ANALYSIS OF VENUES FOR FOREIGN VISITORS IN TOKYO

Ayako Nagao
Coursera IBM DataScience
Capstone project

1.INTRODUCTION

As the population of Japan is shrinking rapidly, it has been said that our economic growth might shrink as well in the future. The Japanese government has been making great efforts to promote the inbound tourism to boost our economy as it is one of the areas that we can expect the growth.
Tokyo is hosting Summer Olympic and Paralympic Games in 2021. We are expecting around 10 million foreign visitors for both games. It will be a great chance to promote the city as their holiday destination.
So, before our big day comes, I want to know if we have enough venues for them to enjoy their stay. It is important to know which areas do not have enough venues to rest, to eat, to get basic amenities, cash so on.

In this IBM DataScience capstone project, we are to use location data from FourSquare API to solve the problems.
I would like to find out which areas have to be developed better for the foreign tourists staying in Tokyo.

2.DATA

In order to collect useful data,

1) we need to define the basic venues for foreign tourists. I chose 8 venues.
   - Hotels
   - Restaurants (all categories)
   - Convenience stores
   - ATMs/Exchanges (only ATMs which accepts international cards)
   - Cafes
   - Parks
   - Pharmacies
   - Tourist Information Center

2) we need the geographical coordinates of all the stations (all the lines) in Tokyo. Most of the tourists may travel by trains or subways because traveling with cars and taxis in Tokyo are not popular options as there are limited parking spaces and using taxis is pricey. I scraped the station names from Wikipedia and obtained latitude and longitude through GeoPy library. (Table 2-1 / Map 2-1)

	station	latitude	longitude
0	Adachi-Odai Station	35.754801	139.770404
1	Aihara Station	35.606819	139.331686
2	Akabane Station	35.778139	139.720800
3	Akabane-iwabuchi Station	35.782968	139.719853
4	Akabanebashi Station	35.654987	139.743891
...	...	...	...
637	Zōshiki Station	35.550041	139.715202
638	Itabashi Kuyakushomae Station	35.751809	139.709718
639	Shimo-Shimmei Station	35.608871	139.726232
640	Shōin-Jinjamae Station	35.644143	139.655263
641	Yaguchinowatashi Station	35.562720	139.700291

642 rows × 3 columns

Table 2-1 Geographical Coordinates of Stations in Tokyo

Map 2-1 Map of Stations in Tokyo

3) Search each basic venue data from the all the stations using FourSquareAPI. I used 'search' endpoint, which returns maximum 50 data that matches its categoryId. Note that FourSquareAPI returns some data categorized wrongly so it needs to be checked if the data represents its category and it is divided into too many categories so I re-categorized them into one class for each. (e.g. Fast-food restaurant, Japanese restaurants, French restaurant, …… → 'Restaurants'). And as for the ATMs data, in Japan not all the ATMs accepts international cards. In addition to 7-Eleven and Post office (where the international cards can be used), I used the E.net website to scrape the name of the locations of their ATMs and added them to the data. To align with other data, the stations that have more than 50 data because of this addition, I dropped the exceeded data. Now all the station names have 0 to 50 data for each basic venue. (Table 2-2)

