

## 需求分析

### **1.1** 项目需求

**1 抓取并分析租房数据：**

从链家官网抓取北京市、上海市、广州市、深圳市和天津市的所有租房数据。

对比这些城市的整体租房市场，包括租金均价、最高价、最低价和中位数，另外也要计算单位面积租金（元/平方米）的均值、最高值、最低值和中位数。

**2 城市间租房市场的详细比较：**

针对一居、二居和三居类型的租赁情况，统计并比较各城市的租金均价、最高价、最低价和中位数。

**3 城市不同板块租金分析：**

计算并分析各城市不同板块的平均租金情况，探索不同区域间的租金差异。

**4 朝向对租金的影响：**

比较各城市不同朝向的单位面积租金分布情况，分析朝向因素对租金价格的影响。

**5 经济因素与租金的关系分析：**

查询北京市、上海市、广州市、深圳市和天津市的最新人均GDP和平均工资数据，分析这些经济指标与单位面积租金之间的相关性，探索经济水平对租赁市场的影响。

### **1.2** 需求分析

**1全面获取租房数据：**

收集不同城市的尽可能完整的租房信息，确保涵盖所有相关的房源数据。

**2分类并提取租房信息：**

先提取所有信息，再提取、区分各租房信息的城市、具体板块、户型类型和价格等关键信息，以便进行后续分析。

**3获取经济指标：**

调查获取各城市的最新人均GDP和平均月工资数据，为后续的经济分析提供基础。

**4数据对比与展示：**

对各类租房数据进行比较分析，展示不同城市、不同板块、不同户型的租金分布及其他相关统计信息，帮助可视化不同区域和户型的租赁市场情况。

## 模块设计

### **2.1** 项目模块

我们将项目结构划分为主模块、数据抓取、数据处理、数据分析、数据呈现五个模块。

各模块间的主要顺序及关系如下图（图1）：

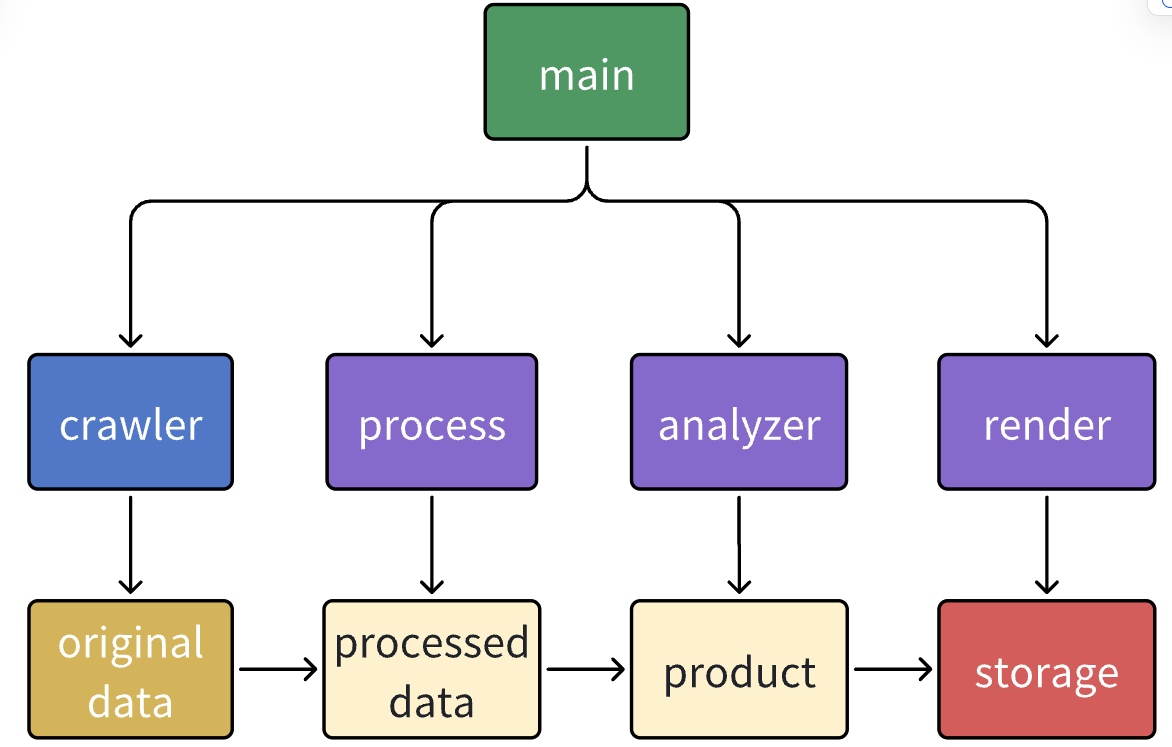


图1 项目模块关系图

### **2.2** 模块解析

**2.2.1 数据抓取**

数据抓取模块的任务是从链家官网抓取五个城市（北京、上海、广州、深圳、天津）所有房源的原始数据，确保数据无重复且完整。此模块将依托 Scrapy 爬虫框架进行数据抓取，使用XPATH路径进行 HTML 解析。抓取到的原始数据将被保存至 .json 格式的数据文件中，以便后续处理和分析。

**2.2.2 数据处理**

数据处理模块的核心任务是对从链家官网抓取到的原始数据进行清洗和预处理，去除无用信息、无效信息及异常值，确保数据的质量和一致性。该模块将使用 Pandas 数据处理库进行数据清洗、格式化和转换操作。处理后的数据将以 CSV 文件形式分类保存，便于后续分析和展示。

**2.2.3 数据分析**

数据分析模块的主要职责是对经过处理后的数据进行深入分析。该模块将根据需求进行不同类型的数据计算、统计和比较，作出数据趋势分析。这部分的分析结果将为后续的报告和展示提供支持。

**2.2.4 数据呈现**

数据呈现模块将基于经过处理和分析的数据，利用 Pandas 可视化库生成各类图表，并根据需求展示不同类别的数据分析结果。呈现的图表将保存为 .png 格式的文件。

**2.2.5 主模块**

主模块负责协调调用上述各个模块，完成整个项目的运行流程。该模块能够根据需求，灵活地配置各个模块，减少重复操作。

## 数据抓取

### 3.1 爬取策略与分类

为了确保获取完整的租房数据，我们采用了分类爬取的策略。由于链家网站每次访问时最多只能返回100页数据（包含3000条信息），因此仅通过直接访问获取到的数据量有限。为了克服这一限制，我们首先不加任何限制条件，获取初步数据。然后根据以下分类策略进一步细分爬取范围：

* **判断页面数量：** 首先判断当前页数是否为100，如果是，则需要进一步分类爬取。此时会依据区域、价格、房型、朝向、楼层等条件进行分区域、分价格等多级筛选。
* **区域筛选：** 首先按区域进行细分，获取不同区域的租房数据。
* **价格筛选：** 在区域筛选的基础上，再根据价格范围进一步筛选房源数据。
* **房型筛选：** 根据房型（一居、二居、三居等）对数据进行分类爬取。
* **朝向筛选：** 在房型筛选的基础上，进一步按朝向对数据进行筛选。
* **楼层筛选：** 最后在朝向的基础上，根据楼层对数据进行筛选。

通过上述多级筛选，避免了链家网站的反爬机制限制，实现了完整租房数据的获取。

### **3.2 反爬策略**

链家具有韧性较强的反爬机制，过于频繁的请求会导致 IP 被封禁或者重定向到验证码页面，得到302响应码。为了绕过这样的限制，我们采用了以下反爬策略：

* **随机 User-Agent：** 每次请求都使用随机的 User-Agent 模拟不同的浏览器，从而避免被检测为机器人。
* **代理池：** 使用[隧道代理服务](https://ip.ihuan.me/)。需要注意的是，每次间隔一定时长的时候都要切换IP，以及在使用代理的情况下，避免设置 cookies 触发反爬机制，不同ip共用cookies比直接访问更有嫌疑。
* **禁用 Cookies：** 为了避免链家根据 cookies 进行个性化推荐而收到重复的爬取结果，禁用了 cookies 设置。
* **请求延时：** 每次请求后都设置 1 秒的延时，避免频繁请求触发反爬机制。
* **重试机制：** 中间件中设置若未能成功获取到数据，系统会不断发起重试请求，增加爬取单一数据项的成功率。
* **自定义重定向重试：** 对于某些代理 IP 被封禁的情况，采用自定义的重定向重试机制，我们取消Scrapy默认的跟随重试，在上一页面进行重试。在 IP 被封禁时，可以通过更换 IP 重新请求，而不会直接跳转到验证码页面。

### **3.3 代码结构与流程**

**3.3.1 start\_requests 方法**

**功能：** 从城市的租房首页开始抓取数据。

**工作原理：** 通过 start\_url 向链家网站发起请求，初步获取城市的租房页面，开始解析。

**3.3.2 parse\_level 方法**

**功能：** 根据不同的层级（城市、区域、价格、房型等）对页面进行解析。

# parse\_level方法解析结构设计

        subdivision\_hierarchy = {

            'city': {

                'should\_subdivide': self.should\_subdivide(total\_pages, total\_houses), # 是否要再分级进行下一步parse

                'next\_level': 'area', # 下一步parse分级

                'extract\_urls': self.extract\_area\_urls # 下一步parse使用的url

            },

            'area': {

                'should\_subdivide': self.should\_subdivide(total\_pages, total\_houses), # 是否要再分级进行下一步parse

                'next\_level': 'price', # 下一步parse分级

                'url\_modifier': self.get\_price\_url, # 下一步parse使用的url

                'max\_pages': self.max\_price\_pages # 规定最大页面数

            },... # parse方法其余部分类似

**工作原理：**

* **解析总页数与总房源数：** 从页面中提取总页数和总房源数（total\_pages 和 total\_houses），并根据这些数据决定是否需要进一步细分。
* **判断是否需要细分：** 通过 should\_subdivide 方法判断是否需要进一步细分。如果页面数为 100 或房源数超过 3000，则进行进一步的分类爬取。
* **按区域、价格、房型等细分：** 根据不同的层级条件（区域、价格、房型、朝向等）递归爬取，逐级细化数据范围，直到没有进一步细分的需要。

**3.3.3 handle\_pagination 方法**

**功能：** 处理分页逻辑。

**工作原理：** 如果当前页面存在多个分页（总页数不为 1），则为每一页生成请求，递归获取所有分页的租房信息。

**3.3.4 parse\_data 方法**

**功能：** 从每个页面中提取租房信息。

for house in response.xpath(“xpath”):

            item = RentHouseItem()

            item["title"] = self.extract\_text(

                house,

                “xpath”,

                "N/A"

            )... # 提取每个数据项的部分属性：无效项记为N/A

**工作原理：**

使用 XPath 提取每个房源的信息，如标题、价格、描述、底部信息等，提前进行一部分预处理，提前规定无效数据部分置为N/A，封装为 RentHouseItem 对象。将提取的数据作为爬取结果返回，供后续存储或进一步分析。

**3.3.5 数据提取与清洗函数**

**extract\_text：** 提取单个文本字段（如房源标题、品牌等）。

**extract\_descriptions：** 提取并清洗房源描述信息。

descriptions = house.xpath(“xpath”).getall()

        return ["".join(text.split()) for text in descriptions if text.strip() and text.strip() not in ["-", "/"]]

# 提取并清理每个文本元素：去除多余的空格，排除无用的字符（"-"和"/"）

**extract\_bottom\_info：** 提取房源底部信息（如是否有电梯、是否有停车位等）。

bottom\_info = house.xpath(“xpath”).getall()

        return [text.strip() for text in bottom\_info if text.strip()] or []

# 将文本元素内的值逐个提取出来

**get\_page\_info：** 获取当前页面的总页数和房源总数，用于判断是否需要分页或进一步细分。

**3.3.6 URL 生成与修改**

**get\_price\_url、get\_room\_url、get\_direction\_url、get\_floor\_url：** 根据不同的条件（价格、房型、朝向、楼层等），生成相应的 URL，用于分页或进一步的细分爬取。

## 数据处理

### **4.1 数据处理模块**

本项目主要采用在/processpy文件夹下使用多个python文件逐步地对源数据进行提取、分析与计算，可以实现包含数据清洗、数据提取、格式化处理等多项功能在内的数据预处理，也可以在处理后的数据上进行关键数据计算，方便后续的图片绘制。

### 4.2 数据预处理

由于链家的原始网页数据字段中，混杂包含着价格、均价、朝向和户型等信息，因此需要我们对这些数据进行数据清洗与格式化处理。

原始原子数据格式如下：

{"title": "整租·高塔小区 3室1厅 南/北", "info": ["延庆", "延庆其它", "高塔小区", "93.20㎡", "南北", "3室1厅1卫", "低楼层（5层）"], "bottom": ["官方核验", "新上", "精装", "集中供暖"], "source": "链家", "price": "1700"},...

* 提取、判断单位面积租金

info表项中包含了该房屋的室内面积，检索到含有“㎡”的字段就进行替换与分割，提取出来面积大小；对应地，price表项中直接包含了该房屋的月租金数值（详情见爬虫源代码），故直接提取，遇到数值缺失的情况，默认记录为0。

提取单位面积租金处理代码如下：（unit\_price\_process.py）

area\_str = next((item.replace("㎡", "").replace(",", "").strip() for item in listing.get("info", []) if "㎡" in item), None)

price\_str = listing.get("price", "0元/月").replace("元/月", "").replace(",", "").strip()

针对需要判断是否有效的表项，设计is\_valid族方法来判断是否有效，无效则返回false

price = float(price\_str.replace("元/月", ""))

return price > 0

...

* 提取户型信息

info表项中包含了该房屋的户型信息，检索到含有“室”的字段就进行替换与分割，提取出来户型前的数字，写入该信息记录的room\_type属性值当中。

提取户型信息处理代码如下：（room\_priceall\_process.py）

info = listing.get("info", [])

room\_type = ""

for item in info:

match = re.match(r"(\d+)室", item)

if match:

num = int(match.group(1))

# 根据房间数量分类

if num == 1:

room\_type = "一居"

elif...

break # 找到后停止

if not room\_type:

room\_type = "未知"

* 提取、判断板块信息

info表项中包含了该房屋的板块信息，检索到info字段中的第二元素进行分割，提取出来对应的城市版块，写入该信息记录的district变量当中。

提取板块信息处理代码如下：（block\_price\_process.py）

info = listing.get("info", [])

if len(info) < 2:

continue

district = info[1].strip() # info对应第二元素即为板块信息

* 提取、判断朝向信息

info表项中包含了该房屋的朝向信息，我们采用关键字搜索的方式，提前规定有效朝向向量组，遍历info中的每一字段，可以保证提取有效朝向。

提取朝向信息处理代码如下：（direction\_unitprice\_process.py）

VALID\_DIRECTIONS = {"东", "南", "西", "北", "南北"}

info = listing.get("info", [])

# 提取朝向

direction = next((item for item in info if item in VALID\_DIRECTIONS), "")

### **4.3 关键数据计算**

**4.3.1 四分位数据计算**

在数据统计的过程中，四分位数的重要性不言而喻：方便地看到数据的集中程度和离散程度、四分位距帮助识别数据中的异常值、不受极端值的影响、可以判断数据的偏斜程度，还可以使得数据的“中间”部分更加清晰。

我们在绘制箱线图时，以及需要直观地得到非复杂数据的分布以及统计信息时，都需要运用到四分位数，因此我们设计了一套计算方法，实现代码如下：（priceall\_process.py）

stats = {

"均值": np.round(np.mean(data\_np), 2),

"标准差": np.round(np.std(data\_np, ddof=1), 2),

"最高值": np.round(np.max(data\_np), 2),

"最低值": np.round(np.min(data\_np), 2),

"中位数": np.round(np.median(data\_np), 2),

"25%分位数": np.round(np.percentile(data\_np, 25), 2),

"75%分位数": np.round(np.percentile(data\_np, 75), 2),

}

**4.3.2 性价比(Cost-Performance Index,CPI)计算**

在任务“地方省市人均GDP和平均工资(ASP)与租房信息对比”中，我们需要一个合适的指标来描述地方省市人均GDP和租房情况，以及地方省市平均工资和租房情况之间的关系，故考虑使用花销-效果指数，在我国俗称为“性价比”，在本项目中，简称为CPI。

我们采用如下函数来实现CPI的计算：（注意CPI在不同章节里代表的含义并不相同，分为CPI for GDP与CPI for ASP，下列两个代码块分别属于cpi\_bar\_render.py与salary\_cpi\_bar\_render.py）

mean\_rent["性价比 (CPI)"] = mean\_rent["人均 GDP（元）"] / mean\_rent["单位面积租金（元/月/㎡）"]

mean\_rent["性价比 (CPI)"] = mean\_rent["平均工资（元）"] / mean\_rent["单位面积租金（元/月/㎡）"]

## 数据展示分析

### **5.1 绘图模块**

本项目采用了Matplotlib完成了几乎所有数据图表的绘制，使用pyecharts完成了户型与均价关系的立体直方图。

Matplotlib是一个广泛使用的 Python 数据可视化库，能够生成各种类型的图表，广泛应用于科学计算、数据分析以及学术研究中。pyplot 模块提供简单的编程接口，用户通过一系列的命令就可以创建图表，并且对图表进行细致的定制，而且可以很方便地与 NumPy、Pandas等数据处理模块配合使用。

Pyecharts 是一个基于 Echarts 的数据可视化库，主要用于生成交互式的图表，能够以 HTML 文件的形式展示。Pyecharts 是基于 JavaScript 的 Echarts 图表库的 Python 接口，支持复杂的交互功能，例如支持鼠标悬停时显示数据点信息、图表的动态更新等。

### **5.2 图像展示分析**

**5.2.1 总体数据情况展示**

展示各个城市的所有租房数据，研究各城市均价和单位面积租金的整体情况，采用直方图和雷达图进行绘制，如图2、图3、图4以及图5。五座城市（北京、上海、广州、深圳、天津）的租房数据叠加在同一张图中，如图4、图5。

