The Battle of Neighborhoods

Wenye Ma

February 2020

1 Introduction and Background

1.1 Business Problem

Tokyo is one of the world's leading metropolises, attracting thousands of tourists every day. Meanwhile, with the 2020 Tokyo Olympics approaching, it is profitable for house owners to transform the houses that are available to bed and breakfast. When dealing with such problem, we should consider the visitor flow rates during the day and night as well as surrounding facilities like restaurants and subway line. Moreover, house price should also be taken into consideration during slack season.

In a nutshell, the issue is to find a suitable place with convenient facilities and relatively low estate price. Also, the accommodation places nearby should not already saturated.

1.2 Target People

1. People with several housing in Tokyo city and have the desire to earn money through bnb.
2. Investor who want to set up an inn or run a chain hotel.
3. Tourists from other cities or countries that want to find a good place to stay when traveling to Tokyo.

Figure 1: Panorama of Tokyo

2 Data Preparation

2.1 Web Scrapping

Considering most of the houses is prepared for tourists from other cities and countries who have a tendency to browse the main spots as many as possible, we only focused on Tokyo's central 6 wards: **Bunkyo**, **Chiyoda**, **Chuo**, **Minato**, **Shibuya** and **Shinjuku**.

The web page zip.nowmsg.com contains the postal code of each district and its corresponding location. We used BeautifulSoup to scrap the page and get a table with four cloums: Ward, Major District, Latitude and Longitude.

Some of the places have different postal codes but share the same location. In such circumstance, we clean the data by only leaving a single line.

	Ward	District	Latitude	Longitude
0	Shinjuku	Agebacho	35.7024	139.7430
1	Shinjuku	Aizumicho	35.6902	139.7199
2	Shinjuku	Akagi Motomachi	35.7056	139.7393
3	Shinjuku	Akagi Shitamachi	35.7057	139.7343
4	Shinjuku	Arakicho	35.6902	139.7229
...	...	...	...	...
178	Minato	Shibadaimon	35.6542	139.7534
179	Minato	Shibakoen	35.6578	139.7476
180	Minato	Shirokane	35.6430	139.7269
181	Minato	Shirokanedai	35.6339	139.7297
182	Minato	Takanawa	35.6321	139.7342

183 rows × 4 columns

Figure 2: Neighborhoods and its location

2.2 Other data

From the report published by Japan Tourism Agency in 2014, we are able to view the tourists in each area during the day (6:00 ∼ 21:59) and night (22:00 ∼ 5:59) respectively. We sorted out the data manually.

In addition, from utinokati.com, we are able to get the information of house price as well as the rent of mansion in the 6 wards.

These two data help us make a more comprehensive decision.

2.3 Foursquare Data

Two groups of data is obtained using Foursquare API:
1. Number of venues and their type and location around every neighborhood.
2. Number of hotels around every neighborhood.

3 Methodology

In first step, we counted the number of hotels around each neighborhood in 500m.

By comparing foreign tourists in each ward, we sifted out a bunch of neighborhoods around which the number of hotels is supposed to be saturated. After dropping some neighborhoods, we only left 104 Neighborhood with hotels nearby undersaturated.

	Ward	Neighborhood	population	Latitude	Longitude	Num_hotel	n
0	Shinjuku	Agebacho	7109	35.7024	139.7430	5.0	0.703334
1	Shinjuku	Aizumicho	7109	35.6902	139.7199	4.0	0.562667
2	Shinjuku	Akagi Motomachi	7109	35.7056	139.7393	2.0	0.281334
3	Shinjuku	Akagi Shitamachi	7109	35.7057	139.7343	1.0	0.140667
4	Shinjuku	Arakicho	7109	35.6902	139.7229	4.0	0.562667
...	...	...	...	...	...	...	...
174	Minato	Minamiaoyama	8726	35.6646	139.7155	4.0	0.458400
175	Minato	Motoazabu	8726	35.6565	139.7292	6.0	0.687600
177	Minato	Roppongi Izumigadentawa(1-Kai)	8726	35.6588	139.7344	5.0	0.573000
180	Minato	Shirokane	8726	35.6430	139.7269	3.0	0.343800
181	Minato	Shirokanedai	8726	35.6339	139.7297	3.0	0.343800

104 rows × 7 columns

Figure 3: Neighborhood with Hotels Nearby Undersaturated

The second step is to analyze each neighborhood.

By taking the mean of the frequency of occurrence of each category, we first group rows by neighborhood. After processing, we got the 104 neighborhoods with under saturated hotels nearby and its latitude and longitude. We then sorted the top 10 most common venues of each neighborhood in descending order.

