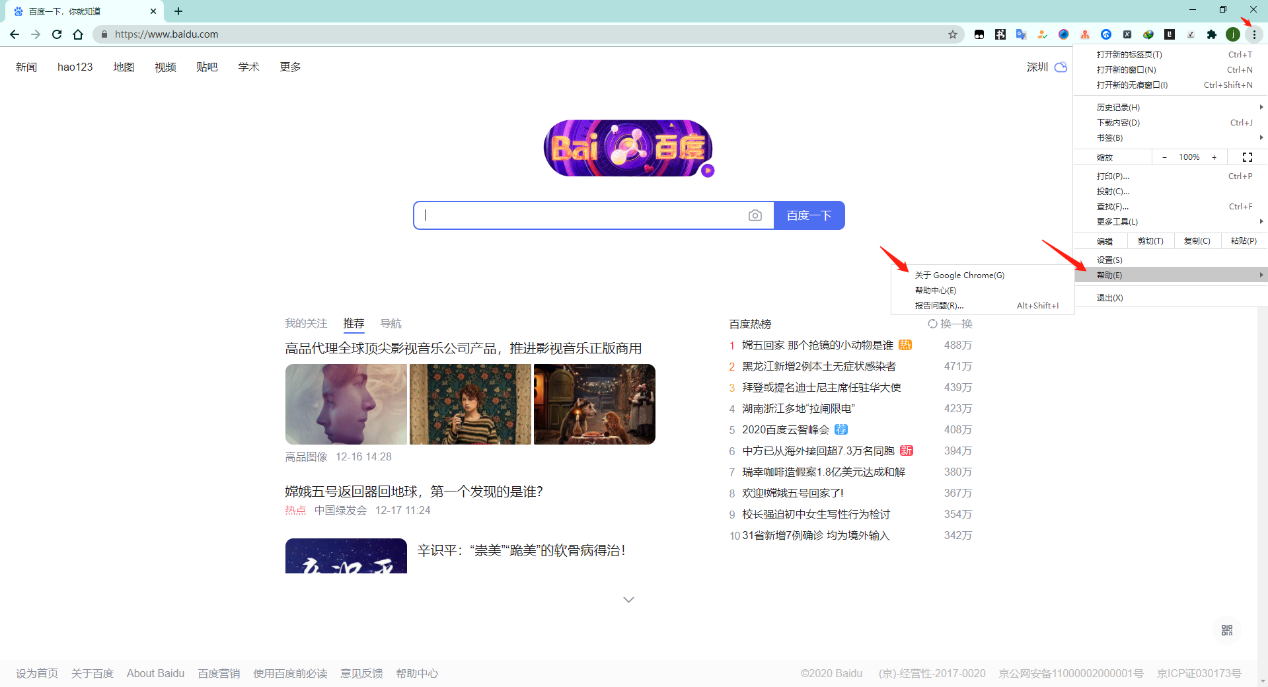
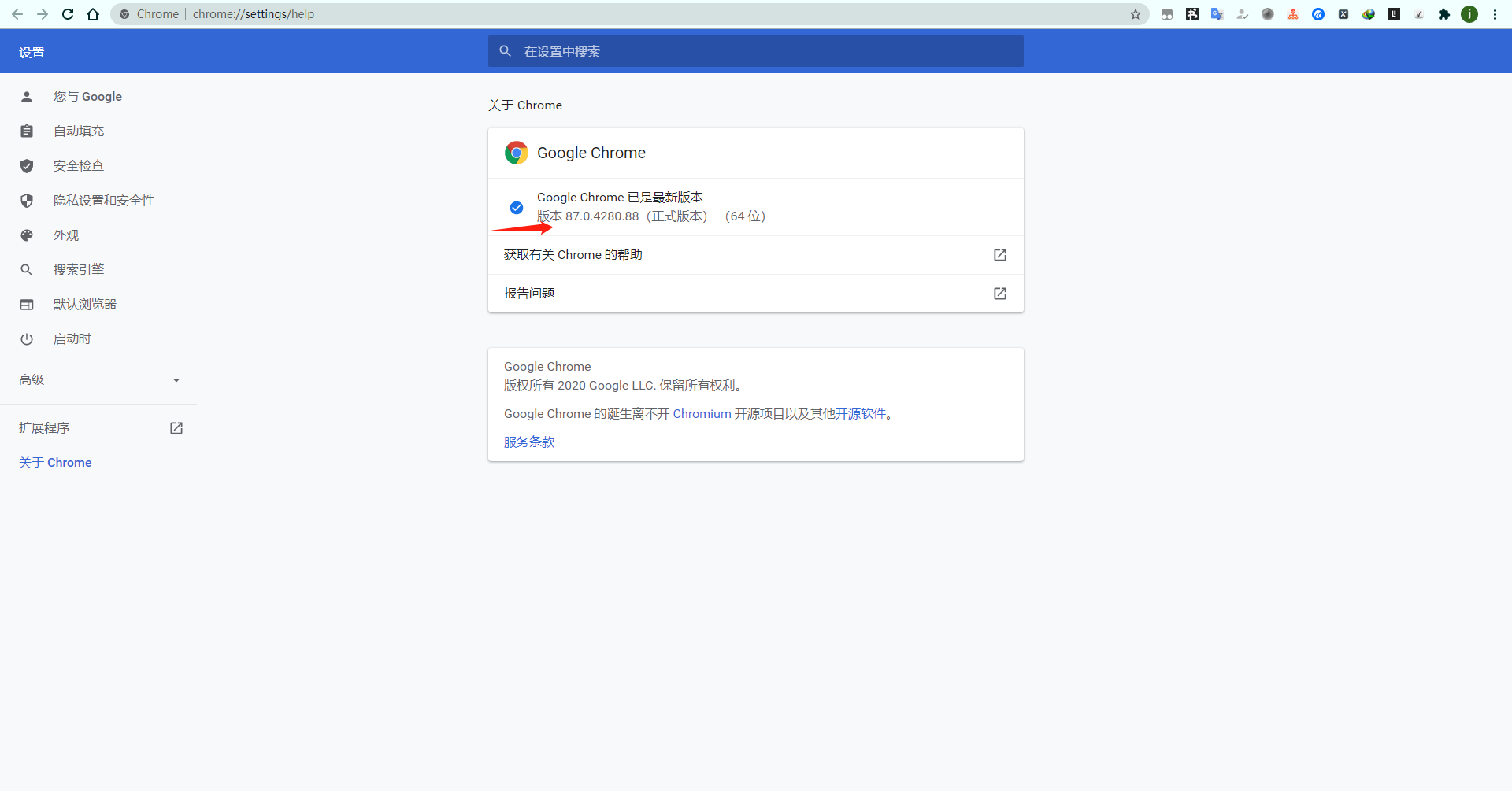
1. 知乎热榜说明书

