Capstone project - The battle of Neighborhoods

Jingwen Dong

August 16, 2020

1. Introduction/Business Problem

1.1 Introduction

Tokyo is the capital and most populous prefecture of Japan. It is the political and economic center. Tokyo has a population of over 10 million and is the most densely populated city in Japan. Such population density has boosted the development of restaurants and other social venues. People there have developed an incredibly rich and varied food culture and ramen is one of the most significant parts. Ramen shops are very common and popular throughout the entire country. Japanese obsession with ramen even extends to a ramen museum.

1.2 Business Problem

Unlike in Toronto and Newyork, where cities are split into neighborhoods, Tokyo consists of 23 special wards. There are around 1000 ramen shops in Tokyo located in its 23 special wards. With the increasing amount of population, demand for restaurants, especially ramen shops has gone up. More investors are looking to open up ramen shops however they need more information to decide the location for the new business.

In this project, we will explore and cluster the 23 special wards in Tokyo based on their population density, direct competitors, as well as other social venues close by with the help of data science technology. With the analytical results, Investors will be able to compare different clusters based on these criteria and then choose the optimal location for starting a new Ramen Restaurant.

2. Data Section

2.1 Data Collection

here are some of the datasets that will be used in my project:

- List of Special Wards: there are 23 special wards in Tokyo and I will scrap the Wikipedia page to read the wards name and population density from the table.
- Location Data: Geocoder Python package is used to find the geographical coordinates of each special ward in Tokyo.
- Venue data: Foursquare API will be used to obtain up to 100 social venues or ramen restaurants located in each special ward in Tokyo in radius of 800 meters.

2.2 Data Preparation

Forming Tokyo dataframe: An empty dataframe is created. Then I looped through the special wards name scrapped from webpage, obtained the coordinates of each wrad using the Geocoder Package, and filled in the Name, Population, Density, Latitude and Longitude coordinates columns to form the Tokyo dataframe like shown below:

	Name	Population	Density(/km2)	Latitude	Longitude
0	Chiyoda	59441	5100	35.682100	139.765579
1	Chūō	147620	14460	35.666255	139.775565
2	Minato	248071	12180	35.644646	139.724120
3	Shinjuku	339211	18620	35.692169	139.700558
4	Bunkyō	223389	19790	35.705571	139.751970

Forming Venues Dataframe: Foursquare API are used to obtain the venues in each ward. There are 1846 venues in total and they are in 230 unique categories. The ward coordinates, venue name, venue coordinates and venue category are selected to form a new dataframe:

Name	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Chiyoda	35.681091	139.767186	Tokyo Station Hotel (東京ステーションホテル)	35.681274	139.765917	Hotel
Chiyoda	35.681091	139.767186	Saryo Tsujiri (茶寮 都路里)	35.681373	139.769289	Tea Room
Chiyoda	35.681091	139.767186	Fukuramu-chan (ふくらむちゃん柱)	35.682498	139.766453	Exhibit
Chiyoda	35.681091	139.767186	KITTE Garden (屋上庭園 KITTEガーデン)	35.679654	139.765169	Roof Deck
Chiyoda	35.681091	139.767186	Tokyo Station Gallery (東京ステーションギャラリー)	35.682293	139.766332	Art Gallery

3. Methodology

3.1 Exploratory Data Analysis

using one hot encoding to convert the venues categories variable to dummy variables of either 0 or 1 so that the categorical variables are easier to work with. The resulting data frame is in shape 1846 * 230 like below:

	Name	ATM	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Chiyoda	0	0	0	0	0	0	0	0	0
1	Chiyoda	0	0	0	0	0	0	0	0	0
2	Chiyoda	0	0	0	0	0	0	0	0	0
3	Chiyoda	0	0	0	0	0	0	0	0	0
4	Chiyoda	0	0	0	0	0	1	0	0	0

By applying the groupby method on wards name and taking the mean values, we are able to get a 23*231 dataframe. Each row in the data frame indicates a special ward and each column indicates the frequency of occurrence of each category in that ward like shown below:

	Name	ATM	Accessories Store	African Restaurant	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant
0	Adachi	0.00000	0.00	0.00	0.01	0.00	0.010000	0.010000	0.000000	0.000000
1	Arakawa	0.00000	0.00	0.00	0.00	0.00	0.000000	0.013889	0.000000	0.000000
2	Bunkyō	0.00000	0.00	0.00	0.00	0.00	0.010000	0.000000	0.000000	0.000000
3	Chiyoda	0.00000	0.00	0.00	0.00	0.00	0.010000	0.010000	0.000000	0.000000
4	Chūō	0.00000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	0.000000

To gain better understanding and to see how values are distributed, I sorted the venues in each ward, and printed out the top 5 common venues along with the occurrence frequency, as well as the ranking index of Ramen Restaurant in each ward.

