**信息管理与信息系统专业课程设计（2020级）**

**×××分析报告**

**学生姓名：**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**学生班级：**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**指导教师：**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**黑龙江科技大学管理学院**

课程设计指导教师成绩评定书

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 评分标准：   |  |  | | --- | --- | | 评定项目 | 成绩 | | 1.选题是否符合专业培养要求、难易度、工作量是否适当（10分） |  | | 2.报告结构是否合理，分析四部分是否全面，有应用价值（10分） |  | | 3.数据挖掘、结果分析是否准确，文档格式是否规范（20分） |  | | 4.学生运用专业知识的熟练程度，分析解决问题的能力（10分） |  | | 5.分析成果是否有独到见解，是否有创新或亮点（10分） |  | | 总分（60） |  |   指导教师（签字）： 年 月 日 |

课程设计答辩及综合成绩评定书

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 评分标准：   |  |  | | --- | --- | | 评定项目 | 成绩 | | 1.叙述是否条理清晰，层次分明，重点突出。（10分） |  | | 2.回答问题是否全面准确，理解深刻，是否语言表达清晰、流利，逻辑性强。（15分） |  | | 3.是否很好地掌握了本专业的理论和知识，是否具有非常强的分析和解决实际问题的能力。（15分） |  | | 总分（40） |  |   综合成绩： 年 月 日 |

目 录

[第1章 绪论 1](#_Toc4890)

[1.1选题背景 1](#_Toc5600)

[1.2目的与意义 2](#_Toc2735)

[1.2.1分析目的 2](#_Toc29641)

[1.2.2分析意义 2](#_Toc25880)

[第2章 数据准备 3](#_Toc16842)

[2.1 数据采集（或数据获取） 3](#_Toc13956)

[2.1.1 数据来源 3](#_Toc8658)

[2.1.2 获取方法 3](#_Toc1448)

[2.2 数据整理 3](#_Toc8791)

[2.2.1整理工具 3](#_Toc1895)

[2.2.2 整理过程 3](#_Toc11217)

[2.2.3 整理结果 3](#_Toc29939)

[第3章 数据挖掘的模型构建与评估 4](#_Toc8923)

[3.1 理论基础 4](#_Toc14310)

[3.2模型构建 4](#_Toc11371)

[3.2.1 特征选择与数据探索性分析 4](#_Toc17434)

[3.2.2 算法选择与模型训练 4](#_Toc11461)

[3.3模型评估 4](#_Toc30027)

[第4章 可视化及结果分析 5](#_Toc7958)

[4.1 可视化需求分析 5](#_Toc14027)

[4.2 可视化工具选择 5](#_Toc4828)

[4.3可视化设计与展示 5](#_Toc11724)

[4.3.1 \*\*趋势分析 5](#_Toc17791)

[4.3.2 \*\*相关性分析 5](#_Toc8978)

[4.4 结论与建议 5](#_Toc16841)

[参考文献 6](#_Toc3842)

**第1章 绪论**

**1.1 选题背景**

近年来，二手房市场经历了一系列的变化。首先，从交易规模上来看，自2017年以来，国内二手房交易规模呈持续增长的趋势，但是却在2021年出现了首次同比下降。据过专业的测算，2021年全国二手房成交金额约为7.0万亿元，同比下降约6%；二手房成交套数约为393万套，同比下降约9%；成交面积约为3.6亿平方米，同比下跌。

在价格方面，2021年二手房市场呈现“前热后冷”的态势，政策“维稳”贯穿全年。15个代表城市在1-11月间的二手房成交面积为11034万平方米，同比增长3.2%。尽管成交量从4月的高峰开始连续下滑，但到11月已结束连跌态势，环比出现小幅回升。同时，百城二手住宅价格在1-11月累计上涨3.36%，上半年价格上涨较为稳定。

不过值得注意的是，不同城市的二手房市场表现各异。例如，深圳的二手房价格在过去一年中下跌了1.5%，这主要是因为深圳实施了二手房指导价制度，变相提高了首付比例并使价格不再公开，从而改变了客户的预期。相比之下，北京、上海和广州的二手房房价则呈现上涨趋势。

情理之中的是，哈尔滨的二手房市场在2023年也呈现了一定的波动。根据最新的数据，哈尔滨二手房的挂牌价格出现了下跌的趋势，从9385元变为9370元。同时，市场上的二手房数量也表现出上下震荡的情况，近三周在售二手房数量分别为90270套、90625套以及92200套。此外，哈尔滨二手房的价格变化也是十分频繁。以2022年1月为例，当时二手房的挂牌价格从9709元上涨到9732元。然而，随后的几个月中，哈尔滨二手房的价格环比下跌了0.80%。

所以不难看出，哈尔滨的二手房市场正处在一个调整期，价格和交易量都存在一定的波动。但房地产市场受多种因素影响，包括但不限于经济环境、政策法规、市场供需等，因此对于市场的预测需要综合考虑各种因素。例如教育资源的平均化和城镇化进程对学区房的优势产生影响。过去，购买学区房的主要目的是为了给孩子提供良好的学习环境。但随着不同学校之间的师资水平差异逐渐缩小，学区房的优势已不再明显。

这个选题的背景是为能更好的了解在全国二手房市场下哈尔滨二手房市场的现状。

**1.2 目的与意义**

**1.2.1 分析目的**

分析哈尔滨二手房的目的主要有以下几点：

1.了解市场行情：通过对哈尔滨二手房市场的挂牌价格、交易量等数据的观察和分析，可以了解当前哈尔滨二手房的市场行情。

2.研究房源分布：比如哪些地区的房源数量比较多，哪些地区的数量较少，以及不同地区的房源的平均面积和户型分布等信息。

3.预测房价走势：基于历史数据和当前的市场情况，可以尝试预测未来哈尔滨二手房的价格走势。

4.发掘潜力区域：通过分析各区域的二手房均价，可以看出哪些区域的二手房价格有上涨趋势，进而挖掘出具有投资潜力的区域。

综上所述，这样的分析更多的是为帮助购房者做出更明智的决策，同时也为投资者提供有价值的参考信息。

**1.2.2 分析意义**

对哈尔滨二手房分析的意义：首先，为购房者提供参考：通过了解哈尔滨二手房市场行情、房源分布等信息，想要购房客户或者意向者就可以更好地选择适合自己的房源，避免盲目购房，同时也为购房者提供了一个很好的了解渠道。其次，是为投资者提供决策依据：对于投资者来说，了解哈尔滨二手房市场的走势和潜力区域可以帮助他们做出更明智的投资决策，提高投资回报率。此外，促进房地产市场的健康发展：通过对哈尔滨二手房市场的分析，可以为政府部门制定房地产政策提供参考，有助于调控房地产市场，促进房地产市场健康稳定发展。然后，对城市规划和发展产生影响：分析哈尔滨二手房市场还可以反映出城市不同区域的居住需求和发展趋势，为城市将来的规划和建设提供有迹可循的信息。最后，有利于提高居民生活质量：通过对哈尔滨二手房市场的深入分析，可以发现二手房市场上存在的问题和不足，从而推动相关部门去改进服务，提高居民的住房满意度和生活质量。总之，这样的分析对购房者、投资者、政府部门以及城市规划和发展都具有重要意义，有助于促进房地产市场的健康发展，提高居民生活质量。