前提：安装好python环境，下载pycharm 或 vscode 编辑器

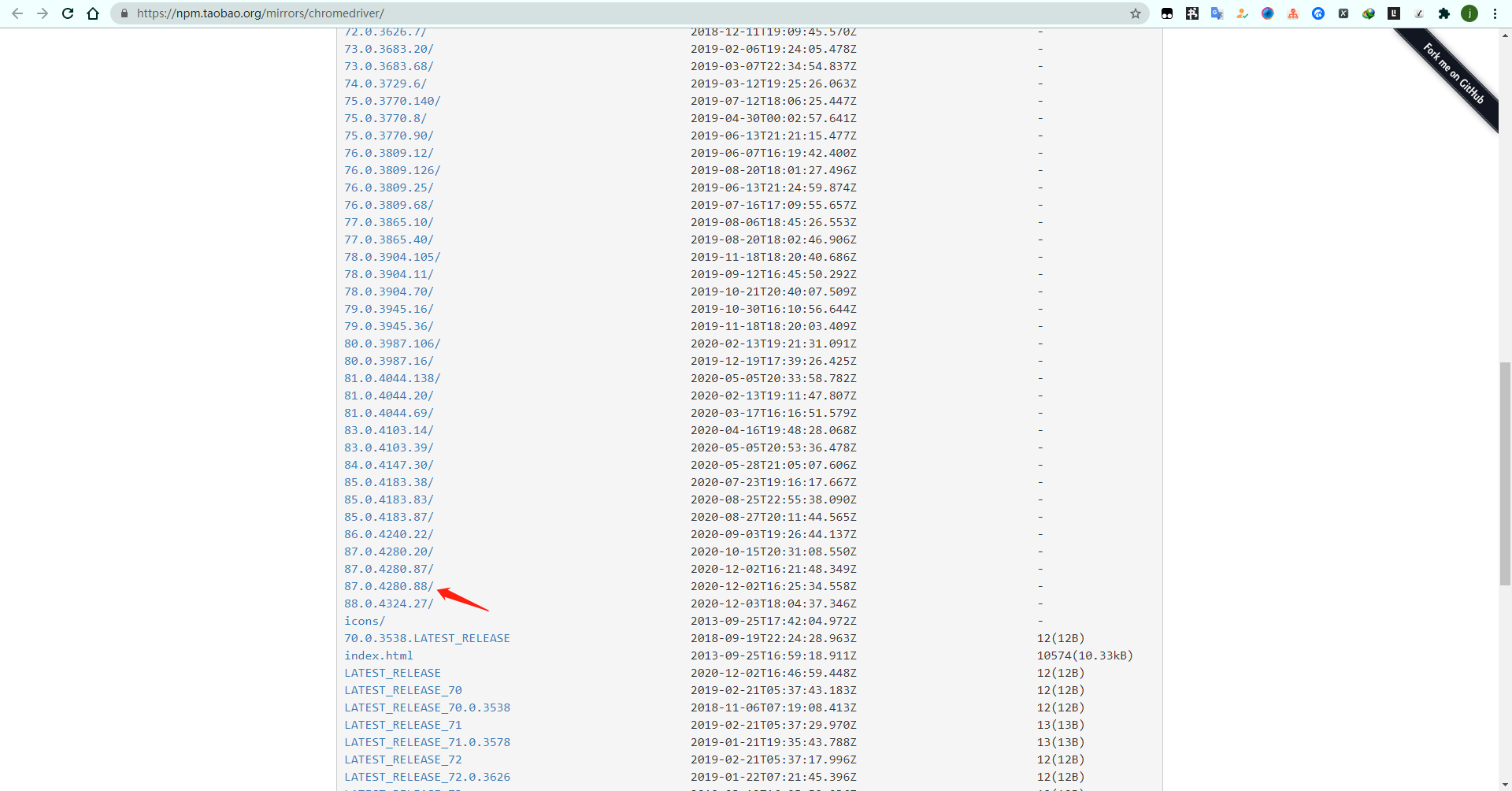
1. 下载chrome浏览器 查看chrome版本



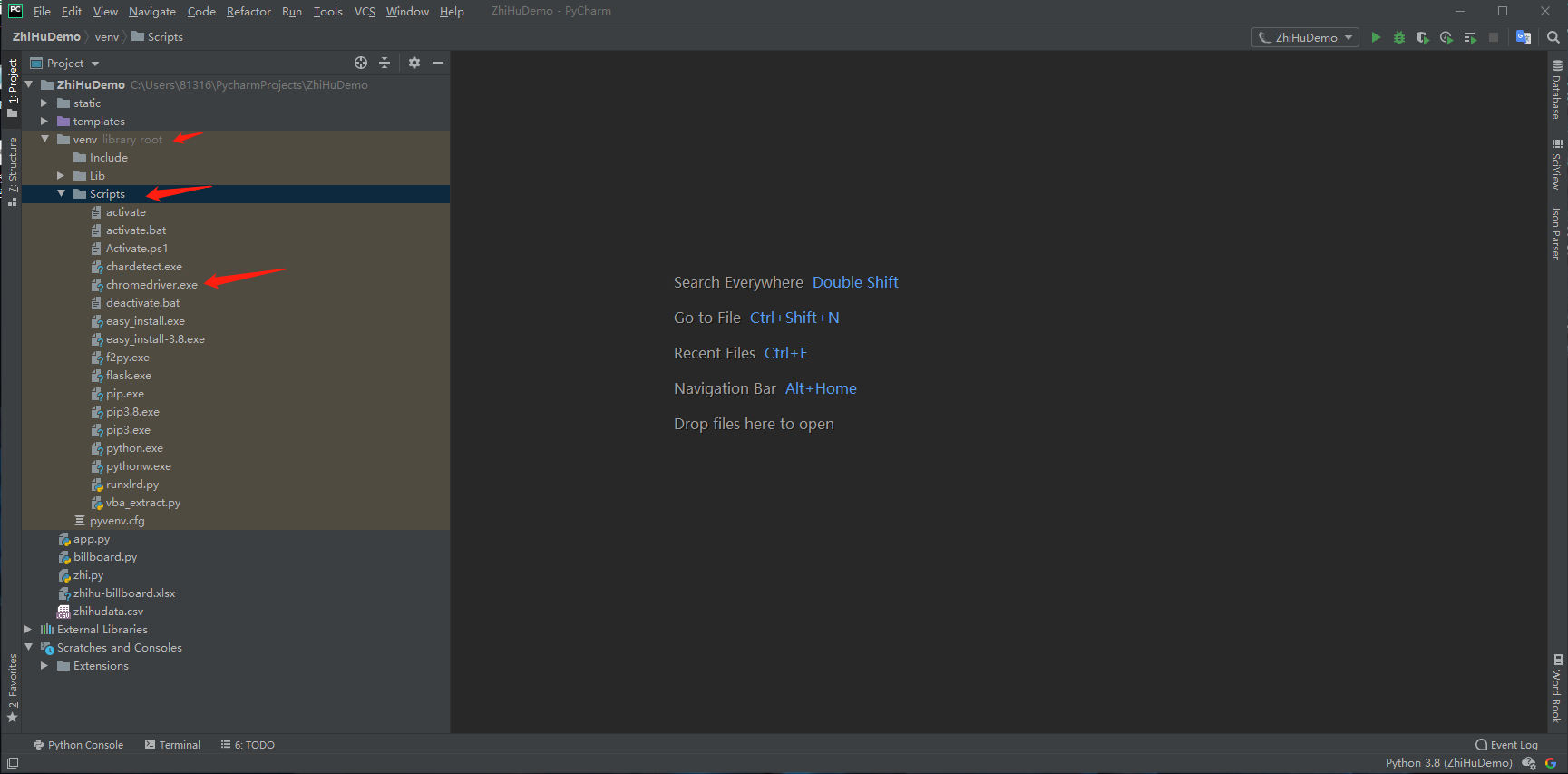


2.下载chromedriver 谷歌浏览器与chrome对应的版本驱动

