Where is the best place to live in Tokyo?

Hirofumi Iwata

July 26 2020

1.Introduction

1.1 Background

Tokyo, where is the capital of Japan is the largest city in Japan. As the Olympic Games are scheduled to be held in 2021, it is the city with the large number of tourists. The central part of Tokyo is divided into 23 wards, each with different characteristics. For example, Ota Ward has many small factories, and Chiyoda Ward has many offices. In Shinagawa Ward, where I live, the west is a residential area, while the east is a business area. With a population of 9 million, Tokyo is a very interesting city with a mix of cultures and characteristics.

1.2 Problem

Tokyo has various regional characteristics, which are not understood by not only foreign tourists and Japanese travelers but also those living in Tokyo.

1.3 Interest

By classifying the twenty three wards of Tokyo based on their characteristics, I analyze what areas are suitable when staying overnight, and which areas when living. And the theme was to give advice to travelers.

2. Data acquisition and cleaning

2.1 Data sources

Data including Japanese postal code, city name, latitude and longitude can be obtained from the following.
https://www.aggdata.com/free/japan-postal-codes
However, this data includes data from all over Japan. Therefore, it is necessary to extract only the necessary part. In addition, the granularity of the data is so fine that multiple city names with the same latitude and longitude may be included.

The other data is regional information obtained from Foursquare.

2.2 Data Cleaning

There are several problems for the datasets.

	postalcode	Neighborhood	state	Borough	Latitude	Longitude
0	490-1401	Rokujocho	Aichi Ken	Yatomi Shi	34.9	137.15
1	490-1402	Gotoyama	Aichi Ken	Yatomi Shi	34.9	137.15
2	490-1403	Toriganjicho	Aichi Ken	Yatomi Shi	34.9	137.15
3	490-1403	Toriganji	Aichi Ken	Yatomi Shi	34.9	137.15
4	490-1404	Ikadaba	Aichi Ken	Yatomi Shi	34.9	137.15

```
df.shape
```

```
(123695, 6)
```

First, the dataset is huge because it contains all addresses in Japan, and there are 123695 rows. So, after making only the data whose "state" is "Tokyo to", I made only the data that "ku"(it means “ward”) was included in "borough".

Second, the data was very detailed and included multiple addresses with the same "longitude" and "latitude". Therefore, the duplication of “longitude" and "latitude" was dropped.

The following data set was obtained by the above processing.

	postalcode	Neighborhood	state	Borough	Latitude	Longitude
108164	130-0000	Ikanikeisaiganaibaai	Tokyo To	Sumida Ku	35.7068	139.8072
108165	130-0001	Azumabashi	Tokyo To	Sumida Ku	35.7096	139.8031
108166	130-0002	Narihira	Tokyo To	Sumida Ku	35.7079	139.8134
108167	130-0003	Yokokawa	Tokyo To	Sumida Ku	35.7048	139.8156
108168	130-0004	Honjo	Tokyo To	Sumida Ku	35.7047	139.8021

```
df_tokyo4.shape
```

```
(651, 6)
```

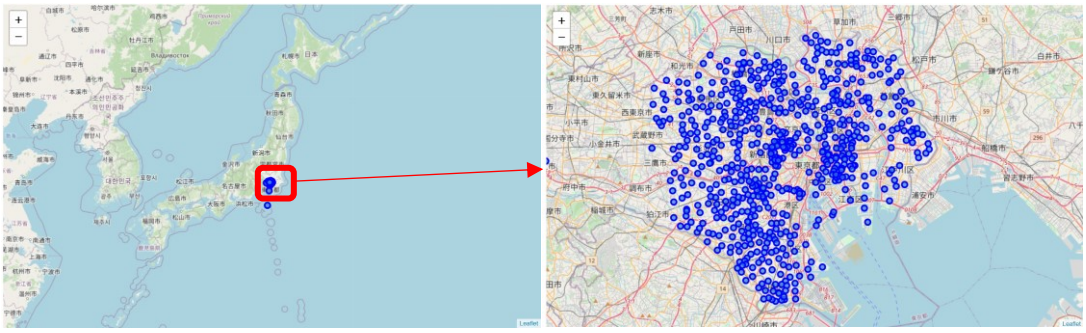
I will work with this dataset and Foursquare geodata to solve the problem.

