

# 中国大学生计算机设计大赛



## 大数据实践赛作品报告

小组成员： 新闻传播学院 何雨馨 李葛洁

作品名称： 应届毕业生上海租房指南

填写日期： 2022 年 12 月 20 日

### 填写说明：

- 1、本文档适用于大数据实践小类；
- 2、正文一律用小四号宋体，1.3 倍行距；一级标题为二号黑体，其他级别标题如有需要，可根据需要设置；
- 3、本文档应结构清晰，突出重点，适当配合图表，描述准确，不易冗长拖沓；
- 4、提交文档时，以 PDF 格式提交；
- 5、本文档内容是正式参赛内容的组成部分，务必真实填写。如不属实，将导致奖项等级降低甚至终止本作品参加比赛。

# 目 录

第 1 章 作品概述.....	1
第 2 章 问题描述.....	1
第 3 章 技术方案.....	6
第 4 章 系统实现.....	7
第 5 章 系统评测.....	8
第 6 章 安装使用/平台呈现.....	9
第 7 章 作品总结.....	12

# 第1章 作品概述

据教育部统计，应届毕业生数量逐年增长，2022 年达到新高，毕业生的去向成为值得关注的问题。无论就业或升学，住宿都是不可避免的话题，越来越多的应届毕业生正在采用租房的方式开启人生新阶段。然而面对租房，初出茅庐的毕业生们仍然有许多困惑。

本实践作品基于 2022 年最新应届毕业生现状，立足现实需求，关注应届毕业生的租房状况。上海，不仅是全国各大高校毕业生的选择，而且是南京大学毕业生就业的重要地区，仅次于江苏省。同时，上海高昂的生活成本、复杂的交通状况、繁杂的信息也让许多毕业生在寻找租房房源时产生困惑与烦恼，这是值得我们关注的问题。

本小组试图通过对数据的海量收集，并结合应届毕业生的实际情况为逐梦东方明珠的青年学子们提供租房指导。

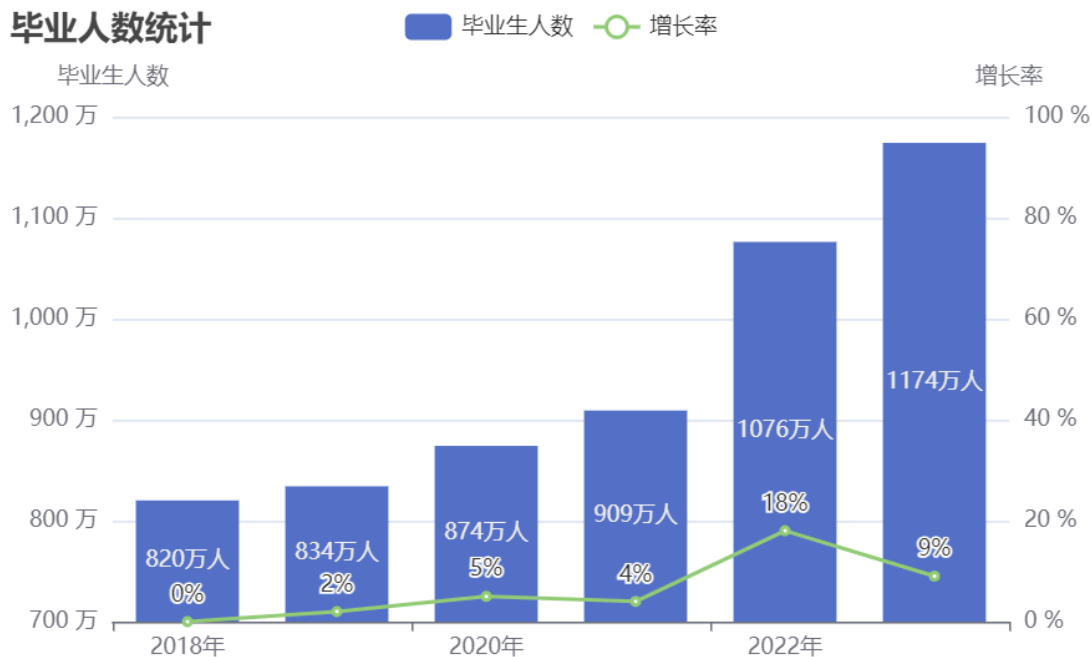
在技术方面具体而言：首先我们广泛收集二手资料，使用 echart 和 excel 统计绘制应届毕业生数量、收入、城市流向图表；其次我们使用 python 语言中 requests 函数库爬取租房市场上几个较大平台的上海市租房数据，如链家、房天下、安居客等，具体关注上海租房信息中的位置、交通、朝向、价格等在租房中考量的重要因素，运用 finebi 处理绘制出一张综合图表，直观展示上海租房现状；最后，结合应届毕业生实际情况与上海租房信息，我们使用 python 进行数据处理，试图回答应届生关注的问题——“我们能租到什么样的房？”“离实现租房自由还有多远？”。

