IBM Applied Data Science Capstone

Opening a Restaurant for Clients with Certain Food Restrictions in Tokyo

By: Mahmoud Mohamed
April 2021

1. Introduction

As of 2021, Tokyo prefecture has an estimated population of 13,960,236. The Greater Tokyo Area is the most populous metropolitan area in the world, with more than 37.393 million residents as of 2020.

Although there are over 148,000 (Annual Report on Food Sanitation 2016) restaurants in Tokyo, most of them focus on serving the Japanese client like Ramen restaurants, Tonkatsu restaurants, ... etc. With the increasing number of foreigners in Tokyo in the past years, few restaurants started to focus on serving clients with food restrictions, but still not enough.

The aim of this report is to explore best locations for opening a restaurant for clients with certain food restrictions in Tokyo. For opening such a unique restaurant, picking the proper place is a key factor for such risky investment as it affects its chances of success or failure. Depending on the available data, I would like to give some insights about it.

1.1. Business Problem

The objective of this Capstone project is to analyse and select the best locations in Tokyo to open a new Halal/ Kosher/ Vegan restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in Tokyo, should the investor open such a restaurant.

1.2. Target Audience

- ✓ Developers and investors looking to open or invest in a niche restaurant in Tokyo as there are many foreigners visiting and living in Japan and some of them have some food restrictions.
- ✓ Business Analysts or Data Scientists, who wish to analyse Tokyo neighbourhoods using python, jupyter notebook and some machine learning techniques.
- ✓ Someone curious about data that want to have an idea, how beneficial it is to open a niche restaurant and what are the pros and cons of this business.

2. Data

First, we need some information about the wards of Tokyo, we can scrap Wikipedia to get that information:

Data source:
https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards

We can then clean the data by getting rid of useless columns and rows and also by renaming the column into easy and meaningful names.

Second, we need the latitudes and longitudes of the wards, we can get it using *geopy.geocoders.Nominatim* from *geopy* python library.

Third, by using Foursquare API we will get all the venues in each neighbourhood.

Now we are ready to start analysing our data and exploring different possibilities for the potential new restaurants.

3. Methodology

We import the data from Wikipedia, edit columns names and clean it.

	Romaji	Kanji	Population	Pop. density per km2	Area in km2
0	Chiyoda	千代田区	59441	5100	11.66
1	Chūō	中央区	147620	14460	10.21
2	Minato	港区	248071	12180	20.37
3	Shinjuku	新宿区	339211	18620	18.22
4	Bunkyō	文京区	223389	19790	11.29
5	Taitō	台東区	200486	19830	10.11

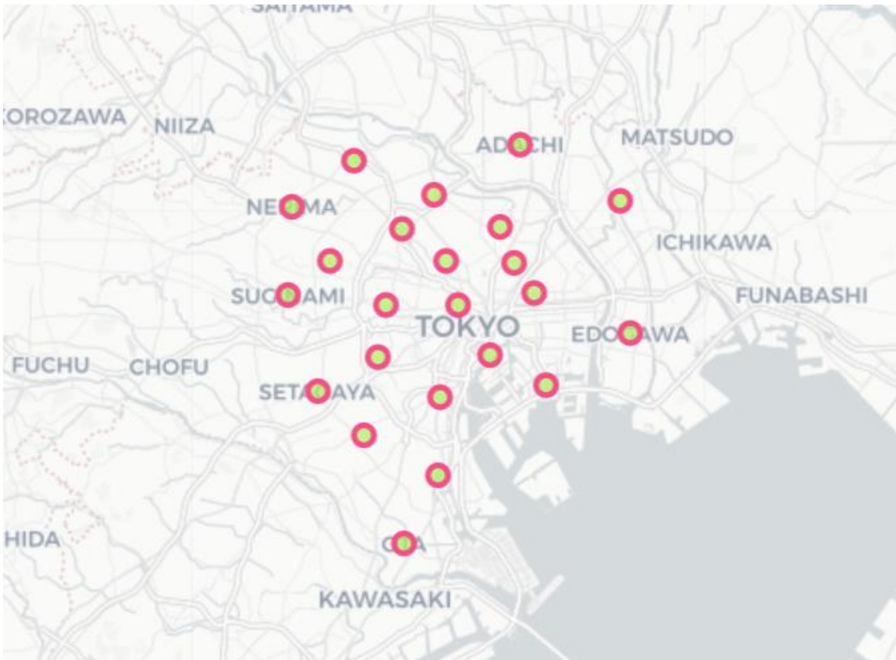
Romaji columns represents the Latin spelling of the ward name, while Kanji column represents the name in Chinese characters.

