

# Sectioning Airbnb Neighborhoods in Tokyo: Survey the Surroundings

## 1. Description of the Problem and Discussion of the Background

Despite rising tension with the government authority in recent years, Tokyo has been a growing market for property owners renting out their space for the public; 民泊 or minpaku as they would call in Japanese. Airbnb, just as in other major cities, is among the popular platforms to find one, especially for foreigners who possess little or no local language understanding.

When arriving at Airbnb, people tend to wonder what are some good spots nearby. That hidden underrated sushi bars, authentic local ramen stalls, or maybe those grammable Japanesque cafes are waiting to be discovered. Quick Google search will do, but what if we could explore what kind of places are there in which area before even booking the place? We probably could save more time on deciding where to go and even plan our trip better. This light project aims to answer that question. It's not only targeting travelers, but also might be useful for those interested in running the lodging or related business in near proximity, or even for people who consider living in the world's most populous metropolitan city and wonder what the neighborhoods look like.

## 2. Data Preparation

Inside Airbnb has one of the most updated databases for any Airbnb cities in the world, and it's free; thanks to their scraping tool who managed to extract texts from relatively complex HTML structure of the Airbnb webpage. After grabbing only the necessary data for the analysis, the data consisting of 13367 listings (as of November 2019) would look like this:

	name	neighbourhood	latitude	longitude	price
0	La Casa Gaienmae C Harajuku, Omotesando is nearby	Shibuya	35.67152	139.71203	4160
1	Oshiage Holiday Apartment	Sumida	35.71721	139.82596	10948
2	Private apt in central Tokyo #203	Nerima	35.74267	139.65810	4270
3	Cozy flat #203, local area YET 10 mins to shib...	Setagaya	35.66344	139.65593	7007
4	Private apt in central Tokyo #201	Nerima	35.74264	139.65832	4051
5	Kero-kero house room 1	Arakawa	35.73818	139.77009	7663
6	RELAX AT HOME:Shimo 5'/Shibuya 10'	Setagaya	35.66832	139.65159	8977
7	Classy room @Shinjuku, Takadanoba	Shinjuku	35.70865	139.69681	5145
8	4F - Near Shinjuku & Shibuya w/Free WiFi	Shibuya	35.67968	139.67949	9853
9	WOMAN ONLY LICENSED ! Cosy & Cat behnd Shibuya	Setagaya	35.65833	139.67153	7007

Dataset 1: Tokyo Airbnb listings

Then I aggregated the data to find out the number of listings, average price, and latitude/longitude coordinate of each neighborhood. To add some more spice, I did a quick BeautifulSoup scraping on the Special Wards of Tokyo Wikipedia page containing population and area data of the city's administrative 23 wards.

	Neighborhood	No. of Listings	Average Price	Latitude	Longitude	Population	Area_km2
0	Adachi	214	6728	35.776307	139.790275	674067	53.25
1	Arakawa	344	12019	35.733565	139.780470	213648	10.16
2	Bunkyo	142	21831	35.719431	139.756299	223389	11.29
3	Chiyoda	207	17991	35.696185	139.773024	59441	11.66
4	Chuo	505	21333	35.678838	139.777576	147620	10.21
5	Edogawa	263	17960	35.707953	139.871731	685899	49.90
6	Itabashi	324	15075	35.758386	139.691179	569225	32.22
7	Katsushika	353	13276	35.741874	139.853582	447140	34.80
8	Kita	324	12778	35.753275	139.738731	345063	20.61
9	Koto	224	12372	35.686533	139.818622	502579	40.16
10	Meguro	86	11924	35.637254	139.696739	280283	14.67
11	Minato	577	17468	35.659335	139.739242	248071	20.37
12	Nakano	421	14866	35.701728	139.671702	332902	15.59
13	Nerima	90	7683	35.750107	139.646873	726748	48.08
14	Ota	607	11958	35.565070	139.719751	722608	60.66
15	Setagaya	331	12937	35.647916	139.655358	910868	58.05
16	Shibuya	946	18541	35.668130	139.693488	227850	15.11
17	Shinagawa	194	14670	35.614495	139.724895	392492	22.84
18	Shinjuku	2252	17364	35.698916	139.705109	339211	18.22
19	Suginami	350	11729	35.697518	139.635722	570483	34.06
20	Sumida	1199	17695	35.707154	139.808853	260358	13.77
21	Taito	1922	16923	35.713836	139.788075	200486	10.11
22	Toshima	1492	15770	35.732756	139.715963	294673	13.01