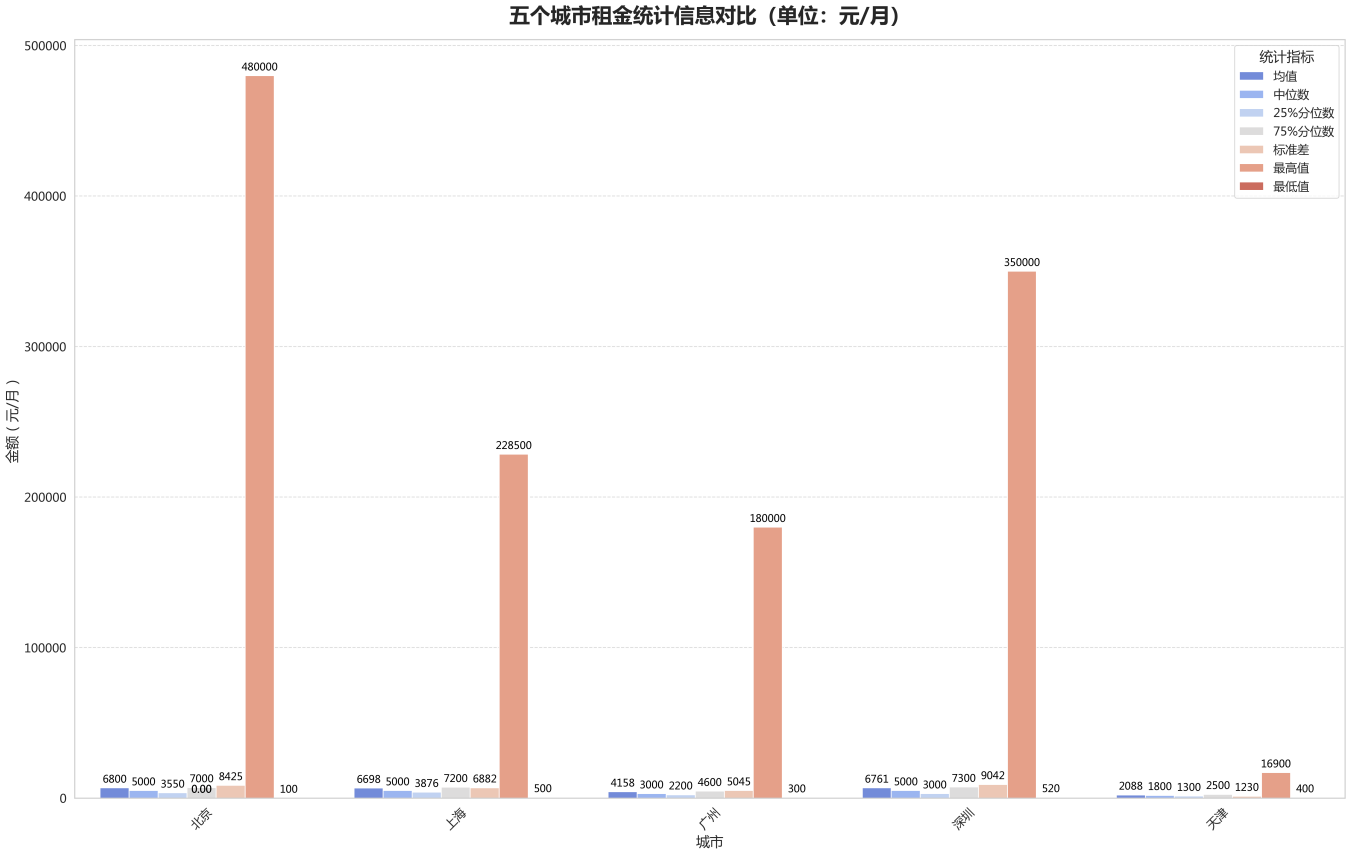


图2 五个城市租金统计信息对比图

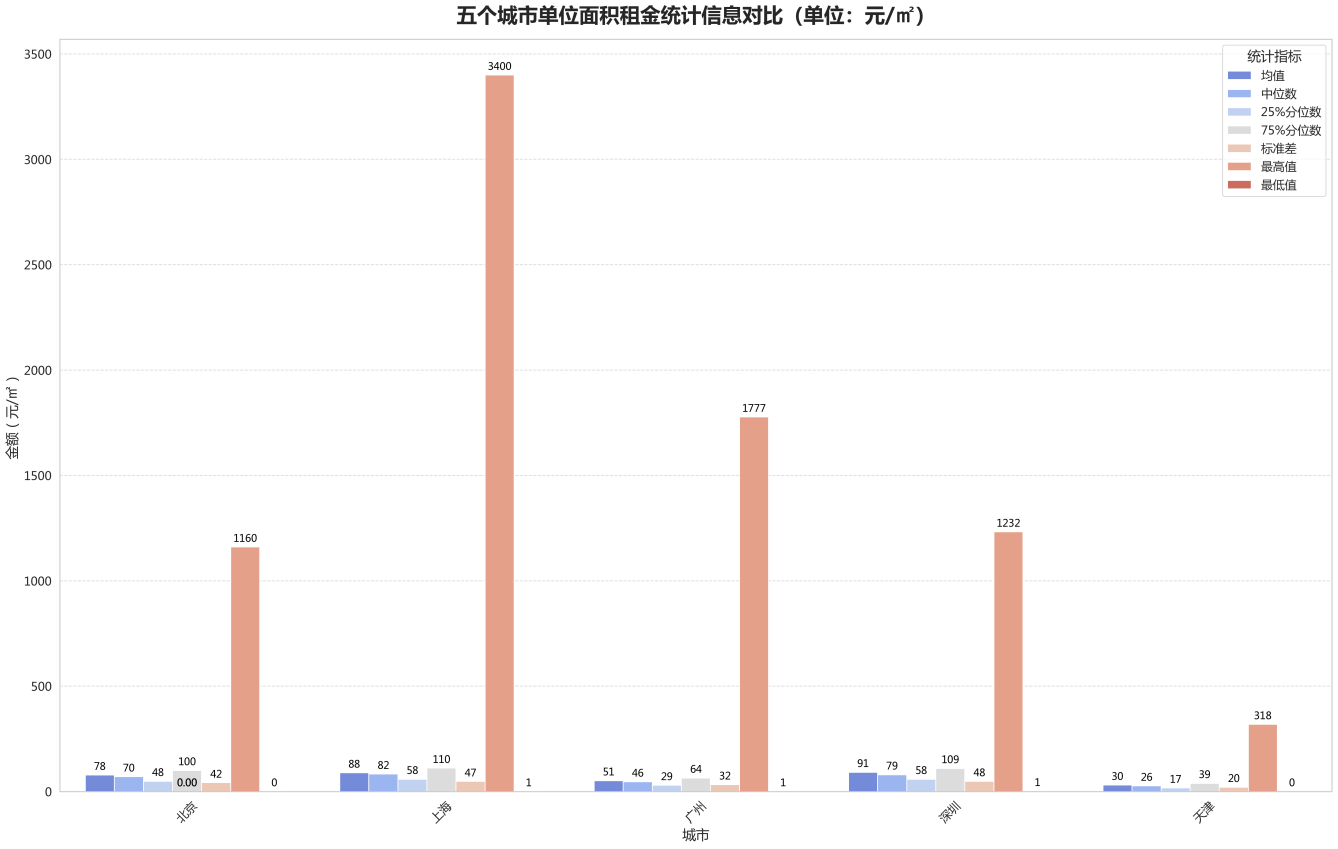


图3 五个城市单位面积租金统计信息对比图

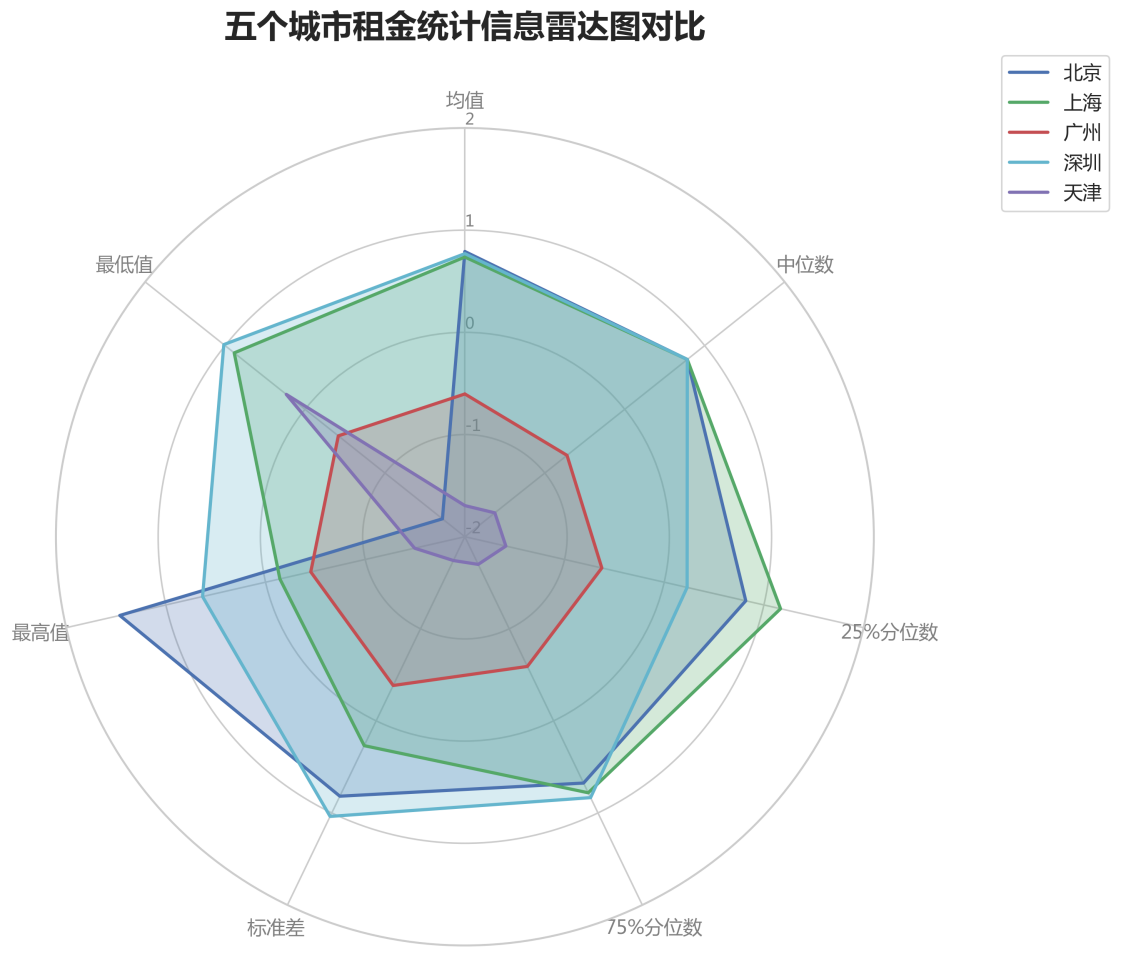


图4 五个城市租金统计信息雷达图对比

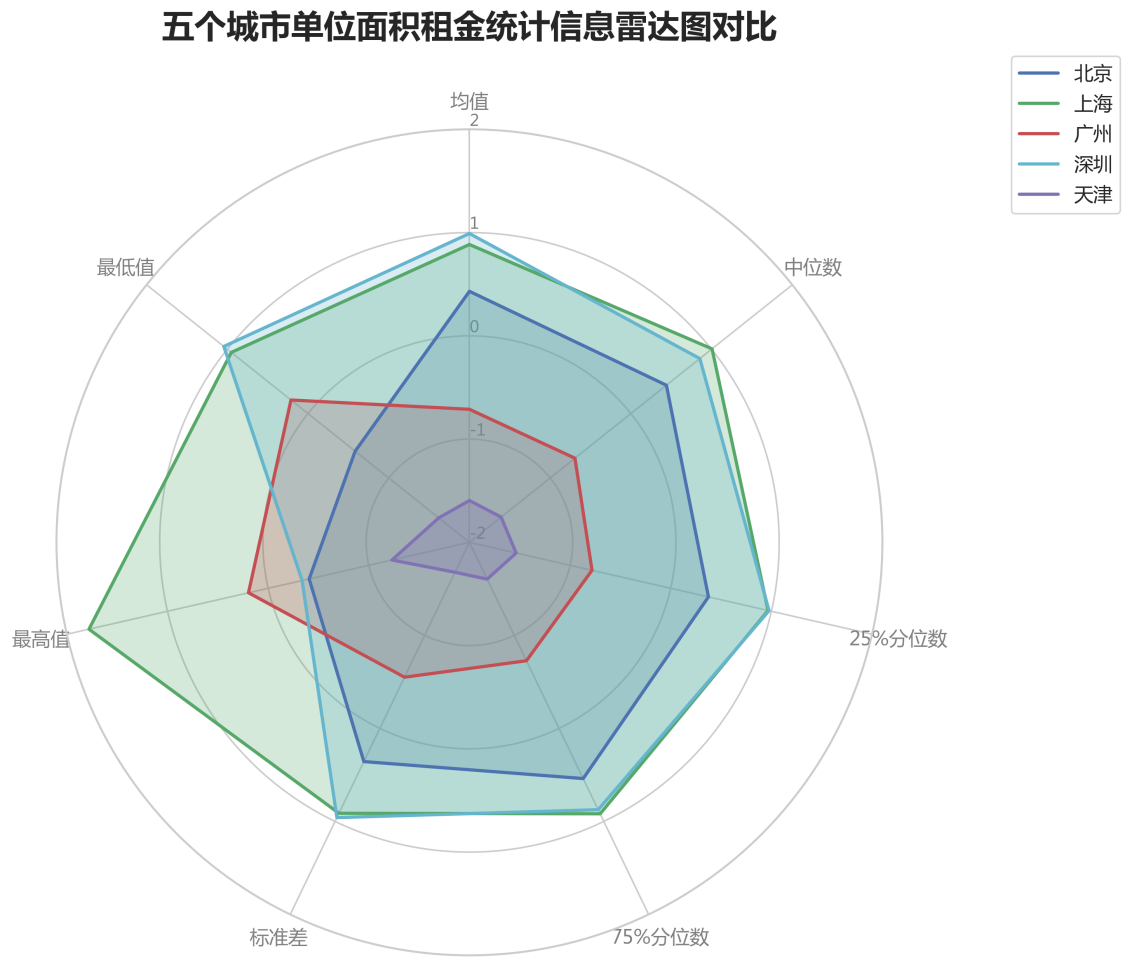


图5 五个城市单位面积租金统计信息雷达图对比

**5.2.2 总体数据情况解析**

图2-图5这几张柱状图和雷达图，展示了五个城市的租金统计信息与单位面积租金的分布特点。

从整体分析可以看出，北京和上海的租金水平最高，且存在显著的高租金房源拉升均值的情况。具体来说，北京的租金最高值达到480,000元/月，显著高于其他城市，同时均值与中位数差距较大，反映了租金分布的不均衡性。上海的最高租金为228,500元/月，也表现出类似的特点，且标准差较高。深圳的最高租金居于北京与上海之间，均值和中位数较接近，显示深圳租房市场相对稳定。广州和天津则表现出较为集中且低廉的租金分布，其中天津的均值最低，仅为2,088元/月。

从单位面积租金的统计来看，上海的单位面积租金最高，达到3,400元/㎡，体现了其房屋利用率和市场价值的领先优势。北京和深圳次之，但深圳的单位面积租金分布更均匀。广州和天津的单位面积租金最低，市场整体负担较小。

雷达图则清晰展现了各城市的统计指标差异，北京、上海和深圳在最高值和标准差上相互接近，而广州和天津在中位数、均值等指标上表现出更接近普通租客需求的市场特征。综合来看，北京、上海和深圳适合高端市场，广州和天津更贴近大众需求。

在图4中，北京具有出乎意料的租金最低值，不排除是地下室等过于狭小、不宜居的场所；天津的租金最低值相对较高的同时，最高值相当低，说明天津市场极端值非常少，租房市场相当有序。而在图5中，我们可以看到几乎各项指标都呈现出上海≈深圳＞北京＞广州＞天津，和我们得出的总体结论相似；此外当我们研究单位面积租金时，所有城市的图谱都比较均衡，可以看到租金和单位面积租金两个指标虽然都是租金，但是实际差异相当大。

**5.2.3 多城市户型数据情况展示**

图6-图8这组三维柱状图详细展示了五个城市中一居室、二居室、三居室及四居室以上房型的租金分布情况，通过对比各户型的均价、最大值、最小值及中位数，直观呈现不同城市在各类房型上的租金差异。

