

Introduction/Business Problem

1.1. Description of the Problem

Tokyo was chosen as the hosts for the summer 2020 Olympic and Paralympic games. This is was set to take place in August 2010. However, due to the Coronavirus pandemic, the ceremony has been postponed to July 23rd , 2021. The aim of this project is to help visitors visualize the area in terms of restaurants that will fit their unique needs or requirements.

1.2. Discussion of the Background

The city has previously hosted the 1964 Summer Olympics as well as three G7 Summits (1979, 1983 and 1993). Tokyo is one of the most populous cities in the world and boasts much more restaurants than any other place in the world, offering unique Japanese, other Asian cuisine as well as international cuisine.

Tokyo is one of the most populous cities in the world and boasts much more restaurants than any other place in the world, offering unique Japanese, other Asian cuisine as well as international cuisine.

It will be a bit of a challenge for visitors to find the closest restaurants to them as many websites can be confusing and may not always offer the best recommendations.

The question is; How can we leverage Foursquare API data and Machine Learning to help us find the best restaurants for each visitor in Tokyo?
In this project, Foursquare location data and clustering methods will be used to group the different suburbs (or districts) according to their restaurant information.

2. Description of the Data

For this project, we will need the following data:

i) A list of Special Wards (districts) in Tokyo

Data Source: https://en.wikipedia.org/wiki/Special_wards_of_Tokyo

Description: We will use the above website source to get the list of all the Special Wards of Tokyo and some other information about them.

ii) Restaurants in each of the Special Wards (districts) in Tokyo

Data Source: Foursquare API

Description: These will help us get all the restaurants in each neighborhood. This can also be filtered to show the most popular restaurants in each area.

3. Methodology

This section discusses the data analysis that I conducted in order to arrive at the final results.

3.1. Scraping Data

I started off by scraping the data obtained from Wikipedia. I used the Pandas read function to create a data frame. I also dropped unnecessary columns from the list such as “No.” and “Flag”.

	Name	Kanji	Population(as of October 2016)	Density(/km2)	Area(km2)	Major districts
0	Chiyoda	千代田区	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	Chūō	中央区	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	Minato	港区	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	Shinjuku	新宿区	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	Bunkyo	文京区	223389	19790	11.29	Hongō, Yayoi, Hakusan

3.2. Add Geospatial Data to Data Frame

Then, I enabled geopy functions by installing the conda-forge geopy package. I used the nominatim function to add geospatial data to the data frame, i.e. the latitude and the longitude .

	Name	Kanji	Population(as of October 2016)	Density(/km2)	Area(km2)	Major districts	Latitude	Longitude
0	Chiyoda	千代田区	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.693810	139.753216
1	Chūō	中央区	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	35.666255	139.775565
2	Minato	港区	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055
3	Shinjuku	新宿区	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693763	139.703632
4	Bunkyo	文京区	223389	19790	11.29	Hongō, Yayoi, Hakusan	35.718810	139.744732
5	Taitō	台東区	200486	19830	10.11	Ueno, Asakusa	35.717450	139.790859
6	Sumida	墨田区	260358	18910	13.77	Kinshichō, Morishita, Ryōgoku	35.700429	139.805017