Dataset 2: Aggregated attributes for each neighborhood

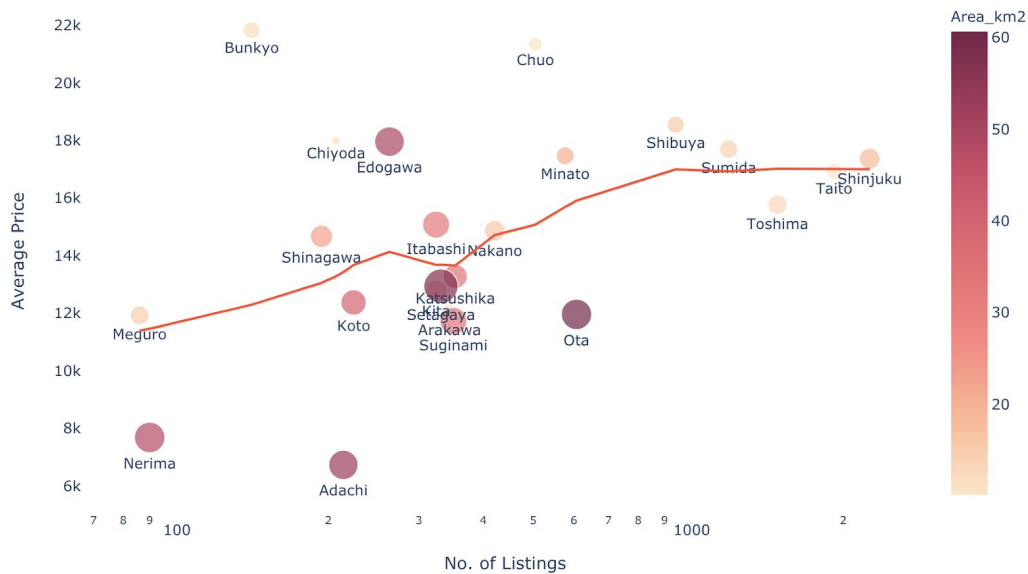
Based on the coordinates, I retrieved a maximum of 100 recommended places for each area within walking distance radius (less than 1km) using Foursquare API. There might be some areas that are less concentrated than the others, therefore having fewer spots surrounding. I found it interesting to see the obtained results as I didn't know some quite nice places around despite living within the area for years now. Displayed below are the first few rows of 2216 records collected. We would dig more on the venue category to have a glance at the surroundings and answer the project question. So let's get going to further analysis.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2011	Sumida	35.707154	139.808853	蕎麦前 小まつ	35.699438	139.812544	Soba Restaurant
781	Kita	35.753275	139.738731	Key's Cafe	35.753933	139.737430	Café
1430	Setagaya	35.647916	139.655358	MERCI BAKE (メルシーベイク)	35.643533	139.655419	Cupcake Shop
2159	Toshima	35.732756	139.715963	BAKE CHEESE TART	35.730315	139.711410	Pastry Shop
1563	Shibuya	35.668130	139.693488	LINE CUBE SHIBUYA	35.664100	139.698520	Concert Hall
1904	Suginami	35.697518	139.635722	川名肉店	35.693494	139.639707	Deli / Bodega
487	Chuo	35.678838	139.777576	Snoopy Town mini (スノーピータウンミニ)	35.681313	139.769319	Gift Shop
406	Chuo	35.678838	139.777576	Pokémon Cafe (ポケモンカフェ)	35.680620	139.774372	Café
1757	Shinjuku	35.698916	139.705109	TOHO Cinemas (TOHOシネマス 新宿)	35.694856	139.702018	Multiplex
1704	Shinagawa	35.614495	139.724895	SPORU (スポル大井町)	35.608368	139.731697	Athletics & Sports