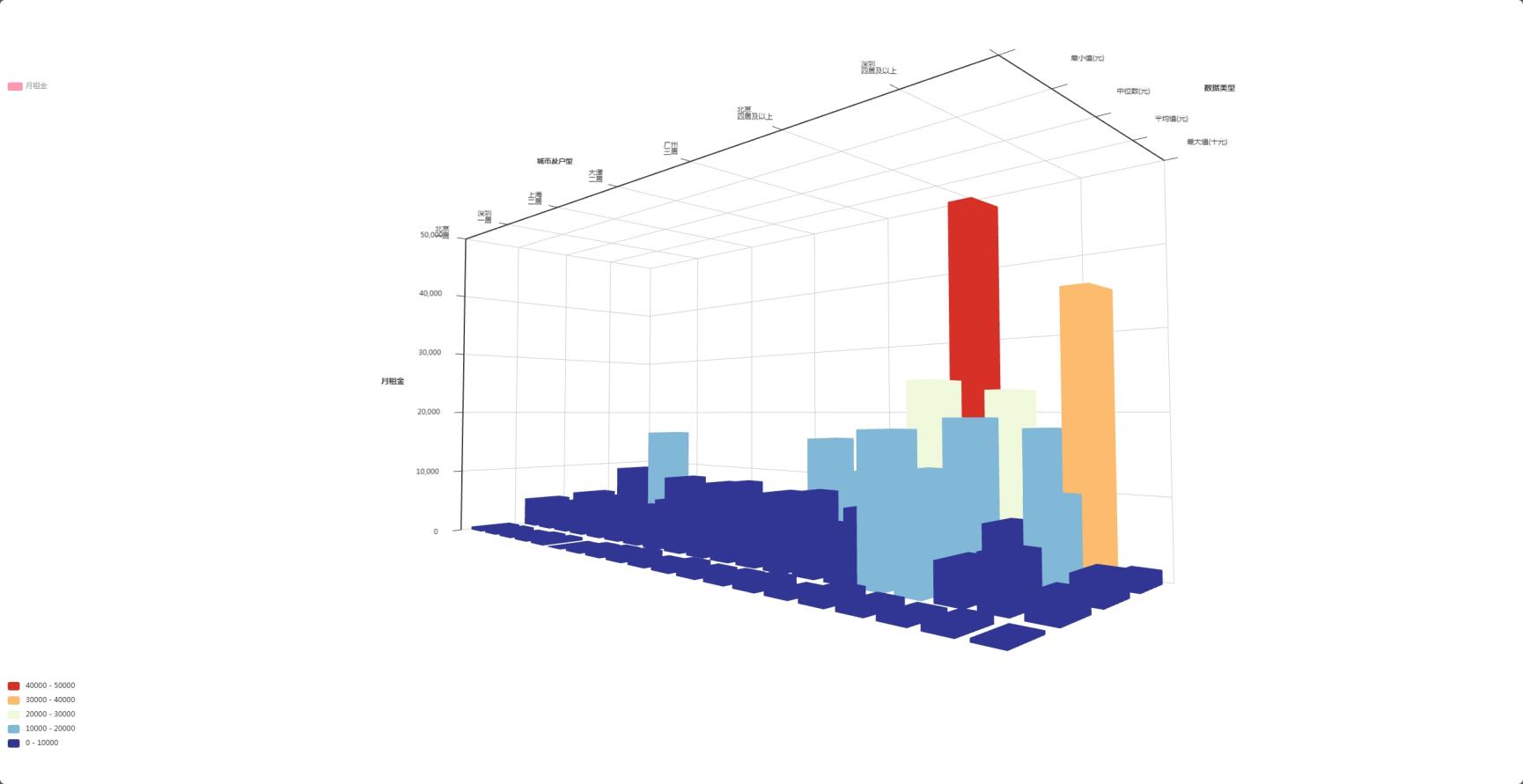


图6 五个城市房型的租金分布情况（斜视图）

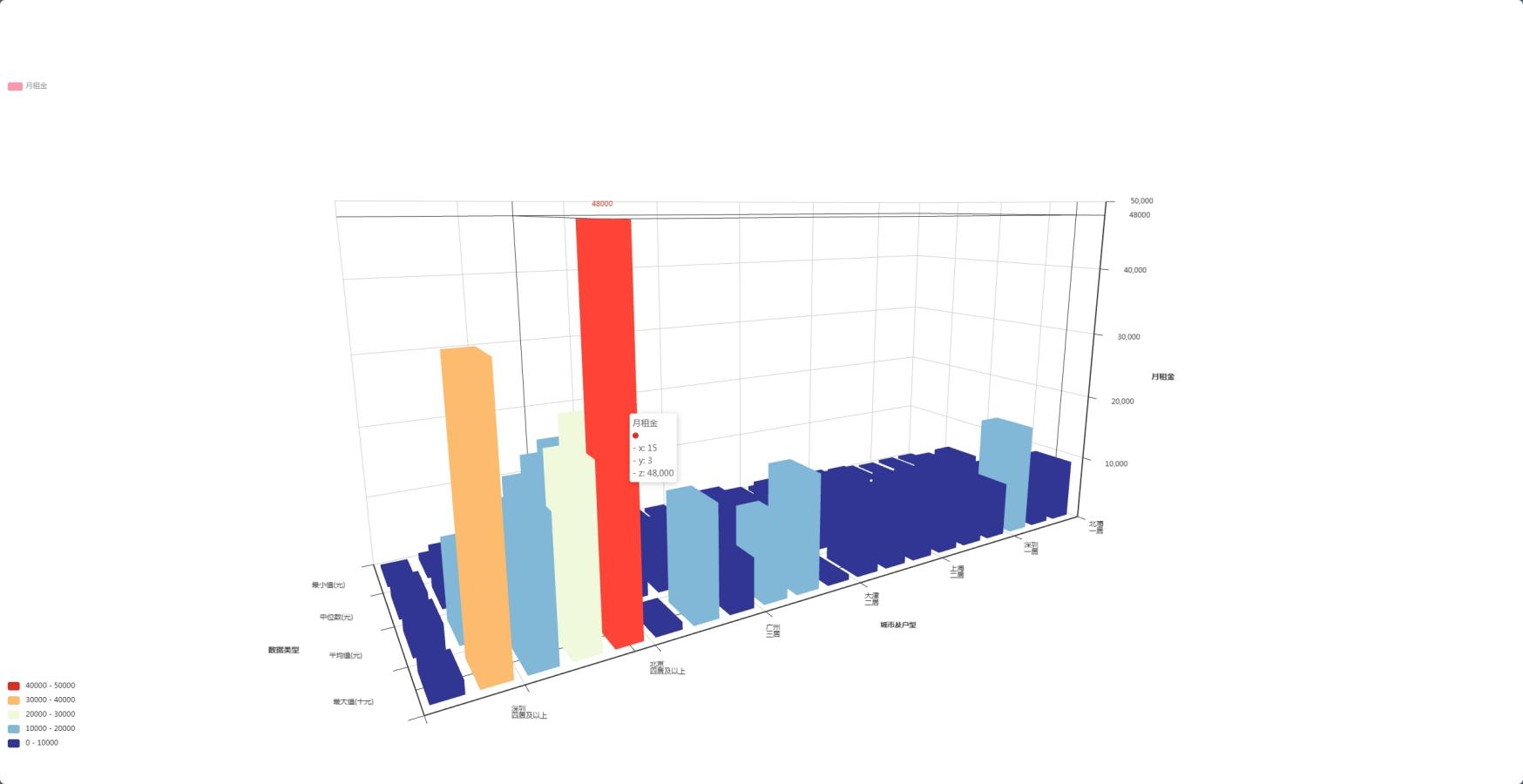


图7 五个城市房型的租金分布情况（俯视图）

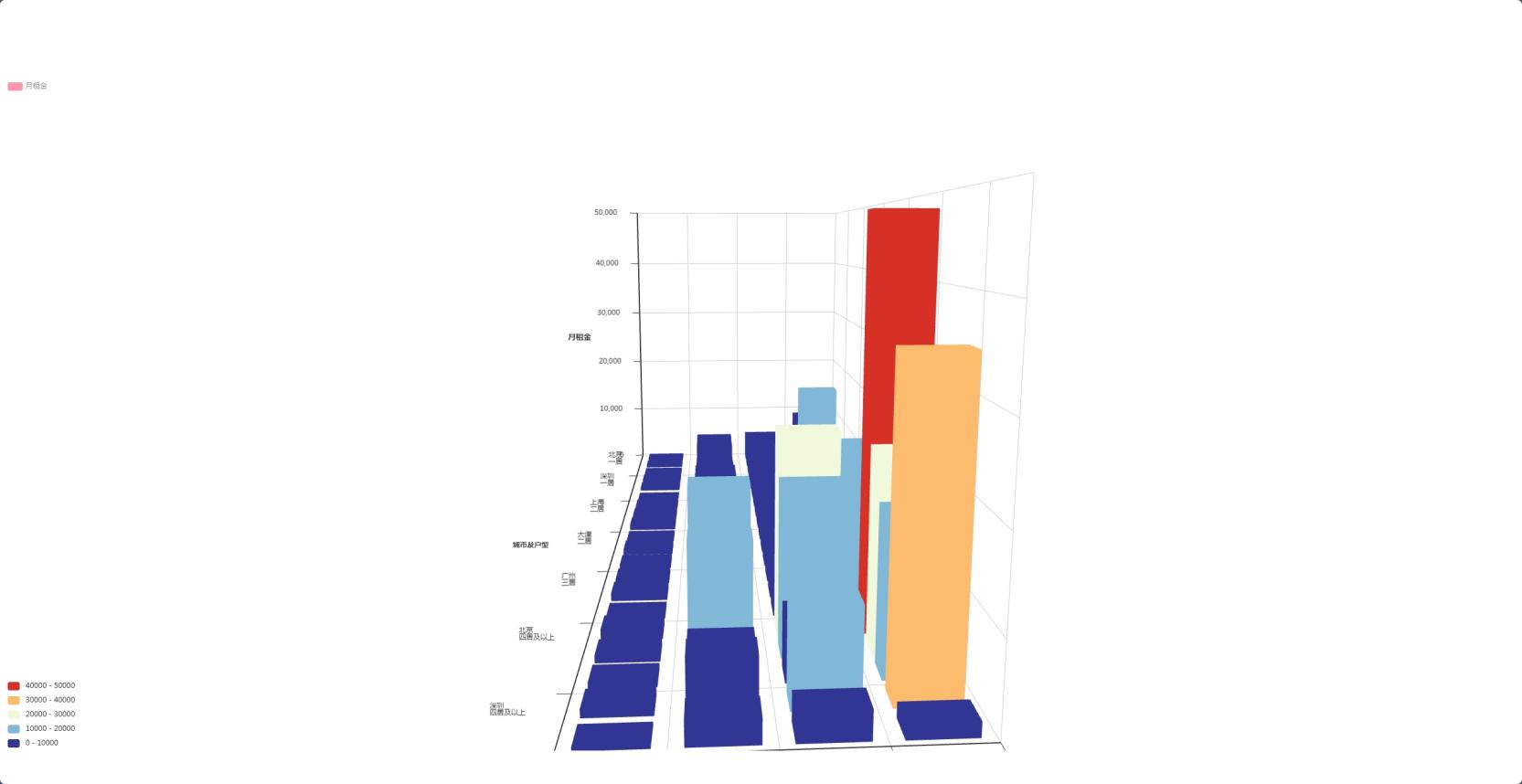


图8 五个城市房型的租金分布情况（侧视图）

**5.2.4 多城市户型数据情况分析**

三维柱状图展示了不同城市和房型的月租金分布，从图中可以看出，北京和上海的高端房型（如面积较大的四居室及以上）租金水平最高，其中北京的最高租金达到480,000元/月，其次是上海，租金呈现明显的层级差异；深圳的高端房型租金也较高，但略低于北京和上海。广州和天津则整体租金水平较低，尤其是在小户型（如一居室）方面租金更加亲民。

此外，不同城市的租金分布呈现出显著的区域性特征，北京、上海和深圳的租金跨度大，而广州和天津的租金跨度较小，显示出后者市场更均衡、更贴近大众需求。综合来看，北京、上海和深圳主导高端市场，租金因房型差异显著，广州和西安更具性价比，适合经济型住房需求。

针对单一户型，北京市、上海市的月租金都要高于其他城市。而一线城市广州和天津的月租金差异不大。针对单一城市，一居、二居、三居的月租金均呈现递增趋势。居室数量越多，平均租金也会越高。