最终，我们以图表+文字的形式为应届毕业生展示上海租房现状，提供大数据分析下的租房参考。希望能够让毕业生少走弯路、少信息差，多选择权、多幸福感。

# 第2章 问题描述

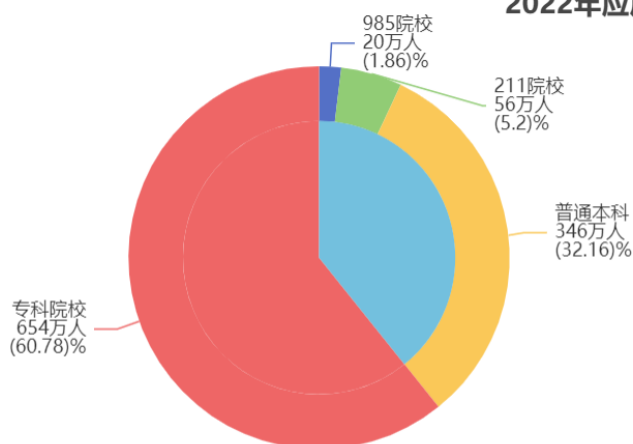
根据教育部的公开资料，2018 年至 2023 年，高校应届毕业生人数逐年增长，且增长率在 2022 年较前一年提升了 18%，达到 1076 万人。2022 年应届毕业生中，422 万人学历为本科及以上（共 422 万人）。

