Capstone Project Report –

The Battle of Neighborhoods (week 2)

by Mingkai Liu
2019-12

Table of contents

1. Introduction: Background and Business Problem

1.1 Background

Melbourne is an international city and the capital of Victoria State, Australia. The city is famous for all types of events and cultural festivals, attracting over 10 millions of tourists from Australia and around the world in 2019 (source: https://www.finder.com.au/press-release-jan-2019-10-million-tourists-to-visit-australia-in-2019). In addition, Melbourne has dominated Australia's population growth (http://worldpopulationreview.com/world-cities/melbourne-population/), which is estimated at around 4.967 million by 2020 and will increase to around 5.39 millions by 2025. A large number of visitors and new population bring in increasing demands and opportunities for catering business. Among the international tourists and local population growth, people of Asian ethnics have become a major source, with Chinese visitors and immigrants being the dominant contributor.

1.2 Business problem

This report will try to find potential places for the stakeholders to start a **Chinese restaurant** targeting **mid/high-end customers** in Melbourne. Ideally, the location should be within 20 min driving distance to Melbourne CBD. The location should be chosen in a neighbourhood with a middle to high level of population and an above mid-level income. The location of the new restaurant should be chosen in a neighbourhood with low crime rate and easy accessibility.

2. Data

To find the optimal location for the new Chinese restaurant targeting mid to high-level customers, we need to understand the social-geo-economical distribution of Melbourne suburbs. The selection of the restaurant location requires some comprehensive analysis of the suburbs' data
  https://www.fsrmagazine.com/expert-takes/8-factors-choosing-new-restaurant-location

In this project, we will try to analyse the most important ones:

1. *Visibility*. To analyse the visibility, we will use the data of local population http://data.id.com.au/idforecasts/ as well as the venue information returned from Foursquare. Direct counting of walkers' traffic is difficult, however, this can be estimated indirectly based on the local population and the number of venues, since the higher the population, the more venues, the larger number of people in that neighbourhood, and the more visible the restaurant will be.
2. *Affordability*. Since the target is to locate a restaurant for mid/high-level customers, it is natural to find a location close to suburbs that have a mid to high-level income. The income data can be found at http://house.speakingsame.com/suburbtop.php?sta=vic&cat=Median%20household%20income&name=Weekly%20income&page=1

3. *Parking and Accessibility*. There is no need to explain why parking is important, especially for restaurant targeting mid/high-level customers. The information about parking can be found using the Folium map. Ideally, the restaurant should be within walking distance from a car park or located inside a shopping mall, where free or cheap parking is available.

4. *Crime Rates*. Safety is important, it is essential to make sure the location selected for the restaurant has a reasonably low crime rate. The crime rate data for different suburbs can be found at https://www.crimestatistics.vic.gov.au/explore-crime-by-location

5. *Surrounding Businesses and Competitor Analysis*. The data can be found using Foursquare. For restaurant location, there is a fine balance between 'assemble' and 'isolation'. People like to go to places with a lot of choices of restaurants, however, it is also important to avoid too many similar restaurants crowding at the same place. Therefore it is important to make sure that the location is nearby a 'food' area to be seen by more customers, yet avoid in close proximity to a similar restaurant that can cause direct competition.

3. Methodology

In this project, we will provide a bird-eye view on the geo-social-economical information of Melbourne suburbs, in order to find a potential spot for a Chinese restaurant targeting mid to high-level customers.

First, we will collect and clean the required data for geo-social-economical analysis, including crime rate, income and population distribution.

Second, we will perform basic statistics to understand the distribution of the data. Then we will filter and categorize the data based on income, population and distance to CBD. We will only filter suburbs with above middle levels of income, population, population growth, and located within 15 km distance from the CBD.

Third, we will visualize how different groups of data are distributed and clustered geographically using Folium map. In order to maximize the potential income of the restaurant, we will use machine learning to find a centroid of the clusters of filtered suburbs, taking into account the weight of income and population (weighted k-mean clustering). This centroid will be the starting point for further analysis and searching.

Finally, we will use Foursquare API and Folium map to examine the restaurants around this centroid. Whether there is an assembly of "food street" or "food area" to attract customers? Whether there is any similar Chinese restaurant as competitors? Whether there is a car park area nearby? We will recheck if this suburbs is safe based on the crime rate data. After answering these questions, we will suggest the potential locations to the stakeholders

4. Result and Analysis

4.1 Data acquisition and preprocessing

We use the data of the local population and its prediction from http://data.id.com.au/idforecasts/ .
The income data can be found at
http://house.speakingsame.com/suburbtop.php?sta=vic&cat=Median%20household%20income&name=Weekly%20income&page=1.
The crime rate data for different suburbs can be found
at https://www.crimestatistics.vic.gov.au/explore-crime-by-location. The data are downloaded as .xlsl
files and .csv files from the websites and imported as Pandas Dataframes.

We have removed the column for income ranking and population ranking information, which is not
necessary and could be found out later easily if required. We have to remove spaces at the beginning of
the suburb names and merge the data of income and population based on suburb names.