**5.2.5 单一城市户型数据情况展示**

通过折线图进行城市内的比较，可以更直观地观察随着居室增加的变化情况，如图9、图10、图11、图12、图13。

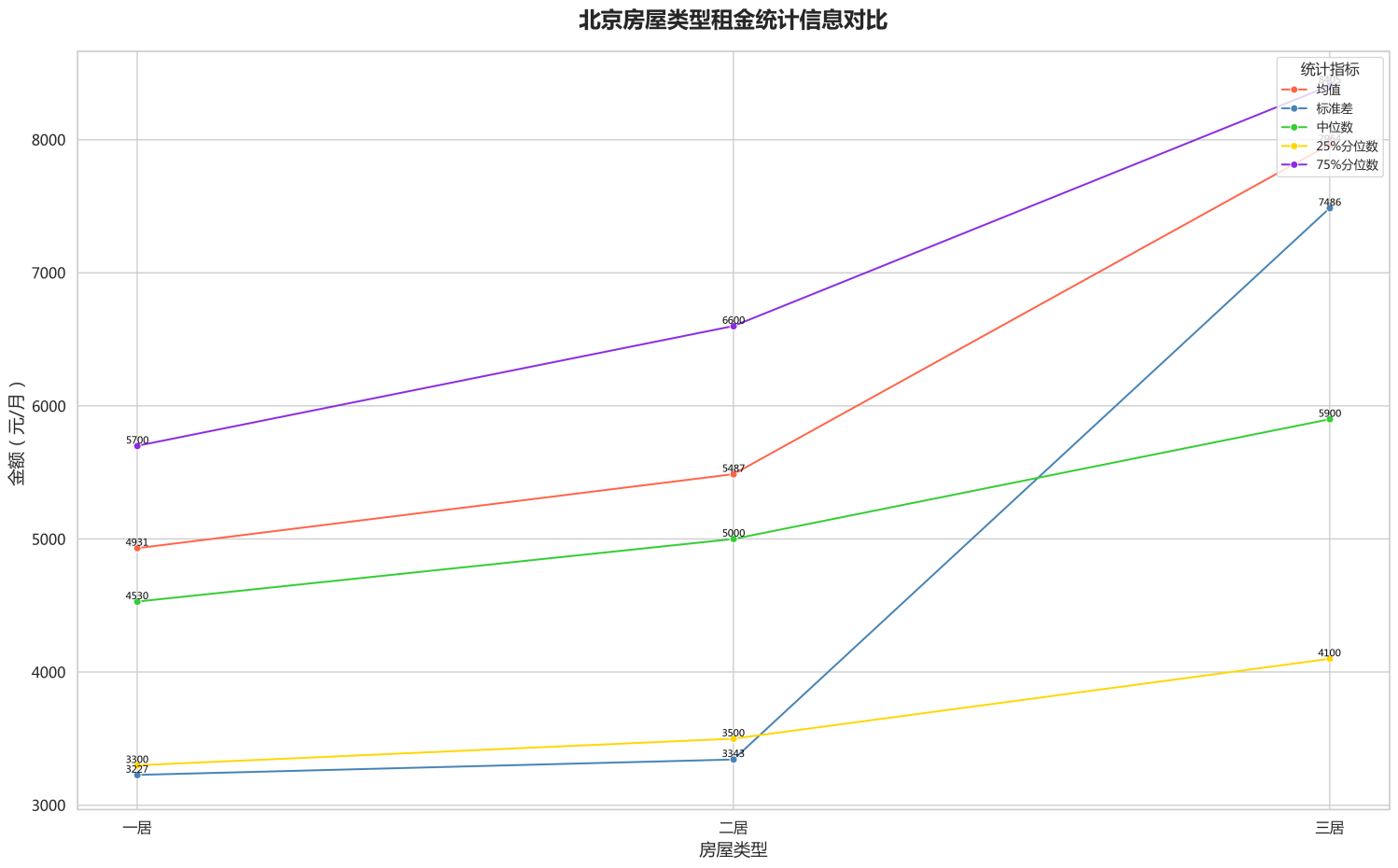


图9 北京房屋类型租金统计信息图

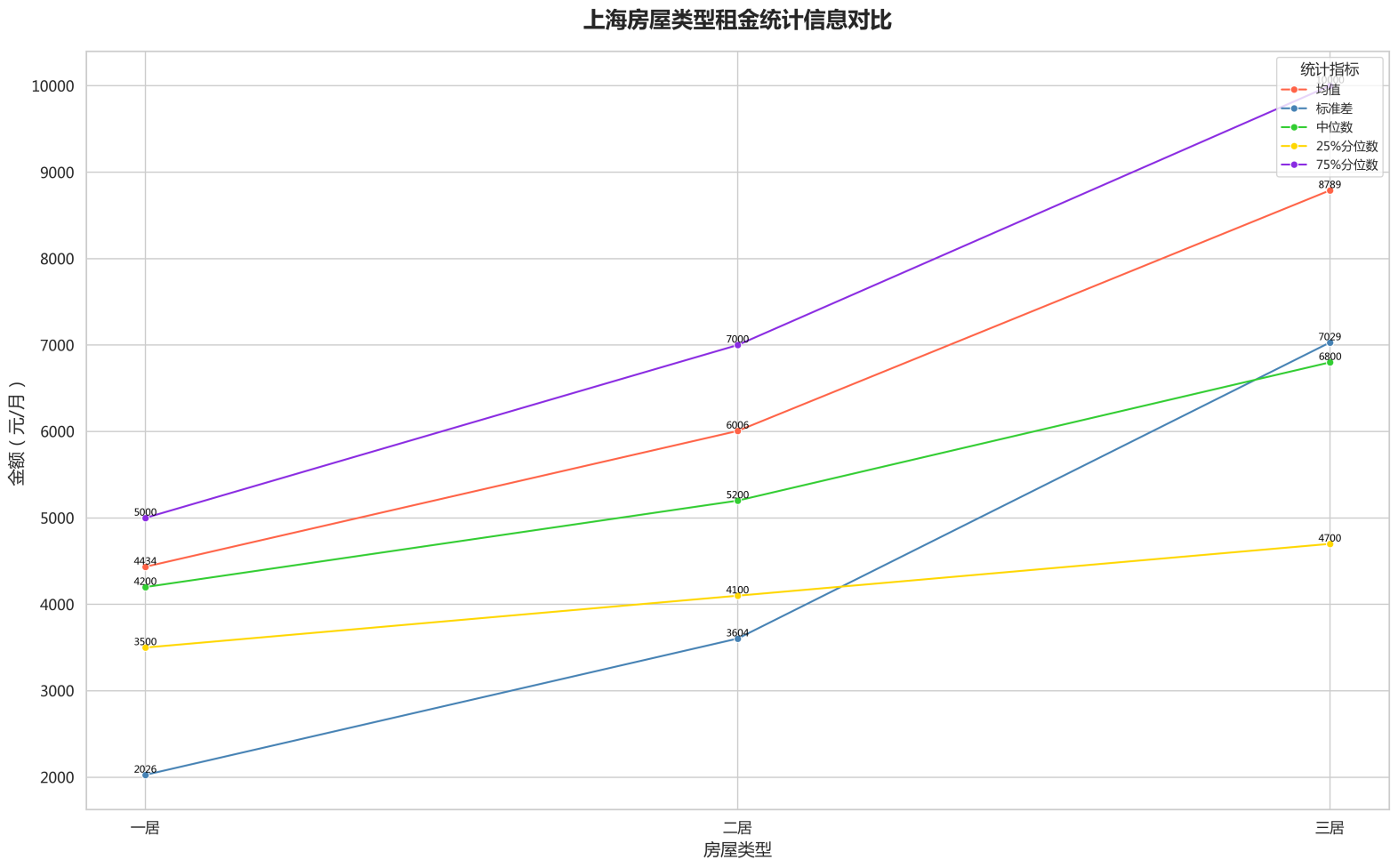


图10 上海房屋类型租金统计信息图

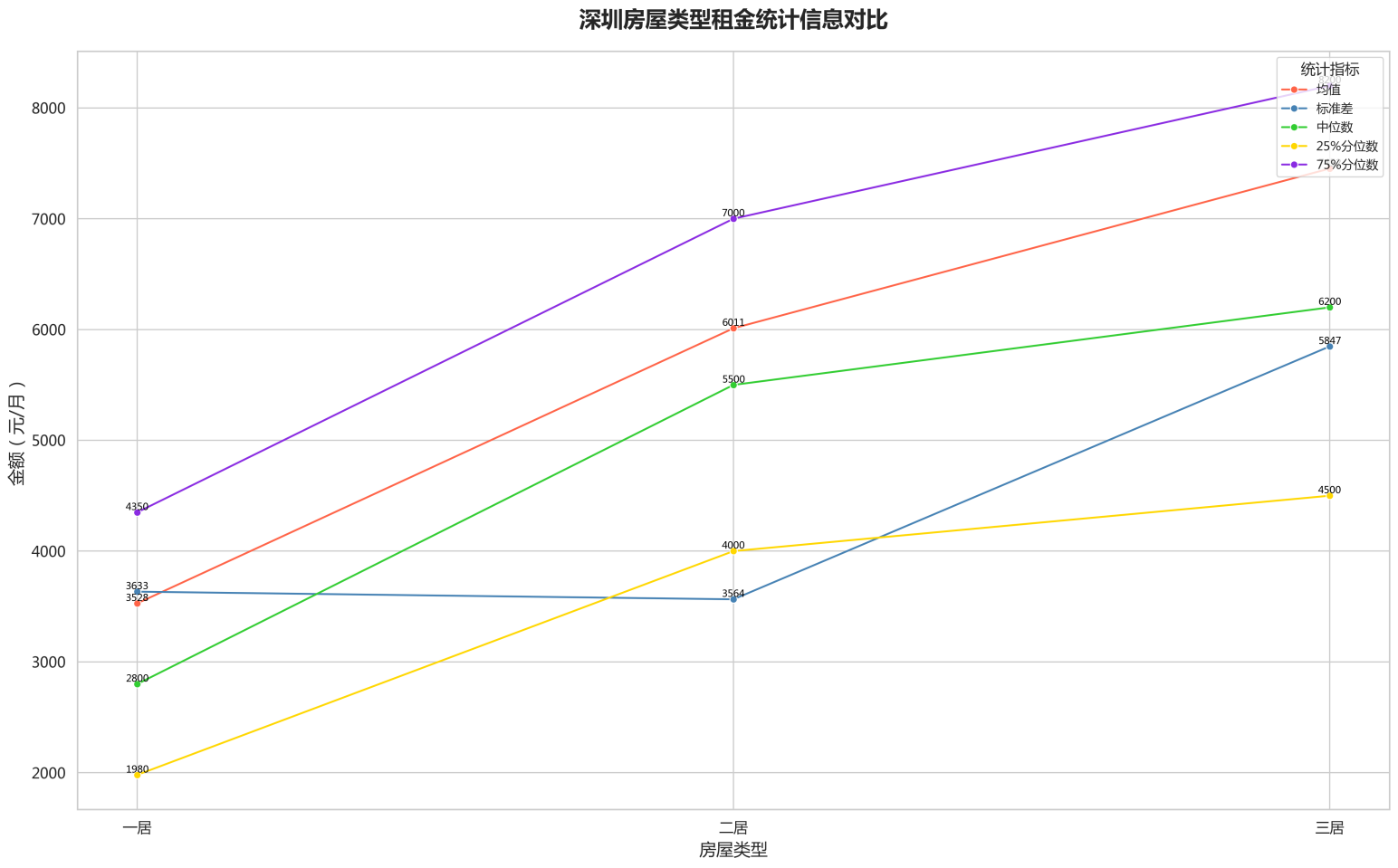


图11 深圳房屋类型租金统计信息图

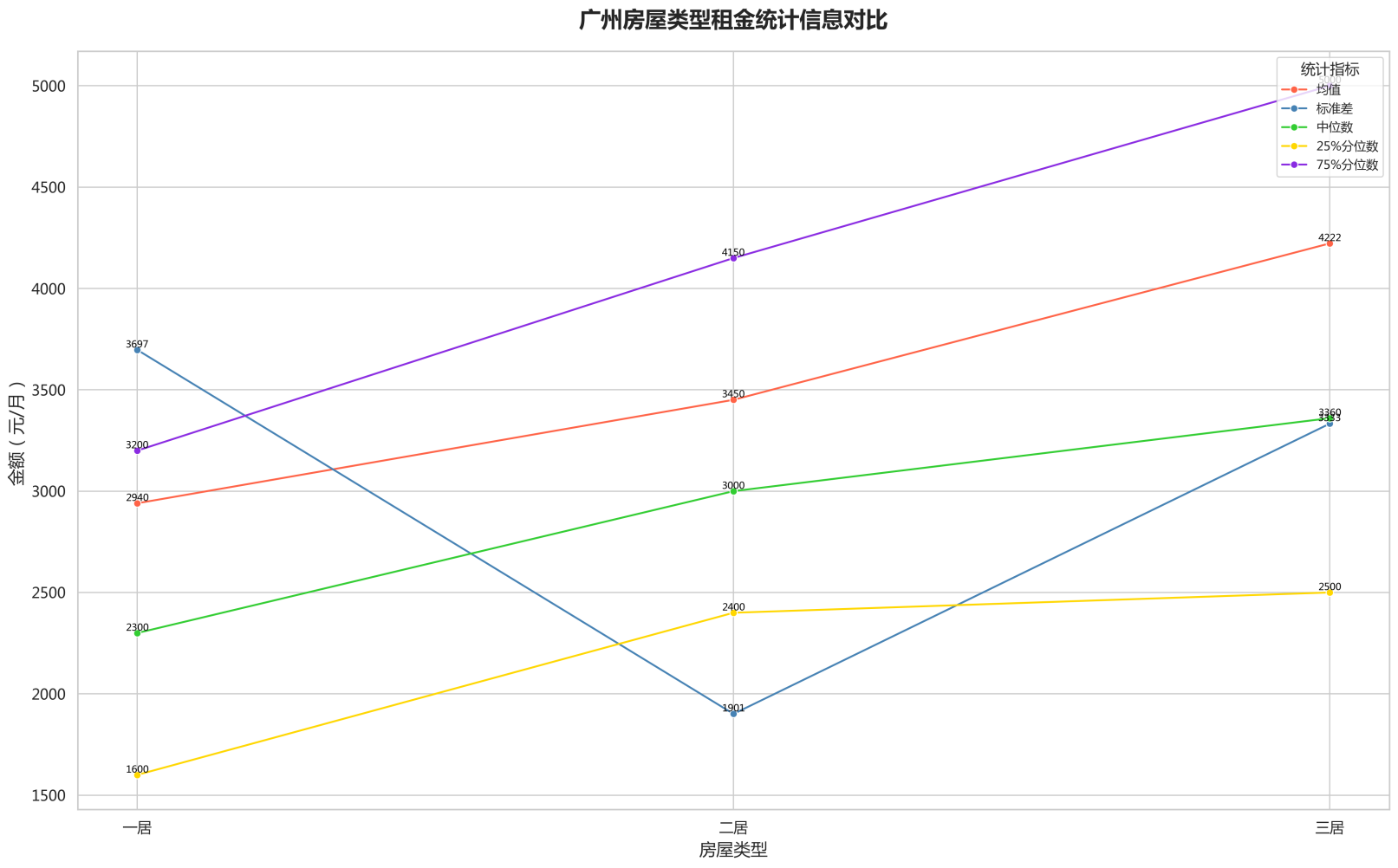


图12 广州房屋类型租金统计信息图

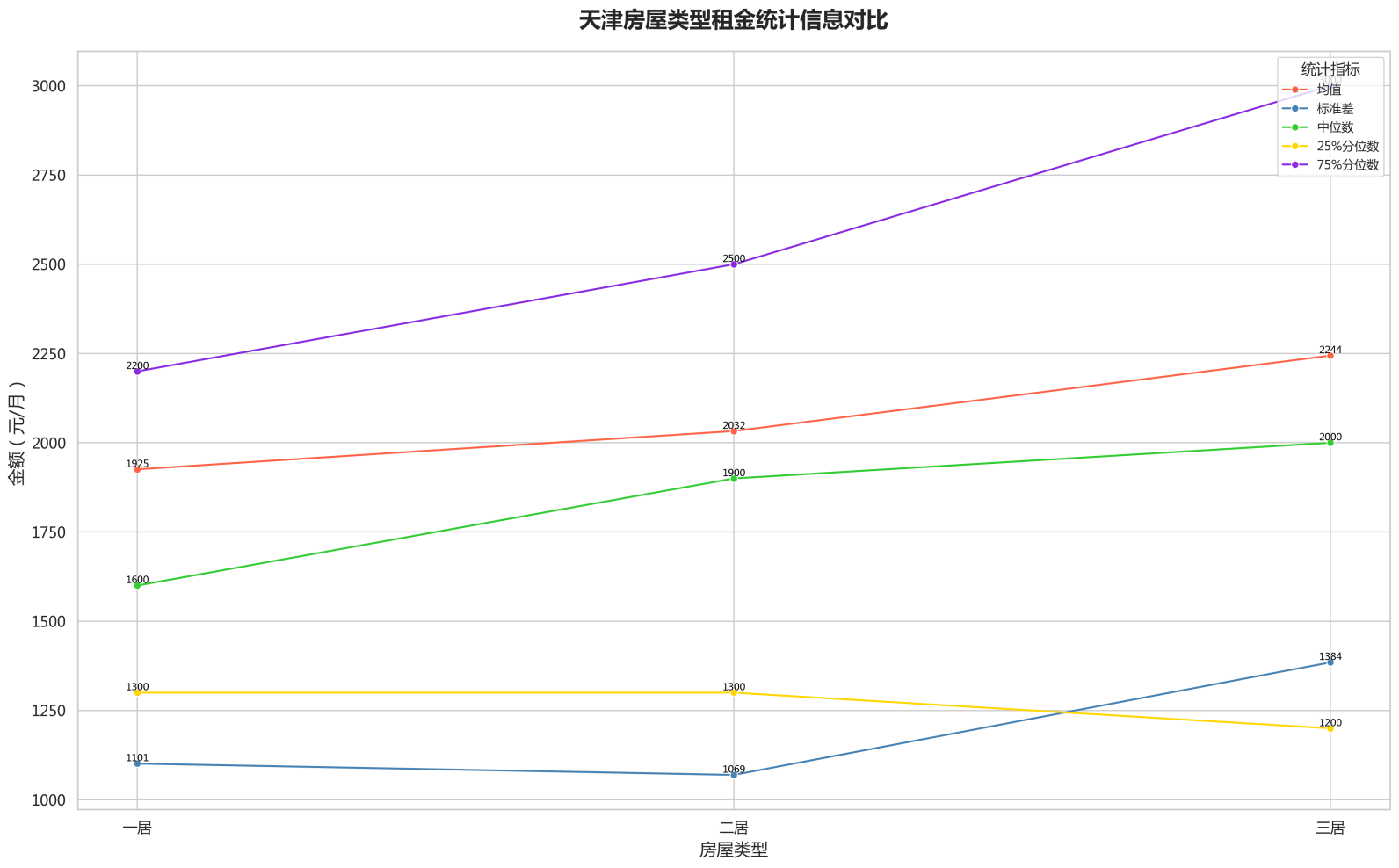


图13 天津房屋类型租金统计信息图

**5.2.6 单一城市户型数据情况分析**

这五张折线图分别展示了北京、上海、广州、深圳和天津在不同房型（如一居、二居、三居）的租金统计数据，包括均值、最大值、最小值、中位数及分位数变化。

北京和上海的租金水平较高且随房型增加显著上涨，三居室的高端房源拉升整体价格，尤其是75%分位数与标准差显示市场不均衡；深圳的租金次之，表现出三居室的高需求，同时波动性较大；广州的租金水平中等偏低，分布均衡，房型差距不显著，适合普通租户需求；天津的租金水平最低，且房型间涨幅最小，显示市场稳定性高且亲民，经济型租房优势明显。

北京、上海、深圳的大房型租金涨幅最大，广州和天津则更适合经济性需求的租户。

**5.2.7 城市板块租金数据情况展示**

本项目在绘图前对板块租金进行排序，利用直方图展示各板块的租金信息，以便进行对比和分析，如图14、图15、图16、图17和图18。

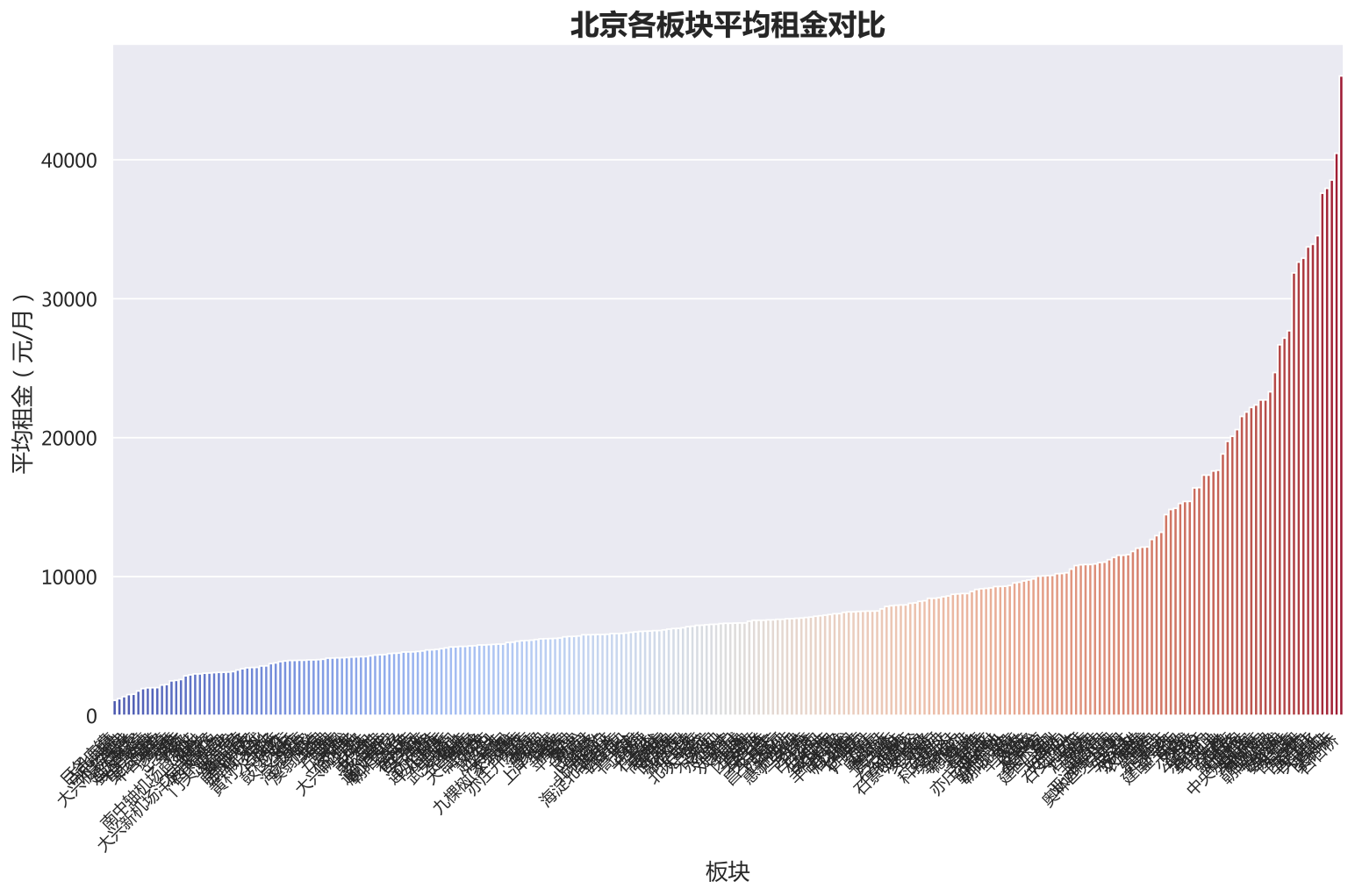
由于板块数量较多，直方图在展示大量板块的租金信息时显得不够直观，难以快速获取整体分布规律，因此改为采用地图上的区域租金热力图进行展示。热力图不仅能够更直观地反映各区域的租金水平，还能结合地理位置展示租金的空间分布特征，有助于分析租金的区域差异和热点。例如，通过图19、图20、图21、图22和图 23的热力图，观察不同区域的租金高低分布，不仅可以直观对比各区域的租金水平，还可以将租金信息与交通、商圈、城市功能区等地理要素关联分析，从而为租房者或市场决策者提供更加清晰、全面的参考。