<https://npm.taobao.org/mirrors/chromedriver/>



3.将chromedriver放置flask项目的虚环境里面



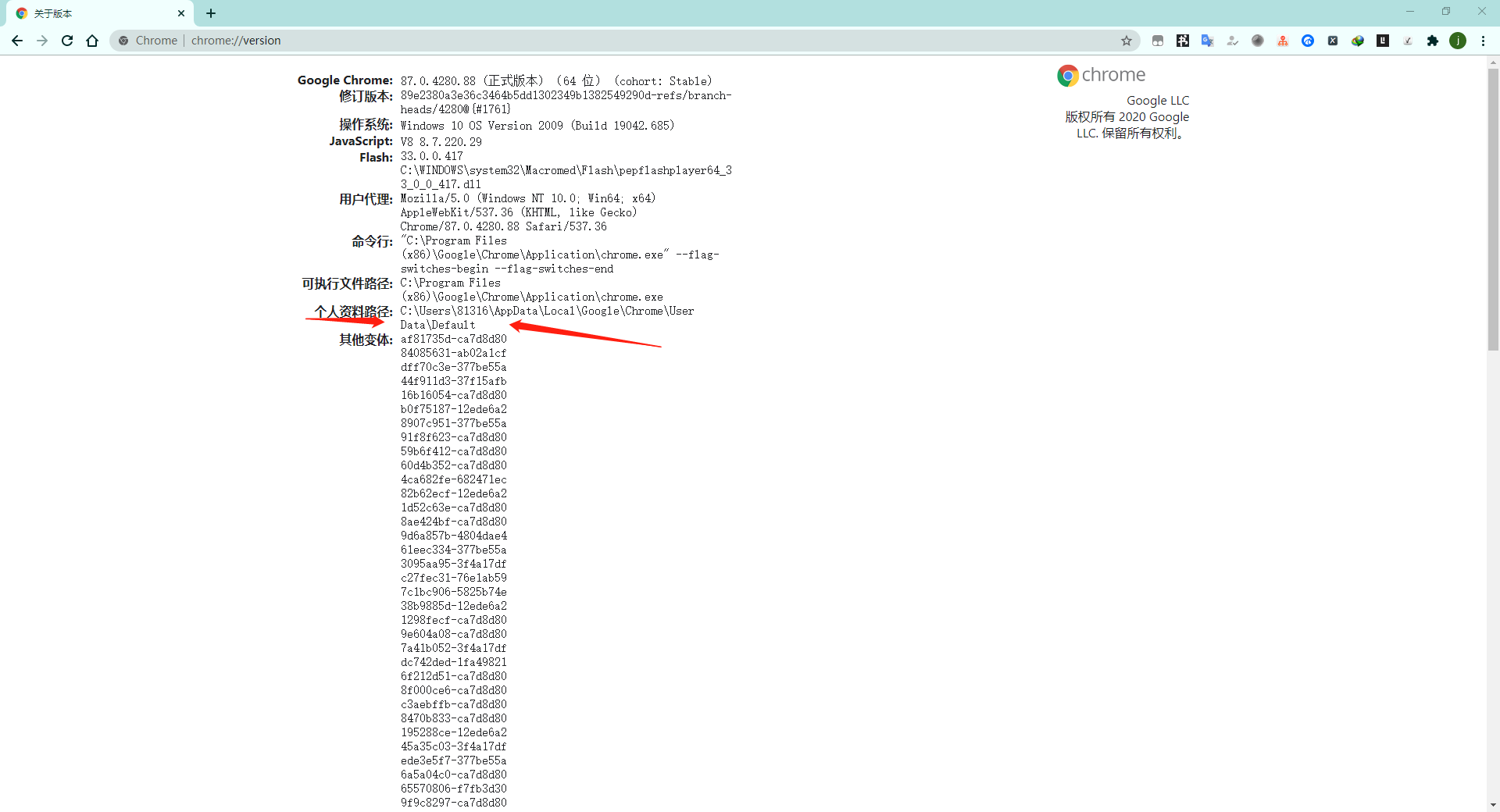
4.修改爬虫代码

原因:知乎热榜的爬取必须要登录操作,为了减少麻烦，我直接让selenium使用本地浏览器，因为我之前已经登录过了，浏览器保存了知乎的登录cookies,所以来实现知乎免登录的功能

