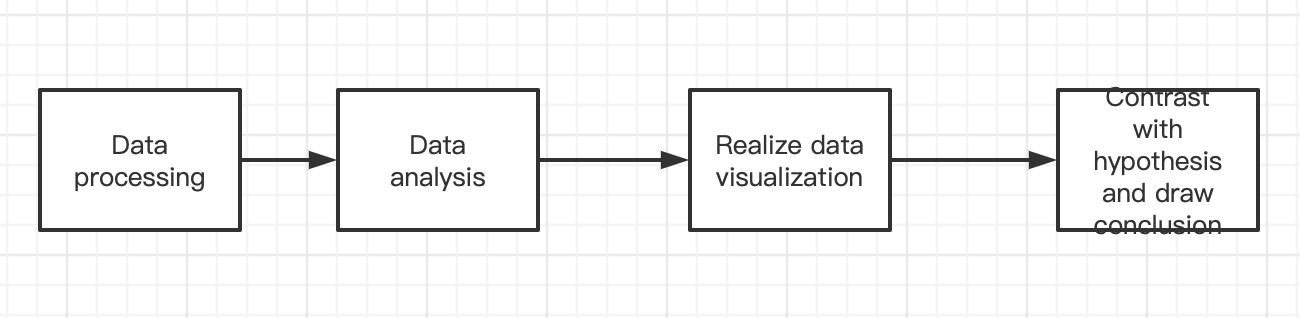
**Changes in the real estate market in different cities in China due to the impact of objective factors in recent years**



**1.Contribution:**

Yunjia zhang:

Completed the real estate market analysis in Nanjing area, put forward two hypotheses, and carried out data analysis to solve them.

All the code content has been placed in the folder: Nanjing

**2.Hypothesis:**

**Hypothesis a):**

The most active area of a city's real estate transactions is in the most prosperous commercial center of the entire city.

**Hypothesis b):**

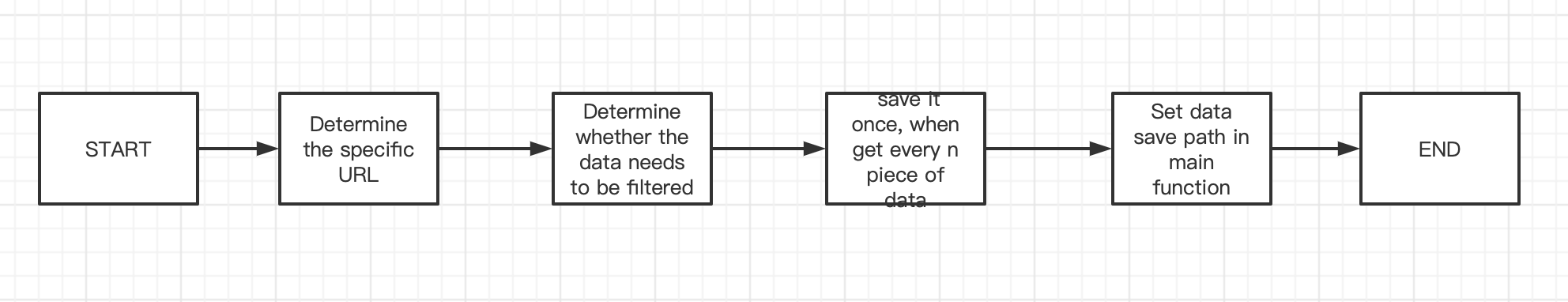
The housing price of a city is affected by a combination of many factors, such as the type of housing or the area of the house. I think that in cities with relatively developed economies, the demand for office buildings is the highest in the real estate market, followed by more spacious residential buildings.

**3.Why we choose this topic?**

The reason for choosing this topic is that we can obtain real-time market data that can be open sourced. In this case, data analysis will be more effective. The city I chose to analyze is Nanjing, and in Nanjing, China, real estate is still the most popular investment field, and its popularity is enduring. Combining the above two points, I think this is a very realistic topic.

