# Beijing Housing Market Analysis: Exploring Determinants through Online Transactions

**By Luyi Huang, Mengyang Zheng**

**May 1st 2021**

## 1. Abstract

With the booming economy and the urban centralization, Beijing has witnessed an inflation of housing prices. The surge of housing prices induces the real estate bubble and unaffordable life costs for residents in Beijing. In addition to this surge in housing price, the uneven urban planning  also brings side effects: the sharp inequality of business development and real estate layout. This unhealthy cluster of business districts, which will further deform the house price distribution in Beijing. We noticed that adding clusters is an important method when analyzing this uneven distributed house market. Meanwhile, it is noticeable that the population of people largely relies on migrant workers. This population structure will experience seasonal change because workers may leave Beijing during certain months in one year, such as during Chinese lunar new year. Based on this phenomenon, we notice that there might be a seasonal fluctuation in the housing price in Beijing from the construction period or house price demand. The paper utilizes the clustered standard error fixed effects model to consider this clustering phenomenon. Results derived from this paper may offer insights into the housing market in Beijing.

## 2. Introduction

Over the last decades, the house market in Beijing has busted. The house price has been rocketing and thousands of buildings are erected. The long existing core in Beijing( the Forbidden City, now DongCheng and Xicheng district) and the uneven distribution of facilities from previous urban planning combine to form the uneven distribution of house price and house types.  In other words, the house market in Beijing tends to have a spatial heterogeneity phenomenon, which will further deteriorate the deformation of the housing market and solidify the spatial heterogeneity. This indicates the necessity of diving into the house market in Beijing to improve the further urban planning development and house market.

The values of amenities and environment on the house market are complex to interpret. The non market valuation is the first and most important step to understand the relation between amenities and house price. However, the past research has been mainly focused on the house price market in the United States while the research on China was not advanced enough. The East Asian countries are significantly different in the house market. Instead of having urban decentralization, East Asian countries, especially China, have urban centralization and the whole city has a radiant layout, the CBD and political department are in the central point of the city while resident districts are on the outer side. Transportation such as railways help transits the labor force from the outer circle to the inside central point. This geographical characteristic makes the house price have a decreasing trend while the distance to the central point increases. In addition to this distance factor, the house price also largely relies on the transportation sight.

Based on the house market in Beijing, previous research has focused on the house attributes and environmental factors to evaluate the price of amenities and predict the house price in Beijing. Based on this result, this paper will further analyze the house price in Beijing through Rosen hedonic model with clustering into consideration. This paper is as follows: 1) Literature Review, 2) Data Description, 3) Basic hedonic model , 4) Panel Analysis using fixed effects, 5) Likelihood prediction using Logit model and Poisson Regression, 6) Conclusion.

## 3. Previous Work

Previous work related to the Beijing housing market has been conducted by multiple economists. In the Beijing house pricing paper (Xiao and others 2017)[1], they managed to run the hedonic regression over close amenities, house size, room types and numbers, and environmental factors. They found that house size, bedroom numbers, living room numbers, and district distribution positively and heavily impact the per

---

[1] Yixiong Xiao, Xiang Chen, Qiang Li, Xi Yu, Jin Chen, Jing Guo, *Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data*. International Journal of Geo-Information, 6, 358, 2017.

square meter price of the house in Beijing. They also found that the housing price is largely differentiated by district and nearby amenities.

However, their work mainly focuses on the geographical location of the houses but factors or attributes to the house itself are limited. Consumers and investors are not only concerned about the geographical benefits of the houses but also the quality of them.

More specifically, education plays an important role in China's house market. Haizhen Wen (2014) points out such high importance in his paper[2]. Others like Helbich (2013), provide the analysis of region effect for the house market[3].

Multiple econometric experiments were conducted over the same dataset from other Kaggle users. One user named Xiaoming Jim utilized the common linear regression over all variables and his result was not convincing[4]. First, he only performs regression without using a time variable (or creating a cross-sectional data in a selected year), as a result, readers cannot obtain time-related information. Second, the results of district level coefficients can be disturbing since the house prices across districts have small differences in a range of 0 to 4000 yuans. A house in Haidian or Chaoyang is worth way more than a house in any other suburban districts. The conjecture is that another strong collinearity exists in the data so the coefficients look odd.

Our goal is to find proper justifications and identifications to our regression models as much as possible and potentially extend the level of the research by using proper techniques.

## 4. Data Description

The raw data of the Beijing housing price is extracted from the Kaggle website created by Qichen Qiu[5]. It is obtained from Lianjia-a famous online real estate agency and it contains 296,273 transactions from 2011 to 2017 after dropping missing values, invalid information, scam online postings, and extreme outliers.

[2] Wen, Haizhen, et al. "Do Educational Facilities Affect Housing Price? An Empirical Study in Hangzhou, China." *Habitat International*, vol. 42, pp. 155–163., doi:10.1016/j.habitatint.2013.12.004, 2014.
[3] Helbich, Marco, et al. "Data-Driven Regionalization of Housing Markets." *Annals of the Association of American Geographers*, vol. 103, no. 4, pp. 871–889., doi:10.1080/00045608.2012.707587, 2013.
[4] Xiaoming Jim, 2021. https://www.kaggle.com/xiaomingjim/beijing-house-price-regression
[5] Qichen Qiu, 2018. https://www.kaggle.com/ruiqurm/lianjia

The paper will initially use relevant variables below to find determinants of Beijing housing market:

1. Tradeyear: time variable of the year where the transaction completed
2. Price: dependent variable of the price per square meter
3. Id: each transaction is recorded as a unique id
4. Cid: community (administered by a resident committee) id
5. District: Dongcheng (1), Fengtai (2), Tongzhou (3), Daxing (4), Fangshan (5), Changping (6), Chaoyang (7), Haidian (8), Shijinshan (9), Xicheng (10), Pinggu (11), Mentougou (12), Shunyi (13)
6. Square: the footage of the house in square meters
7. Bedroom, livingroom, kitchen, bathroom: the number of bedrooms, livingrooms, kitchens, and bathrooms for each house.
8. Buildingtype: tower (1), bungalow(2), plate and tower(3), plate (4)
9. Renovationcondition: other (1), rough (2), simplicity (3), hardcover (4)
10. Buildingstructure: unknown (1), mixed (2), brick and wood (3), brick and concrete (4), steel (5), steel-concrete (6)
11. Buildingyears: years a house has been built
12. Ladderratio: how many elevators a resident has on average on the same floor
13. Elevator: if the house has an elevator
14. Subway: if the house has a subway nearby
15. Communityaverage: average price of houses in a community

Table 1 (all tables and figures are presented in the Appendix) shows the summary statistics of the non-categorical variables. The average house price per square meter is 43960.314 yuan with a standard deviation of 21691.025. The average size for each house is 82.6 square meters with a standard deviation of 35.69. A typical house has 2 to 3 bedrooms, 1 to 2 living rooms, 1 kitchen, and 1 to 2 bathrooms. On average, 2-5 residents share an elevator on the same floor and floor number ranges from 6 to 20. The average building years is 15.634 and the community average is about 63734.972 yuan.

