**租房通**

项目报告

**院 （系）： 软件与微电子学院**

**课 程： Python**

**小 组： 第21组**

**组 长： 查鑫池**

**组 员： 于子烜 唐龙基 单利剑**

**目录**

[一、项目简介 3](#_Toc26390389)

[二、项目流程介绍 3](#_Toc26390390)

[2.1市场调研 3](#_Toc26390391)

[2.2数据获取 3](#_Toc26390392)

[2.3数据处理 4](#_Toc26390393)

[2.4数据分析 4](#_Toc26390394)

[2.5编程实现 4](#_Toc26390395)

[三、项目功能介绍 5](#_Toc26390396)

[3.1初版项目 5](#_Toc26390397)

[3.1.1实现的功能 5](#_Toc26390398)

[3.1.2存在的缺陷 6](#_Toc26390399)

[3.2终版项目 7](#_Toc26390400)

[3.2.1改进和扩展的功能 7](#_Toc26390401)

[四、项目数据分析 10](#_Toc26390402)

[五、人员分工 11](#_Toc26390403)

[六、参考文档 11](#_Toc26390404)

# 一、项目简介

本项目为租房项目，旨在提供一种更直观、更便捷、多选择的选房模式，从而减少选房、看房的成本。在对地点检索、地图找房、地铁找房等多种找房模式进行调研后，我们最终认为：地图找房相对来说，是一种更高效、直观的选房方式，并且涵盖了其他选房模式的部分功能。因此，本项目基于地图找房的模式，提供价格筛选、房源筛选、路径规划等功能。

# 二、项目流程介绍

## 2.1市场调研

我们首先对主流的租房网站：赶集网、安居客、房天下、链家网、58同城进行了调研。了解并切身体验了它们地图找房、地铁找房、地点检索、学校找房等功能，最终认为：地图找房功能相比于其他找房功能，更加直观的体现了房源的所在位置，以及房源周围的设施、交通等情况。在功能上，地图找房的功能也涵盖了其他找房模式。因此，将地图找房作为本项目的模式。

详细内容见“市场调研报告”。

## 2.2数据获取

本项目通过编写三个爬虫文件，分别对58、链家和赶集网的房源数据进行爬取。58和链家的初始数据包含”title”，”location”，”price”，”url”四个字段；赶集的初始数据包含”title”，”location”，”price”，”url”，”source”五个字段。

爬虫文件以及获取的原始数据见“爬虫及数据”文件夹。

## 2.3数据处理

58和链家的原始数据新增了source字段，58的source字段值为“58”；链家的source字段值为“lj”；赶集原始数据自带source字段，值为“3”。

将三个来源的数据合并后，新增了lng（经度）和lat（纬度字段）。并将极个别无法获取经纬度信息的数据剔除，形成最终数据。将经纬度的信息直接写进数据中，会极大提升加载速度。

数据处理文件及中间数据，见“数据处理与分析”文件夹。

## 2.4数据分析

对合并后的数据进行离散化、分组处理，以及描述性统计分析。

数据分析文件，见“数据处理与分析”文件夹。

## 2.5编程实现

编程主要分为两部分，一部分是关于数据获取、处理和分析的编程，该部分主要用Python语言实现。其中主要涉及到BeautifulSoup、pandas等模块库的使用。

