

基于 python 爬虫技术对连云港二手房市场现状分析与研究

李春光

(江苏财会职业学院, 江苏连云港 222061)

摘要: 房价一直是关系民生的最根本问题,早在 2013 年 11 月 12 日十八届三中全会公报中,就明确提出,要稳定房价。如何才能做到稳定房价,让二手房市场健康发展,政府的正确决策与政策导向很重要!为了引导市场,政府也通过各种手段准确获得二手房市场的准确信息,以推断二手房市场的走向,加以引导,使其健康发展。但二手房市场是庞大的,每日推送的数据是海量的,想快速获得准确的统计数据,单凭人力统计,方法老套,数据费时费力;若通过机器学习,使用爬虫技术对二手房网站数据进行批量爬取,然后分析,则为一种全新的技术手段。本研究,运用 Python 第三方库 Requests 来爬取二手房网站的房源数据,并利用 Pandas 库对爬取的数据进行结构化处理,运用 tableau 和 python 的可视化包对二手房源数据进行多维度的深入分析及可视化呈现;通过对大量线上数据分析,推断出连云港市二手房源分布状况、市场热点及价格走势等信息,一者,帮助购房者、房产中介等在市场活动中高效决策;二者,为政府政策干预及对二手房市场进行有效监管,提供决策参考!

关键词: 连云港;二手房;爬虫技术;数据可视

一、研究背景

国务院印发《扩大内需战略规划纲要(2022—2035年)》提出:促进居住消费健康发展。坚持“房子是用来住的、不是用来炒的”定位,加强房地产市场预期引导,探索新的发展模式,加快建立多主体供给、多渠道保障、租购并举的住房制度,稳妥实施房地产市场平稳健康发展长效机制,支持居民合理自住需求,遏制投资投机性需求,稳地价、稳房价、稳预期。在十八届三中全会公报中明确提出要稳定房价,政府陆续也出台了许多的惠民政策,例如新房限价,其具体内容做法是政府出台新房价格备案制度,盖房时就规定了新房的价格限值,避免开发商新房预售开盘时随意涨价的问题。如此以来,虽然新房价格低了,但是可供出售的新房的数量也减却少了,经常可以看到一二十个人去摇号争购一套房源,新房源仍不远远满足购房者的购房需求。目前,在二手房市场,政府对其双方交易还未有太多的政策限制,致使二手房的价格,依然居高不下,部分学区二手房的价格居然远远高于同地段新房价格。

目前,连云港的二手房价市场现状,也正如此,如何协助政府在“让住房回归其原本的属性”上,发挥引导作用,是居高不下,这就需我们对二手市场做些极为深入的研究,了解市场,需要拥有二手市场最新的一手数据,加以分析,去提供研究报告,帮助投资和协助政府制定切实可行的应对策略,加大、加深对新二手市场的干预、引导力度,引导百姓合理投资,理性消费,促进二手市场的合理、健康发展。

二、爬虫的内容、工具及方法

(一) 爬取对象

爬取对象为 58 同城 2023 年 7 月连云港二手房数据,其中数据包括海州、连云、灌南、灌云、赣榆、东海地区的二手房数据。

(二) 爬虫工具

1.python: 通过 URL 链接,将网站的所有内容和链接进行阅读,按照规律抓取网站有价值的信息并下载,以供分析之用。

2.requests: 是 Python 中一个非常流行的 HTTP 客户端库,它

可以用于发送 HTTP 请求和处理 HTTP 响应。

3.xml: 用 Python 语言编写的库,用于解析和提取 HTML 格式的数据,可以利用 XPath 语法快速地定位特定的元素或节点,将传进去的字符串转变成 _Element 对象。

(三) 爬虫过程

1. 寻找需要爬取的 url 链接,本研究的网站为 58 同城连云港二手房的信息,因此链接为: <https://lyg.58.com/ershoufang/>。其中 lyg 为连云港的缩写,如果想找海州的二手房信息,可以在链接中加入海州的拼音,例如: <https://lyg.58.com/haizhou/ershoufang/>。

2. 发送一个 get 请求并获得整个页面响应数据的源码,然后进行局部信息爬取。

3. 创建 etree 实例对象,防止一些不是正规的 utf-8 编码格式的源码而报错。

4. 检查具体信息的标签所在位置,进行具体的数据清洗和数据爬取。

5. 将爬取的 dataframe 信息存入 csv 文件中。

(四) 爬虫内容

爬取的主要内容有:标题、房型、面积、朝向、每平方米单价、总价和地址。通过数据清洗,获取海州、连云、灌南、灌云、赣榆、东海地区共 471 条二手房信息。爬取内容节选如图

