The Battle of Neighborhoods

Picking up a location in Tokyo for opening a new business(Week 2)

Background

Tokyo is a very busy city, best known for tourist attractions and business innovation. Some investors are willing to open new business in Tokyo, but always not sure about the best location for the new venue. To survive in the competitive competition, one needs to understand the critical factors that contributes to the profitability.

Data

The special wards (特別区 tokubetsu-ku) are 23 municipalities that together make up the core and the most populous part of Tokyo, Japan.

First, I get the information about the special wards through Wikipedia page : https://en.wikipedia.org/wiki/Special_wards_of_Tokyo.

```
response = requests.get('https://en.wikipedia.org/wiki/Special_wards_of_Tokyo').text
soup = BeautifulSoup(response,"lxml")
Table = soup.find("table",{"class": "wikitable sortable"})
```

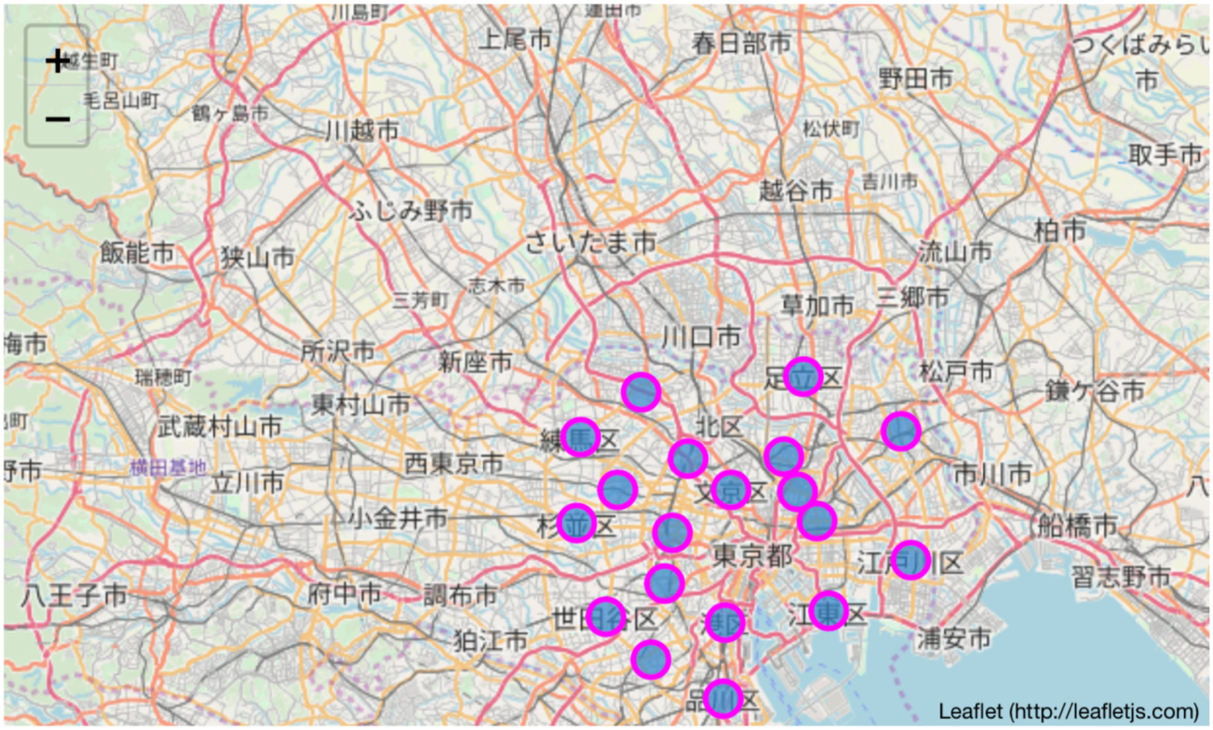
I use BeautifulSoup and pandas to create the initial data frame:

	Name	Population	Area	Major_district
3	Minato	0248,071	12,180	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
4	Shinjuku	0339,211	18,620	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
5	Bunkyō	0223,389	19,790	Hongō, Yayoi, Hakusan
6	Taitō	0200,486	19,830	Ueno, Asakusa
7	Sumida	0260,358	18,910	Kinshichō, Morishita, Ryōgoku

Next, I use the Geocoder library to obtain the coordinator variable.

