

# Capstone Project

## Setting up Chinese Restaurant in the Greater Johor Bahru Region

Author: Y.D. Lim

### Abstract

In this capstone project, a research on a market-entry problem is presented. Johor Bahru, one of the biggest cities in Malaysia, is explored for the favorable locations of opening a Chinese restaurant. Useful data is collected and *k*-means clustering machine learning modelling is performed to sort out the similar neighborhoods. A few most promising neighborhoods and their advantages are discussed, and the potential neighborhoods to operate high-end, mid-end and family-owned (small-sized) Chinese restaurants are recommended, respectively. Lastly, it is hoped that the insights generated from the analysis will be valuable for stakeholders who are interested in opening a Chinese restaurant in the city.

Keywords: Chinese restaurant, Johor Bahru, Malaysia, *k*-means, neighborhoods

# Table of Contents

Abstract .....	1
Table of Contents .....	1
Introduction.....	3
Data .....	4
Methodology .....	6
Result .....	6
Discussion .....	16
Conclusion .....	20
Reference .....	21

# Introduction

Johor Bahru is the capital city of the Johor state in Malaysia. It is the second largest city of Malaysia (following Kuala Lumpur), with a total population of about 1,061,950 and densities of 814/km<sup>2</sup>, 1200/km<sup>2</sup> and 2,259/km<sup>2</sup> in the metro, urban and city areas, respectively [1]. The greater Johor Bahru region consists of Johor Bahru city centre, Pasir Gudang, Pulai, Gelang Patah, Kulai and Tebrau. The region is well-known of its multi-racial cultures and religions, where Malay and Chinese contribute about 48.4 % and 37.6% to the total population. In addition, Johor Bahru is vicinal to the Republic of Singapore and the Senai International Airport, which have enhanced the tourism development in the region. With the support of Iskandar Development Region project [2], the population and economic in the greater Johor Bahru area is expected to grow drastically. Hence, it is worth to explore the area for potential business opportunities.

This report will target stakeholders who are interested in opening a Chinese restaurant in the greater Johor Bahru region. The neighborhoods covered by the region will be identified and analysed. A few potential neighborhoods will be discussed so that the stakeholders may choose the best possible location to open a Chinese restaurant.



Figure 1 The Iskandar Malaysia project development region [3]

## Data

To search for the potential locations to open a Chinese restaurant, the following data is required for the analysis:

- Number of existing restaurants (any) in the neighborhoods
- Number of Chinese restaurants in the neighborhoods, and their distances to the centres of the corresponding neighborhoods

The neighborhoods data is extracted from the geojson file published on the Github repository by leowmjw (<https://github.com/TindakMalaysia/Johor-Maps>) [4], as shown in the DataFrame in Figure 2. The database shows the administrative states and areas of the entire Johor state. The data used in the analysis is described in Table 1.

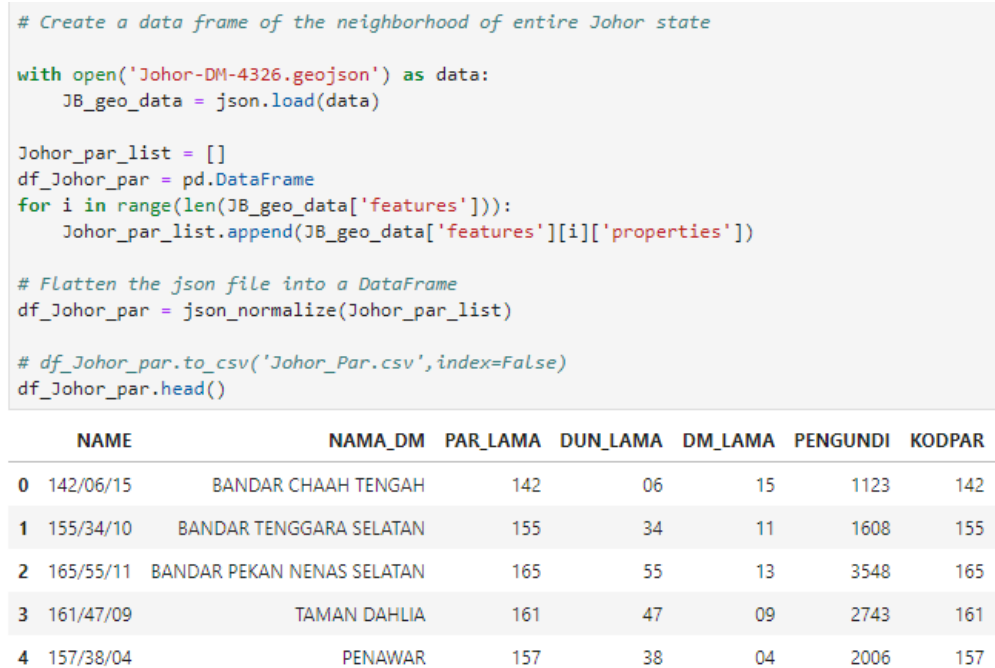


Figure 2 The extracted data of the Johor state from the geojson file[4]

Table 1 Original attributes and their descriptions

No.	Attributes	Description
1	NAMA_DM	Name of the neighborhoods/administrative areas, (‘Neighborhood’)
2	PAR_BARU	Name of the belonging districts of the neighborhoods/administrative areas (‘District’)
3	geometry.type	The type of the geometry border of neighborhood
4	geometry.coordinates	The coordinates plots of the border of neighborhood (‘Latitude’, ‘Longitude’)
5	PENGUNDI	Number of registered voters in the neighborhood, which are also local residents aged 21 or above (‘Number_of_Residents’)

The mean values of the *'geometry.coordinates'* are calculated and treated as the centre coordinates of the corresponding neighborhoods due to the unstable response from the geopy geocoder package. Additionally, to search for the nearby restaurants with respect to the centre of neighborhood, Foursquare API is employed in the analysis [5].

## Methodology

The following methodologies are applied in the analysis:

1. Extract the neighborhood data from the geojson file described in Section Data, calculate the centres coordinates by taking the mean values of the border geometry coordinates of each neighborhood.
2. Searches for the nearby food-related venues in each neighborhood that are within 1 km from the neighborhood centre using Foursquare API.
3. Identifies the similar neighborhoods and those with potential of having higher Chinese population by applying one-hot encoding and *k*-means clustering technique to the food-related venues found in Step 2.
4. Sorted out the Chinese restaurants based on the Foursquare API categories, applies DBSCAN clustering technique to detect the neighborhoods with the highest density of Chinese restaurants (and people whose ethnicity are Chinese).
5. Find the cluster with nearest distance to the city centre.
6. Finalized the potential neighborhoods by considering additional factors such as the number of residents, the distance between the cluster centroids and city centre and the number of Chinese restaurants (localized competition).

