**郊野公园的社会影响分析**

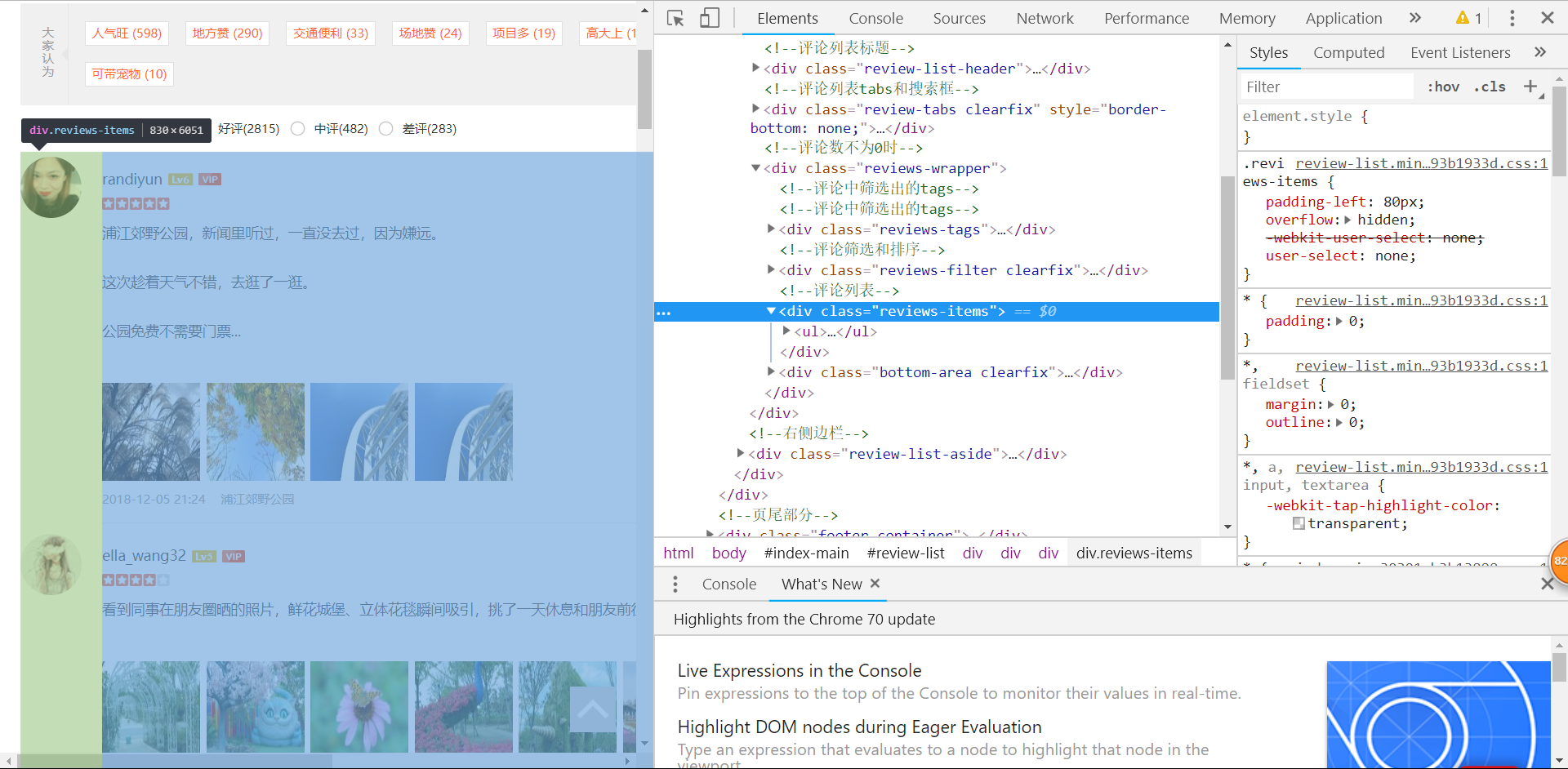
1. 社会影响分析概述

分析的意义 blablablabla

因此，我们决定通过获取目前市场上一个具有影响力的评价网站——大众点评上关于两家公园的用户评价数据，在进行进一步数据处理，最后通过可视化的方式展现郊野公园对于社会的影响力

1. 基于Python语言的数据抓取

为获取大众点评网站上关于两个公园的用户评价数据，我们初步分析了用户评论页面的网页结构，整理了数据采集的思路



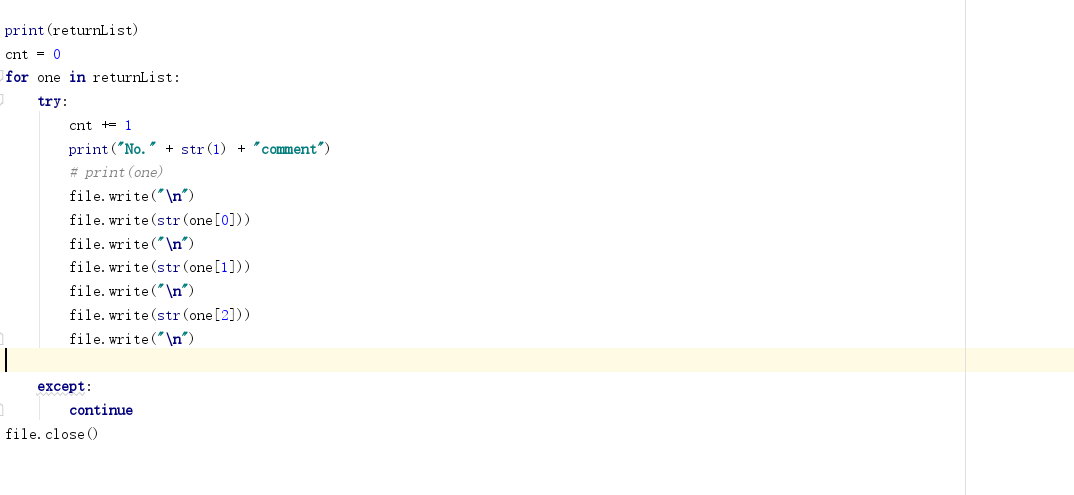
因为大众点评网站采取了优秀的反爬虫机制，在Python的代码中，我们尝试了借用IP代理池，模拟了用户访问