**第2章 数据准备**

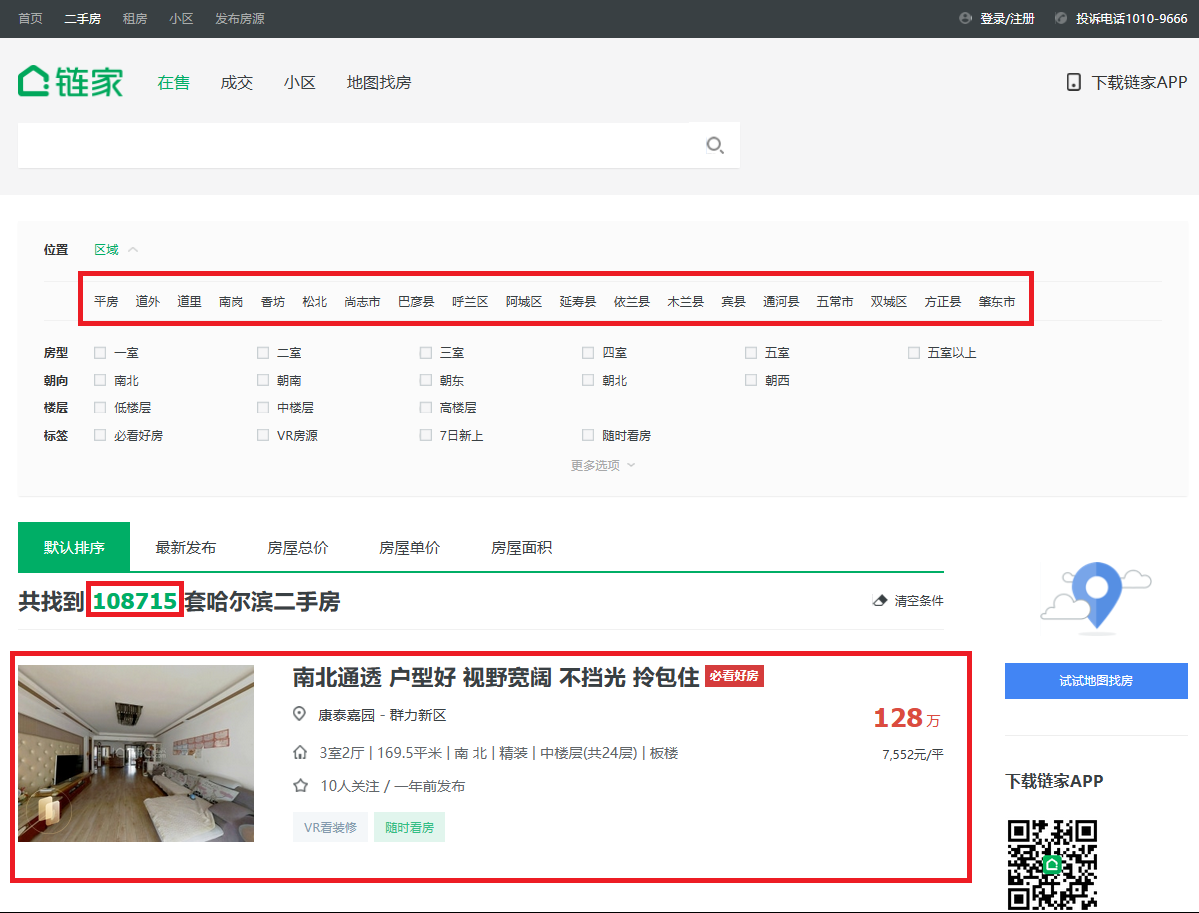
**2.1 数据采集**

通过基于Python的网络爬虫程序，用于爬取链家网站上哈尔滨市的二手房源数据，收集原始数据，作为整个数据分析的基石

**2.1.1 数据来源**

首先对链家网网站结构进行分析。

链家网二手房主页如图2-1所示，位于主页上方的红色方框位置显示目前哈尔滨二手房在售房源的各区域名称，中间的红色方框位置显示了房源的总数量，下方的红色方框为二手房源的信息略缩图，且包含了二手房房源页面的URL地址标签。图2-2中的红色方框显示了二手房主页上的房源页数。



**图 2-1**



**图 2-2**

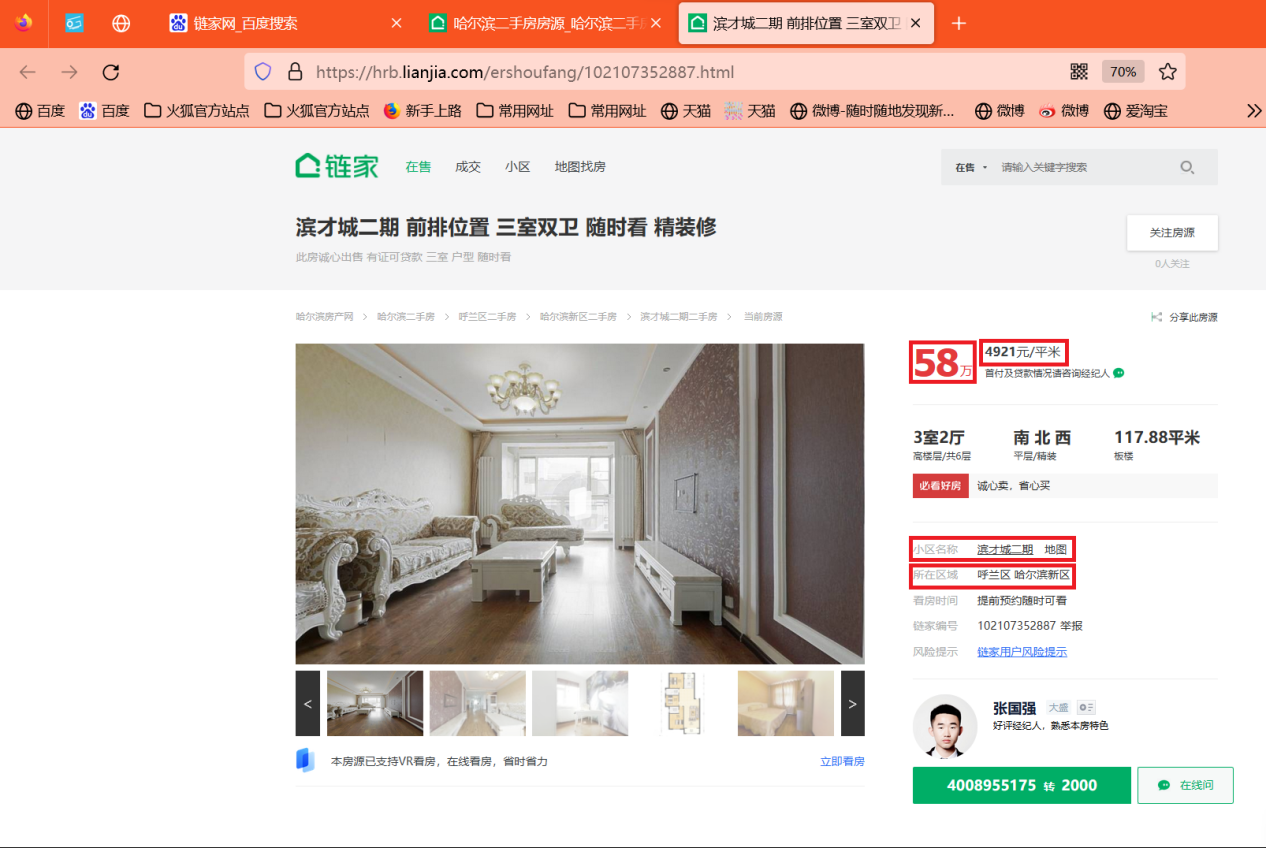
二手房源的信息页面如下，在图2-3、图2-4、图2-5中，含有我们需要采集的目标数据，包括基本信息、房屋属性、交易属性和房源特色四大类。各类信息包括的数据项如下：

基本信息：小区名称、所在区域、总价

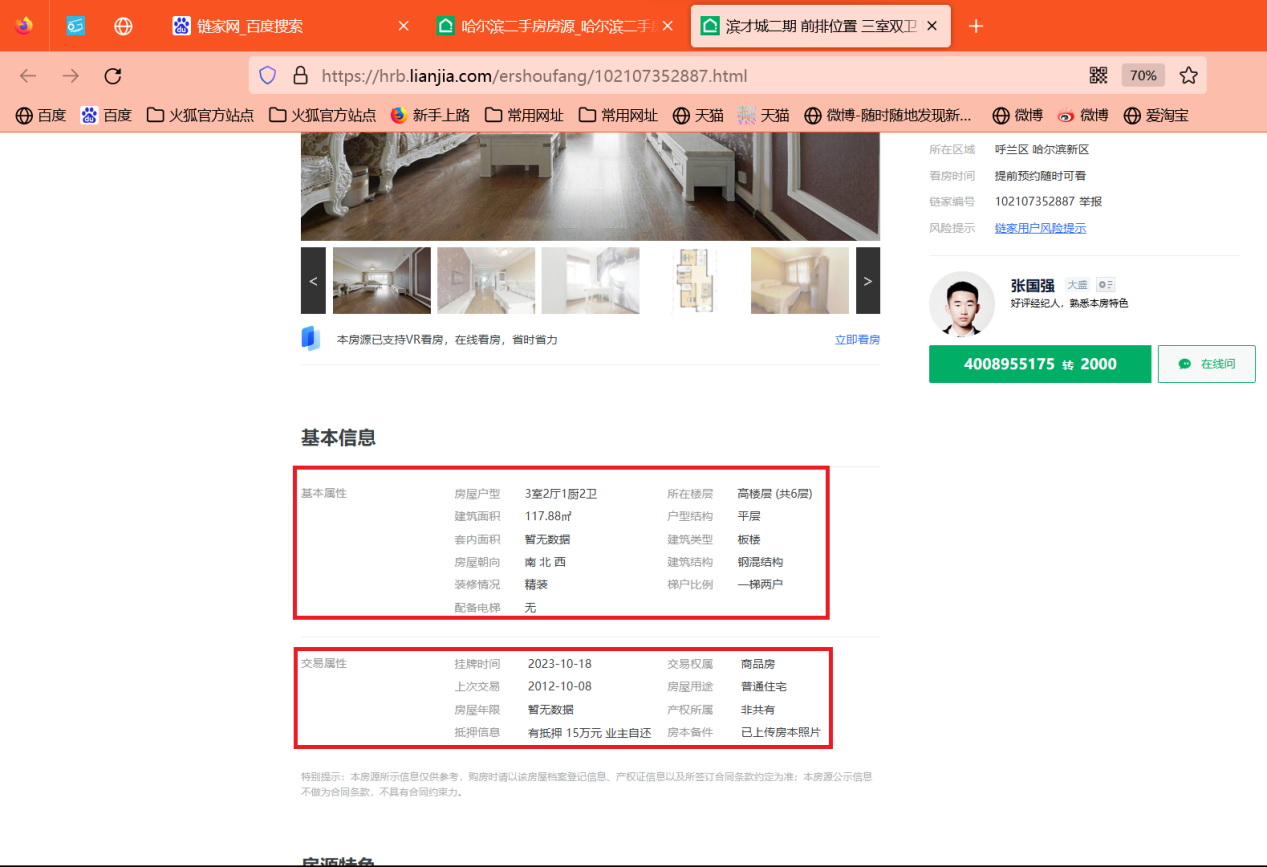
房屋属性：房屋户型、所在楼层、建筑面积、户型结构、套内面积、建筑类型、房屋朝向、建筑结构、装修情况、梯户比例、装配电梯、产权年限。

