

Prediction of Housing Price in Beijing Based on Linear Regression Model

Donghao Shao*

School of weiyu international school, Shanghai,China

*Corresponding author: 1621020444@stu.cpu.edu.cn

Abstract. With the booming of the economy, more people are willing to have their own house and want to live in Beijing. But the increasing price make people feel confused and hard to do the choice. So this content describe the situation of housing price rise up in Beijing in 20 century by collecting the data and analyze the history and predict through linear regression. This action can help citizen to decide whether they should buy a house and where they are supposed to buy. From the content, the research demonstrates the city center is the most popular place and people want to live the place with the transport. What' s more, the housing price in Beijing will increase in the future. It can help people realize the how the house price system operates and let them to make the correct decision.

Keywords: Beijing house price, linear regression, price predict

1. Introduction

In the 2000s, China's housing market met an extraordinary growth, and the purchasing power reached an unprecedented high level. As the result of it, the economy showed a good prospect. Fudan University analysed the GDP data, it showed that housing and its industrial chain account for 17% of national GDP, the real state generated about 8% of total GDP directly in 2020. As the representative of the housing price, Beijing known as the capital city, it contains people from around city such as, Shijiazhuang, Qingdao. This action leads the population in Beijing more over it can hold. As the Atlantis press says "In recent years, housing prices in Beijing have continued to rise, and high housing prices have brought a series of social and economic problems to Beijing. As an ordinary resident, choosing to buy or rent a house in Beijing is a problem that plagues many people. If from the perspective of the rent-to-sales ratio, Beijing house prices are too high and there is a bubble, so the rational perspective should be renting, but, from the perspective of investment, the appreciation of real estate is better, and it is suitable for investment in Beijing." [1].

With reference to some researches on Beijing housing price such as Chen suggested the house price will grow constantly nearly ten years through hedonic regression in his article. [8]. and Song and Wang used the mathematical model as the the base model and find the optimal problems [10]. Also, there is an article from Huang tainba shows a new method of find the rule of exploring the second-hand market through the algorithms.[11] And the article written by Philip Wen, published in October 7,2016. This article shows the developed cities' house increase about 3-5 percentage.[12] The GlobalPropertyGuide analyses the second-hand house and publishes House prices in China; second-hand house price index, Beijing, the main chapter is after 2015 the house market become crazy, the price reach a extremely high level.[13] IORI KAWATE, Nikkei staff explain the reason of booming in the house market which is the high input population from surrounding city need to live in.[14] Chengjiehe the content describe the base of the Beijing house and analyse what is possible happen in future.[15] Yongzhou Hou is an article that claim the economics bubble in house and the predict.[16] The content show the price change in Beijing an the real estate price index. Xinhua report the new policy that limit the house market in Beijing issues new rules to limit house purchases.[17] And Aljazeera a social media publish China's new home prices fall for the first time since December to share their opinion on the first fall on house price in Beijing.[18] And Iga Akito focus on the shanghai and Beijing's house price inspire this article a lot. What is more, Ryan Dong Chen analyze the Beijing house market by the machine learning.

With the research, they all use the model to make a study and make several compares . In that case, it suggests to make the research with linear regression.

Thus the housing price analyse and predict is the main topic of content and it can help those people who want to purchase the house in Beijing to choose the most valuable and suitable living place.

2. Data and Method

2.1 data collection

This paper collects the data from the database in the www.lianjia.com, a website that contain the most horsing price in china.

At first, the machine learning is found that the data is not working because there is some missing details and some data like the room type are not easy to analyze and collect. To know the exact position of the housing I use the distance between the house and four landmarks: the Tainanmen, guomao, xinjiekou, tiantan. And the transport is one of the key reason. Nowadays, the convenience of housing become increasingly important. The data also described the number of the social transport such as subway and bus. This is the signal of the function of a housing. It is relative to the housing price.