**4.Data collection**

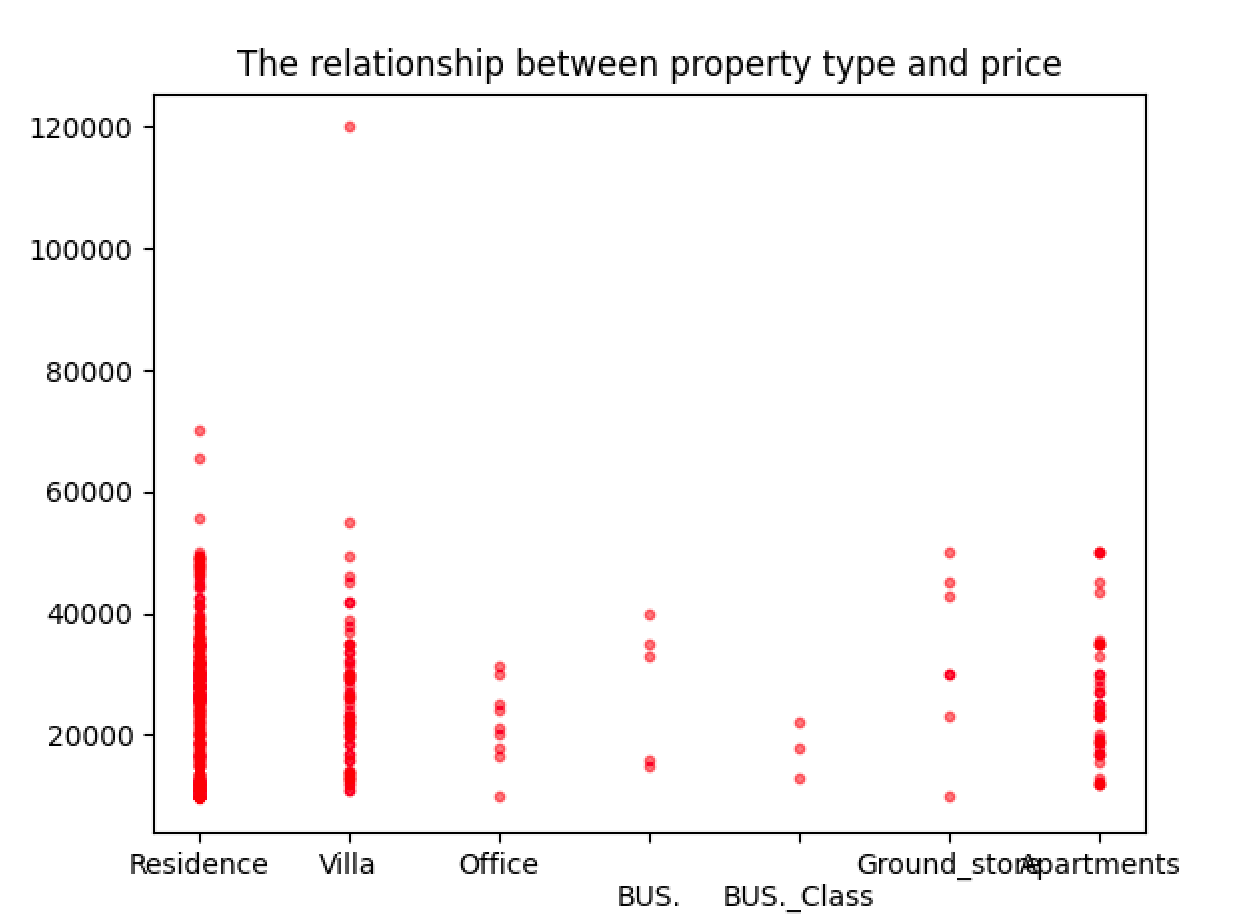
The data used in the hypothesis I put forward is mainly crawled from lianjia website, which is a real estate rental website that provides real listings. It provides first-hand market information. Before crawling, I checked the robots.txt protocol of this website and confirmed that the information I crawled complies with laws and regulations. And the network request made is controlled to ensure that it does not interfere with the normal operation of the website.

