****

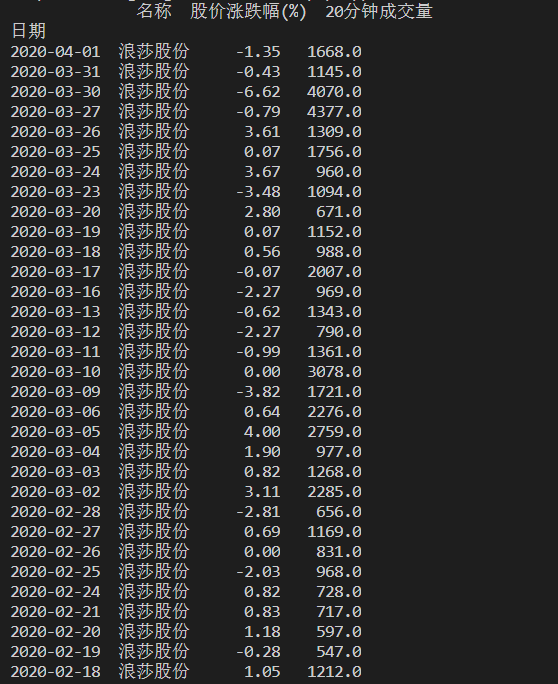
**HARBIN INSTITUTE OF TECHNOLOGY**

2020 年 春 季学期本科生课程考核

（结课报告）

|  |  |  |  |
| --- | --- | --- | --- |
| 考核科目 | Python金融大数据分析 | | |
| 学生所在院（系） | 计算机科学与技术学院 | | |
| 学生所在专业 | 大数据科学与技术 | | |
| 学 生 姓 名 | 彭钰驯 | | |
| 学 号 | 1170300916 | | |
| 考核结果 |  | 阅卷人 |  |

1. 运用相关库获取上市公司 A 的股票基本数据、计算相应的衍生变量数据、通过相关性分析选取合适的衍生变量、可视化呈现数据、并生成 Excel工作簿。
2. 运用Tushare库获取上市公司A从2020-02-03至2020-04-01的股价涨跌幅和前20分钟成交量数据。

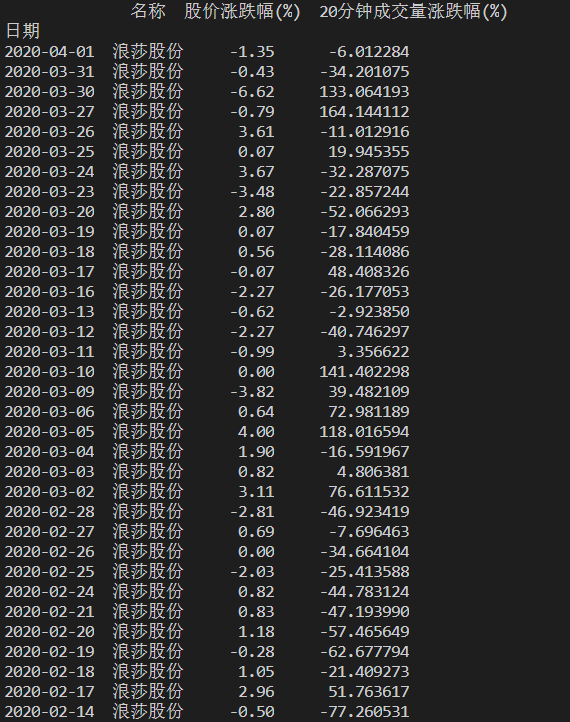


2) 采用当日交易量和昨日交易量的方法:

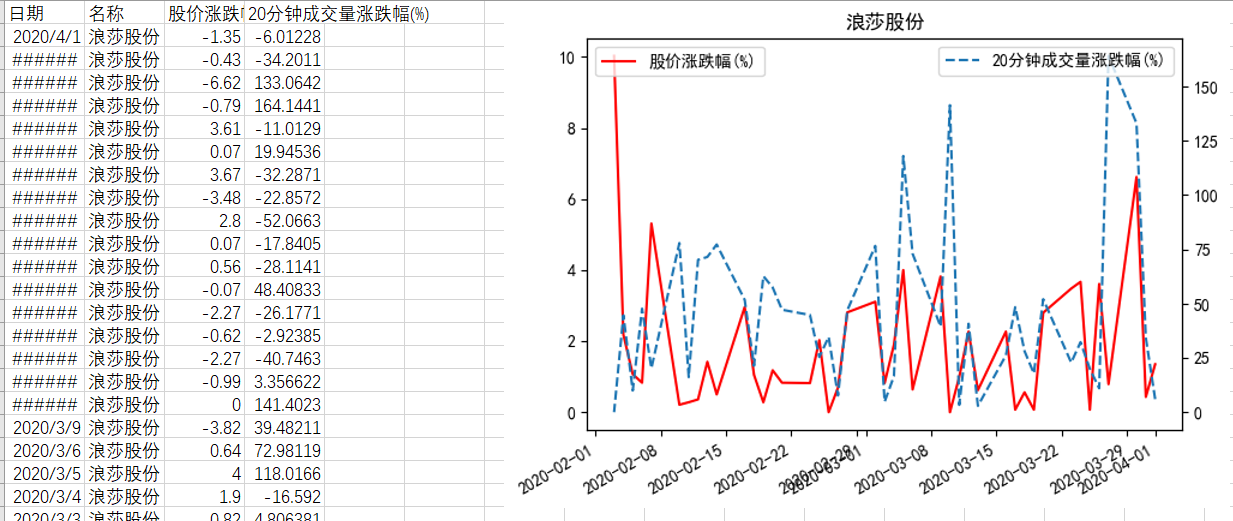
采用当日交易量和多日交易量的平均值的方法:



