

Which Planning Area to Live in the City of Singapore for a Young Couple

1. Introduction

Singapore, is a sovereign city-state and island country located in maritime Southeast Asia. As a high-income economy and developed country, Singapore keeps attracting a large number of young immigrants every year.¹

There are different districts with different characteristics, new immigrants need to consider the shops, tenancy ratio, mode of transports to work and many other factors of a district, where to buy or rent a flat, especially for young couples, is a big interest for the immigrants. This project will have a study based on the statistical data to cluster and analyze different districts of Singapore.

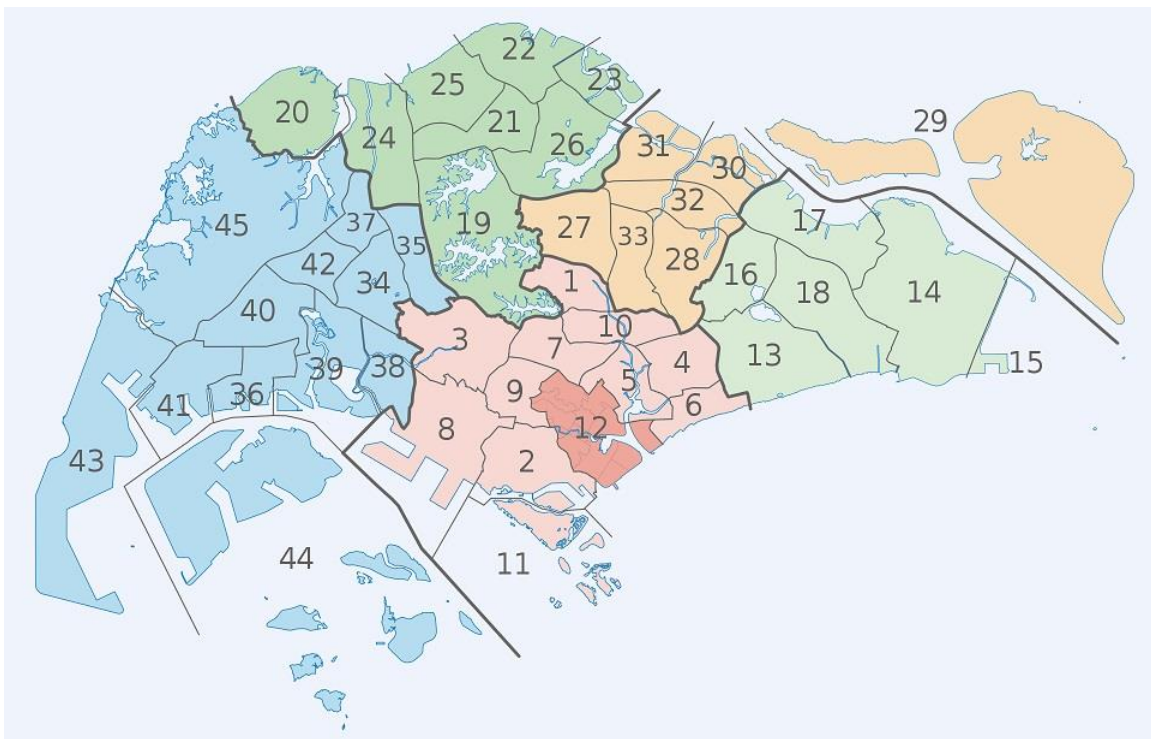


Figure 1: Maps of Planning Areas in Singapore

¹ <https://en.wikipedia.org/wiki/Singapore>

2. Data acquisition and cleaning

2.1 Data sources

OneMap is the authoritative national map of Singapore with the most detailed and timely updated information developed by the Singapore Land Authority. Planning Areas are the main urban planning and census divisions of Singapore delineated by the Urban Redevelopment Authority. First, I used **OneMap API** to get the multi-polygon boundaries of all planning areas of Singapore². Then, I used OneMap API to get the population statistical data of each planning area³. After that, I used Foursquare API to get the venues within each planning area⁴.

