The Research on Factors Influencing Housing Prices -Take Kansas City as an Example

Muqi Li*

High school Attached to Northeast Normal University, Changchun, 130017, China

*Corresponding author: austin02040024@outlook.com

Abstract. This article aims to identify those factors that have an impact in housing prices. The sample of the study is 15 variables, including 19554 pieces of data. Among the 15 variables, there are both factors of the house itself and some external influencing factors. The method of Multiple Linear Regression is used to analyze the significant factors with the housing prices of Kansas City. Based on these 15 variables it does correlate with house prices. The 19554 data substituted into Multiple Linear Regression Model. The result shows that grade, latitude, years of built, waterfront, view, square of basement, square of above, zip code, bathrooms and condition have a significant linear relationship with prices, while the id, date, bedrooms, square of lot, floors and year of renovating fail the significance test. In totally, the volatility of housing prices in Kansas City can be considered by the extent to which these factors affect them.

Keywords: House prices; multiple linear regression model; effecting factors.

1. Introduction

From the perspective of the international situation, the world today is in a situation of great change that has not been seen in a century. As the international economic situation keeps fluctuating, housing prices have also been affected a lot. From 2010 to 2020 the housing price of Kansas City had an increase of 2.4 times. Investment in real estate is often accompanied by a lot of risks. House prices also have the characteristics of volatile and unstable [1]. The saturation of the real estate market has exacerbated the fluctuations in house price [2]. Houses are indispensable part in public's daily life. However, the factors affecting housing prices are unbeknown. That affected public's daily consumption and increased the risk of investment. Therefore, this paper aims to help people evaluate the expected purchase of a house based on the different potential factors that may lead to house prices in Kansas City's housing price research.

Real estate industry is playing an important role in the economic development in a country. Therefore, contribution has been the object of the government and public's focus and research. Hung et al pointed out that the contribution of real estate to Gross Domestic Product (GDP) has been maintained at more than 10%, which is an indicator of China's macroeconomic situation [3]. For the factors affected housing prices, house price data is of great significance for analyzing and predicting the development of the real estate market. Dong et al. have statistic of the national commercial housing sales price index. They have analyzed that GDP, per capital GDP, resident deposit limit, consumer price index for urban resident are some main reasons of housing price fluctuation by cobweb model. They have also analysis government, owner occupied buyers, investment homebuyer and land agent's influence on housing price. They have solved the problem of model's unstable by establishing a nonlinear and non-equilibrium model and reasonable control of price adjustment parameters [4]. However, the paper has just analysis the influence of some external factors on him. Ling et al. have found most papers that about housing price forecast used to use a single model to analysis. However, a single model has shortcomings such as low model accuracy, poor generalization ability, and large influence on the number of samples. In response to the above question, a house price prediction model based on feature selection and integration learning is adopted. They mainly used 4 models in this process [5]. He et al. used a Combined Regression Prediction Model Based on Stacking Integrated Learning to solve the problem of just using a single model [6]. When Zhou et al. conducted research on the price of second-hand housing in a district of Zhuzhou City, Hunan Province, a Long

Short-Term Memory (LSTM) neural network model was built, and the prediction accuracy was improved through the particle swarm algorithm to achieve model optimization [7]. Cui et al. pointed out housing prices is affected by microscopic and microcosmic two aspects. In the paper, they used HPM to decompose the housing value into structure attribute, neighborhood attribute and location attribute [8]. In addition to the factors of real estate itself, external factors and social development also have an important impact on house prices. In the study of China's new first-line house price trends, Ma et al listed the impact of factors such as the city's gross domestic product, the number of employees at the end of the year, the number of resident population and industrial structure on house prices [9]. In the study of house prices, researchers should take into account multiple angles. Including external factors such as lot and GDP, as well as its own factors such as area and decoration [10]. This paper focuses on 15 variables (grade, latitude, years of built, waterfront, view, square of basement, square of above, zip code, bathrooms, condition, id, date, bedrooms, square of lot, floors and year of renovating) that were studied to determine their impact on housing prices.