序号	标题	房型	面积	朝向	单价/万元/元	总价	地址
1	急售! 琅琅美墅对电梯洋房3室毛坯,送车位,手楼元!	3室2厅2卫	120.55㎡	南北	7861元/㎡	94万	连云港-海州-环海路
2	急售! 琅琅美墅一期 毛坯四室 户型方正 南北通透 随时看房	4室2厅2卫	134.37㎡	南北	8805元/㎡	118万	连云港-海州-汇源路
3	2299精装现房二房两卫,每车方便本世纪,业主急售价格可以谈	2室1厅1卫	49.11㎡	东北	4887元/㎡	24万	连云港-海州-海州湾路3号
4	碧桂园凤凰 送2万 精装 中央空调 地下室 随时看房	3室2厅1卫	90㎡	南北	6566元/㎡	65万	连云港-海州-环海路15号
5	急售! 琅琅美墅二期 两室两厅两卫,家具家电齐全,随时看房	3室2厅2卫	123㎡	南北	6383元/㎡	78.5万	连云港-久和-新光路1号
6	5米1号纯洋房精装修,南北通透,落地窗阳光刺眼,随时可看	2室2厅1卫	106㎡	南北	7246元/㎡	76.5万	连云港-海州-中山南路317号
7	急售! 琅琅美墅三期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	109㎡	南北	7890元/㎡	89万	连云港-海州-大港路
8	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	2室2厅1卫	97㎡	南北	8330元/㎡	80.6万	连云港-海州-云和路88号
9	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅1卫	97㎡	南北	7619元/㎡	73.5万	连云港-海州-大港路
10	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	131㎡	南北	6168元/㎡	80.6万	连云港-海州-中山南路116号
11	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	133.78㎡	南北	7849元/㎡	105万	连云港-海州-云和路1号
12	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	124.27㎡	南北	8986元/㎡	112万	连云港-海州-云和路88号
13	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	2室2厅1卫	88㎡	南北	7791元/㎡	68.1万	连云港-海州-云和路116号
14	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	133.78㎡	南北	7849元/㎡	105万	连云港-海州-云和路1号
15	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	124.27㎡	南北	8986元/㎡	112万	连云港-海州-云和路88号
16	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	2室2厅1卫	88㎡	南北	7791元/㎡	68.1万	连云港-海州-云和路116号
17	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	133.78㎡	南北	7849元/㎡	105万	连云港-海州-云和路1号
18	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	124.27㎡	南北	8986元/㎡	112万	连云港-海州-云和路88号
19	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	2室2厅1卫	88㎡	南北	7791元/㎡	68.1万	连云港-海州-云和路116号
20	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	133.78㎡	南北	7849元/㎡	105万	连云港-海州-云和路1号
21	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	124.27㎡	南北	8986元/㎡	112万	连云港-海州-云和路88号
22	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	2室2厅1卫	88㎡	南北	7791元/㎡	68.1万	连云港-海州-云和路116号
23	急售! 琅琅美墅二期 两室两厅两卫,好楼层无遮挡,随时可看	3室2厅2卫	133.78㎡	南北	7849元/㎡	105万	连云港-海州-云和路1号

图 1 爬虫信息节选

三、连云港二手房分析

(一) 住房面积分析

如图 2 住房面积分析图。连云港的 6 个区域的住房面积出入不大,目前平均的二手房面积在 110-120 平方米之间。其中灌南的平均二手房面积最大,为 123.67 平方米,海州的住房面积最小,为 112.63 平方米。



图 2 住房面积分析图

(二) 二手房价格分析

如图3 二手房每平方米价格可以发现, 东海和赣榆的房价接近, 大概在每平方米一万元左右, 灌南、灌云和连云港的价格较低, 在6000到8000之间, 海州的房价最贵, 目前平均价格在1万3左右。

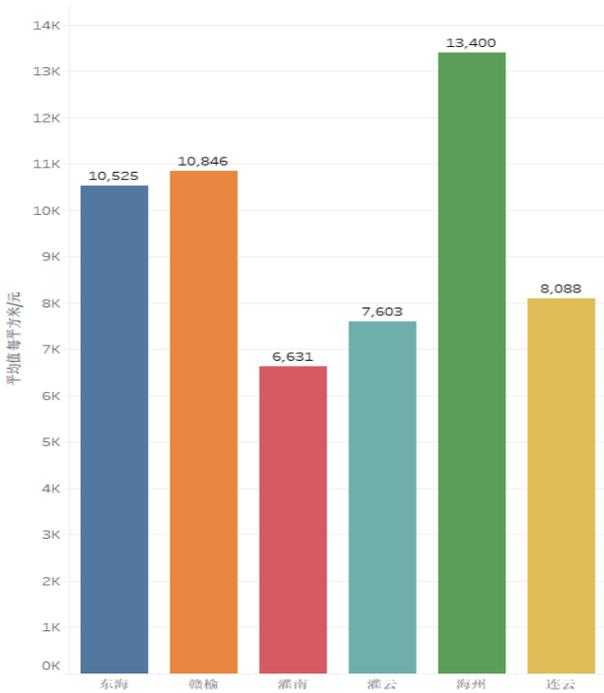


图3 二手房每平方米价格

(三) 住房结构分析

如图3.3 二手住房结构分布图可以看出, 目前的二手房主要以3室2厅1卫和3室2厅2卫为主, 大概占了总统计值的85%, 小部分的房型为4室2厅2卫和2室2厅1卫为主, 合计大概占了15%。

