Uncovering the Drivers of Housing Prices in Beijing: The Influence of Location and Time

Prabhjot Singh, Yunxuan Zhang, Chenyi Zhao, Yelia Ye

Table of contents

Aim

In this study, we investigate the primary factors that influence the cost of homes in Beijing. By analyzing data from Lianjia.com, we explore how the location of a property and the timing of its sale affect its price. This research aims to shed light on the complex dynamics of Beijing's real estate market and provide a clearer picture for individuals looking to understand the value of real estate in this bustling metropolis.

Data

Housing price of Beijing from the Kaggle: Housing price of Beijing from 2011 to 2017

Method

The R programming language (R Core Team, 2019) and the following R packages were used to perform the analysis: knitr (Xie 2014), tidyverse (Wickham 2017), and Quarto (Allaire et al 2022).

Data Analysis

The dataset from Lianjia.com was loaded and inspected for structure and summary statistics. Initial data exploration included reviewing the distributions of key variables such as square footage and price.

Data cleaning processes involved removing irrelevant columns, converting character variables to their appropriate types, and handling missing values. We also removed variables with non-ASCII characters to streamline the dataset for analysis.

Further exploratory data analysis revealed insights into the relationships between various features of the properties and their prices. For example, histograms were used to visualize the distribution of total prices, while scatter plots helped in understanding the relationship between the square footage of properties and their total prices.

Correlation matrices were computed and visualized to identify potential linear relationships between numerical variables, informing subsequent modeling choices.

Visualisation

Upon curating a refined dataset, we embarked on exploratory data analysis to unravel underlying patterns and associations. Figure 1 showcased a histogram of total prices, indicating a right-skewed distribution with a marked concentration of lower-priced properties—a phenomenon underscored by the mean or median price depicted through a central red line. This visualization was pivotal in discerning the varied housing market range within Beijing.

In Figure 2, we computed and visualized a correlation matrix to highlight potential linear relationships, with circle sizes corresponding to the correlation strength. Notably, the matrix revealed a prominent positive correlation between property size and total price, suggesting a direct relationship between square footage and market value.

The value of property features was illustrated in Figure 3 and Figure 4 through box plots comparing total prices based on the presence of subways and elevators. The premium for these features was clearly visible, reflecting the higher prices consumers are willing to pay for added convenience.

Figure 5 delved into the impact of renovation conditions on property values. Through the variance illustrated in the box plots for each condition category, we observed a correlation between renovation quality and pricing, indicating that buyers are inclined to invest more in well-maintained properties.

The market's temporal dynamics were captured in Figure 6, a time series plot that traced the trajectory of average monthly residential sale prices, with the volume of transactions visually represented by the size of the points. This plot served as an indicator of market trends and economic factors affecting housing prices over time.