In summary, this article will use the multiple linear regression model to study the impact of these 18 factors on Kansas City's housing prices.

2. Methods

2.1. Data Source

The data be used in this paper was got from Kaggle. The dataset includes 19554 sets of data and 15 variations of housing prices in Cansas City. The dataset shows some basic facilities index of houses in Kansas City. Most of these variables have linear relationships with house prices. Therefore, the dataset can be used as the sample in this study. The research selected all of them as sample.

2.2. Variable Selection

For the selection of dataset, the most important standard is if there are enough variables to support multiple linear regression model to work out the accurate prediction of housing prices and weight coefficient of each influence factor. Also, the most variables that be chosen must be able to presented by data.

The original dataset includes a large amount of data (Table 1). The picture below shows there are many outliers in bathrooms, floors and many other variations. There are also some repeated data in the square of living room and the square of lot. This paper removed these data. The data be used in the paper includes 15 variations (id, date, bedrooms, bathrooms, square of living rooms, square of lot, floors, waterfront, view, condition, grade, square of above, square of basement, years of built, years of renovated, zip code, latitude, long).

Table 1. Variable introduction

Field	Role	Actions Taken
bathrooms	Predictor	Trim outliers
bedrooms	Predictor	Merge categories to maximize association with target
condition	Predictor	Merge categories to maximize association with target
floors	Predictor	Trim outliers
grade	Predictor	Merge categories to maximize association with target
id	Predictor	Trim outliers
latitude	Predictor	Trim outliers
long	Predictor	Trim outliers

2.3. Method Introduction

The method be used to analysis the factors influencing housing prices is Linear Regression Model. Linear Regression Model is a kind of statistics models, used to describe the linear relationship between one explained variable and multiple explanatory variables. When the regression function

describes the linear relationship between an interpreted variable and multiple interpreted variables, the resulting regression function is a multivariate linear regression model. This paper uses a multiple linear regression model to study the influencing factors of house prices, aiming to explore the relationship between the 20 factors selected in the article and the house price and the weight ratio.

3. Results and Discussion

3.1. Multiple Linear Model Results

Table 2 shows the process of analysis. In the procedure analysis, the model used the method of Forward Stepwise. The house price is predicted through 15 factors selected in the article. Through the collinearity diagnostics there are 3 variables deleted. The VIF values of the rest of the variables are less than 5, and the tolerance is greater than 0.2. This shows that there is no colinear problem with the above variables which can be used as research samples.

Table 2. Model results

	Unstandard		standard	t	p	Collinaerity	
	B	SE	Beta			VIF	tol
Constant	8054035.038	3270733.427	-	2.462	0.014*	-	-
id	0	0	-0.011	-2.588	0.010**	1.026	0.974
view	53690.078	2355.66	0.112	22.792	0.000**	1.385	0.722
grade	138017.325	2231.996	0.439	61.836	0.000**	2.896	0.345
bedrooms	-3805.562	1978.724	-0.01	-1.923	0.054	1.45	0.689
date	129.669	13.991	0.039	9.268	0.000**	1.008	0.992
bathrooms	109517.326	3277.004	0.228	33.42	0.000**	2.674	0.374
yr_built	-3300.519	79.957	-0.264	-41.279	0.000**	2.357	0.424
sqft_lot	0.244	0.056	0.025	4.336	0.000**	1.983	0.504
yr_renovated	23.206	4.144	0.025	5.6	0.000**	1.147	0.872
floors	16764.258	3610.222	0.025	4.644	0.000**	1.616	0.619
zipcode	-513.227	37.08	-0.074	-13.841	0.000**	1.665	0.6
waterfront	626851.953	19688.972	0.145	31.838	0.000**	1.194	0.837
lat	585806.799	11895.992	0.222	49.244	0.000**	1.165	0.858
condition	27666.619	2625.04	0.049	10.54	0.000**	1.248	0.801
long	-162419.477	14480.892	-0.062	-11.216	0.000**	1.757	0.569
sqft_living15	92.663	3.578	0.172	25.898	0.000**	2.531	0.395
sqft_lot15	-0.132	0.087	-0.009	-1.521	0.128	2.021	0.495
R^2	0.661						
Adj R^2	0.66						
F	F (17,19536)=2238.454,p=0.000						
D-W	1.992						
Dependent variable = price							
* p<0.05 ** p<0.01							

