Homework

朱宏橋

Python for Data Science

2017年4月15日

Compare real estate price trend between center area and outer area in Taipei city

Background

The purpose of this analysis is to compare the price trend of real estates between center area and outer area in Taipei city to see if there is any meaningful information. The definition of center area includes 5 districts of Zhong-Zheng, Zhong-Shan, Xin-Yi, Da-an, and Song-Shan districts. The outer area includes 7 districts of Shi-Lin, Nei-Hu, Nan-Gang, Wen-Shan, Wan-hua, Da-Tong, and Bei-tou districts. The data is provided by Ministry of interior administration, ROC.

Data for analysis

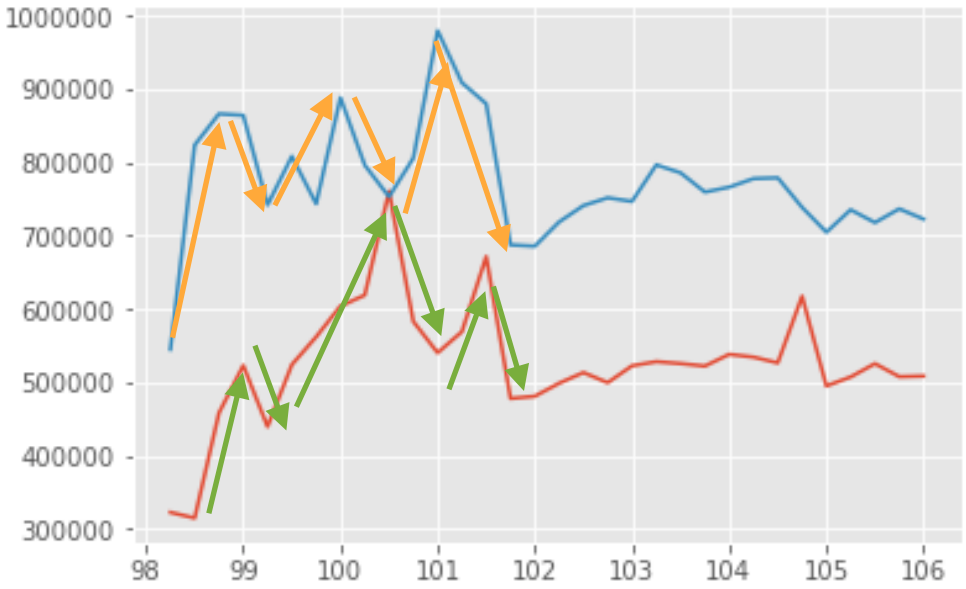
The transaction data is gathered from beginning of 2009 to end of 2016. The raw data's format is .xls file. The raw data contains many fields, but only 5 fields are used for analysis. The fields used for analysis are district, transaction date, transaction object, and unit price. Districts are classified to center or outer area. Transaction objects are limited to houses or apartments. The transactions which only contain land or parking lots are excluded to avoid bias in price. Transaction date is classified to season in a year. For example, 20130101 falls in Season 1 of 2013. In data file, the definition of unit price is NTD/ m^2. Since Ping is mostly used in Taiwan society, the unit price is further transformed to NTD/Ping. 1 Ping = 3.305785 m^2.

In the raw data file, some transactions are recorded to be traded among family members with price lower than market price. Therefore a trimming is applied to data that transaction with unit price is lower than 100000NTD/Ping is discarded.

Homework

Statistical calculation and plots

The analysis approach is rather simple. The data are divided into center area and outer area. Mean value of all transaction records in each season is calculated. Mean transaction prices of each season for both outer and center area are retrieved. Result plots are as following. In the first figure, X-axis data represents season of each year. Y-axis data is unit price, NTD/ping. Blue line is for center area, which has higher price and the orange line is outer area, which has lower price.



In the second figure, the line is meant to represent the slope of price change and is calculated by following formula.

Y = [Prince(Sn)-Price(Sn-1)]/Price(Sn-1)

Like in previous figure, blue line is for center area and orange line is for outer area.

Result and Conclusion

From the visualization result, in the fist figure, from RoC Year 98 (2009) to Year 102(2013) have similar waveform of rise and falls in price trends. Each area has 3 rising periods and 3 falling periods. It seems center area rises first and then outer area follows. When center are reaches its peak, it also starts to decline first and then the outer area follows the trend. If this observation and analysis is valid, we can predict outer area's price by trend by observing center area's status.