## Result

A total of 922 neighborhoods are extracted from the geojson file as described in Section Data. This data is rebuilt to a DataFrame which comprises only the neighborhoods in the 6 districts of the greater Johor Bahru region, namely Johor Bahru (district), Pasir Gudang, Pulai, Gelang Patah, Tebrau and Kulai. Figure 3 shows the first 5 rows of the DataFrame out of 235 rows and 6 columns, where the columns are renamed as presented in Table 1.

(235, 6)

	Neighborhood	District	State	Latitude	Longitude	Number_of_Residents
0	TAMAN DAHLIA	PULAI	JOHOR	1.511644	103.693691	2743
1	SETIA ENAU	TEBRAU	JOHOR	1.580372	103.757401	3168
2	TAMAN DAMAI JAYA	GELANG PATAH	JOHOR	1.506351	103.651338	4668
3	KONG KONG	TEBRAU	JOHOR	1.527090	103.973883	1022
4	TAMAN PUTRI	KULAI	JOHOR	1.659689	103.583704	2295

Figure 3 DataFrame of the greater Johor Bahru region

The choropleth map based on the residents of each neighborhood is plotted in Figure 4. The black-lined circle markers represent the calculated centres of the neighborhoods while the red-dots represent the coordinate of the district as labelled. It is visible that the neighborhood density in the Johor Bahru district is much higher than the remainders.

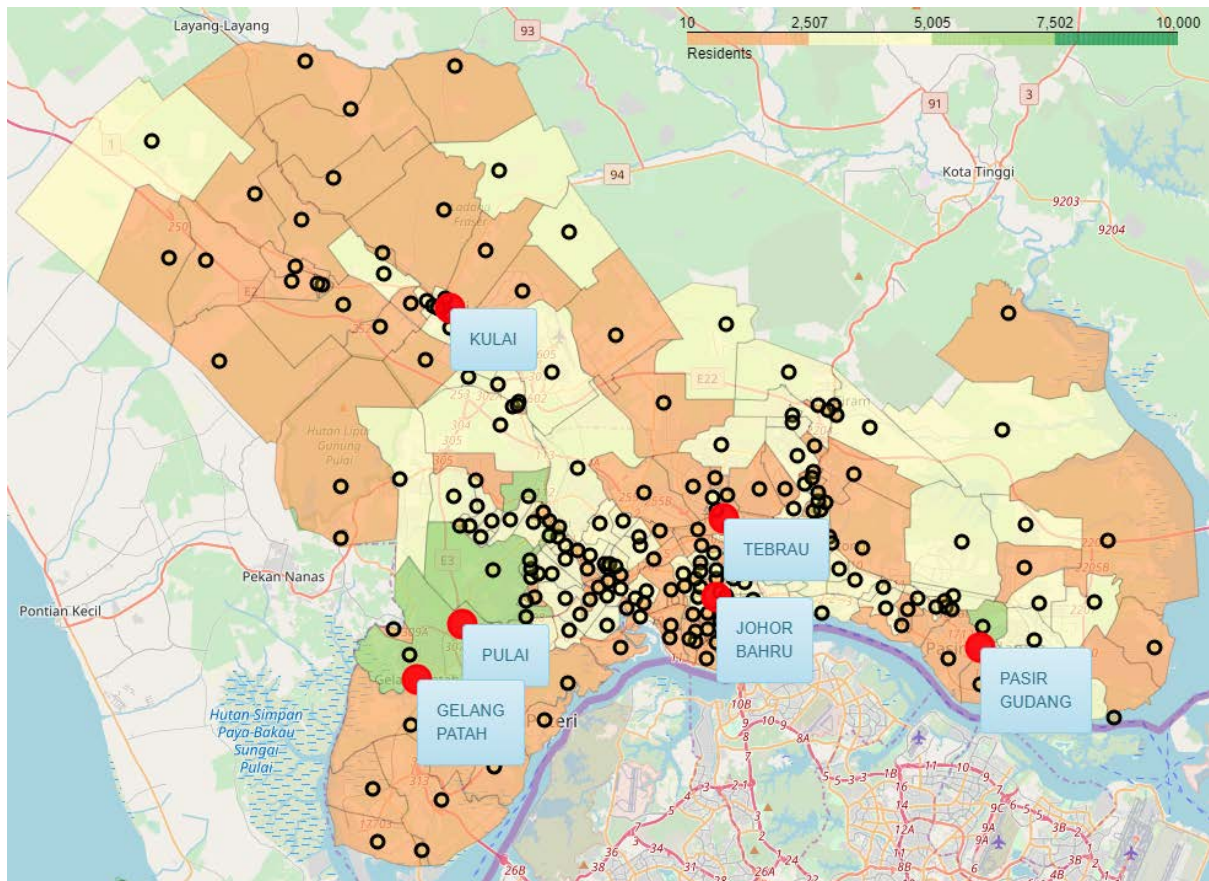


Figure 4 Choropleth map of neighbourhoods of the greater Johor Bahru region

To understand the neighborhoods better, Foursquare API is used to search for the food-related nearby venues around the neighborhoods. This is done by inputting the centre coordinates of the neighbourhood, and the nearby venues are defined as the venues which are within 1.0 KM from the coordinate. A limit of 100 venues is set for the Foursquare API to search for the nearby



venues. 7545 food-related venues are obtained over a span of 304 venue categories. The first 5 entries of each neighbourhood are mapped in Figure 5. From the figure, it can be seen the commercial areas are likely within the Johor Bahru, Pulai and Tebrau districts.

