Analysis HDB Resale Price vs Venues Data in Singapore

Sazili Muhammad

29 December, 2020

1. Introduction

1.1 Background

Majority of Singapore population are living in the public housing, in which the developments are publicly governed and developed by the state Housing and Development Board (HDB) under a 99-year lease. These flats are located in housing estates, which are self-contained satellite towns with well-maintained schools, supermarkets, malls, community hospitals, clinics, hawker centers (food court) and sports and recreational facilities. Every housing estate includes MRT stations and bus stops that link residents to other parts of the city-state.

As compared to most parts of the world, public housing in Singapore is not ostracized by a wide majority of the population and its government, and acts as a necessary and vital measure to provide immaculate and safe housing surrounded by public amenities at affordable prices, especially during its rapid development and industrialization in the early years of independence. It is also meant to foster social cohesion between social classes and races of Singapore, and prevent neglected areas or districts and ethnic enclaves from developing. As such, it is considered a unique part of Singaporean culture and identity, being commonly associated with the country.

There are a large variety of flat types and layouts which cater to various housing budgets. HDB flats were built in mind to primarily provide affordable housing for Singaporeans/Permanent Residents and their purchase can be financially aided by the Central Provident Fund (CPF) in addition to various grants. Due to changing demands, HDB introduced the Design, Build and Sell Scheme to produce up-market public housing developments.

New public housing flats are strictly only eligible for purchase towards Singaporean citizens. The housing schemes and grants available to finance the purchase of a flat are also only extended to households owned by Singaporeans, while Permanent Residents do not get any housing grants or subsidies from the Singaporean government and could only purchase resale flats from the secondary market at a market price. Such policies have helped Singapore reach a home-ownership rate of 91%, one of the highest in the world. In 2008, Singapore was lauded by the United Nations Habitat's State of the World's Cities report as the only slum-free city in the world. [1]

1.2 Problem

HDB flats surrounded by various amenities such as hawker centers, coffee shop, school, carpark, community centers, shopping center/malls, playground, sport center. Those amenities impact to housing resale price fluctuation and hence resale prices among towns in Singapore are not balanced. Some town are highly popular and more matured, makes increasing of resale price faster than other town.

This situation makes house buyer has difficulties and limitation to purchase flats as they need. The challenges also facing to business owner to invest which location has more profitable and more promising. For government and regulation makers will have difficulties to make fair decision to build facilities in certain location.

1.3 Purpose

This project will try to analyze the housing resale price vs various amenities (venues) in different angle. This project will give better insight to house buyer, business owner, or government/regulator to have better decision to invest and improve in various location across the island.

2. Data aquisition and data cleaning

2.1 Data Sources

The dataset for this project are captured from:

Data.gov.sg [2]. This is Singapore government body which release data for public purpose.

Google Maps to capture latitude and longitude.

2.2 Data cleaning and data selection

Originally this data is very detail which comprises resale price from all towns in Singapore from Jan-2017 until Oct-2020 which has in detail block number, flat type, and flat floor. Since this project is not purpose into detail flat address but instead only interested on towns. So that, I clean up and summarize the dataset into average of housing resale price across all Singapore towns for year 2020 (since Jan-Oct).

Note that, all the prices is in Singapore dollars (SGD).

The total Singapore towns are 26 towns, as follows:

	Town	Resale Price
0	ANG MO KIO	400148.74
1	BEDOK	405873.45
2	BISHAN	627302.21
3	BUKIT BATOK	397548.97
4	BUKIT MERAH	562416.43
5	BUKIT PANJANG	432441.11
6	BUKIT TIMAH	696836.38
7	CENTRAL AREA	585125.84
8	CHOA CHU KANG	404028.30
9	CLEMENTI	493022.24
10	GEYLANG	429776.44
11	HOUGANG	449869.59
12	JURONG EAST	407133.80
13	JURONG WEST	407372.67
14	KALLANG/WHAMPOA	475455.08
15	MARINE PARADE	480472.48
16	PASIR RIS	499348.52
17	PUNGGOL	462420.83
18	QUEENSTOWN	584474.49
19	SEMBAWANG	392777.98
20	SENGKANG	450835.38
21	SERANGOON	487274.65
22	TAMPINES	478879.13
23	TOA PAYOH	442313.77
24	WOODLANDS	392932.61
25	YISHUN	379274.21