	station	id	name	venue_lat	venue_lng	categories	classes
0	Adachi-Odai Station	4d21b00af7a9a1437e48389f	Ryota Kuga's Guest House	35.746466	139.771546	Hostel	Hotels
1	Adachi-Odai Station	59a382f5cad1b628d2f030f5	東京ゲストハウス2020	35.743168	139.771271	Hostel	Hotels
2	Adachi-Odai Station	53f45d47498e66d089216755	Tokyo Guest House B&B Hostel (東京ゲストハウス B&B ホステル)	35.743172	139.771286	Hostel	Hotels
3	Aihara Station	4ce533fa5fce5481a53d5aaa	Laxio-Inn (ホテル ラクシオ・イン)	35.610112	139.344740	Hotel	Hotels
4	Aihara Station	5b6c2cf9e65d0c002ced1699	ホテル anniversary	35.611662	139.343505	Hotel	Hotels
...	...	...	...	...	...	...	...
139987	Yushima Station	5b0e7ba6a2a6ce002c79fc2b	Ueno Information Center	35.710779	139.775775	Tourist Information Center	Tourist Information Center
139988	Yushima Station	4ce14895c9a0a0903596246a	Tokyo Tourist Information Center (東京観光情報センター)	35.710783	139.773482	Tourist Information Center	Tourist Information Center
139989	Yushima Station	57368551498ed9b3ec5509db	Park Information Center (公園案内所)	35.714783	139.775900	Tourist Information Center	Tourist Information Center
139990	Yushima Station	5e4256344c4a85000888edc8	General Information Center (総合案内所)	35.716144	139.772131	Tourist Information Center	Tourist Information Center
139991	Zoshigaya Station	5bca6a8c3c858d002c6e93ac	Tobu Tourist Information Center (東武ツーリストインフォメー...)	35.730572	139.710577	Tourist Information Center	Tourist Information Center

139992 rows × 7 columns

Table 2-2 List of Venues for Tourists for each station

3.METHODOLOGY

In this project, we want to find out, which areas must be developed better for the foreign tourists staying in Tokyo.
To answer the question, we need to know, when they arrive in any stations in Tokyo, how many of those basic venues they can find in walking distance. I took 4 steps below.

1) Group the data by stations and count how many each basic venue they have in walking distance.
I transformed the 'classes' column to numerical data (0 or 1) and counted them for each station. (Table 3-1)

	station	Hotels	Restaurants	Convenience stores	ATMs/Exchanges	Cafes	Parks	Pharmacies	Tourist Information Center
0	Adachi-Odai Station	3	50	18	14	16	19	9	0
1	Aihara Station	2	40	10	7	3	11	3	0
2	Akabane Station	22	50	50	31	46	17	30	0
3	Akabane-iwabuchi Station	22	50	50	29	46	14	30	0
4	Akabanebashi Station	45	49	50	41	47	46	49	0
...	...	...	...	...	...	...	...	...	...
637	Ōtorii Station	18	50	41	29	17	43	17	0
638	Ōtsuka Station	39	50	50	38	49	50	49	0
639	Ōtsuka-Teikyō-Daigaku Station	3	48	14	12	7	23	1	0
640	Ōtsuka-ekimae Station	38	50	50	43	49	50	49	0
641	Ōyama Station	11	50	48	40	49	20	25	0

642 rows × 9 columns

Table 3-1 Venues Counts grouped by stations

2) Cluster the stations according to the counts.
   To cluster the stations, I used KMean clustering algorithm from scikit-learn library. After several attempts, choosing 4 center points (the number of the cluster) and scaling the counts between 0 to 1 gave me the most distinct clusters.

3) Visualize the cluster on the map and see where they locate.

Map 3-1 Clusters of Stations (Green: Cluster 0/ Red: Cluster 1/ Light Blue: Cluster 2/ Blue: Cluster 3)