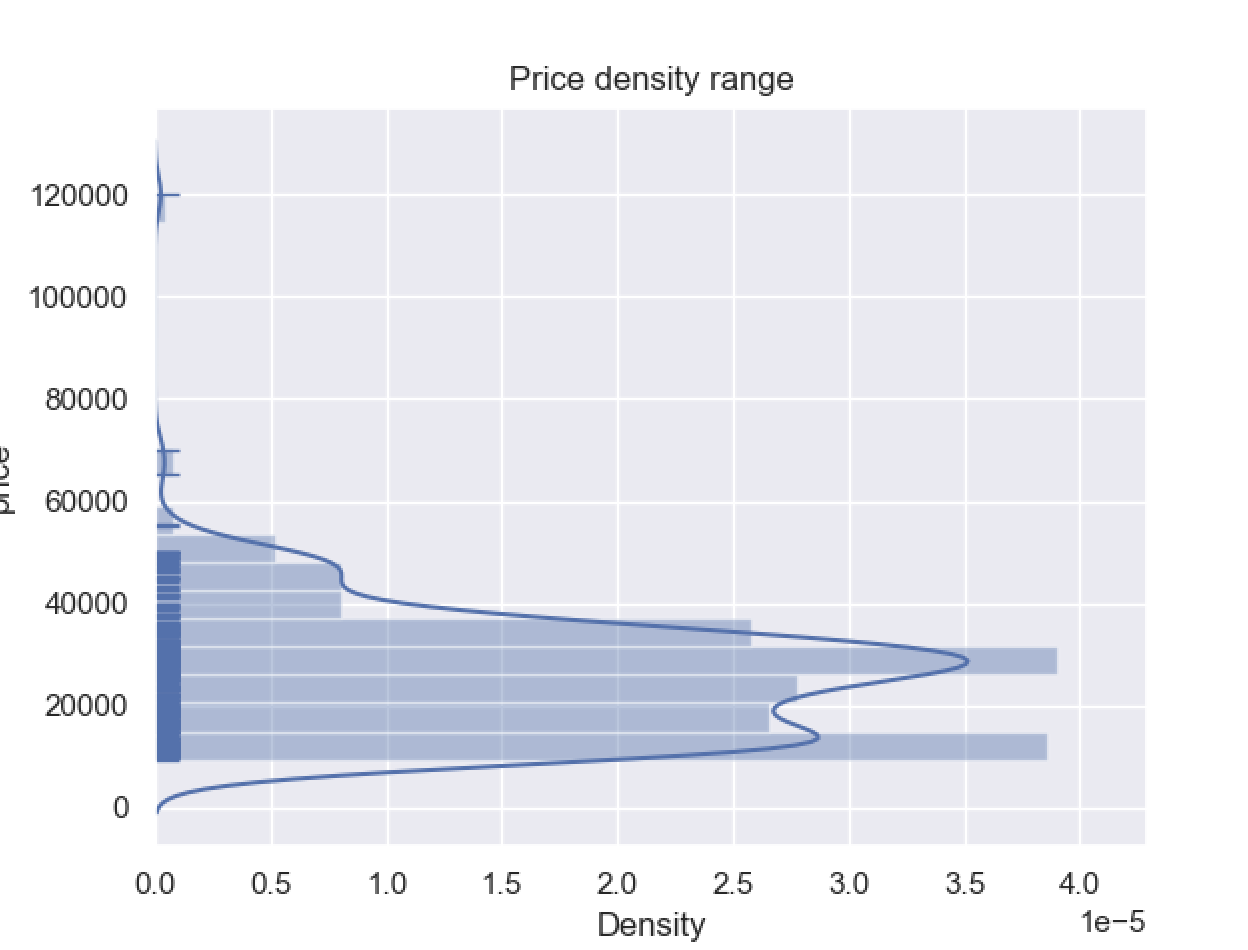


**Data crawling website:** <https://nj.lianjia.com/>

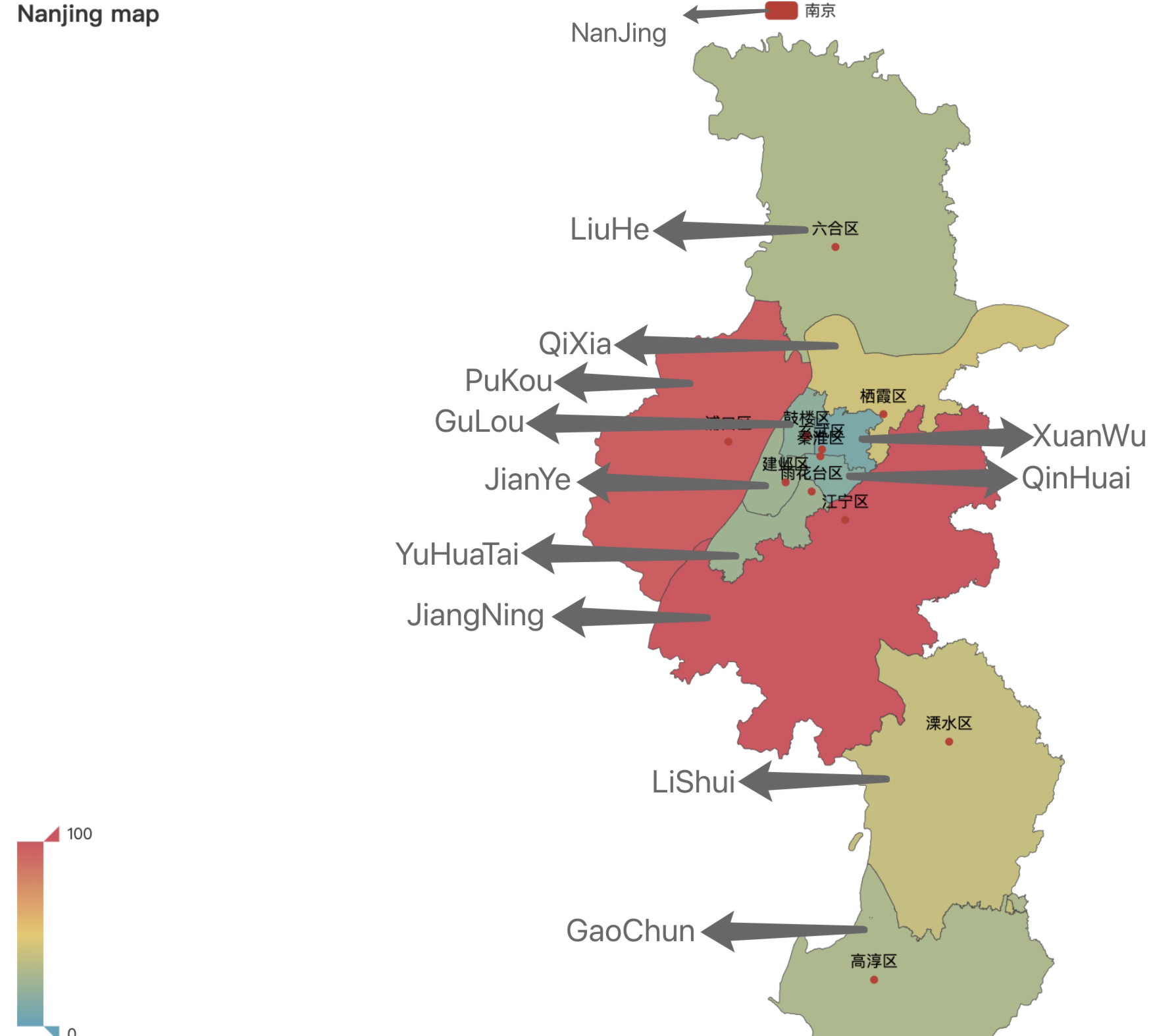
1. **Data analysis：**

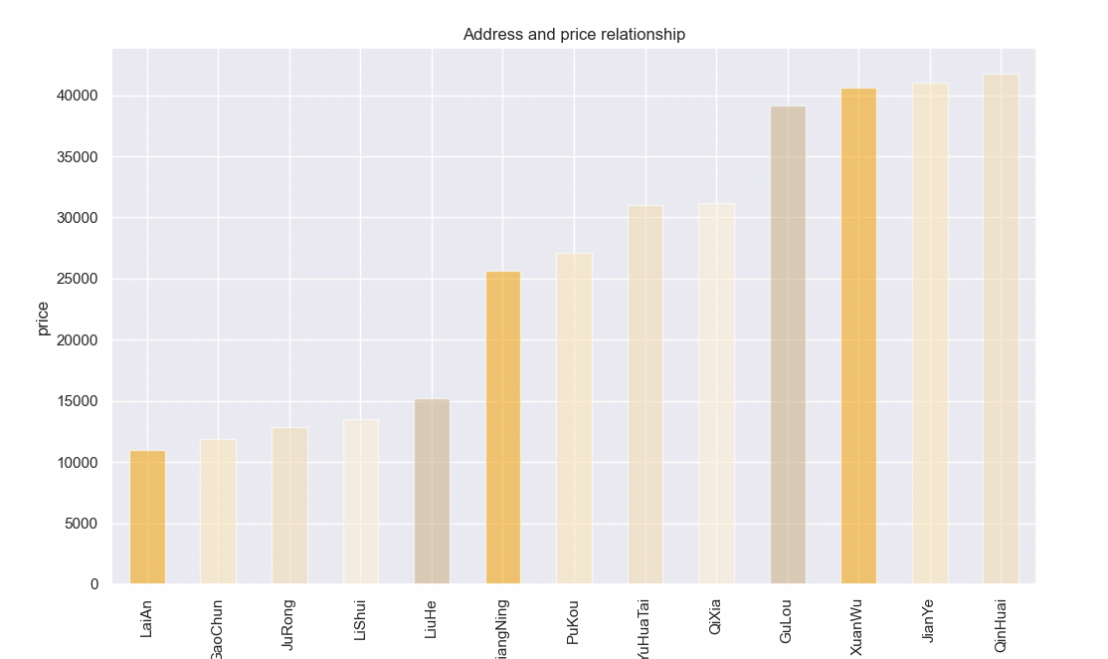