The crime rate data covers the record of the past 10 years and were shown in one column in a mixed
way. We have extracted the crime rate and rearrange them chronologically in different columns and
calculated the averaged crime rate over the past three years from 2017 – 2019.

Since the task is to find a location, we further use geocoder.arcgis to find the corresponding latitudes
and longitudes of each suburbs and Local Government Area, and calculate their distances from the
Melbourne CBD using geopy.distance.vincenty, which is an important factor to consider for the selection
of a restaurant.

For the income and population data, the location information is provided for different *Suburbs*, while
the crime rate information is provided for different *Local Government Areas*. Since the crime rate only
reflects the averaged level of a particular *Local Government Areas,* which includes many different
suburbs, we decided to use this information as a reference at a later stage, after we select the potential
locations and search the corresponding Local Government Areas for the Suburbs, using data from
https://www.greaterdandenong.com/document/32220/statistics-suburbs-postcodes-electorates-within-each-municipality

4.2 Exploratory data analysis

We start our analysis by performing exploratory statistical analysis to the data. This operation helps us
to identify the statistical distribution of the income, population, population growth and crime rate.
Their results are summarized in Table 1 and Table 2, as well as Figure 1.

The income level among different suburbs has a medium level around AUD 1348, while the 25% and 75%
levels mark around AUD 1172 and AUD 1512. The overall distribution follows approximately a normal
distribution (see Figure 1). However, we also found that there are two additional small peaks around
AUD 900 and AUD 1900, indicating the groups of 'poor' and 'rich'.

In contrary, the population and population growth have distributions that are highly skewed, which means there are a few suburbs that have a significantly larger population and population growth.

The crime rates are calculated based on the number of crime incidents in the Local Government Area and have been normalized to the local population. The number showed represents crime incidents per 100,000 population. We find that the crime rate also follows approximately a symmetric normal distribution, with a medium level around 5,000. But there are two outliers which are beyond 10,000.

The medium level and 75% level of the income, population and population growth will be used as the thresholds for characterizing the mid and high level of income, population and population growth in the following analysis.

Table 1.　Statistics of Weekly Income and Population

	Weekly Income	2011	2021	2031	Population Growth (2011-2031)
count	380.000000	380.000000	380.000000	380.000000	380.000000
mean	1352.289474	10458.905263	11930.815789	13165.547368	2706.642105
std	307.826949	9006.705267	10674.965906	12280.329474	5729.019623
min	633.000000	109.000000	233.000000	264.000000	-2081.000000
25%	1172.000000	3748.000000	3991.000000	4270.250000	237.500000
50%	1348.500000	8007.500000	8662.000000	9277.500000	1069.500000
75%	1512.000000	14437.500000	16511.000000	18291.250000	2660.500000
max	2352.000000	49796.000000	54498.000000	74932.000000	53366.000000

Table 2.　Statistics of averaged crime rate over 2017-2019

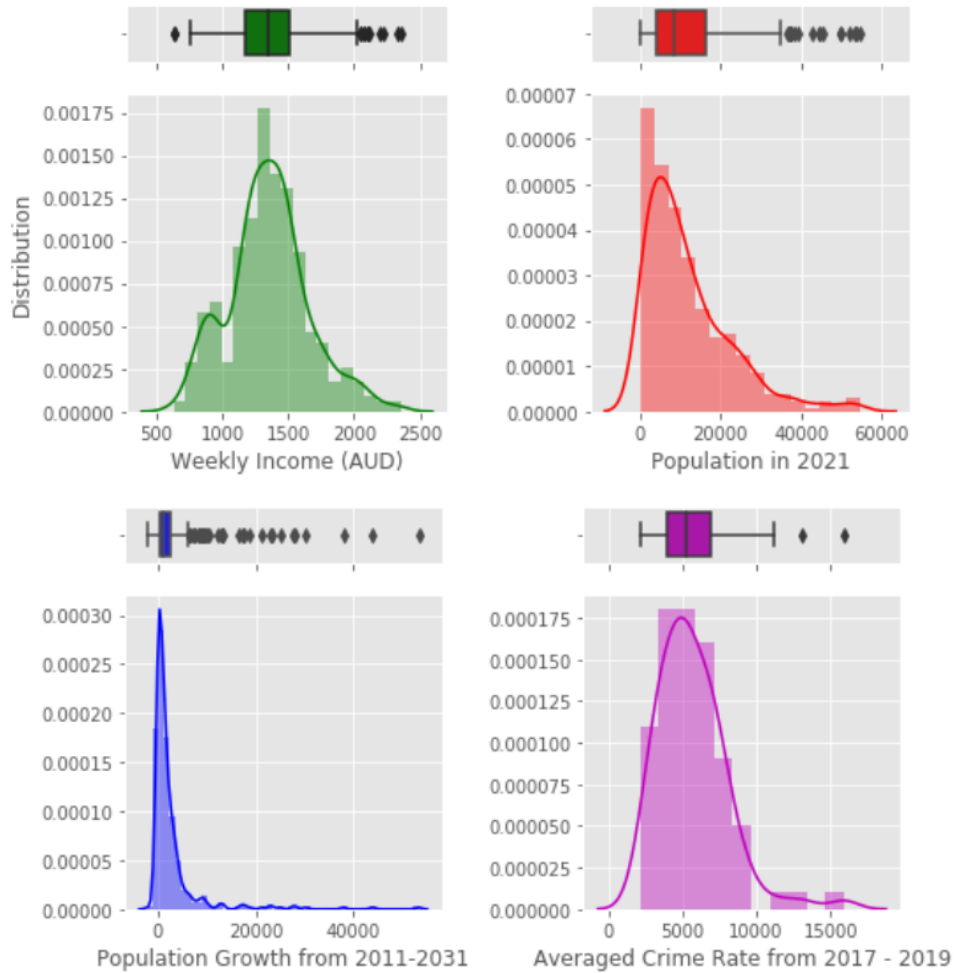|	Mean rate over the past 3 years
count	80.000000
mean	5699.725000
std	2362.270893
min	2160.000000
25%	4026.500000
50%	5262.000000
75%	6940.750000
max	15891.000000