4) Find out any area we can develop better.
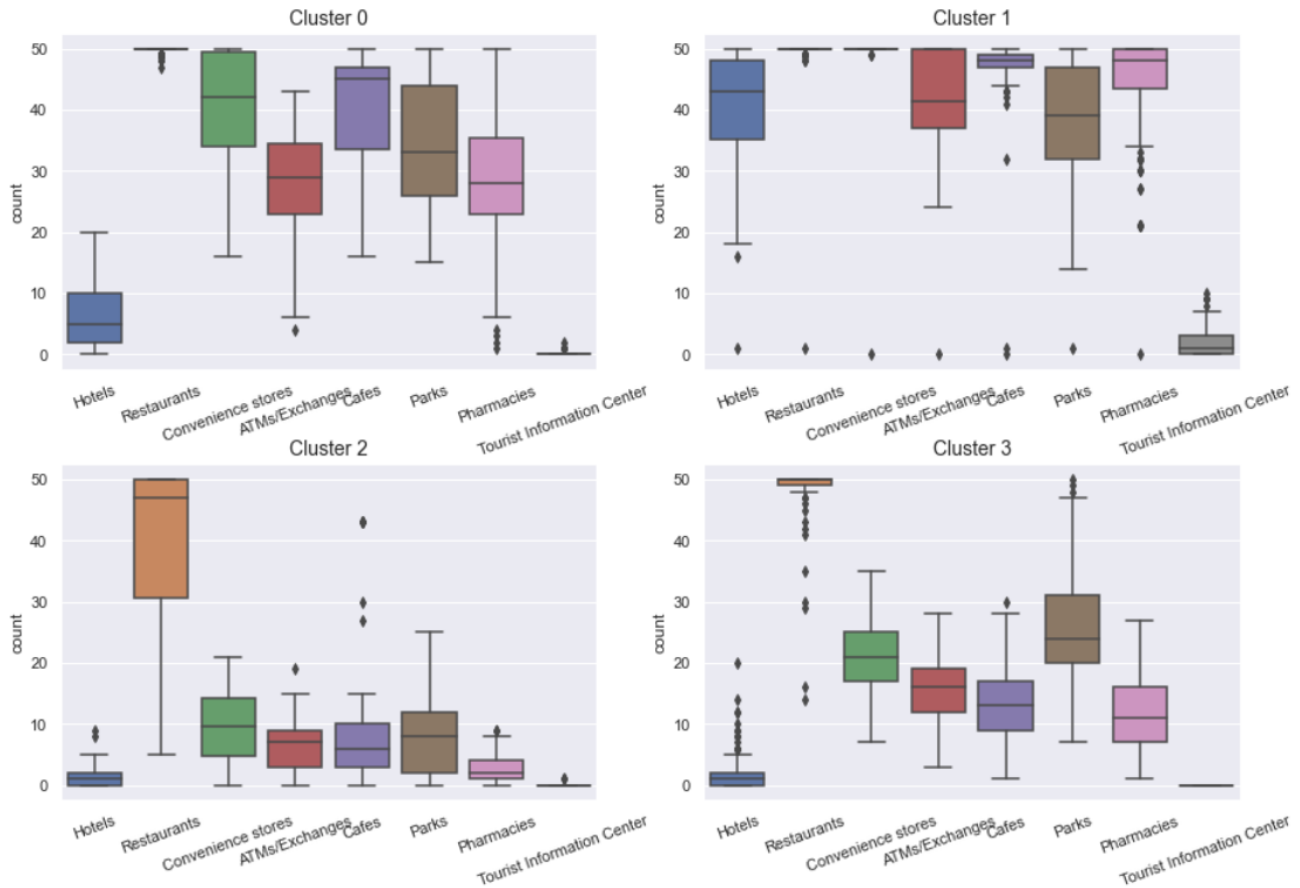   I created boxplots so that we can see which venues are in short in each cluster.

Figure 3-1 Boxplots of Venue Counts for each cluster

4.RESULTS

The stations in **cluster 0** (green on the map) are distributed in the central area just outside of cluster 1. They have relatively high number of most of the venues but hotels and tourist information center. If they have more hotels, we can say they have enough venues for the tourists.

The stations in **Cluster 1** *(red on the map)* are located at the central area which is around 7km radius from the center. This is only area that has nearly 50 or more hotels around each station. Any other venues are also high in number. We can say this area is highly developed for the tourists.