Then the variables listed below are categorical with the left number as the transaction frequency and the right number as the category:

1. District: 15,909 (1); 27,787 (2); 2,207 (3); 14,343 (4); 2,618 (5); 35,188 (6); 100,934 (7); 35,395 (8); 10,410 (9); 10-29,092 (10); 12,094 (11); 1,441 (12); 8,855 (13)

2. Buildingtype: 77,809 (1); 43 (2); 54,600 (3); 163,821 (4)

3. Renovationcondition: 108,464 (1); 4,871 (2); 73,573 (3); 109,365 (4)

4. Buildingstructure: 16 (1); 109,960 (2); 60 (3); 13,426 (4); 153 (5); 172,658 (6)

5. Elevator: 127,091 (0); 169,182 (1)

6. Subway: 117,736 (0); 178,537 (1)

7. Tradeyear: 5,092 (2011); 34,152 (2012); 35,408 (2013); 30,019 (2014); 64,273 (2015); 85,324 (2016); 42,005 (2017)

The next step is to construct a correlation matrix to test multicollinearity. Typically a correlation over 0.8 would consider high and cause collinearity, in Table 2 no high correlation is detected from our correlation matrix.

## 4. Method

### 4.1 Hedonic Model

The hedonic model basically assumes that the price of the house is the product of sets of attributes such as amenities, environment and other factors. The house function can be represented by $y=\alpha+\beta X+\varepsilon$, where y is the price vector, X is the characteristics metric and ß are the relative coefficient for these factors. However, based on the spatial distribution of business districts and the population change in Beijing, we noticed that there might be time fixed effects and district fixed effects, which will bring omitted variable bias without taking these effects into consideration. Thus, we further include these factors in later models.

### 4.2 Fixed Effects with Clustered Standard Errors

Based on the previous hedonic model, there might be a clustering of house prices in Beijing. The advanced school and premature business district are all clustered in specific communities. The radiation of these “central points” makes the house price decrease as the distance towards these points increases. To be specific, the house in the school district will have higher prices and apartments within a certain distance of the business district will also have higher prices.  Meanwhile, we noticed that these clusters might be subjective to common shocks such as Covid-19, stock market or environment changes.We plan to cluster all the observations for each group.

We mainly use clustered standard error to account for this common shock among districts. Adding clustered standard error will "allow for off-the-diagonal terms in the variance-covariance matrix for members of each group to allow for correlated shocks"(Nick, 2020). By applying this, the hedonic function will be

$$y=\alpha+\beta X+T+D+\varepsilon$$

Where T is the time fixed effects and D will be the district fixed effects.

### 4.3 Logit Model within Community and across District

In addition to understanding the house price function, we are also interested in the house price heterogeneity among community and district. In other words, we are interested in the factors which will influence the price variety within a single community and among all municipal districts in Beijing. We utilize an IV method. As for single community analysis, we use the IV variable to indicate the variety of house price by comparing the price per square with the average price per square within the same community. The IV would be 1 if the price of the house is higher than the average price in that community and vice versa. As for comparison across districts, we also utilize a similar IV method. The indicator would be 1 if the price of the house is higher than the average price across all districts within that single purchase month and vice versa.

Given that the IV has a binomial distribution, we thus utilize the logit model. As for within community analysis, given that the houses are all in a single community and the price is compared in a single purchase month, we omit the time effects and district effects and focus more on the house attributes. We choose renovation conditions, building construction, bathroom, kitchen, floor, building years, subway and elevator as the attributes metrics. As for district analysis, we also take the renovation condition, bathroom, kitchen, floor, building years, subway and elevator into consideration. But more importantly, we include district effects to explore the possible spatial heterogeneity.

## 5. Regression Results

### 5.1 Hedonic Analysis

Table 3 shows the hedonic regression result from our data with year and district categorical variables. The model has a R squared of 0.703 which is relatively high in comparison to many other studies. The model should be strongly valid with a F-test score of 17945.938.

Only 4 variables are insignificant and they are all from the dummy variable building structure. However, since the coefficients for these variables are extremely small in a range from 400 yuans to 800 yuans (about 0.9%-1.8% of the average price), such insignificance can hardly affect our results in later studies and in a hedonic setting, these structures are less likely to determine the house prices in Beijing. In addition, the floor variable has a weaker significance compared to other significant variables and it does have a very small coefficient of 12.334 which is marginal to the average house price.

The results for the rest of coefficients are important to our goal of research and these will be discussed next:

Surprisingly, the square variable has a negative impact in house selling-a larger room usually sells for less in price per square meters. One reason could be that Chinese urban consumers and investors usually prefer to purchase a smaller house as its aggregate price will be lower while gaining the same accessibility to public transportation, beneficial policies, educational resources, and so forth as the ones who purchase a bigger house. Therefore, the price can be diminishing when the footage of the house rises.

All the room variables have positive coefficients in hedonic settings and having a bathroom will largely increase the house price. However, this can be disturbing because more bathrooms might result in more bedrooms and living rooms (not with a kitchen because a regular house only has one kitchen no matter how many bedrooms and living rooms it has). In later panel analysis, the room variables will be altered to adjust this.

For building type and renovation conditions, the coefficients fit to the real scenario. A bungalow on average is worth 28,743 yuans more than the tower structure. The bungalow design in Beijing is called Siheyuan (an old-styled historic design) in Chinese and people treat it much more valuable than any other types of buildings. A hardcover renovation is about 2,500 yuans higher than the rough renovated one. Considering a house with 100 square meters, it creates a 250,000 yuans discrepancy, which is how much it usually takes to renovate a house in China.

Another interesting variable we capture is the buildingyears variable since the price of a house increases as the age of the house increases. Since in China, old houses will be torn down by the government and residents usually will receive large compensation in the form of new house(s) or cash money. So an older house will increase in value. Moreover, an old house usually is located in the old part of the city within the downtown area so its price will be higher.