- Overall, the top 5 common venues comprise about 25% - 50% of each ward's total venues. If the majority of the top 5 venues are restaurants or shops that provide food and drinks, then it's reasonable to assume that the restaurant and food service industry in this area is probably saturated. On the other hand, if none of the top 5 common

venues is restaurants, then there might be a potential opportunity to develop more business in this area.

- If Ramen Restaurant is ranked as the top 5 or top 10 most common venues, opening new Ramen Shops might not be something new to the area. Also, existing restaurants might already have regular customers whose tastes won't easily change. It will be challenging for new business owners to attract a sufficient number of customers. However if there are too few ramen restaurants, for example it is the 40th common venue, residents in the area might simply not be a big fan of ramen. Therefore, there might not be too many potential customers.

Based on the two criterias discussed above, a decision can be made by analyzing the two special wards printed below:

```
----Ōta----
              venue  freq
0  Ramen Restaurant  0.16
                 venue  freq
0     Ramen Restaurant  0.16
1            Sake Bar  0.10
2   Chinese Restaurant  0.09
3  Japanese Restaurant  0.07
4  Tonkatsu Restaurant  0.05
```

```
----Setagaya----
                 venue  freq
11  Ramen Restaurant  0.03
                 venue  freq
0  Convenience Store  0.21
1               Café  0.07
2             Bakery  0.07
3       Tram Station  0.07
4           Sake Bar  0.03
```

- In Ota, the top 5 most common venue categories together comprise 47% of Ota's total venues. 4 among those are in the restaurant and food service industry. Besides, Ramen Restaurant itself comprises 16% of Ota's venues and is the top 1 most common venue. With so many restaurants densely-distributed, it is likely that Ota is either a financial, industrial, or commercial area. The cost for opening a business must be high.

- On the other hand, in Setagaya, none of the top 5 most common venues are restaurants. Ramen Restaurant is ranked as the 11th most common venue, which is not bad. The restaurant and food service industry in this area seems less saturated. With Convenience Store ranked as the most common venue, Setagaya might either be a residence area or far away from the city center. Either way, the cost for opening a new business won't be as high as Ota.

Based on the discussion above, without taking population density info into account, Setagaya can be selected as an optimal location for opening up a new Ramen Shop/ Restaurant compared with Ota. Similar analytical skills can be applied to each venue. Since there are 23 venues in total, there must be similarities between them. It will be easier to cluster the venues into different groups based on their features, and apply analytical skills on each cluster to find out the final optimal solution.

A new dataframe is created to display the top 10 venues, as well as the rank index of Ramen Restaurant of each ward. Now this dataframe below will be used for comparing results later.

Name	Ramen Restaurant	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
Adachi	5th most common	Japanese Restaurant	Hotel	Café	Hobby Shop	Ramen Restaurant	Chocolate Shop	Deli / Bodega	Lounge	Tonkatsu Restaurant	Yoshoku Restaurant
Arakawa	5th most common	Convenience Store	Grocery Store	Park	Japanese Family Restaurant	Ramen Restaurant	Sake Bar	Italian Restaurant	Café	Tram Station	Noodle House
Bunkyō	1st most common	Ramen Restaurant	Sake Bar	Baseball Stadium	Hotel	Café	Japanese Restaurant	BBQ Joint	Bed & Breakfast	Italian Restaurant	Theater
Chiyoda	3rd most common	Japanese Restaurant	Hotel	Ramen Restaurant	Coffee Shop	South Indian Restaurant	Hobby Shop	Sushi Restaurant	Café	Dessert Shop	Tonkatsu Restaurant
Chūō	8th most common	Sushi Restaurant	Monjayaki Restaurant	Japanese Restaurant	Soba Restaurant	Italian Restaurant	Coffee Shop	Donburi Restaurant	Ramen Restaurant	Bakery	Tonkatsu Restaurant