交易属性：挂牌时间、交易权属、上次交易、房屋用途、房屋年限、产权所属、抵押信息、房本备件

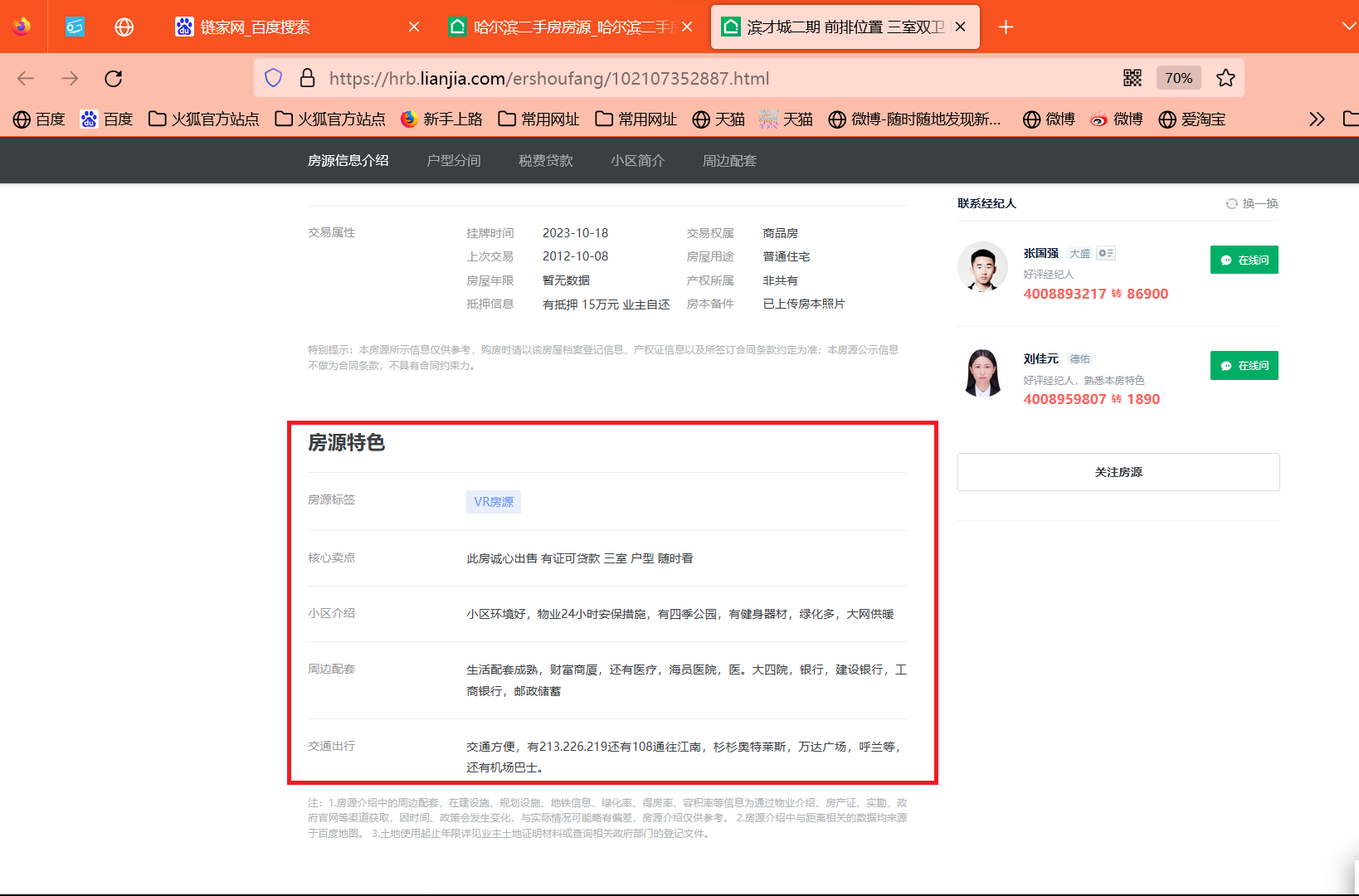
房源特色：房源标签、核心卖点、小区介绍、周边配置、交通出行



**图 2-3**



**图 2-4**



**图 2-5**

**2.1.2 获取方法**

基于Python写一个爬虫代码，对链家网"https://hrb.lianjia.com/ershoufang/+编号.html"实行数据爬取来获取房源信息。

如图2-6中，在主函数main中，程序分为以下几个模块：

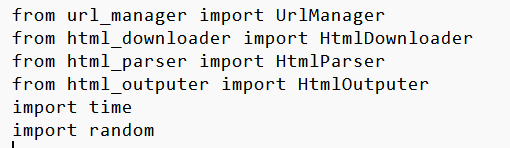
UrlManager：负责管理URL，包括添加新的URL和判断是否有新的URL。

HtmlDownloader：负责下载网页内容。

HtmlParser：负责解析网页内容，提取二手房信息。

HtmlOutputer：负责将提取到的二手房信息输出到文件。

SpiderMain：爬虫主类，负责整个爬虫流程的控制。

****

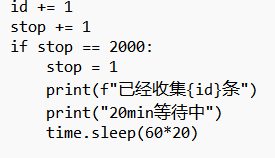
**图 2-6**

如图2-7，首先定义一个名为craw的函数，用来接收root\_url，这个表示链家网站的根URL。在这个函数中，定义一个字典areas，用于存储各个区域的页数。然后遍历这个字典，对于每个区域，构造对应的URL，下载网页内容，解析出房源URL，并将其添加到待爬取URL集合中。同时，将新的房源URL写入到文件new\_url.txt中。当前一个区域的页数爬取完后，会输出提示信息。接下来，程序进入一个循环，从待爬取URL集合中获取新的URL，下载网页内容，解析出房源数据，并将其添加到输出结果中。



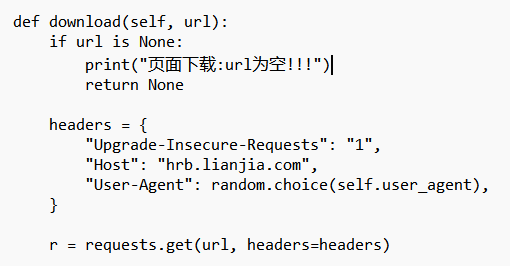
**图 2-7**

图2-8中，为了防止爬虫程序运行过快而触发链家网的反爬虫机制，从而导致后面的数据爬取失败，于是有了以下解决措施：每get到2000条数据时，就将程序暂停20分钟，以此绕过反爬虫机制。至于暂停时间，可以在实际操作中课慢慢调试，找出合适的爬取请求速度，提高程序运行效率。



**图 2-8**

图2-9为负责下载网页内容的部分代码，其中使用到的init方法：初始化类的实例，即设置一个随机的User-Agent列表，用于模拟不同的浏览器访问。还有download方法：是为了接收一个URL参数，然后使用requests库发送get请求获取网页内容。在请求头中设置了"Upgrade-Insecure-Requests"和"Host"字段，以及从User-Agent列表中随机选择一个User-Agent。如果请求成功（状态码为200），则返回网页的HTML内容；否则，打印错误信息并返回None

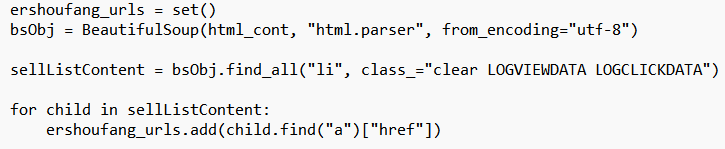
****

**图 2-9**

图2-10为用于解析HTML页面中的房子信息的部分代码，里面主要包含了两个方法：get\_ershoufang\_data和get\_erhoufang\_urls。

get\_ershoufang\_data是为了接收两个参数：html\_cont（HTML页面内容）和id（房子的唯一标识）。该方法首先检查html\_cont是否为空，如果为空则打印提示信息并返回。然后使用BeautifulSoup库解析HTML页面，提取房子的相关信息，如社区名称、区域名称、总价和单价等。最后将这些信息存储在一个列表中并返回。

get\_erhoufang\_urls是为了接收一个参数：html\_cont（HTML页面内容）。该方法首先检查html\_cont是否为空，如果为空则打印提示信息并返回。然后使用BeautifulSoup库解析HTML页面，提取所有房子的链接，并将这些链接存储在一个集合中返回。



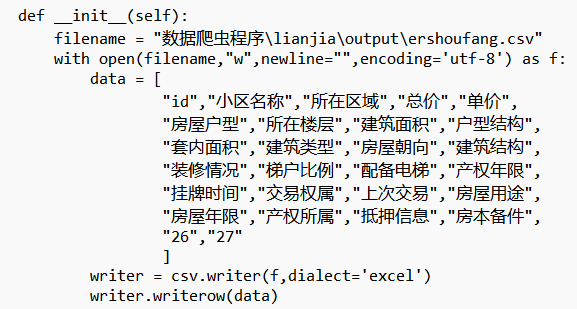
**图 2-10**

图2-11为用于将爬取到的房屋信息数据存储到csv文件的部分代码。它也包含了两个方法：前面提到过得init和collect\_data。

init方法在这里的作用是：在类的实例化时执行，用于初始化csv文件并写入表头信息。首先要定义一个文件名filename，然后以写入模式打开这个文件。然后定义一个二维列表data，里面包含csv文件的表头信息。最后，使用csv.writer对象将表头信息写入文件，并关闭文件。

collect\_data方法：用于将爬取到的房屋信息数据追加到csv文件中。首先判断传入的数据是否为空，如果为空则打印提示信息并返回。然后，定义了与init方法中相同的文件名filename，并以追加模式打开这个文件。然后使用csv.writer对象将传入的数据追加到文件中，并关闭文件。