2.2 Data cleaning

The GeoJSON format returned by OneMap API Planning Area Polygons are not compatible with Foursquare API or Folium Map. I did some string modification and DataFrame group partition to extract the coordinates of multi-polygon and divide them into several polygons, later with some rounding of the polygons so that the format can be accepted by Foursquare API and request URL. Then I swap the latitudes and longitudes of the coordinates so that compatible with Folium format.

Among the planning area data, I also moved out the 'Others' category and Null value.

3. Methodology

For population statistical data of **OneMap API**, I picked average income, average age, tenancy ratio, household dwelling type, education status, marital status and mode of transports to school and work.

I utilized the Foursquare API to explore the venues inside each planning areas. I adopted the method of searching by polygon area. Here is a head of the list Venues name, category and geographic coordinates.

	Planning Area		Venue	Latitude	Longitude	Venue Category
0	ANG MO KIO	Bishan - Ang Mo Kio Park		1.362219	103.846250	Park
1	ANG MO KIO	Aramsa ~ The Garden Spa		1.362292	103.847602	Spa
2	ANG MO KIO	Lower Peirce Reservoir Park		1.370299	103.826565	Park
3	ANG MO KIO	Pond Gardens		1.366381	103.834190	Park
4	ANG MO KIO	Central Delights		1.377691	103.856590	College Cafeteria

Figure 2: Venues in a Planning Area

² <https://docs.onemap.sg/#planning-area>

³ <https://docs.onemap.sg/#population-query>

⁴ <https://developer.foursquare.com/docs/api/venues/search>

Then group the data by planning area and get the mean number of each type of venue.

Planning Area	Ice Cream Shop	Indian Restaurant	Indonesian Restaurant	Indoor Play Area	Intersection	Island	Italian Restaurant	Japanese Curry Restaurant	Japanese Restaurant
ANG MO KIO	0.020000	0.03	0.00	0.0	0.0	0.000000	0.010000	0.0	0.030000
BEDOK	0.040000	0.04	0.00	0.0	0.0	0.000000	0.020000	0.0	0.010000
BISHAN	0.060000	0.01	0.00	0.0	0.0	0.000000	0.010000	0.0	0.020000
BOON LAY	0.000000	0.00	0.00	0.0	0.0	0.000000	0.000000	0.0	0.000000
BUKIT BATOK	0.020000	0.02	0.00	0.0	0.0	0.000000	0.030000	0.0	0.000000
BUKIT MERAH	0.027523	0.00	0.00	0.0	0.0	0.009174	0.009174	0.0	0.045872
BUKIT PANJANG	0.010000	0.03	0.01	0.0	0.0	0.000000	0.020000	0.0	0.010000
BUKIT TIMAH	0.040000	0.05	0.00	0.0	0.0	0.000000	0.060000	0.0	0.040000
CENTRAL WATER CATCHMENT	0.020000	0.03	0.00	0.0	0.0	0.000000	0.010000	0.0	0.010000
CHANGI	0.010000	0.03	0.00	0.0	0.0	0.000000	0.000000	0.0	0.010000

Figure 3: Density of kinds of venues in Planning Areas

In summary of this graph, 256 unique categories were returned by Foursquare, then I created a table which shows list of top 7 venue category for each planning area in below table.

Planning Area	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
ANG MO KIO	Food Court	Chinese Restaurant	Coffee Shop	Park	Fast Food Restaurant	Indian Restaurant	Japanese Restaurant
BEDOK	Chinese Restaurant	Coffee Shop	Food Court	Dessert Shop	Asian Restaurant	Beach	Seafood Restaurant
BISHAN	Chinese Restaurant	Coffee Shop	Food Court	Ice Cream Shop	Café	Thai Restaurant	Park
BOON LAY	Exhibit	Boat or Ferry	Bus Station	Bus Stop	Fishing Spot	Scenic Lookout	Coffee Shop
BUKIT BATOK	Food Court	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Supermarket	Café	Gas Station