	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	Agebacho	Italian Restaurant	Japanese Restaurant	French Restaurant	BBQ Joint	Sake Bar	Ramen Restaurant	Kaiseki Restaurant	Chinese Restaurant	Soba Restaurant	Yakitori Restaurant
1	Aizumicho	Sake Bar	Ramen Restaurant	Convenience Store	BBQ Joint	Café	Climbing Gym	Coffee Shop	Japanese Restaurant	Bar	Chinese Restaurant
2	Akagi Motomachi	Italian Restaurant	Japanese Restaurant	French Restaurant	Sake Bar	Convenience Store	Café	Bakery	BBQ Joint	Steakhouse	Unagi Restaurant
3	Akagi Shitamachi	Convenience Store	Italian Restaurant	Japanese Restaurant	Bakery	Chinese Restaurant	Dessert Shop	Café	Sake Bar	Indian Restaurant	Bar
4	Arakicho	Sake Bar	Convenience Store	Ramen Restaurant	BBQ Joint	Japanese Restaurant	Café	Grocery Store	Bar	Rock Club	Climbing Gym

Figure 4: 10 Most Common Venues of Each Neighborhood

In the third step, we used k-means, a machine learning method to cluster the neighborhood into several clusters.

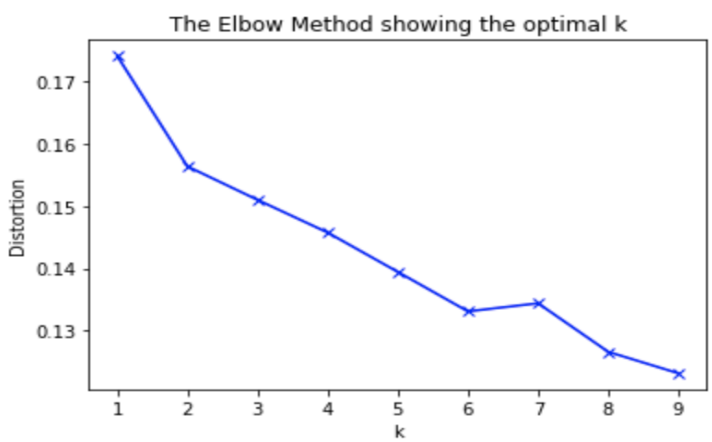

Figure 5: Elbow Method

Through the elbow method, we set k to be 2. After creating a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

We were able to visualize the distribution of each neighborhoods with different clusters.

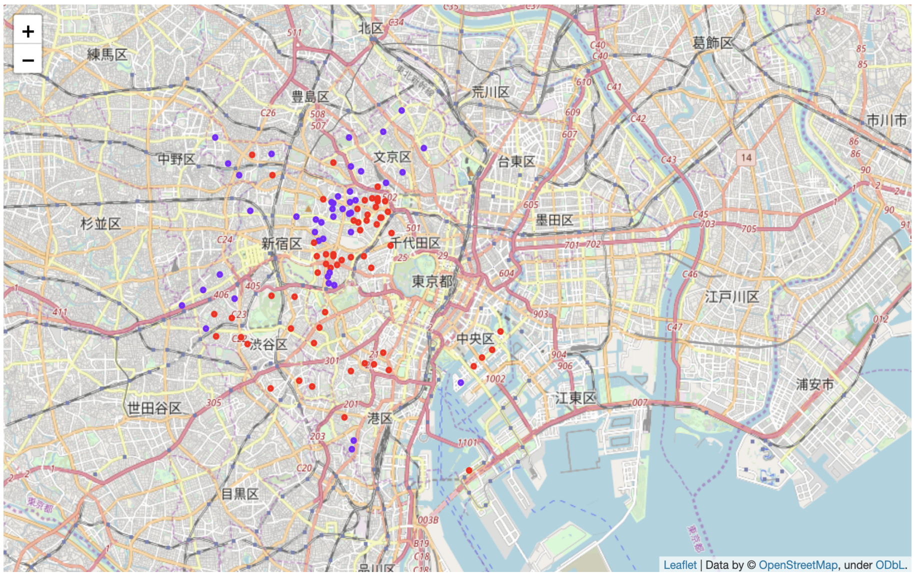

Figure 6: Clustered Map

4 Analysis

Up to now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

And then we have a roughly view on the rental price of every ward. Areas with lower rent is more suitable to be built as bed and breakfast. For people who want to decide whether to change a house he have already owned, rental price can help him to weigh the advantages and disadvantages with renting out his house to local people, because during slack seasons, homes without tenant can become a loss. For people who want to invest in this field (buy a new house), rental price can also reflect the house price of each place.

5 Results

We combine the choropleth map with the clusters and the result is as below.