The stations in **cluster 2** (light blue on the map) are mostly located outskirt of Tokyo. There are a few hotels and not so many of the other venues except restaurants. Many of the stations in this cluster are surrounded by mountains and river. They have much fewer population comparing to the central Tokyo. We cannot say they have enough venues, but we may be able to target the tourists who prefer the nature and quiet surroundings in addition to the proximity to the central Tokyo if we could develop more hotels and tourist information.

The stations in **Cluster 3** *(blue on the map)* are mostly located outside of the central area (further than 12km radius from the center), many of the stations are in residential area. They have high number of restaurants, but other venues are relatively low. We cannot say they have enough venues for tourists.

Let's look closely at the distributions of each venues.

(Hotels) Hotels are located mostly in central Tokyo. **(Restaurants)** we can see most of the stations in Tokyo have 50 or more restaurants in walking distance. **(Convenience stores)** in cluster 1, nearly 100% of stations have 50 or more convenience stores, in cluster 2 which located in outskirt of Tokyo has less but at least 10 stores in average is more than enough. **(ATMs)** One of the issues the tourists complain is difficulty of getting cash as lots of small shops do not accept the credit card. As I mentioned before, in Japan finding ATM sign does not mean you can get cash with international cards. The numbers here are the count of only ATM for both domestic and international cards. Cluster 0 and 1(in the central area) have enough but if they go further away from the center, the less they will find the machine they can use. **(Cafés)** Cafes are a good place to rest during the sightseeing or shopping, use Wi-Fi, having breakfast etc. we ideally have many options. Cluster 2 and 3 have much less cafes compare to the stations in the central area. **(Parks)** if they stay for long period, some tourists may start the day with exercising or if traveling with children, they need some place to play around as Tokyo is crowded city with lots of traffic. The plot shows in cluster 0,1 and 3 they can find enough parks, cluster 2 has less but they may find the other alternatives like trails or outdoor recreation venues. **(Pharmacy)** The reason why I included the pharmacy as a basic venue is, in Japan most of the pharmacy sells wide range of toiletries, baby amenities, even food other than medicines, it would be really helpful during their stay. In recent years we saw the growing number of tourists in the drugstores/some offer tax free price) In the plot, again many of pharmacies are concentrated in the central area. **(Tourist Information Center)** Only 33 tourist information centers were found in Tokyo and majority of them are only in cluster 1. The number is way too small if we want to promote the inbound tourism.

5.DISCUSSION
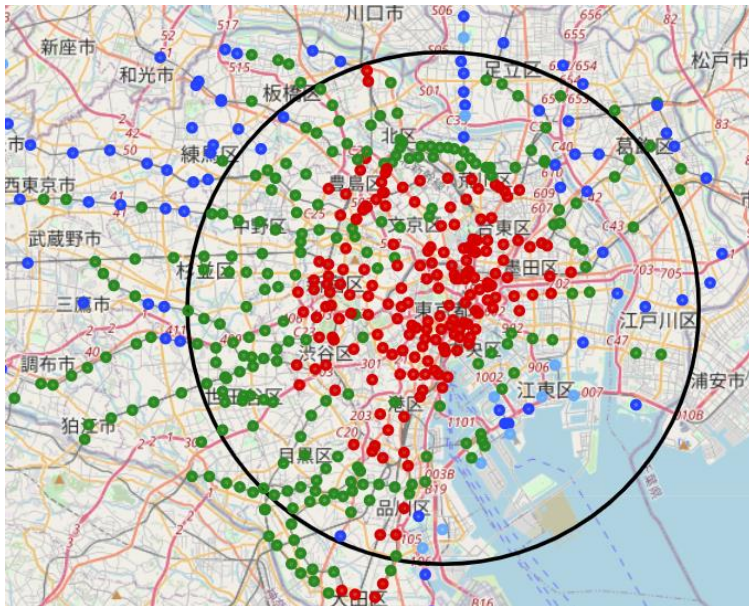
Map 5-1 Stations within 12km radius from Tokyo station