3. Methodology

3.1 Process

In this section, I describe how to proceed with the data analysis.

First, I use Folium to check the distribution of the data prepared in the previous section.The result is as follows, and it was confirmed that the whole of Tokyo was covered.

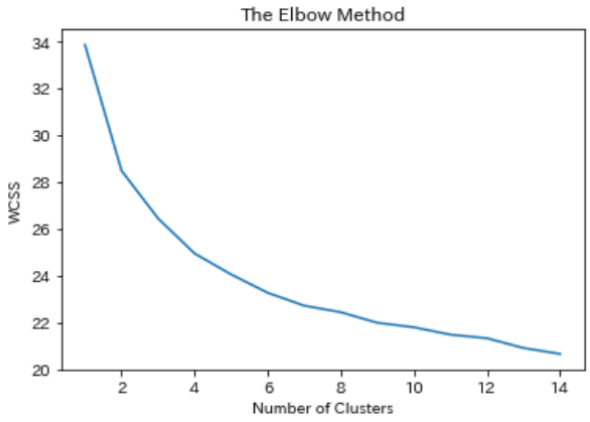


Next, Using Foursquare's API, we have acquired information of venues within a radius of 500m from each point. In a result, the amount of venues is 25,740 and there are 422 uniques categories. From the data obtained here, I identified 10 characteristic venues at each neiborhood. It is like below

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adachi	Convenience Store	Intersection	Donburi Restaurant	Dumpling Restaurant	Bus Station	Fast Food Restaurant	Coffee Shop	Train Station	Ice Cream Shop	Café
1	Agebacho	Japanese Restaurant	Italian Restaurant	BBQ Joint	French Restaurant	Ramen Restaurant	Sake Bar	Kaiseki Restaurant	Yakitori Restaurant	Thai Restaurant	Pub
2	Aioicho	Convenience Store	Intersection	Bus Stop	Bus Station	Grocery Store	Shoe Store	Golf Driving Range	Chinese Restaurant	Café	BBQ Joint
3	Aizumicho	Sake Bar	Ramen Restaurant	Convenience Store	BBQ Joint	Japanese Restaurant	Café	Climbing Gym	Bar	Indian Restaurant	Burger Joint
4	Akabane	Sake Bar	Convenience Store	Ramen Restaurant	BBQ Joint	Soba Restaurant	Yakitori Restaurant	Coffee Shop	Indian Restaurant	Donburi Restaurant	Bar

3.2 Make clusters

Now that we have the data for each location, we will classify this data next. I used the K-means clustering to make clusters. Because it was necessary to adopt a method of unsupervised learning, and with decision tree, because the dataframe has too many elements it was difficult to get the desired result. Using the elbow method, there is no clear part of the value of k to take, so I adopt　k = 5.

