Capstone Project - The Battle of Neighborhoods

Finding Best Residential Areas for Long-Stay Foreign Tourists Coming to Osaka

tm-sho

May 14, 2020

1. Introduction

Osaka is Japan's third largest city with more than 2.7 million people.

In recent years, Osaka has become one of the most popular tourist destinations in Japan particularly among foreign visitors.

Moreover, an increasing number of those tourists are repeated visitors and are looking for long-stay accommodations in rather non-traditional tourist areas.

Once we have overcome the COVID-19 pandemic, the foreign tourists are expected to come back.

Serving the needs of foreign tourists is essential for the future prospects of the city's economy.

1. Introduction

Objective of the Project:

Identify the best residential areas for long-stay foreign tourists through applying some of the machine learning methodologies and techniques.

Target Audiences:

1. Local travel agencies and tourist bureaus who want to attract more tourists to Osaka
2. Hotels and developers who want to invest and open new accommodations in different areas of Osaka

2. Data Sources

The following data were used to solve the problem:

1. *Ward Names of Osaka*: Wikipedia page of Osaka (https://en.wikipedia.org/wiki/Osaka)

   Data on the names of all the 24 wards in Osaka, their population, land area, and population density were scraped from a table of the Wikipedia page of Osaka.

2. *Coordinates of All Wards*: Geopy Client

   The coordinates of the 24 wards were obtained using geocoder class of Geopy client.

3. *Location Data of Popular Venues*: Foursquare (https://api.foursquare.com/v2/venues/)

   Data on popular venues in each ward and their categories were obtained using Foursquare API.

3. Methodology

To segment and cluster the wards of Osaka, K-means clustering was used to classify them into similar groups.

Exploratory Data Analysis and Visualization:

Before segmenting and clustering the wards, however, the summary statistics of the wards were generated.

	Ward	Ward_inKanji	Population	LandArea_SqKm	PopDensity_perSqKm
count	24	24	24.000000	24.000000	24.000000
unique	24	24	NaN	NaN	NaN
top	Sumiyoshi-ku	西区	NaN	NaN	NaN
freq	1	1	NaN	NaN	NaN
mean	NaN	NaN	114149.458333	9.385417	13768.500000
std	NaN	NaN	37605.994748	4.466964	4495.732759
min	NaN	NaN	62872.000000	4.390000	3381.000000
25%	NaN	NaN	82970.500000	6.057500	12333.250000
50%	NaN	NaN	107277.500000	8.375000	14397.500000
75%	NaN	NaN	131381.250000	10.915000	16853.250000
max	NaN	NaN	193282.000000	20.610000	20039.000000

3. Methodology

Population:

Population-wise, Hirano-ku (or Hirano ward) is the largest with about 193 thousand people, whereas Taisho-ku is the smallest with about 63 thousand people.

Osaka, as a whole, has a population of about 2.74 million people.

	Ward	Population
7	Hirano-ku	193282
23	Yodogawa-ku	182254
6	Higashiyodogawa-ku	176943
9	Jōtō-ku	167925
19	Sumiyoshi-ku	153425
10	Kita-ku	136602
8	Ikuno-ku	129641
5	Higashisumiyoshi-ku	126704
18	Suminoe-ku	120629
22	Tsurumi-ku	111501
16	Nishinari-ku	108654
13	Miyakojima-ku	107555
0	Abeno-ku	107000
15	Nishi-ku	103089
2	Chūō-ku	100998
17	Nishiyodogawa-ku	95960
1	Asahi-ku	90854
4	Higashinari-ku	83684
21	Tennōji-ku	80830
12	Minato-ku	80759
3	Fukushima-ku	78348
14	Naniwa-ku	74992
11	Konohana-ku	65086
20	Taishō-ku	62872

3. Methodology

Land Area:

In terms of the land area, the largest ward is Suminoe-ku with 20.61 sq. km, whereas the smallest is the Naniwa-ku with 4.4 sq. km.

The total land area of Osaka is 225.25 sq. km.

