The Battle of Neighborhoods

Introduction/Business Problem

The objective of this project is to segment the different neighborhoods of Tokyo based on the venues they have, especially the restaurants. This could help a hypothetical person who wants to open a restaurant in Tokyo by providing different kinds of neighborhood profiles in reference to the foodservice industry. Finding the right location for a business as a restaurant is a key aspect of business success. Using this data analysis should facilitate the owner to find the optimal location.

Data

- Data from Wikipedia obtained by web scraping about special wards of Tokyo. This data contains a list of Neighborhood names, area, density, and location.
  Data source: https://en.wikipedia.org/wiki/Special_wards_of_Tokyo
- Latitude and longitude of the neighborhoods obtained with Nominatim from geopy.geocoders library.
  Data source: GeoPy's library
- Data from venues and restaurants in the different neighborhoods of Tokyo
  Data source: Fousquare API

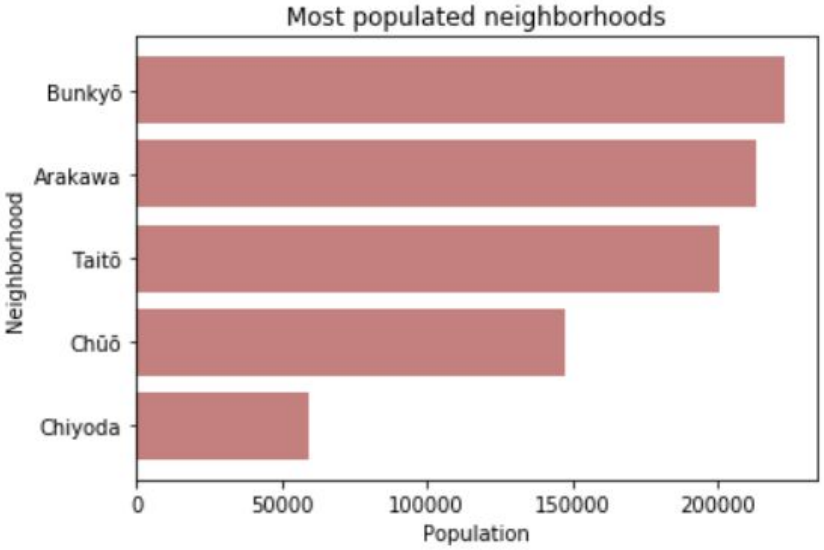
Methodology

- Collect data from Wikipedia of Tokyo neighborhoods by web scraping the site and putting it into a DataFrame

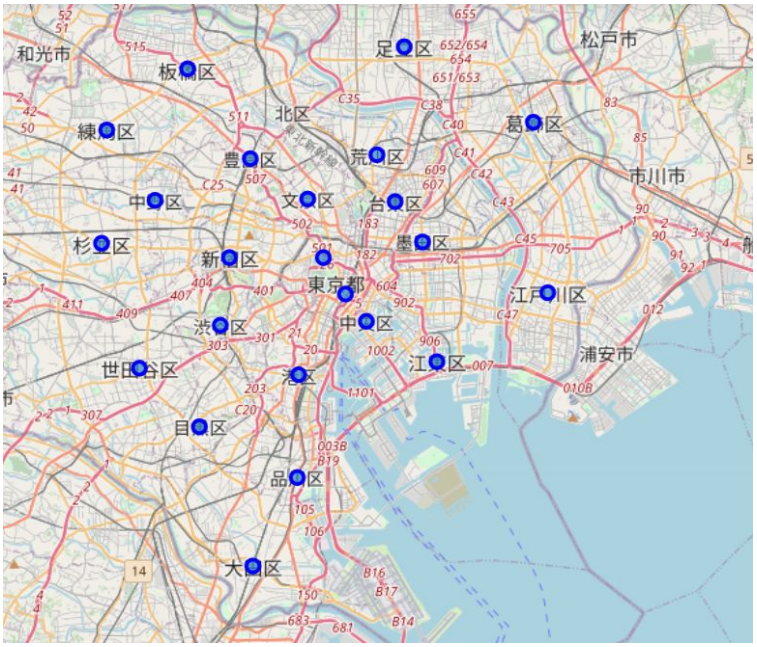
	No.	Flag	Name	Kanji	Population(as of October 2016	Density(/km2)	Area(km2)	Major districts
0	01	NaN	Chiyoda	千代田区	59441	5100	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	02	NaN	Chūō	中央区	147620	14460	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	03	NaN	Minato	港区	248071	12180	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	04	NaN	Shinjuku	新宿区	339211	18620	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	05	NaN	Bunkyō	文京区	223389	19790	11.29	Hongō, Yayoi, Hakusan