1. 通过皮尔逊相关系数分别进行两种前20分钟成交量涨跌幅与股价涨跌幅的相关性分析，并选择出更优的前20分钟成交量涨跌幅。



1. 在进行简单数据优化后，绘制股价涨跌幅与前20分钟成交量涨跌幅的可视化图表(2个Y轴的形式）。



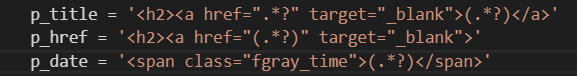
1. 运用 xlwings 库将优化后的数据以及可视化图表导出至 Excel 工作簿。（除了最终的大作业报告，还需提交相应源代码以及Excel 文件）

见文件

1. 运用相关库从新浪财经网获取上市公司A的网页源代码，提取相应的标题、网址、来源和发布日期信息，进行简单的数据清洗，并生成文本文件。具体要求如下：
2. 运用 requests 库从新浪财经中获取上市公司 A 的网页源代码。



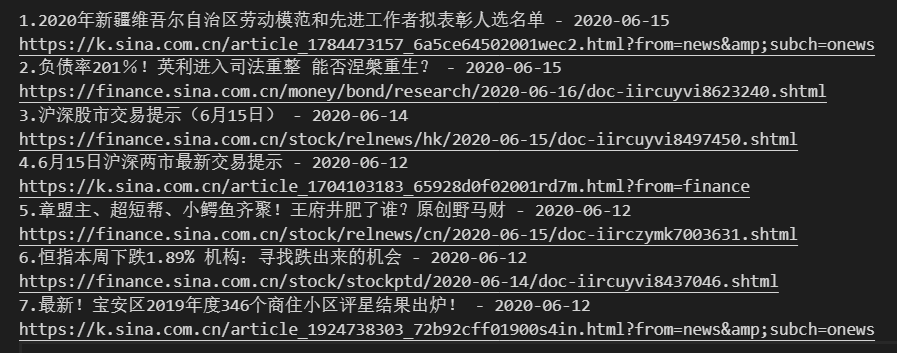
2) 编写正则表达式获取新闻标题、网址、来源和发布日期信息。



1. 对获取的新闻信息进行简单的数据清洗。



1. 尝试 从新浪财经 中获取 上市公司 A 的前5页网页源代码, 同时完成 2 和3 中的内容（新闻信息的提取和简单数据清洗）。



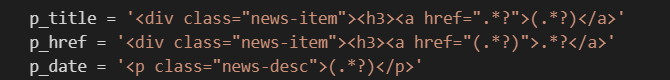
1. 将4中的数据信息保存到一个文本文件中。（除了最终的大作业报告，还需提交相应源代码以及txt文件）

见文件

1. 运用相关库从东方财富个股吧中获取上市公司A的网页源代码，提取相应的标题、网址、来源和发布日期信息，进行简单的数据清洗，并生成文本文件。具体要求如下：
2. 运用 selenium 库从东方财富个股吧中获取上市公司 A 的网页源代码。



1. 编写正则表达式获取新闻标题、网址、 来源 和发布日期等信息。



1. 对获取的新闻信息进行简单的数据清洗。



1. 将数据信息保存到一个文本文件中。（除了最终的大作业报告，还需提交相应源代码以及txt文件）



1. 运用相关库获取上市公司 A 和上证指数的股票价格数据，分析其基本特征，进行正态性检验，并实施贝叶斯回归。具体要求如下：
2. 运用 Tushare 库获取 A 公司和上证指数(000001)从2018- 01-01 至 2019-12-31 的股票价格数据收盘价。
3. 将 A 公司和上证指数的股票价格数据规范化并绘制股票规范化价格数的线图和散点图。
4. 计算 A 公司和上证指数的 对数收益率，并 进行正态性检验 (3种方法选其中2种）。
5. 运用 pymc3 库 对 A 公司和上证指数 的股票价格 实施 贝叶斯回归 (A 公司的股票规范化价格为 x 值,上证指数的股票规范化价格为y值），输出统计结果，并绘制后验分布及轨迹图，和贝叶斯回归线。（除了最终的大作业报告，还需提交相应源代码以及txt文件）