Exploring and Clustering Neighborhoods in Tokyo

Table of contents

1. Introduction / Business Problem

Whether your company is an established, mature business or a start-up, locating in a largely populated city has several benefits. The hustle and bustle of a large metropolis seems to transfer its energy to the businesses and residents. Whether the city is located on the West Coast, the East Coast or in the Midwest, they all share several advantages which majorly includes:

Workforce : A big city has more workers than a small city or town. The workforce may have more varied skills as well. In addition, a big city has a variety of levels of housing from extremely affordable to extremely pricey. The variety means employees from the lowest line staffer to vice presidents will find accommodations that are comfortable and that they can afford.

Infrastructure : Public transportation means it's easier and less costly for employees to get to work. If travel is on the business agenda, a large city will have at least one airport with direct flights to most cities. Infrastructure includes intangibles as well, such as business networks, professional organizations and small business development centers.

Diverse Customer Base : Cities often have a diverse cultural base. Businesses that target a certain segment of customers will do better in a big city that has a larger population of that segment. Businesses that offer a service or product to a narrow niche will find more customers in that niche in a larger population.

In this project we will try to find an optimal location for Grocery stores. Specifically, this report will be targeted to stakeholders interested in opening Grocery stores in Tokyo, Japan.

Since Grocery Stores are often cited as among the types of businesses with the lowest profit margins. It's true. Grocery store profit margins typically range from 1 percent to 3 percent, depending on the items. Grocery stores make their money on volume. They may not make much on any one item, that's why the store kindly provides big shopping carts for their customers. With not much additional effort, the grocery store sells you 20 items or more, making much more profit than they would have if you had bought only one item. Given This, we will employ our skills as a data scientist to identify highly populated neighborhood where competition are bound to be very minimal.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

2. Data Section

Based on definition of our problem, factors that will influence our decision are:

1. number of existing Boroughs/Wards and their Major neighborhoods.
2. number of inhabitants of each Wards that forms the neighborhood.
3. Most Common venues in Those neighborhood to give us an insight to the degree of Competiition if any.
4. We decided to use regularly spaced grid of locations, centered around Wards, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding.

Common Venues and Exploration of the neighborhood will be obtained using *Foursquare API**

Coordinate of Neighborhoods selected based on some criteria will be clustetered to highlight Similarities and Provide Better Understanding of the Neighborhoods.

3. Methodology

In this project we will direct our efforts on Identifying areas of Tokyo that have a a high poulation density.

In the first step we have collected the required data, processed the collected data and obtained the coordinates of the 23 special wards in tokyo using Google Maps API reverse geocoding

Second step in our analysis will make use of folium to visualize and explore Tokyo as well as the wards where Our clients will be siting his Grocery stores. The map will also be used in exploration of 'Population density of Major areas of Tokyo.

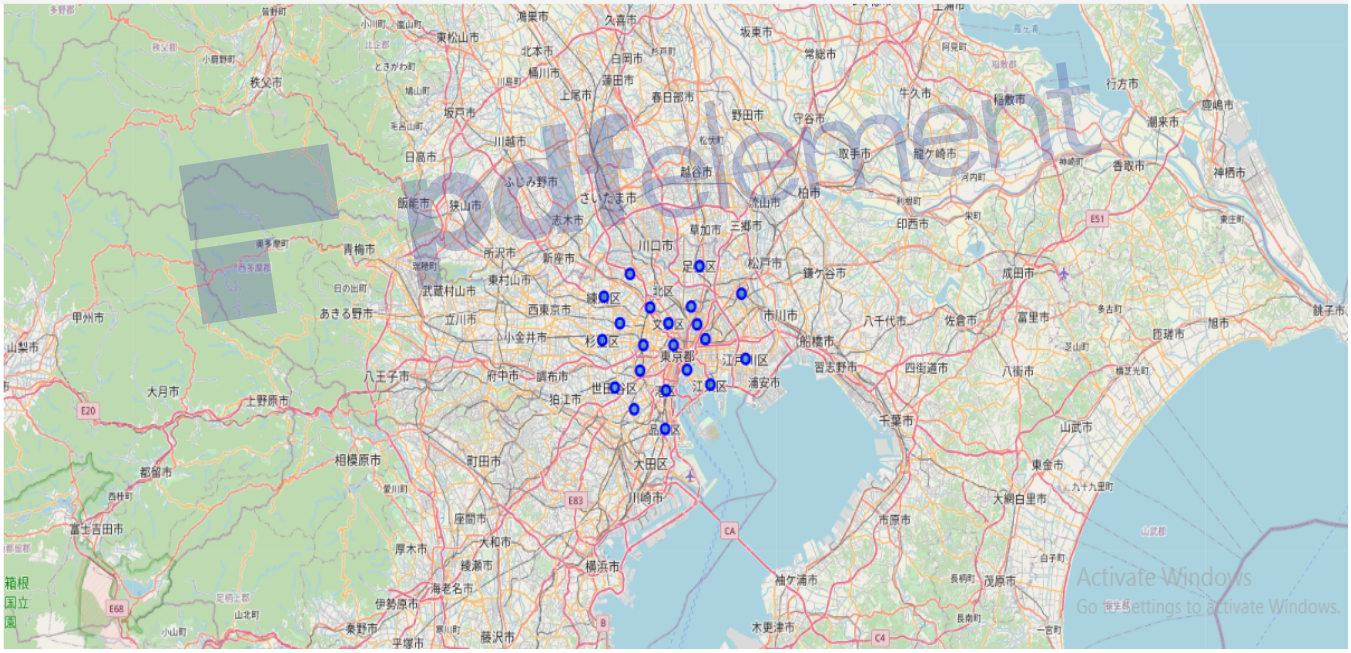
In third and final step we will focus on most promising areas and within these areas create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration The most populated locations as promising areas for a Grocery Business to florish.