Figure 4: Top 7 most common venues in Planning Areas

We have some common venue categories in planning areas. In this reason I used unsupervised learning K-means algorithm to cluster the planning areas. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, I will run K-Means to cluster the planning areas into 6 clusters.

```
kclusters = 6
# drop the population data because there are too much NaN
singapore_grouped_clustering = singapore_grouped.drop(['Planning Area', 'Age_Per', 'Income_Per', '
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(singapore_grouped_clustering)
kmeans.labels_

array([1, 1, 1, 5, 4, 0, 4, 1, 0, 0, 2, 4, 1, 0, 1, 1, 1, 4, 1, 0, 0, 0,
       0, 1, 0, 0, 0, 1, 0, 0, 4, 5, 4, 4, 1, 0, 0, 0, 5, 4, 1, 5, 0, 0,
       3, 5, 1, 0, 5, 1, 5, 3, 4, 4, 1], dtype=int32)
```

Figure 5: Apply K-Means to cluster Planning Areas

Here is my merged table with cluster labels for each planning area.

Cluster Labels	Planning Area	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
1	ANG MO KIO	Food Court	Chinese Restaurant	Coffee Shop	Park	Fast Food Restaurant	Indian Restaurant	Japanese Restaurant
1	BEDOK	Chinese Restaurant	Coffee Shop	Food Court	Dessert Shop	Asian Restaurant	Beach	Seafood Restaurant
1	BISHAN	Chinese Restaurant	Coffee Shop	Food Court	Ice Cream Shop	Café	Thai Restaurant	Park
5	BOON LAY	Exhibit	Boat or Ferry	Bus Station	Bus Stop	Fishing Spot	Scenic Lookout	Coffee Shop
4	BUKIT BATOK	Food Court	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Supermarket	Café	Gas Station

Figure 6: Clusters of Planning Areas

4. Results and Conclusion

This is the visualization result of cluster result by folium map. There are 6 clusters of Singapore planning areas.