Figure 1. The statistical distribution of Weekly Income (AUD), Population (people), Population Growth (people) and Averaged Crime Rate (per 100,000 population) over 2017-2019.

4.3 Geographic distribution of the data

To understand the geographic distribution of the income, population and crime rate, we have used geocoder.arcgis to extract the latitudes and longitudes of the suburbs and local Government areas. After that, we use Folium.map to visualize the geographic distribution.

The geographic distribution of income and population is shown in Figure 2. The size of the red circles represents the size of the population while the opacity represents the income level. The larger the circle, the more population; the redder the circle filling, the higher the income. To exaggerate the contrast, we have rescaled the income and population to the scale of (0,1), with the minimal value and maximal value rescaled to 0 and 1, respectively.

We can see from Figure 2 that apart from the large population in Melbourne central, more population and higher income are clustered in the eastern suburbs of Melbourne. This observation is consistent with the situation in Melbourne. This is important information to be taken into account when choosing locations for new restaurants.

Figure 3 shows the geographic distribution of crime rate. The size of the circle is proportional to the crime rate. It is clear that Melbourne Central and Yarra, the one northeast to Melbourne Central are the

two 'hot spots' of high crime rates. In fact, these two Local Government Areas are the two outliers we found in Figure 1, with the crime rate over 10,000 per 100,000 population. We shall keep this information in mind when we choose location of the new restaurant if we are more inclined to find a safer suburb to start the business.

Figure 2. The geographic distribution of income and population. The size of the red circles represents the size of the population while the opacity represents the income level. The larger the circle, the more population; the redder the circle filling, the higher the income. The bottom one is the zoom-in view.
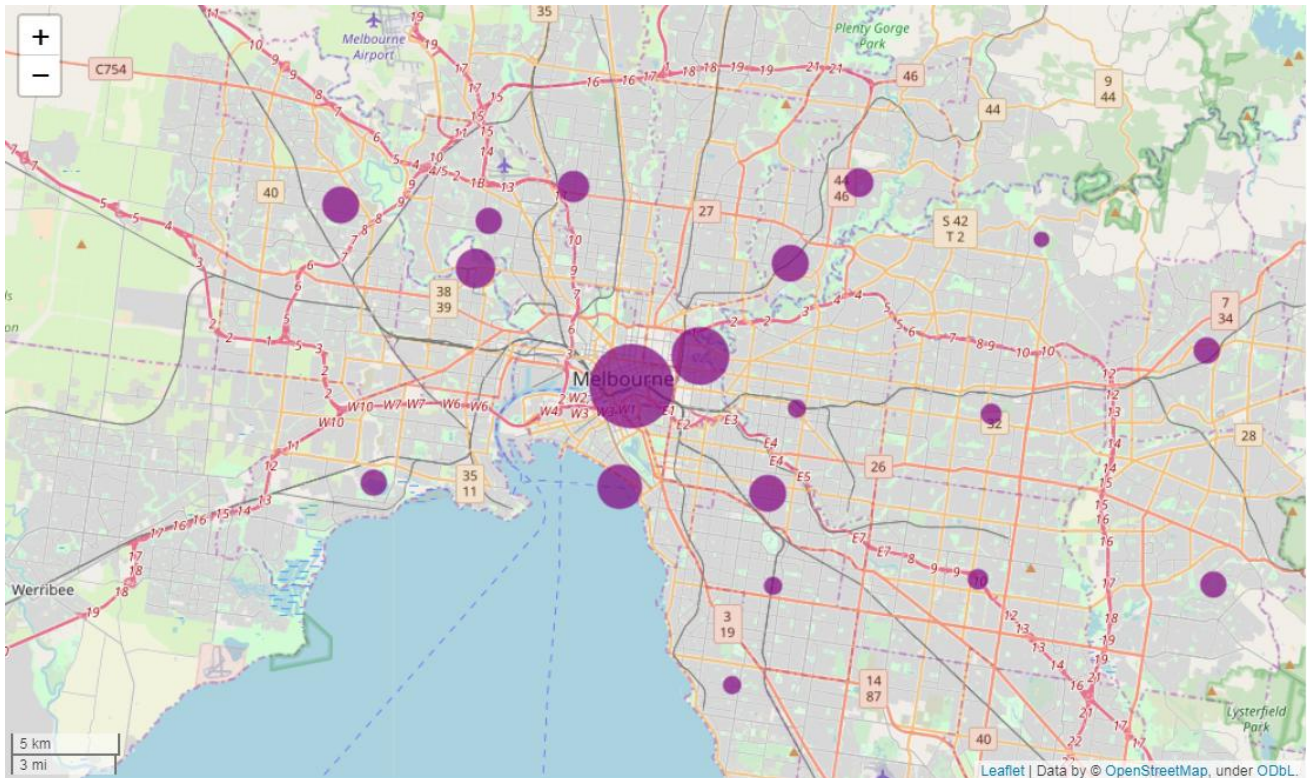
Figure 3. The geographic distribution of crime rate per 100,000 population. The size of the purple circles is proportional to the crime rate