3.2. Model Evaluation

The result shows that grade, latitude, years of built, waterfront, view, square of basement, square of above, zip code, bathrooms, condition are the main influence factors. It is worth mentioning that the percentage of houses' grade accounts about half of the effect that the factors of influencing house prices in Kansas City. It shows the grade of houses influence the value of the houses to a great extent.

The table above shows the evaluation process of model accuracy. Including the degree of fitting and error of the model. The degree of fitting reaches 0.660, more than 0.5. The result shows the model that be used is suitable (Table 3).

Table 3. Model summary

R	R^2	Adj R^2	RMSE	DW	AIC	BIC
0.813	0.661	0.66	213801.347	1.992	535492.611	535634.468

4. Conclusion

The research selected 21614 samples from the dataset, which has 18 variables. During the analysis stage, the article uses a multiple linear regression model, which is used to find out possible relationship between the variables and housing prices. The factors that positively impact on house prices are the square of living room, latitude, square of basement, square of above and bathrooms. The factors that negatively impact on house prices are the square of lot, long, zip code, years of built. In totally, grade, latitude, years of built, waterfront and view are the main influence factors.

The above research can give investors who plan to buy a house and invest in the real estate market a reference perspective, and predict future house prices to a certain extent. However, there are some shortcomings in this article, such as the inability to predict the impact of games between different roles in the market on housing prices. At the same time, there are also many limitations in predicting the market with a single model. In order to optimize the above problems, a combination of multiple models can be used. To evaluate the fit of the model, the R square can be used to measure the fit of the model to the data. In addition, it is necessary to check whether the hypothetical conditions of the model are met, such as the normality, independence, variance, etc.

In a nutshell, the housing prices in Kansas City keep increasing in a slow speed. The houses in Kansas City are good choices to invest or keep them as finance products. people who want to buy the houses in Kansas City should follow the principle that choose the houses with good conditions of the factors that are positively correlated with house prices and avoid the houses with the factors that are negatively correlated with house prices.

References

[1] Liu X. Real estate investment decision and cost control analysis. Residential business, 2023, (08).

[2] Wei H. Analysis on the risk of investment decision of real estate enterprises and its countermeasures. Financial client, 2023, (04).

[3] Huan Yan, Ni Zha. LASSO-ARIMA house price forecast based on urban development in Beijing. 2021 (7th) National Statistical Modeling Competition for College Students, 2021.

[4] Wang Dong, et al. Housing price composition and analysis model based on cobweb theory. 2010 Third International Conference Education Technology and Training, 2010.

[5] Ling Li. Housing price prediction model based on ensemble learning algorithm. IT. 2022.

[6] He Mao, et al. Research on Lasso-GBDT portfolio housing price forecasting Model based on Stacking integrated learning. Suzhou Business Theory, 2023, (06).

[7] Cui Huang. Study on the influencing factors of housing price in Shenzhen based on multi-scale geographical weighted regression. Working paper, 2023 (06).

[8] Zhou Li. Research on forecasting method of regional second-hand house price based on PSO-LSTM. Modern information technology, 2024, 8(05).

[9] Ma Chen. Analysis of the trend characteristics and influencing factors of housing prices in China's new first-tier cities. Regional finance research, 2019, (06).

[10] Stein L J C. Leverage and House-Price Dynamics in U.S. Cities. Social Science Electronic Publishing, 1999, 30(3): 498-514.