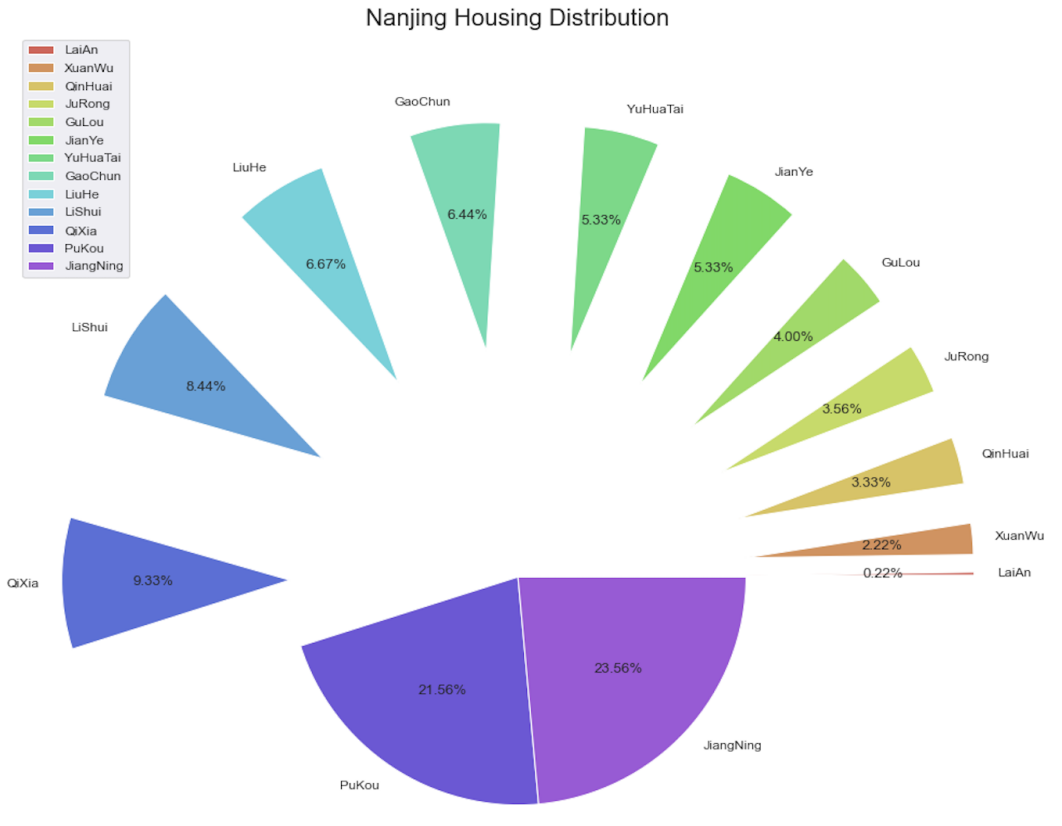
First I conducted an overall analysis, generate a scatter chart to show the general trend of different house types in different price ranges. And I analyze the unit price density per square meter of houses from the entire city to determine the price of most houses and the overall price distribution.





