**一、爬虫文件**

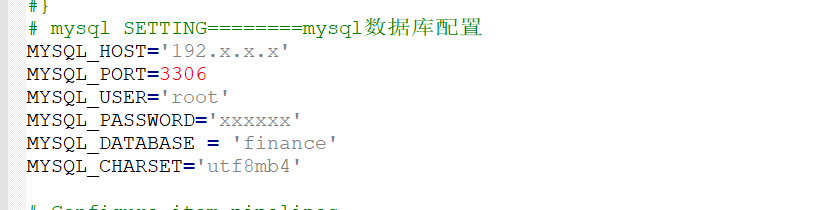
环境配置：https://blog.csdn.net/weixin\_36467887/article/details/147804202

爬虫笔记：https://blog.csdn.net/weixin\_36467887/article/details/147893769

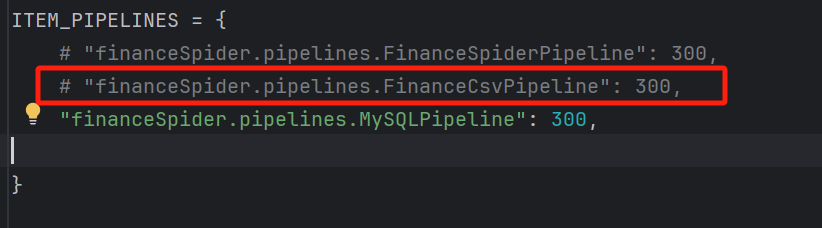
<https://blog.csdn.net/weixin_36467887/article/details/147956670>