- Clean the data leaving only relevant fields and ad latitude and longitude using GeoPy's for each neighborhood

	Neighborhood	Population	Density	Area	Latitude	Longitude
0	Chiyoda	59441	5100	11.66	35.6938	139.753
1	Chūō	147620	14460	10.21	35.6663	139.776
2	Minato	248071	12180	20.37	35.6432	139.74
3	Shinjuku	339211	18620	18.22	35.6938	139.704
4	Bunkyō	223389	19790	11.29	35.7188	139.745
5	Taitō	200486	19830	10.11	35.7174	139.791
6	Sumida	260358	18910	13.77	35.7004	139.805
7	Kōtō	502579	12510	40.16	35.6492	139.813
8	Shinagawa	392492	17180	22.84	35.5993	139.739
9	Meguro	280283	19110	14.67	35.6213	139.688
10	Ōta	722608	11910	60.66	35.5612	139.716



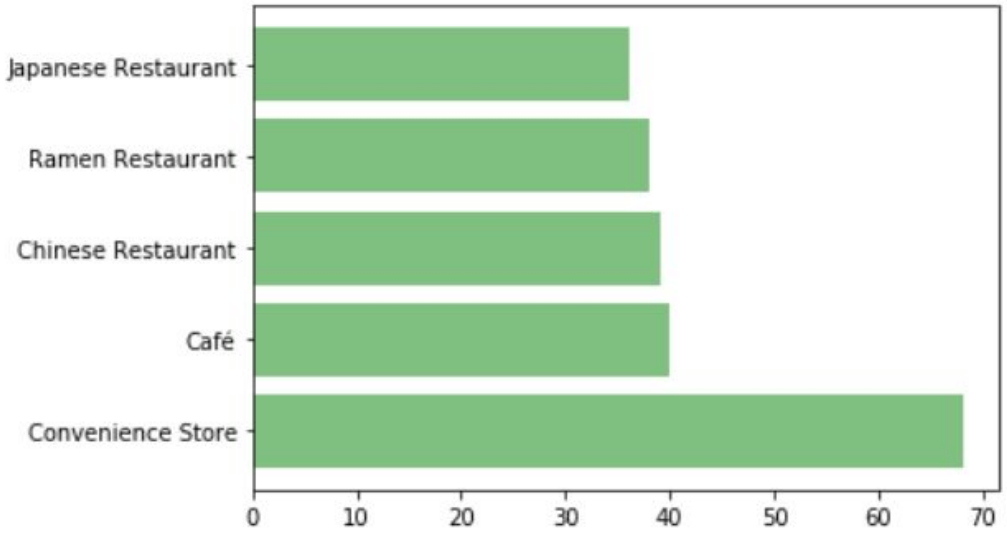
- Visualize neighborhoods



- Get venues for each neighborhood

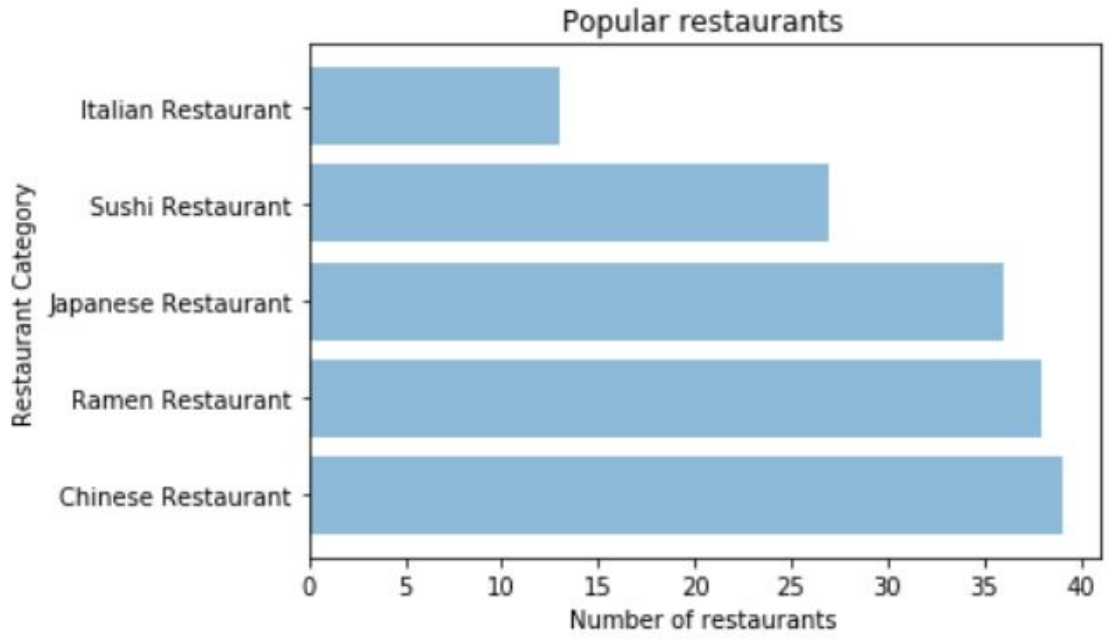
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chiyoda	35.69381	139.753216	Jimbocho Kurosu (神保町 黒須)	35.695539	139.754851	Ramen Restaurant
1	Chiyoda	35.69381	139.753216	Shimizumon Gate (清水門)	35.692685	139.752681	Historic Site
2	Chiyoda	35.69381	139.753216	Kitanomaru Park (北の丸公園)	35.691653	139.751201	Park
3	Chiyoda	35.69381	139.753216	Nippon Budokan (日本武道館)	35.693356	139.749865	Stadium
4	Chiyoda	35.69381	139.753216	Mori no Butchers (森のブッチャーズ)	35.694770	139.755980	Gastropub

	index	Venue Category
0	Convenience Store	68
1	Café	40
2	Chinese Restaurant	39
3	Ramen Restaurant	38
4	Japanese Restaurant	36
...	...	...
165	Theater	1
166	Breakfast Spot	1
167	Diner	1
168	Bus Station	1
169	Arts & Crafts Store	1



- Filter only the venues related to foodservice

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chiyoda	35.69381	139.753216	Jimbocho Kurosu (神保町 黒須)	35.695539	139.754851	Ramen Restaurant