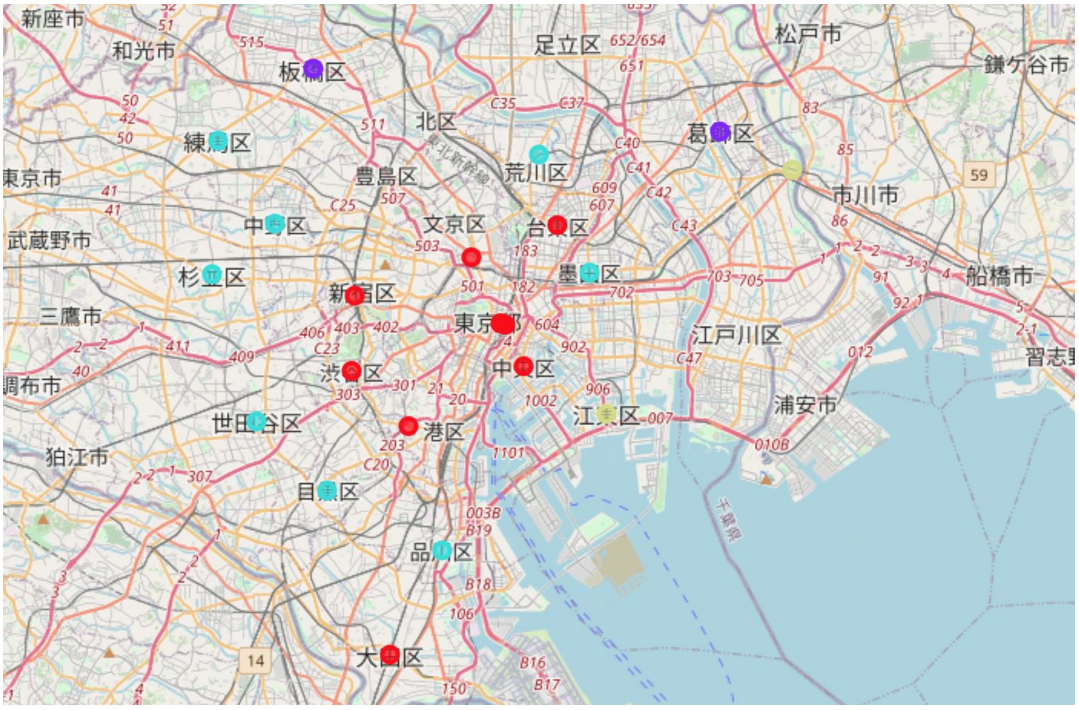
3.2 Model Development

To partition the 23 special wards in Tokyo into groups of individuals that have similar characteristics, K-Means clustering was applied. This strategy allows the business to target specific groups of wards so as to more effectively making business decisions. 4 was selected as the number of clusters that the special wards were partitioned into. I ran KMeans method to generate the cluster labels for each ward. Combining the cluster labels with the top 10 venues, a new dataframe is generated:

	Name	Population	Latitude	Longitude	Cluster Labels	Ramen Restaurant	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Chiyoda	59441	35.681091	139.767186	0	3rd most common	Japanese Restaurant	Hotel	Ramen Restaurant	Coffee Shop	South Indian Restaurant
1	Chūō	147620	35.666255	139.775565	0	8th most common	Sushi Restaurant	Monjayaki Restaurant	Japanese Restaurant	Soba Restaurant	Italian Restaurant
2	Minato	248071	35.644646	139.724120	0	6th most common	BBQ Joint	Japanese Restaurant	Café	French Restaurant	Sushi Restaurant
3	Shinjuku	339211	35.692153	139.700300	0	6th most common	BBQ Joint	Sake Bar	Bar	Café	Japanese Restaurant
4	Bunkyō	223389	35.705571	139.751970	0	1st most common	Ramen Restaurant	Sake Bar	Baseball Stadium	Hotel	Café

3.3 Data Visualization

To visualize this resulting clustering in an easier way, folium was used to generate a map and label each special ward based on its cluster label. The resulting map looks like this:



4. Results

I examined each cluster to determine the categories that distinguish each cluster.