## 毕业生人数统计



- 985院校
- 211院校
- 普通本科
- 专科院校

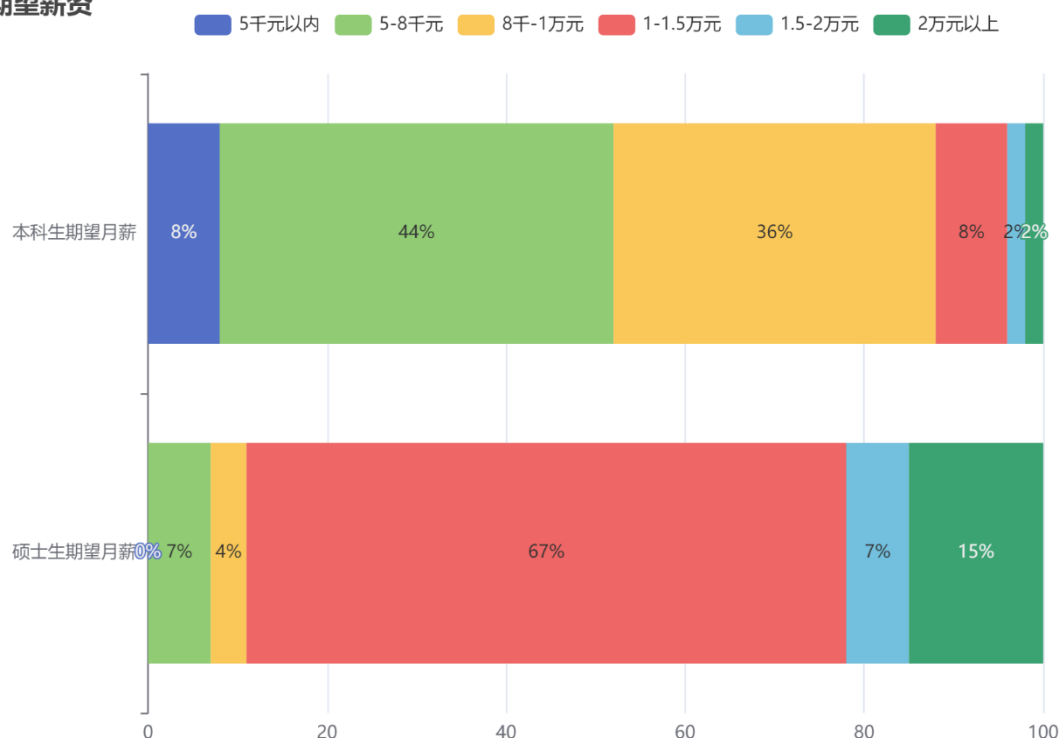
## 2022年应届毕业生学历分布



数据来源：教育部 图表绘制：echart

对于本科及以上学历的应届生，他们所期望的月薪也有所不同。整体而言，本科生期望月薪大多低于一万元，高于一万元仅占比 12%，大多集中在 5-8 千元（占比 44%）；硕士研究生对月薪的期待大多高于一万元，低于一万元仅占比 11%，大多集中在 1-1.6 万元（占比 67%）【数据来源：中智咨询《2021-2022 年应届生求职就业与薪酬调研报告》】。

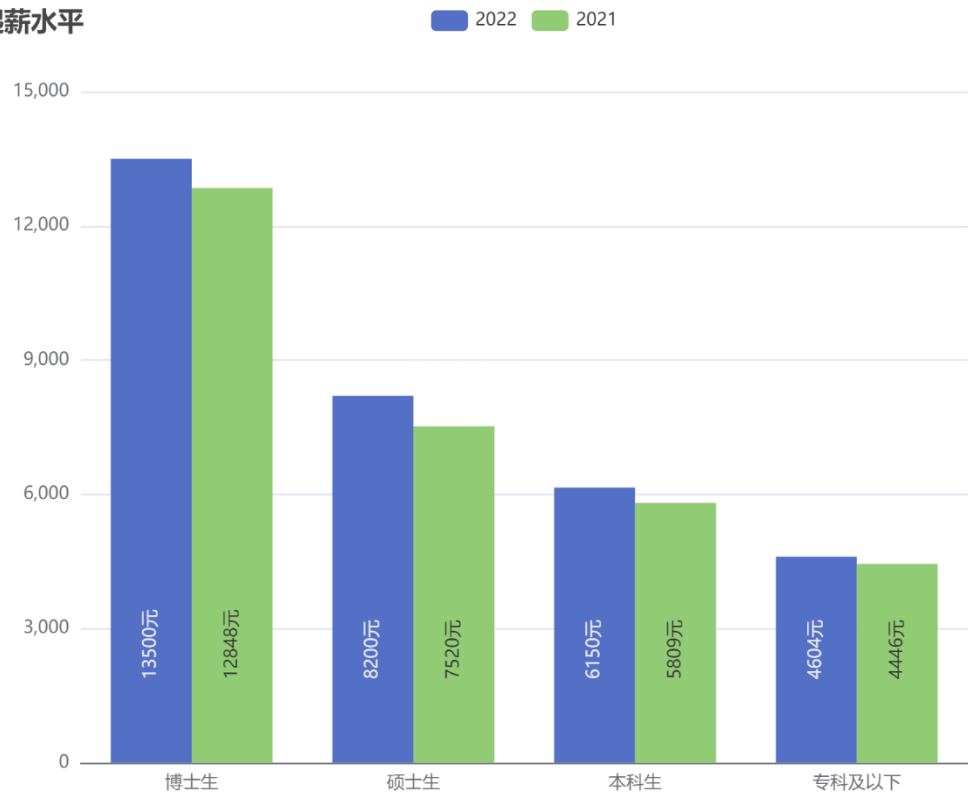
## 期望薪资



数据来源：中智咨询《2021-2022 年应届生求职就业与薪酬调研报告》 图表绘制：echart

而与期望相比，2022 年应届生的实际起薪差距并不大。本科生起薪为 6150 元/月，硕士生为 8200 元/月，相较 2021 年有所增长，但都并未突破万元大关【数据来源：中智咨询《2022 年应届生招聘和薪酬调研报告》】。不同专业的应届生起薪情况也有所不同，整体而言理工医科学生起薪高于文学艺术类学生【数据来源：太和顾问《20222 年企业应届生政策专项调研》】；不同城市应届生的平均年薪也有所不同，北京、上海、深圳的平均年薪居于前三位——其中，上海的新发职位在所有城市中占比最高，达到 10.6%【数据来源：猎聘】。

