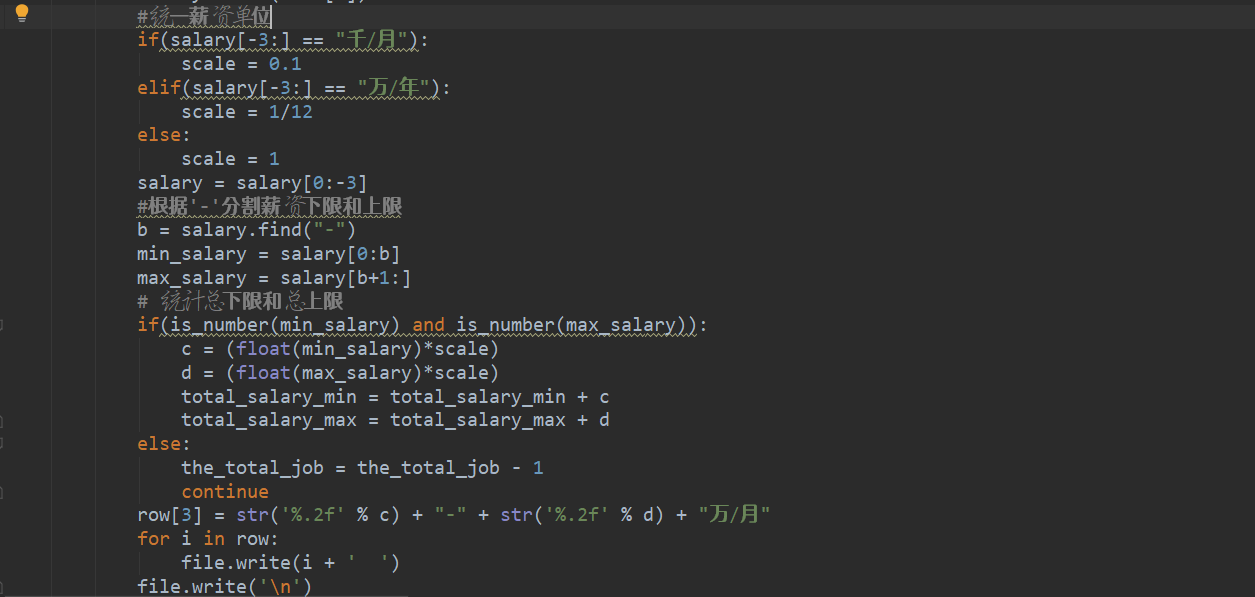
**51job网站爬取北京Python工程师**

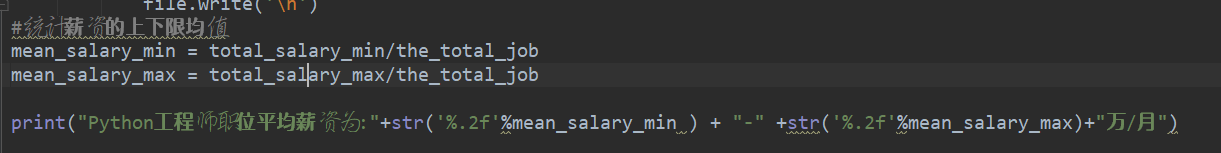
一．功能概述

在51job网站上，爬取2020年发布的Python开发工程师的职位的薪酬，计算北京地区改职位的平均薪酬；此题利用python的request发出网页请求，获取网页内容，利用chardet包获取网页字符编码，利用re和BeautifulSoup4的包，解析网页内容，提取关键字段信息。  
因为原始网页中，薪资栏的单位不统一，大部分是万/月，也有千/月，还有万/年，还有的会没标明薪资。  
所以为统计平均薪资，因为原始网页中给的是薪资范围，薪资下限和薪资上限。  
所以统计的时候也按照上下限统计的。平均下限，平均上限了统一薪资，设置了一个scale系数，统一成万/月。

