ISSN 2616-5902 Vol. 7, Issue 6: 1-6, DOI: 10.25236/AJBM.2025.070601

Data-Mining Algorithms on the Basis of House Prices

Menglei Liu[1,a], Xiaowen Yu[2,b,*]

[1]Hefei University of Technology, Hefei, China
[2]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation
[a]1003165980@qq.com, [b]yuj.s@edu.spbstu.ru
[*]Corresponding author

Abstract: *Data mining is a core component of the digital age, especially in the real estate industry, where both home buyers and developers need a large amount of data analytics to support their decisions. Usually, buyers see a wide range of brands as well as different types and prices are extraordinarily confused, in today's serious inflation, home buyers are generally insufficient budget, so how to use a lower price to buy a preferred housing is an urgent matter. In this paper, we use descriptive statistical analysis to analyze the current offerings of 231 developers in Hefei, Anhui Province, and the results can be used as a reference for home buyers to purchase properties.*

Keywords: *Data-Mining, Descriptive Statistical Analysis, Property Developer, Homebuyer, Hefei City, Anhui Province*

1. Introduction

The rise and fall of house prices are determined by supply and demand. There are two types of buyers, the first is the investor who wants to invest to increase the value of his investment and the other is the consumer who wants to have a home [1]. In the long run, house prices are ultimately determined by consumers; when supply exceeds demand, no matter how many interfering factors there are, house prices will always be in a downward trend; when supply is less than demand, house prices are bound to be in an upward trend [2-3]. Similarly, China's housing prices also follow the above supply and demand, China's property marketization began in 1998 with the housing renovation project, and has been rising steadily since 2000, paving the way for China's housing prices to rise, like a sweeping historical scroll, recording the changes of the times and the pulse of development [4].

The number of home buyers in China in 2025 falls off a cliff compared to the period before 2025. What factors have contributed to this state of affairs in China's real estate sector? This paper analyses the reasons for the downturn in China's property sector in terms of four factors. China has less than 8 million newborns in 2023, a trend of negative population growth, more than 21% of the population over the age of 60, the elderly have a very low demand for housing, and finally the urbanization rate has exceeded more than 65%, and the remaining rural population is mostly elderly people and children who do not have the ability to buy a house, and their contribution to the growth of the urbanization rate is almost 0. Therefore, from the demographic factor, the incremental demand for new housing in urban areas is almost dried up [5].

For economic factors, China's economic growth has continued to slow down since the end of 2024, the domestic GDP growth rate has declined, the per capita income has decreased significantly, the inflation rate has continued to rise, and the property market has been directly affected, with some families choosing to postpone their home purchase plans or simply give up on purchasing a home due to unstable incomes or pressure of indebtedness [6].

For market factors, the 779 million square meters of commercial real estate for sale in China, combined with the 4,666 million square meters under construction, is enough to house 300 million people, which clearly fits the basic economic principle that supply exceeds demand [7]. The head of the real estate enterprise debt ratio ordinary more than 90%, some even exceeded 100%, insolvency has become the new normal. The second-hand property market is relatively hot. Compared with new properties, second-hand properties can usually be purchased in the same condition at a lower price, and they can also enjoy mature supporting facilities, which has impacted the new property market [8].

For the conceptual factors, more and more young people no longer buy a house as a life goal, more focus on the quality of life as well as personal development, in the face of high prices and mortgage pressure, a lot of young people choose to lie flat, renting has become a more popular choice [9]. Young people are not fixated on working in one city, and the increased uncertainty of where they will work makes them less likely to consider buying a house in one area.

Therefore, in this paper, we propose some methods on real estate data mining for the purpose of analyzing the characteristics of real estate data in specific provinces in China, which can provide some references for home buyers.

2. Datasets

This paper collects data from 2022-2024 on mainstream property developers in all areas of Hefei, Anhui Province, China [10]. We have selected six factors that are most important to home buyers as the basic indicators, including the property developer, the size of the property, the opening price of the property, the highest price at which the property was sold, the lowest price at which the property was sold in the previous square footage, and the number of properties sold as of 6 February 2025, as well as the number of homes sold.

The brand influence, quality of buildings, greening rate, reputation and location of different property developers vary greatly, affecting consumer acceptance and expected prices, which in turn affects property prices and sales.

