SUUMO in Central Tokyo

Hiromasa Ito

Contents

The Purpose

Living in the central Tokyo is a dream for most people, which means there is a high demand for apartments there. However, since brand-new apartments are quite expensive, when it comes to purchasing, second-hand apartments are often a more realistic choice than new ones.

So, this analysis focuses on second-hand apartments in central Tokyo.

The major question in this analysis is:

Which types of second-hand apartments are ideal for investment?

Sub-questions to support the major question are:

- Analysis of the most valuable areas and popular types of apartments
- Analysis of factors that affect apartment prices significantly
- Analysis of undervalued apartments

Data

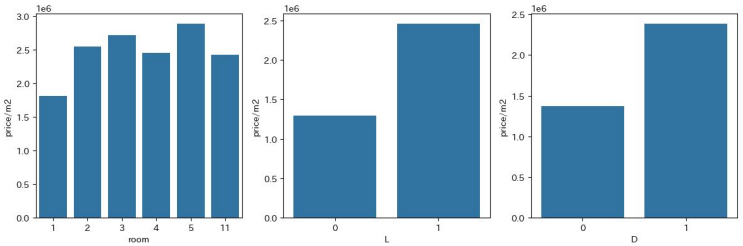
Scraped from SUUMO

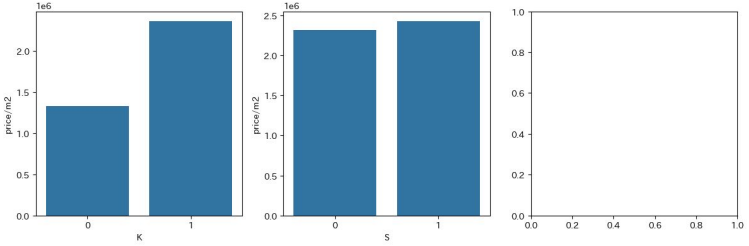
https://suumo.jp/kanto/

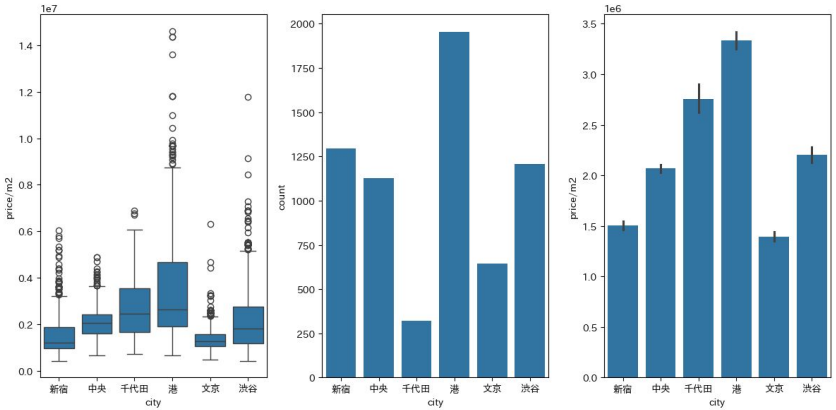
Downloaded from Statistics of Tokyo

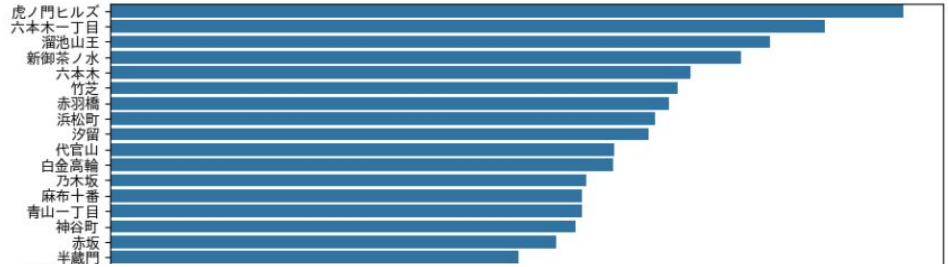
https://www.toukei.metro.tokyo.lg.jp/juukiy/jy-index.htm

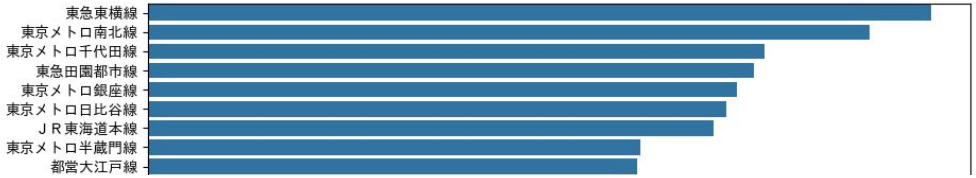
EDA











- Room 2 & 3 are high price
- Living, Dining and Kitchen are important
- “港” is the most expensive city
- “虎ノ門ヒルズ” is the most expensive station
- “東急東横線” is the most expensive line

Machine Learning Model

- **Regression Model (Multiple Linear Regression)**

I evaluated the model using R-squared, and got feature coefficients to understand which features increase or decrease the apartment price.