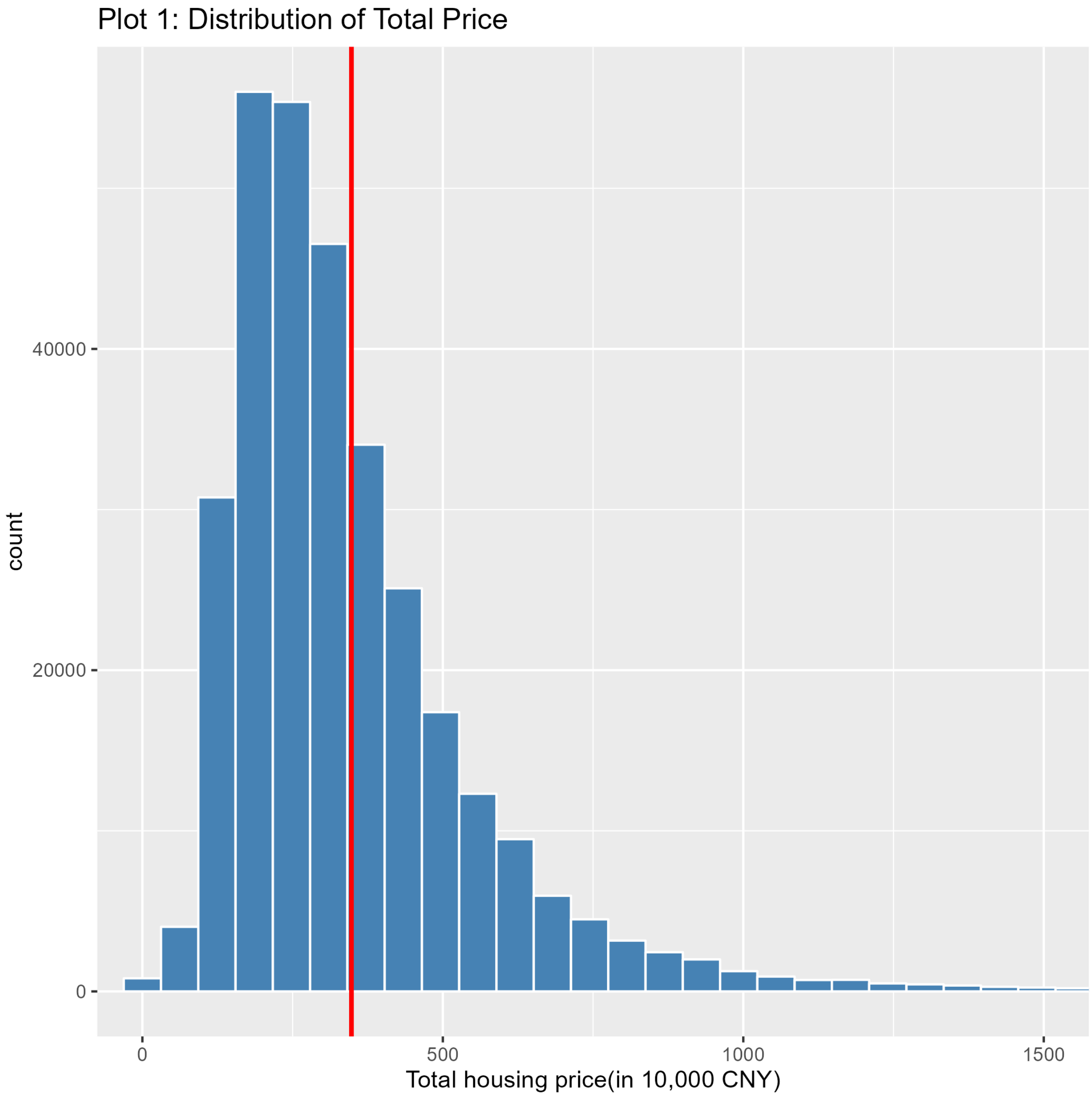

Figure 1: Housing Price of Beijing from 2011 to 2017

Figure 2: Correlation Between Variables

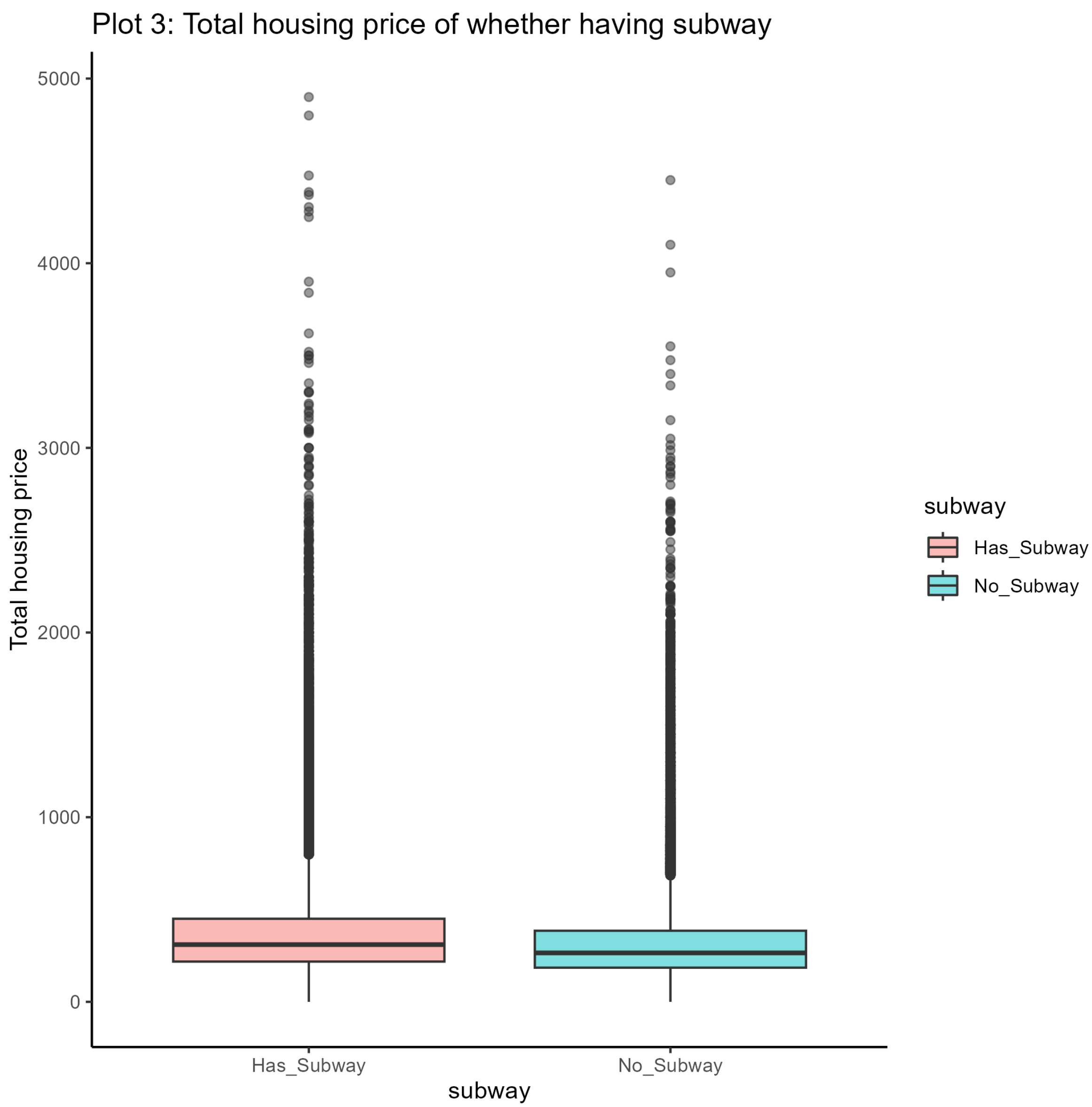

Figure 3: Total Housing Price of Whether Having Subway

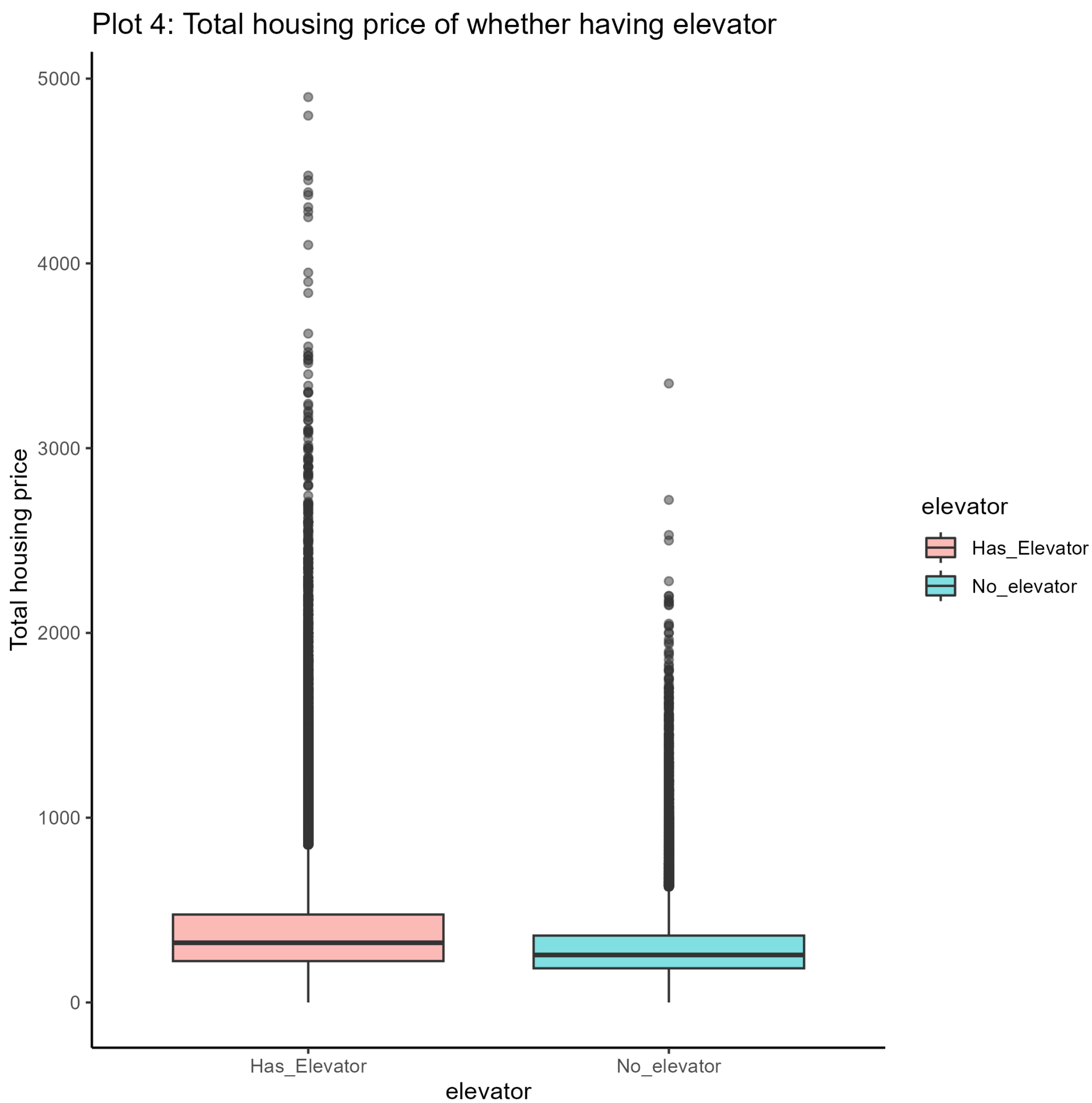