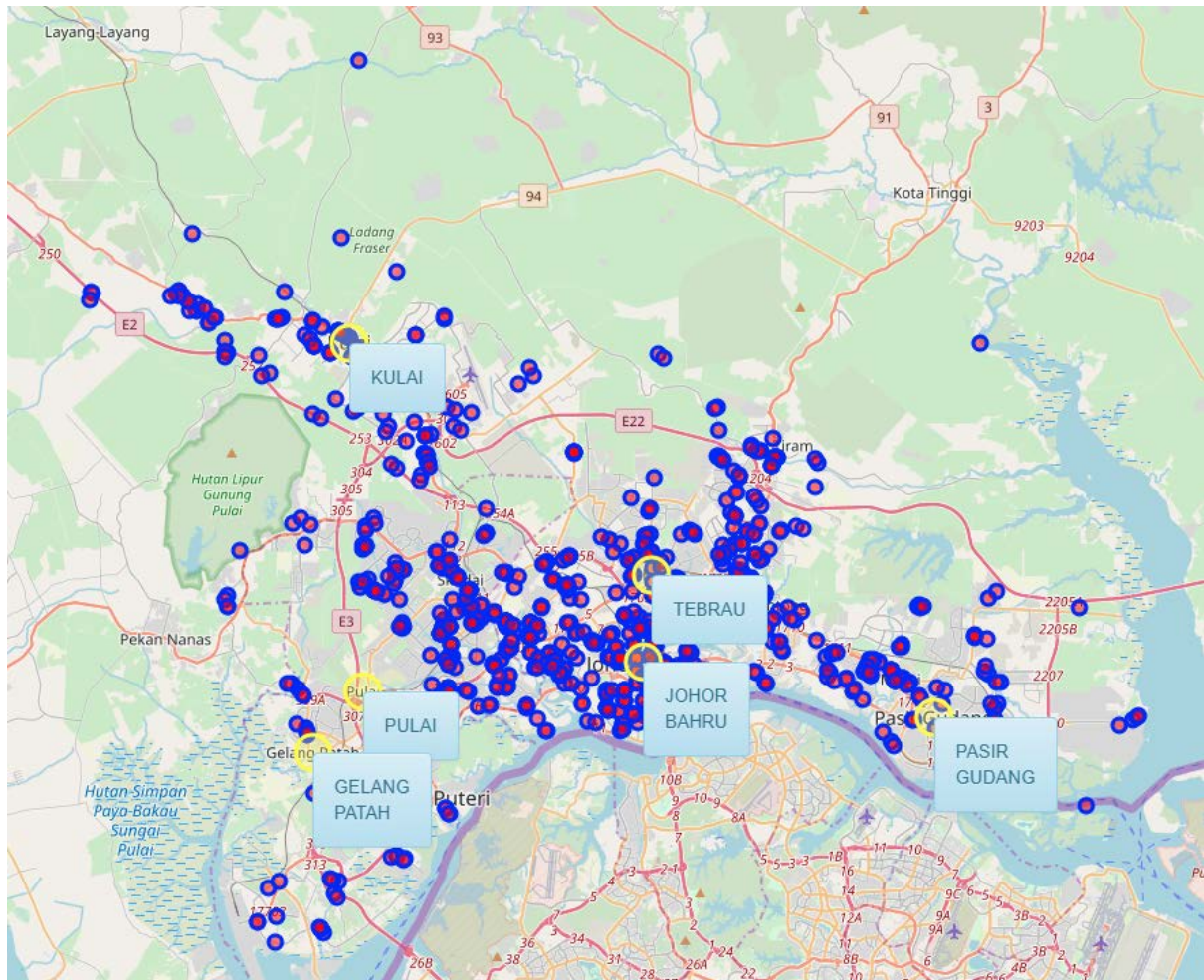


Figure 5 Plot of first 5 nearby venues of each neighbourhood

One-hot encoding is applied to the nearby venue categories. By grouping the venue categories with their mean frequencies of appearance in the neighbourhood, the most common venues can be sorted out. Figure 6 displays the top three most common venues of each neighbourhood.

	Town_Neighborhood	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	ABAD JAYA 1	JOHOR BAHRU	Chinese Restaurant	Hotel	Malay Restaurant
1	ABAD JAYA 2	JOHOR BAHRU	Chinese Restaurant	Hotel	Malay Restaurant
2	ARA CEMERLANG	TEBRAU	Chinese Restaurant	Asian Restaurant	Café
3	AYER BEMBAN	KULAI	Malay Restaurant	Shopping Mall	Piano Bar
4	AYER MOLEK	PULAI	Coffee Shop	Café	Malay Restaurant

Figure 6 Three most common venue categories



After that, the neighbourhoods are clustered based on the *k*-means clustering technique as shown in Figure 7. The identical neighborhoods are grouped under the same cluster where Table 2 listed the information of them.

