

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，政府陆续也出台了很多的惠民政策，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。如此一来，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、lxml：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、总房价和地址。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	每平方米/万元	总房价	地址
0	秒杀价！独幢别墅带电梯洋房三卫，送车位，手慢无！	3室2厅2卫	120.5㎡	南北	7967元/㎡²	96万	连云-海州-环海路
1	急售！步梯四期 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	5805元/㎡²	78万	连云-海州-汇锦路
2	222万！精装修 三卫两室，停车方便无遮挡，业主急卖价格可以谈	2室1厅1卫	49.11㎡	东北	4887元/㎡²	24万	连云-云山-海清路3号
3	绿地世纪城 送8平方 精装三室 中下层 随时看房	3室2厅1卫	99㎡	南北	6566元/㎡²	65万	连云-海州-环海路12号
4	天惠花园电梯房3楼三室两厅两卫，家具家电齐全，随时看房	3室2厅2卫	123㎡	南北	6383元/㎡²	78.5万	连云-久和新光路1号
5	长江一号东户精装2室，南北通透，落地阳光采厅，随时可看	2室2厅1卫	106㎡	南北	7246元/㎡²	76.8万	连云-云山-山西路317号
6	东岸二室精装品牌家具家电东户采光好全新装修随时看房	3室2厅2卫	109㎡	南北	7890元/㎡²	86万	连云-云山-大港路
7	急售！步梯二期毛坯两室两卫朝南客厅大，好楼层无遮挡	2室2厅1卫	97㎡	南北	8330元/㎡²	80.8万	连云-云山-云和路88号
8	独幢 首付少！奶油风全装拎包入住！带中央空调！有钥匙	3室2厅1卫	97㎡	南北	7619元/㎡²	73.9万	连云-云山-院前路89号
9	新湖澜湾中上楼层，亏本急卖，真心性价比！	3室2厅2卫	131㎡	南北	6168元/㎡²	80.8万	连云-云山-平山北路118号
10	独幢独院，西院牙毛坯二室两卫精装修 好楼层 送车位。	3室2厅2卫	133.78㎡	南北	7849元/㎡²	105万	连云-云山-云宿路1号
11	福港好莱坞二期纯西户双阳台精装修，品牌家具配置！	3室2厅2卫	124.27㎡	南北	9898元/㎡²	123万	连云-云山-云和路88号
12	单价7千多！景城名都中间楼层两室两厅两卫采光好有钥匙看房	2室2厅1卫	88㎡	南	7591元/㎡²	66.8万	连云-久和-五丰路16号
13	美麓湖畔，飞机户型，南北通透，采光无遮挡，送车位	3室2厅1卫	120.5㎡	南	7718元/㎡²	93万	连云-海州-环海路
14	绿地悦澜山，中间楼层，采光无遮挡，户型好，南北通透	3室2厅2卫	139㎡	南北	8274元/㎡²	115万	连云-海州-环海路
15	绿洲好房急售84万！单价5100！大三居随时看，首付低月供少	3室2厅2卫	125㎡	南北	5600元/㎡²	70万	连云-久和-东方大道23号
16	单价5800！步梯花园4室 房东诚意出售 随时看房	4室2厅2卫	134.7㎡	南	5776元/㎡²	77.8万	连云-海州-汇锦路
17	89.9w！精装三房三卫 飞机户型 带暖气送家电拎包入住	3室2厅1卫	112.66㎡	南	7980元/㎡²	89.9万	连云-久和-新光路8号
18	四季花城东区精装三室南北通透精装修有钥匙随时看房	3室2厅1卫	118㎡	南北	7034元/㎡²	83万	连云-云山-院前路89号
19	首付低！房东急售！全新装修中央空调家具家电齐全拎包入住随时看	3室2厅1卫	97㎡	南北	7619元/㎡²	73.9万	连云-云山-大港路
20	急售！美麓湖畔 送车位，飞机户型 中层，有钥匙看房	3室2厅2卫	120㎡	南北	7667元/㎡²	92万	连云-海州-环海路
21	景城名都急售83.8万！单价7200！新装修两室两卫江南庭院	2室2厅1卫	88.31㎡	南	7429元/㎡²	65.6万	连云-久和-五丰路16号
22	今日新上一室70平产教育燃气 好贷款电梯中华豪庭双	1室1厅1卫	47.89㎡	南北	5764元/㎡²	27.6万	连云-云山-中华西路
23	降价啦！快来看看业主工作明确自住商品房精装修送双车位可刀	3室2厅1卫	99㎡	南北	8586元/㎡²	85万	连云-久和-五丰路16号