图14 北京各板块平均租金对比图

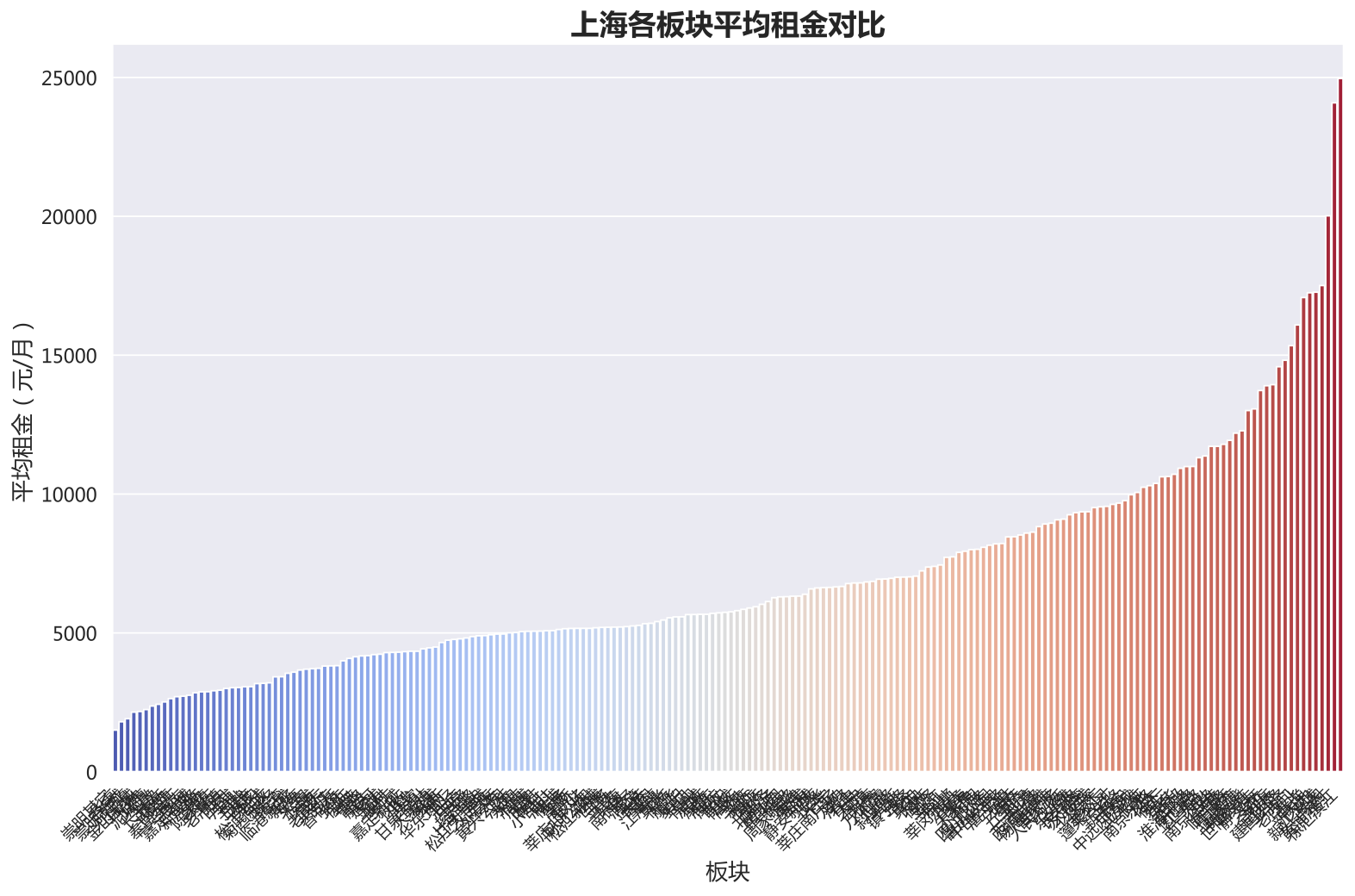


图15 上海各板块平均租金对比图

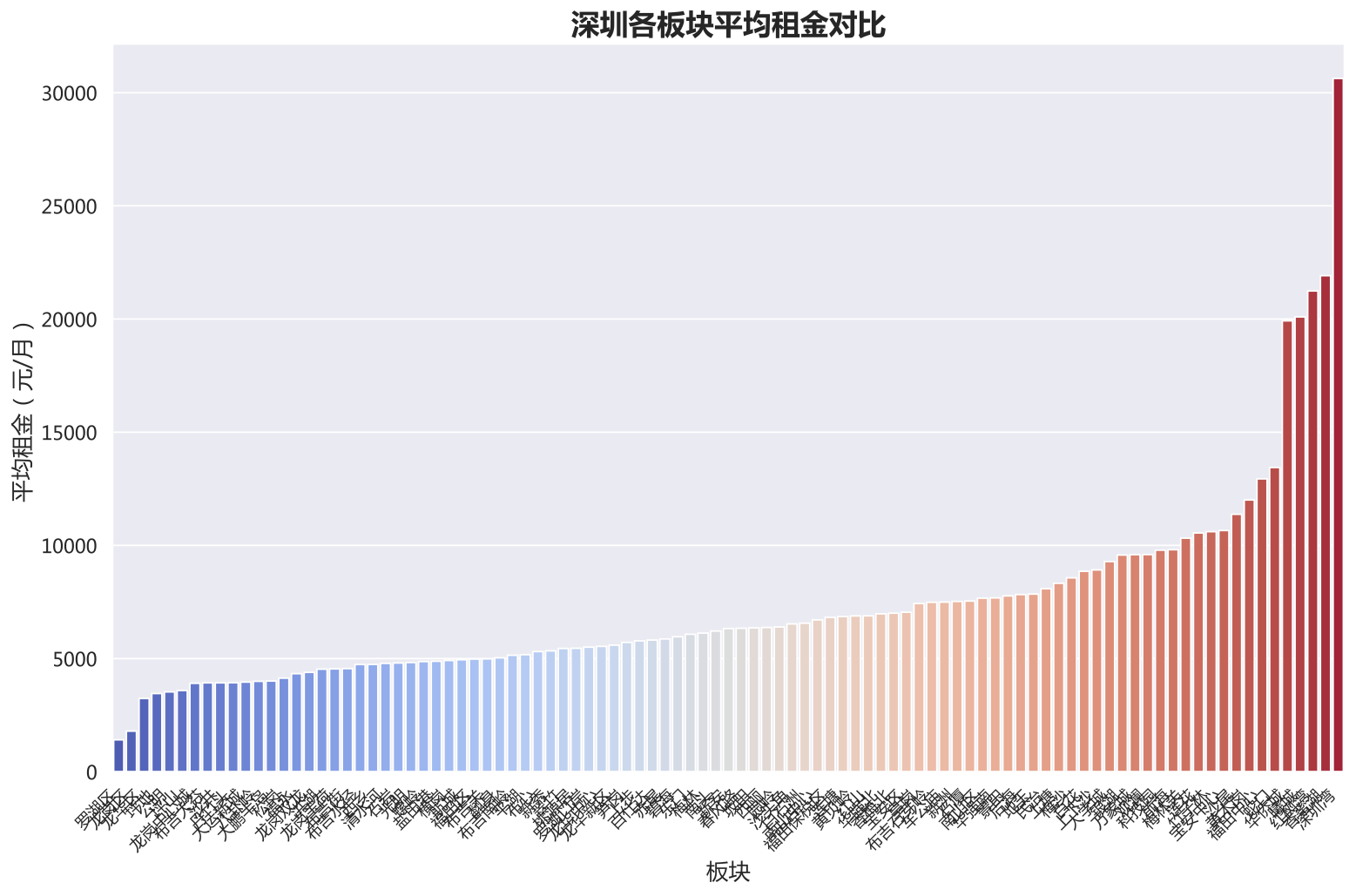


图16 深圳各板块平均租金对比图

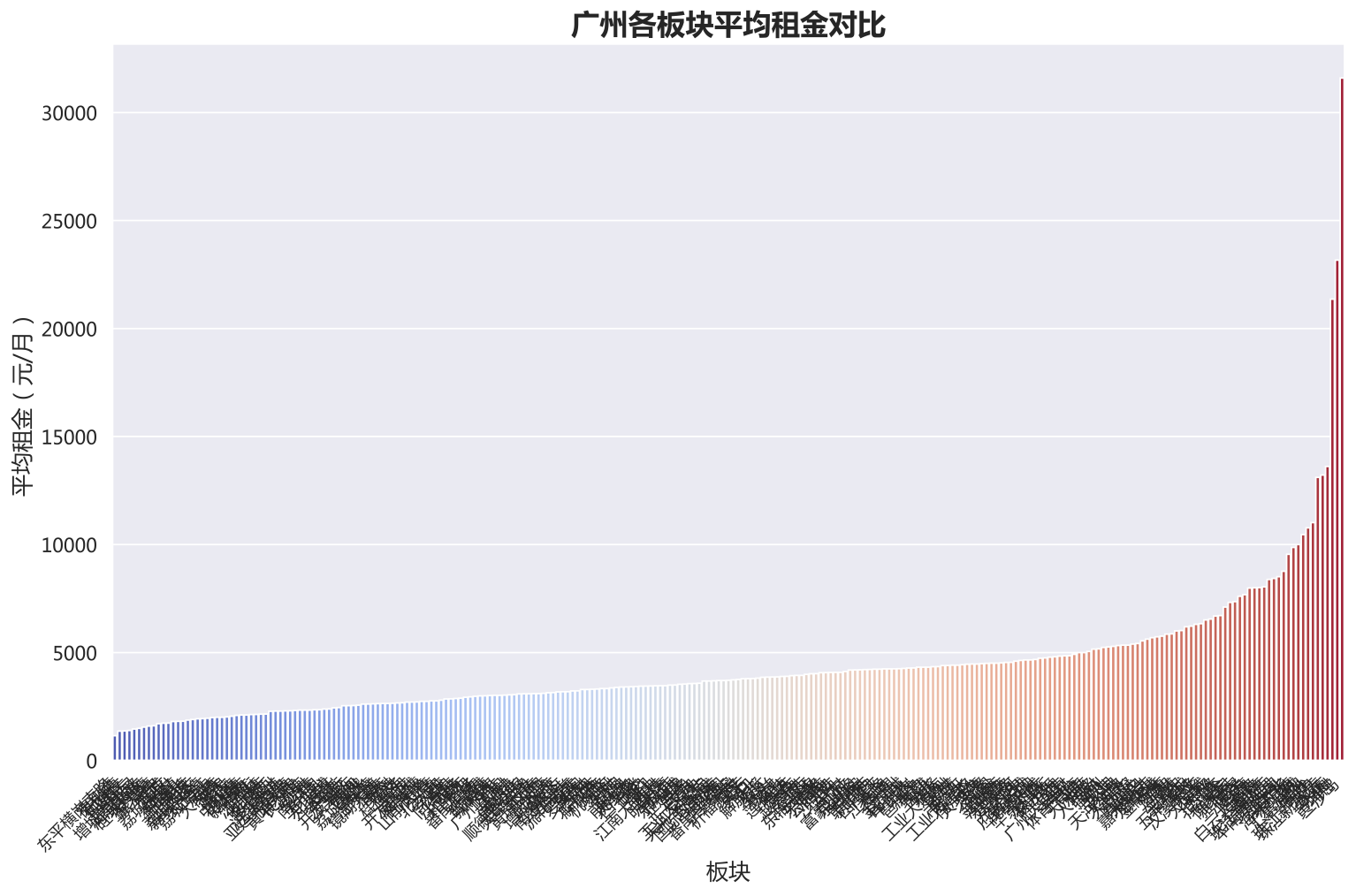


图17 广州各板块平均租金对比图

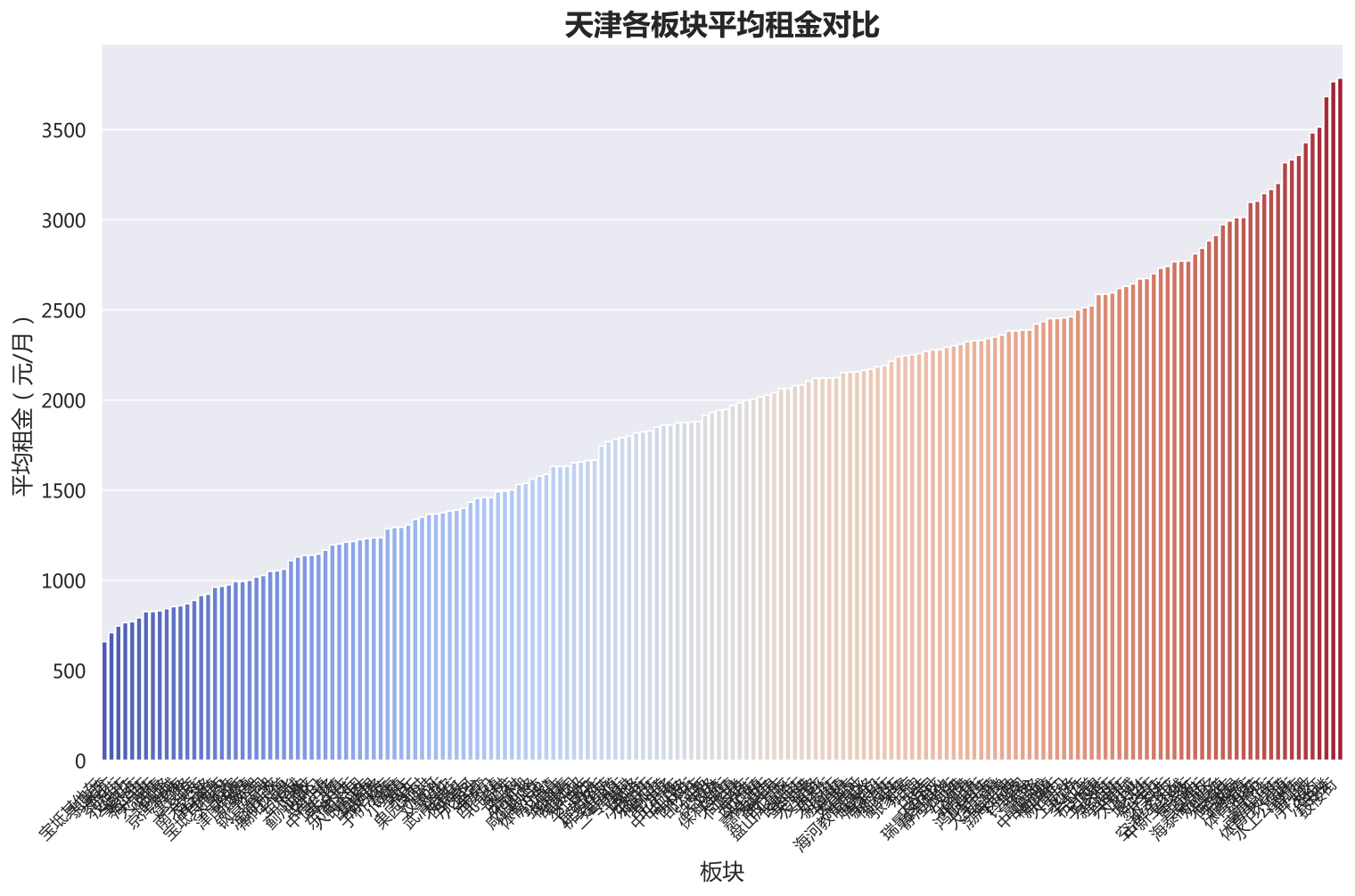


图18 天津各板块平均租金对比图

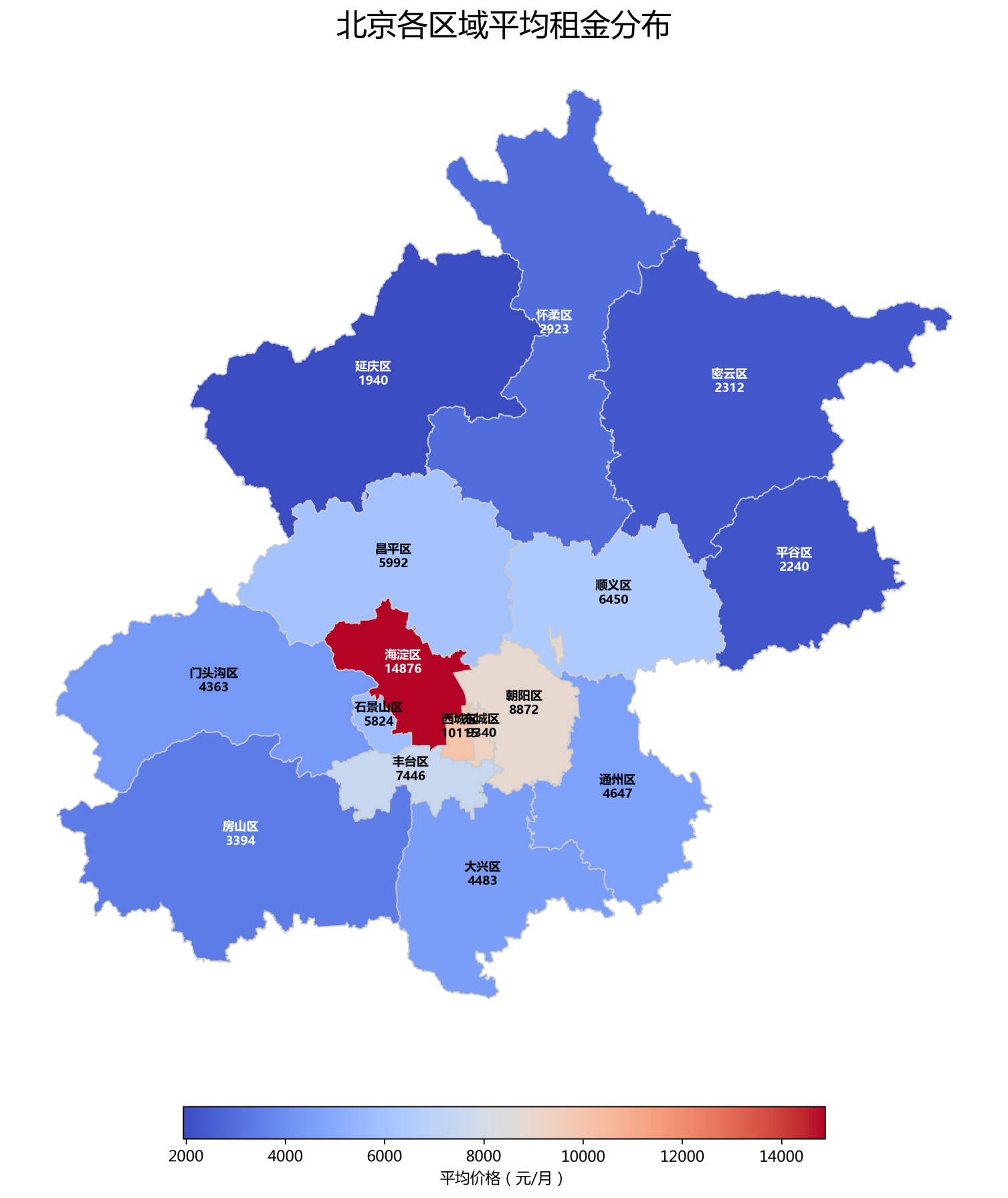


图19 北京各区域平均租金对比图

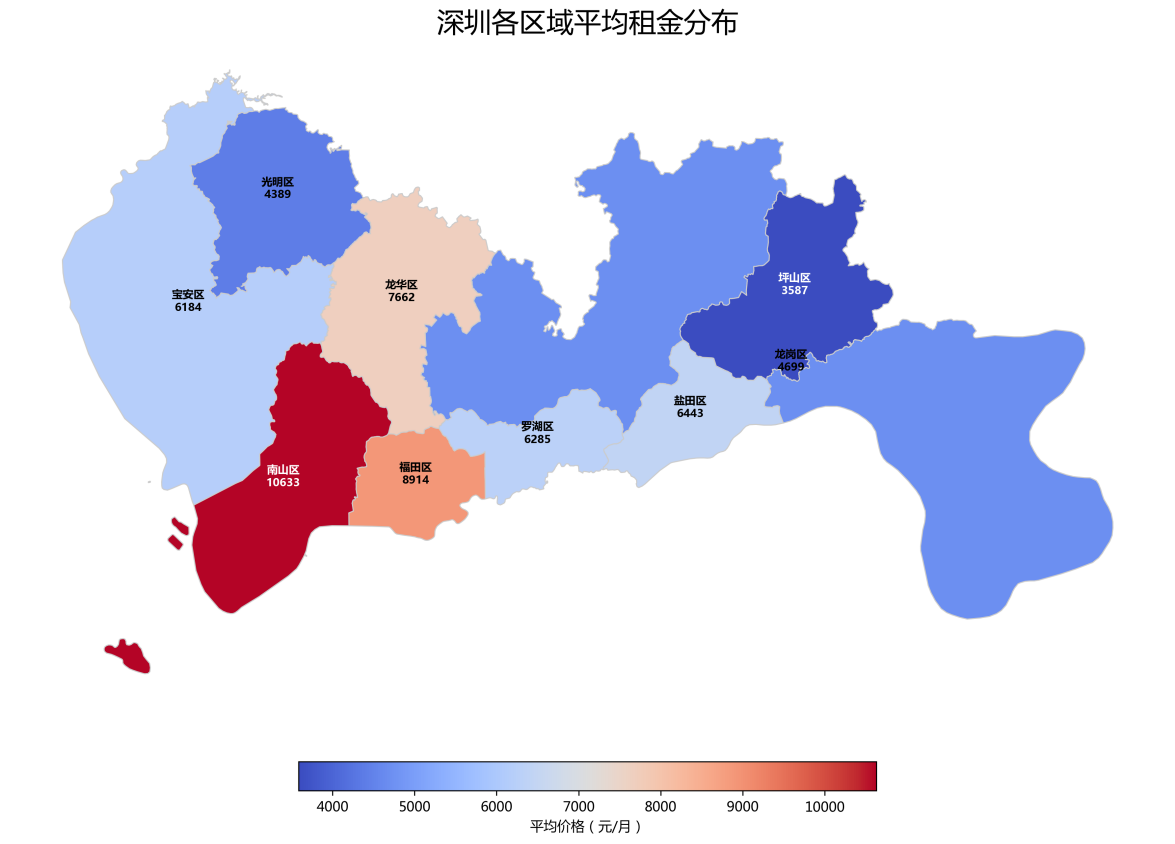
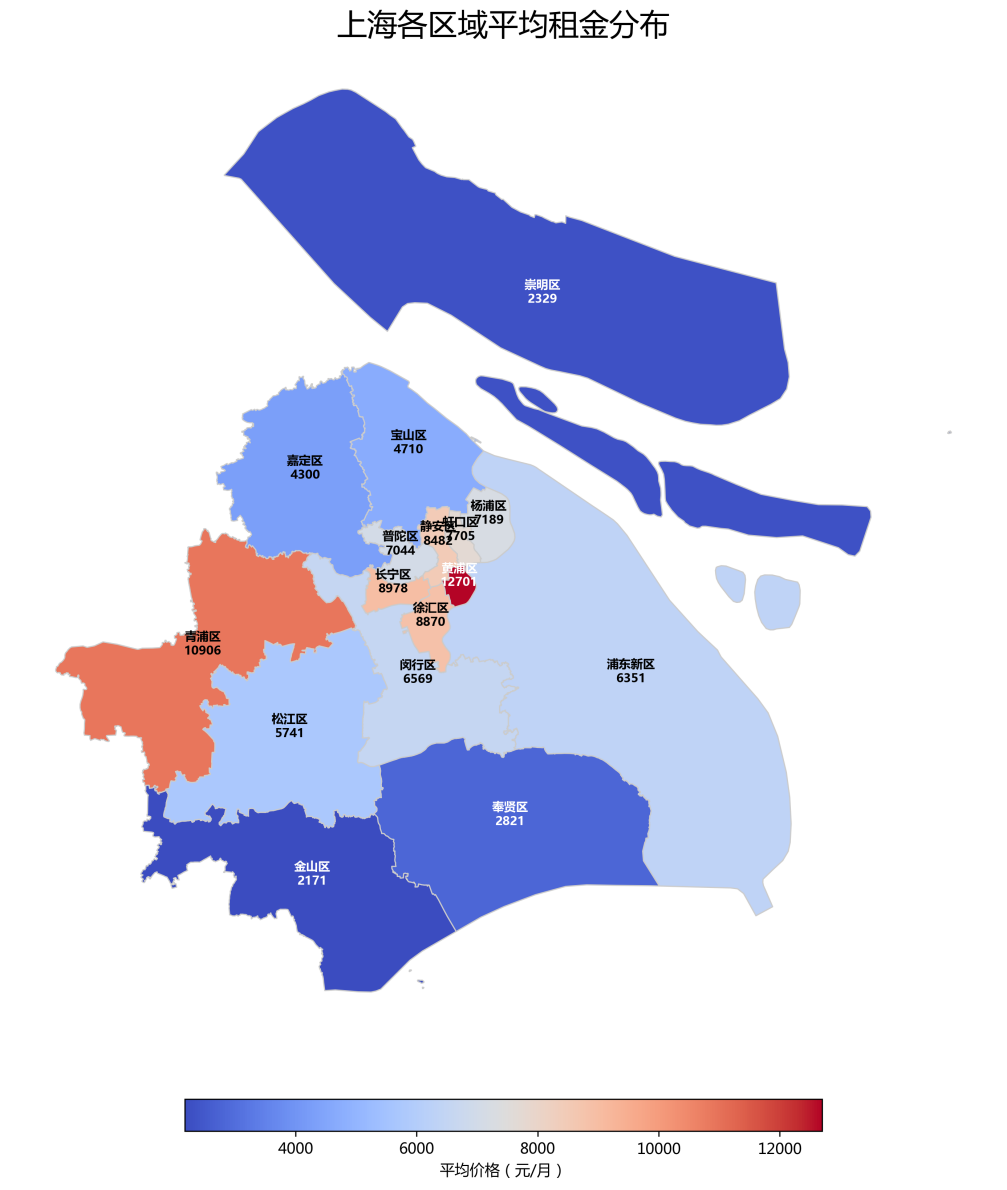


图20 上海各区域平均租金对比图

图21 深圳各区域平均租金对比图

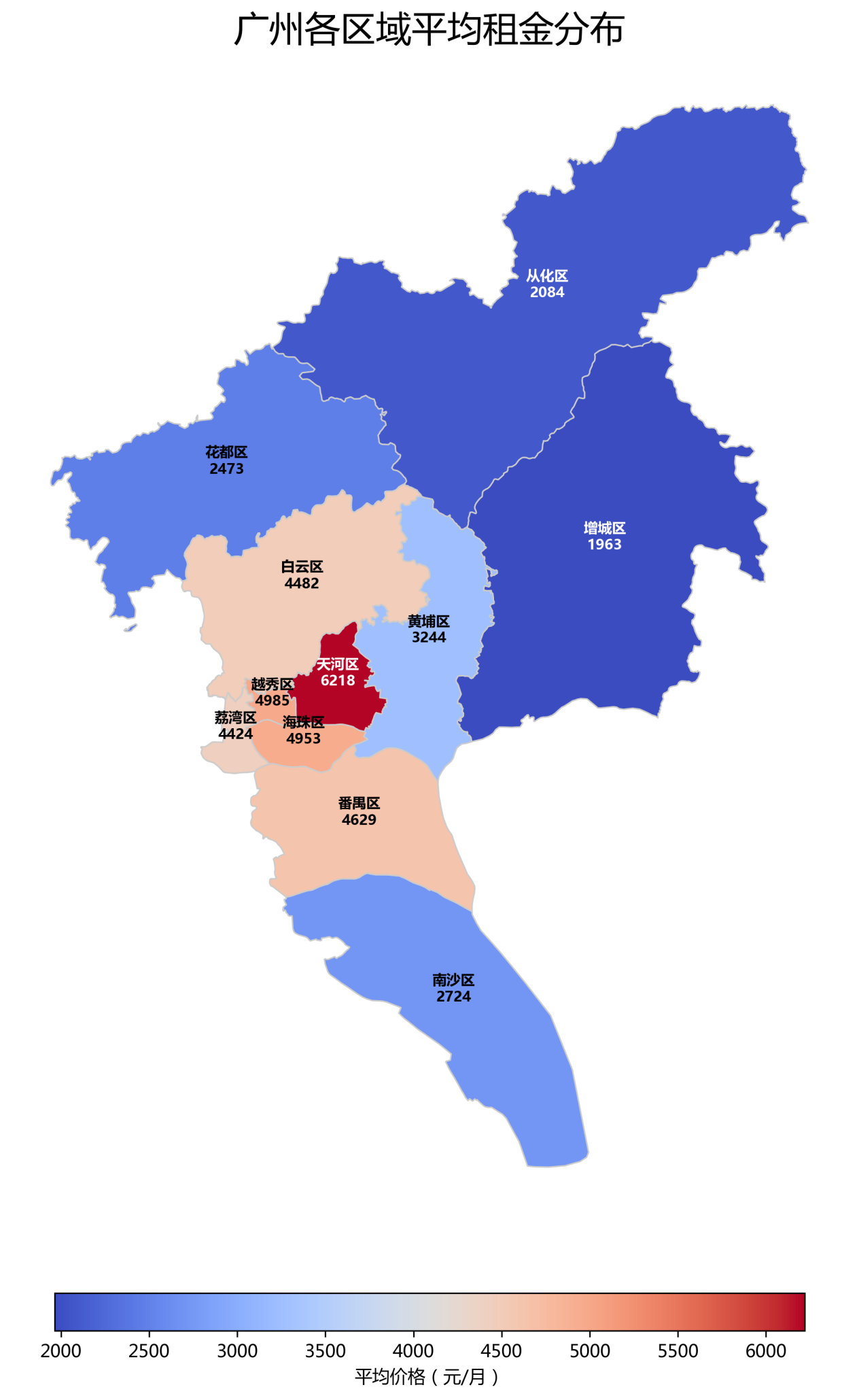


图22 广州各区域平均租金对比图

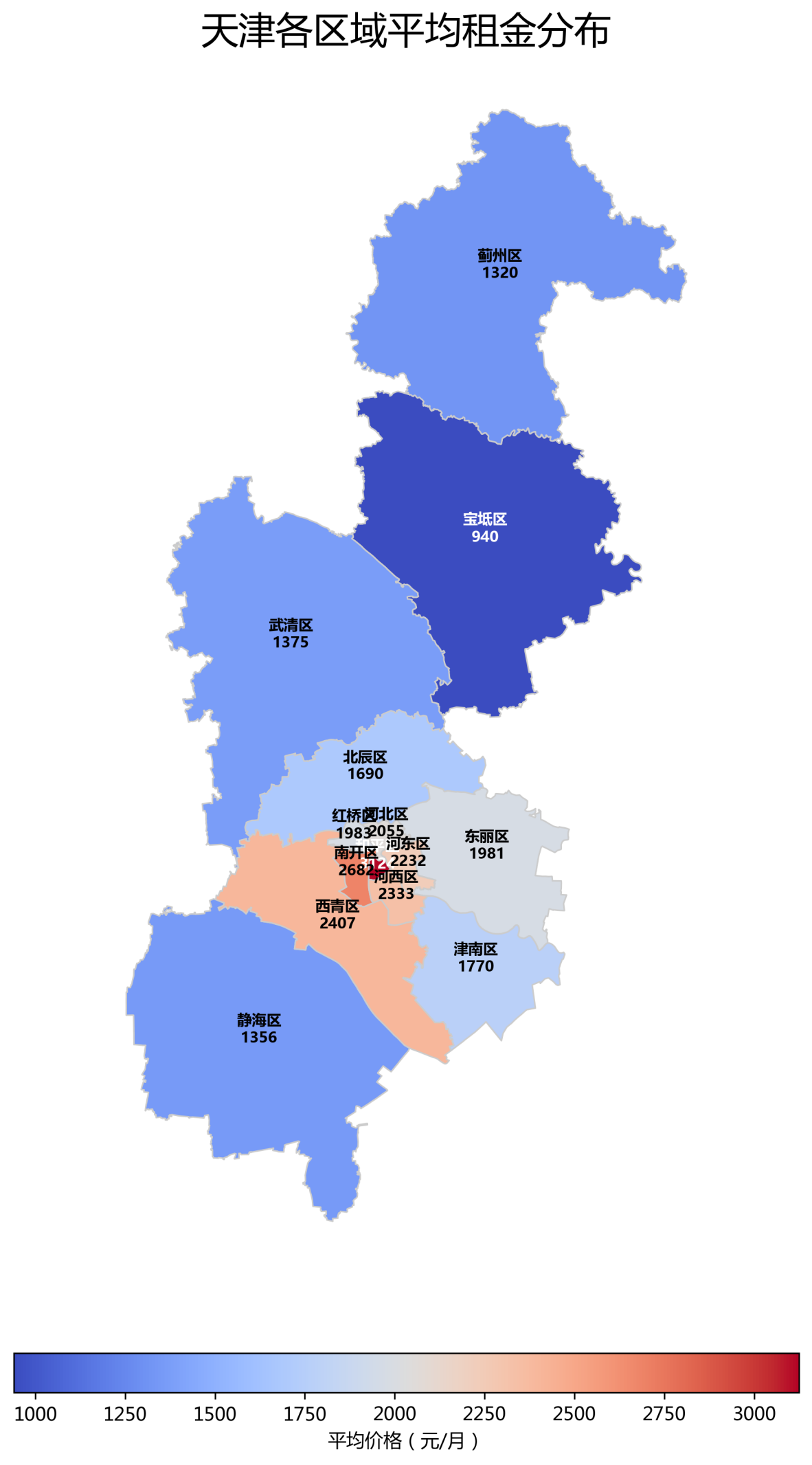


图23 天津各区域平均租金对比图

**5.2.8 城市板块租金数据情况分析**

从这五张图可以看出，北京、上海、广州、深圳和天津的区域租金分布都体现了一个普遍规律：越靠近城市中心、资源丰富、经济发达、交通便利的地段，租金水平就越高。