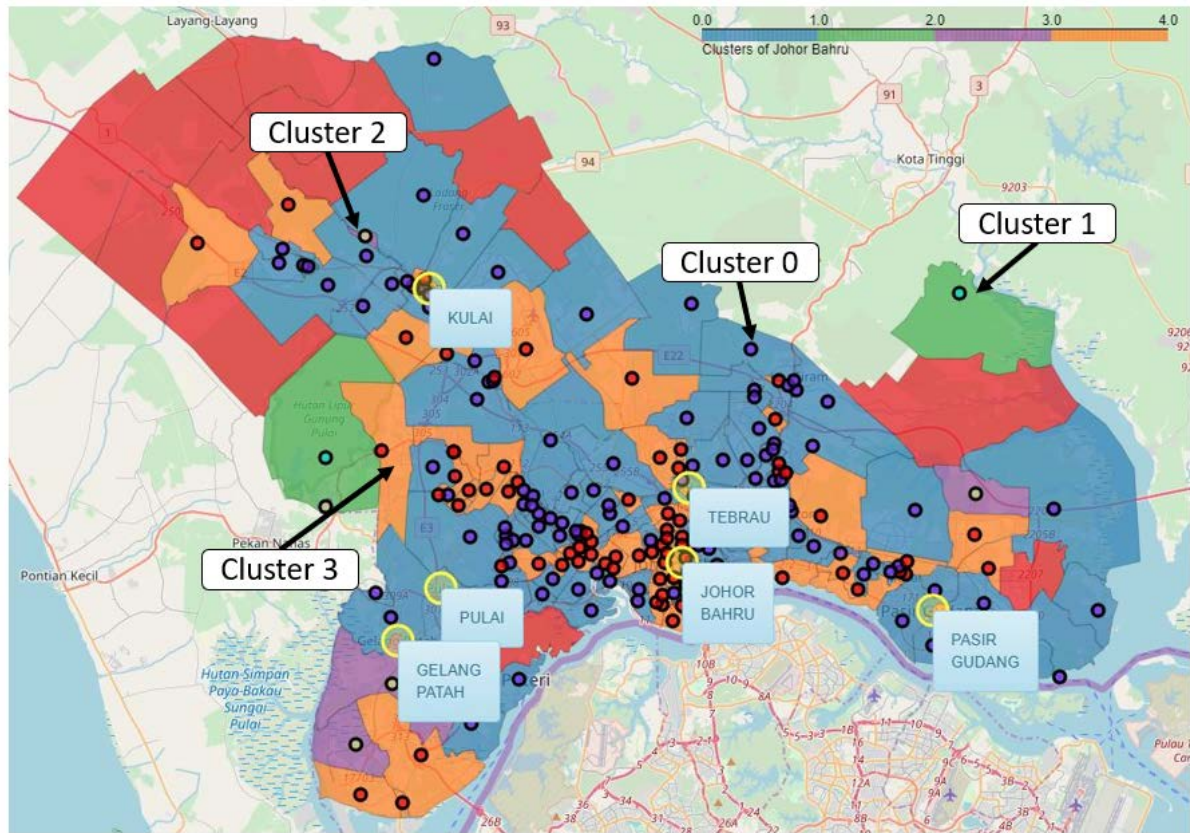


Figure 7 Clusters of neighbourhoods

Table 2 Number of top 3 common venues of each cluster

Cluster No.	Area Color	Number of Neighborhoods	Number of Top 3 Common Venues
0	Blue	141	Chinese Restaurant: 33 Malay Restaurant: 19 Asian Restaurant: 19
1	Green	2	Harbor/Marina: 2 Zoo Exhibit: 2 Food: 19
2	Purple	5	Asian Restaurant: 4 Zoo Exhibit: 2 Food & Drink Shop: 2
3	Orange	74	Malay Restaurant: 60

			Asian Restaurant: 19
			Asian Restaurant: 7

From Table 2, most Chinese restaurants are nearby to the neighbourhoods in Cluster 0. Hence, the neighborhoods in this cluster are believed to have higher Chinese population compare to the people of other races.

Table 3 Number of Chinese restaurants in Cluster 0

	Frequency
Chinese Restaurant	428
Asian Restaurant	345
Dim Sum Restaurant	15
Szechuan Restaurant	6
Taiwanese Restaurant	5
Hunan Restaurant	4
Cantonese Restaurant	3
Chinese Breakfast Place	2
Hakka Restaurant	1
Dongbei Restaurant	1
Chinese Aristocrat Restaurant	1

After filtering the neighbourhoods in Cluster 0 according to the categorization provided by Foursquare API [6], it is discovered that there are 811 Chinese restaurants from 115 neighbourhoods which are distributed within 1 KM from the corresponding neighbourhood centres. The categories and number of Chinese restaurants are tabulated in Table 3 and the distribution of which is plotted in Figure 8.

However, the neighborhoods in Cluster 0 may contain residential areas and commercial areas. The residential area may have a lower density of Chinese restaurant (lower number of Chinese restaurants within 1.0 KM from the neighborhood), while the commercial areas would have higher density of Chinese restaurants. To separate the possible commercial and residential areas from these 115 neighborhoods, DBSCAN clustering technique is applied to the dataset by fitting both the neighborhood coordinates and the coordinates of Chinese restaurants to the machine learning training. Figure 9 illustrated the ‘outliers’ from the result, which are the

possible residential areas or neighborhoods with low density of Chinese restaurants. As seen from Table 4, the number of Chinese restaurants in these neighborhoods is 5 or below.