In previous regression done by the other Kaggle user, he includes the community average variable. This will result in obvious endogeneity problems because our dependent variable is price per square meters and the average price within the community is calculated from the dependent variable-A higher price per square meters will generate a higher community average. Therefore, the coefficients for his regression model look odd and unreliable. After dropping the variable, the district categorical variables have expected coefficients and a proper ranking order where district 10 (Xicheng) has the highest average price ,district 1 (Dongcheng) has the second highest, and suburbs such as district 12 (Mentougou) and district 13 (Shunyi) have the lowest prices.

**5.2 Panel Analysis**

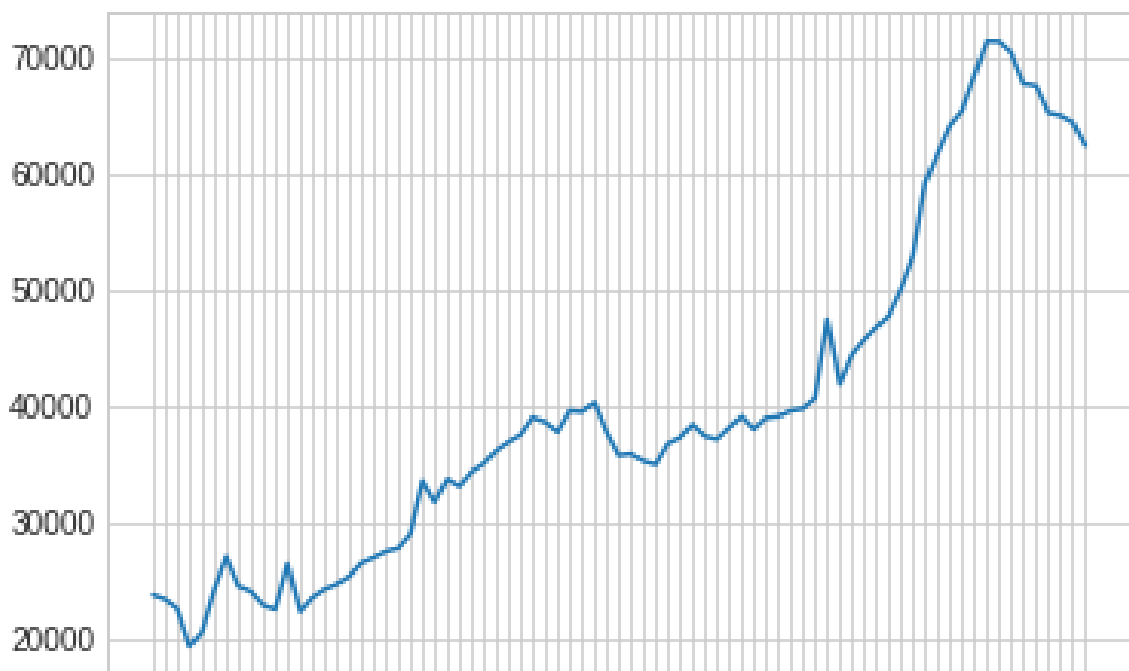


Figure 1: Monthly Beijing House Price from 2011 to 2017

From the figure above, we noticed that the house price in Beijing has been steadily increasing from 2011 to 2017 with seasonal fluctuation. Thus, it is important for us to take these trends into consideration.

Table 4 shows the fixed effects regression with seasonal effects, year effects and district effects. This model also uses clustered standard errors based on district in case of common shock possibility. As for the regression result, We notice that almost all of the results are significant and the Durbin-Watson test indicates that there might not be a significant collinearity problem.

By adding in the seasonal/month effect in the model, houses sold in the second half year are typically worth more than houses sold in the first half of the year. More support are needed to see whether the Beijing housing market has an off season and a peak season.

Another finding is that the coefficient for the bathroom has changed to weakly negative. In this setting, an additional unit bathroom only results in 217 yuans reduction in house prices. This seems odd and we need more theory support.

Moreover, the floor variable now is insignificant with a p-value of 0.107 so it is safe to say that floor numbers can hardly affect the price per square meters of a house.

Other coefficients and significant levels are consistent with the hedonic regression results.

**5.3 Likelihood Analysis**

Two likelihood regressions were performed: Table 6 shows the logit likelihood regression using community average dummy (if the house price is higher than the community average price, then the dummy is 1) as a dependent variable and Table 7 shows the logit likelihood regression using monthly average.

From Table 6, we are not interested in the district effect because the community average dummy already includes the location effect by comparing the house price to the average price from all houses nearby.

The year effect is not significant and the year increase does not affect the possibility to purchase a house. However, the seasonal effect is strong. Buyers are less likely to purchase in months of May, June, July, and August in comparison to January and they are more likely to purchase in months of September, October, November, and December. The pattern is mostly consistent with the panel seasonal effect and it is another indication there is an off-peak season in Beijing's house market.

For renovation conditions, building structure and subway, these categorical variables are insignificant in affecting a buyer's choice to purchase a house or not.

From Table 7, the district effect is consistent with our panel analysis. Houses in low averaged priced districts are less likely to be purchased and high averaged priced districts are most likely to be purchased holding other factors unchanged.

## 6. Conclusion and Future Improvements

For Beijing’s house market, the district effect or the geographical effect is the most important determinant. The social and educational benefits a house could bring are far more essential than the quality of the house for consumers and investors in Beijing real estate’s market.

In this scenario, the layout and the renovation conditions of the house in Beijing has a limited effect on the house price. In fact, bigger houses are less preferable for most buyers because they have a very constrained budget to maximize with and the marginal utility is quickly diminishing if the footage of the house continues to increase when the basic needs for living are satisfied.

We also found a strong seasonal effect in Beijing’s house market. The price off season is the first half year and the price peaking season is the second half year. So the demand will be more in the first half year and will be less in the second half year (163,351 sold in 1st half and 132,922 sold in 2nd half).

However, the district effect is highly complex and we are not able to decompose it in this project paper. For future improvement, researchers might want to analyze the relationships between non-time variables and the geographical variables in a much more proper and rigorous way. Also, different cities should have different levels of district effect. For example, more urbanized cities may have a stronger district effect than others and in less-urbanized cities, people focus more on the quality of houses instead of other benefits.

## 7. References

1.Yixiong Xiao, Xiang Chen, Qiang Li, Xi Yu, Jin Chen, Jing Guo, *Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data*. International Journal of Geo-Information, 6, 358, 2017.