2.2 Linear Regression

Linear regression is an algorithm used to predict, a relationship between two various features. In linear regression, there are two kinds of various being examined: the dependent variable and the independent variable. This is the definition of linear regression and the content will claim the different aspects of linear regression.

The linear regression can make the predict more clear, the content take a brief look at four techniques to prepare for the model, sometimes there is missing message and implement when people search the data but it is enough to get a flavor of the computation.

2.2.1 Simple Linear Regression

As simple linear regression when people have a easy input, we can use statistics to analyze the coefficients.

It requires you to calculate statistical attributes based on data such as mean, standard deviation. All data must be available to traverse and calculate statistics. The data are useless in this process. So when the machine learning operate people should avoid this happen.

2.2.2 Ordinary Least Squares

When people have multiple inputs, people can use the ordinary the smallest squares method to estimate the value of the coefficients.

It is unique to implement the Ordinary Least Squares procedure themselves unless a training in linear algebra. It is more possible that people will call a procedure in a linear algebra library. This procedure is extremely speedy to operate.

2.2.3 Gradient Descent

When there are one or more inputs you can use a process of optimizing the values of the coefficients by minimizing the error of the model on your training data.

This operation is called Gradient Descent and works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

Gradient descent is often taught using a linear regression model because it is relatively straightforward to understand. In practice, it is useful when you have a very large dataset either in the number of rows or the number of columns that may not fit into memory.

2.2.4 Regularization

There are some extensions to the training of linear models, called regularization methods. These methods are designed both to minimize the sum of the squared errors of the model on the training data (using the ordinary least squares method) and to reduce the complexity of the model.

3. Results and Discussion

3.1 Descriptive Results

As shown in table 1 it is five of the data that are most representative. These data show the size of house, the followers, the price, the floor and the location.

Table 1. Descriptive Results

Lng	Lat	DOM	followers	Totalprice	square	ladderratio	low	bottom	end	top
116.7	40.0	1464	106	415	131	0.217	0	0	0	0
116.45	39.9	903	126	575	132	0.667	0	0	0	0
116.56	39.8	1271	48	1030	198	0.500	0	0	0	0
116.43	40.0	965	138	297	134	0.273	0	1	0	0
116.42	39.9	927	286	392	81	0.333	0	0	0	0

Table 2. Growth rate

Dis_TianAnMen	Dis_TianTan	Dis_GuoMao	Dis_XinJieKou	growthrate
0.142	0.150	0.105	0.128	1.82
0.070	0.411	0.035	0.104	1.76
0.176	0.149	0.102	0.202	1.96
0.176	0.194	0.164	0.147	1.93
0.045	0.016	0.047	0.081	1.83

Table 3. Ladder ratio

Renovation condition	Building structure	Ladder ratio	Elevator	subway	district	average
3	6	0.217	0	1	7	56021
4	6	0.667	1	0	7	71539
3	6	0.500	0	0	7	48160
1	6	0.273	0	0	6	51238
2	2	0.333	1	1	1	62588

Figure 1 claims that the price is the most important thing all the citizens are watching on it. The second choice is the square of the 0.57, and the type of room is also the topic that the people focus on.