Figure 4: Total housing Price of Whether Having Elevator

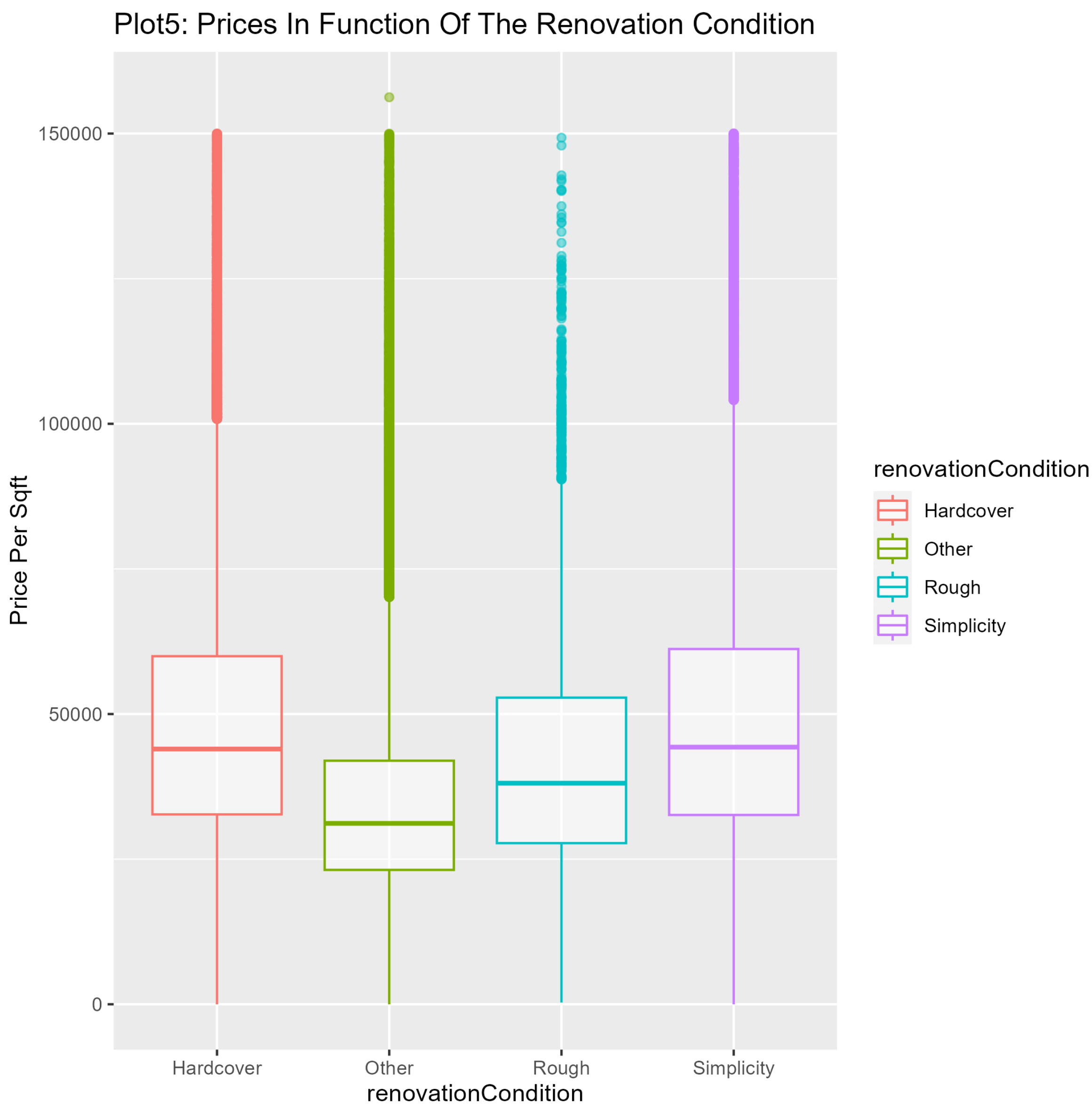

Figure 5: Prices in Function of The Renovation Condition

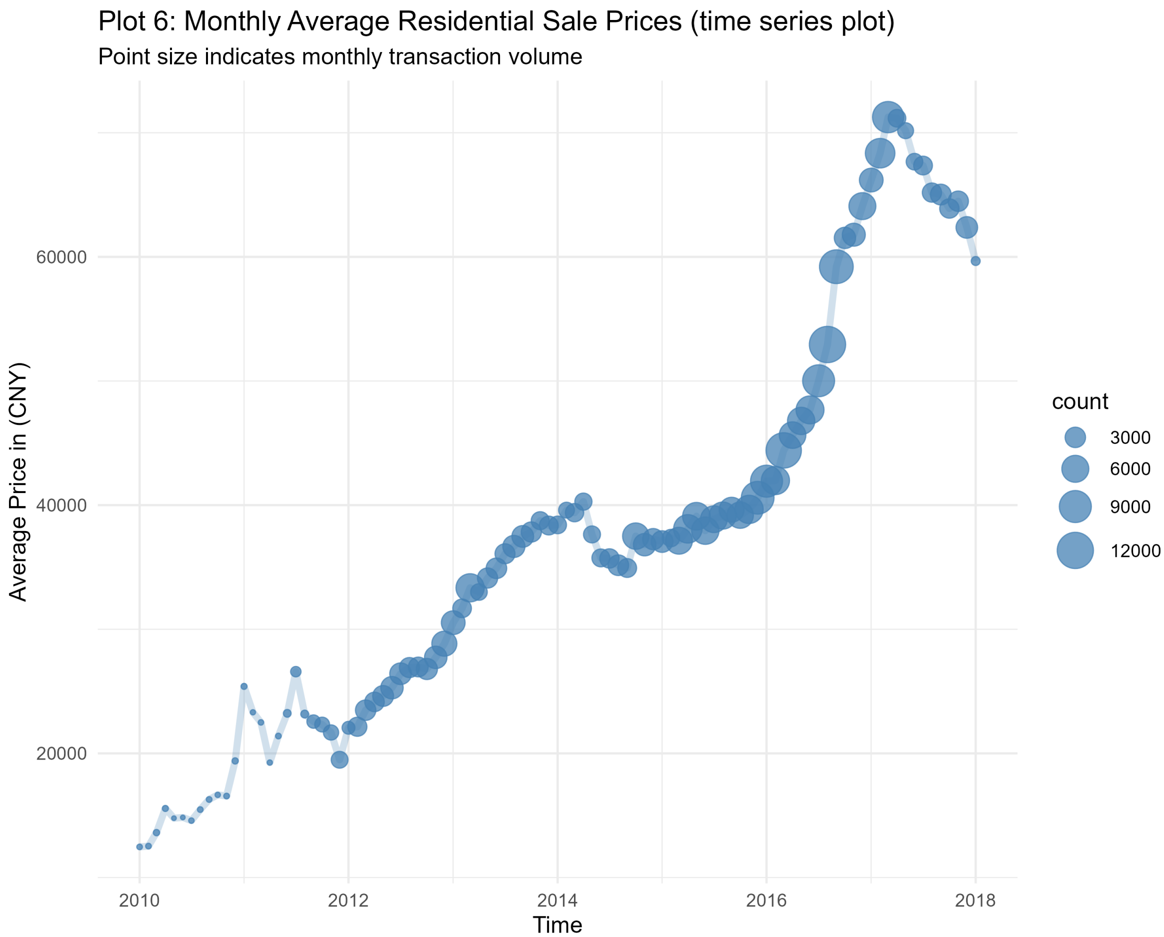

Figure 6: Monthly Average Residential Sale Price

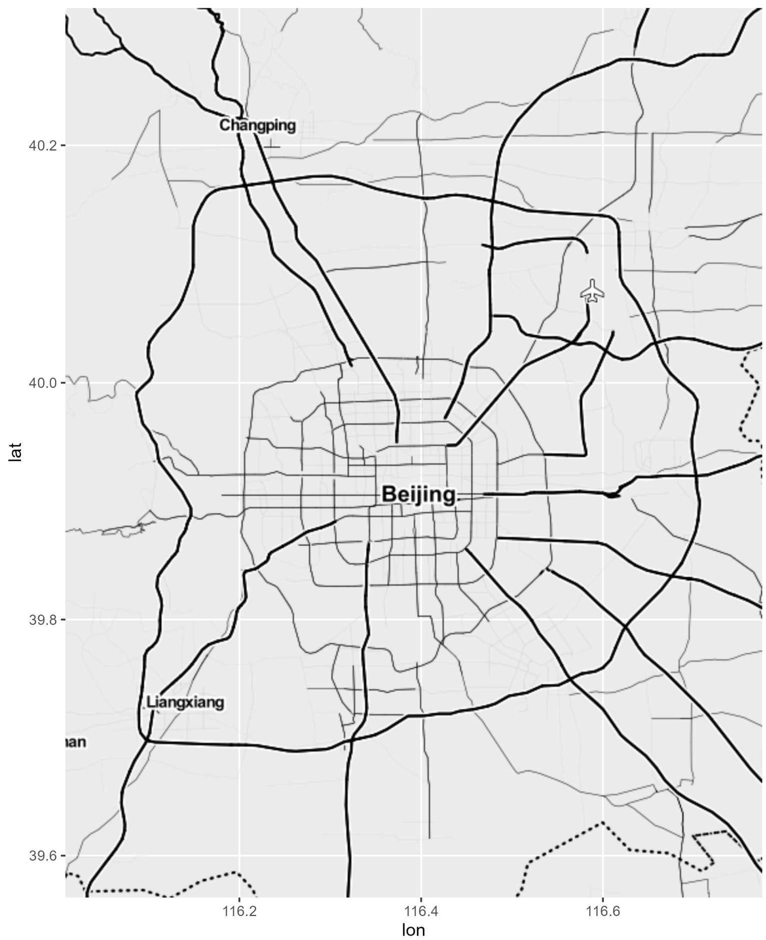

Figure 7: Beijing Map