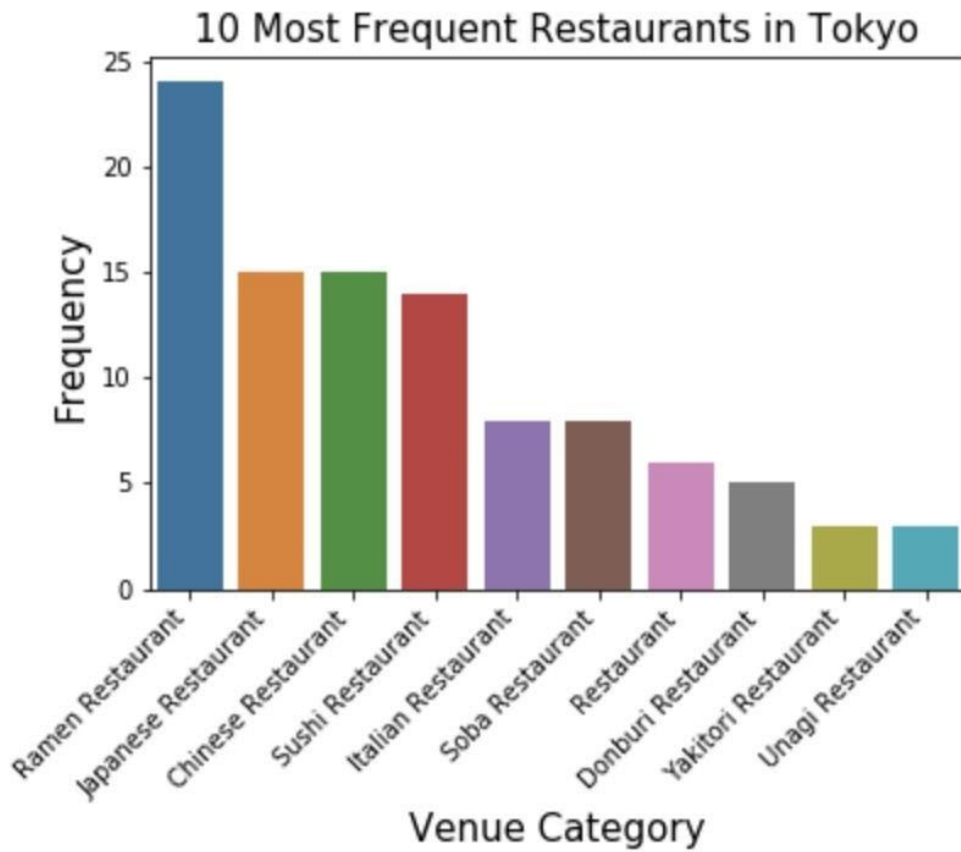
3.3. Visualizing the Special Wards on a Map

The Folium package as well as the data frame was used to visualize the special wards of Tokyo on a map.



3.4. Accessing Nearest Venues

Foursquare API data was used to access data of venues within a 500m radius. Here I also plot the top 10 most frequently occurring restaurants in Tokyo.



3.5. Finding Clusters

To find clusters of restaurant types in the different wards, I first transformed the data frame with the restaurant venues, associated to city districts, by one-hot encoding (0/1), as seen in the picture below.

	Neighborhood	Asian Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Kaiseki Restaurant
1	Chiyoda	0	0	0	0	0	0	0	0	0	0	0
2	Chiyoda	0	0	0	0	0	0	0	0	0	0	0
3	Chiyoda	0	0	0	0	0	0	0	0	1	0	0
4	Chiyoda	0	0	0	0	0	0	0	0	0	0	0
5	Chiyoda	0	0	0	0	0	0	0	0	0	0	0

Next, I used grouping to show the frequency of each category of restaurants in each neighborhood.

	Neighborhood	Asian Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Kaiseki Restaurant
0	Adachi	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000
1	Arakawa	0.0	0.166667	0.166667	0.000000	0.000000	0.000000	0.000000	0.166667	0.000000	0.166667	0.000000
2	Bunkyo	0.0	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Chiyoda	0.0	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000
4	Chūō	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.117647	0.000000	0.117647	0.000000
5	Edogawa	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.250000	0.000000	0.000000	0.000000
6	Itabashi	0.0	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000
7	Katsushika	0.0	0.000000	0.400000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

I then created a data frame to show the most common restaurants in each neighborhood.

	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	Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
1	Arakawa	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Italian Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant
2	Bunkyo	Chinese Restaurant	Szechuan Restaurant	Ramen Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
3	Chiyoda	Ramen Restaurant	Chinese Restaurant	Tempura Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant
4	Chūō	Sushi Restaurant	Soba Restaurant	Italian Restaurant	Japanese Restaurant	Tempura Restaurant	Seafood Restaurant	Yoshoku Restaurant	Kebab Restaurant	Chinese Restaurant	Donburi Restaurant
5	Edogawa	Ramen Restaurant	Indian Restaurant	Italian Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant
6	Itabashi	Chinese Restaurant	Italian Restaurant	Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
7	Katsushika	Donburi Restaurant	Dumpling Restaurant	Ramen Restaurant	Korean Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Dongbei Restaurant	French Restaurant	Indian Restaurant
8	Kita	Restaurant	South American Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
9	Kōtō	Chinese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	Japanese Curry Restaurant

After this, an unsupervised machine learning method called kmeans, was used to divide the neighborhoods into 5 clusters.