这种分布特点由地理位置、配套设施、经济发展水平、房屋质量、供需关系以及区域规划等多种因素综合决定。例如，北京的西城区拥有优质的教育资源（如众多重点学校）、便利的地理位置和浓厚的历史文化底蕴，使其租金水平处于全市的顶端；上海的黄浦区作为中心城区，集繁华的商业区、顶级配套设施和深厚的历史文化底蕴于一体，是经济、文化、交通的核心区域，租金自然高企；广州的天河区作为广州的核心商务区，聚集了高端商业中心、优质教育资源和便捷交通枢纽，吸引了大量高收入人群和企业，租金水平领先；深圳的南山区则是科技创新中心，汇聚了大量高新技术企业、高收入人才和优质教育资源，推动区域租金水平居高不下；天津的教育强区比如南开区、西青区等地，租金较高，外围区域如蓟州区、静海区则租金很低。

此外，高租金区域还可能是别墅区，例如北京的中央别墅区等，这些区域由于高端房屋集聚和特殊区域规划，也带动了整体租金水平的提升。整体来看，这种区域租金差异反映了城市功能分区和经济发展的不均衡性，同时也体现了不同区域吸引力和人群定位的差异。

**5.2.9 朝向租金数据情况展示**

为了更直观地观察价格分布，本项目绘制了各城市不同朝向的价格概率分布曲线。为了避免极端大值对图像的影响，筛除了超出平均值加上3倍标准差的数据，使得分布图更加清晰明了，如图24、图25、图26、图27和图28所示。而为了更精确地分析各朝向的数值分布特征，则采用箱线图进行展示，如图30所示。这样的多种可视化方式结合，不仅提高了数据分析的清晰度，还为朝向与价格关系的深入研究提供了全面的参考。