因此，通过爬虫得到的原始数据就储存在ershoufang.csv。

****

**图 2-11**

**2.2 数据整理**

**2.2.1 整理工具**

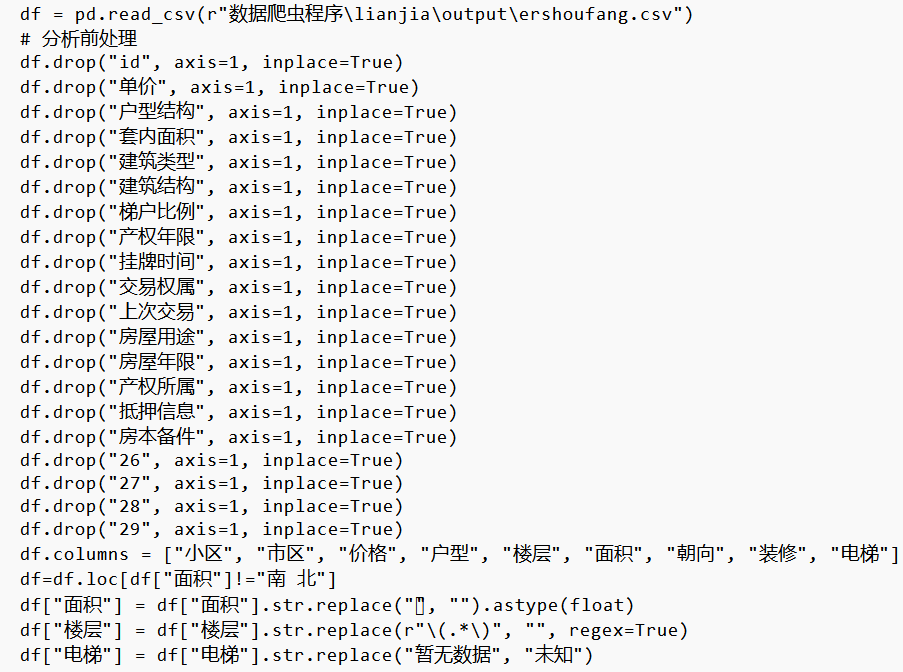
使用传统的统计学软件（Excel）进行数据整理工作，Microsoft Excel是Microsoft为使用Windows和Apple Macintosh操作系统的电脑编写的一款电子表格软件。直观的界面、出色的计算功能和图表工具，再加上成功的市场营销，使得Excel成为最流行的个人计算机数据处理软件。

使用Python进行大数据清洗工作，Pandas是Python中很流行的类库，使用它可以进行数据科学计算和数据分析，Pandas提供功能强大的类库,不管数据处于什么状态，他可以帮助我们通过清洗数据，排序数据，最后得到清晰明 了的数据。除此以外，主要涉及应用的库还有numpy，matplotlib等。

**2.2.2 整理过程**

（1）使用Excel对表格进行样式调整，删除多余的字段。

（2）使用pandas库的read\_csv函数读取csv文件，并将其存储在变量df中。然后，对数据进行预处理，删除不需要的列（如"id"、"单价"等），并将剩余的列重命名为更具描述性的名称（如"小区"、"市区"等）。接下来，对数据进行筛选，保留面积不为"南 北"的行。最后，对某些列进行处理，如将"面积"列中的单位去掉并转换为浮点数，将"楼层"列中的括号去掉，将"电梯"列中的"暂无数据"替换为"未知"。

**图 2-12 数据清洗**

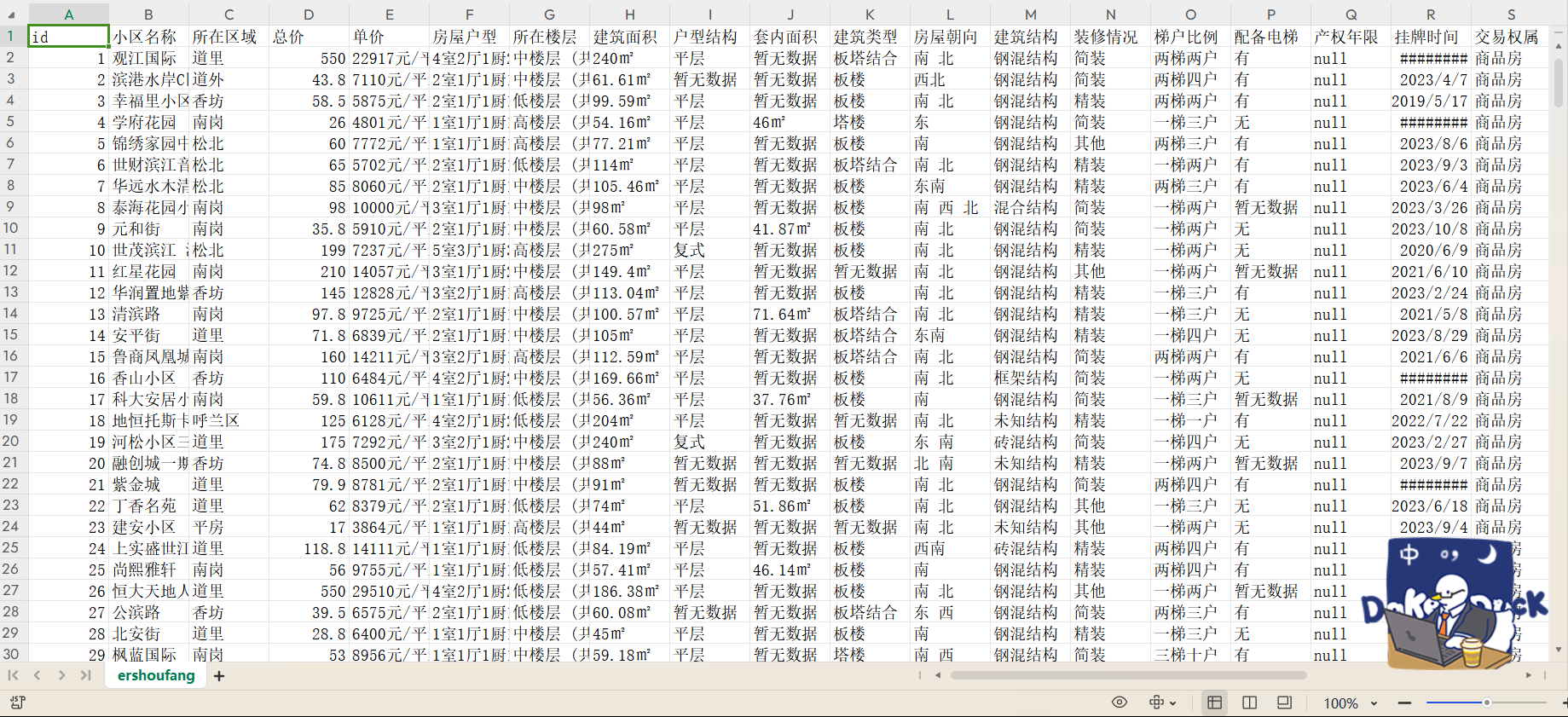
（3）在Python中使用info函数来查看缺失值，如图2-6所示可以看出哪些数据存在缺失值，如某些二手房未填有无电梯、是否装修等存在缺失值。