2.Wen, Haizhen, et al. “Do Educational Facilities Affect Housing Price? An Empirical Study in Hangzhou, China.” *Habitat International*, vol. 42, pp. 155–163., doi:10.1016/j.habitatint.2013.12.004, 2014

3. Helbich, Marco, et al. “Data-Driven Regionalization of Housing Markets.” *Annals of the Association of American Geographers*, vol. 103, no. 4, pp. 871–889., doi:10.1080/00045608.2012.707587, 2013.

4. Xiaoming Jim, https://www.kaggle.com/xiaomingjim/beijing-house-price-regression, 2021

5. Qichen Qiu,  https://www.kaggle.com/ruiqurm/lianjia, 2018

6. Wooldridge, Jeffrey M. *Introductory Econometrics: a Modern Approach*. Cengage Learning, 2020.

# 8. Appendix

**Descriptive Statistics**

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| price | 296273 | 43960.314 | 21691.025 | 5000 | 150000 |
| square | 296273 | 82.6 | 35.69 | 7.37 | 640 |
| bedroom | 296273 | 2.01 | .765 | 0 | 8 |
| livingroom | 296273 | 1.171 | .512 | 0 | 5 |
| kitchen | 296273 | .995 | .099 | 0 | 3 |
| bathroom | 296273 | 1.182 | .422 | 0 | 7 |
| ladderratio | 296273 | .382 | .176 | .014 | 5 |
| floor | 296273 | 13.169 | 7.762 | 1 | 63 |
| buildingyears | 296273 | 15.634 | 8.852 | 0 | 67 |
| communityaverage | 296273 | 63734.972 | 22134.957 | 10847 | 183109 |

Table 1: Summary Statistics of the Cleaned Data

**Matrix of correlations**

| Variables | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) square | 1.000 | | | | | | | | | | | | | | | |
| (2) bedroom | 0.720 | 1.000 | | | | | | | | | | | | | | |
| (3) livingroom | 0.617 | 0.476 | 1.000 | | | | | | | | | | | | | |
| (4) kitchen | 0.083 | 0.095 | 0.126 | 1.000 | | | | | | | | | | | | |
| (5) bathroom | 0.732 | 0.538 | 0.521 | 0.106 | 1.000 | | | | | | | | | | | |
| (6) buildingtype | -0.015 | 0.118 | 0.074 | 0.021 | 0.030 | 1.000 | | | | | | | | | | |
| (7) renovationcond~n | 0.035 | 0.010 | -0.009 | -0.005 | 0.027 | -0.015 | 1.000 | | | | | | | | | |
| (8) buildingstruct~e | 0.166 | -0.062 | 0.058 | -0.011 | 0.121 | -0.538 | 0.050 | 1.000 | | | | | | | | |
| (9) ladderratio | 0.373 | 0.308 | 0.300 | 0.047 | 0.259 | 0.386 | 0.010 | -0.135 | 1.000 | | | | | | | |
| (10) subway | -0.095 | -0.060 | -0.094 | 0.011 | -0.061 | -0.147 | 0.005 | 0.094 | -0.106 | 1.000 | | | | | | |
| (11) district | -0.006 | 0.019 | 0.012 | -0.034 | 0.000 | 0.053 | -0.011 | -0.111 | 0.008 | -0.107 | 1.000 | | | | | |
| (12) floor | 0.162 | -0.085 | 0.035 | -0.018 | 0.073 | -0.735 | 0.048 | 0.680 | -0.141 | 0.121 | -0.073 | 1.000 | | | | |
| (13) elevator | 0.193 | -0.067 | 0.056 | -0.011 | 0.133 | -0.633 | 0.055 | 0.818 | -0.148 | 0.104 | -0.077 | 0.782 | 1.000 | | | |
| (14) buildingyears | -0.363 | 0.005 | -0.222 | 0.024 | -0.253 | 0.147 | 0.055 | -0.446 | -0.148 | 0.160 | 0.079 | -0.405 | -0.460 | 1.000 | | |
| (15) tradeyear | 0.004 | 0.010 | -0.067 | 0.012 | 0.005 | 0.007 | 0.649 | 0.020 | 0.005 | -0.010 | -0.037 | 0.027 | 0.020 | 0.144 | 1.000 | |
| (16) communityaver~e | -0.148 | -0.051 | -0.095 | 0.027 | -0.066 | -0.093 | 0.009 | 0.067 | -0.112 | 0.313 | 0.094 | 0.047 | 0.081 | 0.279 | -0.017 | 1.000 |

