Finding an appropriate location in Tokyo for foreign F&B owner

Neo Wing Tat Cheung

May 20, 2020

1. Introduction

1.1 Background

Tokyo is one of the most populous metropolitan in the world. Just Tokyo alone has over 15 million people. In addition, there are few hundred thousands people commute from suburb to Tokyo business area to work every day. Tokyo is a one of the most famous gourmet city in the world. Besides traditional Japanese food like sushi, ramen, tempura, other cuisine like Italian, Chinese and French are ranked in top tier globally. There are 226 Michelin-starred restaurants in Tokyo totally, which are the most in the cities around the globe.

Tokyo have 5 major business wards Chiyoda (千代田区), Chuo (中央区), Shinjuku (新宿区), Shibuya (渋谷区) and Shinagawa (品川区). Hundred thousands of white collar workers consume food, beer, coffee, etc in these business districts. There are numerous options to choose. Competition in this area is extremely high. If you are a foreign restaurant owner and planning to open a new venue in Tokyo the first time, it may be difficult to decide where to open. Would you choose the most populated area or should you choose less competitive suburb district. On the other hand, rent, the highest fixed cost expense in F&B business is also an important factor in making business decision. In this project, we will examine all these factors and advice business owners where they should open a new restaurant.

1.2 Problem

Data that might contribute to determining business owners where to open new venue might include cost (land price) and competition analysis (venues categories and occurrences in each category). This project aims to examine characteristics in consideration of land cost of all 23 wards and advice business manager where to open or avoid to open new venue base on distributions of venue categories.

2. Data Acquisition and cleaning

2.1 Data Sources

Wikipedia provides all Tokyo 23 wards and major neighborhoods names as a table format in the following page. Population, density and size data are also included in the dataset.

https://en.wikipedia.org/wiki/Special_wards_of_Tokyo#List_of_special_wards'

	No	Flag	Name	Kanji	Pop	Density	Size	Neighborhood
0	01		Chiyoda	千代田区	0059,441	05,100	011.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	02		Chūō	中央区	0147,620	14,460	010.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	03		Minato	港区	0248,071	12,180	020.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	04		Shinjuku	新宿区	0339,211	18,620	018.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	05		Bunkyō	文京区	0223,389	19,790	011.29	Hongō, Yayoi, Hakusan
5	06		Taitō	台東区	0200,486	19,830	010.11	Ueno, Asakusa
6	07		Sumida	墨田区	0260,358	18,910	013.77	Kinshichō, Morishita, Ryōgoku
7	08		Kōtō	江東区	0502,579	12,510	040.16	Kiba, Ariake, Kameido, Tōyōchō, Monzennakachō,...
8	09		Shinagawa	品川区	0392,492	17,180	022.84	Shinagawa, Gotanda, Ōsaki, Hatanodai, Ōimachi,...
9	10		Meguro	目黒区	0280,283	19,110	014.67	Meguro, Nakameguro, Jiyugaoka, Komaba, Aobadai
10	11		OtaŌta	大田区	0722,608	11,910	060.66	Ōmori, Kamata, Haneda, Den-en-chōfu
11	12		Setagaya	世田谷区	0910,868	15,690	058.05	Setagaya, Shimokitazawa, Kinuta, Karasuyama, T...
12	13		Shibuya	渋谷区	0227,850	15,080	015.11	Shibuya, Ebisu, Harajuku, Daikanyama, Hiroo, S...
13	14		Nakano	中野区	0332,902	21,350	015.59	Nakano
14	15		Suginami	杉並区	0570,483	16,750	034.06	Kōenji, Asagaya, Ogikubo
15	16		Toshima	豊島区	0294,673	22,650	013.01	Ikebukuro, Komagome, Senkawa, Sugamo
16	17		Kita	北区	0345,063	16,740	020.61	Akabane, Ōji, Tabata

To retrieve JPY/sqm land price, we access the following page.
We are making an assumption here that higher per unit price will lead to higher rent.

https://utinokati.com/en/details/land-market-value/area/Tokyo/

Location data like latitude and longitude of each district will be retrieved through Geopy client method. We further use Foursquare API to retrieve top 100 venues that are within a radius of 500 meters in each 23 wards' neighborhood.