4.4 Filter suburbs based on income, population and distance

In order to target more potential customers with above the average income, we further filter the suburbs into 3 groups based on the statistic information.

★ Group 1: large population (>75%, i.e. population > 16511), high population growth (>75%, i.e. population growth > 2660) and high average income (>75%, i.e. income > 1512)
★ Group 2: large population (>75%, i.e. population > 16511), high population growth (>75%, i.e. population growth > 2660) and medium average income (>50%, i.e. income > 1348)
★ Group 3: medium population (>50%, i.e. population > 8660), medium population growth (>50%, i.e. population growth > 1069) and high average income (>75%, i.e. income > 1512)

The result is shown in Figure 4. The red, blue and green circles represent suburbs that are filtered and based on the condition of group 1, group 2 and group 3. The purple circle with yellow filling represents the Melbourne CBD. We find the following distribution patterns:

★ For Group 1 (red circles), the suburbs are mostly on the east side of Melbourne, close to CBD.
★ For Group 2 (blue circles), the suburbs are mostly on the west side of Melbourne CBD, quite far away from the CBD.
★ For Group 3 (green circles), the suburbs are mostly on the east, northeast and southeast sides of Melbourne CBD.
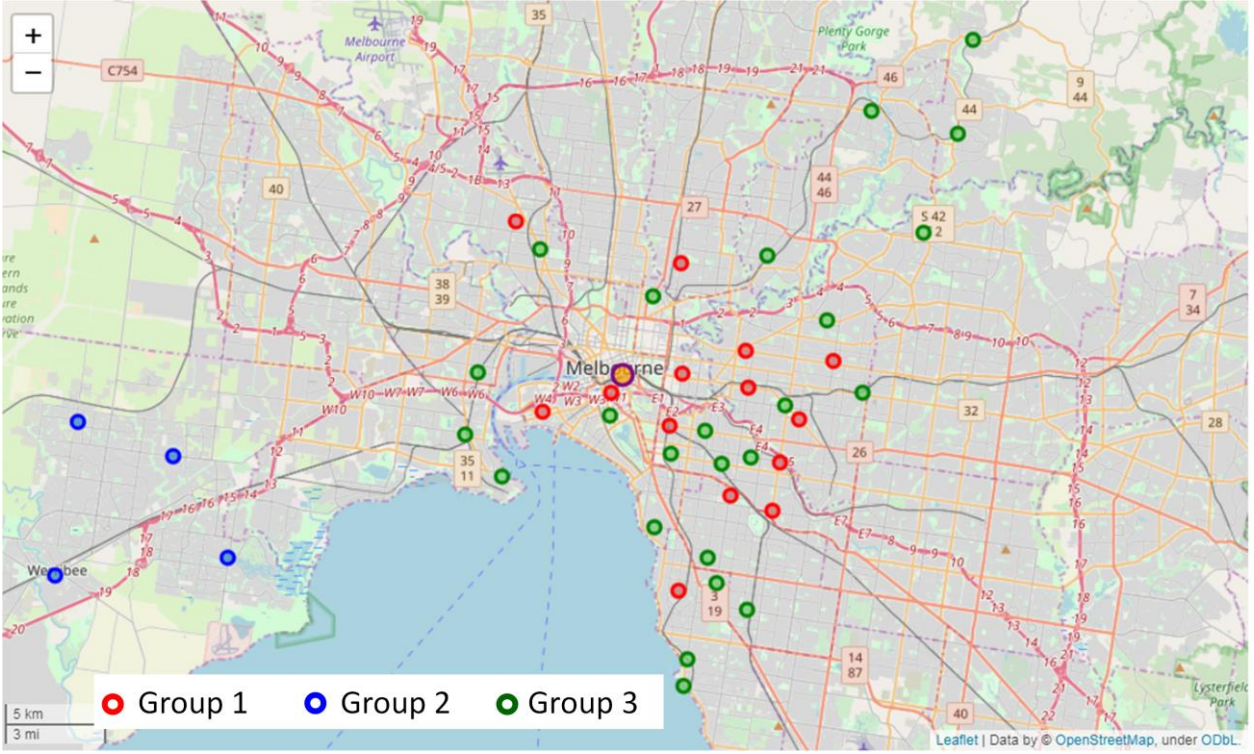
Figure 4. Filtered suburbs that satisfy the conditions of Group 1, Group 2 and Group 3.