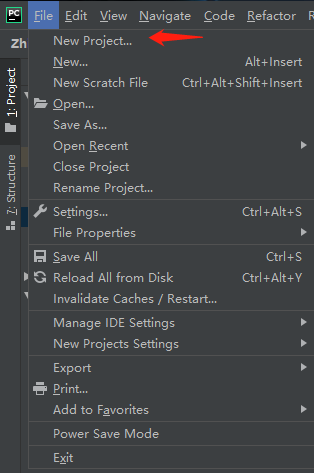
options.add\_argument('--user-data-dir=C:/Users/81316/AppData/Local/Google/Chrome/User Data/Default')

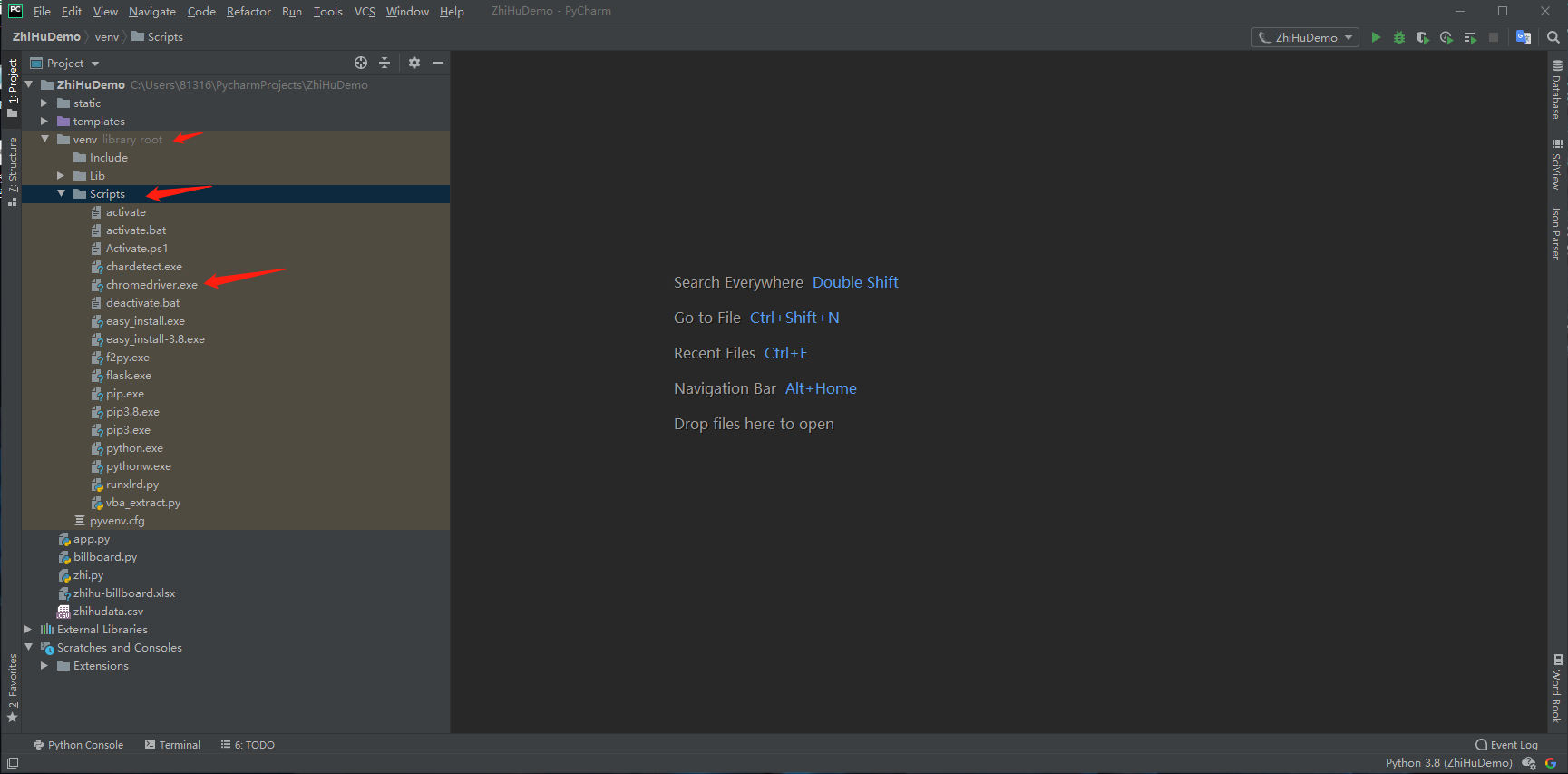
