

# 基于Python的北京房价影响因素分析及购房建议

史贺莹, 刘芳远

北方工业大学经济管理学院, 北京

收稿日期: 2023年10月26日; 录用日期: 2023年11月13日; 发布日期: 2023年12月28日

## 摘要

北京作为中国经济、政治和文化中心, 房价一直处于较高水平, 受到人们的广泛关注。本文基于Python爬虫功能, 通过爬取链家网站的2023年北京房屋数据, 对房价排名前十、后十的小区进行影响因素分析, 并对房屋地理位置、房屋装修、房屋面积等影响因素的影响度进行展开分析。研究发现, 地理位置是影响北京市房价波动的主要因素, 究其原因, 不同地理位置的经济发展水平与对教育重视程度皆有差异。此外, 房屋面积、格局、装修条件等内部因素对房价同样有一定影响。再结合分析的结果, 提出购房建议, 希望能够为购房者提供参考, 以便能够更好的选房、购房。

## 关键词

Python, 房价影响因素, 购房建议

# Python-Based Analysis of Factors Affecting House Prices in Beijing and Suggestions for Buying a House

Heying Shi, Fangyuan Liu

School of Economics and Management, North China University of Technology, Beijing

Received: Oct. 26<sup>th</sup>, 2023; accepted: Nov. 13<sup>th</sup>, 2023; published: Dec. 28<sup>th</sup>, 2023

## Abstract

Beijing as China's economic, political and cultural centre, housing prices have always been at a high level and have attracted widespread attention. Based on the Python crawler function, this

paper analyzes the influencing factors of the top ten and bottom ten communities of house prices by crawling the 2023 Beijing housing data of the Lianjia website, and analyzes the influence of house geographical location, house decoration, housing area and other influencing factors. The study found that geographical location is the main factor affecting the fluctuation of housing prices in Beijing. The reason for this is that the level of economic development in different geographical locations and the degree of attention to education are different. In addition, internal factors such as housing area, pattern and decoration conditions also have a certain impact on house prices. Combined with the results of the analysis, we put forward suggestions for buying a house, hoping to provide reference for buyers so that they can better choose and buy a house.

## Keywords

Python, House Price Influencing Factors, House Purchase Suggestions

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

房地产一直是我国的支柱型产业,与人们的生活息息相关。房屋支出对于居民来说占比较大,因此房价也逐渐影响到居民的日常生活,并且对社会的稳定有一定的影响。北京是目前人口聚集度最高的城市,各区域间不同的特点,使得北京各区域间房价差距较大。因此研究北京房价波动的影响因素有着普遍意义。

通过对北京近十年的房价统计可以得出,从2011年开始房价一直处于持续上升趋势,截至2023年9月底,北京平均房价已涨至60,565元/m<sup>2</sup>,与2011年相比增长了196%。在2019年房价波动最为严重,由于疫情这一外部因素,导致经济发展速度变慢,房产建设等进度变缓,但根据后几年的平均房价来看,疫情对于房价的影响已几近消退。20年至23年北京平均房价基本稳定在60000~70000元/平方米之间。经过查阅文献,发现房价的波动受到多种因素影响,由此利用北京的房价特点展开调查、分析哪种因素对于房价的影响更大。

## 2. 影响房价的因素

房价会因各种影响而发生波动,并且根据影响因素不同产生的波动程度也不同。郭金[1]的研究认为房价的影响因素包含经济因素、社会因素、行政政治因素以及房产的内在和环境因素。“太平洋房屋”也认为房子自身条件、外部环境、政策因素以及市场供需状况等都对房价有着不同程度的影响。

经济因素包括城市的发展是否繁荣,是否能够提供更好的就业,居民的收入和消费水平的不同,都对房价有着不同程度的影响。一般来说一个地区的经济条件越好,房地产行业发展就越繁荣[2]。

社会因素对于房价的波动有着两面性的影响,其中包含着城市人口的自然出生率、外来人口的迁入迁出等,这些因素都会影响房屋需求,但这对房地产的发展来说这也是把双刃剑。一个地区资源毕竟是有限的,过于密集的人口会导致资源人均占有率下降,换言之过密集的人口会导致人均收入等降低、资源稀缺性加剧,影响房屋购买力。但另一方面人口的增加也会导致这一地区的房屋需求上升、房屋购买力加大,从而拉动这一地区的房地产的发展,房价也会因此上升[3]。

政策因素主要是政府出台的一些土地、货币等政策,比如19年发布的《中华人民共和国城市房地产管理法》;20年提出稳妥推进房地产税立法;20年提出的《中华人民共和国土地管理法》;22年发布