The apportioned floor area refers to the floor area of the common parts of the entire building that is shared by the owners of the entire building [11]. Including elevator shafts, pipe shafts, stairwells, etc., as well as for the whole building services public rooms and administration of the building floor area, calculated as the horizontal projected area. The size of a commercial property is one of the core factors in determining the price of a property. Usually, the larger the size, the higher the total price of the commercial property, which is a key indicator of the value of the property.

The opening price is a key concept in the financial markets, especially in real estate. The opening price refers to the first sale price of a commercial property when it begins to sell. In areas where the property market is highly competitive, developers, in order to attract more home buyers and gain market share, will launch some relatively low-priced commercial properties in order to stand out from the crowd. Listings with certain defects in terms of flooring, orientation, and floor plan are often priced below market. Real estate projects are nearing their end, with a small number of properties remaining, and due to factors, such as a mismatch between buyer demand and these properties, developers will lower their prices in order to achieve the goal of clearing their properties as quickly as possible, and to quickly get their money back. As the auction price of land in different cities is different, and is in the core of the city, near the transport hubs, the housing stock with good facilities has the scarcity and convenience, the price is often at a high level.

The reason for selecting these price data in this paper's dataset is to be able to show the initial positioning of property prices as well as the range of fluctuations, reflecting the market's judgement of the value of the property, and the relationship between supply and demand, which is essential for analyzing house price data.

For conservative homebuyers, the ability to deliver on time is critical, and the amount of the house payment is so large that it has led to a series of mortgage breaks. Transaction volume is higher in core locations but almost zero in remote locations. transaction volume reflects the popularity of the property and the actual sales situation, which can reflect to a certain extent the supply and demand relationship in the market, which also affects the property prices.

In this paper, data mining is required, where descriptive statistical analyses are performed, correlation coefficients between variables are usually analyzed and relationships between variables are visualized. Based on the results obtained from the analysis, it can provide some reference for home buyers who are going to buy a home in Hefei City in the near future.

3. Methods

Descriptive statistical analysis is a method of data analysis designed to summaries and describe data in order to better understand its characteristics and trends.

Four different descriptive analyses are used in this paper:

- Concentration trend analysis;

- Discrete trend analysis;

- Distribution pattern analysis;

- Frequency analysis.

Concentrated trend analysis includes the mean, median, and plurality. The mean reflects the average level of the data and provides a reasonable price for home buyers to refer to. The median can avoid the interference of extreme values. When analyzing house prices, the median can more reasonably reflect the middle price level. The median is a more reasonable reflection of the middle price when analyzing house prices. The plurality indicates the most frequent data and is used to analyze the most popular commercial properties in different areas.

Dispersion analysis, which consists of variance, standard deviation, and extreme deviation. Variance measures how much the data deviates from the mean; the higher the variance, the more dispersed the data. The standard deviation is the square root of the variance, which is the same unit as the original data, making it easy to understand, and is used in this paper to analyze house price fluctuations. The extreme deviation is the difference between the maximum and minimum values, which can visually reflect the range of fluctuations and the degree of dispersion of the data.

Distribution pattern analysis included skewness and kurtosis. Skewness describes a measure of the symmetry of the data distribution, which is approximately symmetrical when the skewness is zero; positively skewed when the skewness is greater than zero; and negatively skewed when the skewness is less than zero. Kurtosis is a measure of the height of the peak of the data distribution. Compared with the normal distribution, a kurtosis greater than 3 is a peaked distribution, and a kurtosis less than 3 is a flat-peaked distribution. It can help home buyers quickly understand the degree of concentration and dispersion of the data.

Frequency is the number of occurrences of data in a certain interval or taking a certain value, which is used to draw a frequency histogram, which can visually present the distribution of data in order to help home buyers make the right decision.