注意个人资料路径：必须是斜杆/ 不能是反斜杠\

谷歌浏览器地址栏输入 chrome://version/ 查看个人资料路径，复制自己谷歌浏览器个人资料路径直接替换

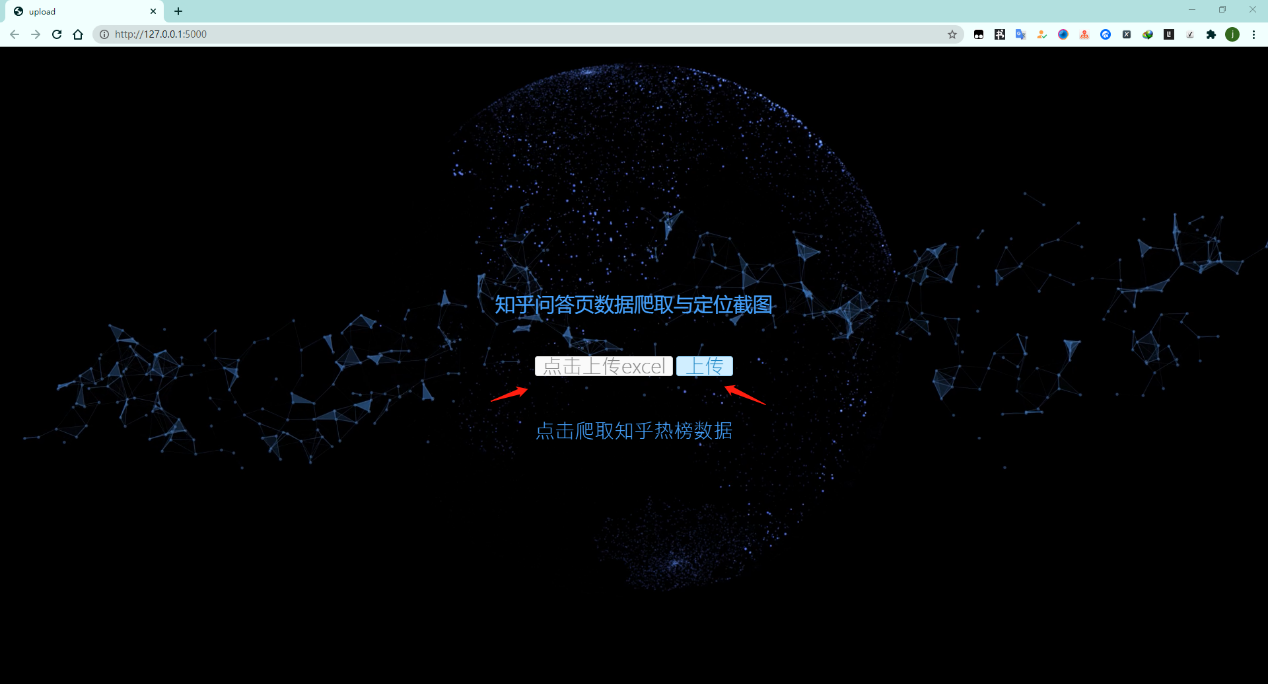


使用pycharm新建flask项目



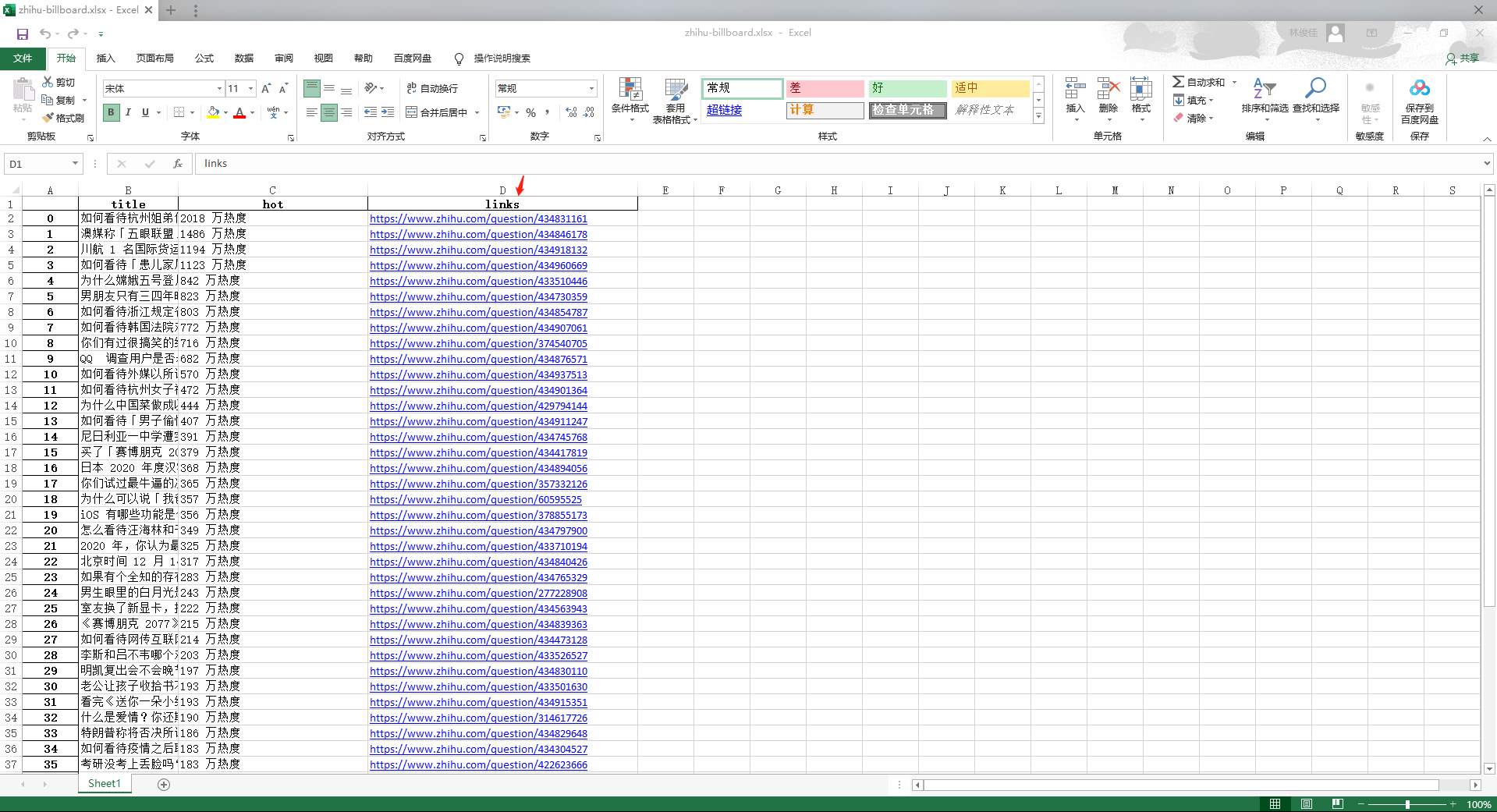