Futhermore, we shall employ the FourSquare API to explore these densely populated areas. This areas will be Anayzed to further obtain the 5 most common venues.

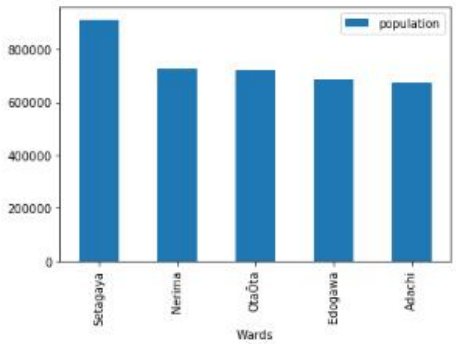
Finally, we will further Analyze these Neighborhoods and ceate clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final exploration and search for optimal venue locations by stakeholders.

3.1. Exploratoration and Analysis

map of Tokyo showing each wards and how populated they are using latitude and longitude values

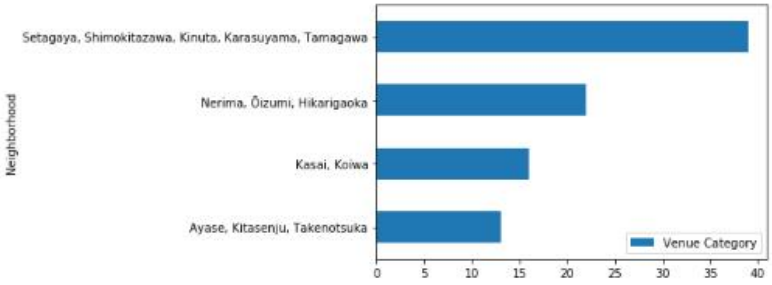


selecting the top 5 populated cities to be used for further Analysis



Exploring venues in tokyo's 5 most populated wards¶

Neighborhood	Venue Category
Ayase, Kitasenju, Takenotsuka	13
Kasai, Koiwa	16
Nerima, Ōizumi, Hikarigaoka	22
Setagaya, Shimokitazawa, Kinuta, Karasuyama, Tamagawa	39



The table above shows the number of venue categories as returned by Four square Api in each Tokyo wards and its corresponding Districts

10 most Common Venues in The 12 Neighborhoods returned by Four square Api in The selected Wards

	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	Ayase, Kitasenju, Takenotsuka	Convenience Store	Drugstore	Park	Restaurant	Italian Restaurant	Japanese Restaurant	Pharmacy	Discount Store	Supermarket	Sake Bar
1	Kasai, Koiwa	Convenience Store	Grocery Store	Bakery	Bath House	Supermarket	Italian Restaurant	Bridge	Ramen Restaurant	Deli / Bodega	Clothing Store
2	Nerima, Ōizumi, Hikarigaoka	Convenience Store	Intersection	Café	BBQ Joint	Bus Stop	Historic Site	Karaoke Box	Metro Station	Chinese Restaurant	Discount Store
3	Setagaya, Shimokitazawa, Kinuta, Karasuyama, T...	Convenience Store	Café	Bakery	Tram Station	Restaurant	Ramen Restaurant	Used Bookstore	Fast Food Restaurant	Drugstore	Diner