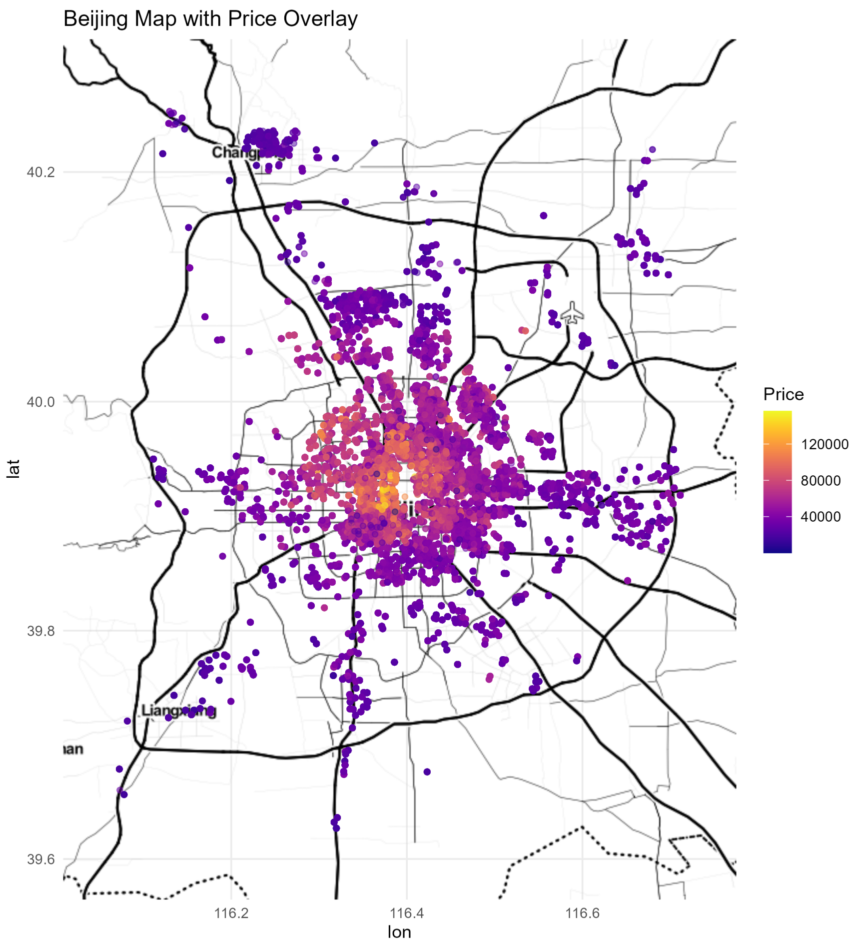

Figure 8: Map with Housing Price

Geospatial distribution and its influence on property values were elucidated through Figure 7 and Figure 8. A base map of Beijing provided the context for a subsequent overlay of price data, revealing distinct