**图 2-13 info查看缺失值**

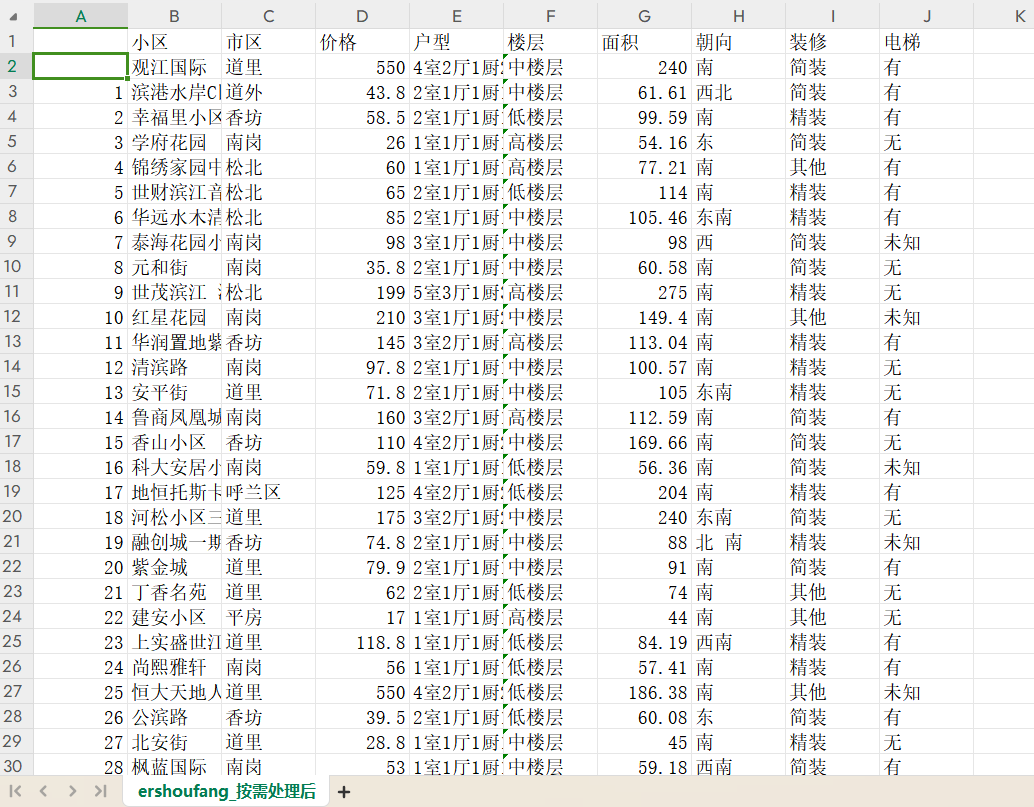
**2.2.3 整理结果**

如图2-7是经过爬虫获取的二手房原始数据，且未经过筛选。



**图 2-14 整理前的部分csv文件**

在经过最后的整理筛选中，共获取如图2-8中的小区、市区、价格、户型、面积、朝向、装修、电梯8个字段。

**图 2-15 整理后的部分csv文件**

**第3章 数据挖掘的模型构建与评估**

**3.1 理论基础**

我这里使用的是多元回归模型。  
 回归分析是一种用于研究变量之间关系的统计方法，它可以是线性回归、非线性回归或多项式回归等。回归分析的目标是寻找自变量与因变量之间的数学关系，以预测或解释因变量的变化。简单回归分析通常包括一个自变量和一个因变量，而多元回归分析涉及多个自变量对一个或多个因变量的影响。回归分析可以用于建立线性或非线性关系，但通常假定因变量与自变量之间存在某种形式的关联。

多元回归模型是回归分析的一种特定形式，它涉及多个自变量对一个或多个因变量的影响。在多元回归中，我们可以考虑多个自变量同时对因变量的影响，从而更全面地解释因变量的变化。多元回归可以是线性多元回归，其中自变量与因变量之间是线性关系，也可以是非线性多元回归，其中关系是非线性的。多元回归模型通常以以下形式表示：Y = β0 + β1X1 + β2X2 + ... + βnXn，其中Y是因变量，X1、X2、...、Xn是自变量，β0、β1、β2、...、βn是回归系数。

**3.2 模型搭建**

数据收集阶段是数据分析的第一步，它涉及到数据的采集、调查、实验或数据集的获取。在这个阶段，我们需要确保所收集到的数据是可靠的，具有代表性，并且包含了因变量和自变量。

接下来是数据探索与预处理阶段。在这一阶段，我们会进行探索性数据分析（EDA），以深入了解数据的特征和关系。这包括查看数据的分布情况，探索特征之间的相关性，检查异常值的存在，处理缺失值以及进行数据的转换。通过EDA，我们可以为后续的建模工作提供基础。

特征工程是数据分析中的重要环节。在这个阶段，我们可能需要选择、构建或转换自变量。特征选择是一个关键任务，因为不是所有自变量都对因变量有影响，而且过多的自变量可能会消耗计算资源。在特征工程中，我们可能需要进行特征缩放、创建交互项、进行独热编码等操作。

数据拆分是将数据集划分为训练集和测试集的过程。训练集用于建立模型，而测试集则用于评估模型的性能。通过将数据集拆分为训练集和测试集，我们可以更好地评估模型的泛化能力。

在选择回归模型时，我们需要根据问题的性质和数据的特点来选择合适的模型。常见的回归模型包括线性回归、多项式回归、岭回归、Lasso回归、决策树回归、随机森林回归等。选择正确的模型是成功的第一步。

模型拟合阶段使用训练集来拟合所选的回归模型，即估计回归系数。这一过程涉及寻找最适合训练数据的模型参数。

**3.3 模型评估**

最后是模型评估阶段。在这一阶段，我们使用测试集来评估模型的性能。常见的评估指标包括均方根误差（RMSE）、决定系数（R-squared）、平均绝对误差（MAE）等。通过评估结果，我们可以确定模型的准确性和适应性。

**第4章 可视化及结果分析**

**4.1 可视化需求分析**

分析爬取的房源信息，判断购房者购买需求，了解购房者对二手房的看法。使用绘图工具，绘制出二手房的楼层、房屋朝向占比、房屋的面积、装修情况等消费者购买在购房时在意因素的各项占比。

**4.2 可视化工具选择**

Pyecharts库是一个基于Web技术的Python数据可视化库，它提供了丰富的图表类型，使数据分析和展示变得更加直观和易于理解。与Jupyter Notebook相结合使用，可以轻松创建交互式数据可视化文档。以下是Pyecharts库和Matplotlib库的一些主要特点：

（1）Pyecharts库：Pyecharts库允许创建各种类型的图表，例如折线图、柱状图、饼图、地图等，以此来展示数据的不同方面。

（2）与Jupyter Notebook集成：可以在Jupyter Notebook中使用Pyecharts库来创建图表，不需要离开笔记本环境，这使得数据分析更加流畅和直观。

（3）富媒体展示：Pyecharts可以生成富媒体格式的图表，包括HTML，图片和SVG等。这样就可以在文档中嵌入漂亮的图表，进而更好地传达数据的含义。

（4）Markdown支持：与Jupyter Notebook一样，Pyecharts库也支持Markdown语法，这就意味着可以直接在图表的旁边添加解释性的文本，帮助读者更进一步理解数据。

（5）互动性：Pyecharts库允许你创建交互式图表，通过鼠标悬停和点击等方式与数据进行互动，进一步增强了数据分析的效果。

（6）Matplotlib库：Matplotlib库是Python可视化程序库的领导者，经过数十年的发展，它仍然是python使用者最常用的话图库。它用于生成高质量的图形，包括线图、散点图、直方图等。它提供了广泛的自定义选项，以确保使用者可以创建满足特定需求的图形。

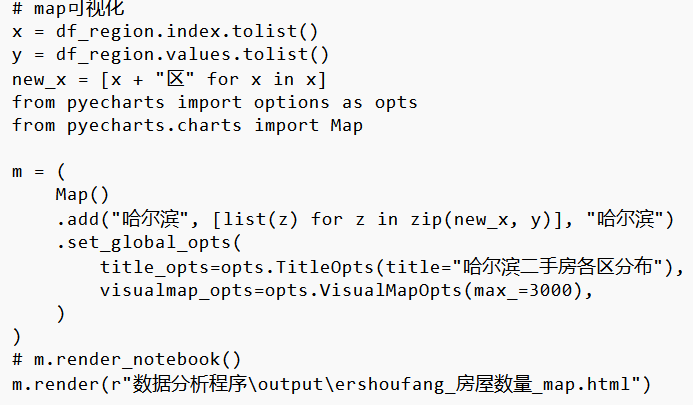
（7）满足大部分绘图需求：Matplotlib和Seaborn库通常能够满足数据分析中的绝大多数绘图需求，无论是探索性数据分析还是最终的结果展示。

Pyecharts库和Matplotlib库都是数据可视化的有力工具，可以在数据分析和报告中起到关键作用。Pyecharts库特别适用于创建交互式和富媒体的图表，而Matplotlib和Seaborn库则适合生成各种类型的静态图形。

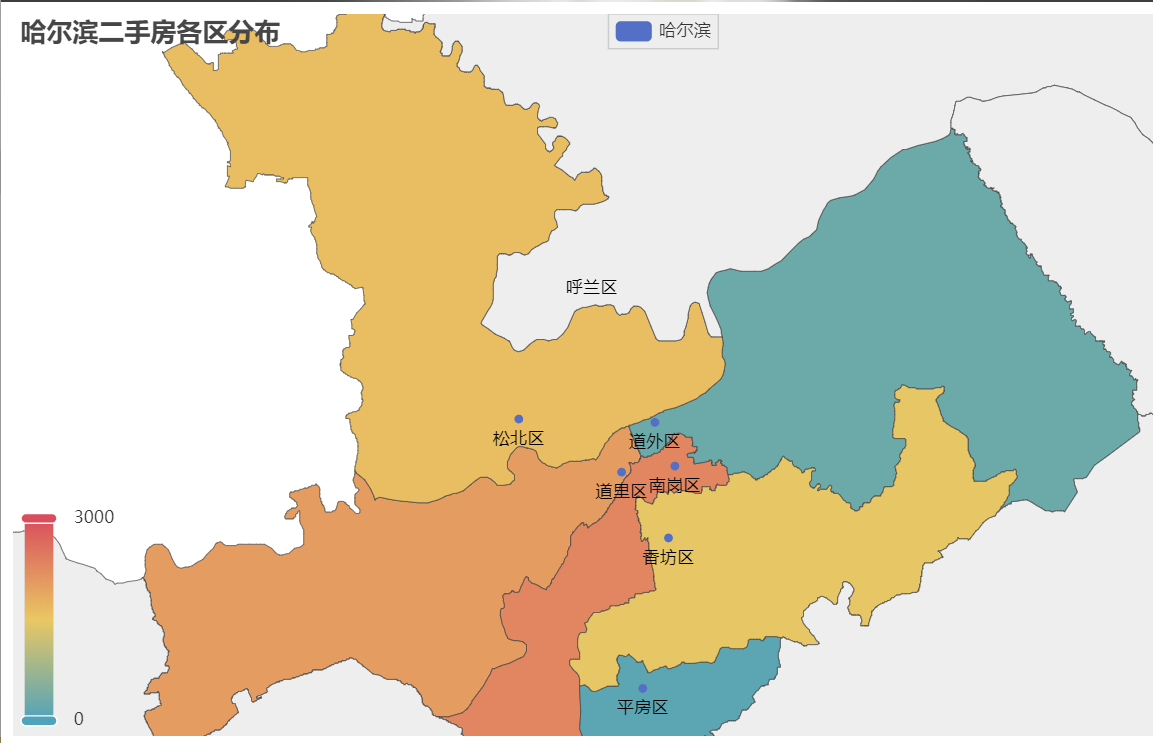
**4.3 可视化设计与展示**

**4.3.1 数量分布**

如下图4-1，使用了Matplotlib库，绘制出哈尔滨二手房数量各区分布图，其中，南岗区和道里区的二手房数量都属于遥遥领先的地区，结合2021年哈尔滨市最富有十大区县中排名中，南岗区以1029.80亿元位居榜首，道里区以718.90亿元紧随其后，可以得出的结论是，在这经济雄厚的两个区，而且是二手房的高端市场，其房源库存量是非常大的。



