黔南民族职业技术学院

**实 验 报 告**

实验课程 数据挖掘技术

实验项目 Python数据挖掘与分析

实验地点 408实训室

指导教师 姜永成

班 级 19大数据技术与应用(5)班

学生姓名 窦友俊

学 号 202003170536

教师评分

日 期

**期末实验 Python数据挖掘与分析**

**一、实验目的**

1. 掌握数据挖掘的基础知识，能够深刻理解并熟练运用Python进行数据处理的过程。

2. 掌握数据获取与收集方法。

3. 学会数据清洗和整理的方法。

4. 了解数据统计方法。

5. 能够利用可视化工具进行数据的展示。

1. **实训任务**

**任务1：【分析网页url】**

爬取的网站名称为：链家

爬取内容：房子的详情

网页地址为：https://gy.lianjia.com/ershoufang/pg/

网址规律：

第一页网址为：https://gy.lianjia.com/ershoufang/pg/1

第二页的网址为：https://gy.lianjia.com/ershoufang/pg/2

以此类推可以发现该网址分为三部分，每翻一页网址的变化都为/pg/ +1

|  |
| --- |
| 源码如下 |
| url |
| url1捕获 |

**任务2：【利用scrapy框架方法爬取链家二手房数据】**

从网页的分析来看，所需爬取的数据为：

“链接”，“房名”，“地点”，“价格”，“面积”，“朝向”，“上市日期”，“装修”，“几室几厅”