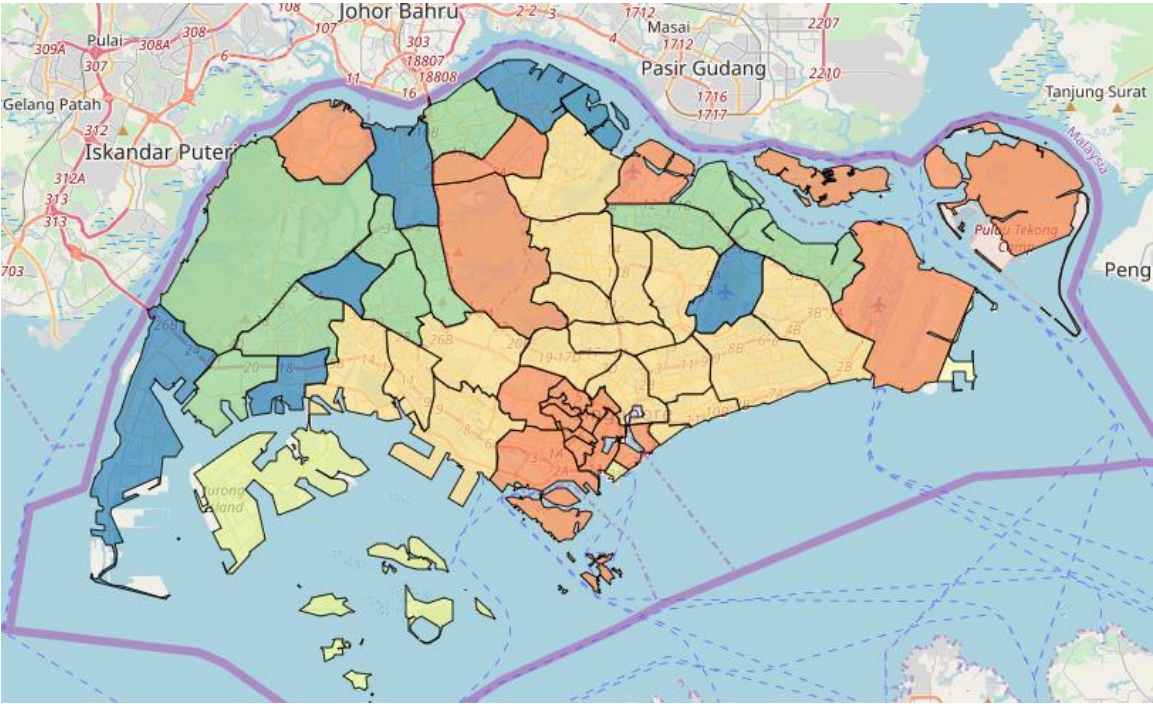


Figure 7: Cluster Map of Planning Areas

We merge those population statistics with cluster information in our main table.

Cluster Labels	Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	ANG MO KIO	1.060903	0.921914	0.119718	0.237560	0.123397	0.585885	0.641115	0.720930	Food Court	Chinese Restaurant	Coffee Shop
1	BEDOK	1.032859	0.956226	0.159574	0.228261	0.094978	0.598147	0.602281	0.672986	Chinese Restaurant	Coffee Shop	Food Court
1	BISHAN	1.020555	1.027919	0.213793	0.353791	0.083333	0.599459	0.606818	0.590278	Chinese Restaurant	Coffee Shop	Food Court
5	BOON LAY	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Exhibit	Boat or Ferry	Bus Station
4	BUKIT BATOK	0.978717	0.985199	0.176707	0.341686	0.057078	0.590517	0.608828	0.745283	Food Court	Coffee Shop	Fast Food Restaurant

Figure 8: Cluster Table with most common venues and statistical data

Then list out full table of each clusters:

Cluster 1:

There are many hotels, airports, piers and many natural reserve sites. Besides, the statistical data is not too much in this cluster. Therefore, we can conclude that the planning areas in this cluster are mainly tourist attraction and other non-residential areas.

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
BUKIT MERAH	1.079188	0.915892	0.113744	0.26	0.194545	0.572112	0.707627	0.711957	Café	Japanese Restaurant	Park
CENTRAL WATER CATCHMENT	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Exhibit	Flower Shop	Chinese Restaurant
CHANGI	0.872823	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Café	Airport Lounge	Sandwich Place
DOWNTOWN CORE	1.148501	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Hotel	Waterfront	Boutique
LIM CHU KANG	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Farm	Campground	Asian Restaurant
MANDAI	0.923035	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Chinese Restaurant	Bus Stop	Bus Station
MARINA EAST	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Satay Restaurant	Scenic Lookout	Seafood Restaurant
MARINA SOUTH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Garden	Hotel	Scenic Lookout
MUSEUM	1.049367	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Japanese Restaurant	Hotel	Café
NEWTON	1.026918	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Hotel	Boutique	Japanese Restaurant
NORTH-EASTERN ISLANDS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Trail	Campground	Pier
ORCHARD	1.075468	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Hotel	Boutique	Japanese Restaurant
OUTRAM	1.102253	0.898256	0.173913	0.20	0.309524	0.609756	0.746988	0.750000	Japanese Restaurant	Hotel	Café
RIVER VALLEY	1.003383	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Japanese Restaurant	Coffee Shop	Shopping Mall
ROCHOR	1.110386	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Café	Indian Restaurant	Hotel
SELETAR	0.969588	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Airport	Café	Restaurant
SINGAPORE RIVER	0.941981	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Japanese Restaurant	Hotel	Café
SOUTHERN ISLANDS	0.942756	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Theme Park Ride / Attraction	Beach	Theme Park
TANGLIN	0.995520	1.141275	0.172414	0.00	0.279412	0.695652	0.277108	0.178571	Hotel	Boutique	Japanese Restaurant

Figure 9: Cluster 1: Tourist Attraction or other Non-Residential Areas such as Airports, Natural Reserve

Cluster 2:

There are many food courts and restaurants in Cluster 2, we can conclude that they are mainly traditional residential areas.