Now, we need the latitudes and longitudes of the wards, we can get it using *geopy.geocoders.Nominatim* from *geopy* python library.

“Kanji” column is used to retrieve locations from geopy. A sample of the new table after merging the coordinates follows:

	Romaji	Kanji	Population	Pop. density per km2	Area in km2	Latitude	Longitude
0	Chiyoda	千代田区	59441	5100	11.66	35.693810	139.753216
1	Chūō	中央区	147620	14460	10.21	35.666255	139.775565
2	Minato	港区	248071	12180	20.37	35.643227	139.740055
3	Shinjuku	新宿区	339211	18620	18.22	35.693763	139.703632
4	Bunkyō	文京区	223389	19790	11.29	35.718810	139.744732
5	Taitō	台東区	200486	19830	10.11	35.717450	139.790859

Create a map of Tokyo with 23 major wards superimposed on top using folium library.



We explore venues in different neighbourhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Chiyoda	35.69381	139.753216	Kanda Tendonya (神田天丼家)	35.695765	139.754682	Tempura Restaurant
2	Chiyoda	35.69381	139.753216	Sushi Masa (九段下 寿司政)	35.695234	139.752227	Sushi Restaurant
3	Chiyoda	35.69381	139.753216	Jimbocho Kurosu (神保町 黒須)	35.695539	139.754851	Ramen Restaurant
4	Chiyoda	35.69381	139.753216	Bondy (欧風カレー ボンディ)	35.695544	139.757356	Japanese Curry Restaurant
5	Chiyoda	35.69381	139.753216	たいよう軒	35.696454	139.754809	Chinese Restaurant

We use of Foursquare API and get the top 100 venues that are in Chiyoda within a r adius of 1000 meters. We get 47 unique venue categories returned by Foursquare le d by Café category.

```
In [18]: print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
100 venues were returned by Foursquare.
```

```
In [19]: print ('{} unique categories in Chiyoda'.format(nearby_venues['categories'].value_counts().shape[0]))
```

```
47 unique categories in Chiyoda
```

```
In [20]: print (nearby_venues['categories'].value_counts()[0:10])
```

```
Café                          13
Ramen Restaurant               8
Japanese Curry Restaurant      7
Chinese Restaurant             6
Tempura Restaurant             4
Bookstore                      4
Coffee Shop                    4
Historic Site                  4
Sushi Restaurant               3
Steakhouse                     2
Name: categories, dtype: int64
```