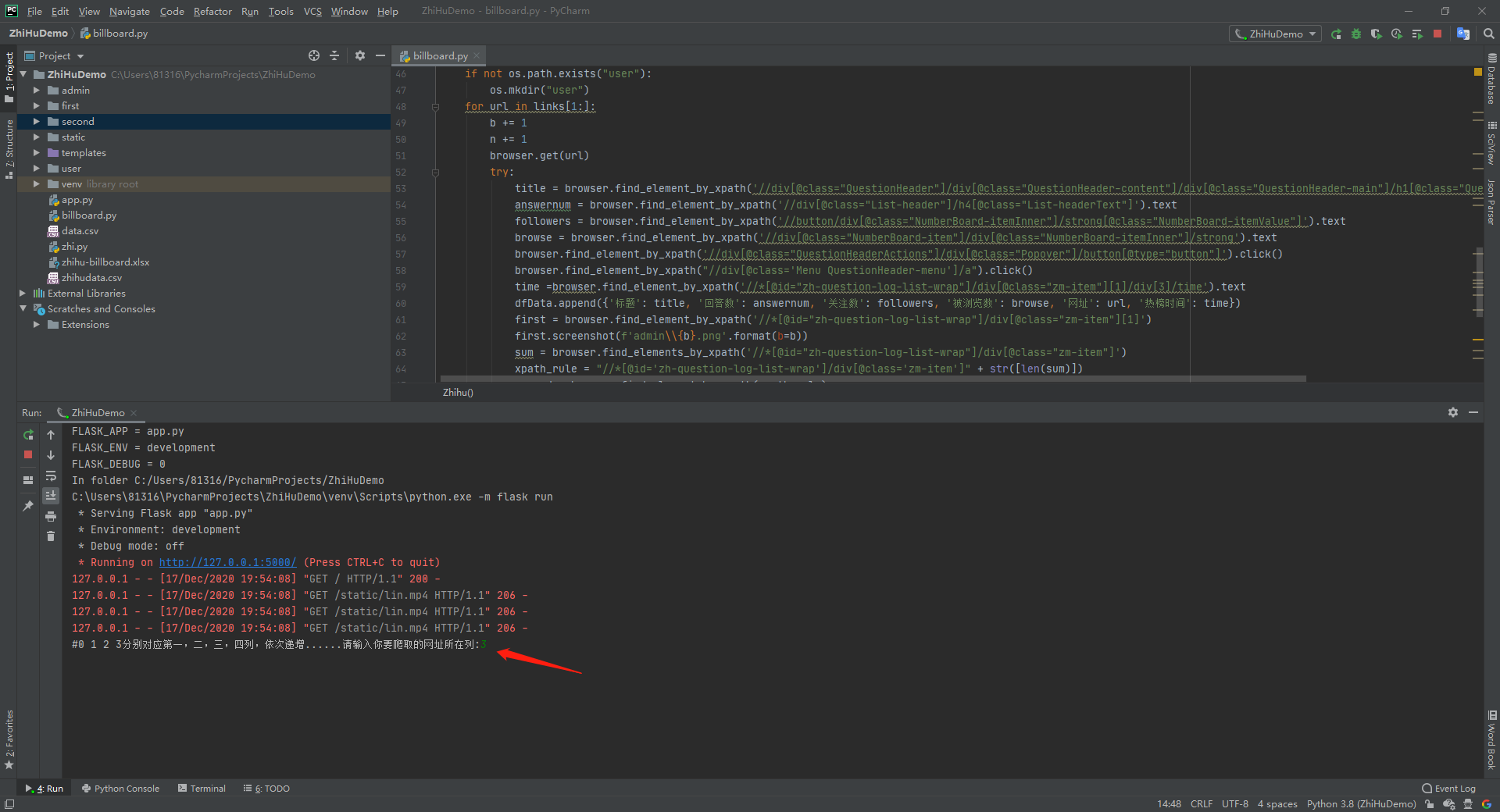
完成后，导入flask的全部项目文件，安装相应库，后就可以启动flask项目了



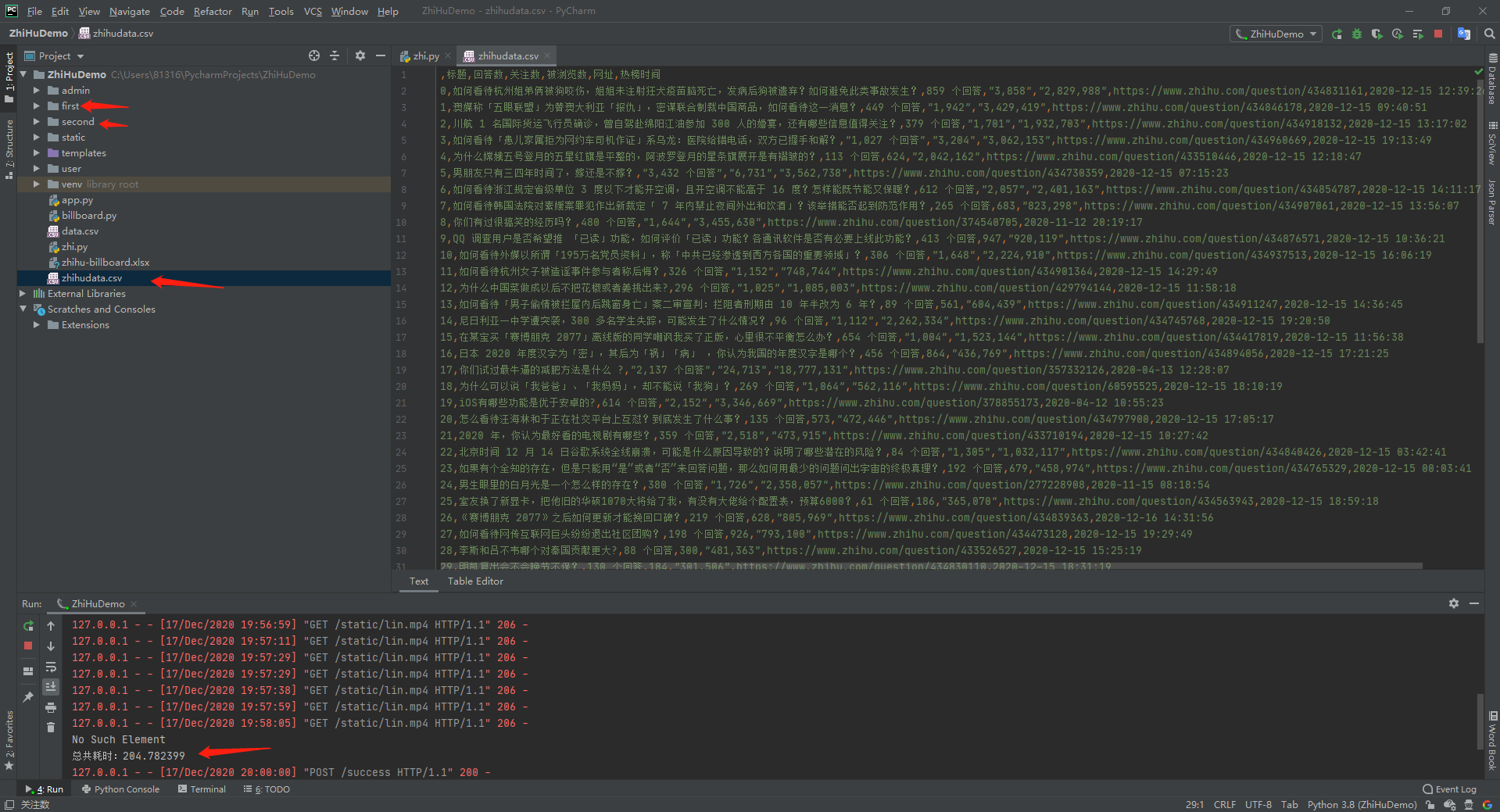
由于我知乎热榜问答页的网址存放在excel表的第四列,所以我在终端这里输入3