|  |
| --- |
| 源码如下 |
| 几捕获 |

**任务3：【数据存储】**

将所爬取的相关数据存入到json

|  |
| --- |
| 源码如下 |
| json捕获 |

**任务4：【数据清洗和导入SQLite】**

将所爬取的相关数据导入到SQLite

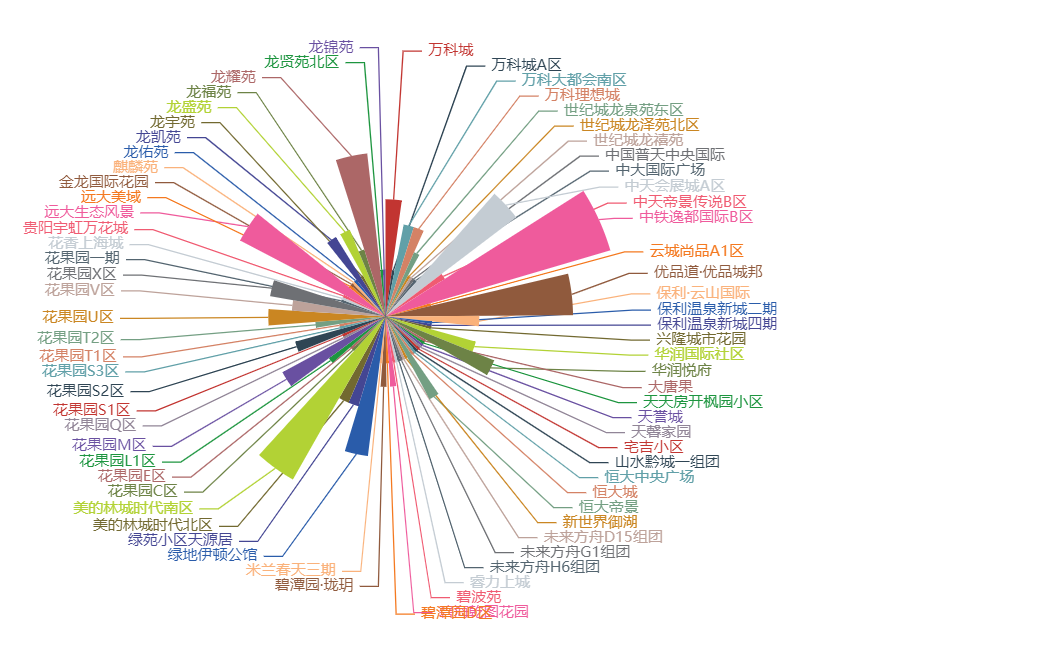
|  |
| --- |
| 源码如下 |
| sql捕获 |

**任务5：【数据分析与可视化展示】**

**可视化（1）:**

使用sql语句查询贵阳各地房产分布。

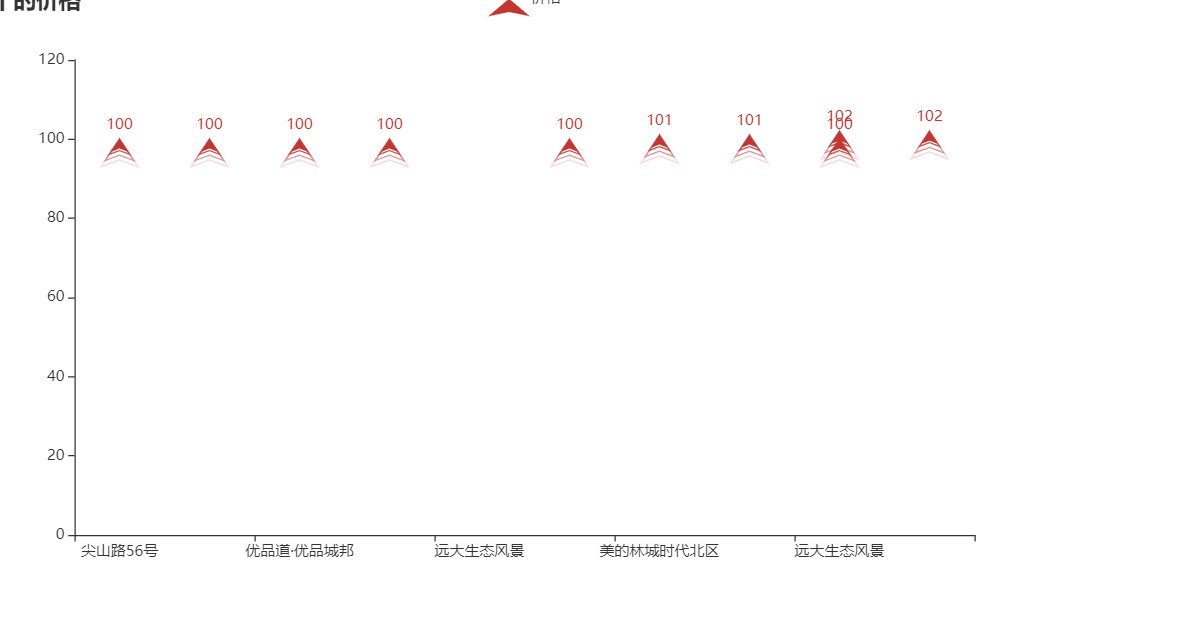
|  |
| --- |
| 源码及图形如下 |
| **发布捕获** |

****

**可视化（2）:**

SQL查询价格在100万左右的房子，取10条