另一部分是关于页面的编程，该部分主要用到javascript和高德地图API。

# 三、项目功能介绍

## 3.1初版项目

### 3.1.1实现的功能

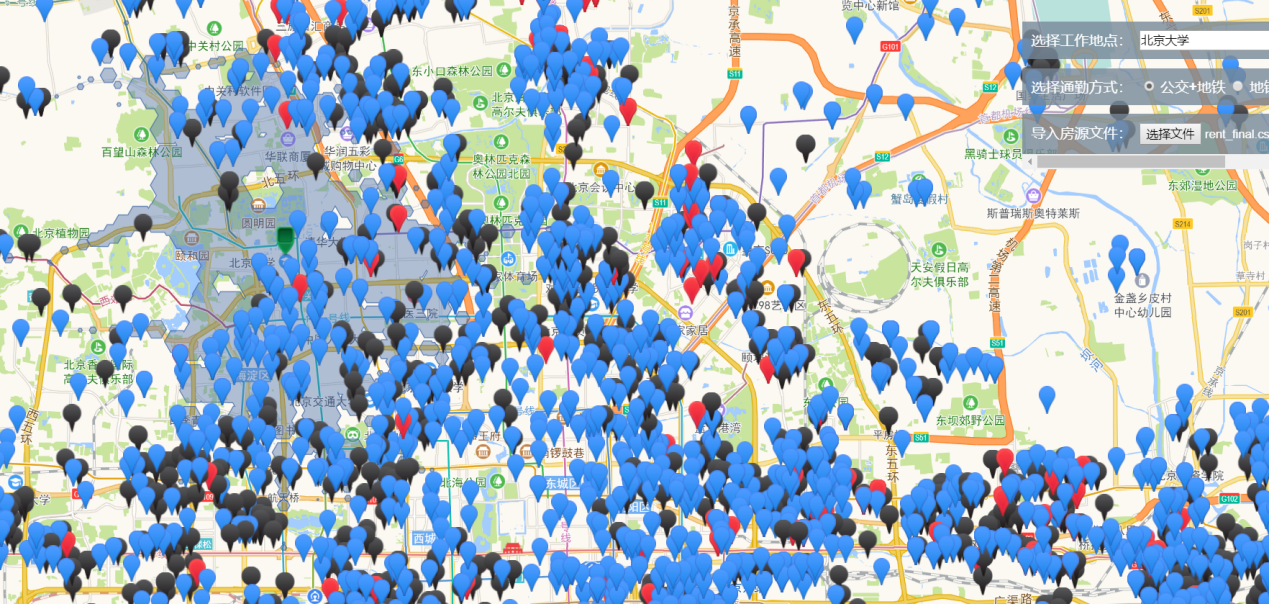
初版项目的演示视频见“[视频-初版功能演示.wmv](file:///F:\\CollegeTime_PKU\\python\\p\\project\\视频-初版功能演示.wmv)”。

根据已有的代码，我们首先制作了第一版租房通（rent\_pre.html）。第一版的租房通已实现的功能有：

①区别显示不同来源的房源

导入rent\_fianl.csv文件后，会加载所有房源的数据。

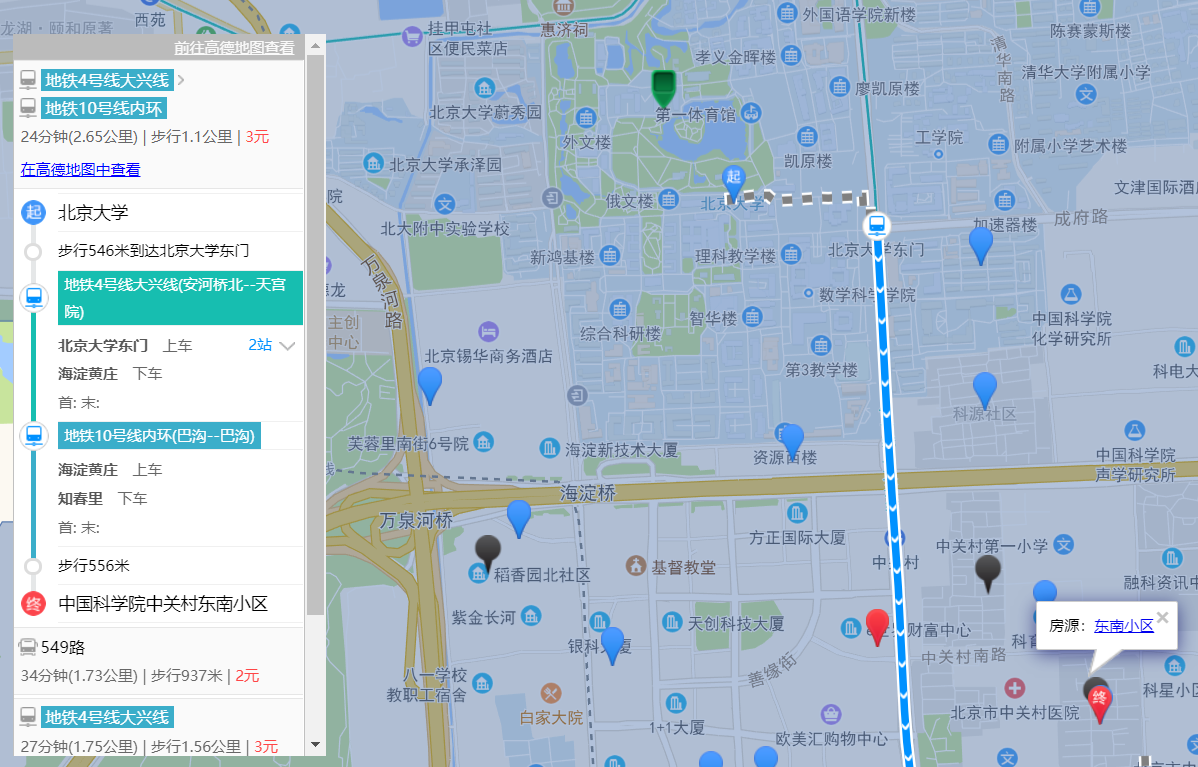
效果图如下图所示。绿色点标记代表工作地点，其余红、蓝、黑不同颜色的点标记分别代表来自58、赶集网和链家的数据。其中，工作地（绿色标记）周围的浅蓝色区域，表示当前通勤方式下，1小时内可到达的范围。