	Name	Neighborhood	Latitude	Longitude
0	Chiyoda	Nagatachō	35.675618	139.743469
1	Chuo	Nihonbashi	35.684068	139.774503
2	Minato	Odaiba	35.619050	139.779364
3	Shinjuku	Shinjuku	35.693763	139.703632
4	Bunkyo	Hongō	35.175376	137.013476
5	Taito	Ueno	35.711759	139.777645
6	Sumida	Kinshichō	35.696312	139.815043
7	Koto	Ariake	35.634556	139.793256
8	Shinagawa	Shinagawa	35.599252	139.738910
9	Meguro	Meguro	35.621250	139.688014

2.2 Data cleaning

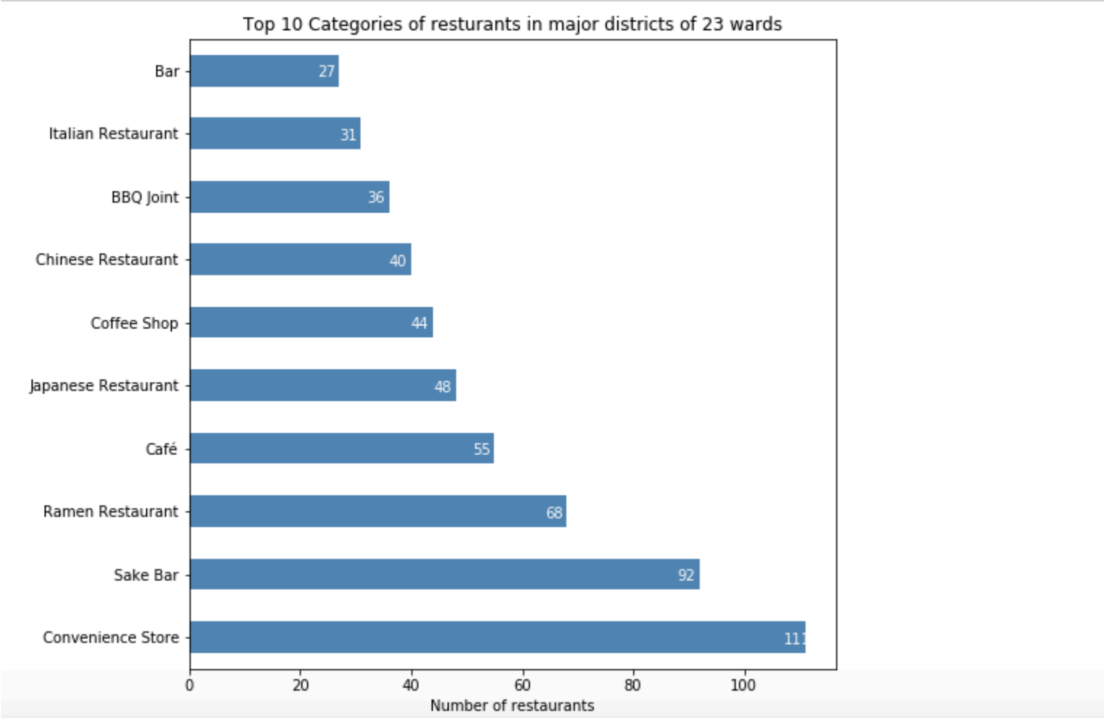
Since we don't use population, density and size in our analysis, we first remove them in the dataset. There are some spelling discrepancies under wards name and neighborhood name between data set and location database. We amend those to ensure that we capture right data.

In land price data set, we only use the average unit price column and omit average trading price column.

3. Exploratory Data Analysis

3.1 Top 10 Commercial Venue Categories

Foursquare API provides various venues data including hotel, park, restaurant, and library, etc in certain area. By summarizing them, we can examine the characteristic of each neighborhood. 1341 venues data points across 203 categories are retrieved thru API. We limit 100 venues per each district should be retrieved within a radius of 500 meters of the targeted district.