而后从网页结构入手，找到用户评论部分，伪装成浏览器访问进行数据采集



并且将数据写入txt文档保存在桌面



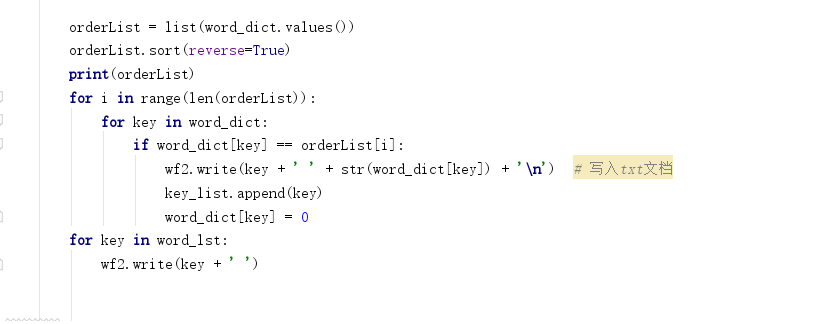


1. 数据的分词处理和过滤

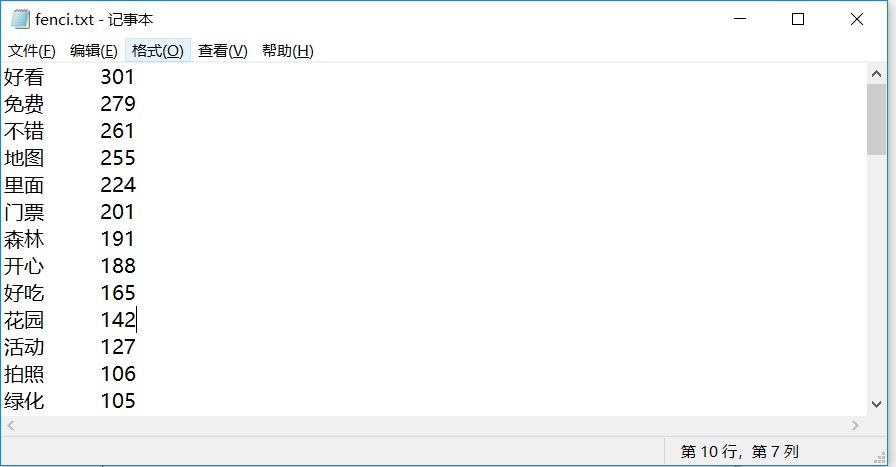
为了进行词频分析，首先要进行分词处理和过滤，分词的目的在于把完整的句子拆成可以表达用户观点情感态度反映社会影响的单个词语，而与此同时也要过滤掉在数据收集中获得的不可用的数据，无意义的词汇和因为数据爬虫时选择原因而导致的某些重复词汇（比如：在选择的评论reviews-items类中包括了收起评论这些会影响后续分析的词汇或者语句）

首先是进行分词，并进行词的归类



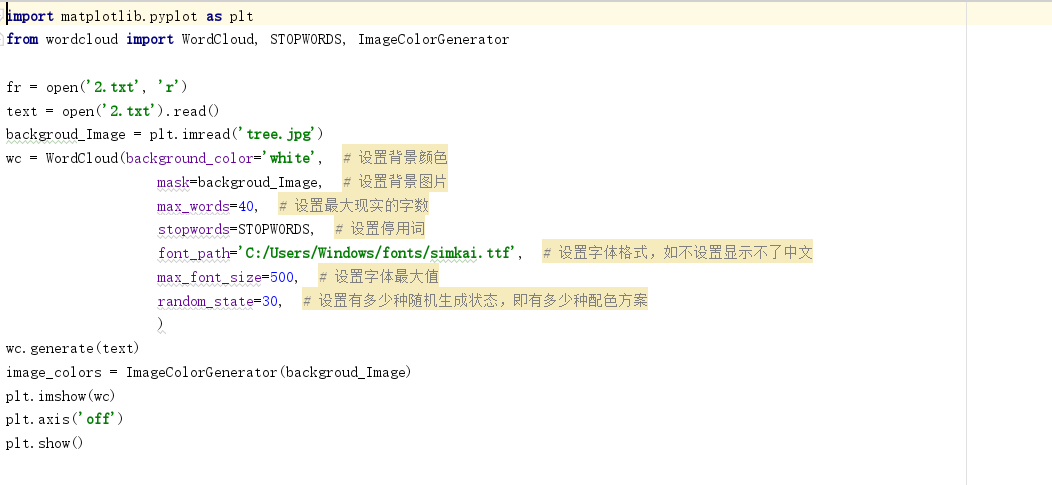


此时可以得到简单的词频分布文件并且进行一定的处理（去除如：公园，浦江等无意义或者会对于用户分析产生影响的词汇）



1. 基于词频分析技术的可视化处理（词云）

为了让结果更加直观，进一步可视化处理采用的是制作词云的方式，利用matplotlib库和wordcloud库将所有分词按词频制作出可视化词云



1. 结论