②根据工作地点和房源位置显示通勤方式

选中某个房源地点后，会显示工作地到该房源的通勤路径。

效果图如下所示。



③信息显示及租房链接跳转

在选中房源后，会显示房源的名称。点击房源的名称，可以跳转到相应的租房链接。



### 3.1.2存在的缺陷

从演示视频和实际操作中可以看出，虽然初版租房通已经实现了部分功能，但是存在一些缺陷：

第一，页面加载数据是逐个加载的，在数据量较大的时候，需要等待较长的加载时间。

第二，在房源数据比较多时，大量的点标记覆盖到了地图的信息，影响美观和获取地图信息。

第三，由于数据量过大，导致在对页面进行拖动、缩放、选择等操作时，会出现卡顿等情况，严重时，可能导致浏览器崩溃，影响体验。

## 3.2终版项目

### 3.2.1改进和扩展的功能

由于初版租房通存在一些缺陷，我们在功能上对其进行了改进或扩展。新的功能如下：

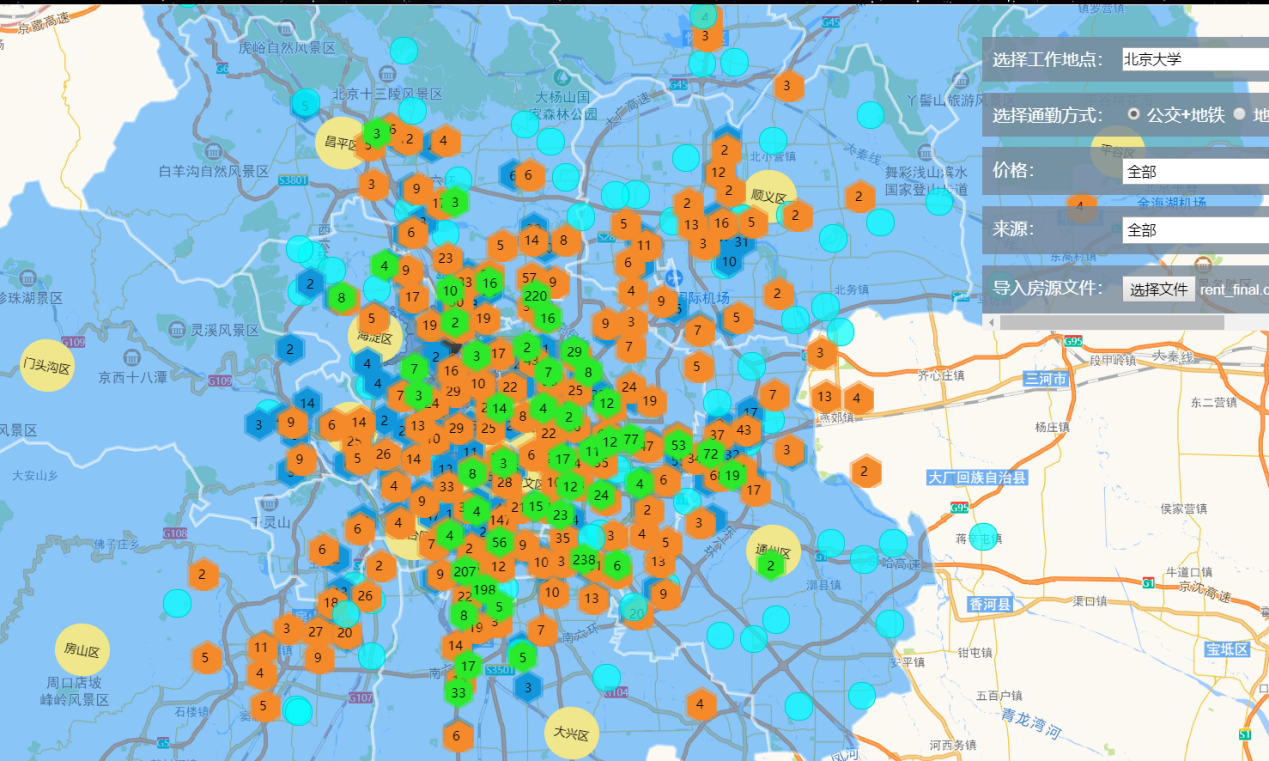
①区别显示不同来源的房源

演示视频：<视频-终版功能①.wmv>

该功能为初版项目的改进。效果图如下所示。

我们用蓝、橙、绿三种不同颜色的六边形代表链家、赶集、58不同房源的数据。并且，我们对不同房源添加了聚合功能，使得在比例尺增大时，点标记的数量会减少，降低了点标记对地图信息的遮掩，增加了美观和可读性。其中六边形中的数字，代表聚合的数量。蓝色的圆圈为相聚太远而无法聚合的单个房源点。此外，由于点数量的下降，在对页面进行拖动、缩放、选择等操作时，卡顿现象有所改善。