3.2 K means clustering of neighborhood

There are 23 wards in Tokyo. Some wards are more business focus. Some are more residential. Certain types of restaurants, for example high end Japanese cuisine may be suitable to open in business area because there are more business dinners. Family type restaurants may be suitable opened in residential area where there are more kids around.

One hot encoding helps to turn all 203 categories into a form of numbers for comparison purpose.

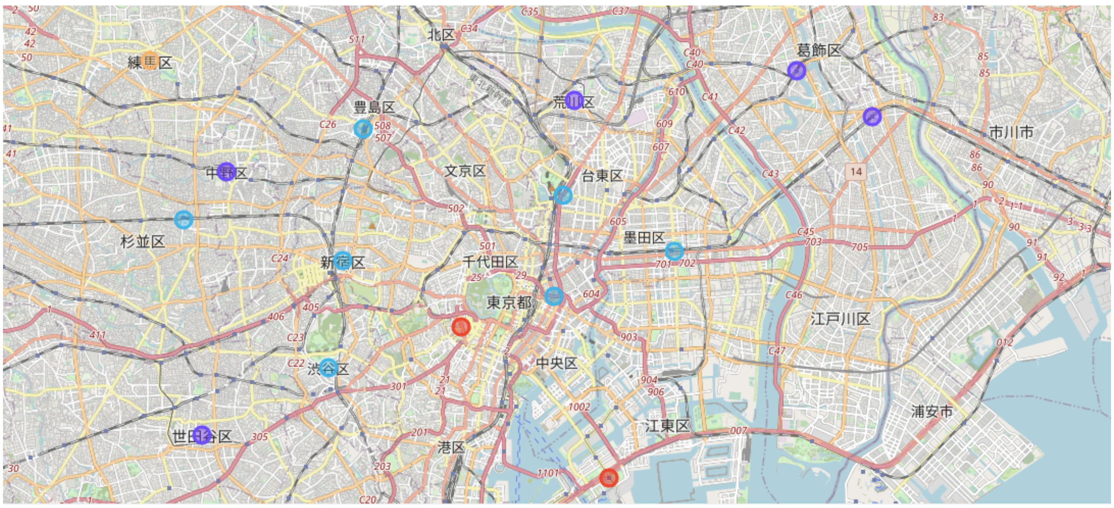
	Neighborhood	ATM	Accessories Store	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Auditorium	Auto Dealership	Auto Garage
0	Aoto	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021739	0.00	0.000000	0.000000	0.000000
1	Ariake	0.016129	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
2	Koiwa	0.000000	0.000000	0.000000	0.014706	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
3	Akabane	0.010000	0.000000	0.000000	0.010000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
4	Arakawa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
5	Ayase	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
6	Hongō	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
7	Ikebukuro	0.000000	0.000000	0.000000	0.000000	0.010000	0.000000	0.020000	0.00	0.000000	0.000000	0.000000
8	Itabashi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.052632
9	Kinshichō	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
10	Kōenji	0.000000	0.000000	0.010000	0.010000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
11	Meguro	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.037037	0.000000
12	Nagatachō	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
13	Nakano	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
14	Nerima	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
15	Nihonbashi	0.000000	0.000000	0.000000	0.000000	0.011628	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
16	Odaiba	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.025641	0.000000	0.000000
17	Setagaya	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
18	Shibuya	0.000000	0.020833	0.000000	0.000000	0.020833	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
19	Shinagawa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000
20	Shinjuku	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.01	0.000000	0.000000	0.000000
21	Ueno	0.000000	0.000000	0.011236	0.000000	0.000000	0.022472	0.000000	0.00	0.000000	0.000000	0.000000
22	Ōmori	0.030000	0.000000	0.000000	0.020000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000