起薪水平



数据来源：中智咨询《2022 年应届生招聘和薪酬调研报告》 图表绘制：echart

2022年中国不同城市企业对不同专业应届生的起薪情况（单位：元）										
区域	管理学	经济学	法学	哲学	理学	农学	历史学	工学	医学	艺术学
整体	6500	6700	7000	5600	7000	7000	6000	7990	7650	6000
北上广深	6900	7100	7300	6000	8000	8200	6250	8000	8750	7000
新一线城市	6550	6300	7000	5750	7800	8000	6000	7980	6800	6000
其他地区	6000	6000	6800	5500	6500	5000	4800	6500	6000	4500

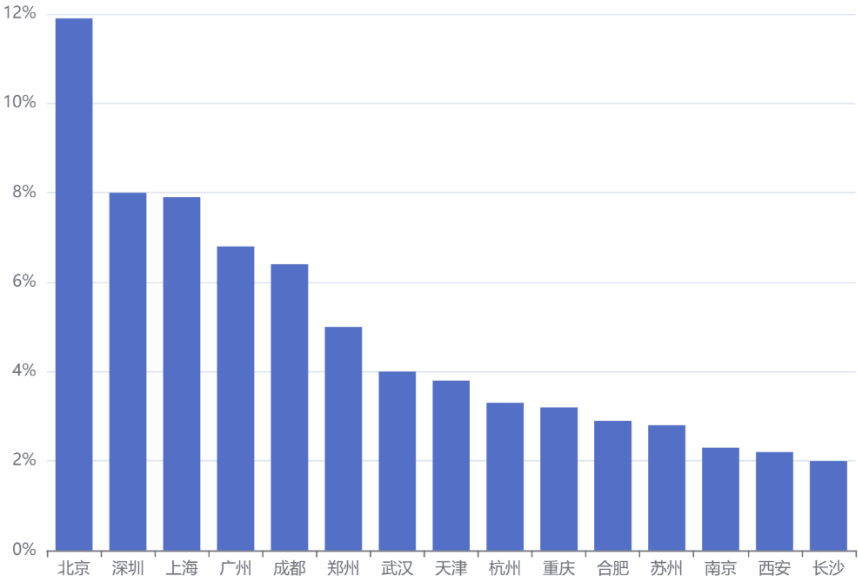
数据来源：太和顾问《2022 年企业应届生政策专项调研》 图表绘制：excel

2022年中国不同城市应届生职位的平均年薪与新发职位占比		
城市	平均年薪（万元）	新发职位占比（%）
北京	20.4	10
上海	19.7	10.6
深圳	18.9	6.5
杭州	17.9	4.5
西安	16.4	2.2
南京	15.9	2.5
武汉	14.3	2.8
苏州	14.3	2.8
东莞	13.7	1.6
成都	13.6	4.3
广州	13.4	5.9
长沙	13	1.8
合肥	12.7	1.9
青岛	12.5	1.6
天津	12.4	1.4
宁波	11.9	1.5
济南	11.8	1.3
重庆	11.5	2.4
佛山	11	1.8
郑州	11	1.6