另外，我们修改了加载房源数据的方式，使得点的加载速度有了很大提升，减少了等待时间。



②根据工作地点和房源位置显示通勤方式

该功能与初版没有太大变化。

③信息显示及租房链接跳转

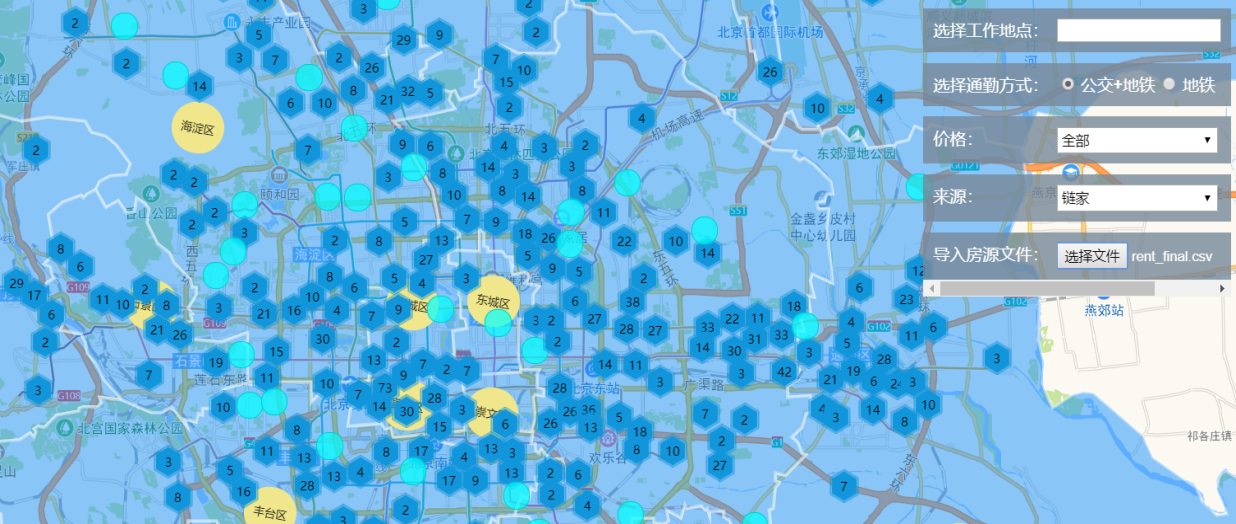
终版项目对该功能进行了扩展，使其可以提供更多的房源信息。



④房源筛选功能

演示视频：<视频-终版功能④.wmv>

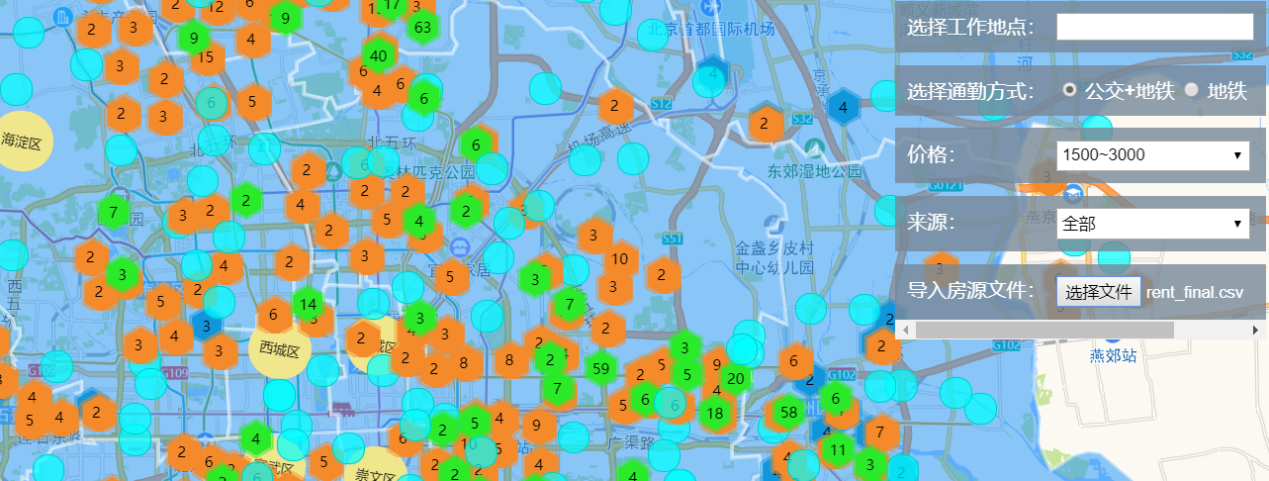
该功能为新增功能。在导入房源数据文件前，可以提前选择房源的来源，使其只显示特定来源的数据。效果如下图所示。



⑤价格筛选功能

演示视频：<视频-终版功能⑤.wmv>

该功能为新增功能。与房源筛选功能类似，在导入房源数据文件前，可以提前选择房源的价格区间，使其只显示特定价格区间的数据。效果如下图所示。



⑥其他改进

除了上述功能之外，我们还增加了北京市行政区的划分。整个北京市以浅蓝色覆盖，并且显示行政区之间边界，以此来辅助对租房位置的选择。

# 四、项目数据分析

由于项目对房源数据进行了收集和处理，因而我们顺便对这些数据做了简单分析。

①总体的描述性统计

描述性统计见下表。从表中可以得知：本数据集共有7576个房源。这些房源中，最便宜的月租420元，位置在玉泉营三环新城6号院；最贵的月租350000元，位置在西城西单四合院西琉璃厂街小区。均价为4600元左右；超过一半的房源其月租低于均价。

|  |  |
| --- | --- |
|  | **price** |
| **count** | 7,576.00 |
| **mean** | 4,588.73 |
| **std** | 9,107.98 |
| **min** | 420.00 |
| **25%** | 2,500.00 |
| **50%** | 3,500.00 |
| **75%** | 5,200.00 |
| **max** | 350,000.00 |

②各区域分析

房源数量最多的前十个区域如下：

|  |  |