In order to use Foursquare to find venues data, I capture latitude and longitude for each Singapore town from Google Maps, as shown below:

	Town	Latitude	Longitude
0	ANG MO KIO	1.3691	103.8454
1	BEDOK	1.3236	103.9273
2	BISHAN	1.3526	103.8352
3	BUKIT BATOK	1.3590	103.7637
4	BUKIT MERAH	1.2819	103.8239
5	BUKIT PANJANG	1.3774	103.7719
6	BUKIT TIMAH	1.3294	103.8021
7	CENTRAL AREA	1.2789	103.8536
8	CHOA CHU KANG	1.3840	103.7470
9	CLEMENTI	1.3162	103.7649
10	GEYLANG	1.3201	103.8918
11	HOUGANG	1.3612	103.8863
12	JURONG EAST	1.3329	103.7436
13	JURONG WEST	1.3404	103.7090
14	KALLANG/WHAMPOA	1.3245	103.8572
15	MARINE PARADE	1.3020	103.8971
16	PASIR RIS	1.3721	103.9474
17	PUNGGOL	1.3984	103.9072
18	QUEENSTOWN	1.2942	103.7861
19	SEMBAWANG	1.4491	103.8185
20	SENGKANG	1.3868	103.8914
21	SERANGOON	1.3554	103.8679
22	TAMPINES	1.3496	103.9568
23	TOA PAYOH	1.3343	103.8563
24	WOODLANDS	1.4382	103.7890
25	YISHUN	1.4304	103.8354

Combine original data from Data.gov.sg and Google maps (latitude/longitude), hence getting final dataset as follows:

	Town	Resale Price	Latitude	Longitude
0	BUKIT TIMAH	696836.38	1.3294	103.8021
1	BISHAN	627302.21	1.3526	103.8352
2	CENTRAL AREA	585125.84	1.2789	103.8536
3	QUEENSTOWN	584474.49	1.2942	103.7861
4	BUKIT MERAH	562416.43	1.2819	103.8239
5	PASIR RIS	499348.52	1.3721	103.9474
6	CLEMENTI	493022.24	1.3162	103.7649
7	SERANGOON	487274.65	1.3554	103.8679
8	MARINE PARADE	480472.48	1.3020	103.8971
9	TAMPINES	478879.13	1.3496	103.9568
10	KALLANG/WHAMPOA	475455.08	1.3245	103.8572
11	PUNGGOL	462420.83	1.3984	103.9072
12	SENGKANG	450835.38	1.3868	103.8914
13	HOUGANG	449869.59	1.3612	103.8863
14	TOA PAYOH	442313.77	1.3343	103.8563
15	BUKIT PANJANG	432441.11	1.3774	103.7719
16	GEYLANG	429776.44	1.3201	103.8918
17	JURONG WEST	407372.67	1.3404	103.7090
18	JURONG EAST	407133.80	1.3329	103.7436
19	BEDOK	405873.45	1.3236	103.9273
20	CHOA CHU KANG	404028.30	1.3840	103.7470
21	ANG MO KIO	400148.74	1.3691	103.8454
22	BUKIT BATOK	397548.97	1.3590	103.7637
23	WOODLANDS	392932.61	1.4382	103.7890
24	SEMBAWANG	392777.98	1.4491	103.8185
25	YISHUN	379274.21	1.4304	103.8354

Hence, I will use this dataset above for further analysis for this project.

3. Methodology

The methodology in this project, I am using clustering approach based on the most common venues in various towns in Singapore. The venues data is getting from Foursquare website.

First, I will Singapore town data including the latitude and longitude.

Then, I will get venues data from Foursquare for each town within 1 km and limit to 1000 venues. This data collection based on the latitude and longitude from each town.

I create bar chart to show total venues for each town and bar chart to show housing resale price for each town, side-by-side, to visualize the correlation between total venues effected the housing resale price.

Another analysis, I am using clustering which finding the best cluster number using Elbow Method to find optimal k.

