Pisco Med Publishing

Original Research Article

A Study on Spatial Visualization of Residential Prices in Yibin City with Big Data Technology

Bin Zhao , Prof.Chao Chen

School of Computer Science and Engineering, Sichuan University of Science & Engineering

Abstract: In this paper, five counties and districts of Yibin City are selected as key research samples. With the support of ArcGIS, python, echorts, pyechards and other software technologies, this paper studies the data acquisition, cleaning and visualization of second-hand houses sold in Yibin City in 2021 on the chain home platform, scientifically and objectively understands the law of housing price differentiation in Yibin City, explores the impact of various influencing factors on housing prices, promotes urban planning and construction, and promotes the steady development of the real estate market.
Keywords: Arcgis; Data visualization; Residential prices; Real Estate.

1. Introduction

Yibin City has 10 county-level administrative units, and five administrative districts, namely Xuzhou District (Yibin County), Cuiping District, Nanxi District, Changning County and Jiang'an County, are selected as the focus of this study.

Data acquisition and visualization research has a wide range of applications, involving traffic visualization, biomedical visualization, stock market technology analysis, and many other fields[1-2].

The conversion of data into graphical representations, supplemented by interactive means to enhance human cognitive ability of data, decision aids, etc. demonstrates a powerful empowering effect[3]. Therefore, big data technology is introduced into the analysis of house prices and presented in a visual way to explore the variation of house prices in cities. Yibin is developing rapidly. The vigorous introduction policy of "Yibin University City" makes Yibin real estate warm. Therefore, it is of great significance to do visual research.

2. Data Acquisition

2.1 Yibin second-hand house price data collection

Real estate information data: By simulating a browser client web crawler, the server data is requested in bulk. Specifically, it crawls the Yibin second-hand property information on the Chain Real Estate website, including address, unit price per square meter, name, etc.

The web crawler gets a total of 18 dimensions of information such as property category title, broker, household type, area, unit price, address, and construction time.

Geographic information data: Committed to perfect visualization, longitude and latitude are added to each property information based on the address field information obtained by the crawler. We use Map Location tool to convert address information to latitude and longitude coordinates in batch[2], and use the Chinese coordinate offset standard BD-09 coordinate system to improve the reliability of accuracy. Thus, the sample data with both price and geographic attributes are generated.

2.2 Data Processing

The cleaning of house price data mainly includes: merging of list elements, processing of data missing values, outliers, duplicate values, etc. The processing of outliers and missing values needs to be sequential, and the outliers cannot be used as the basis for filling in the missing values, so the processing of outliers is performed first.

Grubbs test: A set of measurement data, if the individual data deviates from the average value, then this data is called "suspicious value". If a statistical method such as Grubbs (Grubbs) method of judgment, the "suspicious value" from the group of measurement data excluded from the calculation of the average value, then the "suspicious value" is called "outliers (Gross error） ". The calculation steps are as follows:

（1）Calculate the average of the sample unit prices:

$$\mu=\frac{(X_1+X_2+\cdots+X_n)}{n}\#(1)$$

（2）Calculate the sample standard deviation:

doi: 10.18282/l-e.v10i9.3194

$$s = \sqrt{\frac{\sum (X_i - \mu)^2}{n-1}} (i = 1,2\cdots n) \#(2)$$

（3）Calculation of the Grubbs test statistic:

$$G_i = \frac{(X_i - \mu)}{s} (i = 1,2\cdots n) \#(3)$$

（4）The calculated value G_i is compared with the critical value given by the Grubbs table $G_p(n)$, and if the calculated value is greater than the critical value it is considered as an outlier.

（5）The threshold value $G_p(n) = 3$ is set according to the house price data, so when $G_i > G_p(n)$, X_i is discriminated as an abnormal value, and the opposite is normal.