	Ward	LandArea_SqKm
18	Suminoe-ku	20.61
11	Konohana-ku	19.25
7	Hirano-ku	15.28
17	Nishiyodogawa-ku	14.22
6	Higashiyodogawa-ku	13.27
23	Yodogawa-ku	12.64
10	Kita-ku	10.34
5	Higashisumiyoshi-ku	9.75
20	Taishō-ku	9.43
19	Sumiyoshi-ku	9.40
2	Chūō-ku	8.87
9	Jōtō-ku	8.38
8	Ikuno-ku	8.37
22	Tsurumi-ku	8.17
12	Minato-ku	7.86
16	Nishinari-ku	7.37
1	Asahi-ku	6.32
13	Miyakojima-ku	6.08
0	Abeno-ku	5.99
15	Nishi-ku	5.21
21	Tennōji-ku	4.84
3	Fukushima-ku	4.67
4	Higashinari-ku	4.54
14	Naniwa-ku	4.39

3. Methodology

Population Density:

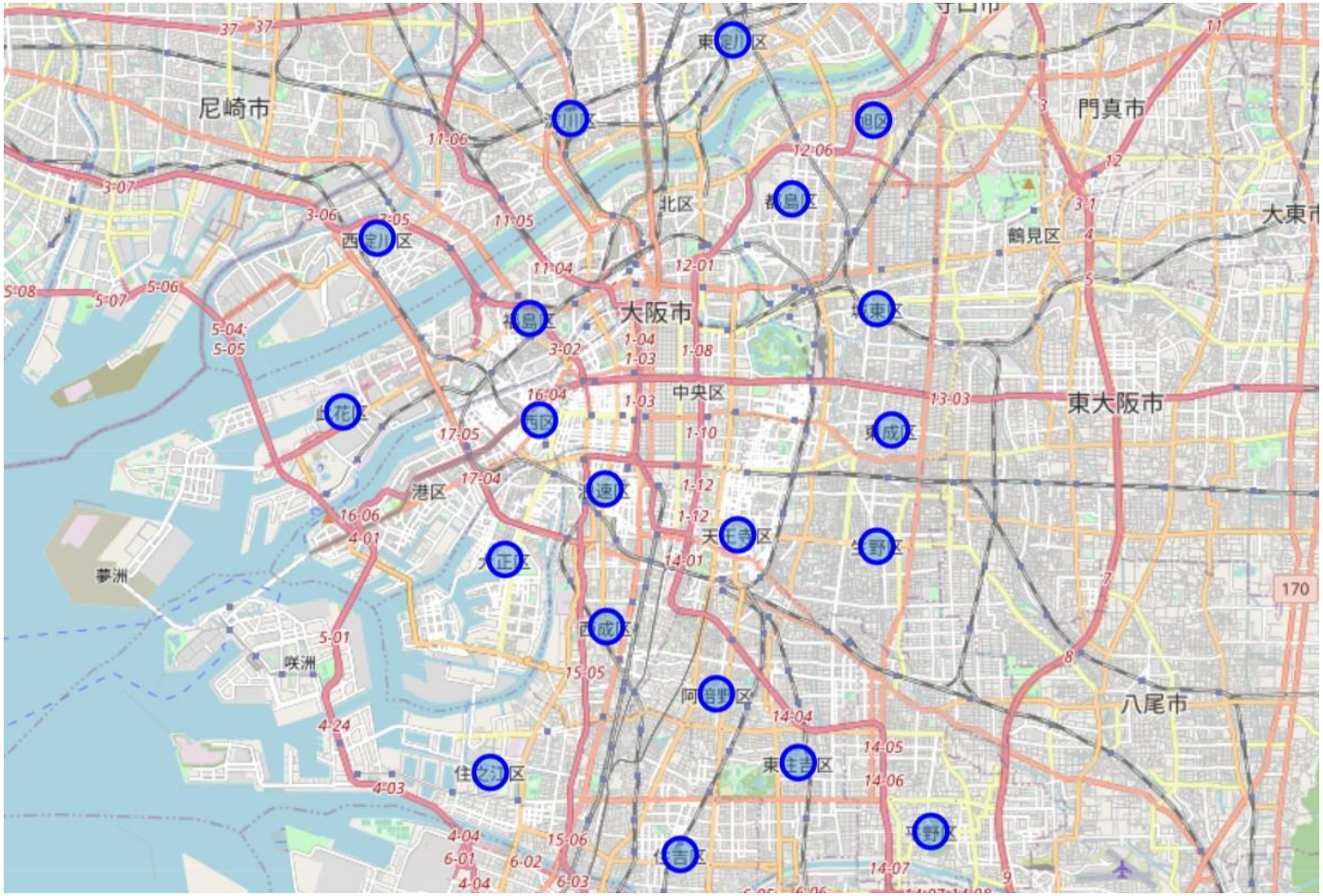
The average population density in Osaka is 12,162 people per sq. km.