Clustering Neighborhoods in Tokyo's 5 Most populated Wards

Run k-means to cluster the neighborhood into 3 clusters

	Wards	population	Neighborhood	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	Setagaya	0910,868	Setagaya, Shimokitazawa, Kinuta, Karasuyama, T...	35.646096	139.65627	1.0	Convenience Store	Café	Bakery	Tram Station	Restaurant	Ramen Restaurant	Used Bookstore	Fast Food Restaurant	Drugstore	Diner
1	Nerima	0726,748	Nerima, Ōizumi, Hikarigaoka	35.74836	139.638735	1.0	Convenience Store	Intersection	Café	BBQ Joint	Bus Stop	Historic Site	Karaoke Box	Metro Station	Chinese Restaurant	Discount Store
2	OtaŌta	0722,608	Ōmori, Kamata, Haneda, Den-en-chōfu	-36.9673561	174.9407229	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Edogawa	0685,899	Kasai, Koiwa	35.678278	139.871091	2.0	Convenience Store	Grocery Store	Bakery	Bath House	Supermarket	Italian Restaurant	Bridge	Ramen Restaurant	Deli / Bodega	Clothing Store
4	Adachi	0674,067	Ayase, Kitasenju, Takenotsuka	35.783703	139.795319	0.0	Convenience Store	Drugstore	Park	Restaurant	Italian Restaurant	Japanese Restaurant	Pharmacy	Discount Store	Supermarket	Sake Bar

4. Results Section

We reached the end of the Analysis where we have identified and explored Venues in the most poulated areas of tokyo as stated in the business requirements. i have used Data from web sources like Wikipedia,Python, Geopy and Foursquare API, and below are the summary of the results obtained

1. Setagaya, Nerima, OtaŌta, Edogawa, Edogawa and Adachi was identified as the most populated wards in tokyo with a total population of 910868, 726748, 722608, 685899 and 674067 Respectively.

2. Setagaya toped the list comprising of 4 Major Districts which includes Shimokitazawa, Kinuta, Karasuyama, Tamagawa.

3. Using Foursquare API to explore these populated wards, a total of 18 Neighborhoods were returned, which were to be considered favorable foor establishing a Grocery Business. These Neighborhood was Analyzed and a total of 84 venue categories were identified of which the Setagaya Ward alone had a maximum of 38 venue categories.

Exploring the 5 most common Venues in Each ward and its Major district, we found that for Neighborhoods like

1. **Ayase, Kitasenju, Takenotsuka (Adachi Ward):** Venues identified in order of their frequencies are Convienience store, Drug Store, Japanese Restaurant, Park and Supermarkets.

2. **Kasai, Koiwa (Edogawa Ward):** Venues identified in order of their frequencies are Convienience Store, Grocery store, Deli, Italian Restaurant and Bath house.

3. **Nerima, Ōizumi, Hikarigaoka (Nerima Ward):** Venues identified in order of their frequencies are majorly Convienience Store, Cafe and Chinese Restaurant.

4. **Setagaya, Shimokitazawa, Kinuta, Karasuyama, Tamagawa(Setagaya Ward):** Venues identified in order of their frequencies are majorly Convienience Store, Cafe, Restaurants and Bakery.

Finally, These Wards were seperated into 3 clustered with k=means clustering yielding the bekow outcomes:

Cluster 1 : Adachi

Cluster 2 : Setagaya, Nerima

Cluster 3 : Edogawa

5. Discussion and Recommendation

According to the Result of the Analysis SETAGAYA and NERIMA are ideal to provide The Largest prospective customer Base to ensure a better profitability ratio, the work force required to operate large Grocery Stores and Accommodation that suites the different cadres of the employees owing to its large population Density when compared to other wards.

Furthermore, Since these areas also do not feature Grocery stores/Supermarket even at its 10 most common venues, the area will provide the least competition for an upcoming Grocery store. In addition it can be observed that common venues in the area are Convenience Store, restaurants, cafe, and Bakery which from intuition will appreciate the Establishment of Nearby Grocery store from which they can procure their Necessary Raw Materials, facilities and equipments required for Daily use. This will help curtail Transportion expense incurred on purchasing the above enumerated items from a farther Distance.

6. Conclusion

Purpose of this project was to identify ideal, Populated and less competitive areas in Tokyo in order to aid stakeholders in narrowing down the search for optimal location for a Establishment of Grocery stores. By Exploring Populated Wards and Neighborhoods and Analyzing most common venues from Foursquare data we have first identified general Wards that justify further analysis (Setagaya, Nerima, OtaŌta, Edogawa, Edogawa and Adachi), and then generated extensive collection of Neighborhood and venues which satisfy some basic requirements regarding establishment of Grocery stores. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal business locations will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), Availability of other commercial businesses, Convenience stores, Refreshment and recreational center, Accommodation cost and availability, Prices, Social and economic dynamics of every neighborhood etc.