	Name	Population	Area	Major_district	Latitude	Longitude
3	Minato	0248,071	12,180	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055
4	Shinjuku	0339,211	18,620	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693763	139.703632
5	Bunkyō	0223,389	19,790	Hongō, Yayoi, Hakusan	35.718810	139.744732
6	Taitō	0200,486	19,830	Ueno, Asakusa	35.717450	139.790859
7	Sumida	0260,358	18,910	Kinshichō, Morishita, Ryōgoku	35.700429	139.805017

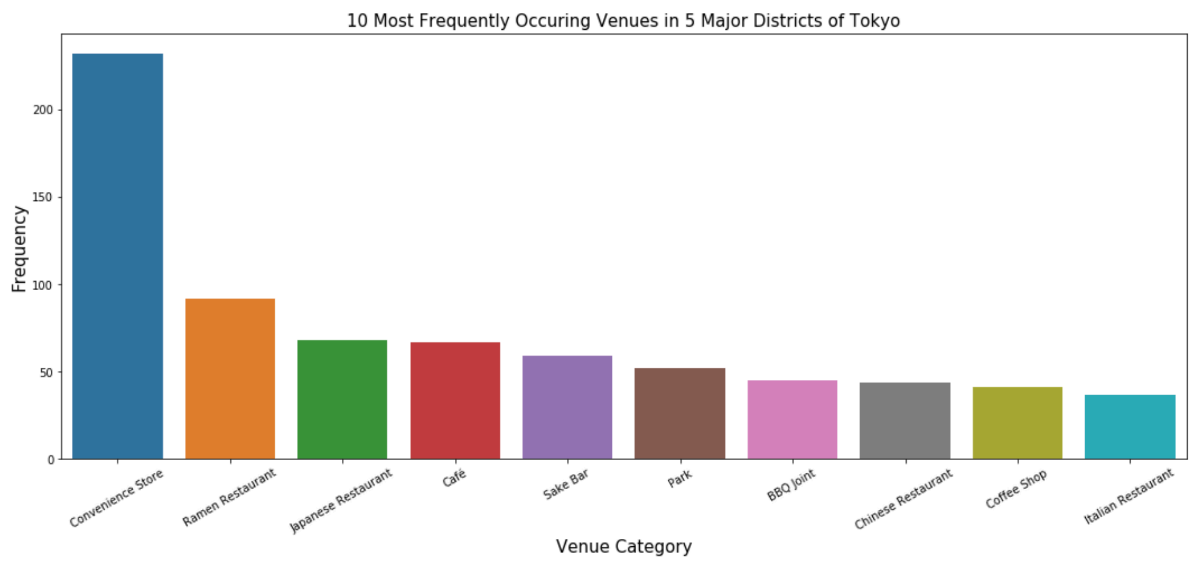
And then I generate a map to display the venues in our dataset:



Generally speaking, the store in Tokyo is dispersed geographically. Then, I want to check the type of store in Tokyo.

Then, I make use of Foursquare API to obtain the 100 most common venues within 1 kilometer of each major district. Here we have used a pipeline to channel the json file from search to Pandas DataFrame. Note the last column is 'Venue Category.' The API requires GPS coordinates of the neighborhoods.

	District	Dist_Latitude	Dist_Longitude	Venue	Venue_Lat	Venue_Long	Venue_Category
0	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Matsushimaya (松島屋)	35.640579	139.737529	Wagashi Place
1	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Maison Kayser (メゾンカイザー)	35.643569	139.735952	Bakery
2	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Imafuku (今福)	35.645379	139.734142	Sukiyaki Restaurant
3	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Ramen Jiro (ラーメン二郎 三田本店)	35.648053	139.741625	Ramen Restaurant
4	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Maruichi Bagel (マルイチベーグル)	35.645429	139.733250	Bagel Shop
5	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	David's Deli	35.644743	139.737651	Kosher Restaurant
6	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Kamezuka Park (亀塚公園)	35.643334	139.740522	Park
7	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	三味	35.646812	139.743341	Japanese Restaurant
8	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Patisserie Passion de Rose	35.645100	139.734464	Dessert Shop
9	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	Muho (夢呆)	35.644777	139.735465	Soba Restaurant



The convenience store tops the chart and exceed the other types by more than two times. And the rest nine types are all restaurants.

Then, I use One Hot Encoding to get more information about the venue categories.

	District	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Unagi Restaurant	Used Bookstore	Vegetarian / Vegan Restaurant	Video Game Store
0	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	0	0	0	0	0	0	0	0	0	...	0	0	0	0

Finally, I try to cluster these districts into 4 clusters based on the venue categories through K-Means clustering from Scikit-learn package.

This map only contains three different clusters, after checking each of the cluster, we know that the missing one is cluster 1:

	Major_district	Latitude	Longitude	Cluster Label	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
11	Ōmori, Kamata, Haneda, Den-en-chōfu	-36.967356	174.940723	1	Fast Food Restaurant	Yoshoku Restaurant	Electronics Store	Fried Chicken Joint	French Restaurant	Forest	Food & Drink Shop	Flea Market	Fishing Store	Fishing Spot

The reason is that the geocoder generates a wrong coordinator here. The follow-up work could include correcting the coordinator in our code.