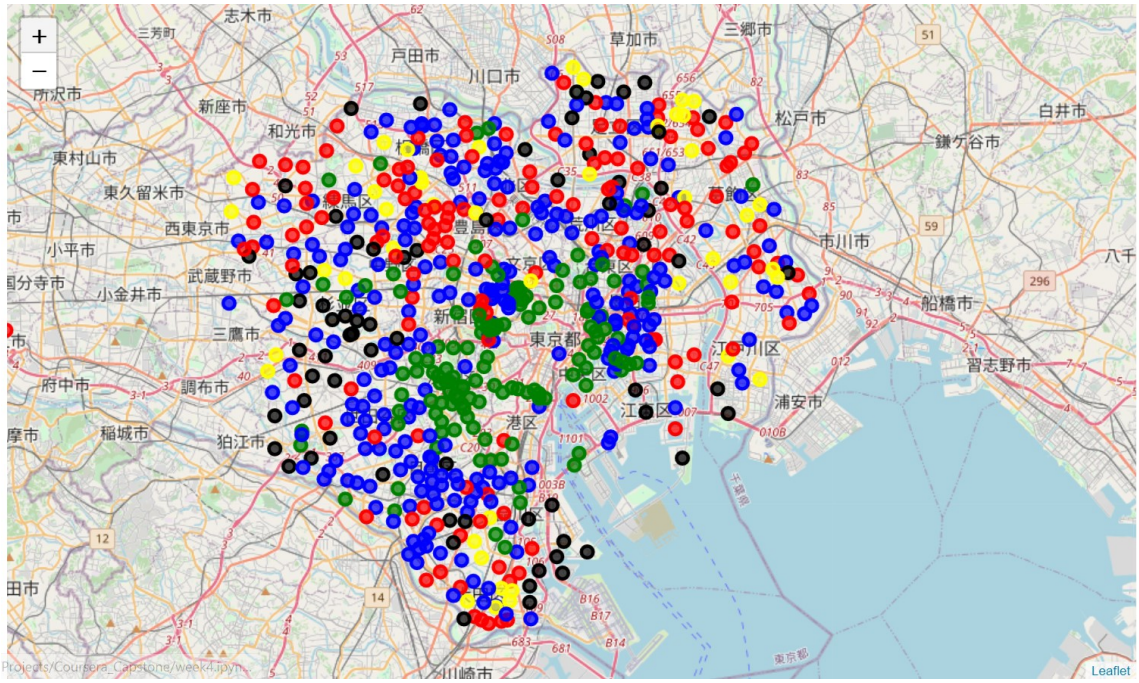


After using k-means clustering to classify into five clusters, merge the data. As a result, the following data frame was obtained.

	postalcode	Neighborhood	state	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Com V
108164	130-0000	Ikanikeisaiganaibaai	Tokyo To	Sumida Ku	35.7068	139.8072	1.0	Convenience Store	Café	Sake Bar	Japanese Restaurant	Ra Resta
108165	130-0001	Azumabashi	Tokyo To	Sumida Ku	35.7096	139.8031	1.0	Convenience Store	Café	Soba Restaurant	Park	Coffee
108166	130-0002	Narihira	Tokyo To	Sumida Ku	35.7079	139.8134	2.0	Convenience Store	Café	Scenic Lookout	Coffee Shop	Interse
108167	130-0003	Yokokawa	Tokyo To	Sumida Ku	35.7048	139.8156	1.0	Convenience Store	Ramen Restaurant	BBQ Joint	Japanese Restaurant	Interse
108168	130-0004	Honjo	Tokyo To	Sumida Ku	35.7047	139.8021	1.0	Convenience Store	Sake Bar	Grocery Store	Coffee Shop	Superm
108170	130-0011	Ishiwara	Tokyo To	Sumida Ku	35.7013	139.8041	0.0	Convenience Store	Bus Stop	Grocery Store	Chinese Restaurant	BBQ

3.3 five clusters

The five clusters obtained are listed below.



In this map, Cluster0 is red, Cluster1 is blue, Cluster2 is green, Cluster3 is yellow, Cluster4 is black. Looking at the map, it can clearly be seen the difference between the center(Cluster2) and the rest. Differences in other parts are hard to find, but it seems that blue(Cluster1) is located in the middle, black(Cluster4) is the outer, and red and yellow are others.

4. Results

4.1 The shape of five clusters

In each cluster, the value with the highest frequency of appearance is displayed for 1st, 2nd, and 3rd Common Venue.

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Convenience Store	Bus Stop	Supermarket
1	Convenience Store	Ramen Restaurant	Sake Bar
2	Japanese Restaurant	Convenience Store	Japanese Restaurant
3	Convenience Store	Grocery Store	Bus Stop
4	Convenience Store	Park	Convenience Store

Next, display the second most frequent venue. Here, "-" indicates that it did not exist.