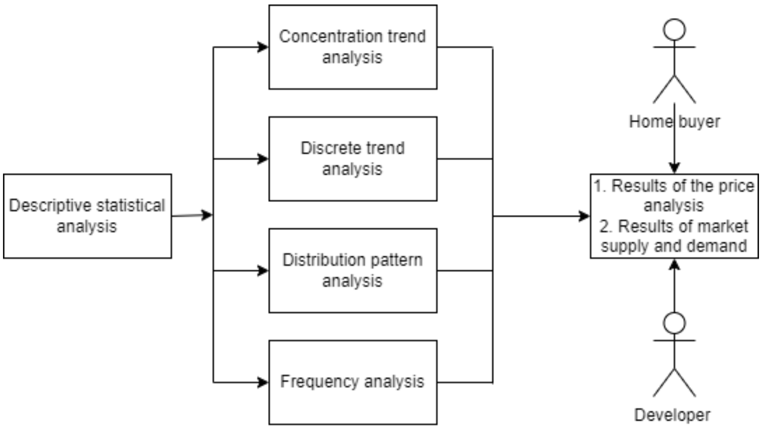
Descriptive statistical analysis
Concentration trend analysis
Discrete trend analysis
Distribution pattern analysis
Frequency analysis
Home buyer
1. Results of the price analysis
2. Results of market supply and demand
Developer

Figure 1: Impact of Descriptive Statistical Analysis Results on Developer and Homebuyer Decisions

As shown in Figure 1, by using four different descriptive statistics methods, we can get the results of analyses about the prices of commercial properties as well as the relationship between supply and demand in the market, which can help home buyers and developers to make the right decisions based on the results.

If analyses show that home prices have been on the rise and will continue to be on the rise, then home buyers will need to accelerate their purchases in consideration of the increased costs and may even accept higher prices. Developers will then need to consider covering their properties and waiting to sell, with the aim of increasing the pricing of new properties to make higher profits.

When supply exceeds demand in the market, home buyers will have more choices, and likewise more room for bargaining, and will be able to choose their favorite properties more comfortably. When the market supply is less than demand, home buyers face more intense competition and need to consider raising their budgets or lowering their purchasing standards, and even after this competition has heated up, there will be developers who will try to get a chance to buy a home through the lottery, scalping, and so on.

For developers, when demand is high, they increase investment, speed up project development and increase the supply of housing to enter new regional markets and make high profits. When demand is low, developers have to consider the number of sites to be purchased, slow down the pace of development and optimize the design of housing models to differentiate themselves from the competition.

Both homebuyers and developers should pay attention to the results of price analyses, market supply and demand, and the current volume of commercial property transactions in order to guide their next decisions based on the real situation.

4. Experiments

In this paper, four different descriptive statistics methods are selected for experimentation based on the characteristics of the dataset, including a frequency distribution chart of area, a price distribution pattern chart, a volume dispersion chart, and a volume concentration trend chart.

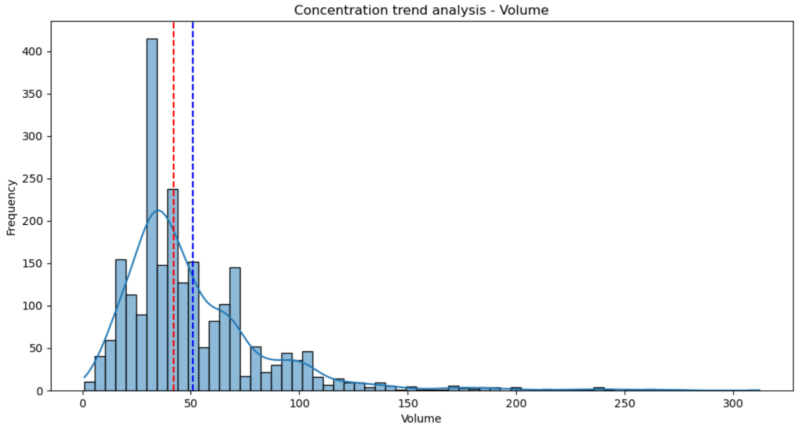

Figure 2: Concentration trend analysis

As shown in Figure 2, we can see the concentration trend in the frequency of volume. It can be analyzed that property developers' current inventory is sufficient, and only by appropriately lowering prices or holding other promotional activities can they reduce the current inventory of commercial properties as soon as possible and carry out capital return.

The volume zones corresponding to the peaks in the graph are the zones with the highest concentration of home sales in the market. Real estate companies can set reasonable sales targets based on the concentration trend of turnover and taking into account their own market share and development plans.