Table 2: Correlation Matrix for Hedonic Regression

**Linear regression**

| price | Coef. | St. Err. | t-value | p-value | [95% Conf | Interval] | Sig |
|---|---|---|---|---|---|---|---|
| square | -87.774 | 1.322 | -66.37 | 0 | -90.366 | -85.182 | *** |
| bedroom | 626.77 | 46.01 | 13.62 | 0 | 536.592 | 716.948 | *** |
| livingroom | 1873.48 | 55.543 | 33.73 | 0 | 1764.617 | 1982.342 | *** |
| kitchen | 1697.051 | 229.787 | 7.39 | 0 | 1246.675 | 2147.427 | *** |
| bathroom | 2553.915 | 77.001 | 33.17 | 0 | 2402.995 | 2704.836 | *** |
| 2.buildingtype | 28743.087 | 2434.39 | 11.81 | 0 | 23971.75 | 33514.423 | *** |
| 3.buildingtype | 2137.522 | 74.678 | 28.62 | 0 | 1991.155 | 2283.889 | *** |
| 4.buildingtype | 3498.335 | 91.822 | 38.10 | 0 | 3318.367 | 3678.304 | *** |
| 2.renovation~tion | -1150.792 | 181.95 | -6.32 | 0 | -1507.408 | -794.175 | *** |
| 3.renovation~tion | -1028.074 | 79.303 | -12.96 | 0 | -1183.505 | -872.643 | *** |
| 4.renovation~tion | 1376.953 | 74.683 | 18.44 | 0 | 1230.576 | 1523.33 | *** |
| 2.buildingstruct~e | -522.907 | 2959.71 | -0.18 | .86 | -6323.855 | 5278.041 | |
| 3.buildingstruct~e | 18396.41 | 3603.969 | 5.10 | 0 | 11332.73 | 25460.089 | *** |
| 4.buildingstruct~e | -510.487 | 2961.345 | -0.17 | .863 | -6314.64 | 5293.665 | |
| 5.buildingstruct~e | -782.094 | 3110.377 | -0.25 | .801 | -6878.347 | 5314.158 | |
| 6.buildingstruct~e | 407.385 | 2960.537 | 0.14 | .891 | -5395.183 | 6209.954 | |
| ladderratio | 5337.521 | 149.694 | 35.66 | 0 | 5044.126 | 5630.916 | *** |
| 1.elevator | 3906.182 | 98.959 | 39.47 | 0 | 3712.225 | 4100.14 | *** |
| 1.subway | 3874.686 | 48.042 | 80.65 | 0 | 3780.525 | 3968.846 | *** |
| 1b.district | 0 | . | . | . | . | . | |
| 2.district | -21444.228 | 120.245 | -178.34 | 0 | -21679.906 | -21208.55 | *** |
| 3.district | -30401.137 | 271.179 | -112.11 | 0 | -30932.64 | -29869.634 | *** |
| 4.district | -29122.549 | 141.149 | -206.32 | 0 | -29399.197 | -28845.901 | *** |
| 5.district | -40480 | 255.286 | -158.57 | 0 | -40980.354 | -39979.647 | *** |
| 6.district | -29403.813 | 121.057 | -242.89 | 0 | -29641.081 | -29166.546 | *** |
| 7.district | -17082.571 | 103.23 | -165.48 | 0 | -17284.899 | -16880.243 | *** |
| 8.district | -4195.97 | 114.743 | -36.57 | 0 | -4420.863 | -3971.077 | *** |
| 9.district | -22971.15 | 153.473 | -149.68 | 0 | -23271.953 | -22670.347 | *** |
| 10.district | 7612.092 | 118.42 | 64.28 | 0 | 7379.992 | 7844.192 | *** |
| 11.district | -29235.277 | 147.89 | -197.68 | 0 | -29525.138 | -28945.416 | *** |
| 12.district | -30823.533 | 340.672 | -90.48 | 0 | -31491.241 | -30155.824 | *** |
| 13.district | -32335.799 | 162.545 | -198.93 | 0 | -32654.382 | -32017.216 | *** |
| floor | 12.334 | 5.563 | 2.22 | .027 | 1.43 | 23.238 | ** |
| buildingyears | 156.426 | 3.713 | 42.13 | 0 | 149.148 | 163.703 | *** |
| 2012.tradeyear | 2559.252 | 177.847 | 14.39 | 0 | 2210.677 | 2907.826 | *** |
| 2013.tradeyear | 12386.895 | 177.55 | 69.77 | 0 | 12038.902 | 12734.888 | *** |
| 2014.tradeyear | 12040.15 | 185.526 | 64.90 | 0 | 11676.524 | 12403.776 | *** |
| 2015.tradeyear | 14114.803 | 183.434 | 76.95 | 0 | 13755.277 | 14474.329 | *** |
| 2016.tradeyear | 27404.295 | 182.096 | 150.49 | 0 | 27047.393 | 27761.198 | *** |
| 2017.tradeyear | 44002.442 | 187.468 | 234.72 | 0 | 43635.01 | 44369.874 | *** |
| Constant | 27248.74 | 2974.544 | 9.16 | 0 | 21418.716 | 33078.763 | *** |

| | | | |
|---|---|---|---|
| Mean dependent var | 43960.314 | SD dependent var | 21691.025 |
| R-squared | 0.703 | Number of obs | 296273.000 |
| F-test | 17945.938 | Prob > F | 0.000 |
| Akaike crit. (AIC) | 6397934.088 | Bayesian crit. (BIC) | 6398358.049 |

**** p<.01, ** p<.05, * p<.1*

Table 3: Hedonic Linear Regression

| | | | |
|---|---|---|---|
| **Dep. Variable:** | price | **R-squared:** | 0.705 |
| **Model:** | OLS | **Adj. R-squared:** | 0.705 |
| **Method:** | Least Squares | **F-statistic:** | 1.572e+04 |
| **Date:** | Fri, 30 Apr 2021 | **Prob (F-statistic):** | 0.00 |
| **Time:** | 07:40:05 | **Log-Likelihood:** | -3.1978e+06 |
| **No. Observations:** | 296273 | **AIC:** | 6.396e+06 |
| **Df Residuals:** | 296227 | **BIC:** | 6.396e+06 |
| **Df Model:** | 45 | | |