的《北京市“十四五”时期住房保障规划》等。

房屋的内部与环境因素, 房产本身的房型、设施, 还有小区的环境、物业以及承包商等都是影响房价的关键。

### 3. 数据获取与研究方法

#### 3.1. 爬取对象

影响房价的因素繁杂多样, 作为消费者很难把大量的房价数据分析出来。在找寻各类房产网站后发现链家网的规模较大、数据信息更多, 适合爬取如今的房价、房源、房型等房屋信息。因此本文利用 Python 爬虫功能, 爬取链家网站上的北京房屋数据, 并对房屋数据进行分析。

#### 3.2. 爬取流程

##### 3.2.1. 分析 URL

查找链家网页, 获取到新的 URL——<https://bj.lianjia.com/ershoufang/pg{i}>, pg 代表爬取页码, 通过修改 i 可以爬取不同页码的房屋信息, 依据新的 URL 爬取网页。

##### 3.2.2. 请求网页

在设计代码时首先向目标站点发送请求(输出一个 request), 等待目标站点服务器的响应, 随后获取响应内容, 如果服务器能正常响应, 会得到一个回答。关键代码见图 1。

```
import requests
import csv
from lxml import etree
url = 'https://bj.lianjia.com/ershoufang/pg{i}'
for i in range(1,10):
    html = requests.get(url=url,headers = headers)
    html = etree.HTML(html.text)
```

Figure 1. Request the key code of the web page

图 1. 请求网页关键代码

##### 3.2.3. 解析网页 提取数据

通过 xpath 路径获取网页信息, 利用 HTML 进行解析, 解析出的数据包含房屋标题、房屋格局、房价等信息, 最后将解析出来的数据打包成一个 CSV 文件, 进行保存[4]。关键代码见图 2。

```
title = html.xpath('//*[@id="content"]/div[1]/ul/li/div[1]/div[1]/a/text()')
print({title}, {houseinfo}, {place_1}, {place_2}, {total}, {price})
f.write(f'{title},{houseinfo},{place_1},{place_2},{total}万,{price}\n')
```

Figure 2. Parsing web pages and extracting key data codes

图 2. 解析网页、提取数据关键代码

## 4. 北京房价主要影响因素分析

### 4.1. 数据预处理

通过 Python 爬取出的二手房屋数据, 对其进行数据预处理, 删除异常值, 并整合数据, 以便更好的进行对比分析。具体如下。

1) 删除缺失地理位置、房屋面积、户型、房价等关键数据的房屋信息, 保留缺失房屋建造年份等非关键数据的房屋信息。

- 2) 保留房屋总价、房屋平均单价的单位信息, 以方便进行数据分析与绘制图表。
- 3) 查询小区所处区域, 并将同一区域的房屋信息整合成同一表格, 以方便进行数据的对比分析[5]。

#### 4.2. 北京房屋单价排名前十和后十小区分析

对数据预处理后的 270 条房屋信息以每平方米房价由高到低为依据排序, 并从中选取了北京市排名前十名和后十名社区的进行对比。如表 1、表 2。

通过对于北京房价排名前十、后十的小区的研究, 找到了影响北京房价波动的主要因素。

排名前十的社区分别为长河湾、黑窑厂西里、复兴门外大街、世华龙樾四期、车公庄西路 20 号院、和平里、阳春光华家园、东花市北里中区。通过查询地理位置发现排名前十的社区均位于海淀区、西城区、东城区、朝阳区四大繁华区域。

**Table 1.** Ten communities with high housing unit prices in Beijing

**表 1.** 北京市房屋单价前十小区

小区名称	所属区域	房屋价格	平均单价
长河湾	海淀区西直门	3540 万	165,059 元/平
黑窑厂西里	西城区陶然亭	825 万	152,807 元/平
复兴门外大街	西城区木樨地	899 万	150,587 元/平
世华龙樾四期	海淀区清河	2860 万	147,120 元/平
和平里七区	东城区和平里	548 万	140,730 元/平
车公庄西路 20 号院	海淀区甘家口	1350 万	135,952 元/平
车公庄西路 20 号院	海淀区甘家口	1540 万	133,914 元/平
和平里六区	海淀区和平里	1900 万	133,147 元/平
阳春光华家园	海淀区万柳	1608 万	132,794 元/平
东花市北里中区	东城区东花市	960 万	131,977 元/平