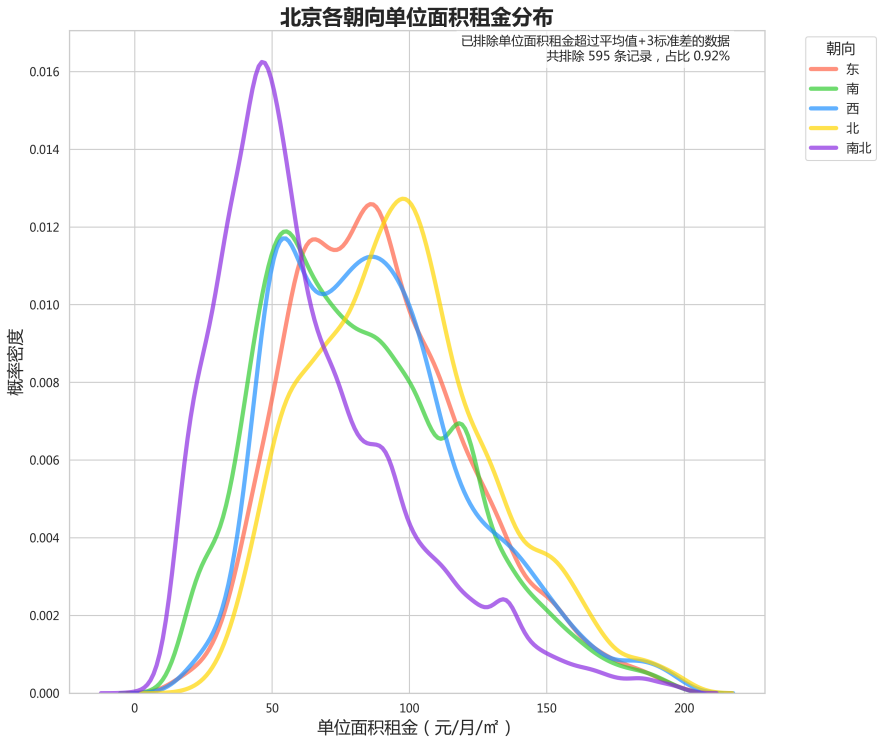


图24 北京各朝向单位面积租金分布图

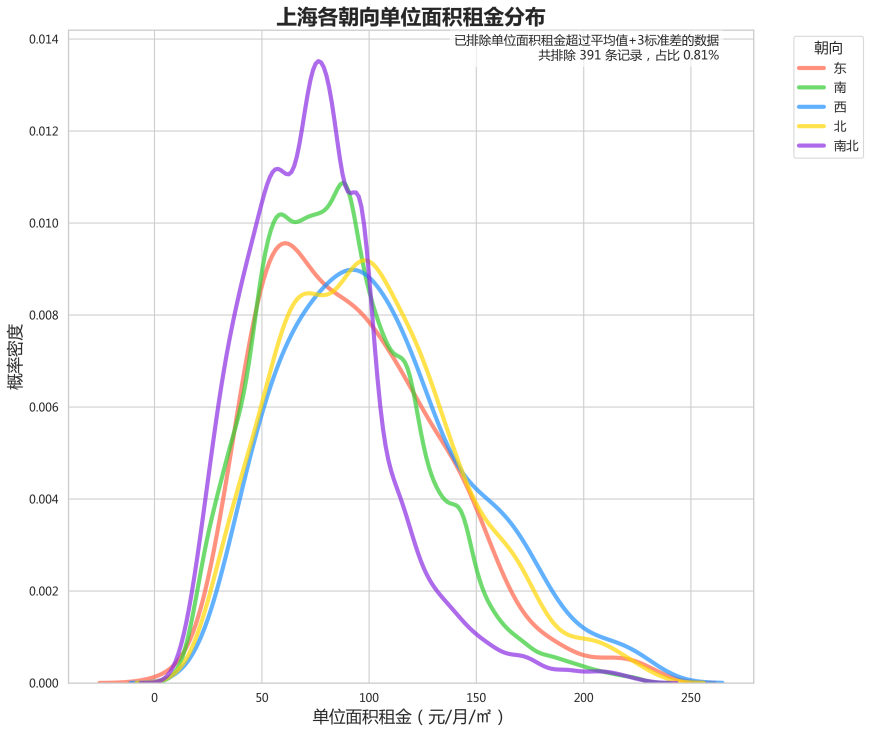


图25 上海各朝向单位面积租金分布图



图26 深圳各朝向单位面积租金分布图

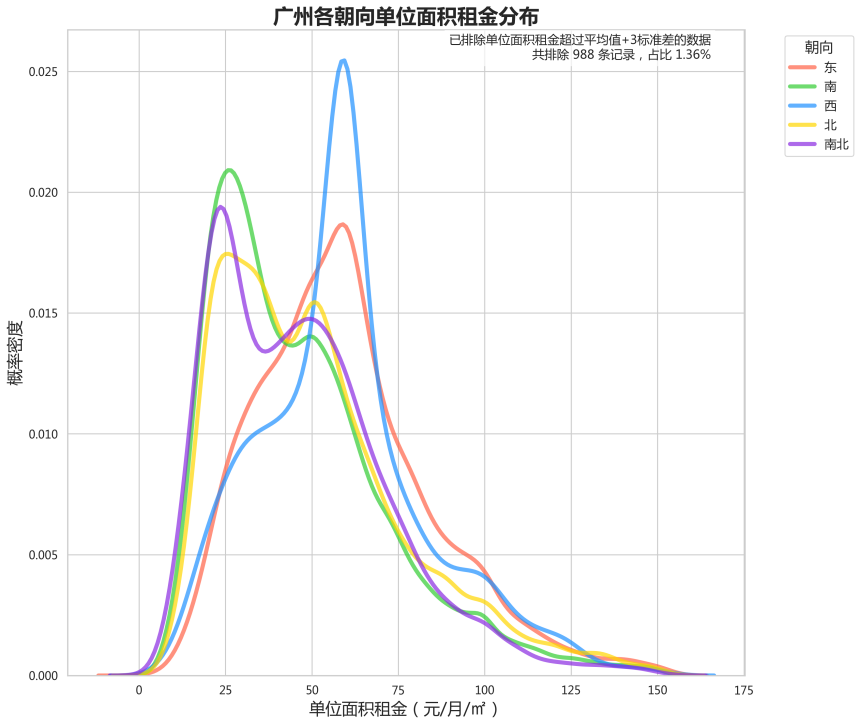


图27 广州各朝向单位面积租金分布图

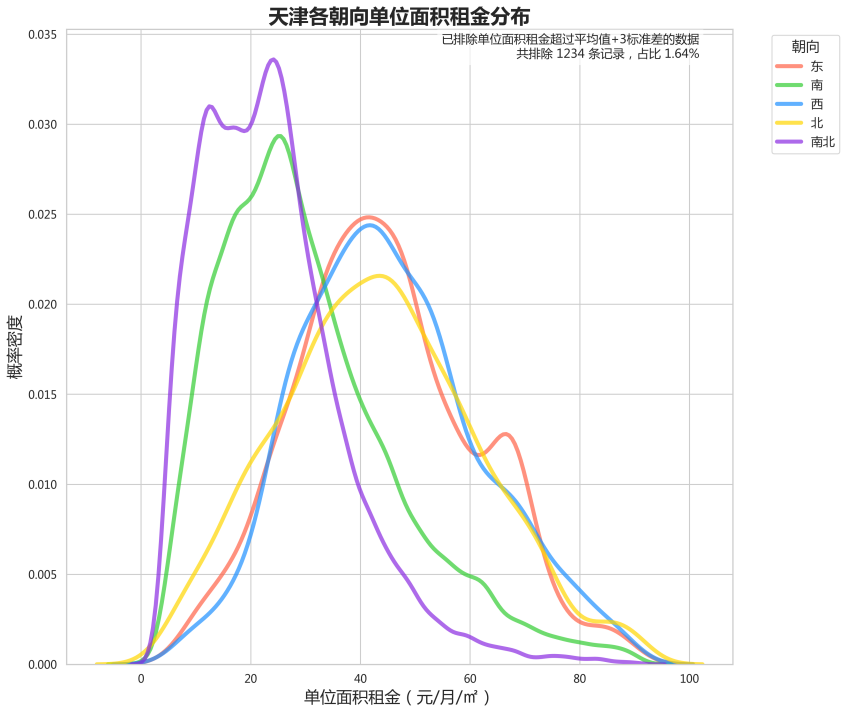


图28 天津各朝向单位面积租金分布图

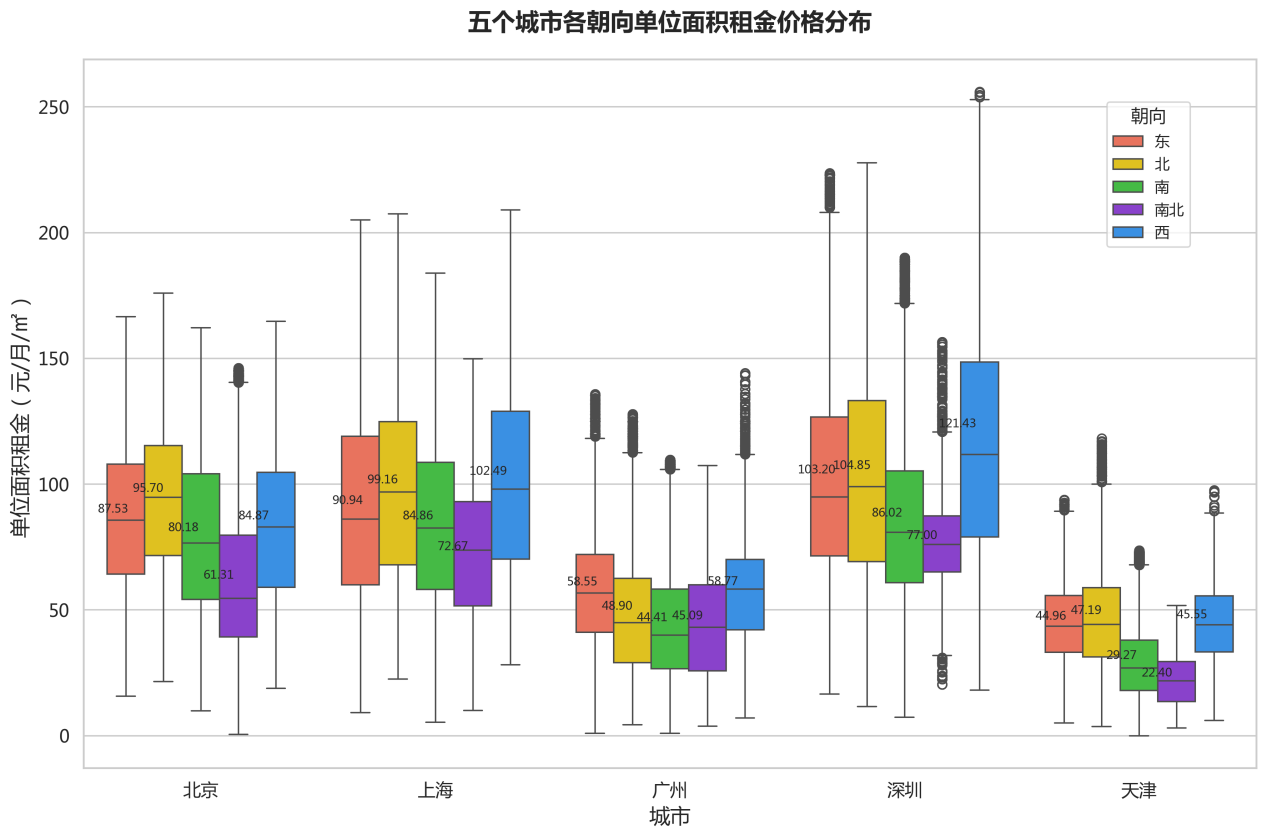


图29 五个城市各朝向单位面积租金价格分布箱线图

**5.2.10 朝向租金数据情况分析**

从上图可以看出，北京和天津的朝北租房单价最高，而上海、广州、深圳的朝西租房单价最高，传统观念中南北朝向的好房反而在租金上并未占据优势，通常为最低。这种现象可以从多个方面进行解释。  
 首先，北京和天津地处北方，冬季寒冷，当前正值冬季，人们对房屋的保暖性能和采光需求更高，而朝北的房屋可能在设计上特别注重避风和保温性能，并且实际光照条件可能优于预期，因此租金较高；而南北朝向的房屋通风较强，在寒冷季节可能更不符合租户的舒适性需求。

其次，上海、广州和深圳位于气候温暖潮湿的南方，夏季高温潮湿，人们偏好避免阳光直射的房屋，朝西的房屋在傍晚采光较好，同时避免了夏季正午烈日的直射，因而更受欢迎；南北朝向房屋由于通风过强或夏季过热可能不符合租户需求，导致租金较低。

此外，这种租金分布还可能与城市规划和建筑布局有关系，南北朝向的房屋由于受欢迎可能已被优先租赁，剩余房源为市场上的次优选择，从而使南北朝向的平均租金水平降低，这可能是一种“幸存者偏差”。同时，市场供需关系、建筑楼层、房屋高度等未被纳入统计的因素也会对租金分布产生一定影响。

因此，上述现象既与区域气候特点密切相关，也与市场偏好和统计特点等多因素综合作用密不可分。

**5.2.11 各城市人均GDP与租金数据情况展示**

根据国家统计局发布的2023年人均GDP数据，以及各城市的单位面积租金水平，我们计算了五个城市的租房性价比指数（CPI），定义为人均GDP除以单位面积租金（见四:4.3一节），CPI值越高表示性价比越高。