Some common characteristics that can be drawn from **Cluster 1** are:

- restaurants densely-distributed, 5 to 9 venue categories are restaurants

	Name	Population	Ramen Restaurant	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	Chiyoda	59441	3rd most common	Japanese Restaurant	Hotel	Ramen Restaurant	Coffee Shop	South Indian Restaurant	Hobby Shop	Sushi Restaurant	Café	Dessert Shop	Tonkatsu Restaurant
1	Chūō	147620	8th most common	Sushi Restaurant	Monjayaki Restaurant	Japanese Restaurant	Soba Restaurant	Italian Restaurant	Coffee Shop	Donburi Restaurant	Ramen Restaurant	Bakery	Tonkatsu Restaurant
2	Minato	248071	6th most common	BBQ Joint	Japanese Restaurant	Café	French Restaurant	Sushi Restaurant	Ramen Restaurant	Italian Restaurant	Coffee Shop	Bakery	Yoshoku Restaurant
3	Shinjuku	339211	6th most common	BBQ Joint	Sake Bar	Bar	Café	Japanese Restaurant	Ramen Restaurant	Movie Theater	Bookstore	Pub	Coffee Shop
4	Bunkyō	223389	1st most common	Ramen Restaurant	Sake Bar	Baseball Stadium	Hotel	Café	Japanese Restaurant	BBQ Joint	Bed & Breakfast	Italian Restaurant	Theater
5	Taitō	200486	5th most common	Coffee Shop	Hostel	BBQ Joint	Japanese Restaurant	Ramen Restaurant	Café	Dessert Shop	Sukiyaki Restaurant	Hotel	Yoshoku Restaurant
10	Ōta	722608	1st most common	Ramen Restaurant	Chinese Restaurant	Convenience Store	Sake Bar	Japanese Restaurant	Grocery Store	Tonkatsu Restaurant	Café	Dumpling Restaurant	Bed & Breakfast
12	Shibuya	227850	7th most common	Café	Record Shop	Sake Bar	Coffee Shop	BBQ Joint	Concert Hall	Ramen Restaurant	Bar	Nightclub	Japanese Restaurant
15	Toshima	294673	3rd most common	Japanese Restaurant	Hotel	Ramen Restaurant	Coffee Shop	South Indian Restaurant	Hobby Shop	Sushi Restaurant	Café	Dessert Shop	Tonkatsu Restaurant
16	Kita	345063	5th most common	Japanese Restaurant	Hotel	Café	Hobby Shop	Ramen Restaurant	Chocolate Shop	Deli / Bodega	Lounge	Tonkatsu Restaurant	Yoshoku Restaurant
20	Adachi	674067	5th most common	Japanese Restaurant	Hotel	Café	Hobby Shop	Ramen Restaurant	Chocolate Shop	Deli / Bodega	Lounge	Tonkatsu Restaurant	Yoshoku Restaurant

- Strong competitors, Ramen Restaurant ranked within the Top 8 venues
- Red markers, located next to or around city center

Some insights that can be drawn for **Cluster 2**:

- Only 1 or 2 categories are restaurants
- Convenience Store is the most common venue
- Ramen Restaurant not in the top 10 venues list
- Purple markers, wards are far away from the city center

	Name	Population	Ramen Restaurant	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	Itabashi	569225	24th most common	Convenience Store	Park	Intersection	Bus Stop	Supermarket	Chinese Restaurant	Furniture / Home Store	Bowling Alley	Rest Area	Japanese Restaurant
21	Katsushika	447140	11st most common	Convenience Store	Grocery Store	Donburi Restaurant	Coffee Shop	Discount Store	Train Station	ATM	Bakery	Pharmacy	Ice Cream Shop

Characteristics found for **Cluster 3**:

- 3 to 5 categories are restaurant, moderately- distributed.
- Convenience Store is the most common venue
- Ramen Restaurant ranked within the Top 15 venues.
- Blue markers, not too far away or too close to city center