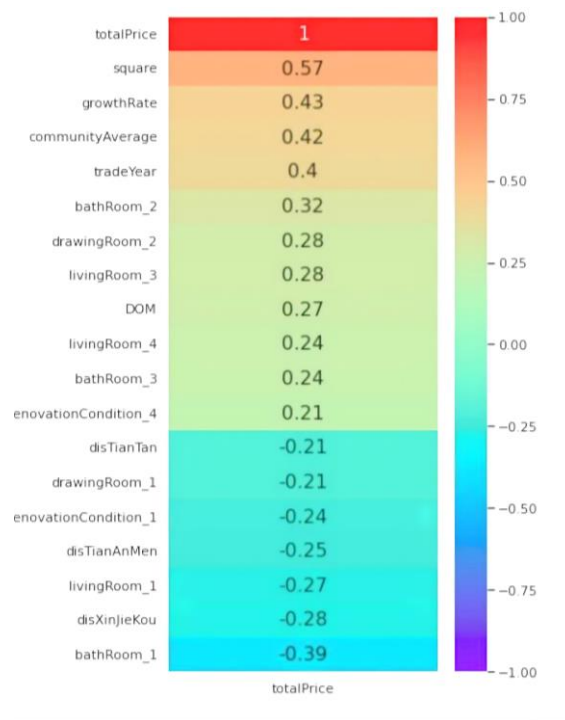


Figure 1. floor table

As shown in figure 2, it demonstrates the different part of the component of the housing. The main character is the price and the name. Then housing type is analyzed a lot. It shows that people always focus on the price at most. And the research also collect the trading time.

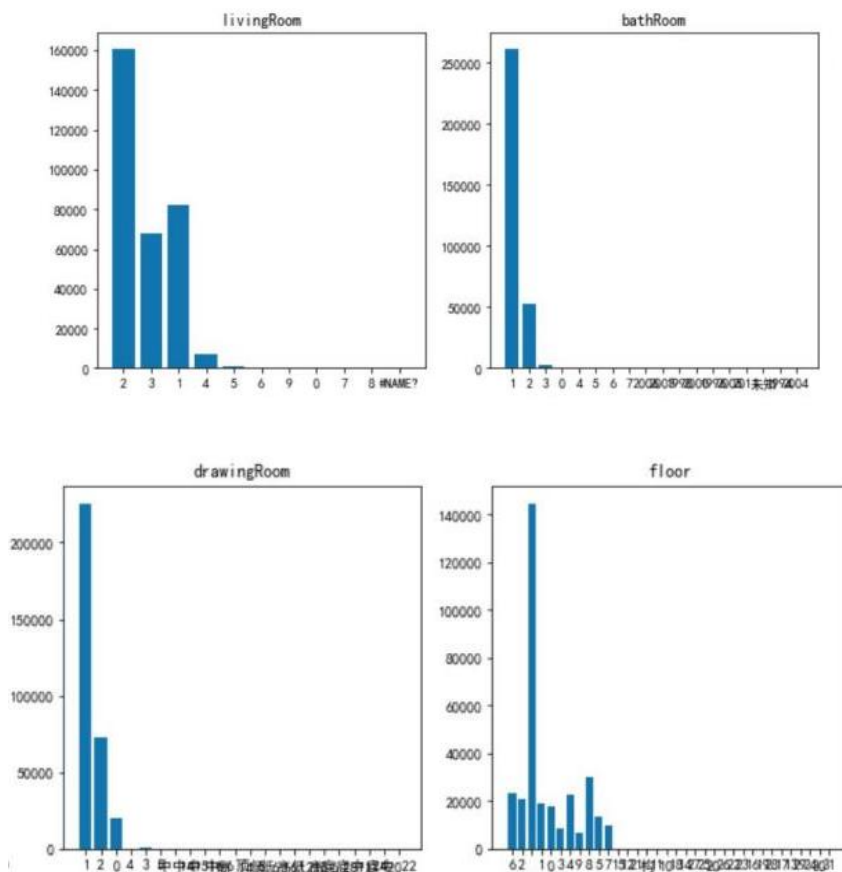


Figure 2. different part of the component of the housing

As the programme operate, the time period table show the trend of the purchasing time. The trend is increasingly rising up. Then it reach the peak in 2017. The increasing trend is extremely stable (see in figure 3).