Table 1. Partially processed data

No	Type	Area	Price	Face	Building year
1	3 bedrooms, 2 baths	155㎡	8259 RMB/㎡	South	2009
2	3 bedrooms, 1 bath	90.6㎡	5611 RMB/㎡	Southeast	2003
3	3 bedrooms, 2 baths	84.5㎡	8876 RMB/㎡	South	2021
4	3 bedrooms, 2 baths	82㎡	7781 RMB/㎡	Southeast	2021
5	2 bedrooms, 1 bath	80㎡	8525 RMB/㎡	Northwest	2013

3. Real Estate Visualization Analysis

3.1 House price data grouping

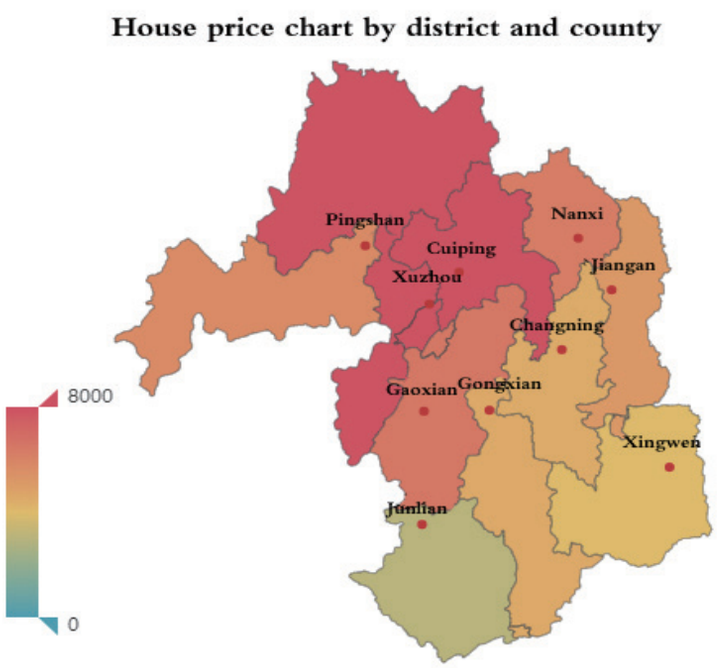

Figure 1. Heat map of the average price of housing

The main basis for classifying property data is clustering and grouping according to the importance of factors affecting house prices, for the overall average second-hand house price in Yibin in the sample data is 7652 Yuan/㎡. However, the overall average price is not sensitive to outliers and does not fully show the house price in Yibin. Therefore, the overall average price cannot be used as a reference price for purchasing a house in Yibin.Therefore, the sample data clustering is divided into five administrative districts, namely Xuzhou District, Cuiping District, Nanxi District, Changning County and Jiang'an County, and the other districts and counties are studied in general.

The greatest advantage of the box plot is that it is not affected by outliers and can accurately and stably depict the discrete distribution of the data, and it also facilitates data cleaning. Figure 1 shows the quartile plot of second-hand house prices in Yibin.

The city house price heat map shows the high and low level of house price information in the form of a map classification to help home buyers make reasonable decisions within the current fluctuation range according to their needs. As shown in Figure1.

According to Figure 1, grouping Yibin properties by district and county, it can be clearly seen that Cui Ping and Xuzhou District (Yibin County) have higher house prices in general, with Cuiping District in the first place in the city. The box plot divides the house prices into four equal parts, such as 25% of the properties in Cuiping District differ by 2,900 yuan from 3,500 yuan to 6,400 yuan, 25% of the properties differ by 1,400 yuan from 6,400 yuan to 7,800 yuan, 25% of the properties differ by 1,600 yuan from 7,800 yuan to 9,400 yuan, and the remaining 25% differ by 3,600 yuan from 9,400 yuan to 13,000 yuan. The median is relatively high, and the overall distribution of different prices is relatively balanced, which can meet the purchase of different income groups.

Despite the impact of the epidemic, the total GDP of the districts and counties still belong to the positive growth state. The GDP of Cuiping District is more than twice as much as that of Xuzhou District, which is in second place. The introduction of the Yibin University City also provided strong impetus for the housing prices in Cuiping District. Changning and Jiang'an counties are relatively stable with a small price gap, so if you don't consider other factors but just need to consider them first.

3.2 Spatial exploratory analysis

The visual representation of urban geospatial data can visually reflect the city's population, economy and other livelihood data on