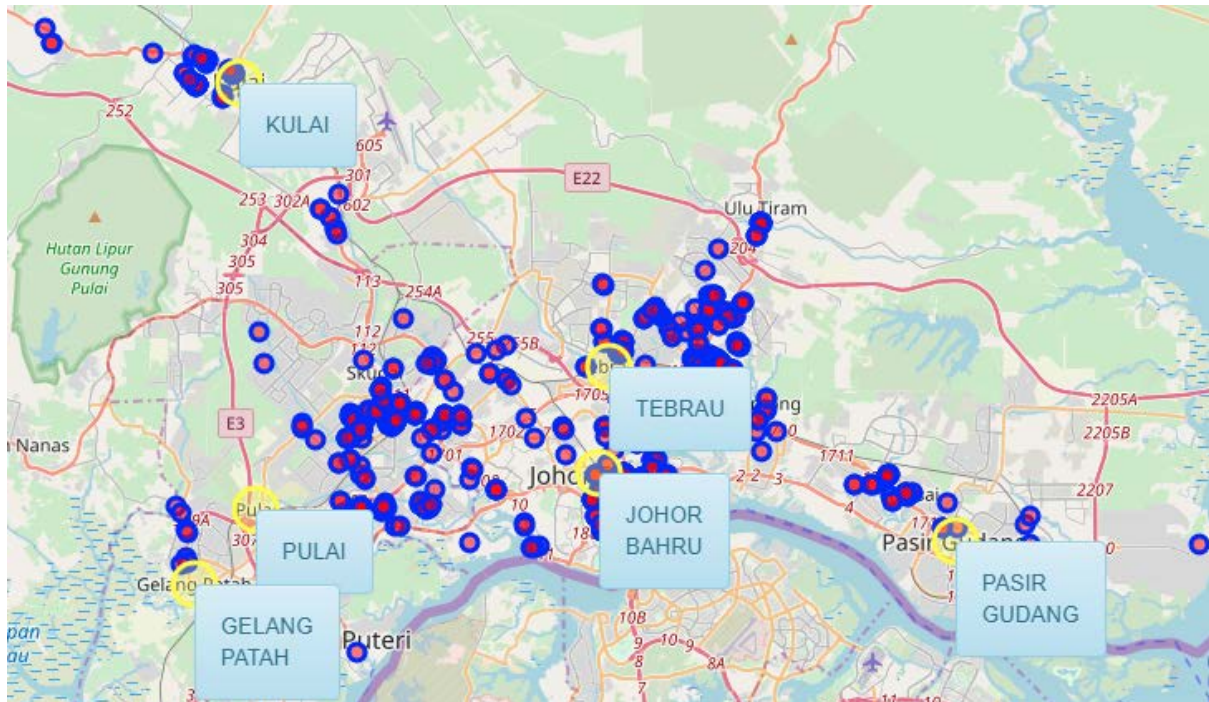


Figure 8 Plot of the Chinese restaurants in the neighbourhoods of Cluster 0

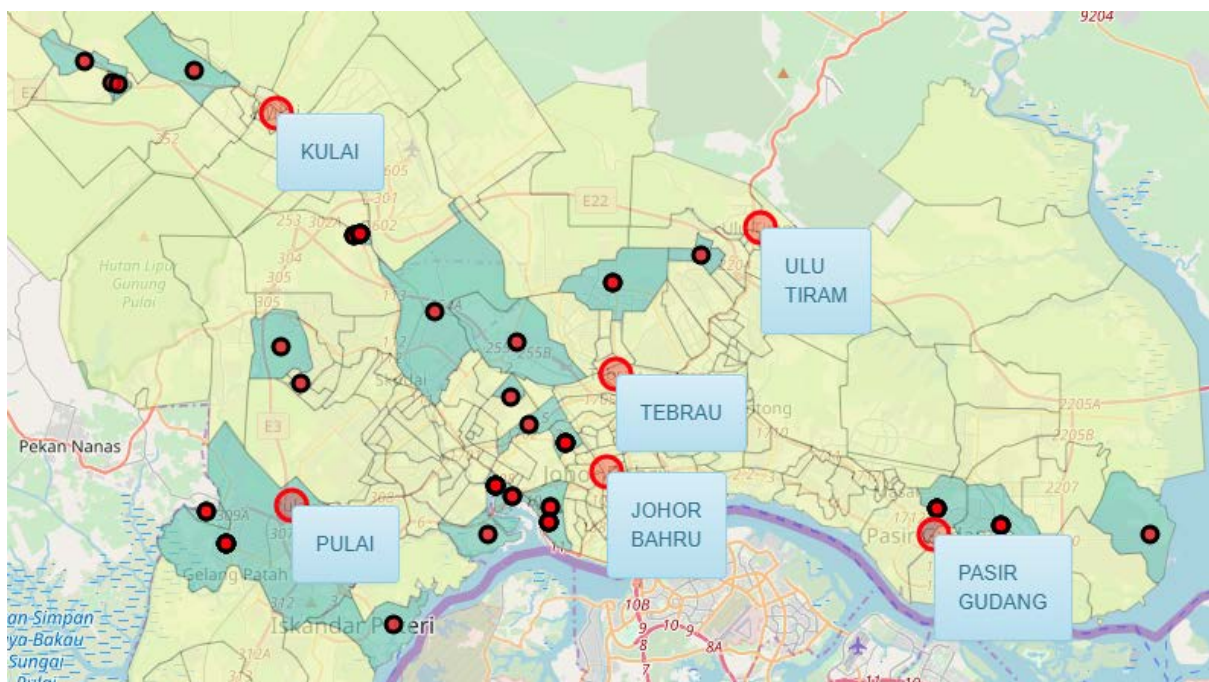


Figure 9 The 'outliers' neighborhoods identified from DBSCAN



Table 4 Information of ‘outliers’ of DBSCAN

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance	Venue_Category
PUTERI WANGSA 1	TEBRAU	1	961.00	Chinese Restaurant
TAMAN JAYA	GELANG PATAH	1	952.00	Chinese Restaurant
TAMBATAN	PULAI	4	867.00	Asian Restaurant, Chinese Restaurant, Chinese ...
BUKIT SERENE	PULAI	2	851.00	Chinese Restaurant, Asian Restaurant
PASIR GUDANG	PASIR GUDANG	4	833.25	Chinese Restaurant, Asian Restaurant, Asian Re...
LEMBAH KEMPAS	PULAI	1	832.00	Asian Restaurant
KAMPONG UNGKU MOHSIN	JOHOR BAHRU	4	815.50	Asian Restaurant, Asian Restaurant, Chinese Re...
PENGKALAN RINTING	PULAI	5	807.00	Asian Restaurant, Asian Restaurant, Chinese Re...
KOTA KULAI	KULAI	1	795.00	Chinese Restaurant
KAMPONG PULAI	GELANG PATAH	2	793.00	Asian Restaurant, Dim Sum Restaurant

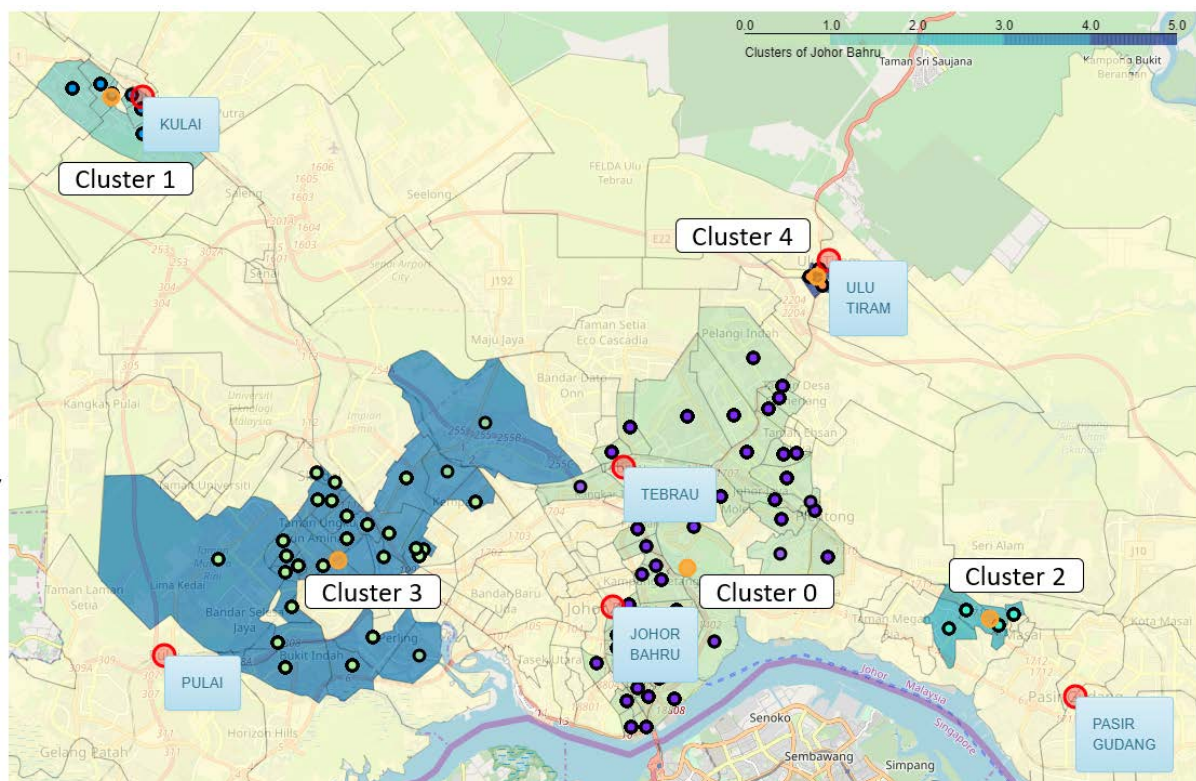


Figure 10 The high Chinese restaurant density neighborhoods clustered from DBSCAN

Figure 10 presents the clusters from DBSCAN based on the density of Chinese restaurants in the region. Each cluster is coloured differently, and the cluster centroids are plotted on the map as orange circle markers. The information of the clusters is tabulated in Tables 5 to 9.