The area circled on the map shows 15km radius from the center. I would like to focus on this area for the development because this is the maximum distance where the tourists can go to the center of Tokyo within 50 to 60 minutes by train. As we can see in the circle, there are some stations categorized same as the stations in suburban area. *The area we have to consider developing better is cluster 2(light blue) and cluster 3(blue) in this*

circle. Those are the stations which are close to the central Tokyo but have not been developed well. Those stations have very few hotels and other venues in walking distance even quite close to the central area. Let's look at few examples in this circle.
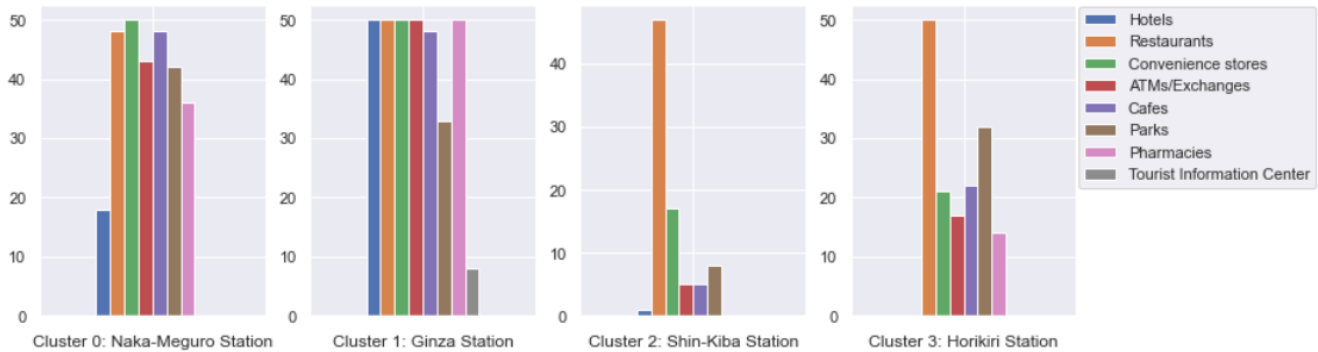

*Figure 5-1 Example Data for each cluster*

I just picked up 4 stations for each cluster in the circle as examples and we can see the huge difference. The stations in Cluster 0 and 1 have all the venues they need for the tourists. But the stations in Cluster 2 and 3 have close to zero hotels and the number of other venues is also small.

6.CONCLUSION

In the previous Discussion section, I circled the area with 12km radius from the center in Tokyo and found around 30 stations falling short of basic venues within (Table 6-1, Map 6-1). This is where we have rooms for the development. 12km seems too long to travel to downtown but Tokyo is a city with huge population (around 140 million in 2020) and as we saw it in the data, many venues are concentrated in the central area. Therefore, it is important to distribute the venues in wider area and develop the better transportation so that they can have options to travel to downtown easy and fast but can stay in less crowded area.