**Analysis of hypothesis a)**

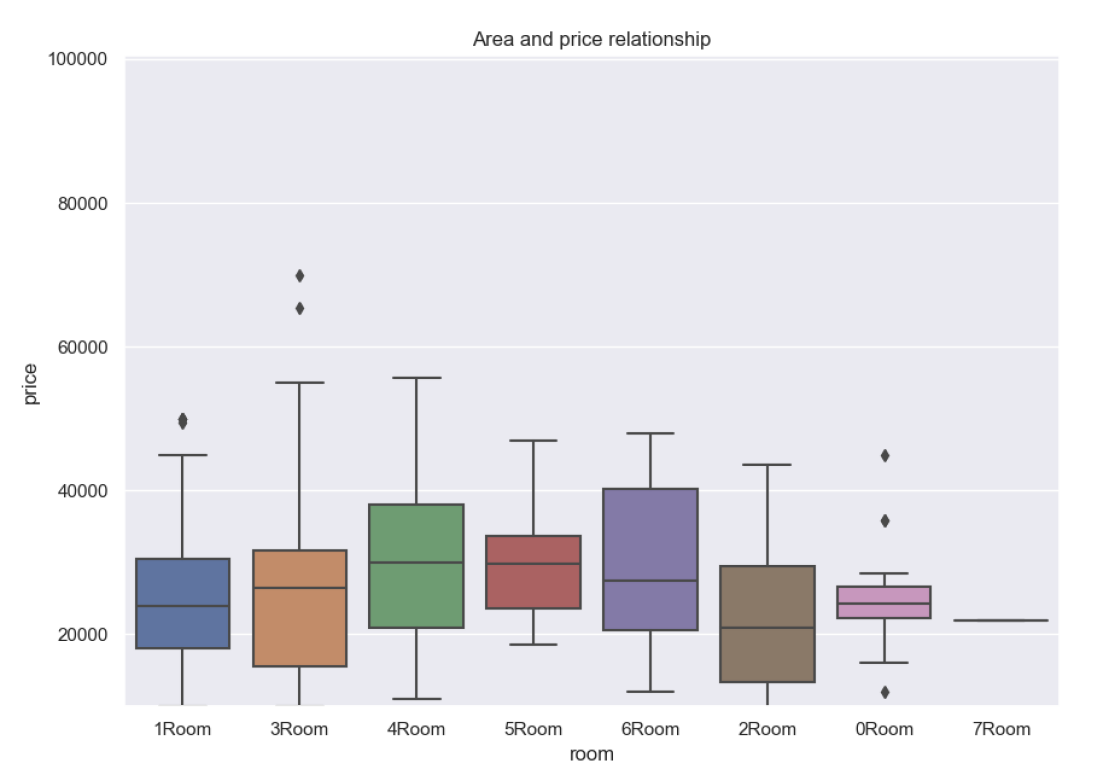


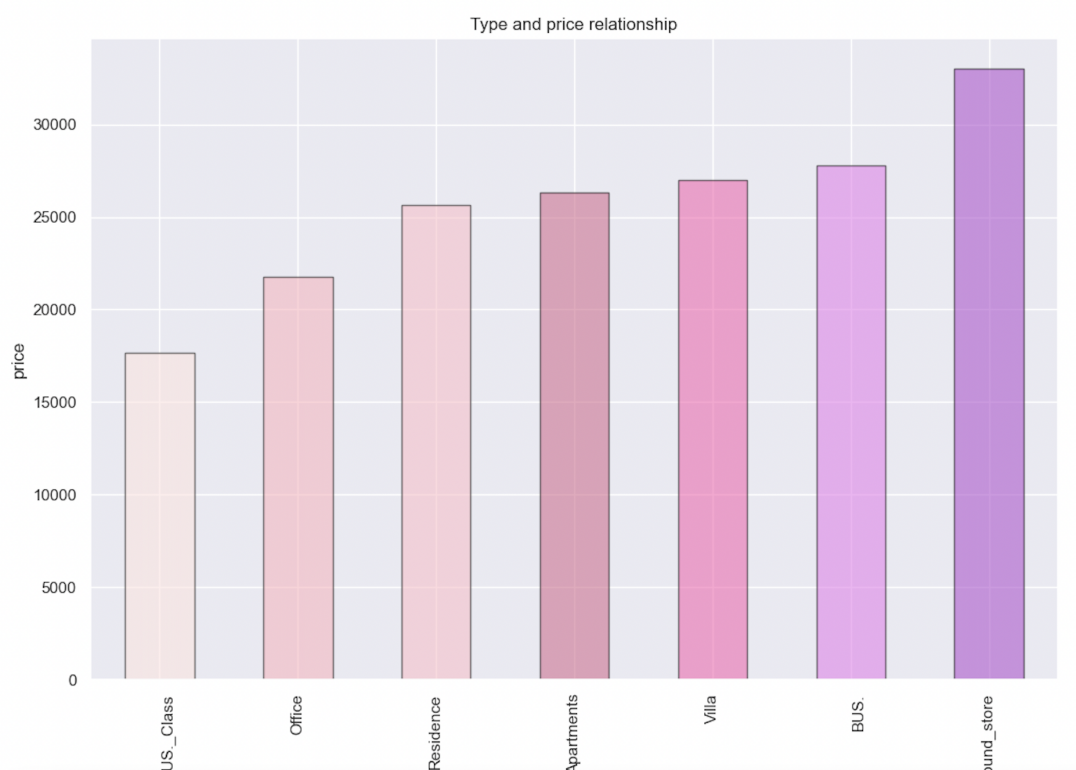


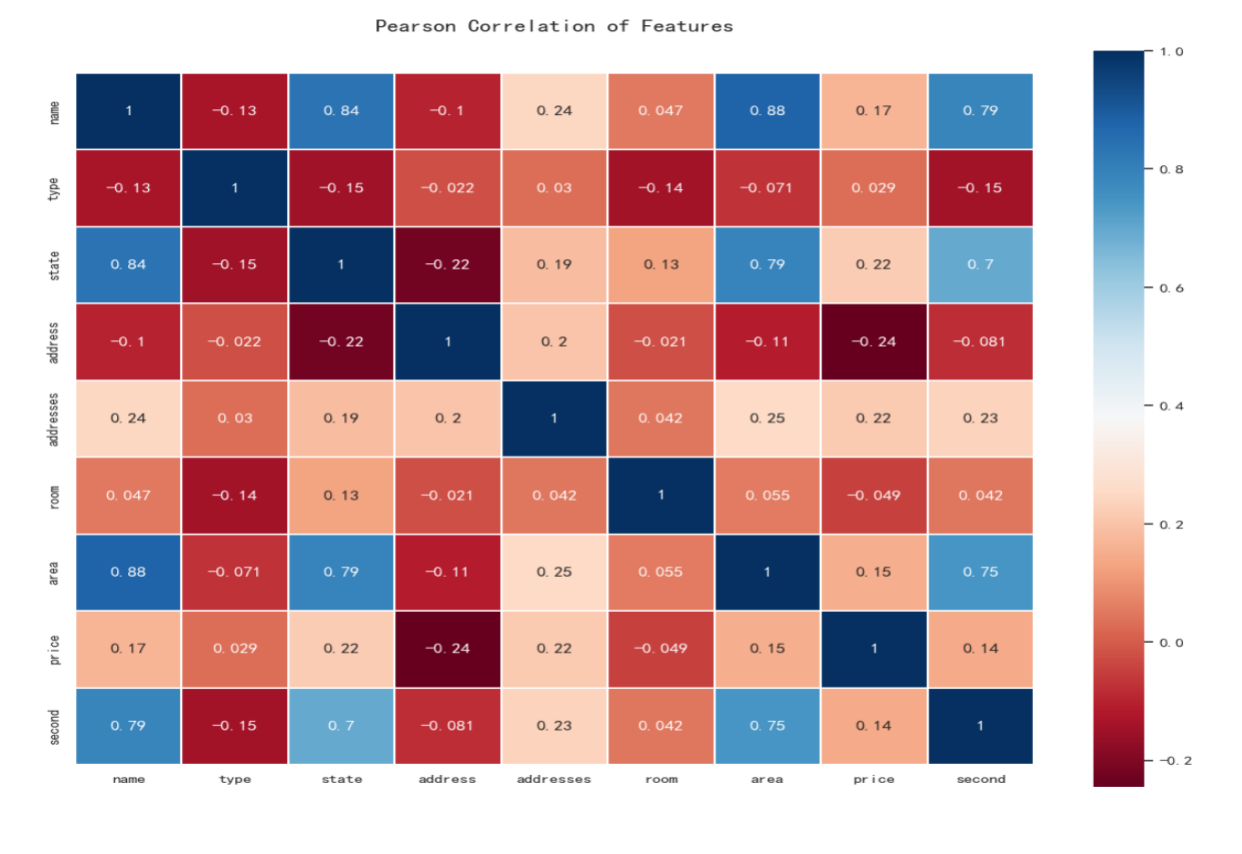
I first analyzed the number of sold houses in each district of Nanjing, and then by calling pyechart, combined with the analysis results, I judged the popularity of real estate through the depth of color. The darker the area, the higher the rate of house purchase. From the above figure, we can judge that the hypothesis is wrong. The economic centers of Nanjing are in Gulou District and Qinhuai District, but the colors of these two areas are very light, but the sales rate in the surrounding suburbs is higher. The same result can be seen from the above pie chart. Based on this result, I checked some information and found that the city's high-frequency real estate transaction zone is located in a high-tech development zone. I also analyzed the housing prices in the area and found that the housing prices in the prosperous areas are the highest.