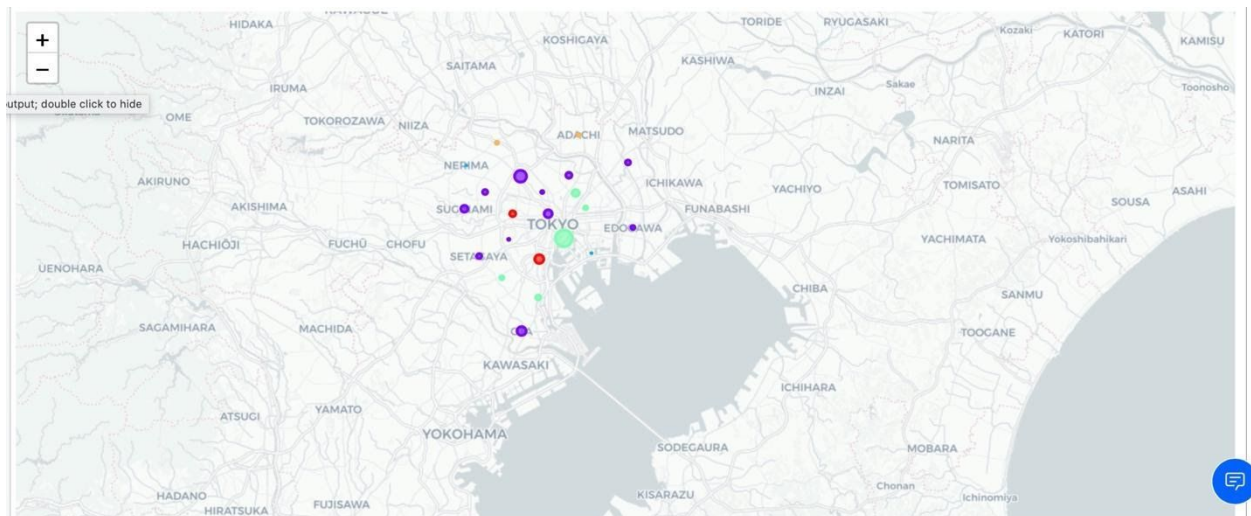
4. Results and Discussion

Below, one can see the tabulated results or data frame with the results.

	Cluster Labels	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	4	Adachi	Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
1	1	Arakawa	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Italian Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant
2	1	Bunkyo	Chinese Restaurant	Szechuan Restaurant	Ramen Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
3	1	Chiyoda	Ramen Restaurant	Chinese Restaurant	Tempura Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant
4	3	Chuo	Sushi Restaurant	Soba Restaurant	Italian Restaurant	Japanese Restaurant	Tempura Restaurant	Seafood Restaurant	Yoshoku Restaurant	Kebab Restaurant	Chinese Restaurant	Donburi Restaurant
5	1	Edogawa	Ramen Restaurant	Indian Restaurant	Italian Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant
6	4	Itabashi	Chinese Restaurant	Italian Restaurant	Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
7	1	Katsushika	Donburi Restaurant	Dumpling Restaurant	Ramen Restaurant	Korean Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Dongbei Restaurant	French Restaurant	Indian Restaurant
8	4	Kita	Restaurant	South American Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
9	2	Koto	Chinese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant	Italian Restaurant	Japanese Curry Restaurant

In the table it will be seen that the clusters are stated for each neighborhood. We can also see the most common venues or restaurants for each cluster.

In the below map, one can see a plot of each cluster, specifically color coded. Areas belonging to the same cluster will have the same color.



We can now finally look at restaurants in each cluster, which was the main aim of this project.

Cluster1

	Kanji	Major districts	Latitude	Longitude	Cluster Labels	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
2	港区	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...	35.643227	139.740055	0	Soba Restaurant	French Restaurant	Indian Restaurant	Japanese Restaurant	Kaiseki Restaurant	Kebab Restaurant	Yakitori Restaurant	Kosher Restaurant
3	新宿区	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...	35.693763	139.703632	0	Yakitori Restaurant	Yoshoku Restaurant	Unagi Restaurant	Thai Restaurant	Japanese Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant

In the first Cluster, the Soba and Yakitori Restaurants are the most common.