Joto-ku is the most densely populated ward with 20,039 people per sq. km, whereas Konohana-ku is the least densely populated ward with 2,381 people per sq. km.

	Ward	PopDensity_perSqKm
9	Jōtō-ku	20039
15	Nishi-ku	19787
0	Abeno-ku	18440
4	Higashinari-ku	18433
13	Miyakojima-ku	17690
14	Naniwa-ku	17082
3	Fukushima-ku	16777
21	Tennōji-ku	16700
19	Sumiyoshi-ku	16322
8	Ikuno-ku	15489
16	Nishinari-ku	14743
23	Yodogawa-ku	14419
1	Asahi-ku	14376
22	Tsurumi-ku	13648
6	Higashiyodogawa-ku	13334
10	Kita-ku	13211
5	Higashisumiyoshi-ku	12995
7	Hirano-ku	12649
2	Chūō-ku	11386
12	Minato-ku	10275
17	Nishiyodogawa-ku	6748
20	Taishō-ku	6667
18	Suminoe-ku	5853
11	Konohana-ku	3381

3. Methodology

A map of Osaka was also created with its wards superimposed on it.



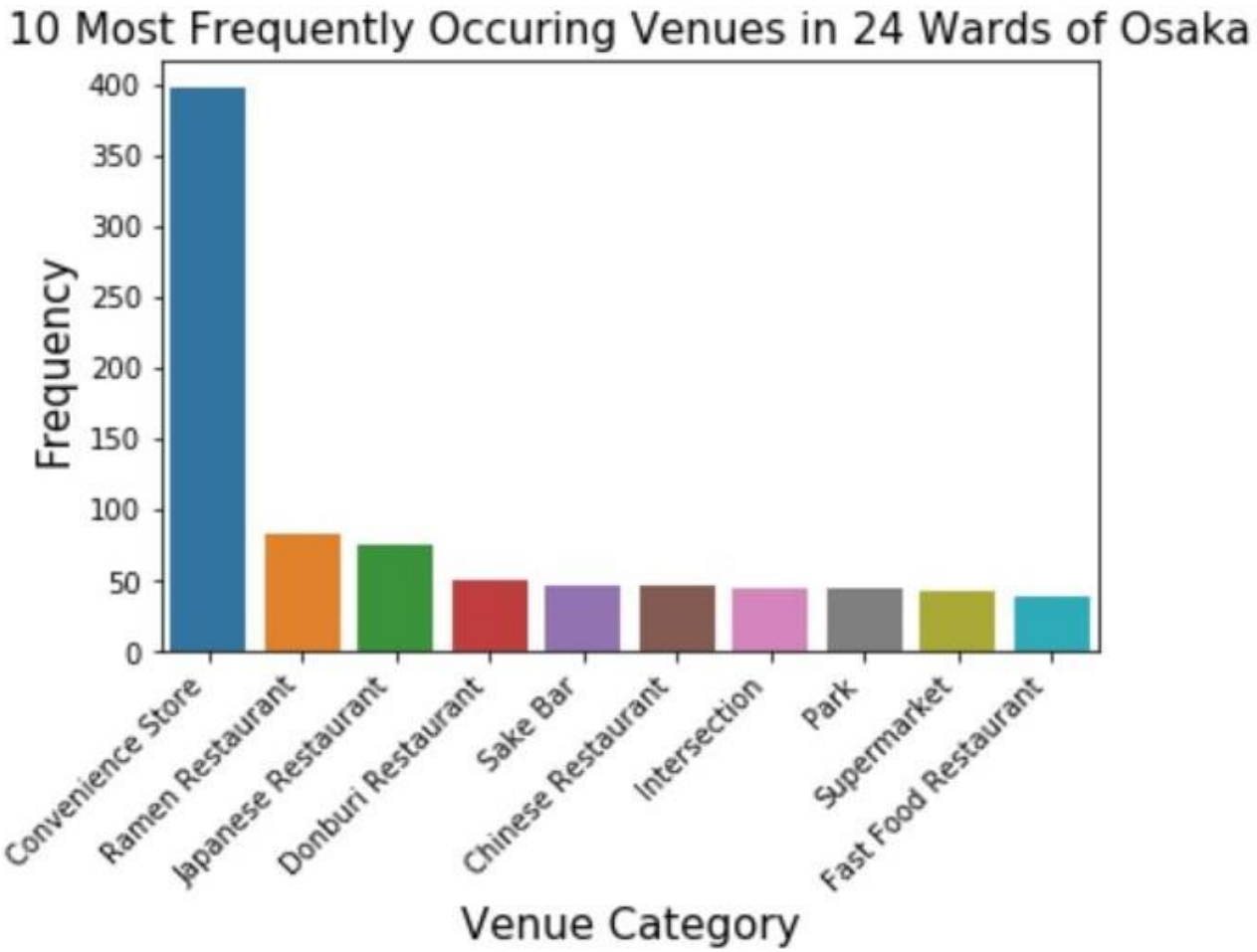
3. Methodology

Cluster Analysis:

To figure out the characteristics of the wards and classify them into similar groups based on the characteristics, cluster analysis was conducted.

As the information about the characteristics, data on up to 100 popular venues (such as restaurants and supermarkets) within a radius of 1 km of the center of the ward was retrieved for each ward, using the Foursquare API.