图 2.1 爬虫信息节选

三、连云港二手房分析

（一）住房面积分析

如图 3.1 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。

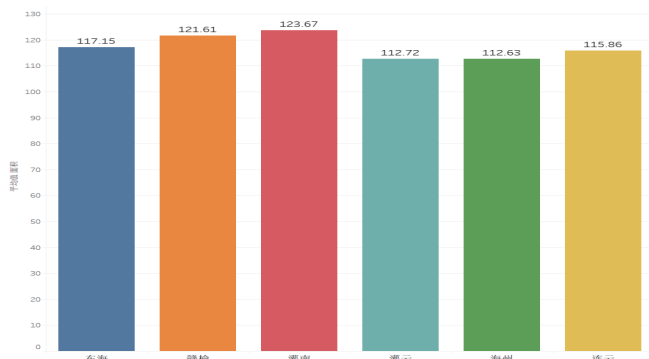


图 3.1 住房面积分析图

(二) 二手房价格分析

如图 3.2 二手房每平方米价格可以发现，东海和赣榆的房价接近，大概在每平米一万元左右，灌南、灌云和连云的价格较低，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。

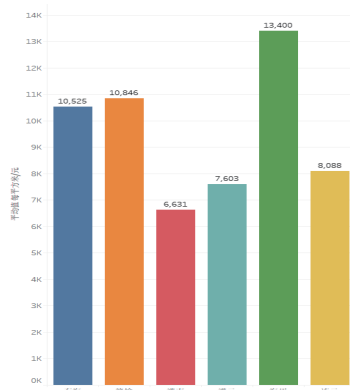


图 3.2 二手房每平方米价格

(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%。

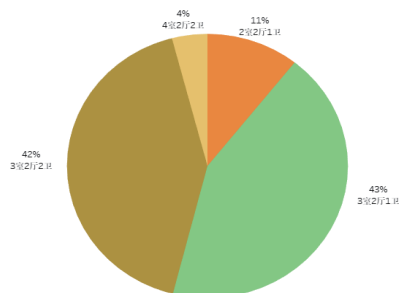


图 3.3 二手住房结构分布图

(四) 总结分析

如图 3.4 江苏各地区平均房价及 DGP 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。

3.4 江苏各地区平均房价及 DGP 分布表

时间	地区	房价	2022 年 GDP 排名
2023 年	南京	26062 元 / m ²	2

2023 年	苏州	21302 元 / m ²	1
2023 年	无锡	14701 元 / m ²	3
2023 年	常州	14556 元 / m ²	5
2023 年	南通	14280 元 / m ²	4
2023 年	扬州	13335 元 / m ²	7
2023 年	泰州	12225 元 / m ²	9
2023 年	连云港	10212 元 / m ²	13
2023 年	徐州	10149 元 / m ²	6
2023 年	宿迁	9923 元 / m ²	12
2023 年	盐城	9784 元 / m ²	8
2023 年	淮安	9324 元 / m ²	11
2023 年	镇江	8845 元 / m ²	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”。鉴于此，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。如此以来，选择多了，大家就不用再去买高价二手学区房，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] 金浩智. 三亚市商品住房特征价格实证研究 [D]. 华东理工大学, 2017.
- [3] 乔林, 孔淑红. 我国不同发展层次城市房价影响因素的差异分析及对策研究 [J]. 特区经济, 2012 (05): 210-212.
- [4] 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.
- [5] 陈秉宇. 中型城市房价影响因素探析 [J]. 北方经贸, 2021 (07): 73-75.