| | **coef** | **std err** | **t** | **P> \|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | -5663.6706 | 2962.034 | -1.912 | 0.056 | -1.15e+04 | 141.833 |
| **C(buildingtype)[T.2]** | 2.889e+04 | 2424.722 | 11.915 | 0.000 | 2.41e+04 | 3.36e+04 |
| **C(buildingtype)[T.3]** | 2950.9340 | 71.772 | 41.115 | 0.000 | 2810.263 | 3091.605 |
| **C(buildingtype)[T.4]** | 4576.6002 | 85.509 | 53.522 | 0.000 | 4409.006 | 4744.195 |
| **C(renovationcondition)[T.2]** | -1419.9260 | 181.247 | -7.834 | 0.000 | -1775.165 | -1064.687 |
| **C(renovationcondition)[T.3]** | -1430.7674 | 79.253 | -18.053 | 0.000 | -1586.100 | -1275.435 |
| **C(renovationcondition)[T.4]** | 986.9710 | 74.640 | 13.223 | 0.000 | 840.679 | 1133.263 |
| **C(buildingstructure)[T.2]** | -771.6748 | 2948.513 | -0.262 | 0.794 | -6550.679 | 5007.329 |
| **C(buildingstructure)[T.3]** | 1.777e+04 | 3590.291 | 4.949 | 0.000 | 1.07e+04 | 2.48e+04 |
| **C(buildingstructure)[T.4]** | -783.3859 | 2950.138 | -0.266 | 0.791 | -6565.573 | 4998.801 |
| **C(buildingstructure)[T.5]** | -1101.9577 | 3098.623 | -0.356 | 0.722 | -7175.171 | 4971.256 |
| **C(buildingstructure)[T.6]** | 122.5113 | 2949.319 | 0.042 | 0.967 | -5658.071 | 5903.093 |
| **C(elevator)[T.1]** | 3640.0840 | 98.460 | 36.970 | 0.000 | 3447.104 | 3833.064 |
| **C(subway)[T.1]** | 3901.8581 | 47.824 | 81.587 | 0.000 | 3808.124 | 3995.592 |
| **C(trademonth)[T.2]** | -1310.0288 | 99.361 | -13.184 | 0.000 | -1504.774 | -1115.283 |
| **C(trademonth)[T.3]** | 171.0861 | 84.351 | 2.028 | 0.043 | 5.761 | 336.411 |
| **C(trademonth)[T.4]** | 207.7027 | 101.796 | 2.040 | 0.041 | 8.186 | 407.219 |
| **C(trademonth)[T.5]** | 434.7980 | 99.921 | 4.351 | 0.000 | 238.956 | 630.640 |
| **C(trademonth)[T.6]** | 584.3963 | 97.346 | 6.003 | 0.000 | 393.600 | 775.193 |
| **C(trademonth)[T.7]** | 1326.3326 | 93.093 | 14.247 | 0.000 | 1143.873 | 1508.792 |
| **C(trademonth)[T.8]** | 2444.5023 | 89.194 | 27.407 | 0.000 | 2269.685 | 2619.319 |
| **C(trademonth)[T.9]** | 4535.1873 | 91.478 | 49.577 | 0.000 | 4355.892 | 4714.482 |
| **C(trademonth)[T.11]** | 3938.8710 | 95.466 | 41.259 | 0.000 | 3751.761 | 4125.981 |
| **C(trademonth)[T.12]** | 5071.8695 | 88.971 | 57.006 | 0.000 | 4897.489 | 5246.251 |
| **C(tradeyear)[T.2012]** | 3996.0003 | 178.490 | 22.388 | 0.000 | 3646.165 | 4345.836 |
| **C(tradeyear)[T.2013]** | 1.424e+04 | 178.760 | 79.647 | 0.000 | 1.39e+04 | 1.46e+04 |
| **C(tradeyear)[T.2014]** | 1.377e+04 | 186.100 | 73.972 | 0.000 | 1.34e+04 | 1.41e+04 |
| **C(tradeyear)[T.2015]** | 1.566e+04 | 184.354 | 84.957 | 0.000 | 1.53e+04 | 1.6e+04 |
| **C(tradeyear)[T.2016]** | 2.916e+04 | 183.663 | 158.779 | 0.000 | 2.88e+04 | 2.95e+04 |
| **C(tradeyear)[T.2017]** | 4.616e+04 | 189.450 | 243.651 | 0.000 | 4.58e+04 | 4.65e+04 |
| **C(districtname)[T.Chaoyang]** | 1.298e+04 | 80.895 | 160.431 | 0.000 | 1.28e+04 | 1.31e+04 |
| **C(districtname)[T.Daxing]** | 1150.6509 | 117.887 | 9.761 | 0.000 | 919.596 | 1381.706 |
| **C(districtname)[T.Dongcheng]** | 2.995e+04 | 120.286 | 248.964 | 0.000 | 2.97e+04 | 3.02e+04 |
| **C(districtname)[T.Fangshan]** | -1.087e+04 | 241.763 | -44.943 | 0.000 | -1.13e+04 | -1.04e+04 |
| **C(districtname)[T.Fengtai]** | 8873.2720 | 99.873 | 88.845 | 0.000 | 8677.523 | 9069.021 |
| **C(districtname)[T.Haidian]** | 2.579e+04 | 94.028 | 274.288 | 0.000 | 2.56e+04 | 2.6e+04 |
| **C(districtname)[T.Mentougou]** | -855.8327 | 328.693 | -2.604 | 0.009 | -1500.061 | -211.605 |
| **C(districtname)[T.Pinggu]** | 1051.8500 | 125.681 | 8.369 | 0.000 | 805.519 | 1298.181 |
| **C(districtname)[T.Shijingshan]** | 7224.6528 | 137.240 | 52.643 | 0.000 | 6955.667 | 7493.639 |
| **C(districtname)[T.Shunyi]** | -2444.5780 | 140.829 | -17.359 | 0.000 | -2720.598 | -2168.558 |
| **C(districtname)[T.Tongzhou]** | -240.9380 | 259.685 | -0.928 | 0.354 | -749.914 | 268.038 |
| **C(districtname)[T.Xicheng]** | 3.769e+04 | 101.733 | 370.522 | 0.000 | 3.75e+04 | 3.79e+04 |
| **kitchen** | 2160.6813 | 227.915 | 9.480 | 0.000 | 1713.974 | 2607.388 |
| **bathroom** | -217.1155 | 54.369 | -3.993 | 0.000 | -323.678 | -110.553 |
| **buildingyears** | 196.3260 | 3.497 | 56.134 | 0.000 | 189.471 | 203.181 |
| **floor** | 8.7854 | 5.457 | 1.610 | 0.107 | -1.910 | 19.481 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 47816.214 | **Durbin-Watson:** | 1.724 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 165052.045 |
| **Skew:** | 0.808 | **Prob(JB):** | 0.00 |
| **Kurtosis:** | 6.280 | **Cond. No.** | 7.27e+03 |

1

Table 4: Panel Model Regression with Seasonal Effect with Clustered Standard Errors

| | | | |
|---|---|---|---|
| **Dep. Variable:** | price | **R-squared:** | 0.727 |
| **Model:** | OLS | **Adj. R-squared:** | 0.727 |
| **Method:** | Least Squares | **F-statistic:** | 1.161e+13 |
| **Date:** | Fri, 30 Apr 2021 | **Prob (F-statistic):** | 1.50e-76 |
| **Time:** | 07:49:39 | **Log-Likelihood:** | -3.1862e+06 |
| **No. Observations:** | 296273 | **AIC:** | 6.373e+06 |
| **Df Residuals:** | 296167 | **BIC:** | 6.374e+06 |
| **Df Model:** | 105 | | |