Table 5 Information of Cluster 0

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance
JOHOR JAYA BARAT	PASIR GUDANG	26	661.230769
SRI TEBRAU (1)	JOHOR BAHRU	25	822.840000
TAMAN MAJU JAYA	JOHOR BAHRU	23	614.782609
SRI TEBRAU (2)	JOHOR BAHRU	21	555.809524
JOHOR JAYA TENGAH	PASIR GUDANG	21	709.380952
SRI PELANGI	JOHOR BAHRU	18	588.555556
TAMAN MOLEK	PASIR GUDANG	17	499.352941
PELANGI	JOHOR BAHRU	17	351.764706
TAMAN MELODIES	JOHOR BAHRU	16	624.500000
TAMAN ABAD	JOHOR BAHRU	16	583.187500

Cluster 0 contains the neighborhoods in the Johor Bahru, Tebrau and Pasir Gudang districts. This region is one of the most happening commercial area in the entire Johor Bahru city. The density of Chinese restaurants is high in this region.

Table 6 Information of Cluster 1

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance
TAMAN SELATAN	KULAI	12	728.750000
LENGKONGAN	KULAI	8	890.125000
TAMAN PUTRI	KULAI	7	578.000000
BANDAR KULAI TIMOR	KULAI	6	706.833333
KULAI	KULAI	5	867.000000
BANDAR KULAI TENGAH	KULAI	3	372.000000

Cluster 1 comprises the neighborhoods in a small town of Kulai district. The town area of a district is normally a commercial zone.

Table 7 Information of Cluster 2

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance
RINTING MERANTI	PASIR GUDANG	10	629.000000
BANDAR MASAI UTARA	PASIR GUDANG	7	786.857143
RINTING CENDANA	PASIR GUDANG	5	722.200000
RINTING BALAU	PASIR GUDANG	3	582.666667

Cluster 2 contains the neighborhoods at the west of the Pasir Gudang district. These neighborhoods are a mixed of both residential and commercial zones.

Table 8 Information of Cluster 3

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance
NAKHODA TUTA	GELANG PATAH	15	389.933333
FLAT PERKASA 'A' TUTA	GELANG PATAH	14	666.285714
NUSA INDAH	PULAI	13	700.461538
BUKIT KEMPAS	PULAI	12	691.583333
BELIBIS PERLING	PULAI	11	695.272727
LAKSAMANA TUTA	GELANG PATAH	10	757.500000
HANG JEBAT	GELANG PATAH	10	596.000000
RAWA PERLING	PULAI	10	785.300000
HANG TUAH	GELANG PATAH	9	810.888889
HULUBALANG TUTA	GELANG PATAH	8	923.625000

Cluster 3 consists of the neighborhoods in the Gelang Patah and Pulai districts. These neighborhoods are a mixed of both residential and commercial zones.



Table 9 Information of Cluster 4

Town_Neighborhood	District	Number_of_Chinese_Restaurants	Average_Distance
BANDAR ULU TIRAM SELATAN	TEBRAU	4	238.00
BANDAR ULU TIRAM TENGAH	TEBRAU	4	572.25
BANDAR ULU TIRAM UTARA	TEBRAU	4	583.00

Cluster 4 is the town area of the Tebrau district, the town is named ‘Ulu Tiram’.

The neighborhoods in Clusters 0-4 are potential areas to set up a Chinese restaurant as they are potentially the commercial zones that crowds of Chinese population may pay a visit. However, to acquire customers who are tourists, it is better to set up the business near the Johor Bahru city centre since most of the tourism spots are located around the city centre. Hence, the distances to each of the cluster centroids are plotted in Figure 11.

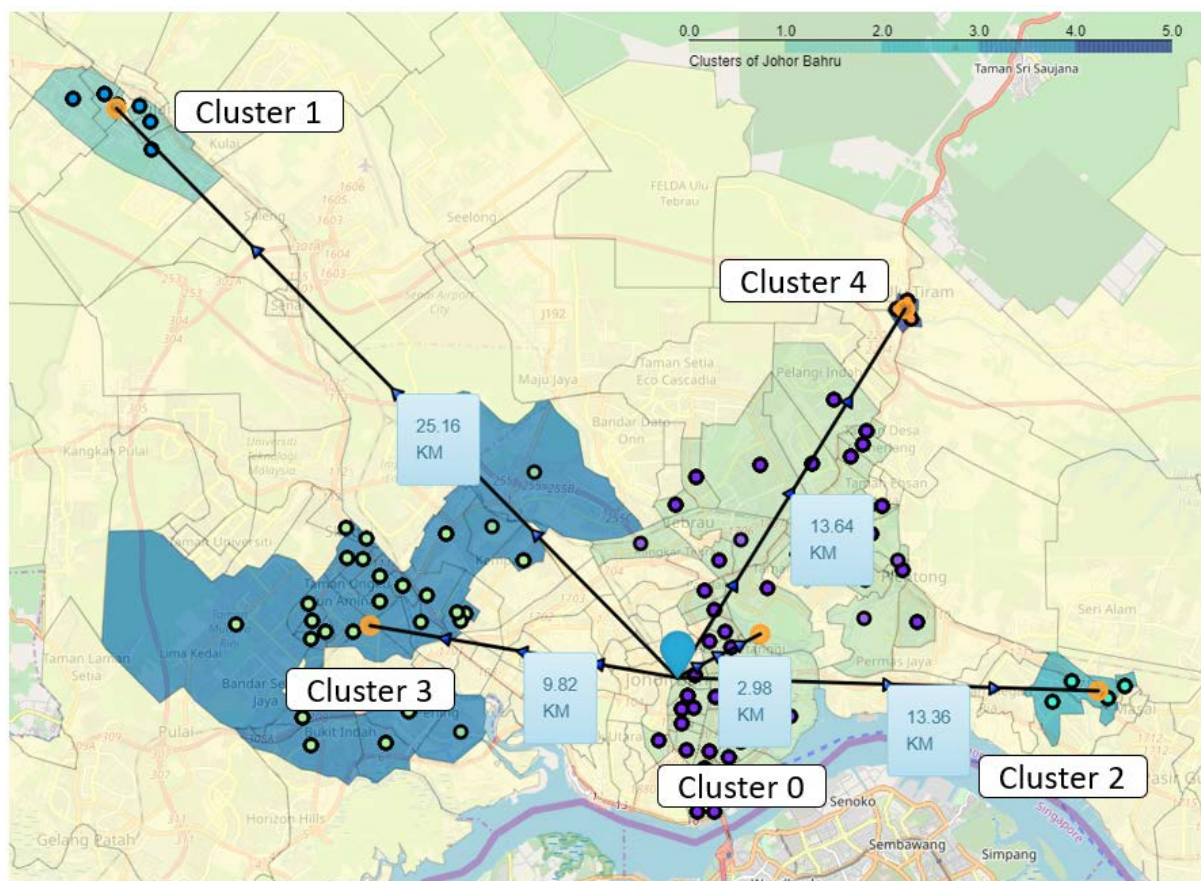


Figure 11 Distances between city centre and centroids of clusters

To this end, it is recommended that the stakeholders should select the neighborhoods in Cluster 0 to set up a Chinese restaurant. An overview of Cluster 0 is displayed in Table 10, where Johor Bahru district has the most nearby Chinese restaurants, due to its vicinity to the city centre and

the Republic of Singapore. The average distance from the neighborhood centres to these restaurants are between 500m to 700m, which is walkable distance.

