

甲方名称:

江苏苏州某大数据公司

甲方需求:

- 采集搜房网（现更名为房天下）全国各城市新房楼盘和二手房小区数据；
- 需要采集的字段如下：

“小区名称”，“城市”，“小区地址”，“项目特色”，“所属区域”，“邮编”，“产权描述”，“物业类别”，“竣工时间”，“开发商”，“建筑结构”，“建筑类别”，“建筑面积”，“占地面积”，“当期户数”，“总户数”，“绿化率”，“容积率”，“物业费”，“停车位”，“开盘时间”，“交房时间”，“小区简介”，“交通状况”，“周边信息”，“经度”，“纬度”，“物业公司”，“价格”，“页面链接”，“地图链接”，“区县”，“商圈”，“装修情况”，“物业地址”，“入住时间”，“物业类型”，“环比上月”，“详情页链接”，“新房/二手房”，“newcode”

- 保证数据完整性，最终只交付数据；

技术难点:

- 访问频率过高会被网站封IP；
- 列表页有页数限制，不能看到所有页；
- 小区/楼盘详情页有多种模板；

实现方案:

- 通过轮换代理IP，严格控制每个IP的两次访问间隔时间，可以有效防止请求被网站拦截；
- 通过“区域”筛选器，尽可能多的让列表页页数减少；
- 分别处理不同模板的数据提取规则；