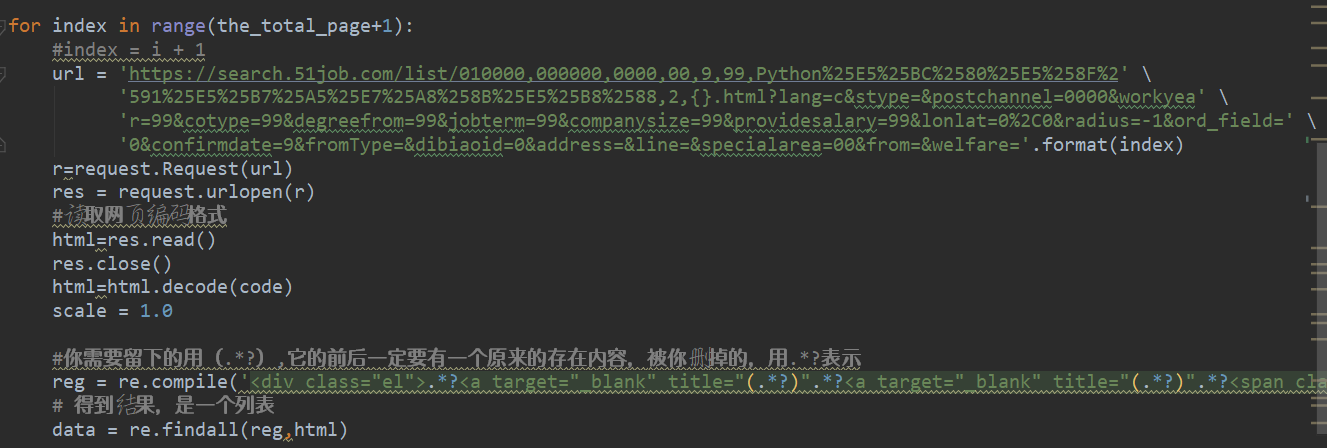
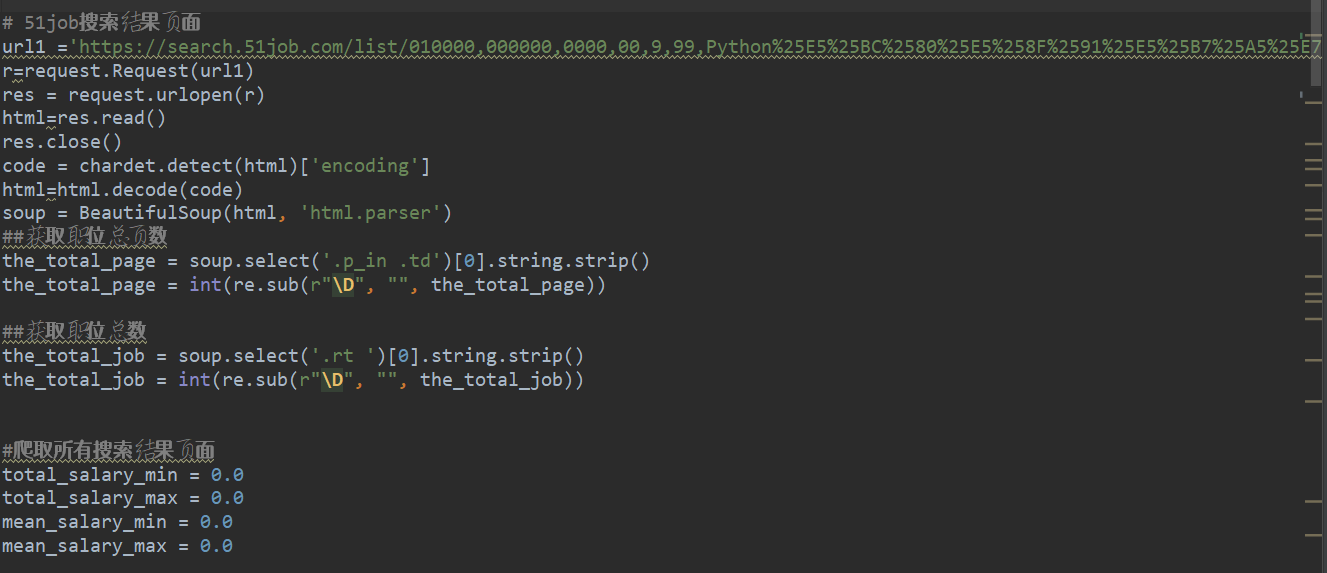
1. 关键函数
2. 统一薪资单位