Table 10 Overview of Cluster 0

	District	Number_of_Chinese_Restaurants	Average_Distance
0	JOHOR BAHRU	248	663.290323
1	PASIR GUDANG	123	643.918699
2	PULAI	33	543.878788
3	TEBRAU	80	649.050000

## Discussion

From the results, several recommendations can be drawn by considering the neighborhoods with residents of more than 2000 only:

- **High-end Chinese restaurants:** For example, Ding Tai Fung [7] and Hai Di Lao [8], they are highly competitive and would most likely prefer to set up their new branch in a crowd of Chinese restaurants in this region (i.e. Number of Chinese restaurants is larger than 10, in a single neighborhood). The neighborhoods in the Johor Bahru district are recommended, as displayed in Table 11 and Figure 12.

Table 11 Recommended neighborhoods for high-end Chinese restaurants

	District	Town_Neighborhood	Number_of_Chinese_Restaurants	Average_Distance	Number_of_Residents
0	PASIR GUDANG	JOHOR JAYA BARAT	26	661.230769	3694
1	JOHOR BAHRU	SRI TEBRAU (1)	25	822.840000	2992
2	JOHOR BAHRU	TAMAN MAJU JAYA	23	614.782609	3060
3	PASIR GUDANG	JOHOR JAYA TENGAH	21	709.380952	2282
4	JOHOR BAHRU	SRI PELANGI	18	588.555556	3579
5	JOHOR BAHRU	PELANGI	17	351.764706	2445
6	PASIR GUDANG	TAMAN MOLEK	17	499.352941	2888
7	JOHOR BAHRU	TAMAN ABAD	16	583.187500	2175
8	JOHOR BAHRU	TAMAN MELODIES	16	624.500000	2449
9	PASIR GUDANG	JOHOR JAYA TIMOR	16	631.437500	3461
10	JOHOR BAHRU	TAMAN KEBUN TEH	15	699.466667	3429
11	TEBRAU	PINANG SAGU	14	675.142857	2454
12	JOHOR BAHRU	SENTOSA	11	763.363636	4886
13	TEBRAU	MOUNT AUSTIN	11	720.272727	2477

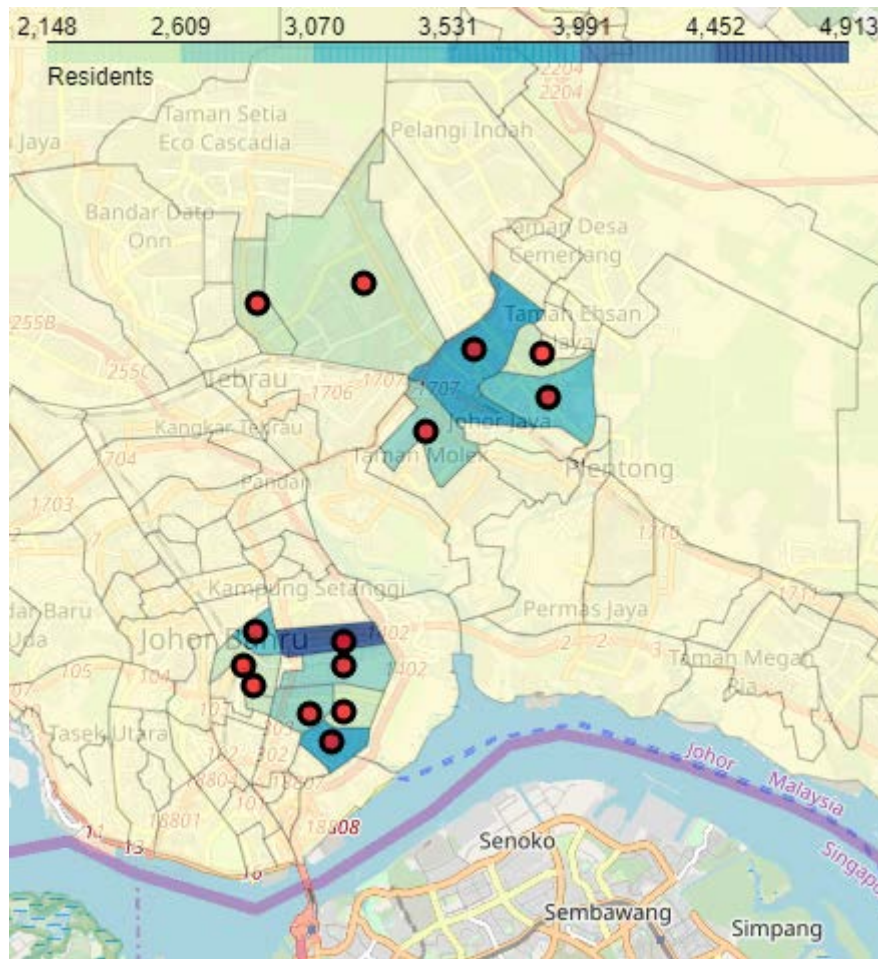


Figure 12 Recommended neighborhoods for high-end Chinese restaurants on map

- **Mid-end Chinese restaurants:** Neighborhoods with moderate competition are recommended, i.e. Number of Chinese restaurants is between 5 to 10, as presented in Table 12 and Figure 13.