数据来源：猎聘 图表绘制：excel

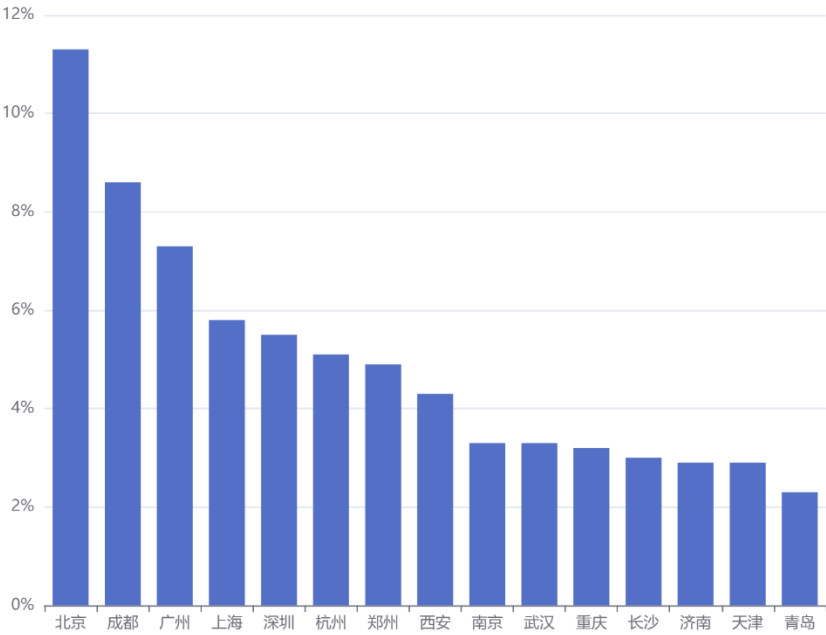
当应届毕业生面对具体的就业城市时，上海不仅招聘需求高，而且整体薪资水平高，所以实际求职人数较多，达到了 6.41 万人次的应届毕业生的净流入量，仅次于北京和深圳【数据来源：中国联通智慧足迹】。而对于南京大学的应届毕业生而言，前往上海的应届毕业生人数仅次于江苏省（整体约占比 49%）——2022 年应届生中，留在上海就业的应届毕业生约占比 12%，其中，本科生占比为 12.19%，硕士研究生占比约为 13.18%，博士研究生占比约为 8.99%【数据来源：南大就业】。

2022年应届毕业生招聘需求



数据来源：智联招聘 图表绘制：echart

2022年应届毕业生求职申请



数据来源：智联招聘 图表绘制：echart

2022年不同城市应届毕业生流动情况 (单位: 万人)		
城市	本地大学生毕业人数	净流入应届生数量
武汉	29.49	-8.44
昆明	19.37	-5.89
南昌	16.1	-5.74
郑州	28.44	-5.71
哈尔滨	14.92	-5.65
长沙	20.19	-4.59
长春	13.8	-3.82
兰州	9.34	-3.81
西安	22.93	-3.46
天津	17.4	-3.41
南宁	12.51	-3.39
贵阳	11.49	-3.19
广州	34	-3.13
南京	21.01	-2.54
合肥	14.42	-1.96
太原	9.72	-1.89
呼和浩特	6.25	-1.8
济南	17.73	-1.57
大连	8.24	-1.5
福州	10.01	-1.4
沈阳	12.78	-1.13
重庆	24.21	-1.1
乌鲁木齐	5.4	-0.84
成都	27.97	-0.6
海口	3.63	-0.56
拉萨	2.57	-0.45
银川	3.1	-0.43
西宁	1.86	0.08
青岛	11.56	0.29
厦门	5.01	0.65
宁波	4.44	2.57
杭州	14.61	5.21
上海	20.7	6.41
北京	24.29	7.15
深圳	4.11	14.54

数据来源: 中国联通智慧足迹 图表绘制: excel

在这样的现状下，应届毕业生在上海如何租到合适的房子成为重中之重。

然而，租房软件套路不断、小红书豆瓣平台信息良莠不齐、初来乍到的懵懂无知都让毕业生们屡撞南墙。应届毕业生们亟需大数据分析，提供租房指南。

## 第3章 技术方案

第一，资料分析阶段。

查找租房平台的年报，通过已有数据了解二手租房市场现状，以及应届毕业生实际的收入、就业城市情况。通过对二手数据的分析考察，首先确定以上海作为小切口展开数据分析，其次筛选出应届生关心的因素:①价格②地理位置（在实际实践中体现为上海行政区划）③房间朝向④交通便利程度（体现为与地铁站的距离），最后我们发现存在很多年轻人一起租房的情况，于是确认在具体计算金额时选择房间数作为最小单位而不是房子。在这个部分，我们使用 echart 和 excel 统计绘制应届毕业生数量、收入、城市流向图表，作为整个方案的背景介绍部分，引入我们想要解决的核心问题。