**注意：**

1、数据存储在csv里面，代码可以直接运行，如果想要存储到mysql，你们要把settings文件中的mysql的配置信息改成自己mysql信息



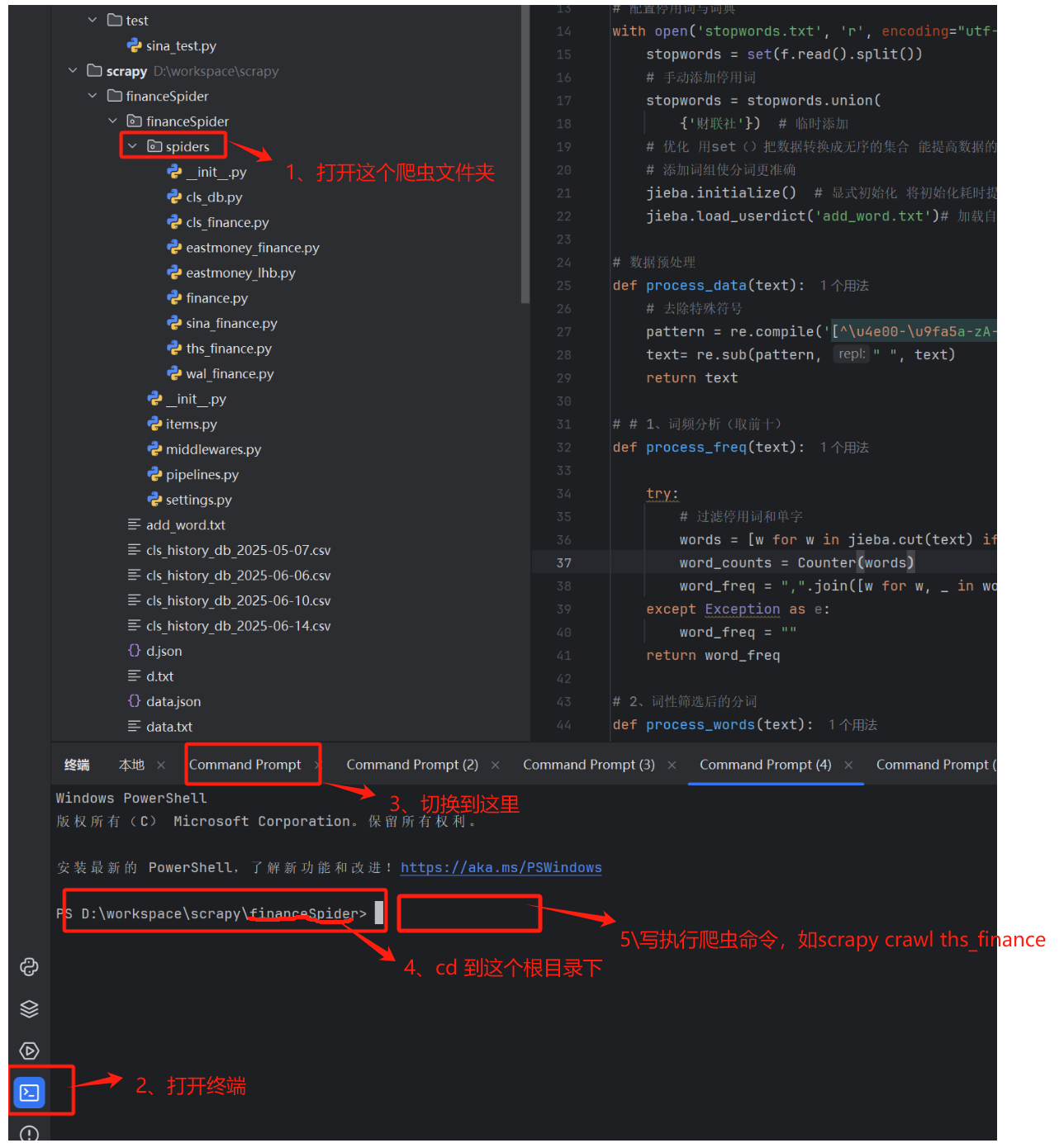
然后在settings中注释掉csv管道，释放mysql存储管道，如下图



2、运行

这套代码的python版本是python3的，请确认您的版本是否在python3以上。

用python工具（如pycharm）打开项目后，别直接在py文件上点run，scrapy运行是在终端(命令行界面)执行命令，不懂的话请按照我的爬虫笔记操作。



**执行需要的依赖包下载**：pip install -i https://pypi.mirrors.ustc.edu.cn/simple/ jieba wordcloud snownlp scikit-learn selenium pymysql bs4

这些可能不全，你看代码报错缺啥包你就安装，如pip install -i <https://pypi.mirrors.ustc.edu.cn/simple/> jieba（module named）

**执行爬虫命令可以放一起运行，如下**：

scrapy crawl eastmoney\_finance

scrapy crawl ths\_finance

scrapy crawl wal\_finance

scrapy crawl sina\_finance

scrapy crawl cls\_finance

；

跑完后数据存储在finance\_news\_{today}.csv（如finance\_news\_2025-06-23.csv

）--源数据集

**二、数据分析文件**

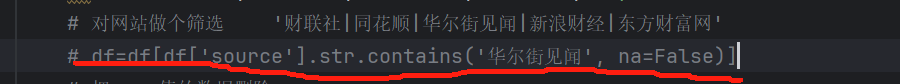
**学习笔记**：<https://blog.csdn.net/weixin_36467887/article/details/148589164>

**文件如下**：

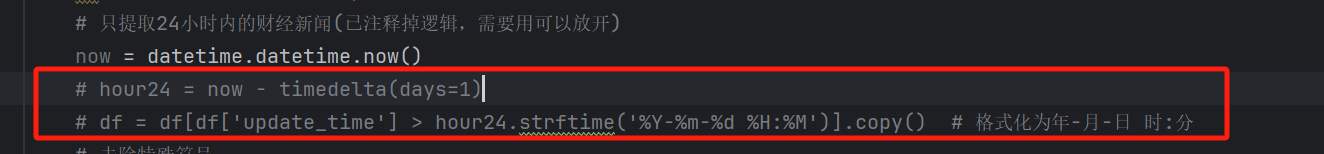


\*\*\*可视化分析文件finance\_visualization有几个功能我注释了，如果需要可以自己打开（删掉前面#），如下：

1. 做网站筛选



1. 筛选24小时内新闻



1. 词云地图形状



**运行**

这两个.py文件运行直接点run，如下



Finance\_nlp运行后数据存储在finance\_nlp\_{today}.csv（如finance\_nlp\_2025-06-23.csv

）

如需定制爬虫、数据分析、NLP处理、远程代码调试以及讲解、大屏可视化、物联网、大数据项目、微信小程序，python代写调试教学等业务，欢迎咨询，价格再议！

【源码和笔记为本人原创，仅作为学习交流使用，请勿用于其他用途，谢谢！】