high and low-value areas across the cityscape.

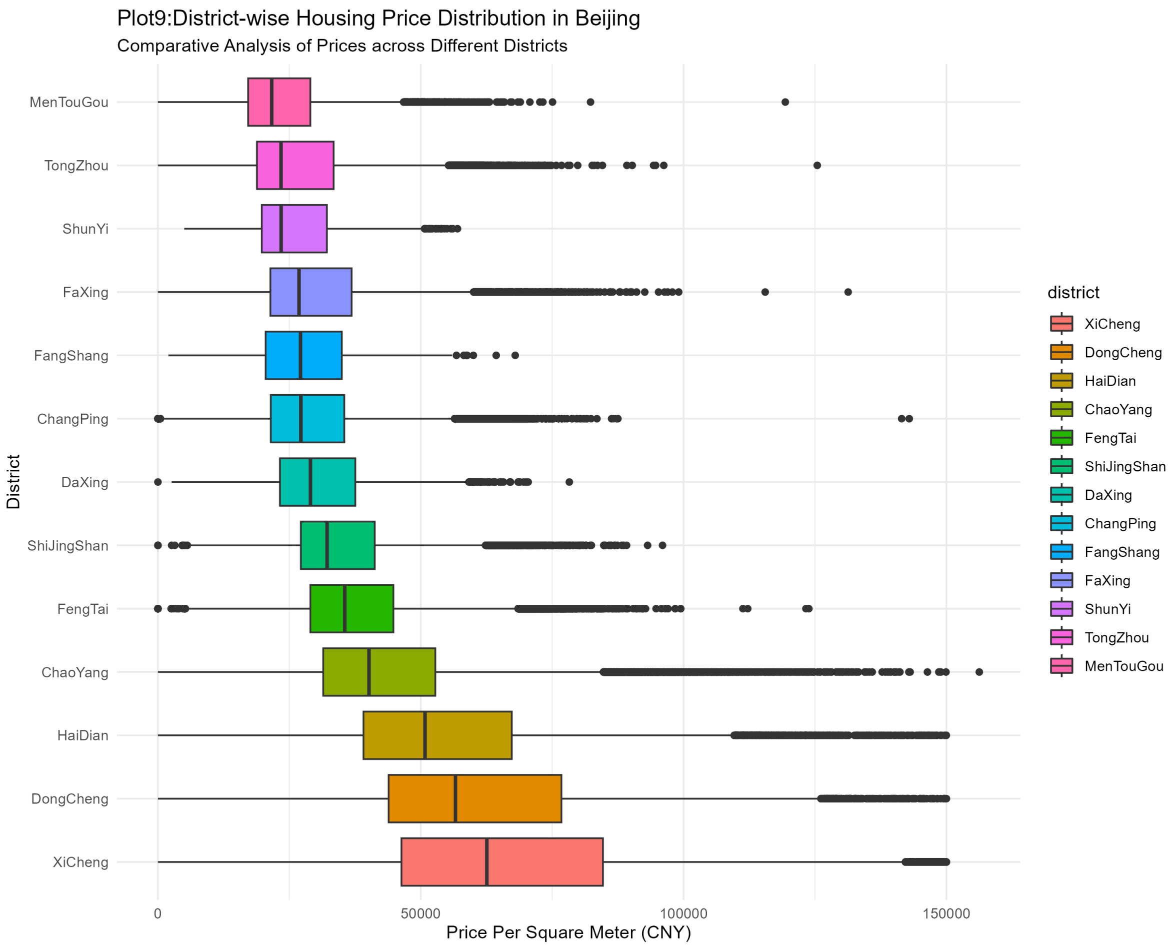

Figure 9: District-wise Housing Price Distribution

Figure 9 presented a detailed district-wise analysis, comparing housing prices across Beijing's various districts through box plots. This granular view highlighted the real estate disparities influenced by district-specific characteristics.

Finally, Figure 10 combined geographic and price data by color-coding each district on a map of Beijing, offering an at-a-glance comparison of real estate valuation across different urban areas.