| | **coef** | **std err** | **t** | **P> \|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | -1379.8868 | 4742.427 | -0.291 | 0.776 | -1.17e+04 | 8952.974 |
| **C(buildingtype)[T.2]** | 2.984e+04 | 7093.143 | 4.207 | 0.001 | 1.44e+04 | 4.53e+04 |
| **C(buildingtype)[T.3]** | 3003.8085 | 551.498 | 5.447 | 0.000 | 1802.197 | 4205.420 |
| **C(buildingtype)[T.4]** | 4650.7917 | 1360.370 | 3.419 | 0.005 | 1686.801 | 7614.783 |
| **C(renovationcondition)[T.2]** | -618.1322 | 633.480 | -0.976 | 0.348 | -1998.366 | 762.102 |
| **C(renovationcondition)[T.3]** | -698.6019 | 302.094 | -2.313 | 0.039 | -1356.809 | -40.395 |
| **C(renovationcondition)[T.4]** | 1699.9276 | 213.699 | 7.955 | 0.000 | 1234.318 | 2165.537 |
| **C(buildingstructure)[T.2]** | -1732.8994 | 2243.146 | -0.773 | 0.455 | -6620.294 | 3154.495 |
| **C(buildingstructure)[T.3]** | 1.697e+04 | 7734.210 | 2.194 | 0.049 | 119.025 | 3.38e+04 |
| **C(buildingstructure)[T.4]** | -1893.3587 | 2082.971 | -0.909 | 0.381 | -6431.763 | 2645.046 |
| **C(buildingstructure)[T.5]** | -2451.7387 | 2638.915 | -0.929 | 0.371 | -8201.442 | 3297.964 |
| **C(buildingstructure)[T.6]** | -834.0126 | 1871.045 | -0.446 | 0.664 | -4910.669 | 3242.644 |
| **C(elevator)[T.1]** | 3608.0346 | 1122.624 | 3.214 | 0.007 | 1162.048 | 6054.022 |
| **C(subway)[T.1]** | 3893.3218 | 667.870 | 5.829 | 0.000 | 2438.158 | 5348.485 |
| **C(districtname)[T.Chaoyang]** | 1.3e+04 | 685.965 | 18.951 | 0.000 | 1.15e+04 | 1.45e+04 |
| **C(districtname)[T.Daxing]** | 1159.2161 | 328.099 | 3.533 | 0.004 | 444.349 | 1874.083 |
| **C(districtname)[T.Dongcheng]** | 2.993e+04 | 652.687 | 45.860 | 0.000 | 2.85e+04 | 3.14e+04 |
| **C(districtname)[T.Fangshan]** | -1.112e+04 | 557.199 | -19.961 | 0.000 | -1.23e+04 | -9908.286 |
| **C(districtname)[T.Fengtai]** | 8947.8351 | 581.383 | 15.391 | 0.000 | 7681.111 | 1.02e+04 |
| **C(districtname)[T.Haidian]** | 2.579e+04 | 544.371 | 47.377 | 0.000 | 2.46e+04 | 2.7e+04 |
| **C(districtname)[T.Mentougou]** | -1238.1853 | 760.799 | -1.627 | 0.130 | -2895.824 | 419.453 |
| **C(districtname)[T.Pinggu]** | 1013.6971 | 597.242 | 1.697 | 0.115 | -287.581 | 2314.975 |
| **C(districtname)[T.Shijingshan]** | 7169.3809 | 605.541 | 11.840 | 0.000 | 5850.021 | 8488.741 |
| **C(districtname)[T.Shunyi]** | -2453.4315 | 352.450 | -6.961 | 0.000 | -3221.353 | -1685.510 |
| **C(districtname)[T.Tongzhou]** | -355.2230 | 180.979 | -1.963 | 0.073 | -749.543 | 39.097 |
| **C(districtname)[T.Xicheng]** | 3.765e+04 | 774.145 | 48.636 | 0.000 | 3.6e+04 | 3.93e+04 |
| **kitchen** | 2459.4915 | 1964.689 | 1.252 | 0.234 | -1821.198 | 6740.181 |
| **bathroom** | -182.3910 | 1116.223 | -0.163 | 0.873 | -2614.432 | 2249.650 |
| **buildingyears** | 197.6180 | 69.363 | 2.849 | 0.015 | 46.489 | 348.747 |
| **floor** | 12.4463 | 123.039 | 0.101 | 0.921 | -255.633 | 280.526 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 49945.684 | **Durbin-Watson:** | 1.822 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 190447.754 |
| **Skew:** | 0.813 | **Prob(JB):** | 0.00 |
| **Kurtosis:** | 6.575 | **Cond. No.** | 7.27e+03 |

Table 5: Panel Model Regression with Time Effect with Clustered Standard Errors

| | | | |
|---|---|---|---|
| **Dep. Variable:** | abovecommunity | **No. Observations:** | 296273 |
| **Model:** | GLM | **Df Residuals:** | 296239 |
| **Model Family:** | Binomial | **Df Model:** | 33 |
| **Link Function:** | logit | **Scale:** | 1.0000 |
| **Method:** | IRLS | **Log-Likelihood:** | -51306. |
| **Date:** | Fri, 30 Apr 2021 | **Deviance:** | 1.0261e+05 |
| **Time:** | 07:57:40 | **Pearson chi2:** | 4.35e+05 |
| **No. Iterations:** | 11 | | |

