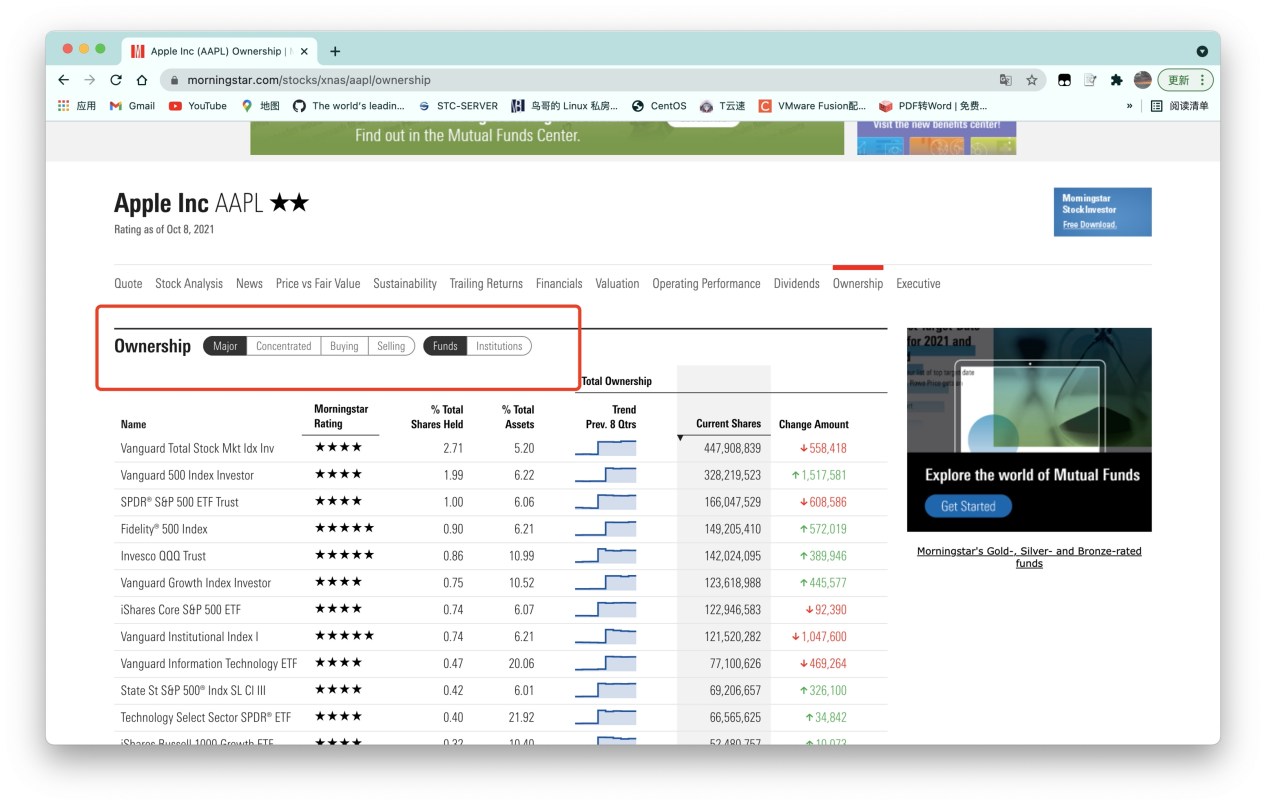
# 01.Download 下载源数据

用途: 下载Funds和Institutions的源数据

打开文件：01 Downloader.py

注意：每个子线程之间的休息时间视网络状况而定

需要下载的内容：

每一只股票Funds和Institutions的Major，Concentrated，Buying以及Selling

操作流程：

同样修改主函数里countries的值运行即可

程序输出：

./RawData/"Country"/Download\_"xx".xlsx

# 02.QC数据清洗

目的：合并funds和Institutions的四个标题为两个sheet并且添加组成新的heads，分开为每个不同的季度

打开文件：02 QC.py

操作流程：

同样修改主函数里countries的值运行即可

程序输出：

./QCReport/"Country"/QCReport\_"xx".xlsx

# 03.Calculate 基础运算

目的: 将 Funds和Institutions重新格式化为所需的格式，并计算 Funds的 %share 和 Institutions的 %share

打开文件：03 Calculate.py

操作流程：

同样修改主函数里countries的值运行即可

程序输出：

./Calculate/"Country"/Calculate\_"xx".xlsx

# 04.ReCalculate 数据补全

目的: 这个程序是用来操作下载的数据的。为避免误处理空数据，提供空数据模板预设数据格式。

打开文件：04 ReCalculate.py

操作流程：

这里需要拷贝一份Calculate跑出来的数据即文件夹到你的桌面，并且要修改里面的存放路径，然后通过修改countries的值运行即可，注意的是countries的值要和你放的文件夹内容的国家要一致。

程序输出：

./Calculate/"Country"/Calculate\_"xx".xlsx

# 05.Sum 数据总结

Purpose: 合并股票symbol的所有符号数据Major，Concentrated，Buying以及Selling

打开文件：05 Sum.py

操作流程：

同样修改主函数里countries的值运行即可

程序输出：

./result/SumReport\_"date".xlsx