第二，数据搜集阶段。

选取了当前租房市场三个顶流平台房天下、链家、安居客，选择包含价格、地理位置、房间朝向、与地铁站距离等因素的房源信息，使用 python 语



言中 requests 函数库爬取信息。

### 第三，数据呈现阶段。

使用 finebi 绘制仪表盘，以图表形式展现上海市各区租金平均数和中位数、上海市出租房朝向情况、上海市出租房朝向对租金的影响、上海市各区每平米房价、上海市人均房租价格分布情况、上海市出租房到最近地铁站的距离分布、上海市出租房到最近地铁站的距离分布对距离的影响。

### 第四，总结回归阶段。

结合应届毕业生实际收入情况与上海租房信息，使用 python 进行处理，回归选题关注的问题。

调研数据显示，2000-3000 元为毕业青年能接受的主要租金范围，提取合租形式且租金在此范围内的租赁信息（整租则按照人均计算），分析此部分的房源情况，总结租房体验，回答“能租到什么样的房？”的问题。

通过计算上海城市 75%分位数租金，即可以租到上海样本小区中的 75%，同时按照租金不超过收入 30%的计算方法，测算出月收入水平，得出需要一个月挣多少钱才能实现租房自由，回答“离实现租房自由还有多远？”的问题。

## 第4章 系统实现

### • 数据从网页爬取&写入 excel 部分

#### 工具使用：

使用函数库：requests

使用函数：requests.get()

功能：将网页源代码爬取以待后续对其中信息的查找

使用函数库：lxml.etree

使用函数：etree.HTML()

功能：将网页源代码以 html 语言方式读取

使用函数：etree\_html.xpath()

功能：在网页中找到数据的 xpath，通过此函数在网页源代码中读取需要的信息

遇到的困难：有时网页数据不是以文本格式存储在源代码中，直接通过 xpath 读取 会输出为地址而不是文本

解决方式：在读取之后加.text 后缀读取其中文本

遇到的困难：读取之后的文本中有多余的空格，换行等不需要的字符

解决方式：在文本之后加 `.split( ' ' )` 后缀，将不需要的字符分割开来只取用其中 需要部分

使用函数库： `openpyxl.Workbook`

使用函数： `Workbook()`

功能：创建一个 excel 文件

使用函数： `workbook.create_sheet()`

功能：在已有 excel 文件中创建一个 excel 表格

使用函数： `sheet.cell()`

功能：在 excel 表格中指定位置输出指定文本

使用函数： `workbook.save()`

功能：将 excel 文件保存在相应位置

代码逻辑：

读取网页源代码→在源代码中找到需要的文本→将文本写入 excel 表格→将表格保存至电脑中

• 将 excel 中数据处理部分：

工具使用：

使用函数库： `openpyxl`

使用函数： `openpyxl.load_workbook()`

功能：读取 excel 文件中数据存储在数组中

代码逻辑：

从处理后的表格中读取“租房价格”以及“房间数量”存储在两个数组当中  
→用租房价格除以对应房间数得到该房平均每人租金存入“平均租金”数组中

→1. 检索“平均租金”数组中 2000-3000 元部分并且输出对应租房信息到.txt 文件中

→2. 对“平均租金”数组进行冒泡排序→输出其从低到高 75%位置的平均租金，除以 0.3 之后输出

## 第5章 系统评测

- 数据来源：房天下 链家 安居客
  - 数据测试：针对最终“实现租房自由”月薪须达到 22222 元进行数据测试
- ① 另取 200 份数据，使用租房金额/房间间数，得到平均每间的租房价格