由于网页中薪资单位不统一，设置了scale函数将全部薪资统一为万/月，便于计算。

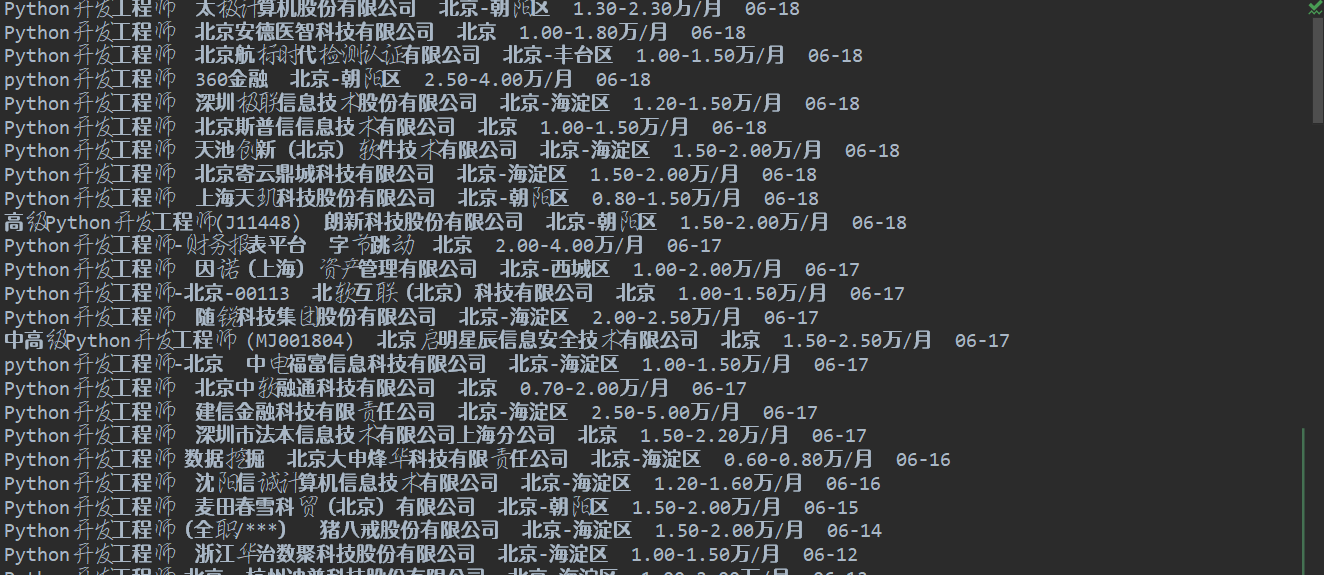
2.平均值统计

因为此网页给的都是工资的上限和工资的下限，所以最后求得的工资上限是由全部的工资上限平均值求得的，工资下限也是如此。

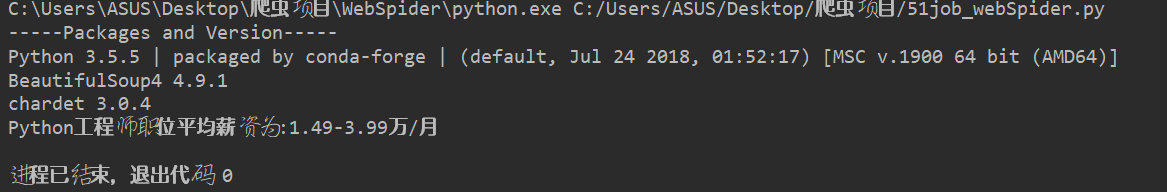
3.爬取数据



1. 数据保存

将爬取到的数据保存到51job.txt中，便于使用。

1. 运行结果



当前爬取算出的平均薪资为1.49-3.99万/月。