**图 4-1 哈尔滨二手房各区分绘制代码**



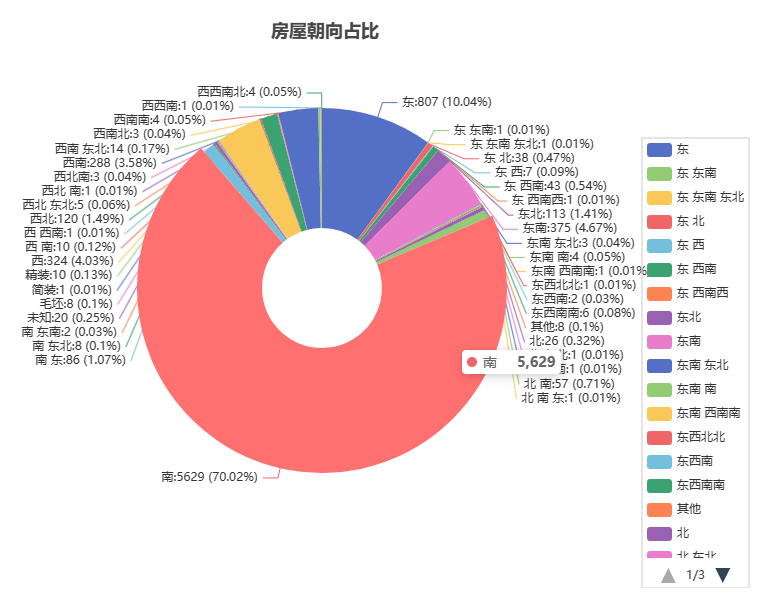
**图4-2 哈尔滨二手房各区分布图**

**4.3.2 房屋朝向占比**

如图4-4，可以看出二手房的房屋朝向占比中，南朝向的占比达到了70.02%，其次是东朝向，占比达到了10.04%。从哈尔滨市的地理方面出发，首先，朝南的房子可以接受更多的阳光照射，这对于提高房间的温度和采光度非常有利。其次，哈尔滨属于高纬度地区，冬季比较长且寒冷，因此充足的阳光对于提高室内温度和生活质量非常重要。最后，朝南的房子还可以避免北风的侵袭，从而使得房间更加温暖舒适。所以结合实际生活的角度出发，朝南的房子在哈尔滨市场上更受欢迎。



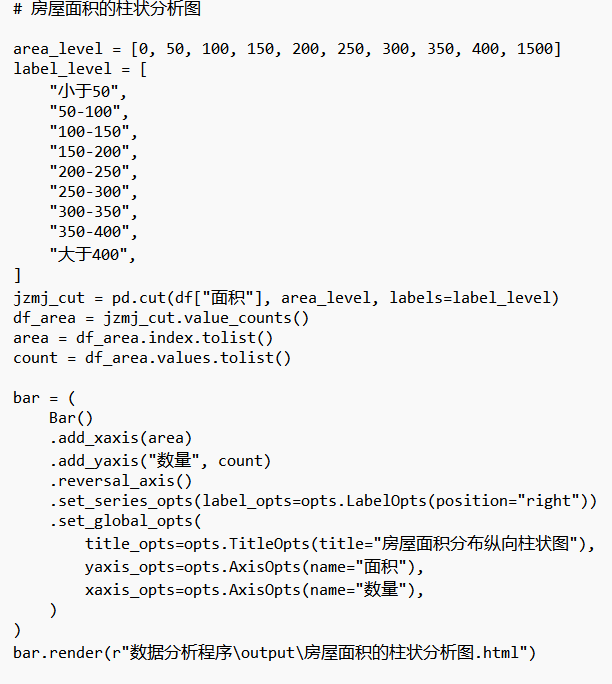
**图 4-3 房屋朝向饼图绘制代码**



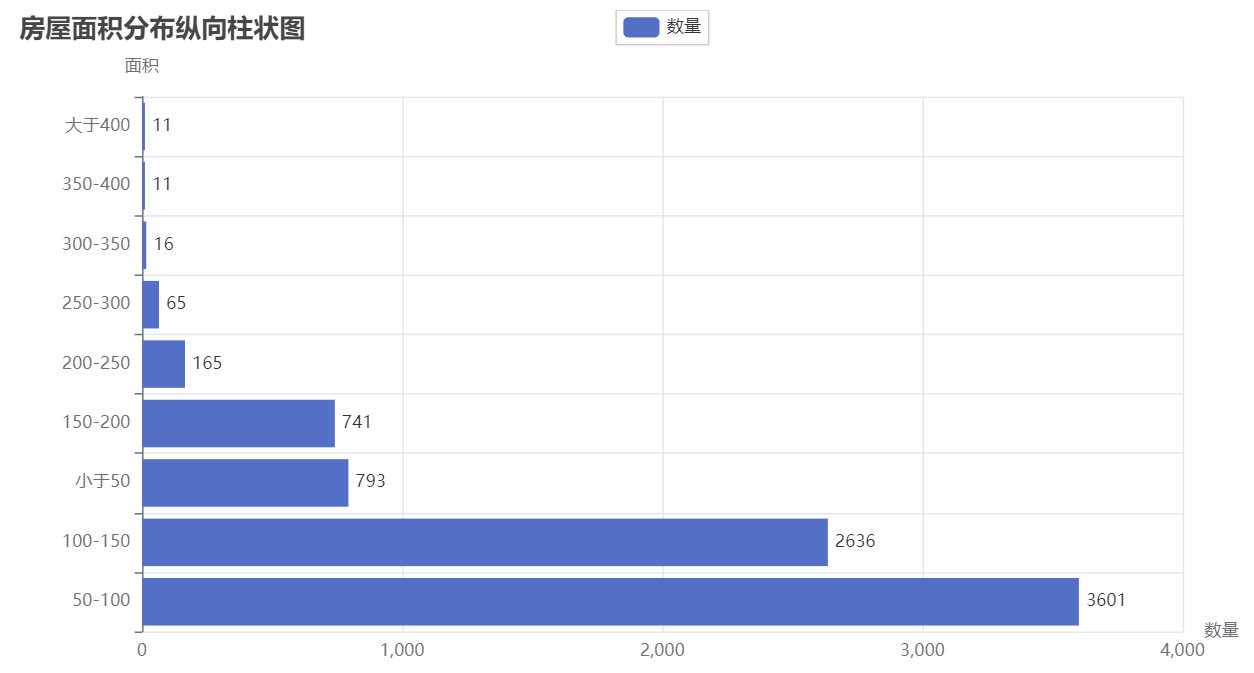
**图 4-4 房屋朝向占比饼图**

**4.3.3 房屋面积**

如下图4-6，可以了解到，在哈尔滨二手房的房屋面积分布纵向柱状图中，50-100平方米的数量占据了将近一半的数量，从哈尔滨当地的二手房源了解到，两室一厅占据了主导地位，当然还有一个因素就是房屋面积与售价成正比，房屋面积越大，售价越高，所以人们都倾向于购买面积比较均衡的房子。



**图 4-5 房屋面积的柱状分析图绘制代码**



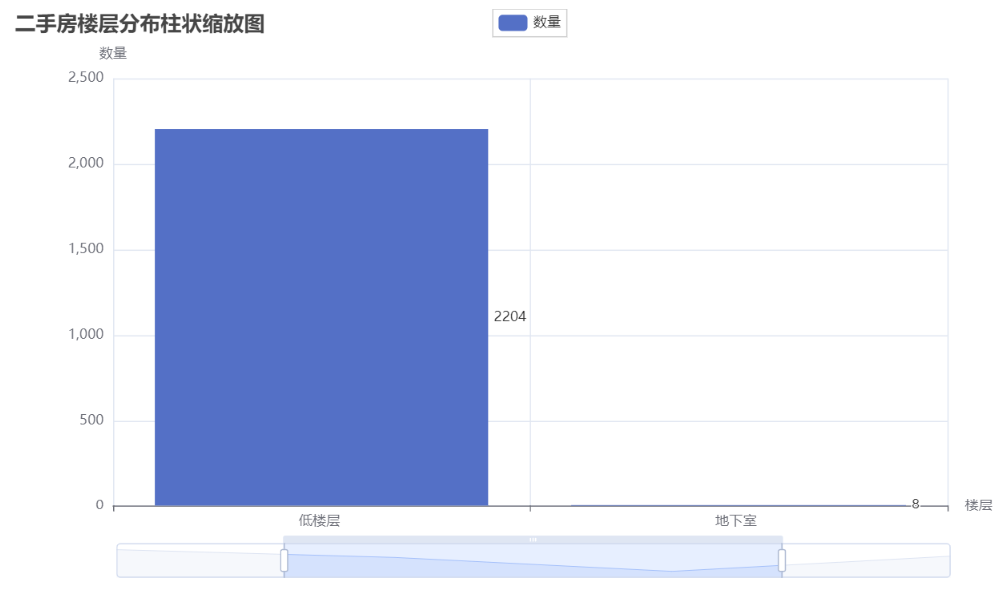
**图4-6 房屋面积分布纵向柱状图**

**4.3.4 楼层分布**

下图4-8所见，很多的二手房源都是低楼层，具体来说，6层以下的房子数量最多，其次是6-12层的房子，而12层以上的房子数量相对较少。这可能与哈尔滨市的建筑结构有关，因为大部分住宅楼都不超过12层。