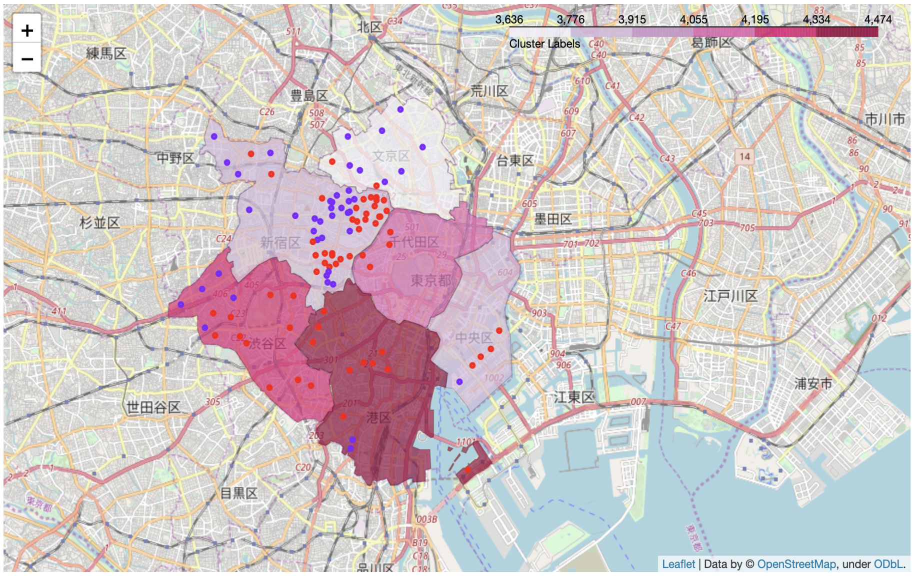

Figure 7: Choropleth Map

6 Discussion

While western restaurants, bars and Café is common as the 1st most common venues of cluster 0, all of the cluster 1 neighborhoods' 1st common venue is convenience store (except 1 bus store).

Undoubtedly, for most of the tourists who go out early and get home late at night (in order to visit places as many as possible), convenience store is an ideal place for them to pick up food and have on the road. (While Italian restaurants, gyms and baseball stadiums are generally for natives.)

Considering the high rent in other wards, purple points (cluster 1) located in Shinjuku and Bunkyo can be sensible choices to set up a bed and breakfast.

To be more specific, we collected the data of 40 neighborhoods in cluster 1. (Unable to find other 3 neighborhoods' data) We also excluded Toyomicho because it is 20 min away from the nearest subway station, which is not suitable for tourists.

7 Conclusion

In general, Shinjuku and Bunkyo are the most suitable places to set bed and breakfasts since they have a lot of neighborhoods with convenient facilities and relatively low rent cost.

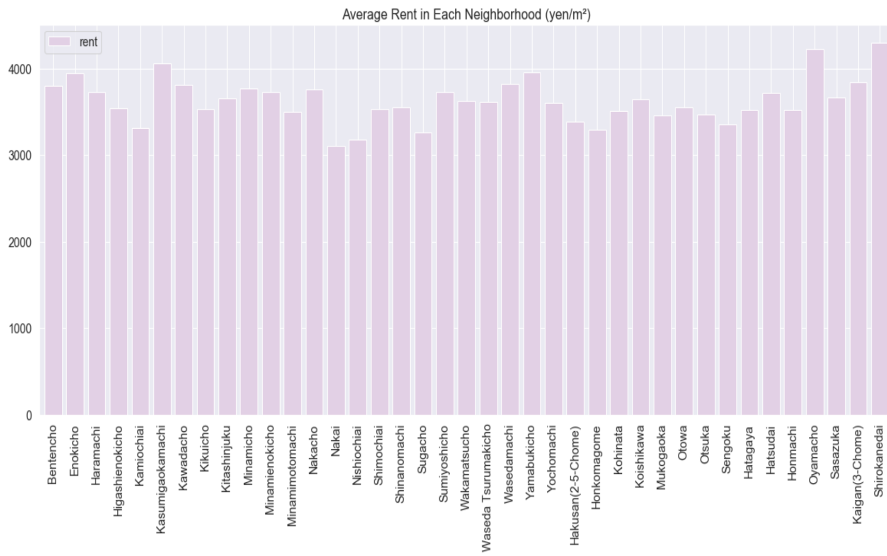

Figure 8: Average Rent in Each Neighborhood

There is no absolute optimal place to set the bed and breakfast, but we indeed can combine the location and the rent price to consider using the $map_{r}esults$ about. (For example, Sugacho is a pretty good place with cheap rent and the location near the center of city)

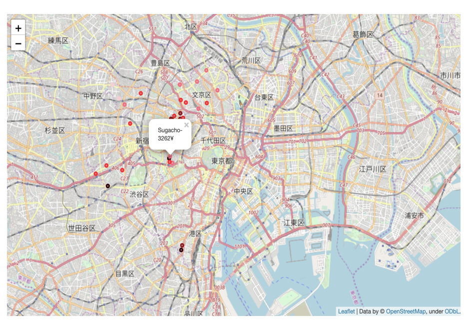

Figure 9: Final Results

Of course, there are still some disadvantages through our analyze, like we didn't consider whether the subway line nearby is just a normal station or a large station.To improve, we can also use other machine learning methods like DBSCAN to get a more precise result.