	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Intersection	Ramen Restaurant	Japanese Restaurant
1	Ramen Restaurant	Café	Ramen Restaurant
2	Convenience Store	Café	Sake Bar
3	-	Park	Park
4	Park	Convenience Store	Intersection

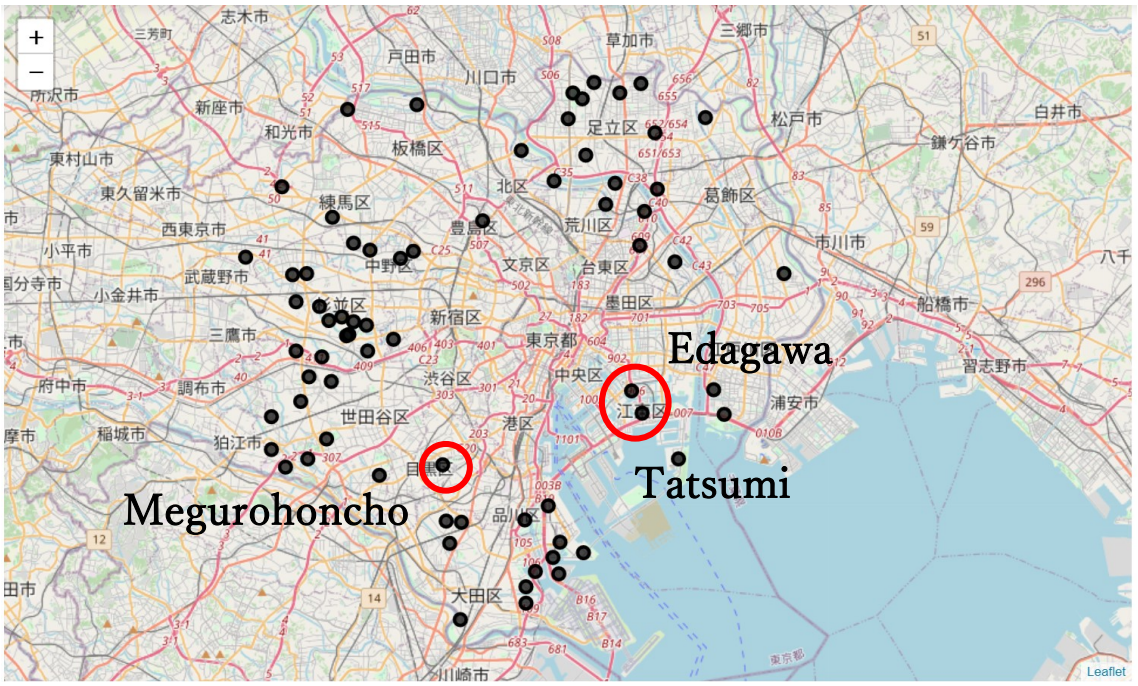
From these results, Cluster2 has many Japanese restaurants, which is clearly different from other areas. There are many convenience stores in other areas. This indicates that there are convenience stores in most areas of Tokyo. (You know that you actually live in Tokyo, but you can discover a convenience store every two minutes.)

Cluster0 is a very livable place with good transportation, many restaurants and supermarkets. Cluster1 is a good place to relax. It may be a place where it is easier to work rather than live. For Cluster3 and Cluster4, there are many parks and it is easy to shop, so it can be said that it is a suitable place for a family to live in. The results are summarized below.

Cluster No	Feature	Neighborhood
0	very livable place	Heiwadai, Okubo, Narimasu, Oi
1	good for work	Kamata, Togoshi, Nerima, Kamiya, Adachi, Sugamo
2	good for stay	Ryogoku, Ueno, Yotsuya, Kabukicho, Shibuya, Tsukiji
3	good for family	Nishioi, Egota, Toneri, Kasuga, Ikebukurohoncho
4	good for family	Seijo, Megurohoncho, Kojiya

4.2 Discussion

Considering living in Tokyo, the more away from the center, the more places for families. However, as a result of the analysis, there were places suitable for families even near the center. The following shows the maps of Cluster3 and Cluster4.



In a result, the five best places to live are: Kasuga, Nishioi, Edagawa, Tatsumi, Megurohoncho.

5. Conclusion

In this study, I used Foursquare geodata to analyze which areas are best when visiting Tokyo for various purposes. As a result, it was found that the areas near the center (Ueno, Yotsuya, etc.) are good for sightseeing and work. On the other hand, when staying for a long time or living alone, a place slightly away from the center (Heiwadai, Narimasu, Oi, etc.) is suitable. When living with a family, it is better to be further away from the center, but some areas are also suitable near the center. I recommend that you remember the following area names for the day when you live in Tokyo someday - Kasuga, Nishioi, Tatsumi, Egawa, Megurohoncho.