**Conclusion:** Although housing prices in prosperous areas will be higher, the more active real estate market does not appear here. So, the previous assumption is wrong

**Analysis of hypothesis b)**







I think that the relationship between supply and demand in the real estate market can be directly reflected in the housing prices, so I analyzed the housing prices according to the different factors of the house.

I first analyzed the relationship between housing types and housing prices, and found that the housing prices of the sub-commercial type are the highest, while the prices of residential areas and offices rank lower among all types. This result shows that the market demand for residential and office buildings is relatively low compared to other categories.

Then I analyzed the box plot of the influence of the number of rooms in the house on the price. It can be seen that the average price of having four or five rooms is the highest, and having more or fewer rooms will affect the price. This result shows that the market demand for housing is relatively spacious.

Finally, according to this heat map of various comparisons of different factors, it can be seen that the factor that has the greatest impact on housing prices is the location of the house, rather than the type and floor space of the house.

**Conclusion:** the basic error of the assumption is that most of the potential buyers of residential houses are indeed more inclined to more spacious residential houses, but this is not as high as the market demand compared with villas and commercial land.

**For this project I refer to the open source code of some packages：**

**(The front-end html file used by pyechart comes from the open source code of the package)**

**<https://cloud.tencent.com/developer/article/1747830>**

**<https://blog.csdn.net/mighty13/article/details/117857200>**

**https://cloud.tencent.com/developer/article/1747830**

**<https://github.com/jiangsir/PythonBasic/blob/master/basic_line_chart.html>**

**<https://pyecharts.org/#/zh-cn/global_options>**