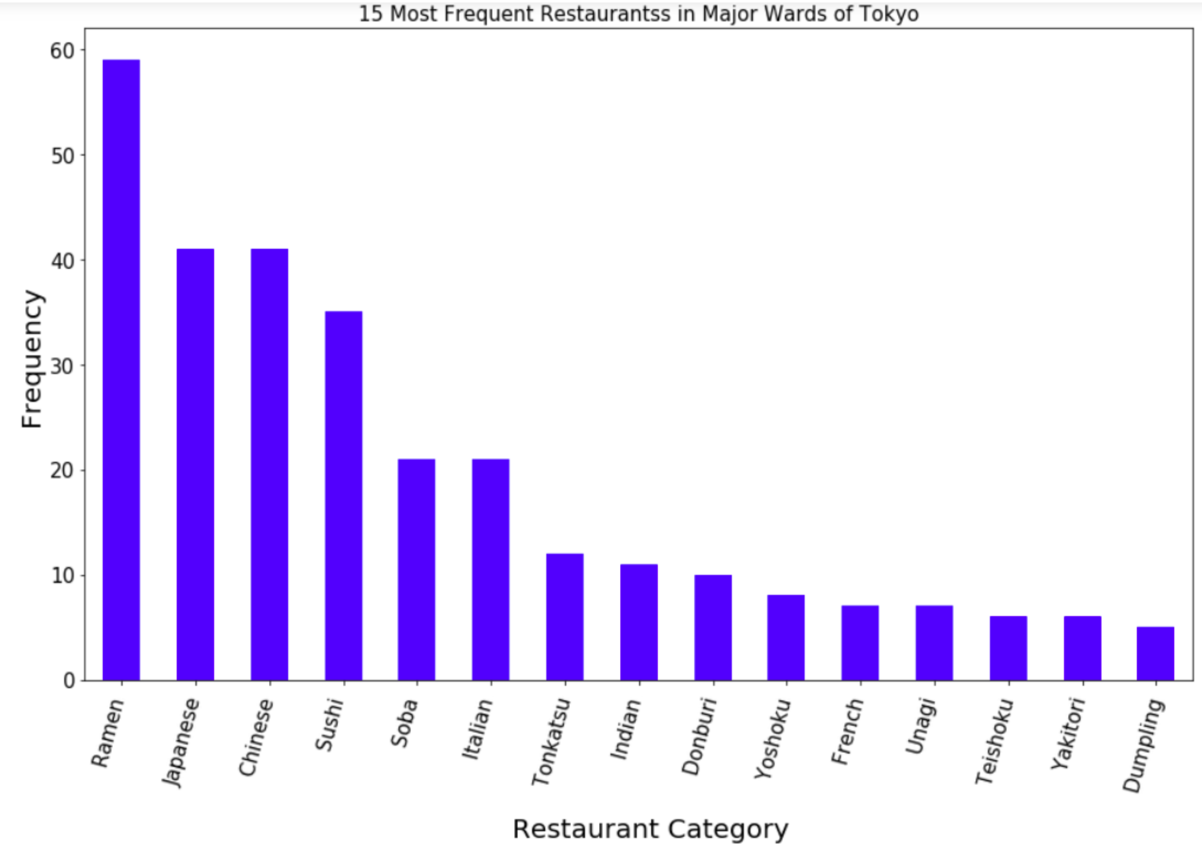
Check the wards with the most restaurants in the collected data.

Check the wards with the most restaurants in the collected data

```
In [31]: Tokyo_Venues_restaurant.sort_values(axis=0, ascending=False)
```

```
Out[31]: Neighborhood
Chūō          59
Ōta           46
Chiyoda       36
Shinjuku      33
Taitō         27
Shibuya       25
Toshima       19
Arakawa       14
Kita          14
Sumida        12
Suginami      11
Nakano        10
Minato        10
Setagaya       8
Katsushika     8
Meguro         7
Shinagawa      6
Edogawa        3
Kōtō           3
Bunkyō         3
Adachi         3
Itabashi       2
Name: Venue Category, dtype: int64
```

After that, we explored the most popular restaurants categories in Tokyo:



We use one hot encoding to represent all restaurants categories numerically. We also use pandas groupby on neighbourhood column and calculate the mean of the frequency of occurrence of each restaurant category.

In [35]:
```
Tokyo_grouped = Tokyo_onehot.groupby('Neighborhood').mean().reset_index()
Tokyo_grouped
```

Out[35]:

	Neighborhood	African Restaurant	American Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Donburi Restaurant	Dongbei Restaurant	Dumpling Restaurant	Fast Food Restaurant	...	Taiwanese Restaurant	Teishoku Restaurant	R
0	Adachi	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	...	0.000000	0.000000	
1	Arakawa	0.00	0.00	0.00	0.000000	0.214286	0.071429	0.000000	0.000000	0.000	...	0.000000	0.071429	
2	Bunkyō	0.00	0.00	0.00	0.000000	0.333333	0.000000	0.000000	0.000000	0.000	...	0.000000	0.000000	
3	Chiyoda	0.00	0.00	0.00	0.000000	0.194444	0.000000	0.000000	0.000000	0.000	...	0.000000	0.000000	
4	Chūō	0.00	0.00	0.00	0.000000	0.016949	0.033898	0.000000	0.000000	0.000	...	0.000000	0.016949	
5	Edogawa	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	...	0.000000	0.000000	