- **Clustering (K-Mean)**

I used the Elbow method to decide the number of clusters, and I found that the k=6 is the best number of clusters. Then, outputted the table and visualized the table with the spider plot.

- **Frequent Pattern Mining (Apriori)**

I set the minimum support for the Apriori algorithm to 0.05, and the minimum confidence threshold for generating rules to 0.5. Then, I performed frequent pattern mining to find out what kinds of apartments tend to be undervalued based on EDA and the coefficients from the regression model by checking three metrics; support, confidence and lift.

Results

Regression

	Feature	Coefficient	Abs Coefficient
0	age	-0.624770	0.624770
1	room	0.150759	0.150759
2	walking time	-0.127047	0.127047
3	city_港	0.123839	0.123839
4	popularity	0.118411	0.118411
5	city_新宿	-0.087199	0.087199
6	count	0.085210	0.085210
7	L	0.078259	0.078259
8	city_中央	-0.071210	0.071210
9	station_渋谷	0.067879	0.067879

Age is the most important element, following room and walking time

If apartments meet conditions below:

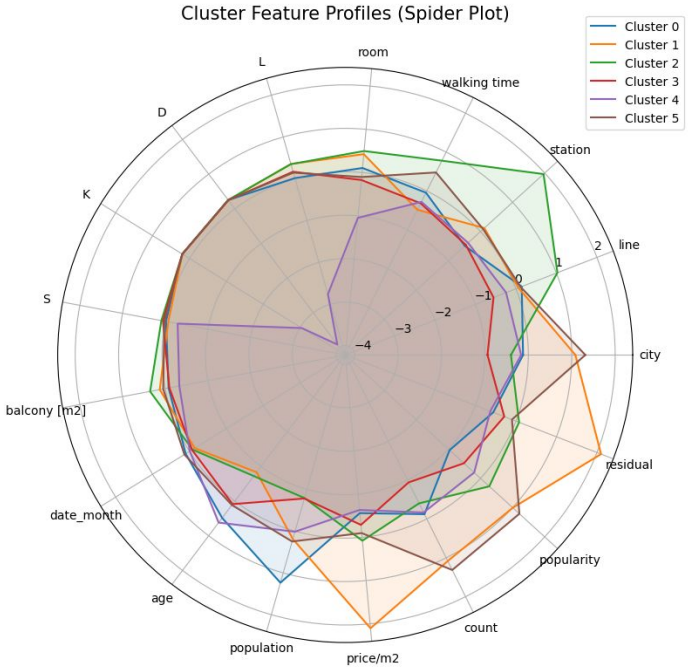
- area [m2]_medium, age_old
- walking time_close, age_old
- room_2, age_old

We can think they are undervalued with a 60% probability

Frequent Pattern Mining

	antecedents	consequents	support	confidence	lift
106	(area [m2]_medium, age_old)	(residual_super undervalued)	0.055115	0.610829	3.047167
79	(age_old, walking time_close)	(residual_super undervalued)	0.061679	0.603886	3.012533
122	(room_2, age_old)	(residual_super undervalued)	0.078168	0.599532	2.990809

Clustering



- Cluster 0: Dense or older area
- Cluster 1: top-tier and high-demand area
- Cluster 2: Far but large apartment
- Cluster 3: low-demand and narrow apartment
- Cluster 4: Poor layout
- Cluster 5: Expensive but not popular

Conclusion

Which types of second-hand apartments are ideal for investment.

Based on the integrated analysis using EDA, regression, clustering and frequent pattern mining, the most promising apartments for investment are:

- Older apartment located close to stations, which are undervalued despite their convenience
- Apartments with 2 rooms or medium floor size, which has balance between cost and functionality
- Apartments in lower popularity areas, which may be overlooked by the market
- Simple layout properties (missing a Living or Dining), which tend to be priced lower but still meet demand for those who do not have a lot of money

These characteristics appeared clearly in undervalued elements in my analytical approaches. Therefore, investors should look for these types of properties in central Tokyo.

Future Works

- Time Series Data

In SUUMO website, there is no time series data. So, it is hard to analyze the trend of apartment. If I have this kind of data, I can know the future price of apartment

- Comparison with the other city

I only considered Central Tokyo this time. However, if I analyze the other big city such as Osaka, Yokohama, I may be able to find the other findings

- Analysis of other building

I analyzed second-hand buildings in this report. However, the purpose of buying other building like a house or new apartment is different from the second-hand apartment. So, I may discover different results