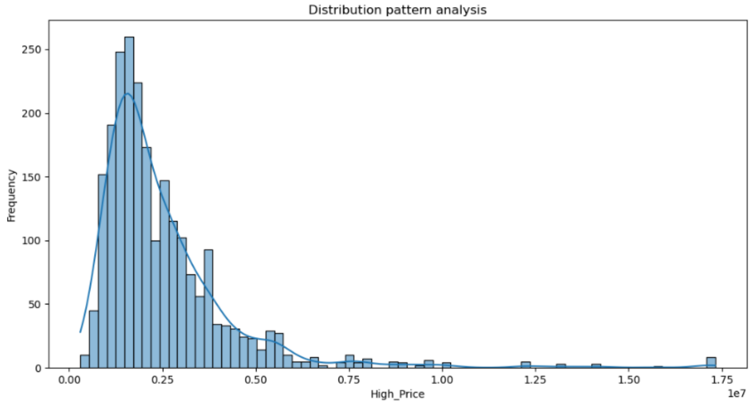

Figure 3: Price distribution analysis

As shown in Figure 3, the distribution of the highest price of commercial housing shows multiple peaks, indicating that there are different price levels in the market, with high-end luxury housing, ordinary commercial housing as well as affordable housing in different grades to meet the needs of different consumer groups.

Investors or home buyers can judge the trend of house prices based on the distribution of prices, the chart shows that the house prices in Hefei City, Anhui Province in the last three years have been in a state of continuous decline after a slight increase, which indicates that there will be a large-scale decline in house prices, and at the same time the market is going downward, the investment or purchase of homes need to be careful in making decisions. It is recommended that non-demand home buyers can continue to wait and see where the industry and property prices are heading, and wait for the market and property prices to be in a more stable state before considering purchasing a home.

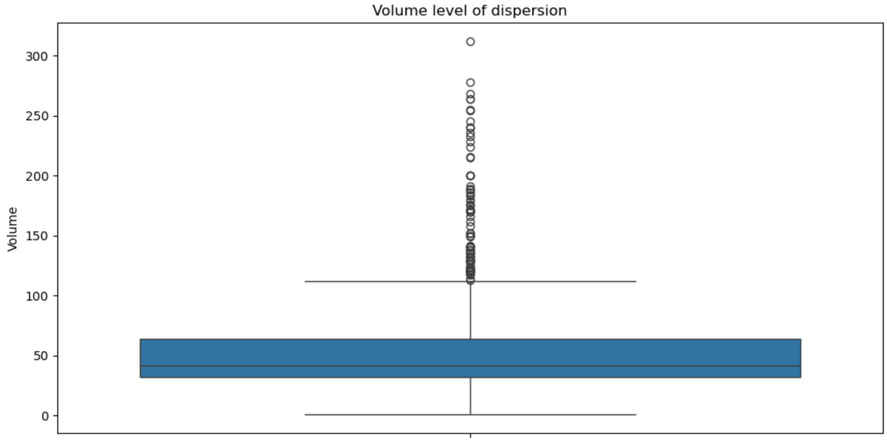

Figure 4: Volume Dispersion Analysis

As shown in Figure 4, the box line chart has a longer box with a larger range of upper and lower limits and more outliers, indicating low housing volume and higher volatility in the market. The turnover of a large number of outliers, some properties are significantly higher than others, it is recommended that buyers with a new need can pay more attention to these properties, because these areas are likely to have significant favorable planning, as well as the government's support policies.