Cluster2

	Kanji	Major districts	Latitude	Longitude	Cluster Labels	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
0	千代田区	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...	35.693810	139.753216	1	Ramen Restaurant	Chinese Restaurant	Tempura Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant
4	文京区	Hongō, Yayoi, Hokusai	35.718810	139.744732	1	Chinese Restaurant	Szechuan Restaurant	Ramen Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant
10	大田区	Ōmori, Kamata, Haneda, Den-en-chōfu	35.561206	139.715843	1	Chinese Restaurant	Ramen Restaurant	Korean Restaurant	Soba Restaurant	Unagi Restaurant	Tonkatsu Restaurant	Japanese Curry Restaurant	Kaiseki Restaurant
11	世田谷区	Setagaya, Shimokitazawa, Kinuta, Karasuyama, T...	35.646096	139.656270	1	Ramen Restaurant	Unagi Restaurant	Szechuan Restaurant	Japanese Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant
12	渋谷区	Shibuya, Ebisu, Harajuku, Daikanyama, Hiroo, S...	35.664596	139.698711	1	Asian Restaurant	Ramen Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant
13	中野区	Nakano	35.718123	139.664468	1	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Italian Restaurant	Yoshoku Restaurant	Kosher Restaurant	Dongbei Restaurant	Dumpling Restaurant
14	杉並区	Kōenji, Asagaya, Ogikubo	35.699493	139.636288	1	Italian Restaurant	Ramen Restaurant	Chinese Restaurant	Dumpling Restaurant	Soba Restaurant	Yoshoku Restaurant	Korean Restaurant	Donburi Restaurant
15	豊島区	Ikebukuro, Komagome, Senkawa, Sugamo	35.736156	139.714222	1	Ramen Restaurant	Chinese Restaurant	Yoshoku Restaurant	Thai Restaurant	Dongbei Restaurant	Soba Restaurant	Japanese Restaurant	Korean Restaurant
17	荒川区	Arakawa, Machiya, Nippori, Minamisenju	35.737529	139.781310	1	Ramen Restaurant	Chinese Restaurant	Donburi Restaurant	Italian Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Dongbei Restaurant

The Ramen restaurants are the most popular in the second cluster. This should be no surprise as Japan is the home of Ramen noodles.

Cluster3

	Kanji	Major districts	Latitude	Longitude	Cluster Labels	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
7	江東区	Kiba, Ariake, Kameido, Tōyōcho, Monzennakachō,...	35.649154	139.812790	2	Chinese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant
19	練馬区	Nerima, Ōizumi, Hikarigaoka	35.748360	139.638735	2	Chinese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	French Restaurant	Indian Restaurant

The third cluster could well be called the Chinese cluster as Chinese restaurants are the most popular here.

Cluster4

	Kanji	Major districts	Latitude	Longitude	Cluster Labels	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
1	中央区	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...	35.666255	139.775565	3	Sushi Restaurant	Soba Restaurant	Italian Restaurant	Japanese Restaurant	Tempura Restaurant	Seafood Restaurant	Yoshoku Restaurant	Kebab Restaurant
5	台東区	Ueno, Asakusa	35.717450	139.790859	3	Sushi Restaurant	Japanese Restaurant	Nabe Restaurant	Chinese Restaurant	Sukiyaki Restaurant	Korean Restaurant	Donburi Restaurant	Dongbei Restaurant
6	墨田区	Kinshichō, Morishita, Ryōgoku	35.700429	139.805017	3	Japanese Restaurant	Tonkatsu Restaurant	Ramen Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant
8	品川区	Shinagawa, Gotanda, Ōsaki, Hatanodai, Ōimachi,...	35.599252	139.738910	3	Donburi Restaurant	Sushi Restaurant	Soba Restaurant	Restaurant	Japanese Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant
9	目黒区	Meguro, Nakameguro, Jiyugaoka, Komaba, Aobadai	35.621250	139.688014	3	Japanese Restaurant	Chinese Restaurant	Sushi Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant

In the third cluster we see Sushi restaurants as being one of the most popular. This is also not a surprise as Japan is mainly known for Sushi.

Cluster5

	Kanji	Major districts	Latitude	Longitude	Cluster Labels	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
16	北区	Akabane, Ōji, Tabata	-0.220164	-78.512327	4	Restaurant	South American Restaurant	Yoshoku Restaurant	Korean Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant
18	板橋区	Itabashi, Takashimadaira	35.774143	139.681209	4	Chinese Restaurant	Italian Restaurant	Restaurant	Yoshoku Restaurant	Kosher Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant
20	足立区	Ayase, Kitasenju, Takenotsuka	35.783703	139.795319	4	Restaurant	Japanese Restaurant	Yoshoku Restaurant	Kosher Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant

In the 5th cluster we see the most common restaurants to be an unclassified Restaurant as well as a Chinese one.

5. Conclusion

The Summer Olympics will start in July 2021, and visitors from all over the world can enjoy various cuisine according to their tastes in Tokyo. As a tourist, it's always best to try out the local cuisine, and they will have no trouble finding Ramen and Sushi restaurants at their home, i.e Japan.

Data Science methodology, Python libraries, Foursquare API and other sources of information on the internet enable us to solve real world problems and challenges. In the case of this project, I was able to provide prospective tourists to Tokyo with restaurant information according to their unique tastes and more especially Restaurant popularity.