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
ANG MO KIO	1.060903	0.921914	0.119718	0.237560	0.123397	0.585885	0.641115	0.720930	Food Court	Chinese Restaurant	Coffee Shop
BEDOK	1.032859	0.956226	0.159574	0.228261	0.094978	0.598147	0.602281	0.672986	Chinese Restaurant	Coffee Shop	Food Court
BISHAN	1.020555	1.027919	0.213793	0.353791	0.083333	0.599459	0.606818	0.590278	Chinese Restaurant	Coffee Shop	Food Court
BUKIT TIMAH	0.999559	1.106425	0.189394	0.038136	0.110638	0.681034	0.327160	0.362903	Bakery	Café	Italian Restaurant
CLEMENTI	1.048825	1.051397	0.142857	0.284281	0.097315	0.615714	0.629073	0.592593	Food Court	Café	Bakery
GEYLANG	1.048308	0.920001	0.174194	0.264865	0.142857	0.585338	0.698656	0.741007	Food Court	Chinese Restaurant	Asian Restaurant
HOUGANG	1.007592	0.974907	0.162234	0.394081	0.053208	0.596413	0.628518	0.741433	Coffee Shop	Chinese Restaurant	Food Court
JURONG EAST	1.010607	0.967259	0.135338	0.316206	0.047059	0.612335	0.641791	0.737864	Chinese Restaurant	Coffee Shop	Food Court
KALLANG	1.074558	0.942921	0.109489	0.221607	0.190083	0.576923	0.678038	0.677419	Thai Restaurant	Chinese Restaurant	Café
MARINE PARADE	1.043112	1.018101	0.180328	0.112583	0.173333	0.594118	0.521739	0.517857	Chinese Restaurant	Noodle House	Coffee Shop
NOVENA	1.030286	0.992437	0.144928	0.140127	0.161290	0.613889	0.562500	0.562500	Chinese Restaurant	Hotel	Food Court
QUEENSTOWN	1.068673	0.984901	0.182540	0.222535	0.101983	0.560567	0.689904	0.732759	Coffee Shop	Café	Ice Cream Shop
SERANGOON	1.037789	1.005414	0.227027	0.274611	0.059585	0.573182	0.592040	0.672131	Chinese Restaurant	Coffee Shop	Noodle House
TAMPINES	0.973641	0.992003	0.177130	0.369811	0.043970	0.595078	0.631264	0.751269	Coffee Shop	Café	Bakery
TOA PAYOH	1.073971	0.947072	0.150000	0.239234	0.142191	0.570594	0.718381	0.730769	Chinese Restaurant	Food Court	Coffee Shop
YISHUN	0.958537	0.927308	0.109422	0.474522	0.049521	0.627057	0.683092	0.775510	Coffee Shop	Chinese Restaurant	Thai Restaurant

Figure 10: Cluster 2: Traditional Residential Area, full of Food Court and Restaurants

Cluster 3:

This is simply a Naval Port

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
CHANGI BAY	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Boat or Ferry	History Museum	Military Base	Smoke Shop

Figure 11: Cluster 3: Naval Port

Cluster 4:

Simply Civil Piers

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
STRAITS VIEW	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Boat or Ferry	Pier	Metro Station	Mexican Restaurant
WESTERN ISLANDS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Boat or Ferry	Pier	Other Great Outdoors	Tourist Information Center

Figure 12: Cluster 4: Civil Piers

Cluster 5:

There are many food courts and restaurants in Cluster 5, especially fast food. The average age is younger than Cluster 2 and the tenancy rate is much lower than Cluster 2. We can conclude that they are emerging residential areas.