Dataset 3: List of recommended spots in each vicinity

### 3. Methodology and Exploratory Data Analysis

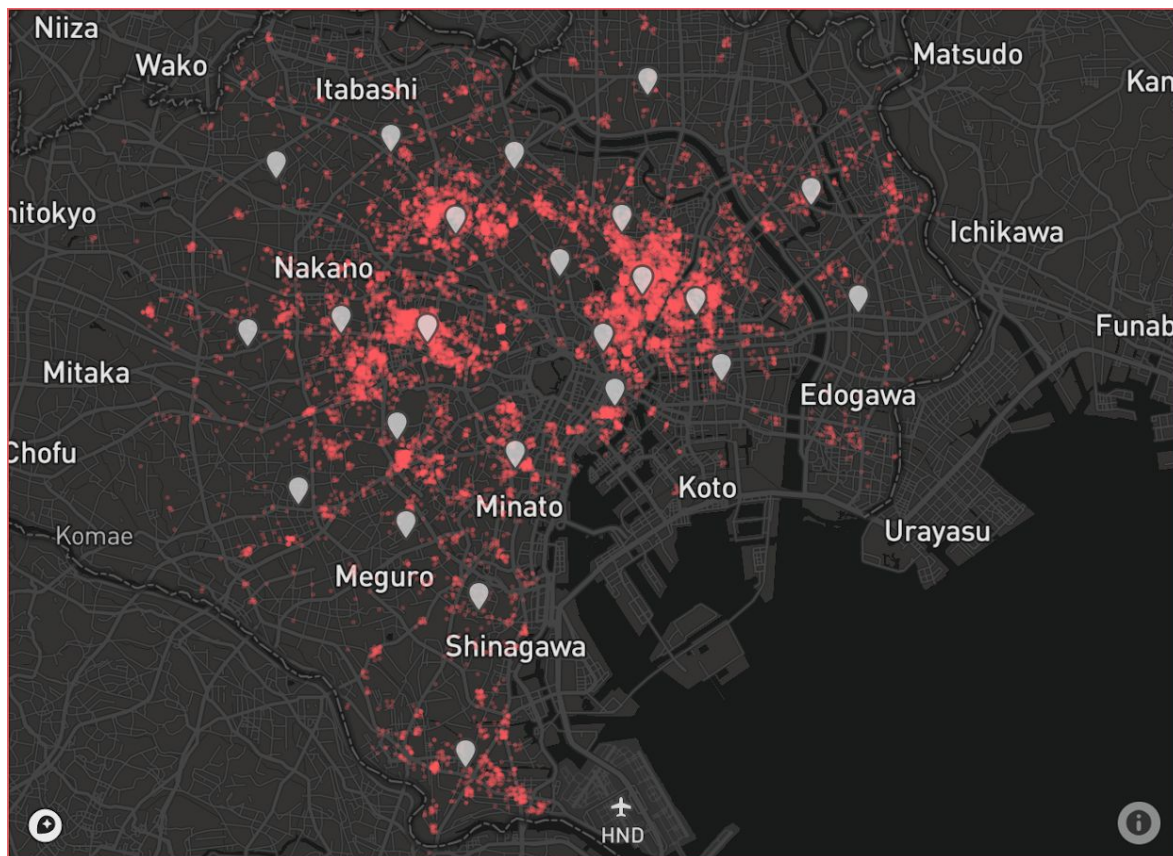
The chart below sums up all the information and gives us a glimpse of the comparison between the wards. Note that the size of the circle indicates the number of population and the horizontal axis is log-scaled.



Overall, Tokyo listings cost 12k~15k JPY (\$110~140) per night on average. It's relatively equal or lower than average in other major cities in the world but it's been actually getting pricier since last year due to rising demand ahead of Olympics Games this summer.

Area and population seem to be well-correlated as **Nerima**, **Adachi**, **Edogawa**, and **Ota** are in roughly equal size and color. It's also interesting to see **Bunkyo** as the most expensive ward although it sits at bottom 3 in the number of listings. Besides, a spike on **Edogawa** price despite being on par with the 2 cheapest regions in terms of population and area, is probably due to its easy access to Tokyo Disney Resort as well as to the city center. 5 outliers are easily identified for having  $\pm 1000\sim$  listings; there seems to be an apparent reason for that which we'll see in the next few lines.

You might be wondering where most of Tokyo's Airbnb listings are concentrated. So here's the map showing the apartment locations and the 23 special wards pointed by the marker icons.



Interactive visualization of this map can be accessed [here](#)

In dataset 3, there are 242 unique categories with 'Convenience Store', 'Ramen Restaurant', 'Japanese Restaurant' (*you don't say*), 'Cafe', and 'Sake Bar' top the list. Random categories from 'Well', 'Belarusian Restaurant', to 'Intersection' are also found in the list.

To figure out how the neighborhood looks like, I listed up the most common venue categories in each ward then used the relative frequency values for the segmenting part. The aggregated table is a sparse  $23 \times 243$  matrix. This time I would just fit the k-means clustering algorithm to find similar

neighborhood groups as it's common and powerful for unsupervised tasks. I choose 4 as the optimum number of k after examining the elbow method and silhouette score. Other features in dataset 2 aren't included in fitting the model to avoid demographic and spatial noise.