As a restaurant, we want to increase our visibility, therefore, we want to find a location is not far away from the centre such that it can attract tourists as well as local residents. We will only include suburbs that are within 15 km radius from the CBD, which is less than 30 min driving. We calculate the distance of the suburbs to the Melbourne CBD based on their latitudes and longitudes using geopy.distance.vincenty, and filter out those outside the 15 km circle.

From the viewpoint of maximizing the potential income, it will be particularly attractive to find a location that can be easily assessed by residents from these filtered suburbs. To do that, we use machine learning to find the 'economic centroid' of the suburb cluster and take into account the weight of population and weekly income. The weight of suburb j in the cluster can be defined as:

$$w_j = I_j P_j$$

where I_j and P_j are the income and population. Then we use the weighted K-mean algorithm to find the 'economic centroid' of the cluster. The mean of the cluster can be calculated as

$$< R >= \frac{1}{\sum_j w_j} \sum_j w_j R_j ,$$

where R_j is the distance from the suburb j to the centroid.

Taking the weight of income and population into account, we find the 'economic centroid', as indicated by the cyan circle. It locates in the south-east of the Melbourne central, around 5 km from the CBD,
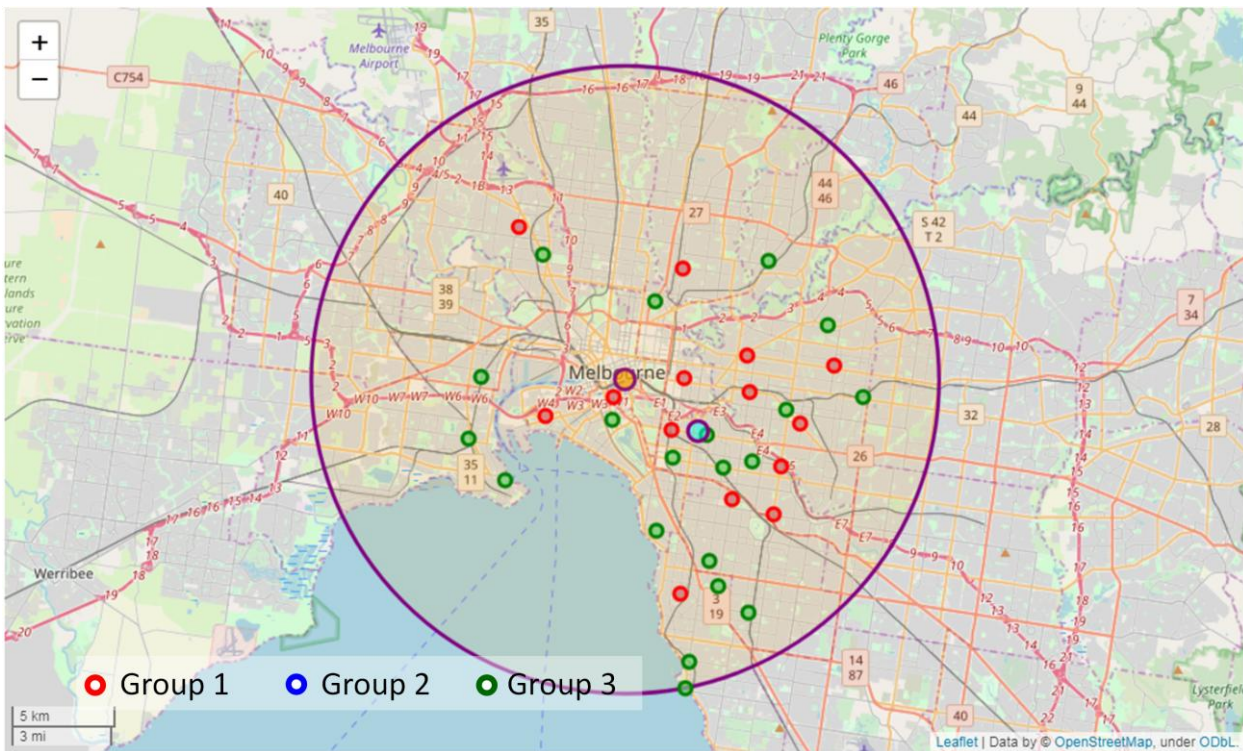
Figure 5. Filtered suburbs that are within 15 km from the CBD. The cyan circle represents the 'economic centroid' of the filtered suburb cluster.

4.5 Restaurants near the economic centroid

The economic centroid will be used as a starting point for further analysis. We search the venues that are within 2km radius of the centroid using Foursquare API. Since we are interested in restaurants that serve a menu of cuisines, we further filter the result and only look for categorical information that has "restaurant" in it, which excluded other types including café, fast foods, steak houses, etc. The results are shown in Table 3.