Table 12 Recommended neighborhoods for mid-end Chinese restaurants

	District	Town_Neighborhood	Number_of_Chinese_Restaurants	Average_Distance	Number_of_Residents
0	JOHOR BAHRU	STULANG	9	829.222222	3198
1	JOHOR BAHRU	LUMBA KUDA	8	742.000000	3069
2	PASIR GUDANG	KEEMBONG	8	665.875000	3918
3	TEBRAU	DESA JAYA	8	595.000000	2746
4	JOHOR BAHRU	BUKIT SENYUM	6	623.166667	2582
5	JOHOR BAHRU	SETANGGI	6	762.333333	2516
6	JOHOR BAHRU	BUKIT CHAGAR	5	929.800000	2358
7	PASIR GUDANG	PERMAS 2	5	651.000000	2713



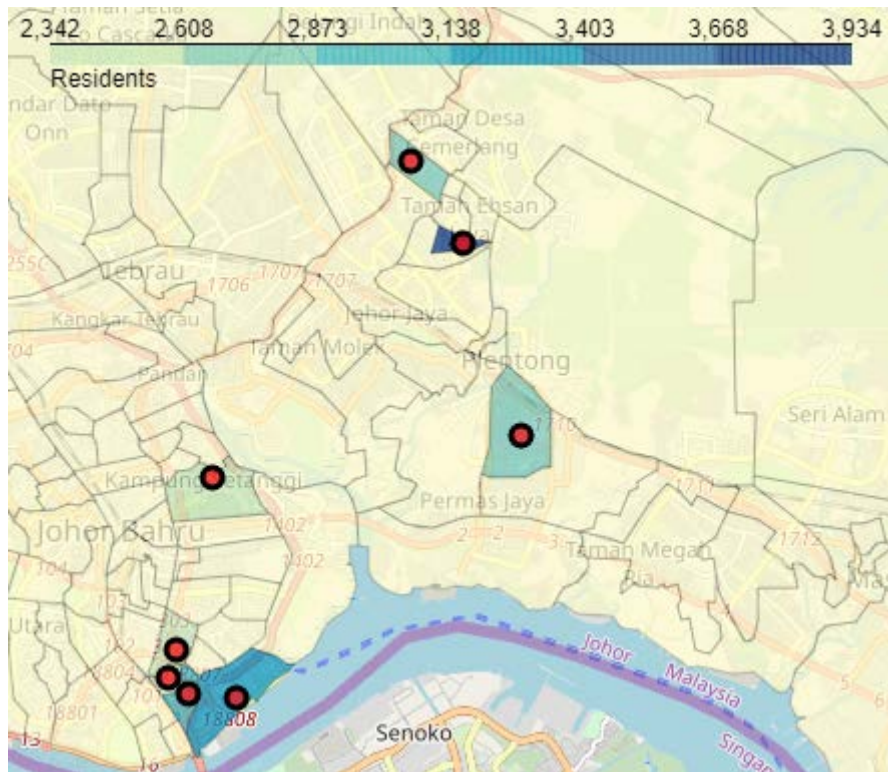


Figure 13 Recommended neighborhoods for mid-end Chinese restaurants on map

- **Family-owned/small-sized Chinese restaurants:** Neighborhoods with low competition are recommended, i.e. Number of Chinese restaurants is lower than 4, as shown in Table 13 and Figure 14.

Table 13 Recommended neighborhoods for family-owned/small-sized Chinese restaurants

	District	Town_Neighborhood	Number_of_Chinese_Restaurants	Average_Distance	Number_of_Residents
0	JOHOR BAHRU	KAMPONG WADI HANA	4	526.750000	2211
1	PASIR GUDANG	BANDAR PLENTONG SELATAN	4	661.750000	4325
2	PASIR GUDANG	TAMAN SAUJANA	4	479.000000	2522
3	TEBRAU	PELANGI GAYA	4	623.750000	2521
4	JOHOR BAHRU	SULAIMAN MENTERI	3	804.666667	2834
5	PASIR GUDANG	BELANTIK	3	854.000000	2507
6	PASIR GUDANG	PANDAN	3	556.333333	3643
7	TEBRAU	PEKAN PANDAN	3	567.333333	2707
8	JOHOR BAHRU	BAKAR BATU	2	785.500000	3827
9	PASIR GUDANG	PERMAS	1	606.000000	4298



Figure 14 Recommended neighborhoods for family-owned/small-sized Chinese restaurants on map

Several factors for consideration:

- **Localized competition:** The average distance can be used as a measure of the adjacency between each of the restaurants in the neighborhood, the lower the average distance, the highest the localized competition.
- **Transport:** The common transport for local residents is vehicle and motor-bike, while for tourists, public transports, taxi or foot are common. Hence, the distance from the city centre is depends on the target audience of the Chinese restaurant.
- **Rental Cost:** The rental price in each neighborhood is different, neighborhoods nearer to the city-centre may have higher rental cost.

- **Regional Development Plan:** The development plan and regulation published by the city council may bring economical effect in certain regions.

## Conclusion

In this report, the greater Johor Bahru region has been explored for the opportunity of opening a Chinese restaurant. The  $k$ -means and DBSCAN machine learning techniques have been applied to find out the potential neighborhoods. With the considerations of the number of Chinese restaurants, the average distance, the number of residents and the distance between city centre and the cluster centroid, a few potential neighborhoods in the Johor Bahru, Pulai, Tebrau and Pasir Gudang district have been selected. For high-end Chinese restaurant, high competition neighborhoods are recommended while for mid-end Chinese restaurant, a few neighborhoods with moderate competition are suggested. More over, neighborhoods with low competition are recommended for family-owned or small-sized Chinese restaurants. Lastly, several factors for further considerations, such as localized competition, transport, rental cost and regional development plan are listed.



## Reference

- [1] D. o. S. Malaysia. Total Population by Ethnic Group, Local Authority Area and State, Malaysia [Online]. Available: [https://web.archive.org/web/20131114184213/http://www.statistics.gov.my/portal/download\\_Population/files/population/04Jadual\\_PBT\\_negeri/PBT\\_Johor.pdf](https://web.archive.org/web/20131114184213/http://www.statistics.gov.my/portal/download_Population/files/population/04Jadual_PBT_negeri/PBT_Johor.pdf)
- [2] I. R. D. Authority. (2016). *Iskandar Malaysia*. Available: <http://iskandarmalaysia.com.my/>
- [3] I. R. D. Authority. (2016). *Our Development Plan*. Available: <http://iskandarmalaysia.com.my/our-development-plan/>
- [4] leowmjw. (2016, 10 February). *Completed migration of Johor Shapefile*. Available: <https://github.com/TindakMalaysia/Johor-Maps/blob/master/2016/MAP/MIGRATED/result/13-Johor-New-DM-4326.geojson>
- [5] F. Inc. (10 February). *Foursquare Developer: Create Magical Real-World Moments for Your User*. Available: <https://developer.foursquare.com/>
- [6] F. Inc. (10 February). *Get Venues Categories*. Available: <https://developer.foursquare.com/docs/api/venues/categories>
- [7] (2020, 12 February). *Ding Tai Fung*. Available: [https://en.wikipedia.org/wiki/Din\\_Tai\\_Fung](https://en.wikipedia.org/wiki/Din_Tai_Fung)
- [8] (2019, 12 February). *Hai Di Lao*. Available: <https://en.wikipedia.org/wiki/Haidilao>