#### 4. Result

The features in dataset 2 are instead useful in comparing the difference between clusters as shown in the table below:

	Cluster Labels	No. of Listings	Average Price	Latitude	Longitude	Population	Area_km2
0	1	274.111111	12314.222222	35.728435	139.760758	563915.222222	38.581111
1	2	995.800000	15589.800000	35.678931	139.717110	349767.000000	23.828000
2	3	872.500000	17314.000000	35.677424	139.750662	227444.666667	14.040000
3	4	228.666667	16076.666667	35.677148	139.705639	395454.666667	22.730000

The first and second cluster look like this,

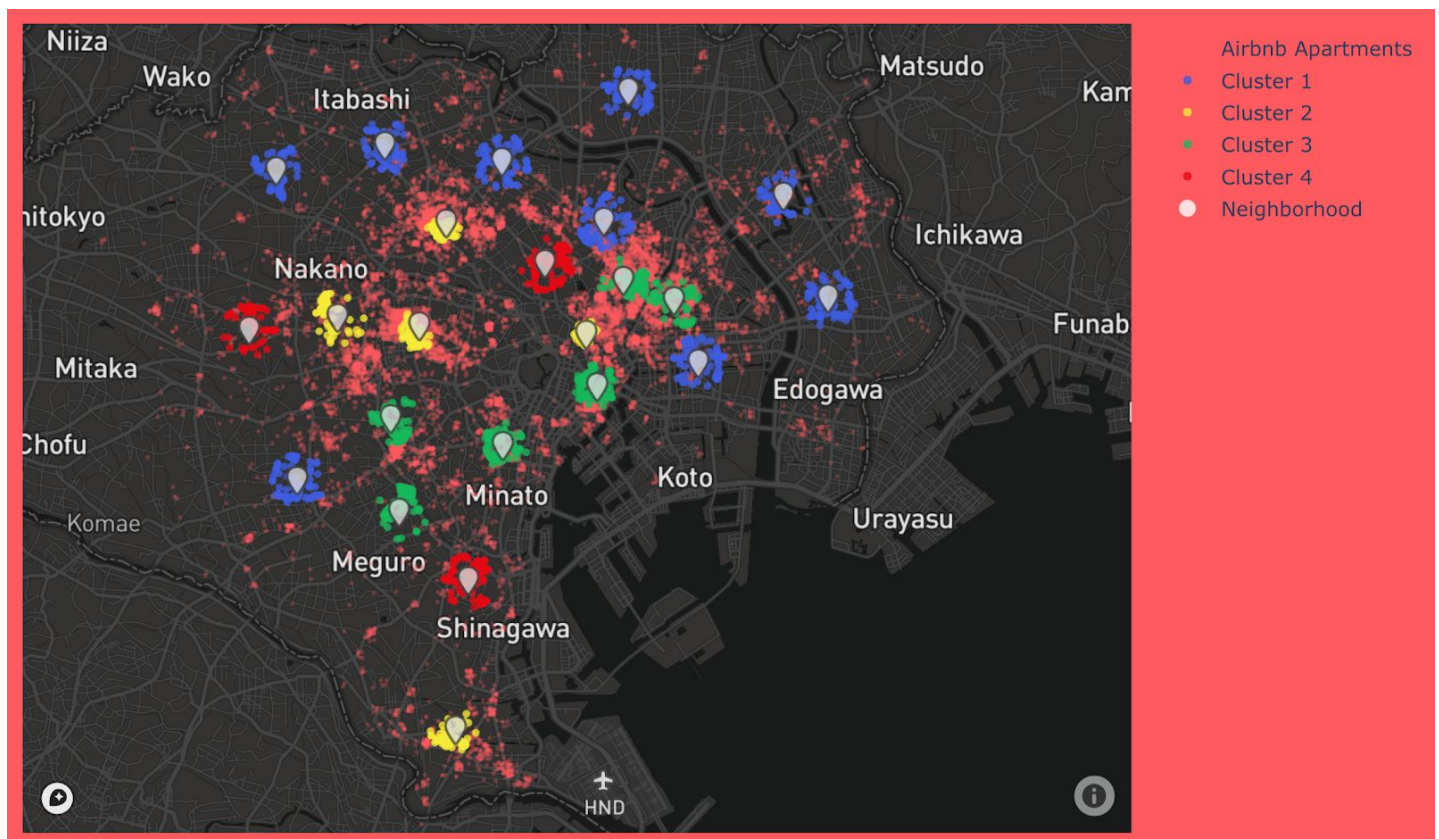
	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adachi	1	Convenience Store	Ramen Restaurant	Discount Store	Pharmacy	Café
1	Arakawa	1	Convenience Store	Ramen Restaurant	Grocery Store	Donburi Restaurant	Café
15	Setagaya	1	Convenience Store	Ramen Restaurant	Grocery Store	Café	Tram Station
5	Edogawa	1	Convenience Store	Intersection	Grocery Store	Park	Electronics Store
6	Itabashi	1	Convenience Store	Ramen Restaurant	Intersection	Supermarket	Fast Food Restaurant
7	Katsushika	1	Convenience Store	Sake Bar	Chinese Restaurant	Supermarket	Sushi Restaurant
8	Kita	1	Convenience Store	Park	Ramen Restaurant	Grocery Store	Intersection
9	Koto	1	Convenience Store	Supermarket	Ramen Restaurant	Japanese Restaurant	Grocery Store
13	Nerima	1	Convenience Store	Italian Restaurant	Grocery Store	Café	Park
18	Shinjuku	2	Bar	Korean Restaurant	Sake Bar	Ramen Restaurant	BBQ Joint
14	Ota	2	Ramen Restaurant	Chinese Restaurant	Sake Bar	Japanese Restaurant	Convenience Store
12	Nakano	2	Ramen Restaurant	Japanese Restaurant	Sake Bar	BBQ Joint	Café
22	Toshima	2	Ramen Restaurant	Japanese Restaurant	Sake Bar	Café	Chinese Restaurant
3	Chiyoda	2	Electronics Store	Sake Bar	Hobby Shop	Ramen Restaurant	Café