Our linear regression model, developed after the exploratory analysis, utilized these processed and visualized data to predict total property prices. The model's efficacy was visually assessed by overlaying actual versus predicted prices on a scatter plot, which served as a testament to

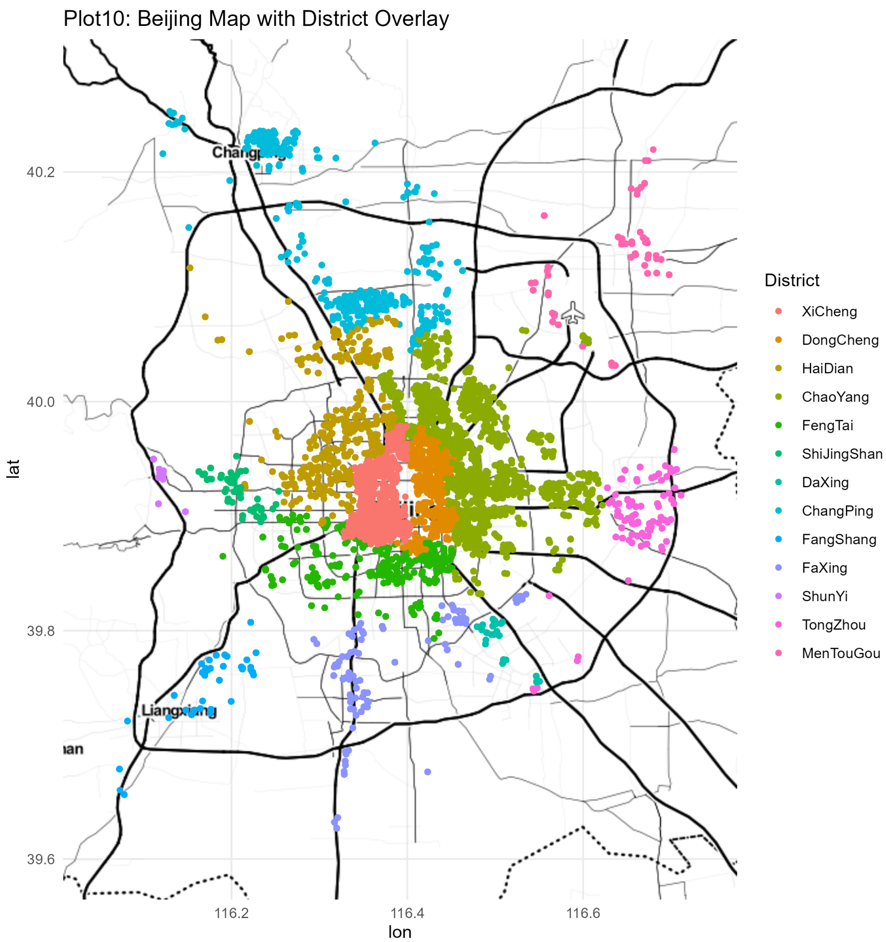

Figure 10: Beijing Map with District Overlay

the model's predictive capabilities. The careful curation and analysis of the dataset provided insights into the market drivers for housing prices in Beijing, which could aid stakeholders in making data-driven decisions.