A	B	C	D	E	F	G	H	I	J	K	L	M	N
小区名称	城市	小区地址	项目特色	所属区域	邮编	产权描述	物业类别	竣工时间	开发商	建筑	建筑类别	建筑面积	占地面积
万科山景城	广州	黄埔区长岭路南	山景地产			70年 住宅	住宅	2009-0	广州市万畅房地产开发有限公司	独栋	多层	高550000平方米	20000平方
桃花源	广州	花都山前大道美麓度假区西	特色别墅 宜居生态地产			70年 住宅	别墅	2009-0	广州市绿雅房地产开发有限公司	独栋	多层	130000平方米	41000平方
金湾明珠	广州	荔湾荔湾区大坦沙岛	景观居所 水景地产 宜居生态地产			70年 住宅	住宅、普通住宅	2016-0	广州市鹏达集团有限公司	塔楼	多层 小	100000平方米	26000平方
尚东嘉御	广州	番禺禺山西路东桥大街68号	大宅 山景地产			70年 住宅	住宅、洋房	2017-1	广州市瑞安房地产开发有限公司	塔楼	超高层	125000平方米	42000平方
碧桂园印象花城	广州	佛山广佛西线石洲出口往陈村方向	不限购			70年 住宅	住宅、洋房	2016-1	佛山顺德碧桂园房地产开发有限公司	塔楼	超高层	310000平方米	11666平方
南沙保利城	广州	南沙市南大道与鸡公山立交交汇处	湖景地产 江景地产			70年 住宅	住宅、公寓	2015-1	广东保利置业有限公司	塔楼	高层	100000平方米	20000平方
南海皇汇云锦	广州	广州周边佛山南海大道与南桂东路	地铁沿线 品牌地产	广州	528000	60年 公寓	住宅、公寓	2015-1	佛山南海越秀房地产开发有限公司	塔楼	高层	超290739平方米	31254平方
敏捷科创中心	广州	白云太和镇北太路以北	广州 品牌地产			70年 住宅	住宅、公寓	2016-0	广州市敏捷数字信息电子科技有限公司	塔楼	高层	150000平方米	34000平方
东方文德广场	广州	越秀文德北路66号	文德路以东 暂无资料			70年 住宅	住宅、商住	2010-0	富春东方地产	塔楼	高层	123000平方米	15000平方
保利悦廷	广州	花都兴南大道以南	天马河以西 暂无资料			70年 住宅	住宅	2016-0	广州保利房地产开发有限公司	塔楼	超高层	200000平方米	40332平方
佛山时代壹桥	广州	佛山南海区灯湖西路7号	湖景地产			70年 住宅	住宅、洋房	2017-1	时代地产佛山公司	塔楼	高层	87536平方米	20463平方
龙湖首开天宸原著	广州	天河奥体路转悦景路800米	山景地产			70年 住宅	住宅、别墅	2017-0	广州市湖品房地产开发有限公司	联排	独栋	19497平方米	62679平方
宝翠园	广州	天河区宝翠路4号	华南植物园斜 景观居所 宜居生态地产 公园地产			70年 住宅	住宅、精装洋	2015-0	广州汉国福房地产开发有限公司	塔楼	小高层	306419平方米	115000平方
碧桂园华府	广州	花都山前大道南航碧桂园	山前特色别墅 打折优惠楼盘			70年 住宅	住宅、别墅	2008-0	广州碧桂园房地产开发有限公司	联排	独栋	350000平方米	160000平方
天马河公馆	广州	花都新街大道望亭路15号	湖景地产			70年 住宅	住宅、别墅	2017-0	佛山南海区华星置业投资有限公司	塔楼	超高层	314468平方米	70679平方
碧桂园凤凰城	广州	增城广园东快速路	广园路、广济品牌地产 特色别墅 花园洋房			70年 住宅	住宅、别墅	2012-1	广州市天马河房地产开发有限公司	联排	双拼	1332000平方	6666700平方
同创·涌江	广州	增城增城大道涌江生活馆	江景地产			70年 住宅	住宅	2015-1	碧桂园物业发展有限公司	塔楼	小高层	121203平方米	42280平方
莱茵水岸	广州	从化从化大道33号	从化汽车站 复合地产 特色别墅 花园洋房			70年 住宅	住宅、别墅	2014-0	广州市嘉利房地产开发有限公司	联排	独栋	430446平方米	242289平方
碧桂园豪园	广州	增城增城公路碧桂园豪园	原石湾旅游度假区 不限购			70年 住宅	住宅、洋房	2015-0	碧桂园地产开发有限公司	塔楼	小高层	100483平方米	20000平方
华发荔湾壹	广州	荔湾区芳德路以西	荔湾涌以北 科技住宅			70年 住宅	住宅、洋房	2016-1	广州华发房地产开发有限公司	塔楼	高层	91218平方米	42143平方
流花君庭	广州	荔湾流花路73号	德宝保龄球馆次 水景地产 复合地产 低密度所 打折优惠			70年 住宅	住宅、住宅	2008-1	广州盈湖房地产开发有限公司	塔楼	高层	64244平方米	9199平方
中国美林湖	广州	山前大道168号	特色别墅 宜居生态地产 打折优惠楼盘			70年 住宅	住宅、别墅	2016-1	美林基业集团	联排	独栋	303996平方米	18211平方
保利·红馆	广州	番禺汉溪大道与新光快速交汇处	科技住宅			70年 住宅	住宅、洋房	2017-1	广州市宏和房地产开发有限公司	塔楼	高层	42793平方米	18211平方
恒大金碧天下	广州	清远清新清连高速旁	（龙颈出口）特色别墅 花园洋房 打折优惠楼盘			70年 住宅	住宅、别墅	2010-0	恒大地产	塔楼	板塔结合	300000平方米	666670平方
南沙湾·石奥	广州	南沙港前大道南沙湾健身会旁	海景地产			70年 住宅	住宅、公寓	2016-1	广州南沙开发建设有限公司	塔楼	多层	72138平方米	39177平方
越秀保利爱特城	广州	黄埔云埔一路以南	暂无资料			70年 住宅	住宅、花园洋	2016-1	广州中耀实业投资有限公司	塔楼	高层	680215平方米	189520平方
清远海沁春天	广州	清远清城区新城广场大峡峡江西路	旅游地产 不限购			70年 住宅	住宅、洋房	2014-0	海伦堡地产	塔楼	超高层	140985平方米	72476平方
越秀·星汇海珠湾	广州	海珠广州市海珠区广纸路	江景地产			70年 住宅	住宅、洋房	2017-0	越秀地产集团	塔楼	超高层	580000平方米	257682平方
凤东·雅轩	广州	海珠革新路凤安街	江景地产			70年 住宅	住宅、洋房	2012-0	广州市和联房地产开发有限公司	塔楼	超高层	34499平方米	16429平方
广物星岛国际	广州	花都区迎宾大道东侧	（广大附中）科技住宅			40年 商住	住宅、洋房	2016-12-01		塔楼	小高层	258000平方米	101000平方
碧桂园国际	广州	花都狮岭大道东8号	复合地产 花园洋房 打折优惠楼盘			70年 住宅	住宅	2017-0	碧桂园地产	塔楼	高层	70000平方米	226780平方
碧桂园·保利星城	广州	荔湾彩虹岗路1号	（广物）住宅 商住 科技			70年 住宅	住宅	2017-0	广州市和联房地产开发有限公司	塔楼	板塔结合	442021平方米	22666平方

示例数据截图



西安鲲之鹏网络信息技术有限公司

选择我们，所有数据都是你的！



公司名称: 西安鲲之鹏网络信息技术有限公司

网 址: <http://www.site-digger.com/>

地 址: 陕西省西安市雁塔区长安中路99号长安文化综合大厦11821室

邮 编: 710061

联系电话: 029 - 87553281

手 机: 13571845363 齐工
13389148466 周工

客 服 QQ: 1649677458 或 312602670

邮 箱: hello@site-digger.com