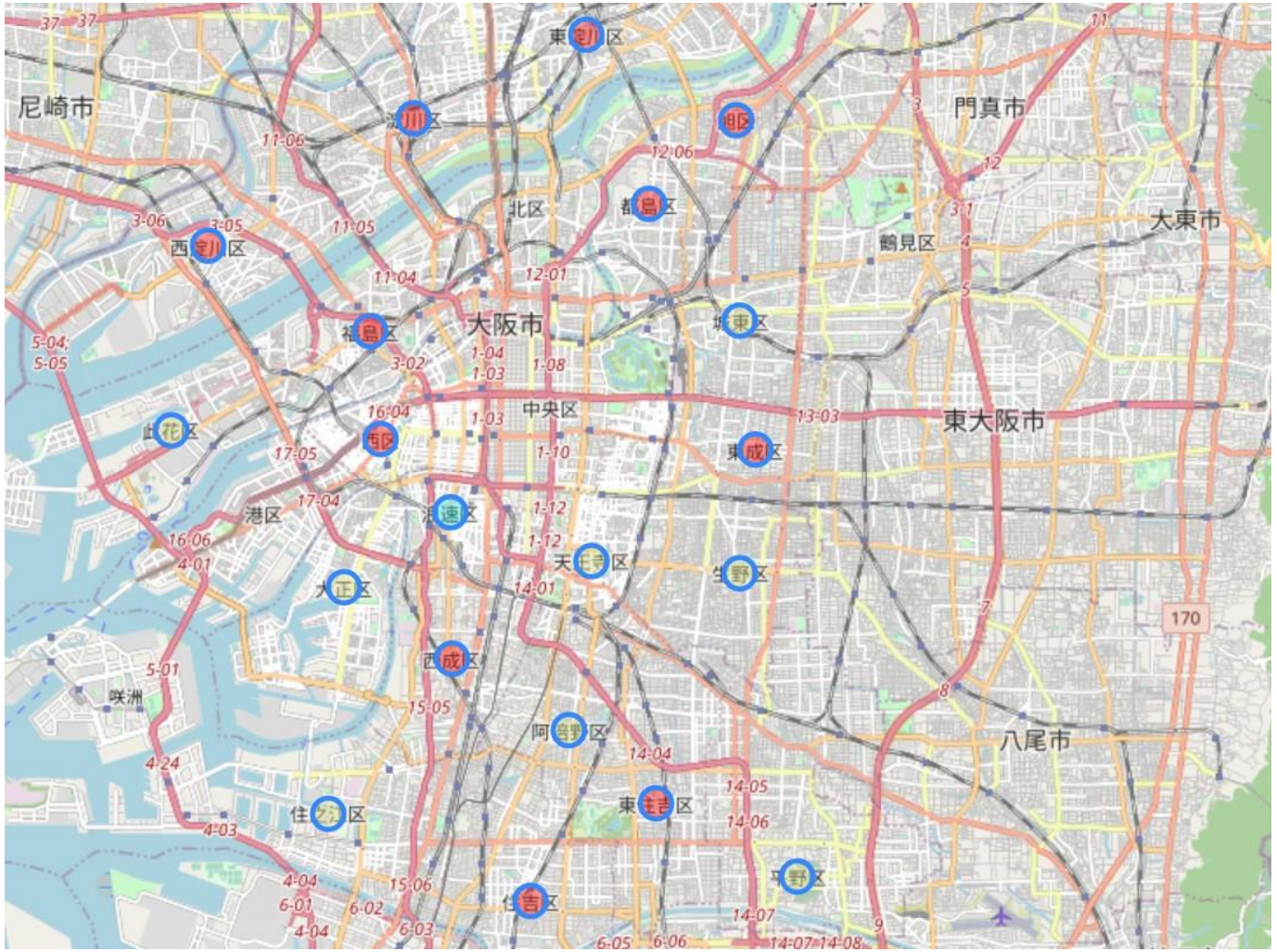
3. Methodology



The convenience store is the most frequently occurring or popular venue category in the 24 wards.

4. Results

Finally, using K-means clustering, all the 24 wards of Osaka were clustered into 4 groups.



4. Results

Cluster 1 has many casual eateries, such as Ramen and Donburi restaurants, and down-to-earth shops, such as grocery and clothing stores – not mentioning convenience stores.

	Ward	PopDensity_perSqKm	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
1	Asahi-ku	14376	34.726483	135.546952	0	Convenience Store	Ramen Restaurant	Udon Restaurant	Train Station	Metro Station	Supermarket	Donburi Restaurant	Noodle House	Intersection
3	Fukushima-ku	16777	34.692104	135.474812	0	Convenience Store	Ramen Restaurant	Fast Food Restaurant	Intersection	Train Station	Donburi Restaurant	Shopping Mall	Supermarket	BBQ Joint
4	Higashinari-ku	18433	34.672912	135.550567	0	Convenience Store	Donburi Restaurant	Sake Bar	Grocery Store	Metro Station	Intersection	Ramen Restaurant	Pharmacy	Park
5	Higashisumiyoshi-ku	12995	34.615662	135.531096	0	Convenience Store	Intersection	Ramen Restaurant	Restaurant	Sushi Restaurant	Bus Stop	Shopping Mall	Pharmacy	Donburi Restaurant
6	Higashiyodogawa-ku	13334	34.740212	135.517432	0	Convenience Store	Ramen Restaurant	Train Station	Donburi Restaurant	Chinese Restaurant	Supermarket	Drugstore	Japanese Restaurant	Fast Food Restaurant