② 与 6666 元（22222 元\*0.3，其中 0.3 指月薪的 30%用于租房）做对比

③ 统计小于等于该金额的房间数

④ 共 143 份，占比 71.5%，相对误差率约为 4.7%

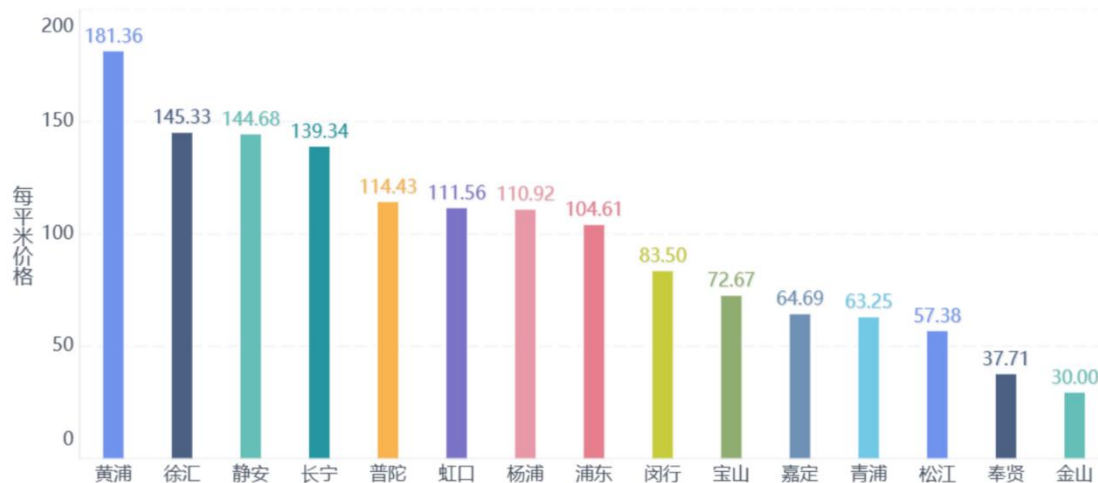
→在误差允许的范围内，该模型有效

- 其他验证：与小红书平台数据对比

上海各区市县租金		
排名	行政区	平均单价/月/M <sup>2</sup>
1	黄浦区	137.49
2	长宁区	130.86
3	静安区	123.01
4	虹口区	117.36
5	徐汇区	111.34
6	普陀区	106.87
7	杨浦区	103.11
8	浦东新区	89.93
9	闵行区	82.66
10	宝山区	81.42
11	嘉定区	79.12

上海租房攻略|应届生毕业生一定要知道

图片来源：小红书收藏超过 1500 笔记截图



结果基本一致。

## 第6章 安装使用/平台呈现

本数据分析结果拟在豆瓣、小红书、微博等年轻人常用社交平台发布。

拟发布内容格式如下：（以小红书平台为例）

标题：上海租房|预算 2-3k 应届生能在上海租到什么房？一图全知道！

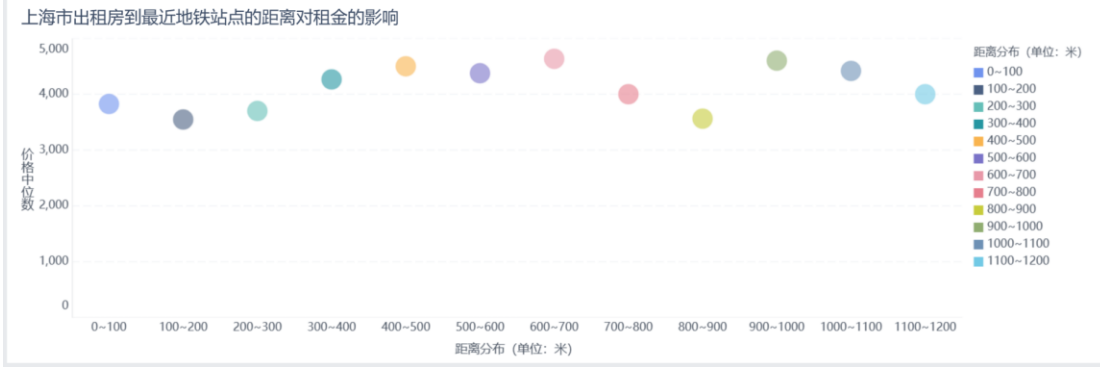
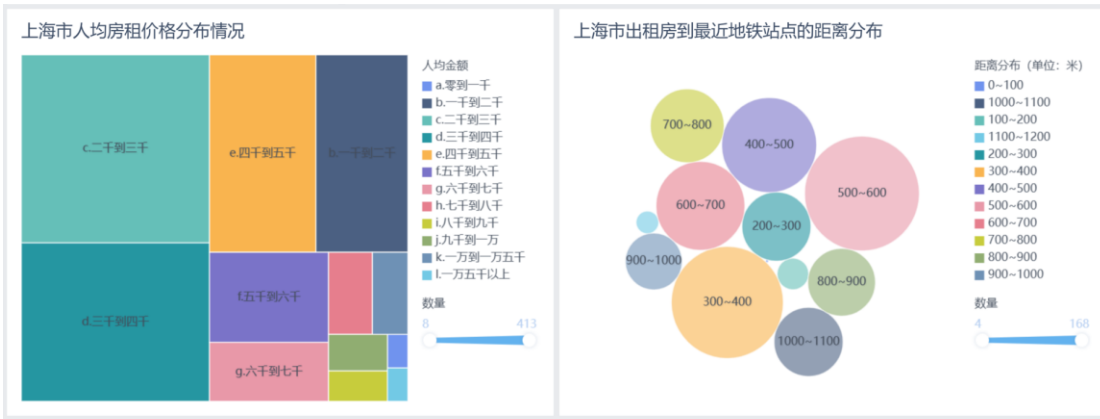
文案：