	Neighborhood	BBQ Joint	Baby Store	Bagel Shop	Bakery	Bangladeshi Restaurant	Bar	Basketball Court	Bath House	Bed & Breakfast	Beer Bar	Beer Garden	Betting Shop	Bike Rental / Bike Share	Bike Shop	Bistro	Bookstore	Bou
0	Aoto	0.021739	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.021739	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.021739	
1	Ariake	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.016129	0.000000	0.000000	0.016129	
2	Koiwa	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.014706	
3	Akabane	0.040000	0.00	0.01	0.010000	0.00	0.030000	0.000000	0.000000	0.000000	0.010000	0.00	0.00	0.000000	0.000000	0.010000	0.010000	
4	Arakawa	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.028571	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
5	Ayase	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
6	Hongō	0.043478	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.043478	0.043478	0.000000	
7	Ikebukuro	0.020000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.020000	0.01	0.00	0.000000	0.000000	0.000000	0.010000	
8	Itabashi	0.052632	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
9	Kinshichō	0.030000	0.01	0.00	0.010000	0.01	0.010000	0.000000	0.000000	0.000000	0.000000	0.00	0.02	0.000000	0.000000	0.000000	0.000000	
10	Kōenji	0.030000	0.00	0.00	0.010000	0.00	0.050000	0.000000	0.010000	0.000000	0.010000	0.00	0.00	0.000000	0.000000	0.000000	0.020000	
11	Meguro	0.000000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
12	Nagatachō	0.000000	0.00	0.00	0.000000	0.00	0.031250	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.031250	0.000000	0.000000	0.000000	
13	Nakano	0.040000	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.040000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
14	Nerima	0.050000	0.00	0.00	0.050000	0.00	0.050000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.050000	0.000000	0.000000	0.000000	
15	Nihonbashi	0.046512	0.00	0.00	0.034884	0.00	0.023256	0.000000	0.000000	0.000000	0.011628	0.00	0.00	0.000000	0.000000	0.000000	0.011628	
16	Odaiba	0.000000	0.00	0.00	0.000000	0.00	0.051282	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
17	Setagaya	0.000000	0.00	0.00	0.050000	0.00	0.000000	0.000000	0.000000	0.000000	0.025000	0.00	0.00	0.000000	0.000000	0.025000	0.000000	
18	Shibuya	0.010417	0.00	0.00	0.010417	0.00	0.010417	0.010417	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.020833	
19	Shinagawa	0.103448	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	
20	Shinjuku	0.050000	0.00	0.00	0.010000	0.00	0.100000	0.000000	0.000000	0.010000	0.010000	0.00	0.00	0.000000	0.000000	0.000000	0.020000	
21	Ueno	0.022472	0.00	0.00	0.000000	0.00	0.000000	0.000000	0.011236	0.000000	0.022472	0.00	0.00	0.000000	0.000000	0.000000	0.011236	
22	Ōmori	0.040000	0.00	0.00	0.000000	0.00	0.010000	0.000000	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	

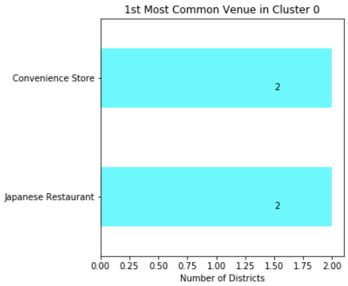
Kmeans clustering method optimizes the data set and groups 23 districts into 5 clusters.

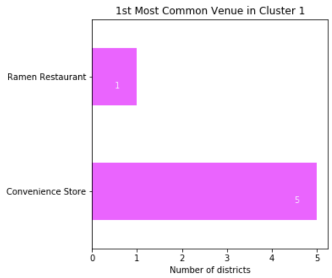
	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	1	Aoto	Convenience Store	Ramen Restaurant	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Sake Bar	Concert Hall
1	0	Ariake	Convenience Store	Bus Stop	Coffee Shop	Italian Restaurant	Plaza	Café	Hotel
2	1	Koiwa	Convenience Store	Ramen Restaurant	Japanese Restaurant	Discount Store	Coffee Shop	Sake Bar	Grocery Store
3	2	Akabane	Sake Bar	Convenience Store	Ramen Restaurant	Soba Restaurant	BBQ Joint	Coffee Shop	Bar
4	1	Arakawa	Convenience Store	Chinese Restaurant	Grocery Store	Park	Ramen Restaurant	Noodle House	Italian Restaurant
5	3	Ayase	Convenience Store	Park	Udon Restaurant	Gym	Donut Shop	Flea Market	Fish Market
6	1	Hongō	Convenience Store	Café	Japanese Restaurant	Mediterranean Restaurant	Deli / Bodega	Rest Area	Pizza Place
7	2	Ikebukuro	Sake Bar	Café	Sushi Restaurant	Yoshoku Restaurant	Japanese Restaurant	Coffee Shop	Dessert Shop
8	4	Itabashi	Convenience Store	Park	Italian Restaurant	Hobby Shop	Rest Area	Plaza	Chinese Restaurant