| --- | --- |
| **位置** | **数量** |
| 西红门 | 342 |
| 亦庄 | 256 |
| 立水桥 | 140 |
| 顺义城 | 105 |
| 定福庄 | 93 |
| 双桥 | 83 |
| 高米店 | 79 |
| 管庄 | 76 |
| 通州北苑 | 73 |
| 朝阳 | 69 |

均价最低的区域前十如下：

|  |  |
| --- | --- |
| **address** | **mean price** |
| 南口 | 1,880.00 |
| 房山其它 | 2,000.00 |
| 温泉 | 2,156.67 |
| 丰台 | 2,170.00 |
| 良乡 | 2,209.20 |
| 长阳 | 2,262.67 |
| 龙泽 | 2,280.00 |
| 城子 | 2,340.00 |
| 八通轻轨沿线 | 2,350.00 |
| 石门营 | 2,366.67 |

③价格离散化处理

对所有房源按照价格分为三档，1500以下为“低”，1500~3000为“中”，3000以上为“高”。各价格区间的房源数量如下：

|  |  |
| --- | --- |
| **price\_grade** | **数量** |
| 低 | 390 |
| 中 | 2485 |
| 高 | 4701 |

# 五、人员分工

|  |  |
| --- | --- |
| **人员** | **分工** |
| 查鑫池 | 选题，统筹协作,PPT制作，项目汇报展示 |
| 于子烜 | 选题，数据获取、处理、分析，编程实现，撰写项目报告 |
| 唐龙基 | 选题，市场调研，辅助撰写项目报告 |
| 单利剑 | 选题，市场调研 |

# 六、参考文档

[1] <https://www.shiyanlou.com/courses/599/learning/>

[2] <https://lbs.amap.com/demo-center/js-api>

[3] <https://www.jianshu.com/p/ed841991e8e2>