Result

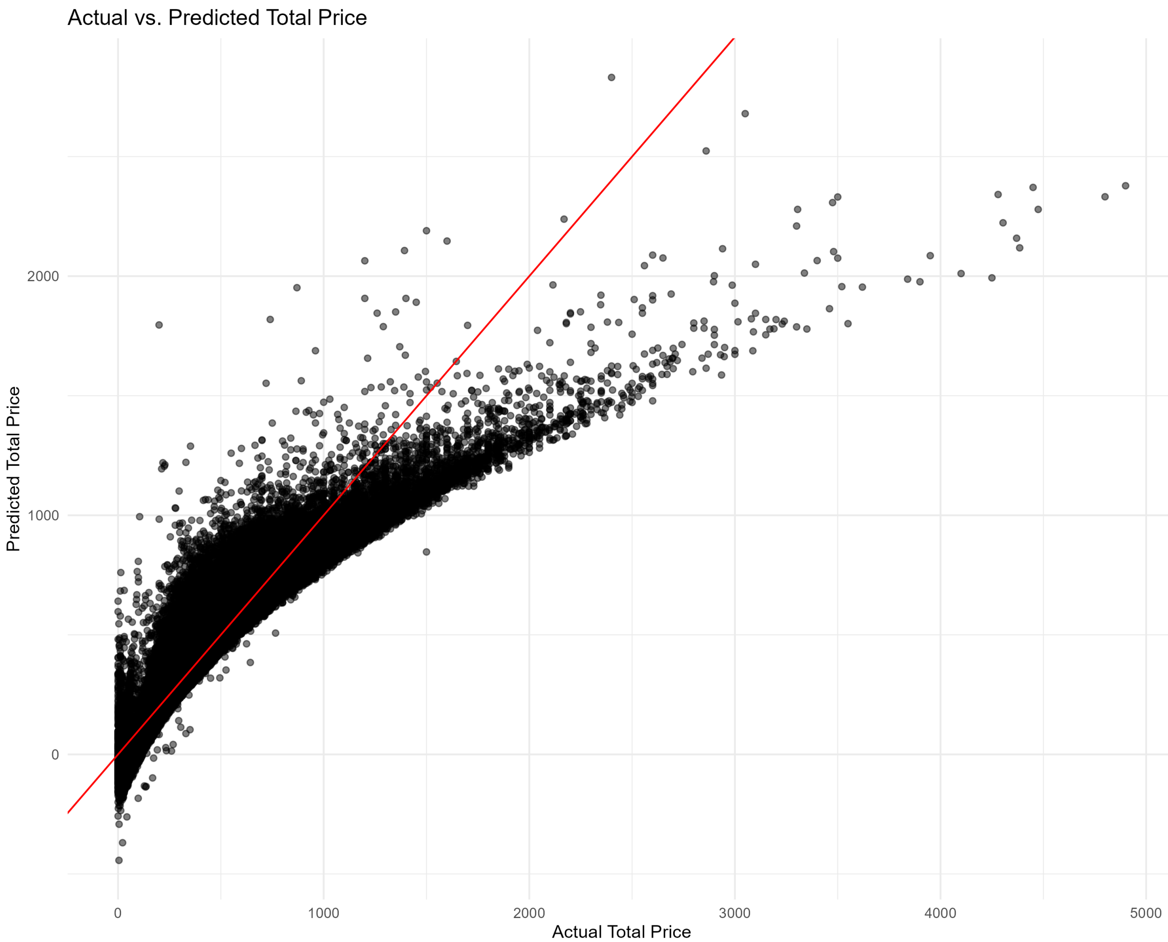

Figure 11: Prediction Model

Table 1 is the statistical summary of our linear regression model, which is built to predict house prices in Beijing, with the response variable totalPrice and various explanatory variables. The model excludes constructionTime and buildingType from the analysis.

The model has a high Multiple R-squared value of 0.8831, indicating a strong fit to the data.

Figure 11 compares actual price and predicted total prices from the model, with a red line representing perfect prediction. Points close to this line are predictions that match closely

Table 1

```
Call:
lm(formula = totalPrice ~ . - constructionTime - buildingType,
    data = beijing_house_price)

Residuals:
     Min       1Q   Median       3Q      Max
-1596.36   -25.86     1.62    29.21  2522.00

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                         7.094e+03  3.970e+02  17.868  < 2e-16 ***
Lng                                -7.155e+01  2.970e+00 -24.090  < 2e-16 ***
Lat                                 1.240e+01  3.415e+00   3.630 0.000283 ***
tradeTime                           1.943e-02  4.687e-04  41.454  < 2e-16 ***
followers                           1.010e-02  4.456e-03   2.267 0.023404 *
price                               7.044e-03  1.378e-05 511.325  < 2e-16 ***
square                              4.493e+00  7.836e-03 573.444  < 2e-16 ***
livingRoom                         -3.137e+00  2.777e-01 -11.294  < 2e-16 ***
drawingRoom                        -8.287e+00  3.450e-01 -24.020  < 2e-16 ***
kitchen                             1.427e+01  1.408e+00  10.134  < 2e-16 ***
bathRoom                           -1.207e+01  4.808e-01 -25.092  < 2e-16 ***
floor                              -2.213e-01  2.913e-02  -7.600 2.97e-14 ***
renovationConditionOther            8.163e-01  4.375e-01   1.866 0.062085 .
renovationConditionRough           -1.522e+01  1.095e+00 -13.900  < 2e-16 ***
renovationConditionSimplicity      -9.305e+00  3.685e-01 -25.254  < 2e-16 ***
buildingStructureBrick/Wood        -2.206e+02  8.732e+00 -25.261  < 2e-16 ***
buildingStructureMixed              5.619e+00  6.951e-01   8.085 6.26e-16 ***
buildingStructureSteel              6.469e+00  6.202e+00   1.043 0.296900
buildingStructureSteel/Concrete     2.215e+01  8.336e-01  26.574  < 2e-16 ***
buildingStructureUnavailable       -1.768e+00  1.150e+01  -0.154 0.877831
ladderRatio                        -2.565e-06  5.440e-06  -0.472 0.637225
elevatorNo_elevator                 4.545e+00  5.982e-01   7.599 3.00e-14 ***
fiveYearsPropertyOwnership > 5years  4.461e-01  2.999e-01   1.488 0.136828
subwayNo_Subway                     2.131e+00  3.076e-01   6.927 4.31e-12 ***
districtChaoYang                    4.417e+01  7.965e-01  55.447  < 2e-16 ***
districtDaXing                      3.773e+01  1.912e+00  19.730  < 2e-16 ***
districtDongCheng                   8.754e+00  1.112e+00   7.871 3.53e-15 ***
districtFangShang                   3.562e+00  2.081e+00   1.711 0.087001 .
districtFaXing                      3.448e+01  1.331e+00  25.906  < 2e-16 ***
districtFengTai                     3.842e+01  1.035e+00  37.110  < 2e-16 ***
districtHaiDian                     1.820e+01  8.194e-01  22.211  < 2e-16 ***
districtMenTouGou                   3.488e+01  1.214e+00  28.723  < 2e-16 ***
districtShiJingShan                 3.899e+01  1.186e+00  32.887  < 2e-16 ***
districtShunYi                      2.662e+01  2.253e+00  11.818  < 2e-16 ***
districtTongZhou                    4.810e+01  1.261e+00  38.153  < 2e-16 ***
districtXiCheng                    -1.401e+01  1.121e+00 -12.503  < 2e-16 ***
communityAverage                    1.013e-03  1.443e-05  70.190  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

with actual prices, whereas points further away represent less accurate predictions. Figure 11 shows a strong positive correlation, but with a certain degree of variance as prices increase, suggesting that while the model predicts lower-priced houses quite accurately, it is less precise with higher-priced ones.