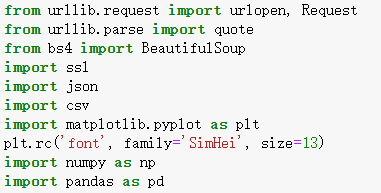
# 豆瓣电影爬虫报告书

## 一、本次项目使用的工具

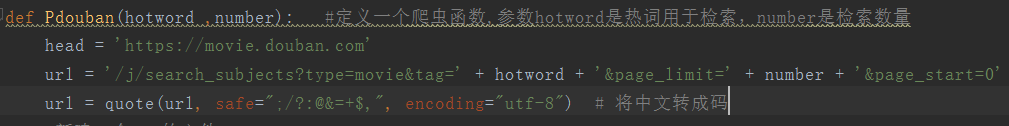
此次项目起初使用**PyCharm**进行调试，后期为了使界面更友好而转移至**Jupyter**并添加了一些可视化功能。

以下是此项目使用的包：



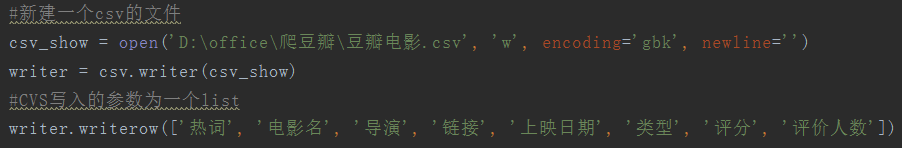
## 研发过程

1. 直接打开‘https://movie.douban.com’并读取输出，没有发现电影的详细信息，通过观察豆瓣网页源码，找到包含电影信息的链接，再通过观察发现链接中包含热词与电影数量的参数，从而决定创建一个拥有两个参数的函数用于检索。



1. 考虑到后期要做可视化，因而创建了一个csv文件，用于存储电影信息。

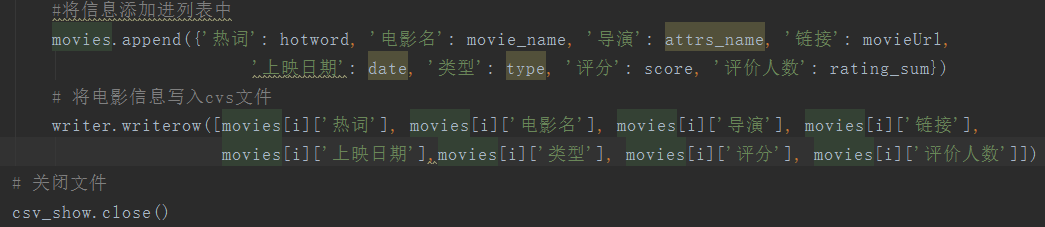
创建csv文件：



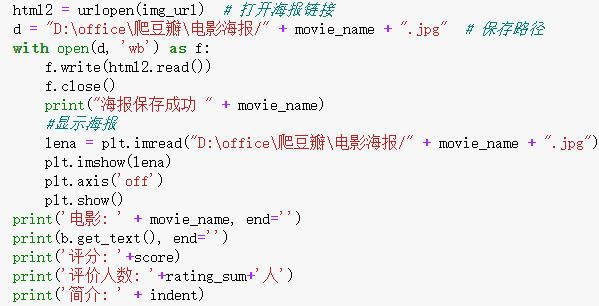
1. 接着通过标签属性定位并获取需要的信息



录入信息：



显示爬取的信息：

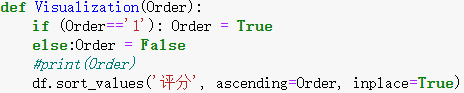


效果：



1. 根据评分排序

创建拥有一个参数的函数，参数控制评分的升序降序



排序效果（升序）：

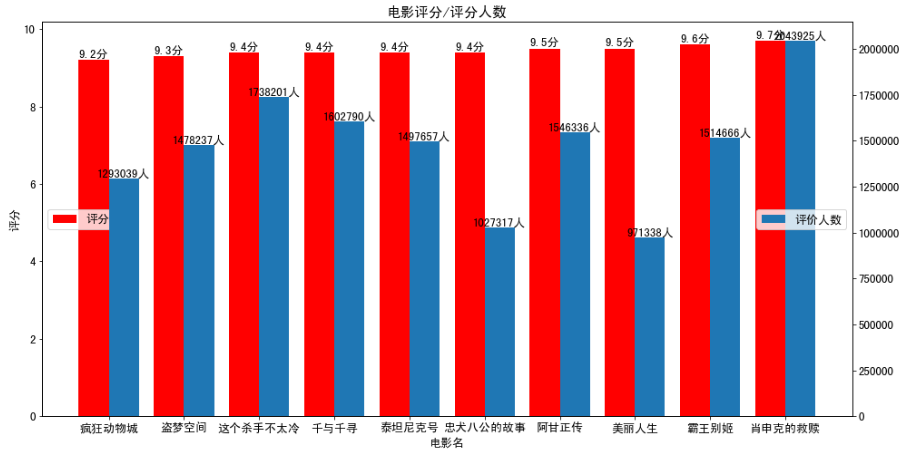


1. 双标柱状图（附加可视化）

代码：



效果：



## 小结

经过这次的项目我对标签与网页结构有了更深刻的理解，而且发现**Jupyter**相比**PyCharm**的界面更友好，不过**Jupyter**在写码的时候没有提示就很难受，要是把俩工具合一起就好了。还有就是对数据处理，包括存储至csv、查询、排序、可视化等操作更加熟练。最后就是发现有好多高评分电影没看过，暑假要去补一补了。