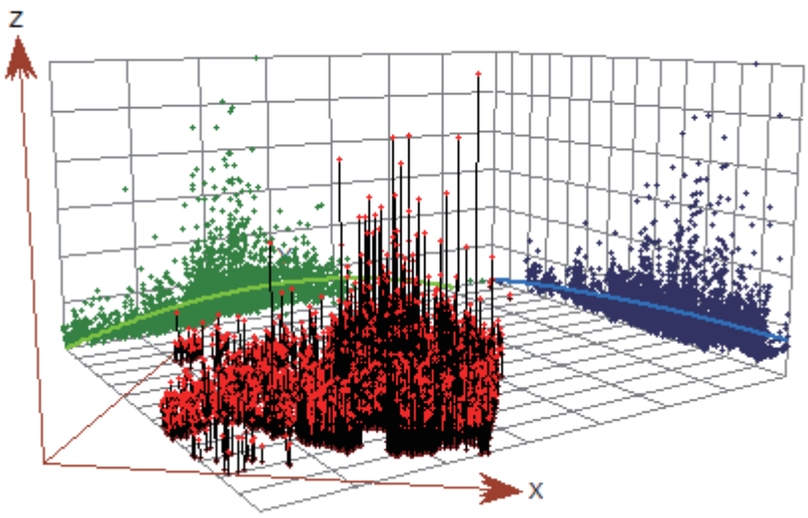

Figure 2. Spatial trend map

the base map. Housing price, as one of its criteria, is also the most concerned by the public.Combined with the spatial trend analysis map to clearly express the distribution of house prices in Yibin. Using the data exploration function of Geostatistical Analyst in Arcgis 10.5 software, the house price data containing geographic coordinates information was added to the trend analysis module, and because the location information was accurate to the neighborhood, multiple properties appeared to have the same coordinate data. To ensure the integrity of the data the overlapping samples are all included.As shown in Figure 2, the x-axis and y-axis represent the east-west and north-south directions of geographic information, respectively, and the z-axis indicates the price of second-hand properties. The red dots represent the price points of the second-hand listings, short black lines of different lengths to the bottom indicate the different geographical locations where the listings reside, and the red price points at the top are projected onto the north-south and east-west grids respectively to form green and blue dots, which more clearly shows the distribution status of the listing data, and the trend analysis fits a trend curve based on the projection on the grid. Figure 2 shows that the listings are mainly concentrated to the downtown location, and there are relatively few listings in the north-south and east-west directions, indicating that the density of listings is higher in the center of Yibin and relatively less in the suburbs. The trend curve fitted by the projection of price points shows an inverted "U" shape, and the price also shows a trend of "high in the middle and low around"[3].

For only focusing on the price information of the property does not highlight the comprehensiveness of the property information. In fact, the house orientation largely determines the price of the property itself. Combining house density, orientation, and price.

The analysis shows that: the overall number of south-facing properties in each district and county is large and evenly distributed, and the most attention is paid to the south-facing properties. The scatter diagram of east facing direction reflects that the scatter is mainly concentrated in Xuzhou district and Cuiping district, and the price is also much higher than other districts and counties at this time, from the regional economic level, Xuzhou and Cuiping are also the leaders of Yibin city economy. It is verified that the house orientation determines the price of the property itself to a large extent. The order of the house orientation is roughly south, southeast, east, southwest, north and west, with the best facing south and the worst facing west.

If you want to have more orientation options, you should give priority to Xuzhou and Cuiping.

4. Conclusion

This paper mainly introduces the method of crawling, cleaning, visualization and spatial integration of property data. The main focus is on the property portal Chain Home as the data source, and with the technical support of Python web crawler to capture the Yibin city property data, to get a total of 18 dimensions of property categories such as household type, area, unit price, and so on, and from this further to get the key geographic information data. The data are processed secondarily using pandas and other technologies and then integrated with charts and spatial exploration methods to represent the data, not just simple graphical data. As shown in the program and the article, the visual representation can clearly and effectively convey the potential information implied by the raw data. We use Echarts, Arcgis, Pyecharts and other technical means to analyze and visualize Yibin property price data in depth, and present the price and its influencing elements in a reasonable visual form. It has obvious advantages over traditional analysis methods for understanding the acquisition of data. At present, the article is only a preliminary rough realization of the analysis and visualization of property data expression, the article and the research method will continue to improve and integrate more powerful analysis tools to promote more quantitative and refined data.

References:

[1] Tuarob Suppawong et al. DAViS: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction[J]. Financial Innovation, 2021, 7(1) .

[2] Qiu Yongjian and Lu Jing. A visualization algorithm for medical big data based on deep learning[J]. Measurement, 2021, 183.

[3] Xia Jiazhi, Li J, Chen S, Qin H-X, Liu S-X. A review of research at the intersection of visualization and artificial intelligence [J/OL]. Chinese Science:Information Science:1-25 [2021-11-18].(in Chinese).