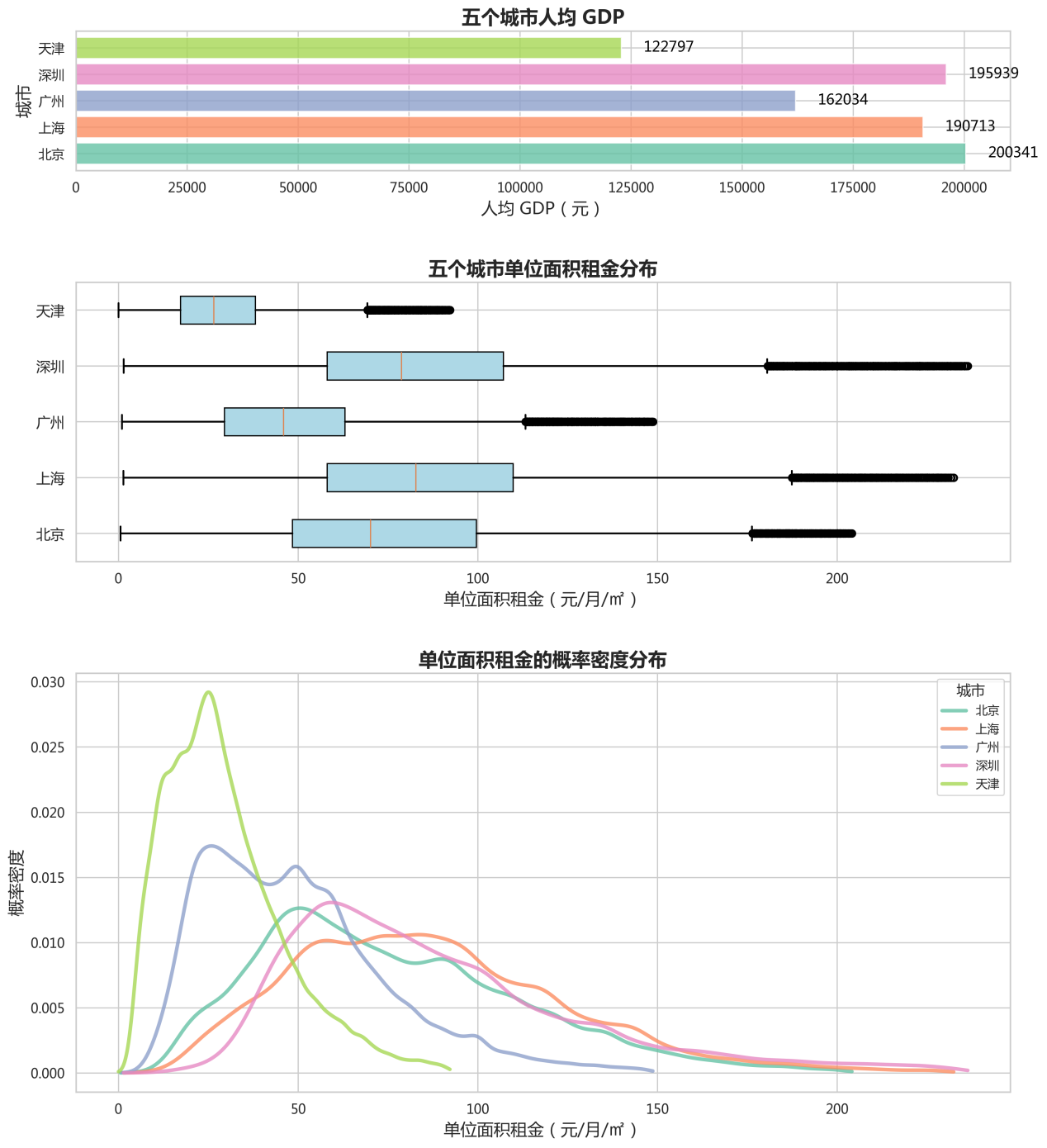


图30 五个城市单位面积租金分布概况

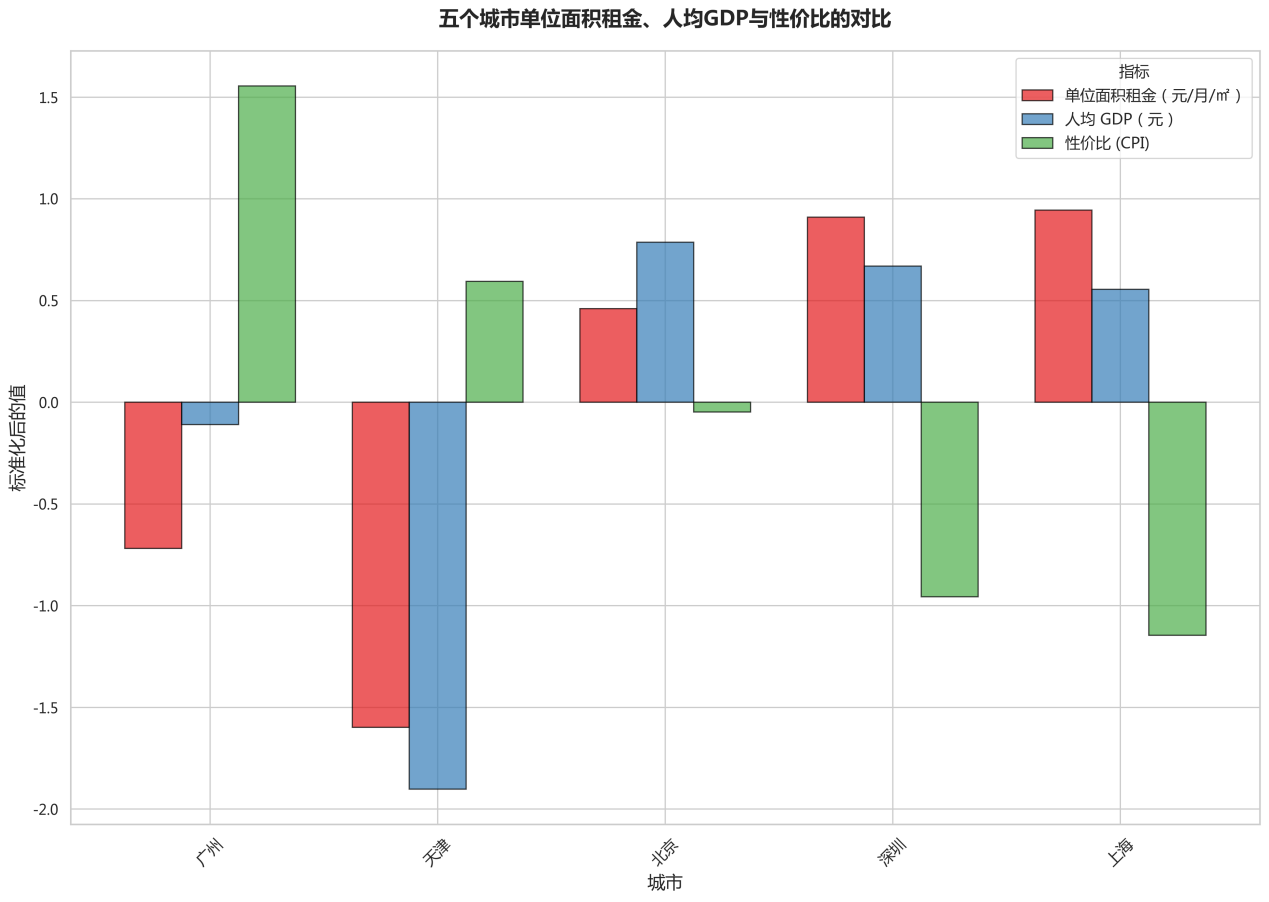


图31 五个城市单位面积租金、人均GDP与性价比对比

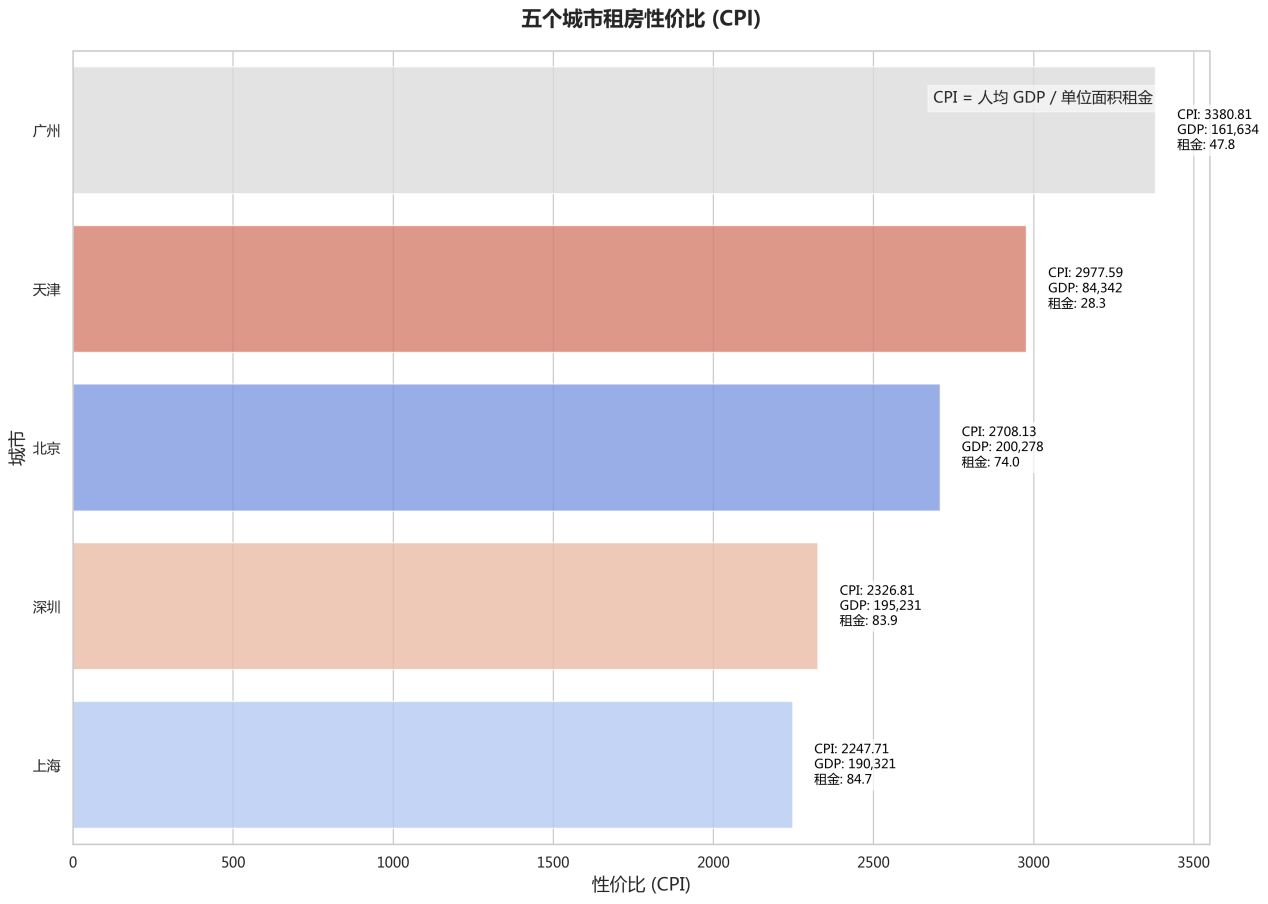


图32 五个城市租房性价比对比

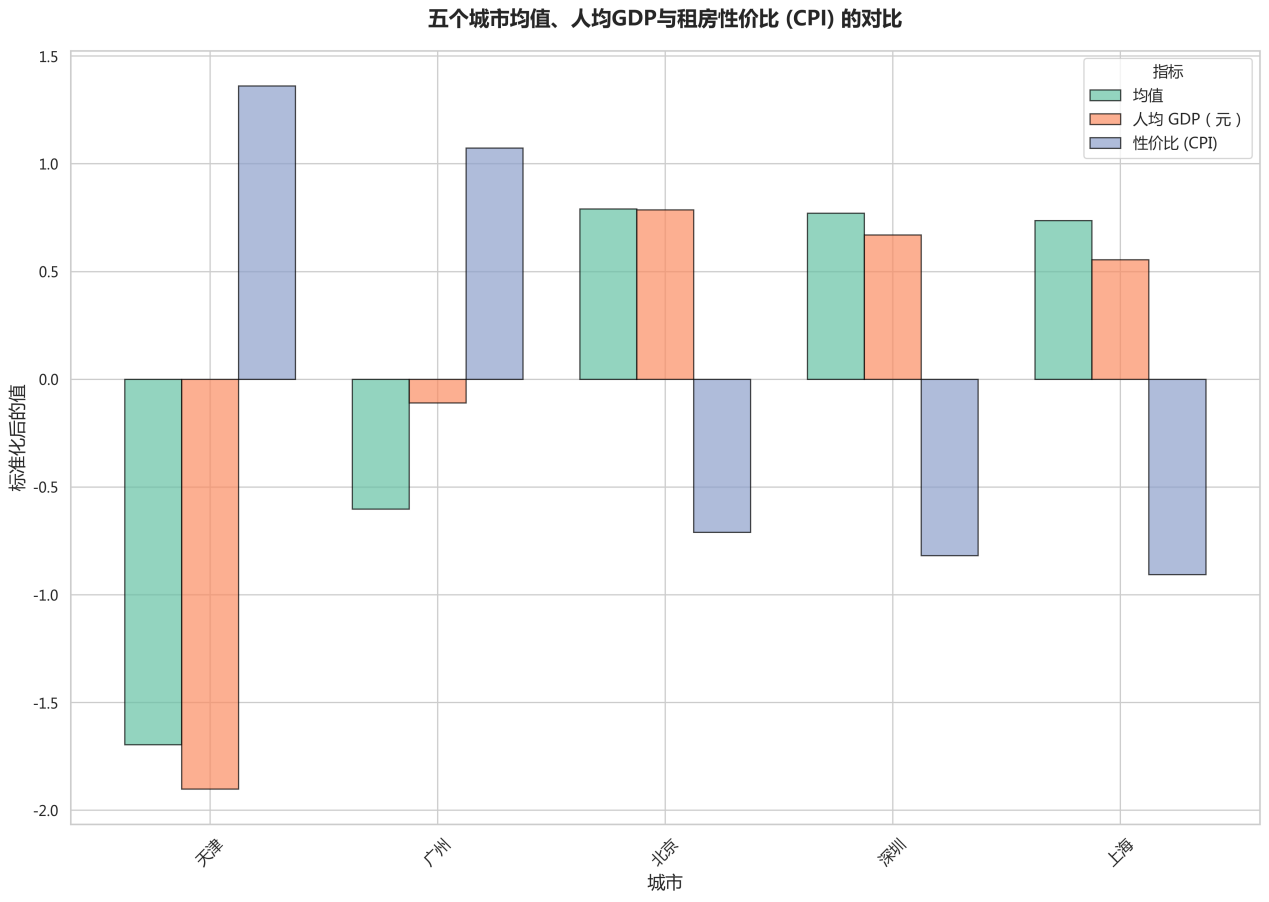


图33 五个城市均值、人均GDP与性价比对比

**5.2.12 各城市人均GDP与租金数据情况分析**

五个城市在单位面积租金、人均GDP及租房性价比（CPI）上的特点和规律。从图30可以看出，北京、上海、深圳的人均GDP较高，但其单位面积租金也较高，因此性价比（CPI）处于较低水平，特别是上海的CPI值最低，这表明上海的租房性价比在五个城市中最差；而广州的人均GDP相对适中，但由于单位面积租金较低，其性价比高于其他一线城市，显示出更高的居住性价比。

图31展示了五个城市的单位面积租金分布和人均GDP水平。天津的人均GDP虽然最低，但其单位面积租金也远低于其他城市，因此总体性价比较高；广州的单位面积租金分布较为集中且低于北京、上海和深圳，因此更适合经济型需求人群；而北京和深圳的租金分布较宽，存在明显的高租金房源拉升整体均值。

图33清晰地说明了五个城市在租金、人均GDP和性价比上的差异。广州在性价比（CPI）上表现最好，且单位面积租金较低，人均GDP适中；深圳和北京的性价比表现较为接近，而上海由于租金过高，性价比显著低于其他城市。

综合来看，图32反映，从单位面积租金角度，性价比排名为广州、天津、北京、深圳、上海；从人均GDP和租金均值关系分析，广州在一线城市中优势明显，性价比更高，而天津虽然人均GDP低，但因租金水平低而保持较高的性价比，适合经济型租房需求。上海虽然经济发达，但租金高昂导致性价比最低，显示居住成本压力最大。

**5.2.13 各城市人均月收入与租金数据情况展示**

根据各地统计局发布的2023年数据，我们获取了北京、上海、广州、深圳和天津的城镇全单位就业人员年平均工资。为统一比较标准，将年平均工资除以12，计算出月平均工资。结合各城市的单位面积租金水平，分析租房性价比指数（CPI），定义为月平均工资除以单位面积租金（见四:4.3一节），CPI值越高表示性价比越高。

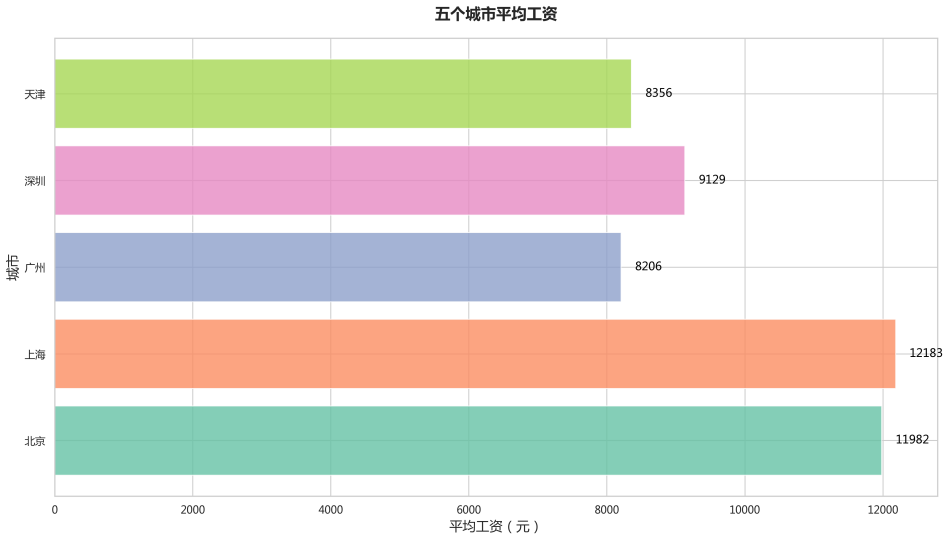
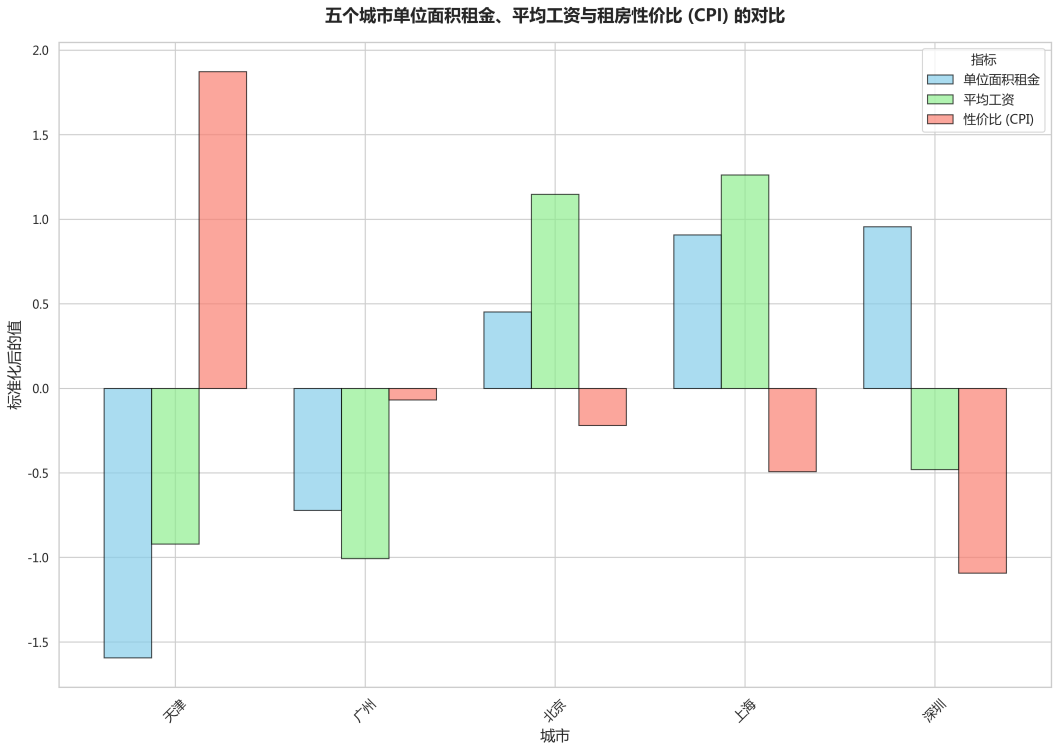


图34 五个城市平均工资对比

图35 五个城市单位面积租金、平均工资与性价比对比

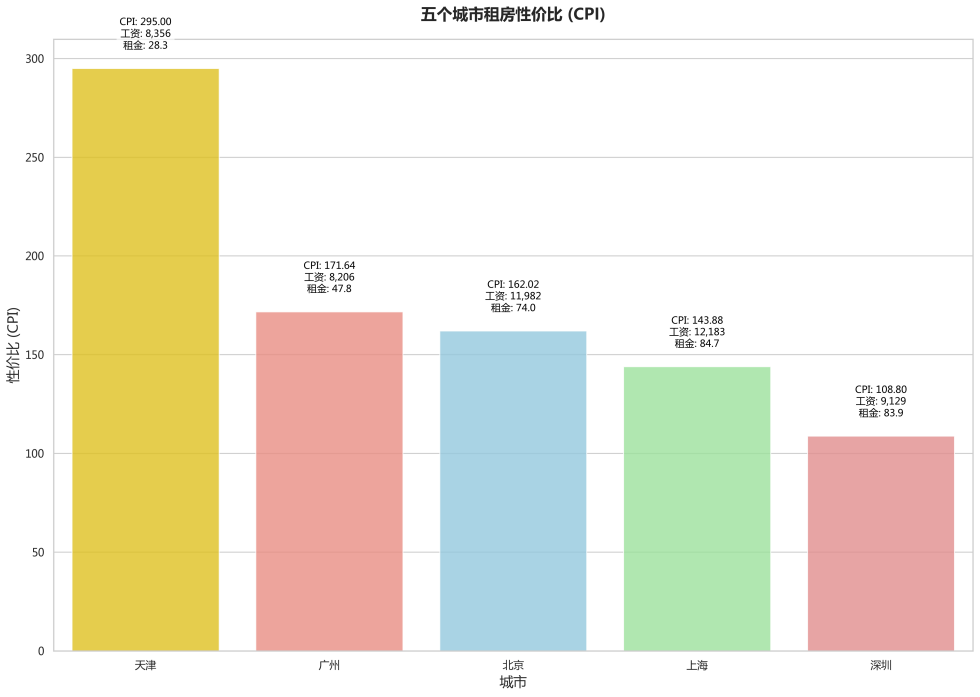


图36 五个城市租房性价比对比

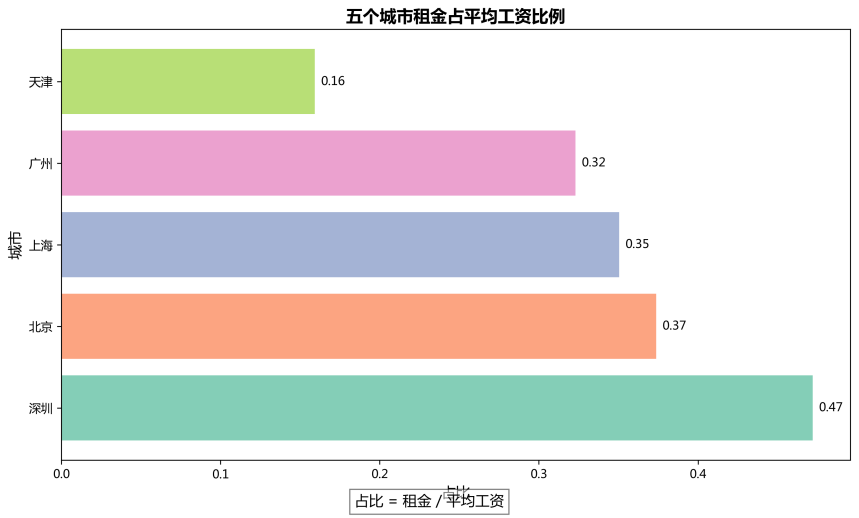
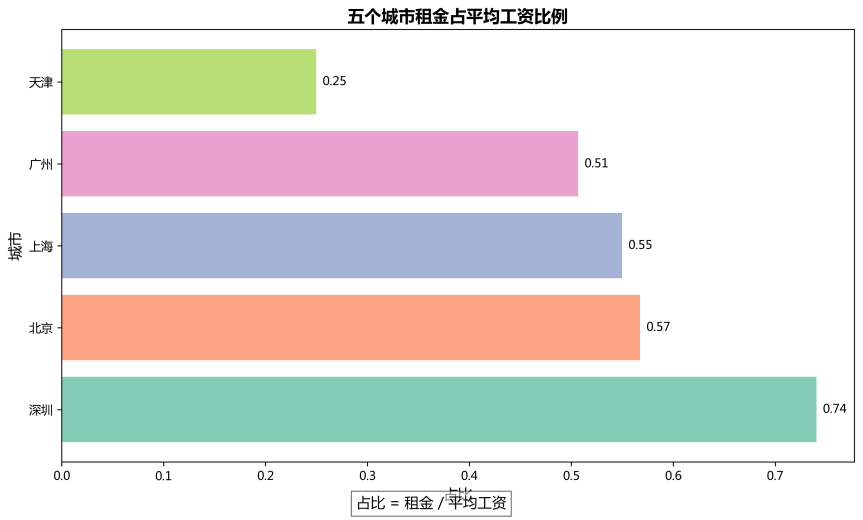


图38 五个城市租金占非私营单位平均工资比例对比

图37 五个城市租金占全单位平均工资比例对比

**5.2.14 各城市人均月收入与租金数据情况分析**

根据图34-36，2023年天津、北京、上海、广州、深圳五个城市的租房性价比和工资占比数据揭示了一些关键的经济特点。首先，天津在租房性价比（CPI）和租金占工资比例方面表现最佳。天津的城镇全单位月平均工资相对较低，但其单位面积租金水平也明显低于其他一线城市，使其性价比显著高于广州、北京、上海和深圳。此外，租金占月工资比例仅为0.16，显示租房负担最小。

由图36与图38，在一线城市中，广州的性价比高于北京、上海和深圳，CPI值达到171.64，租金占工资比例为0.32，表明其租房负担在大城市中相对较轻。北京和上海的性价比接近，CPI分别为162.02和143.88，但租金占非私营单位工资比例分别为0.35和0.37，显示其租房负担适中。深圳的性价比最低，CPI仅为108.80，租金占非私营单位工资比例高达0.47，表明租房压力最大。在图37中，对于收入一般的全单位居民而言，如果选择租平均价格的房子，房租比例相当大，深圳更是达到了0.74，可以看到两类居民贫富差距较大，可供选择的房子价格也相当不同。

结合工资和租金的绝对值分析，天津的月平均工资很低，但由于租金也显著更低，其性价比优势明显。广州虽然工资水平最低，但其租金负担更小，性价比在特大城市中表现最好。相比之下，上海和深圳由于单位面积租金过高，性价比劣势较为显著，特别是深圳，其高租金水平使得租房负担远高于其他城市。

从以上分析可以得出结论，天津作为新一线城市，租房性价比最高，租房负担最小；而广州在一线城市中性价比最佳，租金占工资比例较低，更适合经济型租房需求者。北京和上海的租房性价比中等，租金负担适中；深圳由于高租金和较低的性价比，对租房者的经济压力最大。以上数据为城市居民的租房选择提供了重要参考依据。

## 数据总结

1、(单位面积)月平均租金分布城市顺序: 上海 ≈ 深圳 > 北京 > 广州 > 天津

2、(单位面积)月平均租金分布户型顺序: 三居 > 二居 > 一居

3、靠近市中心的板块，平均月租金 > 远离市中心的板块的平均月租金

4、靠北的城市(北京/天津)朝北方向房屋均价较低；靠南的城市(广州/深圳/上海)朝西方向房屋均价较低，但差距很小。总体上取决于地理位置分布的影响。

5、租房性价比由高到低依次为: 天津 > 广州 > 北京 > 上海 > 深圳

6、租房负担由大到小依次为: 深圳 > 北京 > 上海 > 广州 > 天津