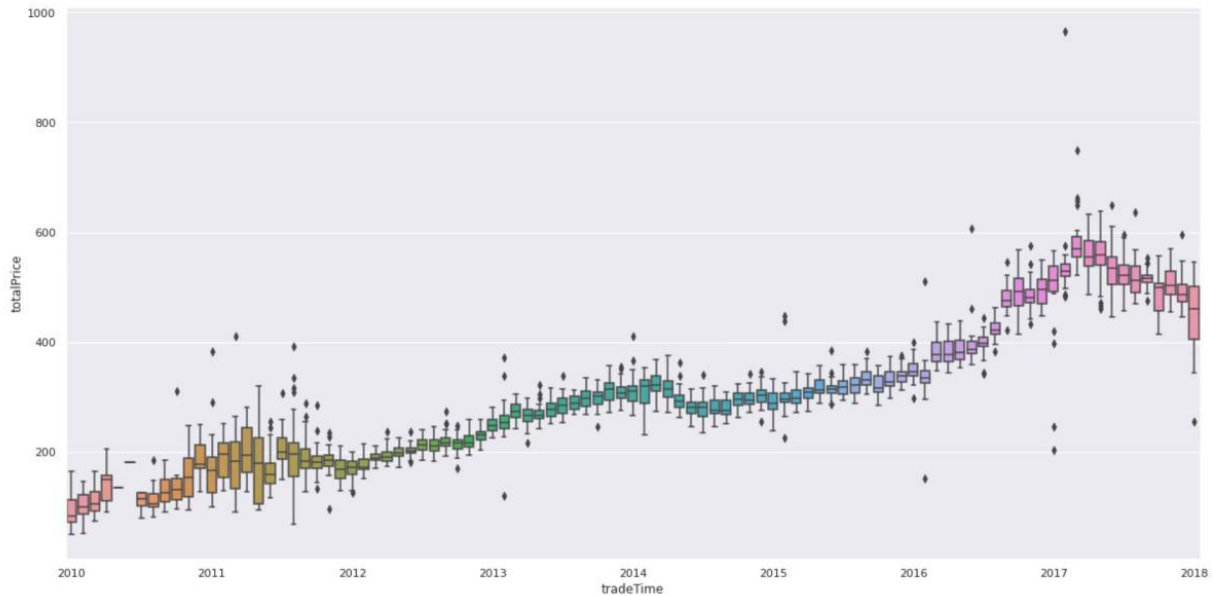


Figure 3. Trade time period

After that, the program shows that the different levels of the key words such as distance between the landmarks, the different functions of the horse. So use the thermodynamic diagram to make the popular place light. Then use the thermodynamic diagram to make the popular place light.