We found 28 restaurants, where Greek restaurants (4) are the most common, followed by Italian restaurants (3), French restaurants (2), Japanese restaurants (2), Mexican Restaurants (2), Vietnamese Restaurant (2). There is one Chinese (Szechuan) restaurant in the area: Dainty Sichuan Restaurant, which is our most direct competitors.

We visualize the locations of the restaurants using Folium map, as shown by the squares in Figure 6. Quite surprisingly, all the restaurants are located on the west side of the economic centroid, mainly on Toorak Road and Chapel Street. This indicates that these two streets are major business streets developed in the suburb.

To avoid direct competition with the existing Chinese restaurant (indicated by the red square in Figure 6), we shall choose locations that are at least 250 m away (indicated by the red circle in Figure 6).

Table 3. Restaurants within 2km distance from the economic centroid. The red rectangular marks the Chinese restaurant *Dainty Sichuan Restaurant*

|	name	categories	lat	lng	postalCode	address
0	Hawker Hall	Asian Restaurant	-37.854527	144.993067	3181	98 Chapel St
1	The Smith	Australian Restaurant	-37.852035	144.995138	3181	213 High St
2	Bistro Thierry	French Restaurant	-37.848141	145.003454	3142	511 Malvern Rd
3	Chez Olivier	French Restaurant	-37.849200	144.990889	3181	121 Greville St
4	Jimmy Grants	Greek Restaurant	-37.825965	144.997876	NaN	427 Church Street
5	Bahari	Greek Restaurant	-37.825490	144.997590	3121	179 Swan Street
6	Lambs on Spit	Greek Restaurant	-37.847021	144.994518	NaN	245 Malvern Rd
7	Lemnos Tavern	Greek Restaurant	-37.853237	145.003340	NaN	445 High St
8	Two Forty Eight	Halal Restaurant	-37.839620	144.995220	3141	248 Toorak Rd
9	Cucinetta	Italian Restaurant	-37.838380	144.990080	3141	4/3 Murphy St
10	La Lucciola	Italian Restaurant	-37.843143	144.994969	NaN	478 Chapel St
11	Da Noi	Italian Restaurant	-37.838694	144.989701	3141	95 Toorak Rd
12	Mr. Miyagi	Japanese Restaurant	-37.854229	144.992694	3181	99 Chapel St
13	Ichi Ichi Ku Izakaya	Japanese Restaurant	-37.834413	144.982481	3141	119 Park St
14	Subi Q	Korean Restaurant	-37.847520	145.000240	3141	407 Malvern rd
15	Fonda Mexican	Mexican Restaurant	-37.825782	144.998363	3121	248 Swan St
16	Fonda Mexican	Mexican Restaurant	-37.853157	144.993141	3181	144 Chapel St
17	Persian Room 248	Middle Eastern Restaurant	-37.839459	144.995309	NaN	South Yarra
18	Woodland House	Modern European Restaurant	-37.853750	145.002530	3181	78 Williams Rd
19	Borsch, Vodka & Tears	Polish Restaurant	-37.852154	144.993322	3181	173 Chapel St
20	abacus.	Restaurant	-37.845723	144.994343	NaN	383 Chapel Street
21	Feast of Merit	Restaurant	-37.825326	144.995048	3121	117 Swan Street
22	Hunky Dory	Seafood Restaurant	-37.837395	144.995963	3141	670 Chapel St.
23	Dainty Sichuan Restaurant	Szechuan Restaurant	-37.839282	144.992979	3141	176 Toorak Rd
24	Colonel Tan's	Thai Restaurant	-37.850482	144.993022	3181	229 Chapel St
25	Veg Out Time	Vegetarian / Vegan Restaurant	-37.846323	144.994222	3141	366 Chapel St.
26	Misschu	Vietnamese Restaurant	-37.839760	144.996440	NaN	276 Toorak Road
27	Master Roll Vietnam	Vietnamese Restaurant	-37.839016	144.992614	3141	169 Toorak Rd
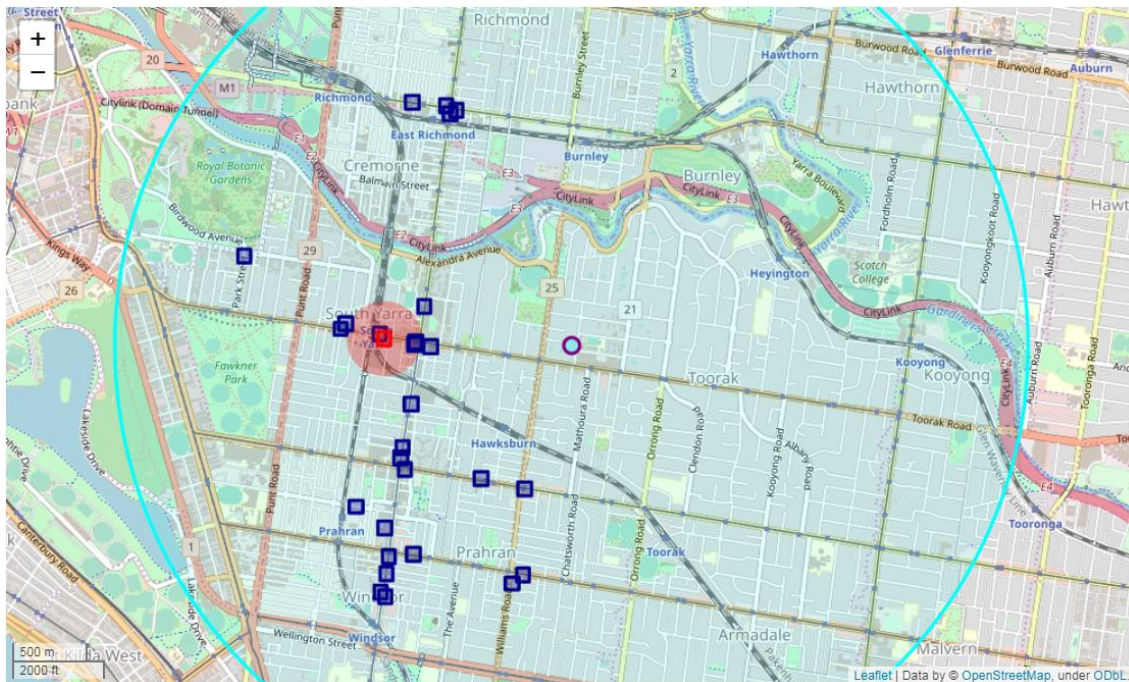
Figure 6. Distribution of restaurants (indicated by squares) within the 2-km distance from the economic centroid (cyan circle). The red square indicates the location of the Chinese restaurant *Dainty Sichuan Restaurant*, while the red circle indicates the 250 m range from the restaurant.