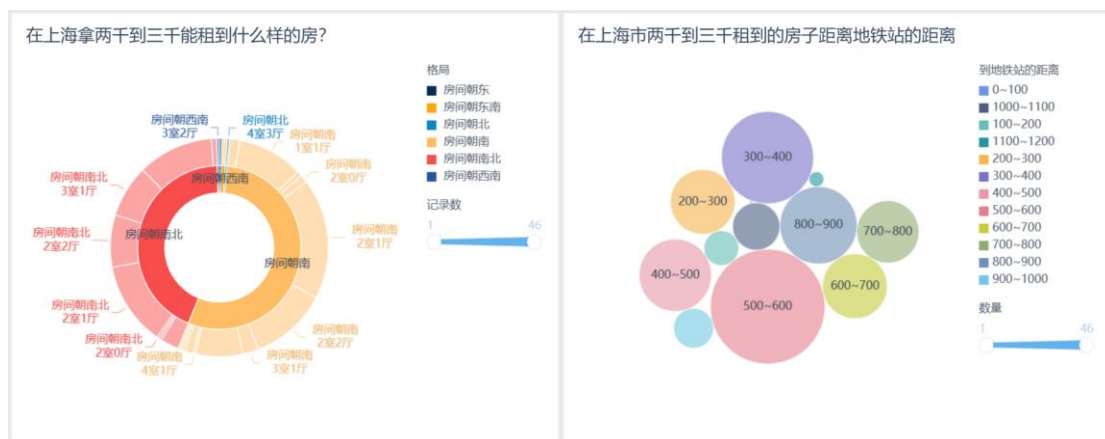
- 刚刚毕业走入社会，投简历、面试、入职……还有最让人头疼的租房。
- 上海那么大，月薪那么少；房子那么多，真诚那么少。
- 每天在不同平台、软件来回切换，频繁查看邮件、短信，只为了找到一个交通方便、地理位置合适、房屋朝向合理的房子，平稳安置一天后的疲乏。
- 最新应届毕业生上海租房指南来啦！
- 你感兴趣的，这里都有！无论是日常通勤地铁的距离，还是早晨第一缕阳光的方向，大数据分析统统都知道！
- 使用 python 爬取房天下、安居客、链家等租房市场上顶流平台的数据，结合年轻人最关注的租房因素，finebi 制图，简明展示上海不同地区租房金额，少走弯路，多做选择！

图片：



（首图）





标签：[#租房#](#) [#上海租房攻略/首选#](#) [#应届毕业生#](#) [#毕业季#](#)

拟发布时间：2023 年 6 月 1 日

补充说明：微博可以在“租房”“毕业生”“上海租房”等超话；豆瓣可以发布在“我们在租房生活”“上海租房”等小组。

## 第7章 作品总结

局限性和需要改进的地方：

- ① 数据清洗力度不够：在实际情境下，房租成交额并不完全等于市场平台上展示的价格，会有相应幅度的上升或下降，可以在收集数据后进行相对应的处理
- ② 数据涵盖不够全面：租房线上平台是一部分，线下中介、私人租赁等渠道并未被涵盖进我们的数据范围内