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
BUKIT BATOK	0.978717	0.985199	0.176707	0.341686	0.057078	0.590517	0.608828	0.745283	Food Court	Coffee Shop	Fast Food Restaurant
BUKIT PANJANG	0.933948	1.019983	0.107143	0.378641	0.046117	0.646958	0.484520	0.664894	Chinese Restaurant	Coffee Shop	Fast Food Restaurant
CHOA CHU KANG	0.924337	1.012545	0.136612	0.425963	0.030612	0.616552	0.624549	0.798561	Coffee Shop	Fast Food Restaurant	Park
JURONG WEST	0.928378	0.999297	0.101449	0.381127	0.062731	0.640598	0.585470	0.795000	Fast Food Restaurant	Food Court	Coffee Shop
PASIR RIS	0.931741	1.024287	0.166113	0.301235	0.049505	0.614853	0.566176	0.744186	Park	Coffee Shop	Fast Food Restaurant
PIONEER	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	Fast Food Restaurant	Food Court
PUNGGOL	0.819803	1.166680	0.065574	0.442771	0.039157	0.771753	0.561587	0.611111	Fast Food Restaurant	Supermarket	Chinese Restaurant
SENGKANG	0.876428	1.067404	0.073434	0.434109	0.035659	0.708817	0.614141	0.667785	Fast Food Restaurant	Food Court	Coffee Shop
WESTERN WATER CATCHMENT	0.929173	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	Fast Food Restaurant	Gym
WOODLANDS	0.901022	0.970771	0.073864	0.424113	0.062589	0.631959	0.652837	0.810160	Food Court	Fast Food Restaurant	Park

Figure 13: Cluster 5: Emerging Residential Area, full of Fast Food Restaurants

Cluster 6:

There are many bus stations and cargo pier related venues. The statistical data is lacking for most of them. They are typical industrial areas in Singapore.

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
BOON LAY	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Exhibit	Boat or Ferry	Bus Station
PAYA LEBAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Bus Station	Bus Line	Bus Stop
SEMBAWANG	0.893986	1.062208	0.139241	0.402597	0.03913	0.668262	0.678832	0.75	Coffee Shop	Chinese Restaurant	Bus Station
SIMPANG	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	River	Seafood Restaurant
SUNGEI KADUT	1.151525	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	Café	Bus Station
TENGAH	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Bus Station	Indian Restaurant	Gas Station
TUAS	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Asian Restaurant	Food Court	Boat or Ferry

From above analysis of 6 clusters in Singapore, we know only Cluster 2 and 5 (traditional and emerging residential areas) are suitable for living. For Cluster 5, the average age is lower, the average income is higher, which will give new immigrants a more energetic environment. The marriage rate is higher and tenancy rate is lower, which means the house renting rate is much lower in Cluster 5. At the meanwhile, the HDB (public housing in Singapore)⁵ rate is higher, which means the available flats with affordable selling price is much more in Cluster 5, so it is more cost-effective no matter the young immigrant couples decide to rent or purchase a flat.

Therefore, we can conclude that Cluster 5, the emerging residential areas are more suitable for young immigrant couples to live in Singapore.

⁵ https://en.wikipedia.org/wiki/Housing_and_Development_Board

5. Discussion

There are some shortage of this study: The usage of Foursquare in Singapore is not as frequent as in USA, so the statistical result of venues may not be too meaningful; Although the official statistical data of Singapore is available in many fields, there is little public data based on planning areas, e.g. the statistical data used in this project is given by interval, so I can only take the mean value of the interval, which is not accurate enough.

There are 2 special distinct minor clusters, so there are only 4 majority useful clusters in the study. We should increase the number of clusters from 6 to 7. However, because of the restriction of threshold scale number of Folium Choropleth map, more than 6 clusters will not have good visual effect of the cluster map. Since this study is based on areas with boundaries instead of spots in other projects, Choropleth map is necessary to visualize the planning areas.

In the future, this study can be improved with more accurate statistics of Singapore, more usage of Foursquare in Singapore, better folium maps and more statistical data of other fields.