**Table 2.** Ten communities with low housing prices in Beijing

**表 2.** 北京市房屋单价后十小区

小区名称	所属区域	房屋价格	平均单价
东亚华欣湾	怀柔区	128 万	14,516 元/平
长海御墅	房山区	210 万	14,955 元/平
万科云溪	密云区溪翁庄镇	380 万	15,903 元/平
当代采育满庭春 MOMA	大兴区大兴新机场洋房别墅区	173 万	19,114 元/平
蓝岸丽舍二期	顺义区	995 万	19,958 元/平
云秀花园	鼓楼街道	206 万	20,908 元/平
中国铁建原香漫谷三区	房山区城关	198 万	23,444 元/平
天恒乐活城南区	房山区阎村	213 万	26,056 元/平
石园南区	顺义区顺义城	275 万	28,200 元/平
绿地花都苑	房山区长阳	252 万	28,321 元/平
加州水郡三期	房山区长阳	410 万	28,339 元/平

排名后十的社区分别为东亚华欣湾、长海御墅、万科云溪、当代采育满庭春 MOMA、蓝岸丽舍二期、云秀花园、中国铁建原香漫谷三区、天恒乐活城南区、石园南区、绿地花都苑、加州水郡三期。通过查询地理位置发现排名前十的社区分别位于怀柔区、房山区、密云区、顺义区等相对偏远郊区。

综上所述, 地理位置是影响北京市房价的主要因素。另外, 前十名社区房屋总价排名与房屋单价排名对比有显著差异, 究其原因, 发现房屋面积、格局、装修条件等内部因素对房价同样有一定影响。

### 4.3. 区域均价差异

#### 4.3.1. 经济发展水平

通过上述分析, 北京房价波动受区域因素影响大。一般来说, 房价走势与经济发展水平密切相关, 一个地区的经济发展水平越高, 人均收入越高, 对房屋需求就越大, 而北京各区域间的经济发展水平各不相同。东城区是四大旧城之一, 地域范围小, 是北京重要的商业中心, 王府井大街地处东城。东城区历史上同西城有东富西贵之说, 是富商的居住地, 东城区的三产业比重已经达到 99%。世界文化遗产故宫, 雄伟的天安门广场属东城辖区; 西城区自古就是达官显贵的地盘, 中南海在西城区辖区内, 大部分国家机关也都建在西城, 金融街是国家金融调控转中心, 占据着国家 60% 以上的金融资产; 朝阳区是北京中心城区的核心, 这里有三里屯、工人体育场和国贸等深受广大消费者青睐的商业街。此外, 朝阳区还拥有许多高端住宅区和写字楼, 因此吸引了许多商务人士和外籍人士居住。这里的教育和医疗资源也非常丰富, 深受市民的欢迎; 海淀区是北京市政治、文化和科技教育的中心。这里有众多高校和研究机构, 因此受到大量学生和知识分子的青睐。此外, 还有中关村科技园区和北京市的许多重点科研项目, 受到了投资者的广泛关注。相比于地理条件、历史发展较好的四大区, 怀柔、房山、密云、顺义等区它们的经济相对较为落后。究其原因, 这些区位于北京市边缘区域, 多山地, 虽能靠山区旅游业发展经济, 但仍处于落后态势, 导致房价相对较低。

为了更好的对比出各区域房价的差距, 对各区域房屋单价的均值进行统计和计算, 东城、西城、朝阳、海淀四大繁华地区的房屋单价均值分别为 120,164 元/m<sup>2</sup>、127,975 元/m<sup>2</sup>、79,253 元/m<sup>2</sup>、106,913 元/m<sup>2</sup> 和 63,832 元/m<sup>2</sup>。而怀柔区、房山区、密云区、顺义区等较偏远地区的房屋单价均值分别为 29,491 元/m<sup>2</sup>、29,640 元/m<sup>2</sup>、23,391 元/m<sup>2</sup>、39,946 元/m<sup>2</sup>。对比两段数据, 可以得出繁华地区的平均房屋单价远高于偏远地区。