**图 4-7 二手楼层分布柱状缩放图绘制代码**

**图 4-8 二手楼层分布柱状缩放图**

**4.3.5 有无电梯有与装修**

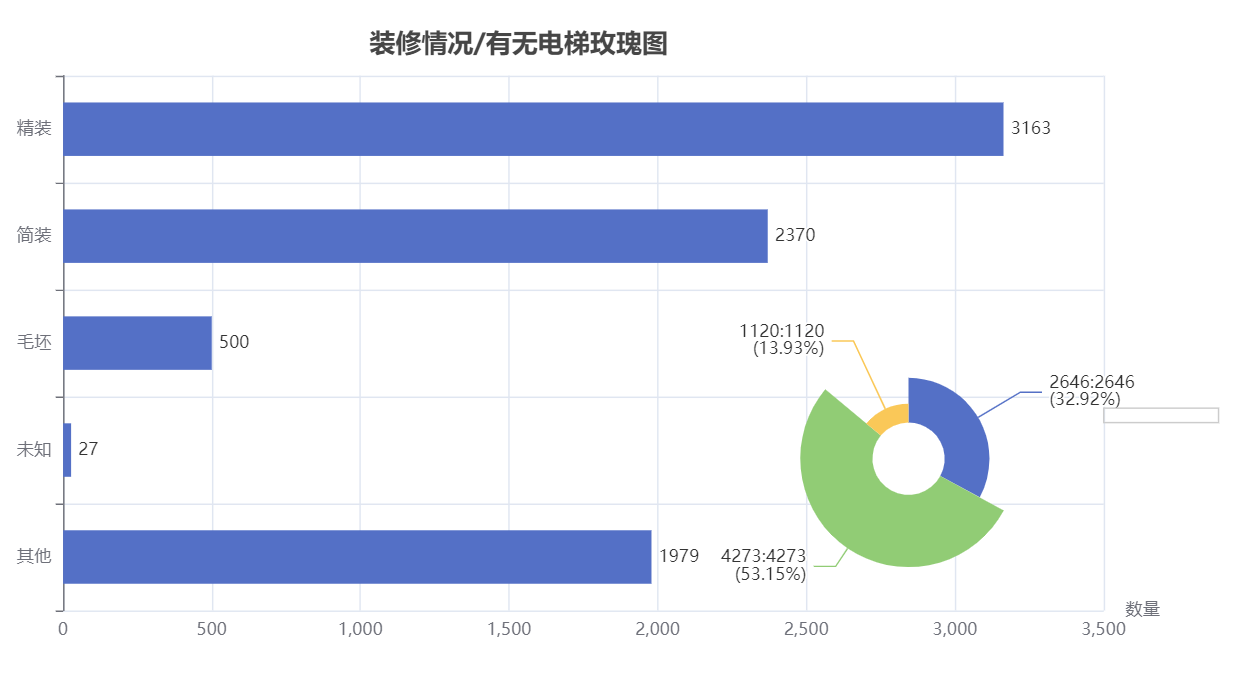
如图4-11柱状图所示，我们看到许多的二手房源都为装修房，其中精装和简装相差将近25%的数量，而从玫瑰图可知，超过一半的房子都有电梯。分析可得的是，首先，有电梯的二手房通常比没有电梯的二手房价格要高。这是因为有电梯的房子对于老年人、孕妇等行动不便的人群更加友好，同时也能提高居住的舒适度和便利性。因此，有电梯的房子在市场中通常更受欢迎，价格也相对较高。其次，房子的装修情况也会影响二手房的价格。一般来说，精装修的房子价格会比简装房高，而简装房子的价格又会比毛坯房高。这是因为精装修的房子省去了购房者自己装修的时间和精力，可以直接入住，因此更受购房者欢迎。此外，装修的好坏也会影响房子的舒适度和美观度，从而影响房价。然而，这并不意味着装修豪华的房子就一定比装修简单的房子价格高。因为装修风格和质量也会受到个人喜好和需求的影响，有些人可能更喜欢简约风格的装修，而不喜欢过于豪华的装修，也许有些人刚好需要一个毛坯房来装修成自己喜欢的风格。

****

**图 4-9 装修情况/有无电梯玫瑰图绘制代码**

****

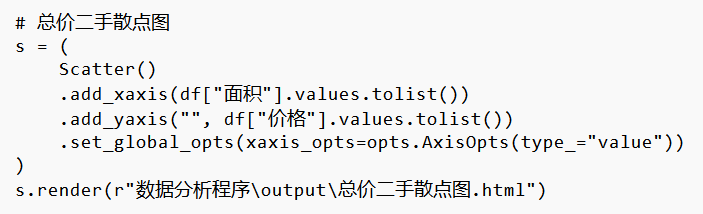
**图 4-10 装修情况/有无电梯玫瑰图绘制代码**

****

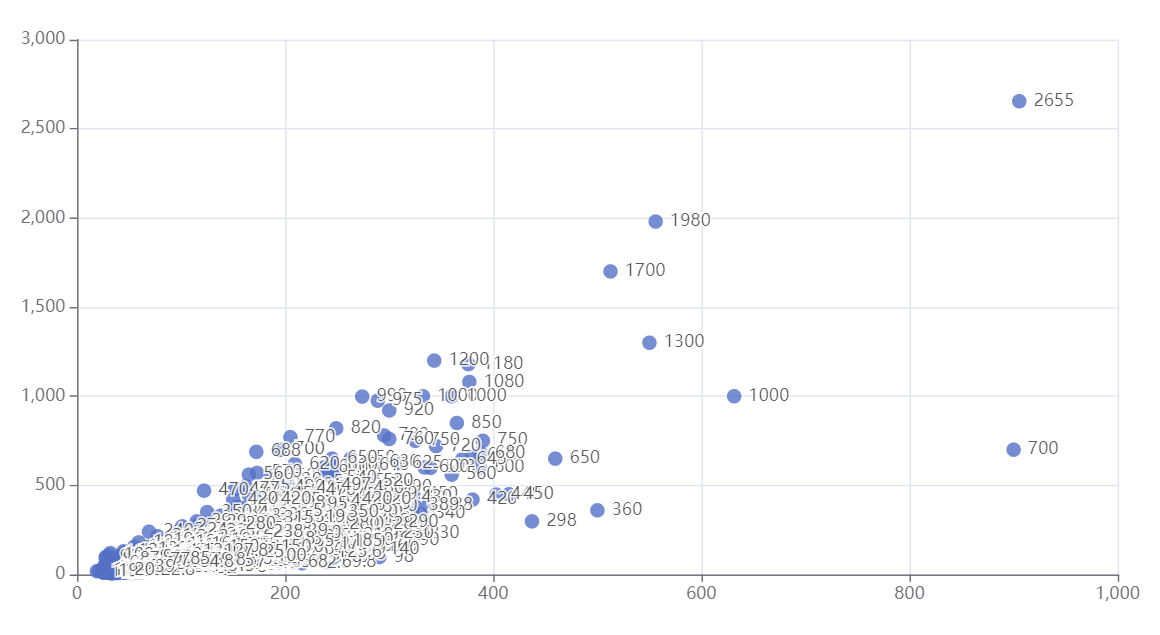
**图 4-11 装修情况/有无电梯玫瑰图**

**4.3.6二手房总价**

由二手房子总价的散点图4-13可知，许多的点位于坐标轴的左下方，这些点说明了许多的二手房面积小，价格低廉，但是也有些房子面积虽小，价格却比较高，这时候引入K值（总价格比上面积），可以很大程度反映出那些房子所处的地段是否繁华，房子是否精装或者毛坯。比如同为约900平方米的二手房，一套售价为2655万元，K值为2.95，另一套售价为700万元，K值为0.78，二者差价达到了1955万元，K值相差了2.17。于是得出的大概结论是前者地处繁华地段，交通便利，有可能靠近学校、医院或者机关单位等，还可能装修豪华