5	Chiyoda	35.69381	139.753216	Kanda Tendonya (神田天丼家)	35.695765	139.754682	Tempura Restaurant
7	Chiyoda	35.69381	139.753216	Bondy (欧風カレー ボンディ)	35.695544	139.757356	Japanese Curry Restaurant
11	Chiyoda	35.69381	139.753216	Genrai Shuka (源来酒家)	35.695671	139.754409	Chinese Restaurant
12	Chiyoda	35.69381	139.753216	咸亨酒店	35.696010	139.756730	Chinese Restaurant



- Create a dataframe with top 10 restaurants based on occurrence for 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	Japanese Restaurant	Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Curry Restaurant	Italian Restaurant	Indian Restaurant
1	Arakawa	Chinese Restaurant	Indian Restaurant	Ramen Restaurant	Donburi Restaurant	Japanese Restaurant	Italian Restaurant	Yoshoku Restaurant	Korean Restaurant	Kebab Restaurant	Kaiseki Restaurant
2	Bunkyō	Chinese Restaurant	Ramen Restaurant	Szechuan Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Kebab Restaurant	Kaiseki Restaurant	Japanese Restaurant	Japanese Curry Restaurant	Italian Restaurant
3	Chiyoda	Chinese Restaurant	Ramen Restaurant	Japanese Curry Restaurant	Italian Restaurant	Sushi Restaurant	Indian Restaurant	Yoshoku Restaurant	Tempura Restaurant	Restaurant	Hotpot Restaurant
4	Chūō	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Unagi Restaurant	Soba Restaurant	Tonkatsu Restaurant	Yakitori Restaurant	Tempura Restaurant	Donburi Restaurant	Spanish Restaurant

- Run the Kmeans algorithm for clustering the neighborhoods

Results

Through a clustering of the neighborhoods, it is easy to see common characteristics of the different areas to make decisions about the location of the new restaurant. In each cluster, similar neighborhoods are listed and there is also demographic data about the neighborhood.