4.6 Safety of the suburb

Before we look further into other detailed information, we check the crime rate of the suburb and see if this part of the city is reasonably safe.

The suburb we consider is south Yarra, which belongs to the Local Government Area of *STONNINGTON*. We check the crime rate information and filter the Local Government Areas that are within 10 km distance from the CBD, and the results are shown in Table 4. The mean crime rate of *STONNINGTON* over the past three years is 7211 per 100,000 population, ranked 5th of the nine Local Government Areas that are within 10 km distance.

Table 4. The crime rate of Local Government Areas that are within 10 km distance from the Melbourne CBD

	Local Government Area	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mean rate over the past 3 years	Latitude	Longitude	Distance to CBD
0	MELBOURNE	26613.6	24126.1	22588.3	20718.4	18581	18033.6	18691.8	16724.2	16129	14820.7	15891	-37.81739	144.96751	0.000000
1	YARRA	11737.9	11068	11411.6	10640.5	11528.4	10881.5	11982.5	11154.9	11513.3	10905.6	11191	-37.80491	145.00338	3.448998
2	PORT PHILLIP	8923.5	8501.6	8439	8954	8500.6	8935.8	9532.6	8964.7	8778.3	8364.7	8702	-37.85874	144.96150	4.619967
3	MARIBYRNONG	8684.6	9314.2	9739.7	8775.2	8155.2	7832.9	8427.5	7310.7	7632.4	7149.3	7364	-37.76920	144.88635	8.928000
4	STONNINGTON	7979.5	6880.9	7185.7	7124.6	6791.7	6989.9	8090.8	7101.9	7091.7	7442.3	7211	-37.86130	145.03776	7.873430
5	DAREBIN	6251.9	6537.6	6834.5	6269.4	7598.1	7259.1	8371.9	7525	7120.9	6968.4	7204	-37.76667	145.05000	9.191426
6	MORELAND	5269.2	5686.3	6165	5671.3	6153.6	6447.8	7057.7	6107.7	6348.6	5805.6	6087	-37.73557	144.93745	9.459637
7	MOONEE VALLEY	5247.7	5341.4	5983.8	5489.2	5443.1	5645.8	6230.9	5493.1	5351.8	4885.4	5243	-37.74941	144.89325	9.986149
8	BOROONDARA	3270.7	3128.8	3444.3	3509.4	3407.3	3233.1	3822.3	3482.9	3058.7	3208.4	3250	-37.82640	145.05299	7.592192

4.7 Parkings in the surroundings

We prefer potential locations that have parking space around. From Folium map, we find parking spots along the two sides of Toorak Road and Chapel Street. These parking spots are either located within shopping centres or beside large markets. Ideally, we would prefer to find the location for the new restaurant that has access to a parking spot within 250 m distance, which is around 5 minutes of walking.

5. Results and Discussion

Combining all the above analysis and the nearby parking and shopping malls from Folium map, we provide 3 potential candidate locations for the stakeholders to examine and decide. The selection of the three candidates is based on three criteria: parking spot nearby, have a large number of customers, no direct competitors within 250 m. The result is visualized in Figure 7.