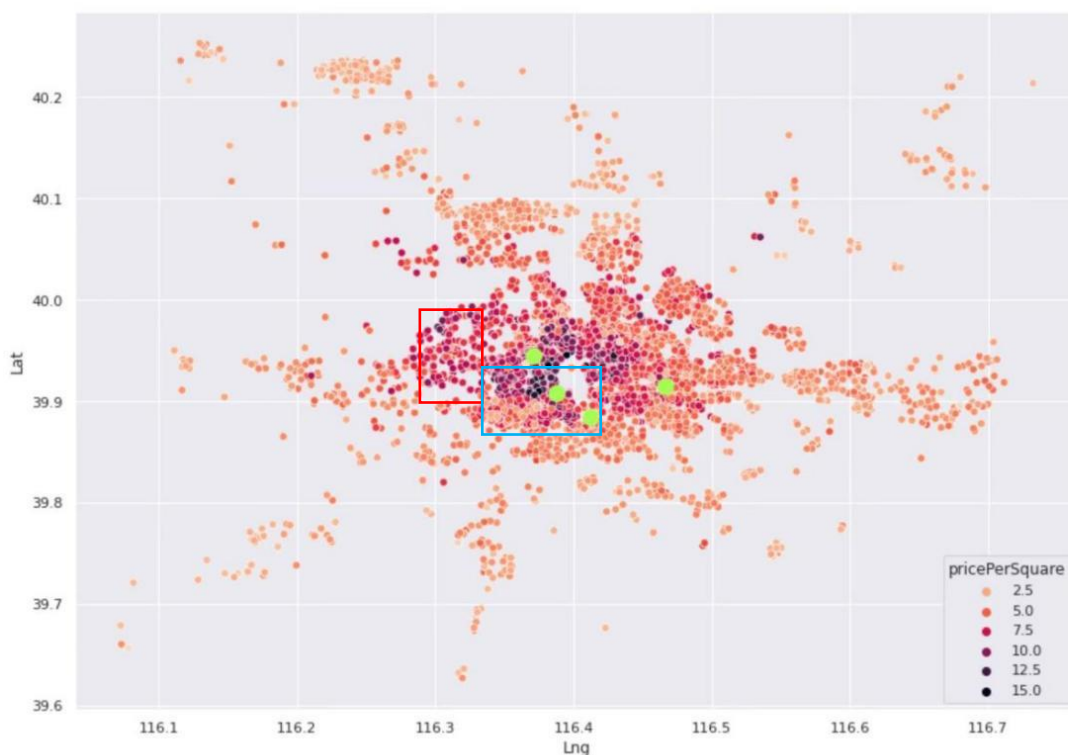


Figure 4. thermodynamic diagram

The figure 4 demonstrates the main housing in Beijing, it is not hard to discover that the fire is in the middle of city. And there are 4 yellow points are landmarks. The horse are distributed around the landmarks, the house density is super high in these place. The other house state are dispersed in wave style with lower density. And in the suburb the house distribute in an extremely spread way.

The red frame square contains two landmarks. The place between two point is filled with black point. It means the houses are most expansive which are popular, people are willing to purchase the house state there. With the research, the research found that one of the landmark is a shopping market. So the shopping center can determined people's decision.

The blue square circle the part of city which below the red frame. It is quite similar with the picture upon, but there is a little difference. The first is that the landmarks is a museum so the population is less than the graph mention before and the transport is not as convenient as others.

In summary, I have analyzed the past 10 years data, I found that the people are more willing to live in the city center than the suburb and they are willing to live in the place with transport. Especially, the house that are close to the landmark like shopping mall and museum.

This place has a good prospect to increase the price.

3.2 Prediction Result

This content claim that the house price will increase in the future about 5 years. Meanwhile, the buyer will increase as well.

The main reason is the fast-paced lifestyle make people need to find a place that close to the working place for not wasting time on the way to work. The second point is there are huge amount of people want to work and dwell in Beijing city, the working population male the housing price rise in a stable trend.

For predicting the housing more correct, this paper uses the machine learning to achieve the goals. Using the linear regression as the base of learning.