	Name	Population	Ramen Restaurant	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
6	Sumida	260358	3rd most common	Convenience Store	Café	Ramen Restaurant	Park	Japanese Restaurant	Supermarket	Hotel	Soba Restaurant	Indian Restaurant	Coffee Shop
8	Shinagawa	392492	11st most common	Convenience Store	BBQ Joint	Furniture / Home Store	Pizza Place	Chinese Restaurant	Bus Stop	Racecourse	Supermarket	Sushi Restaurant	Soba Restaurant
9	Meguro	280283	41st most common	Convenience Store	Bus Stop	Grocery Store	Park	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Playground	Sushi Restaurant	Soba Restaurant
11	Setagaya	910868	10th most common	Convenience Store	Intersection	Café	Japanese Restaurant	Park	Bakery	Bus Stop	Diner	Tram Station	Ramen Restaurant
13	Nakano	332902	2nd most common	Convenience Store	Ramen Restaurant	Italian Restaurant	Café	Park	Grocery Store	BBQ Joint	Soba Restaurant	Noodle House	Diner
14	Suginami	570483	3rd most common	Park	Sake Bar	Ramen Restaurant	Convenience Store	Café	Chinese Restaurant	BBQ Joint	Grocery Store	Intersection	Dumpling Restaurant
17	Arakawa	213648	5th most common	Convenience Store	Grocery Store	Park	Japanese Family Restaurant	Ramen Restaurant	Sake Bar	Italian Restaurant	Café	Tram Station	Noodle House
19	Nerima	726748	7th most common	Theme Park Ride / Attraction	Convenience Store	Intersection	Grocery Store	Fishing Spot	Fast Food Restaurant	Ramen Restaurant	Café	Bookstore	Tennis Court

And finally for **Cluster 4**:

- Only 1 or 2 categories are restaurants
- Most venue categories are outdoor facilities
- Green markers, either next to the harbour or far away from the city center

	Name	Population	Ramen Restaurant	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
7	Kōtō	502579	17th most common	Intersection	Park	Convenience Store	Bus Stop	Rest Area	Dessert Shop	Pool	Baseball Field	Chinese Restaurant	Plaza
22	Edogawa	685899	6th most common	Convenience Store	Park	Intersection	Bus Stop	Baseball Field	Ramen Restaurant	Grocery Store	Gym	Music Venue	Chinese Restaurant

5. Discussion

The restaurant and food service industry seems to be nearly saturated for Cluster 1. With Ramen Restaurant ranked as the top 8 venues, there are already plenty of existing strong competitors. All wards in this cluster seem to be not far away from the city center. This cluster seems to be the financial, commercial, and industrial center of Tokyo, where the cost for opening new business is quite high. With a limited amount of budget, this cluster might not be not the ideal location for a new Ramen Shop.

Being the completely opposite of Cluster 1, Cluster 2 does not contain many restaurants. Most common venues are Convenient Store, Bus/Train Station, and grocery stores. This cluster seems to be away from the city center, and the markers on the map confirms that. The population density of this cluster is in the mid-low level compared with other clusters. Therefore, opening a new business won't cost that much. There will not be too many competitors, however it does not guarantee sufficient numbers of customers.

Cluster 3 contains an equally-distributed amount of restaurants, convenient stores, bus stops and outdoor facilities. This cluster seems to be a mix of residents and commercial areas, where the cost will be in the middle, and still has potential customers. It seems to be an optimal location for a new Ramen Store.Among all the 8 special wards listed in this cluster, Shinagawa and Setagaya seem to be an ideal location as Ramen Restaurant ranked as the 11th and 10th most common venue in these two wards respectively.

Cluster 4 seems to be even further away from the city center, with not many restaurants but fully equipped with outdoor facilities such as baseball fields, parks, and pools. The total area of each ward in this cluster is large, and the population density however is small. Opening and maintaining a new ramen restaurant in this cluster will be quite challenging due to a loosely-distributed population.

6. Conclusion

The purpose of this project is to analyze and cluster the 23 special wards in Tokyo in order to select an optimal location for business owners to open a new Ramen Restaurant. By exploring the social venues, geographical locations, and checking on existing competitors, all wards are partitioned into different clusters.

Cluster 3 is selected as the optimal cluster for opening a new Ramen Restaurant, as small businesses have a better chance to survive and develop in a combination area of commercial and residential, which provides sufficient customers and good location.

The final decision will be made by the business owners depending on their specific requirements, budget, and other concerns.