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Population	Density	Area
4	0	Chūō	Sushi Restaurant	Japanese Restaurant	Italian Restaurant	Unagi Restaurant	Soba Restaurant	147620	14460	10.21
8	0	Kita	Japanese Restaurant	Chinese Restaurant	French Restaurant	Sushi Restaurant	Ramen Restaurant	345063	16740	20.61
10	0	Meguro	Chinese Restaurant	Japanese Restaurant	Italian Restaurant	Sushi Restaurant	Yoshoku Restaurant	280283	19110	14.67
11	0	Minato	Soba Restaurant	Yakitori Restaurant	Indian Restaurant	Kebab Restaurant	Chinese Restaurant	248071	12180	20.37
16	0	Shinagawa	Japanese Restaurant	Donburi Restaurant	Restaurant	Sushi Restaurant	Soba Restaurant	392492	17180	22.84
17	0	Shinjuku	Japanese Restaurant	Yakitori Restaurant	Ramen Restaurant	Thai Restaurant	Chinese Restaurant	339211	18620	18.22
20	0	Taitō	Japanese Restaurant	Sushi Restaurant	Sukiyaki Restaurant	Unagi Restaurant	Chinese Restaurant	200486	19830	10.11

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Population	Density	Area
1	1	Arakawa	Chinese Restaurant	Indian Restaurant	Ramen Restaurant	Donburi Restaurant	Japanese Restaurant	213648	21030	10.16
2	1	Bunkyō	Chinese Restaurant	Ramen Restaurant	Szechuan Restaurant	Yoshoku Restaurant	Hotpot Restaurant	223389	19790	11.29
3	1	Chiyoda	Chinese Restaurant	Ramen Restaurant	Japanese Curry Restaurant	Italian Restaurant	Sushi Restaurant	59441	5100	11.66
5	1	Edogawa	Ramen Restaurant	Italian Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	685899	13750	49.90
7	1	Katsushika	Chinese Restaurant	Donburi Restaurant	Korean Restaurant	Dumpling Restaurant	Ramen Restaurant	447140	12850	34.80
12	1	Nakano	Ramen Restaurant	Indian Restaurant	Chinese Restaurant	Donburi Restaurant	Japanese Restaurant	332902	21350	15.59
14	1	Setagaya	Ramen Restaurant	Restaurant	Unagi Restaurant	Japanese Restaurant	Szechuan Restaurant	910868	15690	58.05
15	1	Shibuya	Chinese Restaurant	Ramen Restaurant	Mexican Restaurant	Seafood Restaurant	Asian Restaurant	227850	15080	15.11
18	1	Suginami	Italian Restaurant	Ramen Restaurant	Asian Restaurant	Chinese Restaurant	Tonkatsu Restaurant	570483	16750	34.06
19	1	Sumida	Chinese Restaurant	Unagi Restaurant	Tonkatsu Restaurant	Japanese Restaurant	Ramen Restaurant	260358	18910	13.77
21	1	Toshima	Ramen Restaurant	Chinese Restaurant	Japanese Restaurant	Yoshoku Restaurant	Soba Restaurant	294673	22650	13.01
22	1	Ōta	Ramen Restaurant	Japanese Restaurant	Chinese Restaurant	Tonkatsu Restaurant	Yoshoku Restaurant	722608	11910	60.66

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Population	Density	Area
6	2	Itabashi	Chinese Restaurant	Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	569225	17670	32.22
13	2	Nerima	Chinese Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Kebab Restaurant	726748	15120	48.08

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Population	Density	Area
0	3	Adachi	Japanese Restaurant	Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	674067	12660	53.25
9	3	Kōtō	Restaurant	Yoshoku Restaurant	Hotpot Restaurant	Korean Restaurant	Kebab Restaurant	502579	12510	40.16

Discussion

- The veracity of data depends exclusively on the data provided by FourSquare, so the results depend on that.
- The clustering analysis only analyzed the venues

Conclusion

The project is a simplification of the analysis that is needed to analyze the optimum location for a restaurant. This simplification only takes into account a few variables and applies a machine learning method to obtain insights from the data. This process consists in understanding the business problem, identifying the data required, processing the data, running the machine learning clustering algorithm, analyzing the results and presenting them.

.

- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
- Conclusion section where you conclude the report.