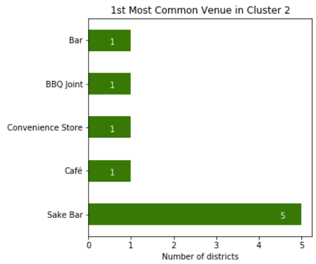
Distribution of cluster groups can be viewed in the following map. Same group is displayed in the same color.



3.3 We further examine each cluster's 1st most common venue ranking







In cluster 0, two districts found Japanese as the most found common venues. Sake bar dominates districts in cluster 2.

Since cluster 1 and cluster 4 districts are both dominated by convenience stores, we examine 2nd most common venue data. Two districts have ramen restaurant ranked the second. One has café and Chinese restaurant.

2nd Most Common Venue in cluster 1
Ramen Restaurant 2
Café 2
Chinese Restaurant 1
Convenience Store 1
Name: 2nd Most Common Venue, dtype: int64

2nd Most Common Venue in cluster 4
Café 1
Park 1
BBQ Joint 1
Name: 2nd Most Common Venue, dtype: int64

4. Discussion and results

In this exercise, we find that convenience store group is the most popular commercial venue category in 23 wards followed by sake bar and ramen shop. Over half of the districts have convenience store ranked as the top venue category. It is not a surprise because convenience store is not only a venue to buy food or drinks but also providing bill payment, cash dispenser facilities. The total numbers of cafe and coffee shops are roughly the same as the numbers of convenience stores. However, they are split into two groups so they are not shown up as the top venue. In terms of foreign cuisine, we find that there are more Chinese restaurants than Korean bbq and Italian cuisines. French cuisine is not ranked in top 10. Shijuku has more bar than other area. You can find record stores and rock clubs in Shibuya where is a popular area for young people. In suburb, there are more park and rest area than resturants.

Using K means for segmentation, 23 neighborhoods are grouped into 5 clusters. The largest cluster group 2 contains 9 districts. Five of them have Sake bar ranked as the most common venues followed. If you would like to open sake bar type restaurant, you can consider open in these area such as Shinjuku, Nihonbashi, Shibuya because there are already many of them in which there is synergy. Both Cluster 1 and 4 have convenience stores as their most ranked venues. Both of them contain 9 districts (Aoto, Koiwa, Arakawa,Hongo, Itabashi..) in total. Convenience store business owners can try to avoid opening new store in these 9 districts where competition is already severe.

Nagatacho and Odaiba are belonged to the same cluster 0. Their land price is at the top level in Tokyo over 2mm yen/sqm, 7 times higher than the lowest one on the list. We consider both districts are similar in terms of distribution of commercial venue type. A business owner may consider to search for business opportunity in both district at the same time.

5. Conclusion

In this project, we use geocoder to retrieve location data then we use foursquare to discover commercial venues category 500 meters radius within the designated district in 23 wards. For easy understanding purpose, we only include one major district in each ward. To explore the data analysis further, we utilize panda one hot encoding to calculate the mean of the frequency of occurrence of each venue category. By using clustering (K means) method, districts are categorized into 5 groups to show their similarity. We present the result in a leaflet map using Folium library. Same color dots are belonged to the same cluster.

To improve the results precisely, we could further include more districts into each ward. For example, Shibuya ward has Shibuya, Ebisu and Daikanyama. We make an assumption that higher land price generates higher rent. It would be more accurate to use actual commercial rental figure rather than average land price because deviation of land price is quite high across different area despite in the same ward.

We removed population, density data in this exercise. However, we could embed this data to calculate venues / population or density in the future to fine tune strategy affected by the extent of population.

Similarly, this type of data can be used to solve similar problem in other metropolitan cities. However, assumptions may need to be adjusted in order to achieve the best result.