★ Candidate A: locations around Como Center, Toorak Rd, South Yarra (as indicated by the magenta square).
   Como Center is a landmark retail space of South Yarra (https://www.comocentre.com.au), showcasing an extensive mix of local and international designer stores. The large number of customers, visitors and close by car parks make it an attractive candidate location for the new Chinese Restaurant. It is also sufficiently far away (>250 m) from the existing Chinese restaurant (as indicated by the red squares).

★ Candidate B: locations around Jam Factory, Chapel St, South Yarra (as indicated by the purple square)
   Jam Factory is a Mall in an old jam production facility featuring a multiscreen movie theatre, eateries & shops (https://thejamfactory.com.au). The large number of movie viewers and close by car parks make it another attractive candidate location for the new Chinese Restaurant.

★ Candidate C: locations around Pran Central Shopping Centre, Chapel St, South Yarra (as indicated by the orange square)
   Pran Central is a heritage-listed seven-storey Edwardian baroque architectural style former department store (https://www.prancentral.com.au), built in 1915 as Read's Store, comprising a shopping centre with apartments above. There are large supermarkets nearby. The large number of customers and close by car parks make it another attractive candidate location for the new Chinese Restaurant.
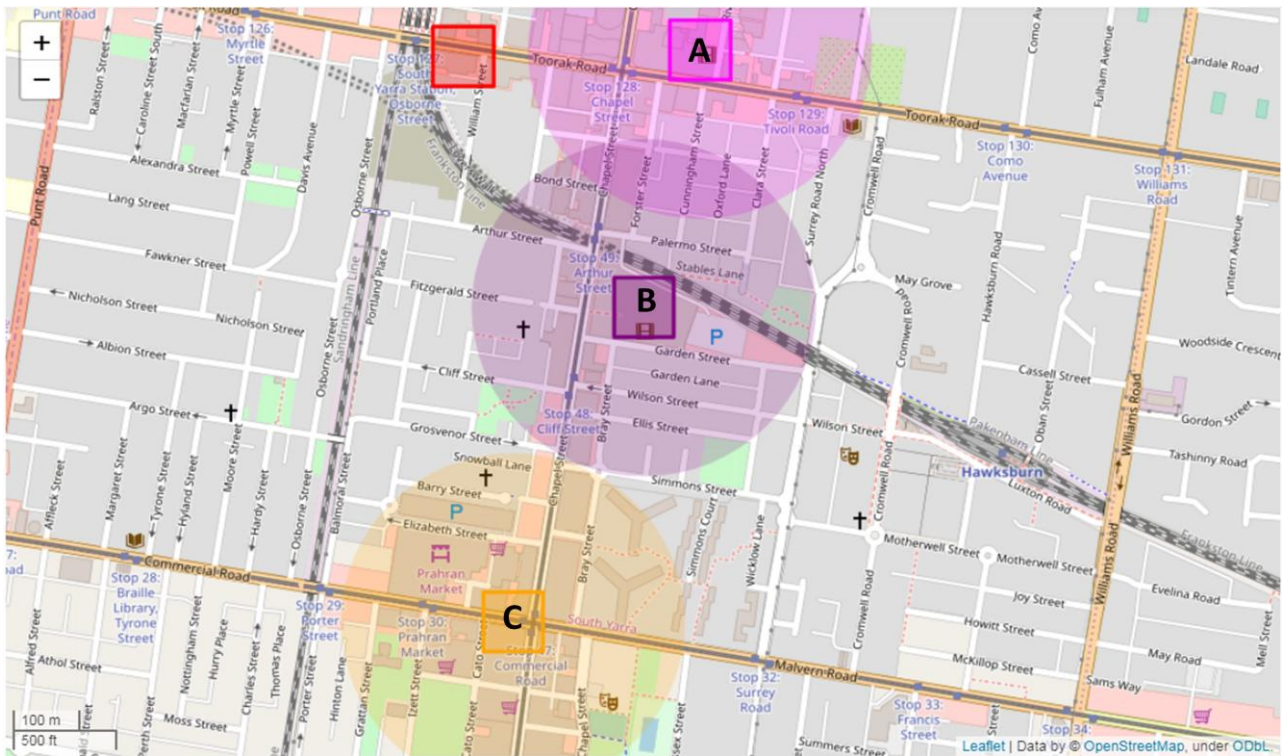
Figure 8. Distribution of the three recommended locations for the new restaurants. The circles indicate the 250 m range. Blue 'P' are parking spots.

6. Conclusion

To conclude, this report provided an analysis of the potential locations for a new Chinese restaurant in Melbourne, targeting mid to high-level customers. We have analyzed the geo-social-economical data of Melbourne suburbs and identified suburbs that have above the medium population and income, with reasonable crime rate. We employed weighted k-mean algorithm to identify the economical centroid of the suburb cluster and analyze the business and competitors around the centroid using data from Foursquare. After taking into account factors such as easy access to parking, large flow of customers and crime rate, we finally provided three potential locations for the stakeholders to further explore and decide the final optimal location based on other factors including community environment, ethnic and cultural, rent and size of location, etc.