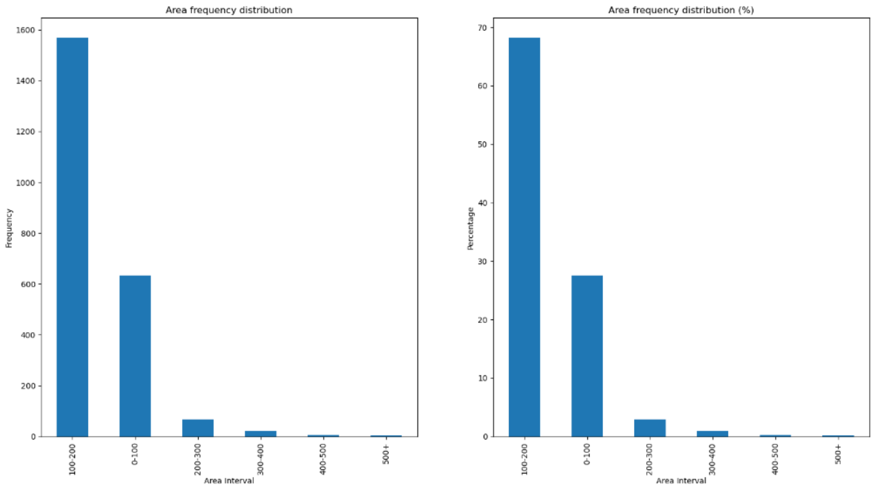

Figure 5: Area frequency distribution analysis

As shown in Figure 5, the peak corresponds to a volume range of 100-200 square meters which is the most concentrated range of housing transactions in the market. Similarly, this shows that the majority of buyers in Hefei City, Anhui Province, choose properties in the 100-200 square meters range, followed by those who are more economically challenged and choose affordable housing in the 0-100 square meters range; this is the market norm and can be used as a benchmark for market analysis. Developers can determine the size and type as well as the price of the products to be developed based on the area frequency distribution, understanding the mainstream market demand for the area range, and combining with their own cost of acquiring land, target customer groups and other factors.

5. Discussion and Conclusion

Through descriptive statistical analyses, this paper draws the following conclusions: small and medium-sized ranges dominate the market, both in terms of turnover and frequency, which perfectly fits the mainstream demand of new and improved homebuying groups. The current distribution of commercial property prices in Hefei City, Anhui Province, shows a right skewed pattern, with most of the properties concentrated in the middle and low-price segments, and prices continue to fall.

The discrete degree of commercial property turnover is obvious, with some unusually high or low transaction figures, indicating poor market stability and localized hot and cold transactions. The fact that the volume of commercial property transactions is more concentrated in a specific volume range indicates that the market has a relatively stable range of transaction sizes under normal circumstances.

Of course, there are some errors in making predictions about the property market just from the conclusions obtained from the descriptive statistical analysis, so we will expand the dataset and use recurrent neural networks to make predictions about house prices in Anhui Province in the next phase.

References

[1] Molloy R, Nathanson C G, Paciorek A. Housing supply and affordability: Evidence from rents, housing consumption and household location[J]. Journal of Urban Economics, 2022, 129: 103427.
[2] Howard G, Liebersohn J. Why is the rent so darn high? The role of growing demand to live in housing-supply-inelastic cities[J]. Journal of Urban Economics, 2021, 124: 103369.
[3] Aastveit K A, Albuquerque B, Anundsen A K. Changing supply elasticities and regional housing booms[J]. Journal of Money, Credit and Banking, 2023, 55(7): 1749-1783.
[4] ZHANG Jinjuan, GUO Haiyan. A Review and Prospect of Data-Driven Research on Real Estate Bulk Appraisal[J]. Journal of Xihua University (Philosophy and Social Science Edition), 2024, 43(3): 13-27.
[5] Zhang L, Ren H, Li C. Study on the development characteristics and spatial and temporal patterns of population ageing in 31 central cities in China[J]. Frontiers in Public Health, 2024, 12: 1341455.
[6] He Y, Wu M, Jiang H. Land supply marketization, economic fluctuations and welfare: A quantitative analysis for China[J]. International Journal of Strategic Property Management, 2024, 28(3): 152–162.
[7] Shen S, Zhao Y, Pang J. Local Housing market sentiments and returns: Evidence from China[J]. The Journal of Real Estate Finance and Economics, 2024, 68(3): 488-522.
[8] Tang C, Liu X, Yang G. A study of financial market resilience in China—From a hot money shock perspective[J]. Pacific-Basin Finance Journal, 2024, 83: 102256.
[9] Soltani A, Zali N, Aghajani H, et al. The nexus between transportation infrastructure and housing prices in metropolitan regions[J]. Journal of Housing and the Built Environment, 2024, 39(2): 787-812.
[10] Data on house prices in Hefei, Anhui Province, China. Anjuke. url: https://m.anjuke.com/hf/xinfang/?from=TW_HOME
[11] Tu Yaping. An introduction to the influence of public supporting facilities and common area on the area of commercial property[J]. Science Chinese, 2015 (08Z): 139.