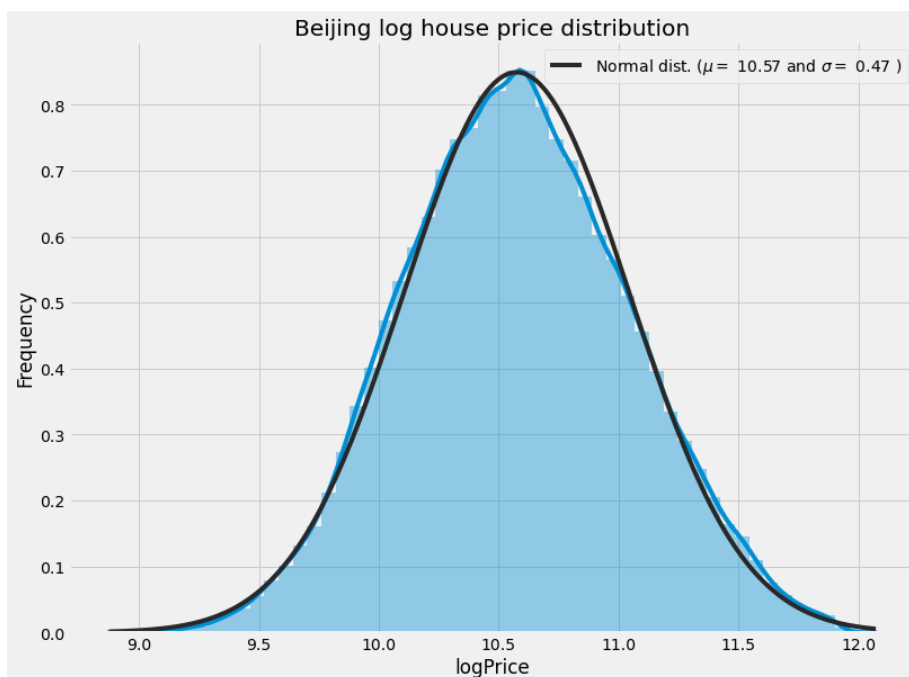


Figure 4. normal distribution

This is the normal distribution of Beijing house. From the Figure 4, we conclude that the medium is 10.5 and the peak is 0.9 located in 10.5. The range is between 9.0 and 12.0. It shows us about the percent price in Beijing is much more higher than other city's house price. For example, the price in Shijiazhuang is 5.5 and in changzhou is 4.

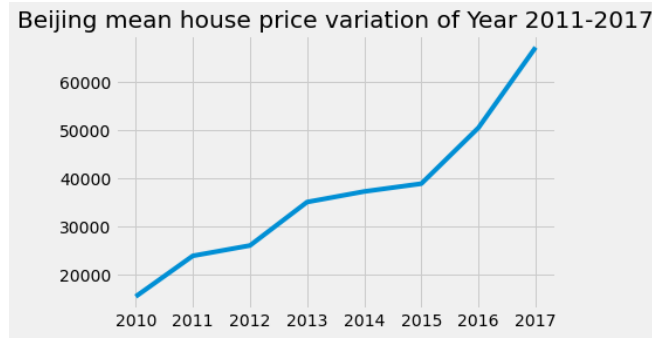


Figure 5. price variation line graph

Figure 5 is the linear graph of past 10 years price. From 2010 to 2017, the price in 2017 has been three times more than the price in 2010. Before the stable rise during 2010 to 2015, the price surge begin at 40000 in 2015 and end at 65000 in 2017. The reason is the policy set up and the economics boom in 2015. It change the market of house, people want to buy the house as the investment. For the price mechanism, the supply of house is limited and the demand rise. As the result, the price increase.

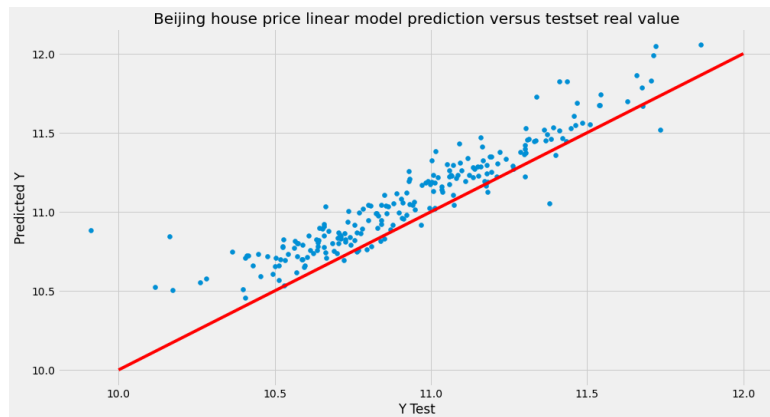


Figure 6. linear regression

Using the linear regression to train the model. This is the first result of the model the orange line is the predict line of price. The blue point is distribute upon the line and disperse. Around the line randomly, do the train again.

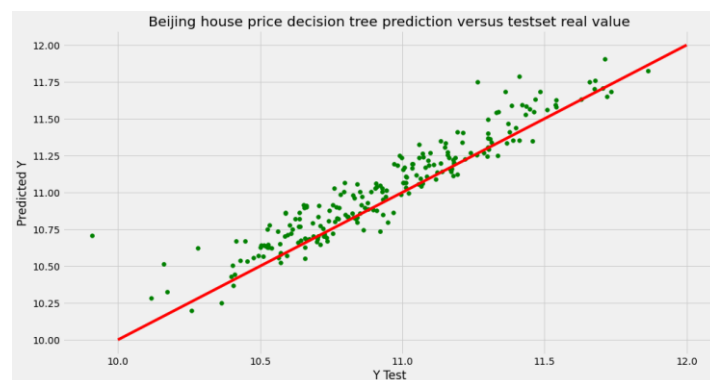


Figure 7. tree prediction

This is the tree model of the housing price in Beijing. It is quite similar to the linear regression but the point is more random, but it is arrange around the line, so i think it catch the line better and more reasonable.