Then we print each neighbourhood along with the top 10 most common venues

```python
In [37]: num_top_venues = 10

for hood in Tokyo_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = Tokyo_grouped[Tokyo_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

```
----Chiyoda----
                      venue  freq
0        Chinese Restaurant  0.19
1          Ramen Restaurant  0.14
2         French Restaurant  0.08
3          Sushi Restaurant  0.08
4  Japanese Curry Restaurant  0.08
5                Restaurant  0.06
6           Soba Restaurant  0.06
7       Japanese Restaurant  0.06
8         Indian Restaurant  0.06
9        Italian Restaurant  0.06


----Chūō----
```

Then we create a pandas data frame representing those frequencies in order

```python
neighborhoods_venues_sorted.head(23)
```

Out[39]:

	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	Japanese Restaurant	Japanese Family Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Kosher Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Curry Restaurant
1	Arakawa	Ramen Restaurant	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Donburi Restaurant	Teishoku Restaurant	Japanese Restaurant	Italian Restaurant	Yoshoku Restaurant	Korean BBQ Restaurant
2	Bunkyō	Chinese Restaurant	Japanese Restaurant	Szechuan Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Family Restaurant
3	Chiyoda	Chinese Restaurant	Ramen Restaurant	French Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Italian Restaurant	Tonkatsu Restaurant	Indian Restaurant	Restaurant

4. Results

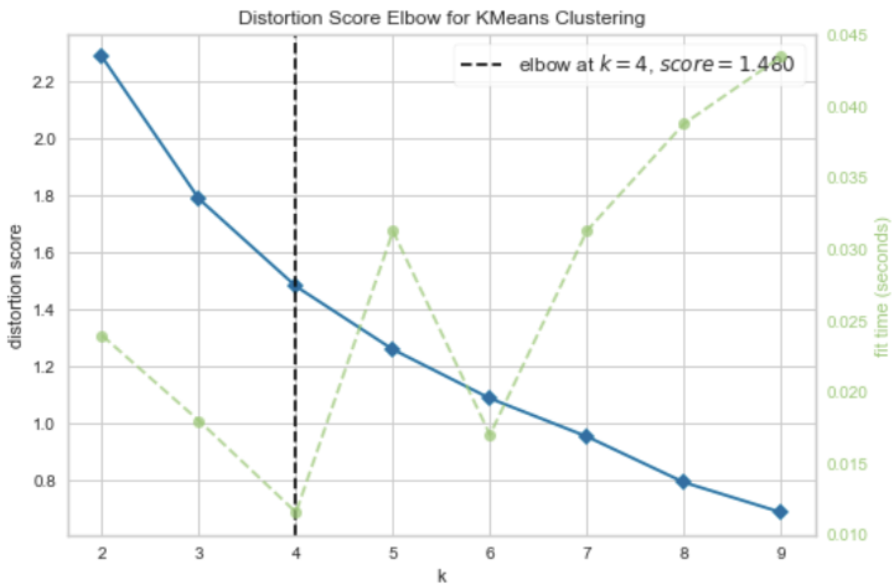
Now we will divide the major neighbourhoods into clusters depending on restaurants categories by using K-Means algorithm, but first we need to decide on the number of clusters, for that we will use the elbow method:

In [40]:
```
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer

Tokyo_grouped_clustering = Tokyo_grouped.drop('Neighborhood', 1)