| | **coef** | **std err** | **z** | **P>\|z\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | -10.6320 | 1.652 | -6.435 | 0.000 | -13.870 | -7.394 |
| **C(buildingtype)[T.2]** | 2.1042 | 0.719 | 2.925 | 0.003 | 0.694 | 3.514 |
| **C(buildingtype)[T.3]** | 0.0200 | 0.026 | 0.780 | 0.436 | -0.030 | 0.070 |
| **C(buildingtype)[T.4]** | 0.1686 | 0.029 | 5.773 | 0.000 | 0.111 | 0.226 |
| **C(renovationcondition)[T.2]** | -0.0951 | 0.063 | -1.512 | 0.131 | -0.218 | 0.028 |
| **C(renovationcondition)[T.3]** | 0.0122 | 0.027 | 0.454 | 0.650 | -0.041 | 0.065 |
| **C(renovationcondition)[T.4]** | 0.2732 | 0.026 | 10.632 | 0.000 | 0.223 | 0.324 |
| **C(buildingstructure)[T.2]** | 1.5839 | 1.312 | 1.207 | 0.227 | -0.987 | 4.155 |
| **C(buildingstructure)[T.3]** | 2.7289 | 1.431 | 1.907 | 0.057 | -0.076 | 5.533 |
| **C(buildingstructure)[T.4]** | 1.6949 | 1.312 | 1.291 | 0.197 | -0.877 | 4.267 |
| **C(buildingstructure)[T.5]** | 0.8610 | 1.353 | 0.636 | 0.525 | -1.791 | 3.513 |
| **C(buildingstructure)[T.6]** | 1.7088 | 1.312 | 1.302 | 0.193 | -0.863 | 4.281 |
| **C(elevator)[T.1]** | 0.0920 | 0.035 | 2.604 | 0.009 | 0.023 | 0.161 |
| **C(subway)[T.1]** | 0.0259 | 0.017 | 1.558 | 0.119 | -0.007 | 0.058 |
| **C(trademonth)[T.2]** | 0.1355 | 0.031 | 4.350 | 0.000 | 0.074 | 0.197 |
| **C(trademonth)[T.3]** | 0.2874 | 0.029 | 9.921 | 0.000 | 0.231 | 0.344 |
| **C(trademonth)[T.4]** | 0.0183 | 0.041 | 0.452 | 0.651 | -0.061 | 0.098 |
| **C(trademonth)[T.5]** | -0.2846 | 0.043 | -6.617 | 0.000 | -0.369 | -0.200 |
| **C(trademonth)[T.6]** | -0.4109 | 0.042 | -9.777 | 0.000 | -0.493 | -0.329 |
| **C(trademonth)[T.7]** | -0.5248 | 0.038 | -13.961 | 0.000 | -0.598 | -0.451 |
| **C(trademonth)[T.8]** | -0.5030 | 0.035 | -14.219 | 0.000 | -0.572 | -0.434 |
| **C(trademonth)[T.9]** | 0.7067 | 0.032 | 22.074 | 0.000 | 0.644 | 0.769 |
| **C(trademonth)[T.11]** | 1.0142 | 0.038 | 26.617 | 0.000 | 0.940 | 1.089 |
| **C(trademonth)[T.12]** | 1.5618 | 0.034 | 46.250 | 0.000 | 1.496 | 1.628 |
| **C(tradeyear)[T.2012]** | 0.4381 | 1.069 | 0.410 | 0.682 | -1.658 | 2.534 |
| **C(tradeyear)[T.2013]** | 2.4914 | 1.011 | 2.464 | 0.014 | 0.510 | 4.473 |
| **C(tradeyear)[T.2014]** | 1.9793 | 1.017 | 1.946 | 0.052 | -0.014 | 3.972 |
| **C(tradeyear)[T.2015]** | 2.0204 | 1.007 | 2.006 | 0.045 | 0.046 | 3.995 |
| **C(tradeyear)[T.2016]** | 7.0711 | 1.001 | 7.068 | 0.000 | 5.110 | 9.032 |
| **C(tradeyear)[T.2017]** | 10.3939 | 1.001 | 10.388 | 0.000 | 8.433 | 12.355 |
| **kitchen** | 0.6877 | 0.086 | 7.983 | 0.000 | 0.519 | 0.857 |
| **bathroom** | -0.4015 | 0.024 | -16.536 | 0.000 | -0.449 | -0.354 |
| **bedroom** | -0.3608 | 0.013 | -28.207 | 0.000 | -0.386 | -0.336 |
| **floor** | -0.0166 | 0.002 | -9.022 | 0.000 | -0.020 | -0.013 |

Table 6: The likelihood of being high price within a community

| | | | |
|---|---|---|---|
| **Dep. Variable:** | abovetime | **No. Observations:** | 296273 |
| **Model:** | GLM | **Df Residuals:** | 296251 |
| **Model Family:** | Binomial | **Df Model:** | 21 |
| **Link Function:** | logit | **Scale:** | 1.0000 |
| **Method:** | IRLS | **Log-Likelihood:** | -1.1825e+05 |
| **Date:** | Fri, 30 Apr 2021 | **Deviance:** | 2.3650e+05 |
| **Time:** | 08:13:02 | **Pearson chi2:** | 3.06e+05 |
| **No. Iterations:** | 11 | | |

| | **coef** | **std err** | **z** | **P>\|z\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | -5.5738 | 0.075 | -74.181 | 0.000 | -5.721 | -5.427 |
| **C(renovationcondition)[T.2]** | -0.5388 | 0.045 | -11.935 | 0.000 | -0.627 | -0.450 |
| **C(renovationcondition)[T.3]** | -0.6027 | 0.014 | -43.335 | 0.000 | -0.630 | -0.575 |
| **C(renovationcondition)[T.4]** | -0.1026 | 0.012 | -8.638 | 0.000 | -0.126 | -0.079 |
| **C(elevator)[T.1]** | 0.6481 | 0.018 | 36.000 | 0.000 | 0.613 | 0.683 |
| **C(subway)[T.1]** | 0.7166 | 0.011 | 63.595 | 0.000 | 0.694 | 0.739 |
| **C(districtname)[T.Chaoyang]** | 3.7106 | 0.045 | 83.267 | 0.000 | 3.623 | 3.798 |
| **C(districtname)[T.Daxing]** | 0.8981 | 0.063 | 14.233 | 0.000 | 0.774 | 1.022 |
| **C(districtname)[T.Dongcheng]** | 5.9270 | 0.052 | 114.737 | 0.000 | 5.826 | 6.028 |
| **C(districtname)[T.Fangshan]** | -3.6070 | 1.001 | -3.603 | 0.000 | -5.569 | -1.645 |
| **C(districtname)[T.Fengtai]** | 2.7490 | 0.046 | 59.220 | 0.000 | 2.658 | 2.840 |
| **C(districtname)[T.Haidian]** | 5.4556 | 0.046 | 118.313 | 0.000 | 5.365 | 5.546 |
| **C(districtname)[T.Mentougou]** | -0.8986 | 0.451 | -1.994 | 0.046 | -1.782 | -0.016 |
| **C(districtname)[T.Pinggu]** | 0.2720 | 0.080 | 3.387 | 0.001 | 0.115 | 0.429 |
| **C(districtname)[T.Shijingshan]** | 2.2766 | 0.053 | 43.096 | 0.000 | 2.173 | 2.380 |
| **C(districtname)[T.Shunyi]** | -1.2998 | 0.175 | -7.427 | 0.000 | -1.643 | -0.957 |
| **C(districtname)[T.Tongzhou]** | 0.0346 | 0.168 | 0.206 | 0.837 | -0.294 | 0.363 |
| **C(districtname)[T.Xicheng]** | 6.9031 | 0.053 | 131.301 | 0.000 | 6.800 | 7.006 |
| **kitchen** | 0.3611 | 0.058 | 6.196 | 0.000 | 0.247 | 0.475 |
| **bathroom** | 0.1778 | 0.013 | 13.864 | 0.000 | 0.153 | 0.203 |
| **floor** | -0.0146 | 0.001 | -14.318 | 0.000 | -0.017 | -0.013 |
| **buildingyears** | 0.0387 | 0.001 | 51.272 | 0.000 | 0.037 | 0.040 |

Table 7: The likelihood of being high price within Beijing during each month