13	Miyakojima-ku	17690	34.712703	135.529697	0	Convenience Store	Park	Ramen Restaurant	Donburi Restaurant	Restaurant	Japanese Restaurant	Fast Food Restaurant	Bakery	Chinese Restaurant
15	Nishi-ku	19787	34.674598	135.476774	0	Convenience Store	Shopping Mall	Train Station	Japanese Restaurant	Ramen Restaurant	Baseball Stadium	Clothing Store	Supermarket	Fast Food Restaurant
16	Nishinari-ku	14743	34.639074	135.490813	0	Convenience Store	Shopping Mall	Grocery Store	Train Station	Chinese Restaurant	Fast Food Restaurant	Donburi Restaurant	Supermarket	Japanese Restaurant
17	Nishiyodogawa-ku	6748	34.705938	135.442936	0	Convenience Store	Grocery Store	Park	Intersection	Clothing Store	Train Station	Japanese Restaurant	Shopping Mall	Fast Food Restaurant
19	Sumiyoshi-ku	16322	34.599765	135.506426	0	Convenience Store	Grocery Store	Chinese Restaurant	Platform	Ramen Restaurant	BBQ Joint	Diner	Noodle House	Fast Food Restaurant
23	Yodogawa-ku	14419	34.726613	135.483397	0	Convenience Store	Ramen Restaurant	Sake Bar	Supermarket	Donburi Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Shopping Mall	Fast Food Restaurant

4. Results

Cluster 2 or Tsurumi-ku features many scenic lookouts, parks, and trails, in addition to bus stops and train stations.

	Ward	PopDensity_perSqKm	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
22	Tsurumi-ku	13648	35.480132	139.693663	1	Bus Stop	Scenic Lookout	Park	Train Station	Trail	Pool	Yoshoku Restaurant	Food Court	Gift Shop

4. Results

Cluster 3 has a large concentration of traditional eating places, such as Sushi, Japanese, and Soba restaurants.