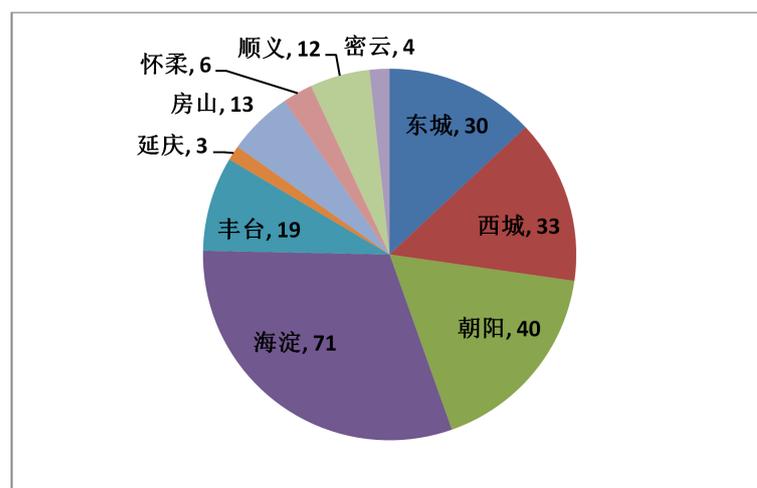


Figure 3. Distribution map of high schools in some areas of Beijing in 2022

图 3. 2022 年北京部分区域高中分布图

### 4.3.2. 教育重视程度

不同区域对于教育的重视程度也不一样, 统计 2022 年北京部分区域的高中数量并制成图表, 如图 3。明显看到东、西、朝、海四区的学校数量远高于其他区域。并且排名靠前的学校大多也分布在这四个区, 由于教育资源的不均衡分布, 导致教育资源良好地区的房价飙升, 区域间房价差异拉大, 这也是学区房诞生的原因。

## 5. 购房与租房建议

购房与租房的第一步是明确需求, 再结合预算、距离来选择房屋。按照年龄以及生活阶段把对购房、租房有需求的人群分为: 学生群体, 职场未婚群体和已婚群体。

对于正在为学业奋斗的学生来说, 一个安静、适合学习的空间是最主要的需求, 并且环境的安全是必不可少的要素, 选择位于安全社区的房屋可以保障个人和财产的安全, 学生上学期间资金不充足, 一般会选择租房而合租是一个不错的选择, 地理位置上靠近校园是最优选择, 不会浪费来回往返时间, 可以为学习和社交提供更大的便利。

已经进入职场的人群, 要利用手中有限的预算选择合适的住所, 这类人群有能力选择租房或者购房, 对于租房人群建议将租金控制在月收入的三分之一以内, 以确保有足够的资金分配给其他支出; 对于购房人群可以选择面积小、首付少、总价低的楼盘, 先买小一点的等赚了钱再选择更好的[6]。在此基础上选择交通便利比如靠近地铁、公交站的房屋或者离工作地点近的房屋, 以减少通勤时间和成本, 居家设施的完备与网络条件的良好, 能够满足日常工作需求, 可以提高生活质量和调节工作心情。

对于已经成家的人群, 舒适性是最主要的需求, 可以满足家庭成员特别是孩子成长和活动的房屋空间, 对于一般的两口、三口之家 100 平方米左右的两室或三室户型是首选, 在资金方面租房人群建议在两人月收入的三分之一内, 而购房人群建议每月还贷额度控制在两人总收入的 30%~40%。环境的安全依旧是一个必要的因素, 安全、宜居的社区有助于孩子的健康成长, 房屋的地理位置最好选在学区附近, 以满足子女的教育需求, 同时有条件的可以综合考虑附近的环境、医疗条件以及房屋的户型、采光等, 成家后若跟老人同住尽量选择有电梯或楼层较低的房屋。

## 6. 总结

在上述研究过程中, 通过 Python 的爬虫功能, 搜集各大房产网站的数据信息, 而利用网络爬虫不仅提高了前期数据收集、分析的效率, 并且爬虫所能获取到的数据, 比人工查找更加全面, 而相对更多、更全面的数据信息也能使图表的对比分析数据更为准确, 因此将爬虫与数据分析相结合为可以大大提高对房价分析的速度及准确性。Python 爬取的数据信息不仅限于房价数据, 还包括不同条件的房子对应的售价。经过对比房屋单价前十、后十的小区, 可以得出经济发展条件、教育重视程度都与地理位置有着密切的关系, 而地理位置是否优越是影响房价高低的主要因素, 房屋本身的面积、格局和装修条件是次要因素。此次的研究结果, 可以帮助消费者更好选择, 也可以为卖家提供价格参考, 促进房地产市场、房地产业的进步, 令我国的经济支柱更加繁荣。

## 参考文献

- [1] 郭金. 基于特征价格模型的北京学区房价影响因素分析[D]: [硕士学位论文]. 曲阜: 曲阜师范大学, 2023.
- [2] 陈将浩. 房价影响因素及 R 语言实现[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2014.
- [3] 吉雪强, 王野, 张跃松. 数字经济、房价与全要素生产率: 调节效应与异质性分析[J]. 经济问题探索, 2023(9): 1-14.
- [4] 李红. 基于 Python 爬虫的太原天气数据分析与可视化研究[J]. 信息与电脑(理论版), 2023, 35(4): 212-214.
- [5] 马腾, 余粟. 基于 Python 爬虫的二手房信息数据可视化分析[J]. 软件, 2023, 44(7): 29-31.
- [6] 刘德炳. 任志强: 不敢做“房价预测帝”了[J]. 中国经济刊, 2013(28): 66-67.