# Instantiate the clustering model and visualizer
model = KMeans()
visualizer = KElbowVisualizer(model, k=(2,10))

visualizer.fit(Tokyo_grouped_clustering)        # Fit the data to the visualizer
visualizer.poof()    # Draw/show/poof the data
```



Now we will divide the neighbourhoods into 4 clusters using KMeans and then create a new dataframe that includes the cluster as well as the top 10 venues for each neighbourhood.

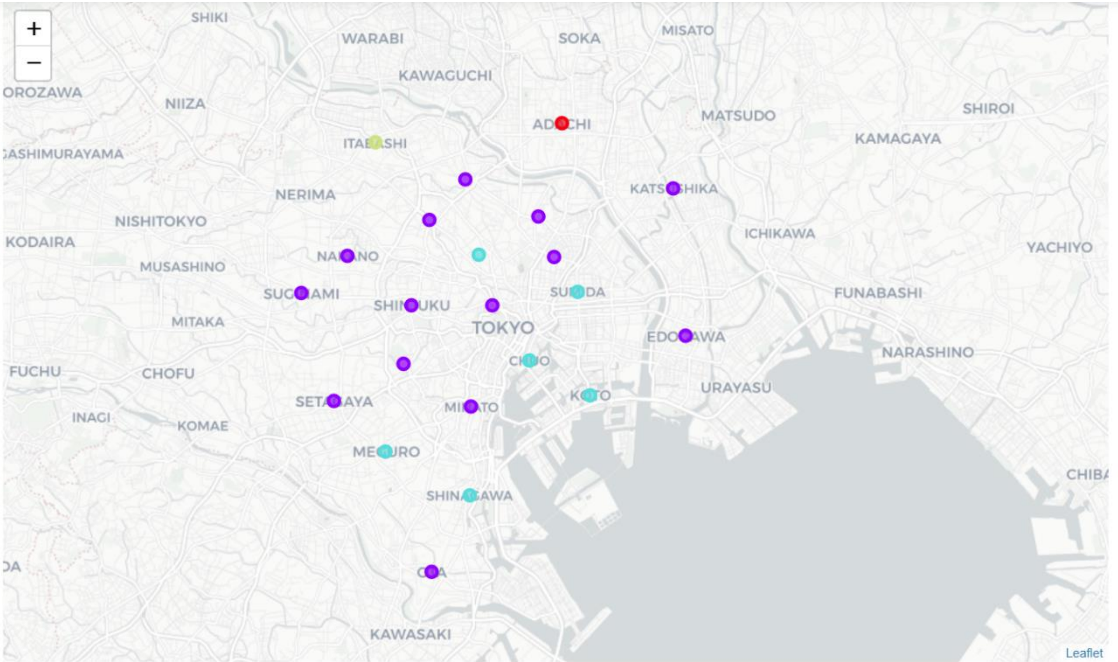
```
# drop NAN rows
tokyo_merged.dropna(axis=0, inplace=True)
tokyo_merged.head() # check the last columns!
```

Out[42]:

	Neighborhood	Kanji	Population	Pop. density per km2	Area in km2	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 Mo Comm Ven
0	Chiyoda	千代田区	59441	5100	11.66	35.693810	139.753216	1.0	Chinese Restaurant	Ramen Restaurant	French Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Itali Restaura
1	Chūō	中央区	147620	14460	10.21	35.666255	139.775565	2.0	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Tempura Restaurant	Donburi Restaurant	Yosho Restaura
2	Minato	港区	248071	12180	20.37	35.643227	139.740055	1.0	Soba Restaurant	Yakitori Restaurant	Indian Restaurant	Kosher Restaurant	Chinese Restaurant	Kebab Restaurant	Kais Restaura
3	Shinjuku	新宿区	339211	18620	18.22	35.693763	139.703632	1.0	Ramen Restaurant	Shabu-Shabu Restaurant	Thai Restaurant	Yakitori Restaurant	Chinese Restaurant	Teishoku Restaurant	Japane Restaura
4	Bunkyō	文京区	223389	19790	11.29	35.718810	139.744732	2.0	Chinese Restaurant	Japanese Restaurant	Szechuan Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Kore BI Restaura

Now we visualize the 4 clusters over folium map

Out[43]:



Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Cluster 1

In [44]:
```
tokyo_merged.loc[tokyo_merged['Cluster Labels'] == 0, tokyo_merged.columns[[1] + list(range(5, tokyo_merged.shape[1]))]]
```

Out[44]:

	Kanji	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	9th Most Common Venue	10th Most Common Venue
20	足立区	35.783703	139.795319	0.0	Japanese Restaurant	Japanese Family Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Kosher Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Curry Restaurant

Cluster 2

In [45]:
```
tokyo_merged.loc[tokyo_merged['Cluster Labels'] == 1, tokyo_merged.columns[[1] + list(range(5, tokyo_merged.shape[1]))]]
```

Out[45]:

	Kanji	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	9th Most Common Venue	10th Most Common Venue
0	千代田区	35.693810	139.753216	1.0	Chinese Restaurant	Ramen Restaurant	French Restaurant	Sushi Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Italian Restaurant	Tonkatsu Restaurant	Indian Restaurant	Restaurant
2	港区	35.643227	139.740055	1.0	Soba Restaurant	Yakitori Restaurant	Indian Restaurant	Kosher Restaurant	Chinese Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Restaurant	French Restaurant	Korean Restaurant
3	新宿区	35.693763	139.703632	1.0	Ramen Restaurant	Shabu-Shabu Restaurant	Thai Restaurant	Yakitori Restaurant	Chinese Restaurant	Teishoku Restaurant	Japanese Restaurant	Tonkatsu Restaurant	Unagi Restaurant	Seafood Restaurant
