数据库设计文档

目录

[表汇总 2](#_Toc511744176)

[基础数据 4](#_Toc511744177)

[均价表（district\_stat） 4](#_Toc511744178)

[区域月度均价表（region\_sell\_stat） 5](#_Toc511744179)

[样本数据信息表(seawater) 6](#_Toc511744180)

[网络抓取 8](#_Toc511744181)

[高德POI表（amap\_poi） 8](#_Toc511744182)

[成交数据表（deal\_price） 9](#_Toc511744183)

# 表汇总

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 表 | 功能说明 | 服务器 | 端口号 | 数据库 | 数据库类型 |
| 1 | region\_sell\_stat | 区域月度均价表 | 222.73.196.143 | 27017 | fangjia\_stat | mongodb |
| 2 | district\_stat | 小区均价表 | 222.73.196.143 | 27017 | fangjia | mongodb |
| 3 | amap\_poi | 高德POI表 | 222.73.196.143 | 27017 | fangjia\_crawl | mongodb |
| 4 | seawater | 样本库 | 192.168.0.105 | 27777 | fangjia | datastore |
| 5 | deal\_price | 成交数据表 | 222.73.196.143 | 27017 | fangjia | mongodb |

# 基础数据

## 均价表（district\_stat）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段描述 | 字段类型 | 备注 |
| \_id | 唯一ID | ObjectId |  |
| city | 城市 | String |  |
| region | 区域 | String |  |
| district\_name | 小区名 | String |  |
| estate\_type | 物业类型 | String | 取值范围(小区，别墅) |
| weekend | 均价日期 | Date | 均价所对应的周日，时分秒都为0 |
| calc\_value | 计算均价 | Integer |  |
| avg\_price | 拟合均价 | Integer | 对外输出价格 |
| credit | 可信度 | double |  |
| value\_type | 计算类型 | Integer | 1：指导价 2：自身样本 3：第三方 4：交易数据 5：周边楼盘， 6：推导 |
| weekend\_markup | 周涨幅 | Integer | 算法：  ((本周均价-上周均价)/上周均价)\* 10000 |
| avg\_markup |  |  | 废弃字段 |
| changed\_markup | 周涨幅 | Integer | 需要兼容字段保持和weekend\_markup一致， |
| markup | 周涨幅 | Integer | 需要兼容字段保持和weekend\_markup一致 |
| sample\_size | 计算均价所使用的样本量 | Integer |  |
| m\_date | 修改日期 | Date |  |
| c\_date | 创建日期 | Date |  |
| **备注**  唯一索引：city + region + district\_name + estate\_type + weekend | | | |

## 区域月度均价表（region\_sell\_stat）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段描述 | 字段类型 | 备注 |
| \_id | 唯一ID | ObjectId |  |
| city | 城市 | String |  |
| region | 区域 | String |  |
| avg\_price | 区域均价 | Integer |  |
| date | 对应年月 | Integer | yyyyMM |
| mon | 同比 | Double |  |
| yoy | 环比 | Double |  |
| updated\_date | 更新时间 | Date |  |
| **备注**  唯一索引：city + region + date | | | |

## 样本数据信息表(seawater)

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段描述 | 字段类型 | 备注 |
| \_id | 唯一ID | ObjectId |  |
| cat | 样本类型 | String | sell:二手房  rent:租房  shop\_sell:商铺出售  shop\_rent:商铺出租  office\_sell:写字楼出售  office\_rent:写字楼出租 |
| city | 城市 | String |  |
| region | 区域 | String |  |
| district\_name | 小区名 | String |  |
| total\_price | 总价 | Integer |  |
| area | 面积 | double |  |
| p\_date | 发布日期 | integer | 格式:yyyyMMdd |
| url | 样本原始url | String |  |
| estate\_type | 物业类型 | String |  |
| height | 楼高 | integer |  |
| floor | 所在楼层 | integer |  |
| room | 室数 | integer |  |
| hall | 厅数 | integer |  |
| toilet | 卫数 | integer |  |
| fitment | 装修 | String |  |
| direction | 朝向 | String |  |
| rent\_type | 出租类型 | String |  |
| user\_id | 用户id | String |  |
| build\_type | 建筑类型 | String |  |
| build\_date | 建造时间 | String |  |
| estate\_charge | 物业费 | String |  |
| status | 状态 | integer |  |
| address | 地址 | String |  |
| block | 版块 | String |  |
| name | 样本名称 | String |  |
| contact\_people | 联系人 | String |  |
| mobile | 手机号 | String |  |
| company | 物业公司 | String |  |
| floor\_info | 楼层信息 | String |  |
| md5\_hash | 主要内容的MD5 | String |  |
| c\_date | 创建时间 | integer |  |
| m\_date | 修改时间 | datetime |  |
| **备注** | | | |

# 网络抓取

## 高德POI表（amap\_poi）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段描述 | 字段类型 | 备注 |
| \_id | 唯一ID | ObjectId |  |
| pname | 省份 | String |  |
| cityname | 城市 | String |  |
| adname | 区域 | String |  |
| adcode | 区域区划编码 | Integer |  |
| uid | 高德地图ID | String |  |
| name | poi名称 | String |  |
| address | 地址 | String |  |
| type | 所属分类名称 | String |  |
| typecode | 分类编号 | String |  |
| big\_type\_code | 大类编码 | String |  |
| mid\_type\_code | 中类编码 | String |  |
| lng | 经度 | Double |  |
| lat | 纬度 | Double |  |
| location | 经纬度索引字段 | Array | [lng,lat] |
| m\_date | 更新时间 | Date |  |
| c\_date | 创建时间 | Date |  |
| **备注**  索引：mid\_type\_code +lng+lat  唯一索引：uid  空间索引:location+ mid\_type\_code | | | |

## 成交数据表（deal\_price）

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 字段描述 | 字段类型 | 备注 |
| \_id | 唯一ID | ObjectId |  |
| city | 城市 | String | 房价网标准城市 |
| region | 区域 | String | 房价网标准区域 |
| district\_name | 小区名 | String |  |
| avg\_price | 均价 | Integer |  |
| total\_price | 总价 | Integer |  |
| house\_num | 楼栋号 | String |  |
| unit\_num | 单元号 | String |  |
| room\_num | 室号 | String |  |
| area | 面积 | Double |  |
| direction | 朝向 | String |  |
| fitment | 装修 | String |  |
| source | 来源 | String | 抓取网站标示 |
| room | 室数 | Integer |  |
| hall | 厅数 | Integer |  |
| toilet | 卫数 | Integer |  |
| height | 总楼层 | Integer |  |
| floor | 所在楼层 | Integer |  |
| trade\_date | 成交日期 | Date |  |
| m\_date | 更新时间 | Date |  |
| create\_date | 创建时间 | Date |  |
| **备注**  唯一索引：city+ region+ district\_name+ source+ trade\_date+ area  已经存在的source值：房途网,新浪,满堂红,链家在线,搜房,中原地产,住在杭州网,爱屋吉屋,美狮,交易中心,网络搜索,我爱我家,呼和浩特房产网,购房网,国大地产,兔博士,家家顺,丹夏网,中安房,麦田,平安好房,链家地产,房天下,上海酷房网,悟空找房,麦田房产,其他,太屋网,其他公司,Q房网 | | | |