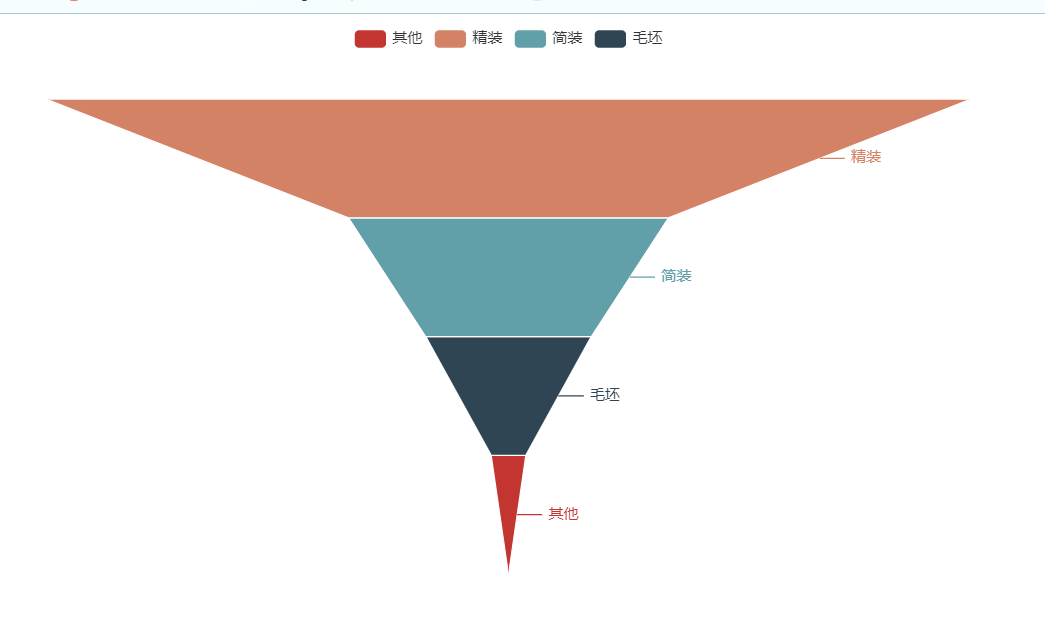
|  |
| --- |
| 源码及图形如下 |
| **价格捕获** |

****

**可视化（3）:**

SQL查询所有房装修情况并统计。

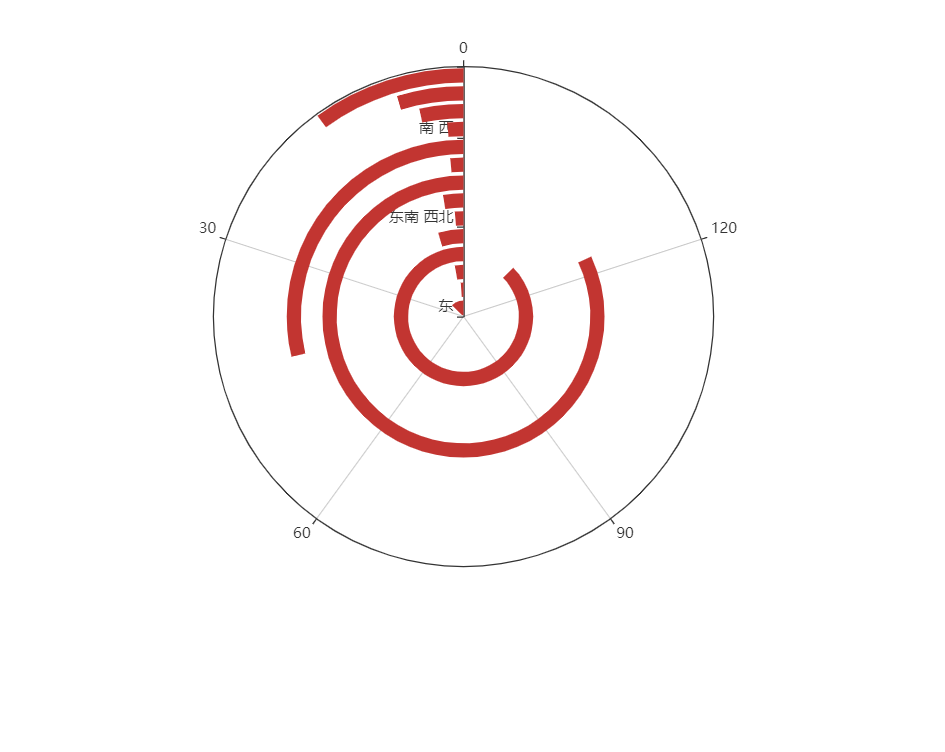
|  |
| --- |
| 源码及图形如下 |
| **装修捕获** |

****

**可视化（4）:**

用SQL查询所有房子查询并统计

|  |
| --- |
| 源码及图形如下 |
| **朝向d捕获** |

****

**可视化（5）:还有一些不一一展示了，**

**程序使用了flask框架，作为数据库内容的展示**

****

**运行 开启SQLite数据库，运行 scrapy 框架，**

**数据库保存到SQlite数据库，-运行testCloud.py文件，**

**写入数据，运行app.py，打开网站，**