因为 0 1 2 3分别对应第一，二，三，四列，依次递增......





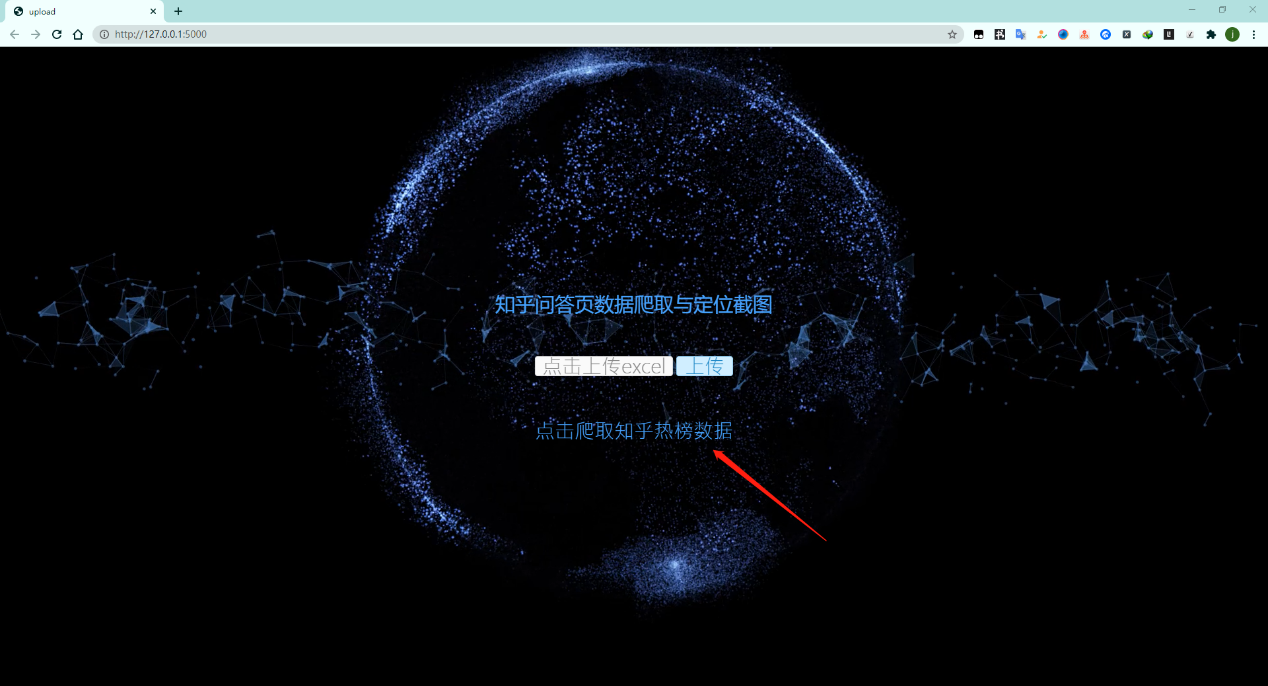
爬取完成后，终端可以看到全部爬完消耗了多少时间，生成first文件夹存放知乎管理员锁定编辑截图 ，second文件夹存放某人添加了问题截图，生成zhihudata.csv存放爬取的数据