5	台東区	35.717450	139.790859	1.0	Ramen Restaurant	Soba Restaurant	Sushi Restaurant	Japanese Restaurant	Chinese Restaurant	Italian Restaurant	Nabe Restaurant	Sukiyaki Restaurant	Monjayaki Restaurant	Restaurant
10	大田区	35.561206	139.715843	1.0	Ramen Restaurant	Japanese Restaurant	Chinese Restaurant	Sushi Restaurant	Tonkatsu Restaurant	Italian Restaurant	Dumpling Restaurant	Yoshoku Restaurant	Udon Restaurant	Vietnamese Restaurant
11	世田谷区	35.646096	139.656270	1.0	Ramen Restaurant	Yoshoku Restaurant	Unagi Restaurant	Japanese Restaurant	Japanese Family Restaurant	Szechuan Restaurant	Fast Food Restaurant	Hotpot Restaurant	Korean BBQ Restaurant	Kebab Restaurant
12	渋谷区	35.664596	139.698711	1.0	Japanese Restaurant	Ramen Restaurant	Italian Restaurant	Chinese Restaurant	French Restaurant	Mexican Restaurant	South Indian Restaurant	American Restaurant	Asian Restaurant	Brazilian Restaurant
13	中野区	35.718123	139.664468	1.0	Ramen Restaurant	Chinese Restaurant	Italian Restaurant	Soba Restaurant	Tonkatsu Restaurant	Donburi Restaurant	Yoshoku Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant
14	杉並区	35.699493	139.636288	1.0	Ramen Restaurant	Italian Restaurant	Soba Restaurant	Tonkatsu Restaurant	Chinese Restaurant	Dumpling Restaurant	Sushi Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean BBQ Restaurant
15	豊島区	35.736156	139.714222	1.0	Ramen Restaurant	Soba Restaurant	Yoshoku Restaurant	Chinese Restaurant	Korean Restaurant	Japanese Family Restaurant	Dongbei Restaurant	Donburi Restaurant	Middle Eastern Restaurant	Tonkatsu Restaurant
16	北区	35.755838	139.736687	1.0	Ramen Restaurant	Soba Restaurant	Japanese Restaurant	Kushikatsu Restaurant	Donburi Restaurant	Teishoku Restaurant	Dumpling Restaurant	Italian Restaurant	Korean Restaurant	Korean BBQ Restaurant
17	荒川区	35.737529	139.781310	1.0	Ramen Restaurant	Chinese Restaurant	Indian Restaurant	Korean Restaurant	Donburi Restaurant	Teishoku Restaurant	Japanese Restaurant	Italian Restaurant	Yoshoku Restaurant	Korean BBQ Restaurant
21	葛飾区	35.751733	139.863816	1.0	Donburi Restaurant	Soba Restaurant	Korean Restaurant	Chinese Restaurant	Dumpling Restaurant	Ramen Restaurant	Indian Restaurant	Yoshoku Restaurant	Korean BBQ Restaurant	Kebab Restaurant
22	江戸川区	35.678278	139.871091	1.0	Ramen Restaurant	Italian Restaurant	Indian Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Restaurant

Cluster 3

In [46]:
```
tokyo_merged.loc[tokyo_merged['Cluster Labels'] == 2, tokyo_merged.columns[[1] + list(range(5, tokyo_merged.shape[1]))]]
```

Out[46]:

	Kanji	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	9th Most Common Venue	10th Most Common Venue
1	中央区	35.666255	139.775565	2.0	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Tempura Restaurant	Donburi Restaurant	Yoshoku Restaurant	Soba Restaurant	Unagi Restaurant	Kushikatsu Restaurant
4	文京区	35.718810	139.744732	2.0	Chinese Restaurant	Japanese Restaurant	Szechuan Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Family Restaurant
6	墨田区	35.700429	139.805017	2.0	Chinese Restaurant	Japanese Restaurant	Unagi Restaurant	Tonkatsu Restaurant	Ramen Restaurant	Sushi Restaurant	Indian Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean BBQ Restaurant