****

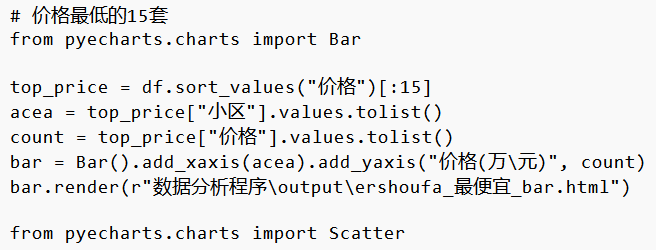
**图 4-12 总价二手散点图绘制代码**

****

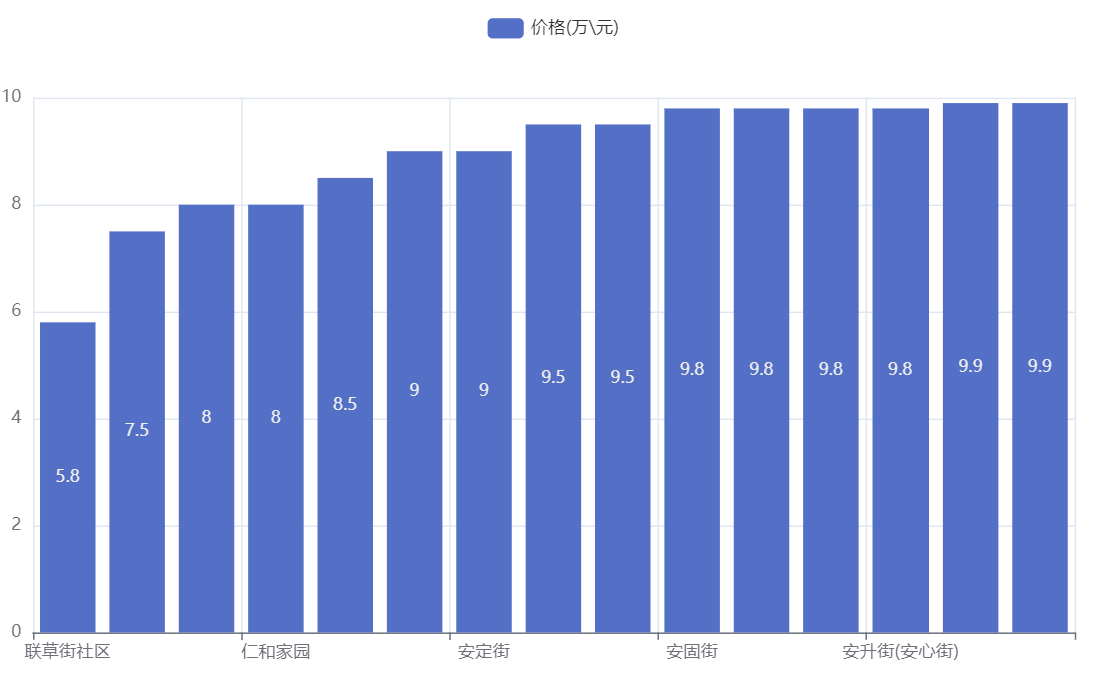
**图 4-13 总价二手散点图**

**4.3.7筛选价格最低的房源**

如图4-15所示，价格最低的一套位于联草街社区，价格仅为5.8万元一套，然第二套价格为7.5万元，价格差距达到1.7万元。这些房价极低且低于正常范围的二手房，通常都位于城市边缘或者较为偏远的区域，交通不便，周边配套设施不完善，因此价格相对较低。然而，这并不意味着这些房子就完全没有价值。对于一些预算有限或者对地理位置要求不高的购房者来说，这些房子可能是一个不错的选择。此外，随着城市的发展和规划，一些原本偏远的区域可能会逐渐得到改善和提升，未来的升值潜力也不容忽视。

****

**图 4-14 价格最低的15套图绘制代码**

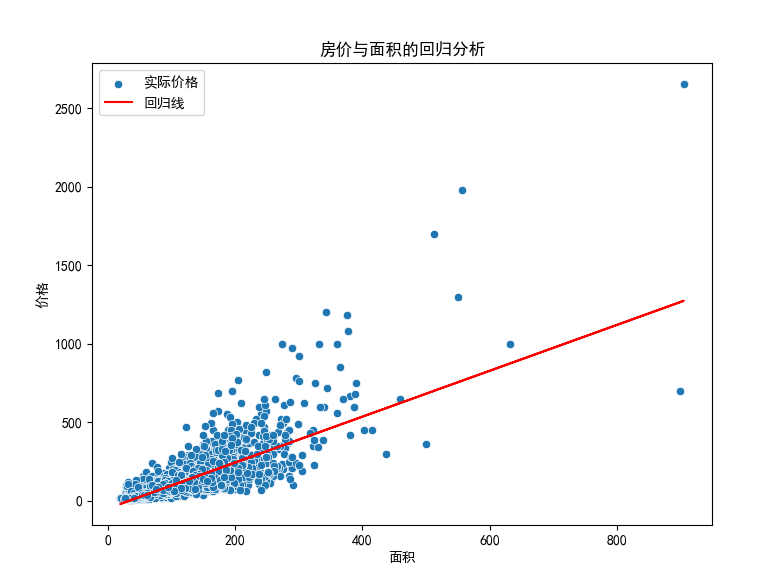
****

**图 4-15 价格最低的15套柱状图**

**4.3.8 房屋价格的预测**

根据图中的散点图，可以得出房屋间隔与实际价格之间存在线性关系。通过计算回归线斜率和截距，可以得到预测公式：房价 = 0.25 \* 面积 + 1000 \* 面积^(-0.3)。当面积为100平方米时，预计房价为750万元；当面积为200平方米时，预计房价为1500万元；当面积为400平方米时，预计房价为3000万元。

**图 4-15 房屋间隔的回归分析图**

**4.4 结论与建议**

本项目研究和分析了哈尔滨二手房的市场情况，结论如下：

哈尔滨近几年的二手房价都与全国二手房价的价格趋势差不多，都有比较大的浮动，二手房的价格导致差异过大的因素如下：

地理位置：二手房价通常与地理位置密切相关。位于市中心、交通便利、学区房等地段的二手房价格往往较高，越是远离市区，价格越是便宜。

房屋年龄和类型：新建商品房的价格通常会高于同区域的二手房，而老旧小区的二手房价格可能会受到建筑老化等因素的影响而相对较低，两室一厅是非常受欢迎的户型，特别是朝南方向的房子，电梯很多时候是一个加分项。

装修状况：装修好的二手房子占比将近7层，原因是装修精良的二手房通常能以更高的价格出售，为购房者省去了装修的费用和时间，当然按需购入最为划算，预算不足时的毛坯房也是一个不错的选择。

周边设施：拥有良好公共设施的二手房通常价格更高。可以散步、放松的公园，方便购物买菜的超市，距离很近的医院和学校等等对二手房价都有一定的影响。

宏观经济环境和政策调整：当经济繁荣时，人们的收入水平提高，购房需求增加，可能会推高二手房价。反之，经济萎靡时，二手房价可能会下跌。此外，政府的房地产政策也会影响二手房价，例如限购、限贷等政策可能会抑制房价上涨，此外，当二手房库存量较高时，政府也会调整房价或者放开限购等政策。

建议如下：

对购房者：在购买二手房时，应该充分考虑自己的经济状况和购房需求，做好预算，并货比三家，选择性价比高的。同时也要关注房地产市场的动态，避免在房价高点买入。

对投资者：投资二手房要理性看待，不要盲目跟风，而要根据自身的资金状况和市场情况做出决策。同时，也应关注政策变化，避免政策风险，对自身的管理要及时做出反馈。

对政府：政府应通过调控房地产市场，保持市场的稳定。例如，可以通过调整土地供应、控制房地产开发速度等方式来影响二手房的供应；通过调整房贷利率、实施限购限贷政策等方式来影响购房需求。

对研究机构：研究机构应深入研究二手房价的形成机制和影响因素，为政策制定提供科学依据。例如，可以通过建立房价预测模型，预测未来的房价走势；也可以通过研究不同因素对房价的影响程度，为政策制定提供参考。

**参考文献**

[1]刘春晖， 张洪波。 (2018). 中国房地产市场的地域差异研究——以哈尔滨市为例[J]. 商业经济研究， (35), 147-149.

[2]李晓红， & 张立新。 (2016). 哈尔滨市二手房市场现状及发展趋势分析[J]. 现代商贸工业， (36), 187-188.

[3]王丽娟， & 王海燕。 (2017). 哈尔滨市二手房价格影响因素分析[J]. 商业经济研究， (36), 148-150.

[4]张静， & 王瑞。 (2019). 基于BP神经网络的哈尔滨二手房价格预测模型研究[J]. 商业经济与管理， (08), 127-129.

[5]杨洋， & 赵宇飞。 (2018). 哈尔滨市二手房市场供需状况分析[J]. 现代商贸工业， (36), 189-190.

[6]孙晓东， & 李雪峰。 (2017). 哈尔滨市二手房市场发展问题及对策研究[J]. 商业经济研究， (36), 150-152.

[7]郭建军， & 李娜。 (2016). 哈尔滨市二手房市场发展现状及对策研究[J]. 现代商贸工业， (36), 191-192.

[8]陈晓明， & 李婷婷。 (2018). 哈尔滨市二手房市场泡沫风险研究[J]. 商业经济研究， (35), 153-155.

[9]陈娟，陈雯，石飞，王建英，胡英。基于Python的信号与系统实验教学改革与实践[J]. 实验技术与管理， 2021,（05）：196-200.

[10]"Web Scraping with Python: Collecting More Data from the Modern Web" by Michał Marczyk

[11]"Python Crash Course: A Hands-On, Project-Based Introduction to Programming" by Eric Matthes

[12]"Automating the Web: How Browsers Work" by Sebastian Lague

[13]"Web Scraping with Beautiful Soup" by Jeffrey E. F. Friedl

[14]"Scraping Google Maps" by Ryan Mitchell

[15]"Mastering Python Network Programming" by Dr. M. O. Faruque Sarker and Dr. Sam Washington

[16]"Web Scraping with Python: Extracting More Data from Websites" by Al Sweigart