	Ward	PopDensity_perSqKm	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
2	Chūō-ku	11386	35.666255	139.775565	2	Sushi Restaurant	Japanese Restaurant	Monjayaki Restaurant	Soba Restaurant	Italian Restaurant	Bakery	Coffee Shop	Ramen Restaurant	Seafood Restaurant
12	Minato-ku	10275	35.643227	139.740055	2	Japanese Restaurant	Chinese Restaurant	Italian Restaurant	Soba Restaurant	Coffee Shop	Sake Bar	Indian Restaurant	BBQ Joint	Bistro
14	Naniwa-ku	17082	34.662830	135.490485	2	Café	Hotel	Sake Bar	Japanese Restaurant	Coffee Shop	Grocery Store	Udon Restaurant	Ramen Restaurant	BBQ Joint

4. Results

Cluster 4 has a lot of convenience stores, supermarkets, and casual restaurants, including Ramen and fast food restaurants.

	Ward	PopDensity_perSqKm	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
0	Abeno-ku	18440	34.627501	135.514095	3	Convenience Store	Grocery Store	Bakery	Supermarket	Park	Shopping Mall	Café	Donburi Restaurant	Sake Bar
7	Hirano-ku	12649	34.603715	135.559027	3	Convenience Store	Chinese Restaurant	Ramen Restaurant	Japanese Restaurant	Italian Restaurant	Donburi Restaurant	Supermarket	Metro Station	Udon Restaurant
8	Ikuno-ku	15489	34.653003	135.547722	3	Convenience Store	Ramen Restaurant	Chinese Restaurant	Drugstore	Intersection	Italian Restaurant	Korean Restaurant	Udon Restaurant	Clothing Store
9	Jōtō-ku	20039	34.693887	135.547769	3	Convenience Store	Supermarket	Ramen Restaurant	Restaurant	Fast Food Restaurant	Grocery Store	Chinese Restaurant	Park	Intersection
10	Kita-ku	13211	35.755838	139.736687	3	Convenience Store	Park	Ramen Restaurant	Intersection	Sake Bar	Japanese Restaurant	Grocery Store	Museum	History Museum
11	Konohana-ku	3381	34.676245	135.435550	3	Convenience Store	Intersection	Train Station	Theme Park Ride / Attraction	Park	Hotel	Bus Stop	American Restaurant	Theme Park
18	Suminoe-ku	5853	34.614132	135.466545	3	Convenience Store	Fast Food Restaurant	Ramen Restaurant	Japanese Restaurant	Drugstore	Donburi Restaurant	Discount Store	BBQ Joint	Light Rail Station
20	Taishō-ku	6667	34.650640	135.469570	3	Convenience Store	Pier	Chinese Restaurant	Shopping Mall	Fast Food Restaurant	Grocery Store	Donburi Restaurant	Ramen Restaurant	Discount Store
21	Tennōji-ku	16700	34.655043	135.518370	3	Convenience Store	Donburi Restaurant	Japanese Restaurant	Historic Site	Coffee Shop	Ramen Restaurant	Sake Bar	BBQ Joint	Tonkatsu Restaurant

5. Discussion

The findings would suggest Cluster 2 or Tsurumi-ku has the most potential for long-stay foreign visitors to Osaka.

Unlike many other wards, it has a lot of green space like parks and trails.

Besides, it is conveniently located and close enough to busy business and shopping districts, as suggested by the frequently occurring Foursquare venue categories of bus stops and train stations.

(Indeed, it takes only five minutes to Kyobashi, and about 20 minutes either to Umeda or Shinsaibashi.)

5. Discussion

The Caveats:

The K-means clustering outcome is influenced by the number of clusters and the selection of initial centroids.

The clustering in this analysis was conducted using the most popular venue data obtained from Foursquare alone and ignores all other factors.

There are other clustering techniques in machine learning, such as mean-shift clustering and density-based spatial clustering of applications with noise (DBSCAN).

6. Conclusion

To identify the best residential areas for long-stay foreign tourists, this project applied the K-means clustering technique to the Foursquare location data of popular venues and found Tsurumi-ku as the best candidate.

As a next step, this project may take into account other potentially important factors, such as crime and cost of living, to improve the precision of the analysis.