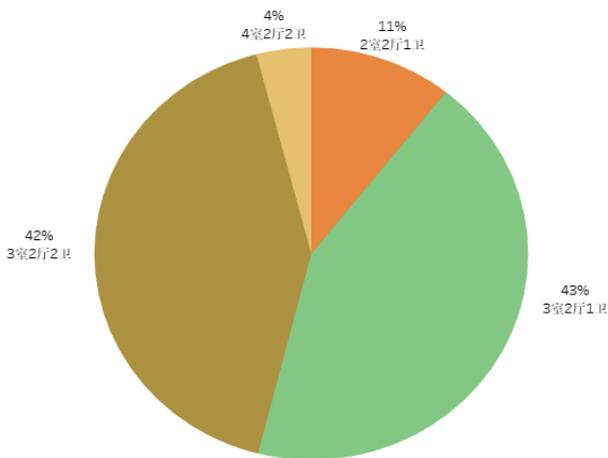


图4 二手住房结构分布图

(四) 总结分析

如图3.4 江苏各地区平均房价及GDP分布表可以看出, 连云港的整体平均房价大概在1万元左右, 在江苏省的房价中排名第8, 但是整体的GDP在2022年排名最后。房价和地区的GDP水平是成正比例分布的, 但是连云港的房价高于应有的GDP水平, 因此连云港的房价更需要政府进行宏观和微观的调控。

表1 江苏各地区平均房价及GDP分布表

时间	地区	房价	2022年GDP排名
2023年	南京	26062元/㎡	2
2023年	苏州	21302元/㎡	1
2023年	无锡	14701元/㎡	3
2023年	常州	14556元/㎡	5
2023年	南通	14280元/㎡	4
2023年	扬州	13335元/㎡	7
2023年	泰州	12225元/㎡	9
2023年	连云港	10212元/㎡	13
2023年	徐州	10149元/㎡	6
2023年	宿迁	9923元/㎡	12
2023年	盐城	9784元/㎡	8
2023年	淮安	9324元/㎡	11
2023年	镇江	8845元/㎡	10

四、建议

房价的主要影响因素是城市的经济发展和房子周围的基础设施, 例如学校、医院、超市、公交等。其中, 就连云港目前而言, 对二手房价格影响最大的是学区房。父母对子女的教育尤为看重, 那些限价的新房, 购买人太多, 摇号买到的, 仅是很少的一部分; 因此, 大多数只能退而求其次, 买那些价格更合适的二手学区房。鉴此, 就上述原因, 本人提出了几点意见:

(一) 增加中小学教育资金投入, 让教育资源均衡化。

父母对子女的教育问题, 在父母看来, 乃重中之重, 若增加中小学教育资金投入, 改善教学设施, 加大师资力量培养投入, 倾向性地改善师资薄弱学校的名师待遇, 吸引名师前往工作, 带动师资薄弱学校师资水平的整体提升; 政策性的引导教育资源均衡化, 让连云港各中小学整体的教育质量均上了台阶, 总体上教育质量无明显差距。如此一来, 选择多了, 大家就不用再去买高价的二手学区房, 整体的二手房价格也就间接降下来供需就更加平衡了。

(二) 招商引资, 发展实体经济, 放水养鱼, 让实体经济更有后劲; 藏富于民, 提高连云港本地居民整体的经济水平消费水平。

降低门槛, 招商引资, 发展实体经济, 增强发展连云港本土特色的本土产业活力, 营造实体经济发展的良好环境, 促进企业做大做强, 促进就业, 提高本地居民收入, 带动人们消费水平的提高, 让“居住”等延伸出的系列问题不再是难事。

(三) 住房政策的制定要统筹考虑, 长久调控。

建立和完善住房供应体系, 依据分层分类供给原则设立机制, 让不同层次需求, 做到有制度保障、有政策支持、有市场选择, 让住房保障不是难事, 让住房改善成为常事; 让二手房市场真正地红火起来。

参考文献:

[1] 余晓龙. 人才落户政策对房价的影响[D]. 浙江工商大学, 2020.
 [2] 王伟华, 袁持平. 因城施策背景下城市房价问题研究——兼析深圳市新房与二手房市场价格传导机制[J]. 价格理论与实践, 2019(12): 35-38.
 [3] 陈秉宇. 中型城市房价影响因素探析[J]. 北方经贸, 2021(07): 73-75.