In conclusion, the machine learning show the trend about the future development in line graph. The price increases a lot in a stable trend.

With the following research, the reason can be concluded into two parts: the first is the population, the citizens live in Beijing will increase in future so the living problem will be more serious, people will meet difficulty to purchase the house. The second problem is the investment side, increasing people will put their money into house to keep the value and be the investment. So, the price will still increase.

4. Conclusion

The article focuses on the Beijing house price, with using the linear regression and the machine learning. At first, the article describe the linear regression about the principle and the use. Then the main research shows the most high price take place in the city center are the area which close to the landmark. Especially, the tiantai and the tiananmen area is the most expansive place in Beijing. With the machine learning, the linear regression model and the tree training model show the upwards trend of the house price. It train the model for several times and at last the result is showed in the content. From the figure 6 and figure 7, the graph demonstrates the housing price will increase in past 10 years. Therefore, the content suggests that people should buy the house immediately for the price increase.

Reference

- [1] Anjue,2022.5.11, 2022.9.1. <https://beijing.anjue.com/market/>
- [2] XinfangSong and ShaoqiangWang <House Investment Model under High Housing Prices in Beijing> 2022.5.7 2022.9.1
- [3] <https://www.atlantis-press.com/article/25876546.pdf>
- [4] Property Price: YTD Avg: Beijing 2020.8.8 2022.9.2
- [5] <https://www.ceicdata.com/en/china/nbs-property-price-monthly/property-price-ytd-avg-beijing>
- [6] Wikipedia <Linear regression> 2022.9.11 2022.9.11
- [7] https://en.wikipedia.org/wiki/Linear_regression
- [8] DongMing Chen. <Predicting Housing Price in Beijing Via Google and Microsoft AutoML> 2020.7.11 2022.9.10
- [9] Philip Wen <Beijing, Shanghai, Shenzhen: the cities where house prices rose by 30 to 40 per cent> October 7,2016 2022.9.10
- [10] IORI KAWATE <China's housing price declines spread to more than half of provinces> October 14, 2021 05:30 JST 2022.9.11
- [11] GlobalPropertyGuide <House prices in China; second-hand house price index, Beijing (2015=100)> May,27,2022 2022.9.11
- [12] ChengjieHe <Driving Forces Analysis for Residential Housing Price in Beijing> 2010.10 2022.9.11
- [13] Yongzhou Hou <Housing price bubbles in Beijing and Shanghai?> 2010 March 2022. 9.11 <https://www.researchgate.net/scientific-contributions/Yongzhou-Hou-2110838276>
- [14] Daniel Raff Susan Wachter <Real Estate Prices in Beijing> 2017.3.7 2022.9.11 <https://realestate.wharton.upenn.edu/wp-content/uploads/2017/03/721.pdf>
- [15] Xinhua <Beijing issues new rules to limit house purchases> 2011/02/17 2022.9.11 <https://www.mfa.gov.cn/ce/ceau/eng/gdxw/t795335.htm>
- [16] Aljazeera <China's new home prices fall for the first time since December> 18 May 2022 11 Sep 2022, <https://www.aljazeera.com/economy/2022/5/18/china-april-new-home-prices-fall-for-the-first-time-since-dec>
- [17] Iga Akito <Two sides to China's booming housing > 2021.3.2 2022.9.11 <https://www3.nhk.or.jp/nhkworld/en/news/backstories/1523/market>
- [18] Ryan Dong Chen <An Empirical Analysis of House Price Bubble: A Case Study of Beijing Housing Market> 2013, Vol1 2022.9.11, <https://core.ac.uk/download/pdf/80828138.pdf>