and here are the remaining third and fourth.

10	Meguro	3	Japanese Restaurant	BBQ Joint	Coffee Shop	Italian Restaurant	Ramen Restaurant
4	Chuo	3	Japanese Restaurant	Soba Restaurant	Ramen Restaurant	French Restaurant	Italian Restaurant
16	Shibuya	3	Café	Italian Restaurant	Japanese Restaurant	Sake Bar	Concert Hall
20	Sumida	3	Ramen Restaurant	Coffee Shop	Café	Scenic Lookout	Bakery
21	Taito	3	Coffee Shop	Dessert Shop	Ramen Restaurant	Japanese Restaurant	Hostel
11	Minato	3	Japanese Restaurant	Steakhouse	Ramen Restaurant	Chinese Restaurant	Pizza Place
17	Shinagawa	4	Convenience Store	Ramen Restaurant	Japanese Restaurant	Chinese Restaurant	Café
2	Bunkyo	4	Café	Convenience Store	Coffee Shop	BBQ Joint	Italian Restaurant
19	Suginami	4	Convenience Store	Ramen Restaurant	Park	Sake Bar	Chinese Restaurant

For those who are not familiar with Tokyo, it should be difficult to see the geographical pattern just from looking at the table above, so the following refined visualization might help.



Interactive visualization of this map can be accessed [here](#)

## 5. Discussion

We have observed interesting findings only from a relatively small dataset from basic geospatial data, Airbnb listings, and Foursquare. In the brief check part, you could have a big picture of which neighborhood that suits your needs. For instance, if you want to stay at a place in the common area and are willing to be charged more than average, set your filter search to one of the top 5 wards in the number of listings. Or those who could compromise spending a bit more transits and commute time for cheaper rents may find places around Nerima and Adachi suitable for them.

The clustering result also turned out to be surprisingly good. Without putting geocoordinate data into k-means fitting, the outcome seemingly found the way to form segments that group neighborhood with its close counterpart. If I may include my own domain knowledge, cluster 1 consists of the outskirts, cluster 2 is traveler-friendly, cluster 3 seems to be all-rounder (business, shopping, tourism), cluster 4 is quite random. Seeing further, each cluster could be characterized as below:

- Cluster 1: Few listings, low price, more population, large area, convenience store dominates the hood
- Cluster 2: Many listings, average price, tourists' popular stay area, ramen restaurants and bars are most common
- Cluster 3: Many listings, high price, few population, small area, variety venues
- Cluster 4: Few listings, average price, a mix between residential, student, and tourists' stay area

## 6. Conclusion

When booking Airbnb or other rental lodgings, it might be useful in understanding the characteristics and the surroundings. You know which neighborhoods that may fit and not fit your preferences, by considering location, average price and/or population density, too. Popular areas like Shinjuku, Shibuya, Ikebukuro, Ueno, and Asakusa are also where the rental apartments are condensed. However, less famous neighborhoods are also worth trying and they cost comparatively less; perhaps you'll find hidden gems and local experiences along the way. Tokyo 23 special wards could basically be sectioned into 4 zones (outskirts, travelers-friendly, all-rounder, random) based on common places nearby as explained above. Indeed it could be improved further by involving more variables, expanding the geospatial information, and utilizing more advanced techniques.