	station	latitude	longitude	cluster		station	latitude	longitude	cluster
0	Adachi-Odai Station	35.754801	139.770404	3	17	Minami-Sunamachi Station	35.668387	139.831870	3
1	Ariake Station	35.634574	139.793294	3	18	Nishi-magome Station	35.586864	139.705982	3
2	Ariake-Tennis-no-mori Station	35.640103	139.788792	3	19	Nishiaraidaishi-nishi Station	35.781512	139.770084	3
3	Ayase Station	35.761651	139.824060	3	20	Ōgi-ōhashi Station	35.763969	139.770826	2
4	Funabori Station	35.684234	139.863792	3	21	Ohanajaya Station	35.747612	139.840249	3
5	Gotanno Station	35.765959	139.809497	3	22	Ōi Keibajō Mae Station	35.595117	139.747054	3
6	Higashi-ojima Station	35.688918	139.849796	3	23	Ryūtsū Center Station	35.581667	139.749124	2
7	Hikawadai Station	35.749860	139.665078	3	24	Shijō-mae Station	35.645729	139.785810	2
8	Hirai Station	35.706421	139.842500	3	25	Shin-Kiba Station	35.646146	139.827286	2
9	Horikiri Station	35.743418	139.817464	3	26	Shinonome Station	35.640730	139.803624	3
10	Horikirishōbuen Station	35.747675	139.827483	3	27	Shiomi Station	35.658929	139.817193	2
11	Kami-Itabashi Station	35.763474	139.676543	3	28	Tatsumi Station	35.645477	139.810739	2
12	Kanegafuchi Station	35.733735	139.820523	3	29	Telecom Center Station	35.617432	139.778911	2
13	Kasai-Rinkai Park Station	35.642211	139.859483	3	30	Tokyo Big Sight Station	35.629763	139.793982	2
14	Kōhoku Station	35.773860	139.770186	3	31	Tokyo Freight Terminal	35.589666	139.759468	2
15	Kokusai-Tenjijō Station	35.634437	139.791755	3	32	Yazaike Station	35.786641	139.768915	3
16	Kōya Station	35.768510	139.770699	3	33	Yotsugi Station	35.732520	139.834320	3

Table 6-1 List of Stations where we can develop better
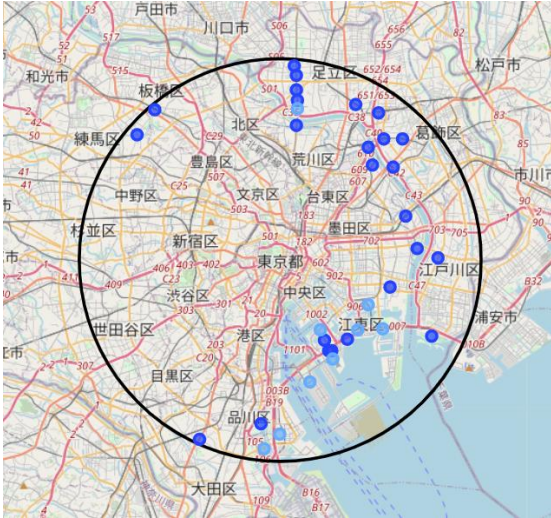
Map 6-1 Stations where we can develop better

And I would like to add that while I was doing this project, I searched many websites regarding the travel experience in Tokyo. I found that It is important to develop the infrastructure for the tourism such as hotels, transportations, restaurants, payment systems etc., but also, we have to solve the issues such as communication in English, flexibility for different culture as well. When we develop the venues or new system for them, it is critical to collect enough data listening to the foreign tourists for issues that they encountered, and we have to make targeted approach to the issues for different groups of tourists. It may give us bigger chances to succeed in the inbound tourism.

Finally, I used FourSquareAPI throughout this project for data collection, because of the limitation of data they can return (up to 50 for each call) for Personal account, the number of the venues is not completely precise (some stations could have more). Also, its categories are sometimes inaccurate, so I corrected as much as I could but there may be some miscategorized data.

All my codes can be found at :
https://nbviewer.jupyter.org/github/belanello/Coursera_Capstone/blob/master/data_collection.ipynb
https://nbviewer.jupyter.org/github/belanello/Coursera_Capstone/blob/master/data_preparation_modeling.ipynb

7.REFERENCE

FOURSQUARE developers
https://developer.foursquare.com/
Railway stations in Tokyo
https://en.wikipedia.org/w/index.php?title=Category:Railway_stations_in_Tokyo
TOKYOCHEAPO
https://tokyocheapo.com/business/withdrawing-cash-japan-atm-credit-card/
statista
https://www.statista.com/topics/4875/travel-and-tourism-industry-in-japan/
TOURISMBUILDER.COM
http://tourismbuilder.com/common-problems-for-tourists-visiting-japan/
E.net
https://pkg.navitime.co.jp/enet_en/address/list?address=13&search=address&lang=en