7	江東区	35.649154	139.812790	2.0	Chinese Restaurant	Indian Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Kosher Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Restaurant
8	品川区	35.599252	139.738910	2.0	Chinese Restaurant	Soba Restaurant	Donburi Restaurant	Japanese Restaurant	Sushi Restaurant	Yoshoku Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant
9	目黒区	35.621250	139.688014	2.0	Chinese Restaurant	Japanese Restaurant	Italian Restaurant	Sushi Restaurant	Soba Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant

Cluster 4

In [47]:
```
tokyo_merged.loc[tokyo_merged['Cluster Labels'] == 3, tokyo_merged.columns[[1] + list(range(5, tokyo_merged.shape[1]))]]
```

Out[47]:

	Kanji	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	9th Most Common Venue	10th Most Common Venue
18	板橋区	35.774143	139.681209	3.0	Chinese Restaurant	Italian Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Kosher Restaurant	Korean Restaurant	Korean BBQ Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Restaurant

5. Discussion

As the analysis showed, the 15 Most Frequent Restaurants in Major Wards of Tokyo don't any include any of our targeted niche restaurants (Halal/ Kosher/ Vegan) categories.

By clustering the major neighbourhood in Tokyo based on restaurants categories, we can see the following:

- Cluster1: Japanese restaurants dominate it. The 5th Most Common restaurant category is Kosher.
- Cluster2: Ramen restaurants dominate it. The 4th Most Common restaurant category in Minato is Kosher, but nothing else at the other neighbourhoods.
- Cluster3: Chinese restaurants dominate it. The 5th Most Common restaurant category in Koto is Kosher, but nothing else at the other neighbourhoods.
- Cluster4: Chinese restaurants dominate it. The 5th Most Common restaurant category is Kosher.

We can also see that Chuo, Ota and Chiyoda cities have maximum number of restaurants, while Nerima and Itabashi have the least.

We should also take in consideration that some of the restaurants can be valid for Halal, Kosher, or Vegan meals. For example:

- Kosher restaurants are also be considered Halal.
- by my life experience in Tokyo, I can tell that most of Sushi/ Seafood/ Kebab/ Indian/ Restaurant’s menus are halal.
- For vegans, many Japanese Curry Restaurants and Hotpot menus are valid for vegans.

Another note is that these analyses depend on the available data from Foursquare and that we did not consider other factors such as population density/ distance from train stations and so on.

6. Conclusion

- There are many real-life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighbourhoods in Tokyo based on the most common food Restaurants in its 23 major districts. The results can help an investor to decide about the district that fit the new restaurant project.
- I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major districts of Tokyo and saw the results of segmentation of districts using Folium map.
- The analysis shows the potential of opening new niche restaurants for clients with food restrictions (Halal, Kosher, or Vegan in this study).
- Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.