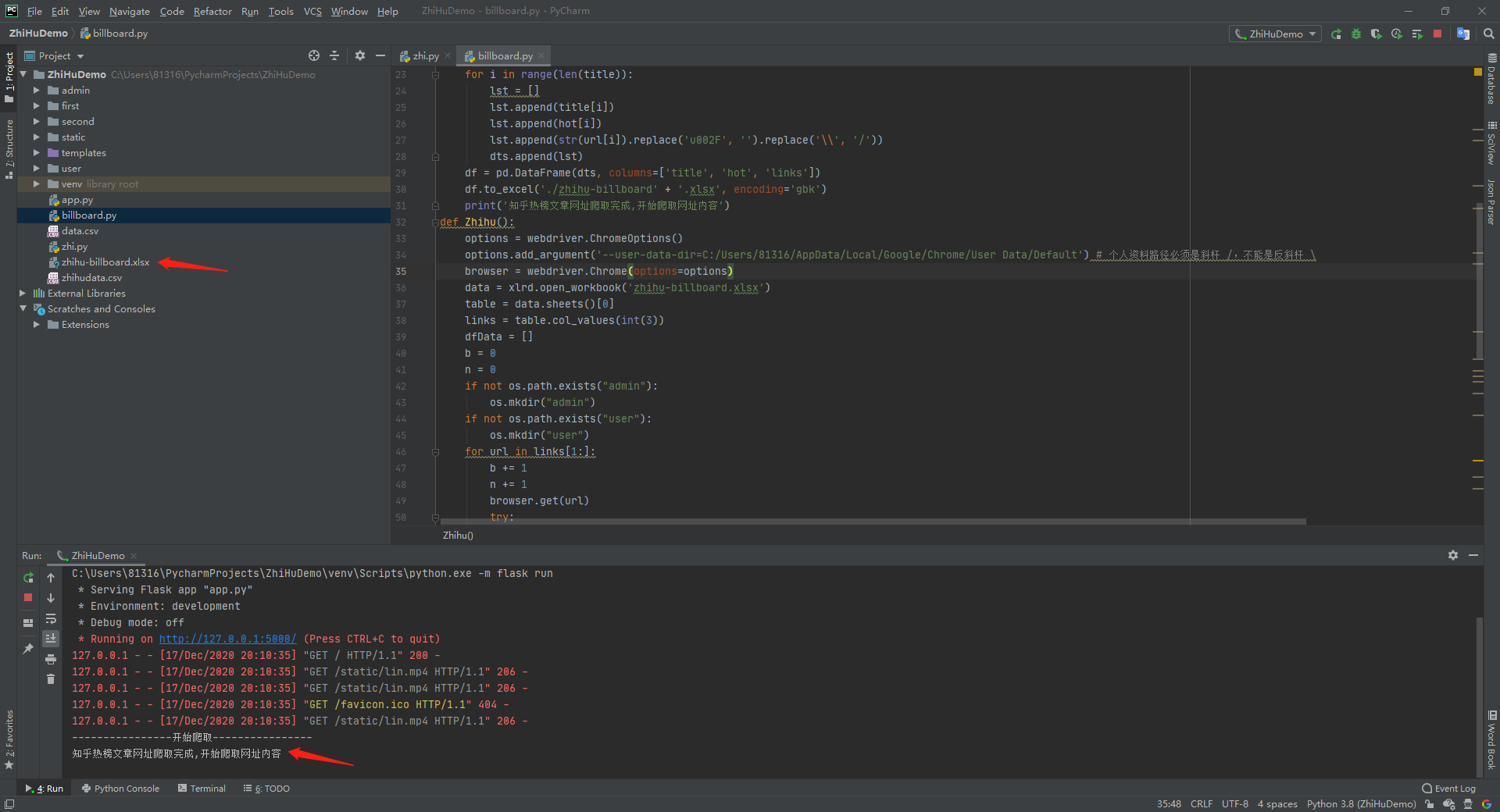
前端界面跳转到这个页面



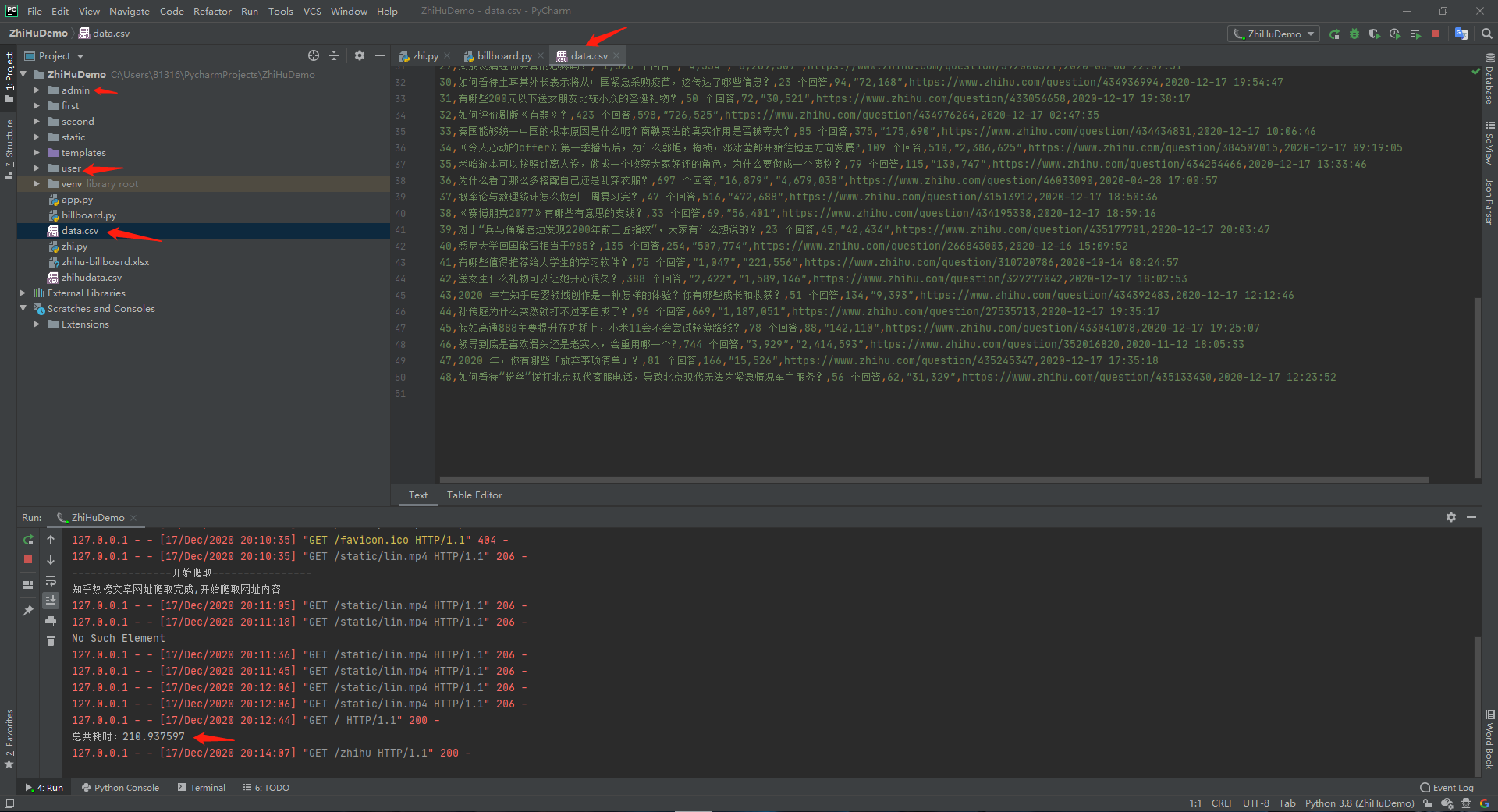
如果用户点击下面图片的按钮 “点击爬取知乎热榜数据”开始爬取 打开chrome浏览器



爬取知乎热榜文章网址爬取完成，生成zhihu-billboard.xlsx，爬虫自动读取zhihu-billboard.xlsx文件里面的网址



爬取完成后，终端可以看到全部爬完消耗了多少时间，生成admin文件夹存放知乎管理员锁定编辑截图,user文件夹存放某人添加了问题截图，生成data.csv存放爬取的数据



爬取完成后前端界面跳转到这个页面