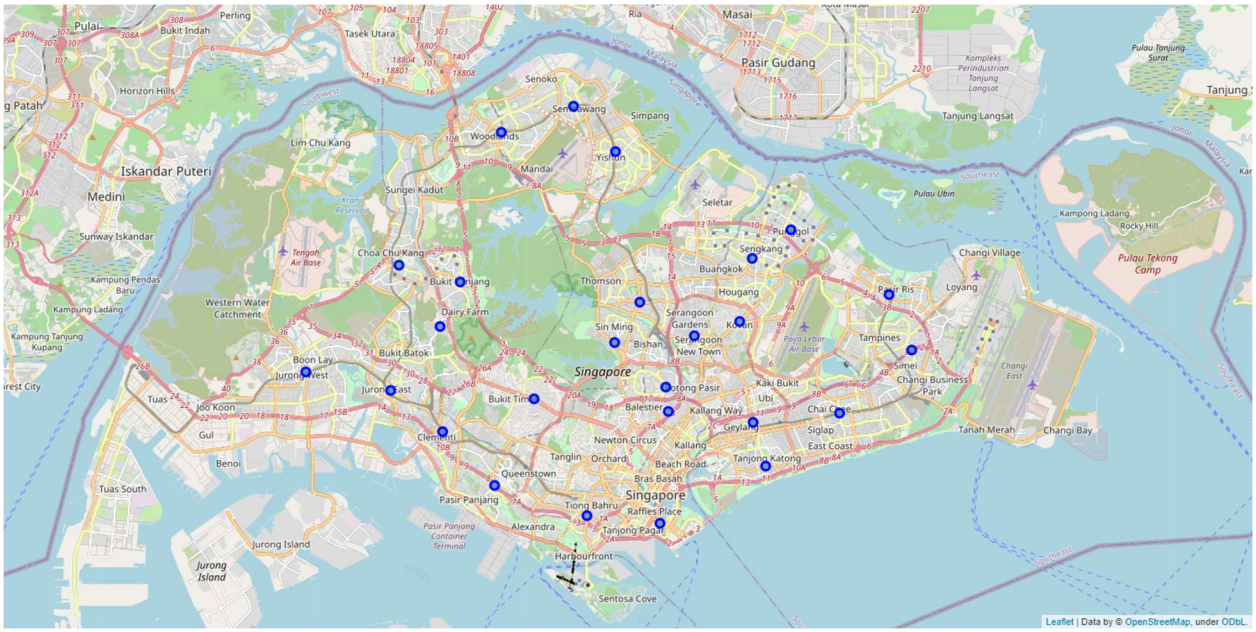
This k-cluster will be used to defined clustering for all towns and visualize in map.

The detail analysis will show as below.

4. Analysis

In section Data, I grabbed the Singapore town data including latitude and longitude.

Now, I will visualize all the Singapore town in the map as per starting analysis.



The venues data collection from Foursquare is processed using this steps below. For each town will search venues in radius 1km which limited maximum up to 1000 venues.

Below is the result of venues data for each town.

	Town	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BUKIT TIMAH	1.3294	103.8021	Peperoni Pizzeria	1.331104	103.807267	Pizza Place
1	BUKIT TIMAH	1.3294	103.8021	La Braceria Pizza & Grill	1.331761	103.807206	Italian Restaurant
2	BUKIT TIMAH	1.3294	103.8021	73@Hillcrest	1.331471	103.807556	Bistro
3	BUKIT TIMAH	1.3294	103.8021	Plank Sourdough Pizza By Baker & Cook	1.323890	103.796797	Pizza Place
4	BUKIT TIMAH	1.3294	103.8021	Lana Cake Shop	1.331789	103.806326	Bakery
...	...	...	...	...	...	...	...
1824	YISHUN	1.4304	103.8354	Warong Penyet Super	1.431995	103.827984	Asian Restaurant
1825	YISHUN	1.4304	103.8354	Chuan Kee Boneless Braised Duck	1.431506	103.828173	Chinese Restaurant
1826	YISHUN	1.4304	103.8354	Sincere Medical & Dental Clinic	1.437493	103.837571	Pharmacy
1827	YISHUN	1.4304	103.8354	Basketball Court @ Blk 761	1.423553	103.831767	Basketball Court
1828	YISHUN	1.4304	103.8354	Kim San Leng (金山岭)	1.434334	103.843033	Coffee Shop

1829 rows × 7 columns

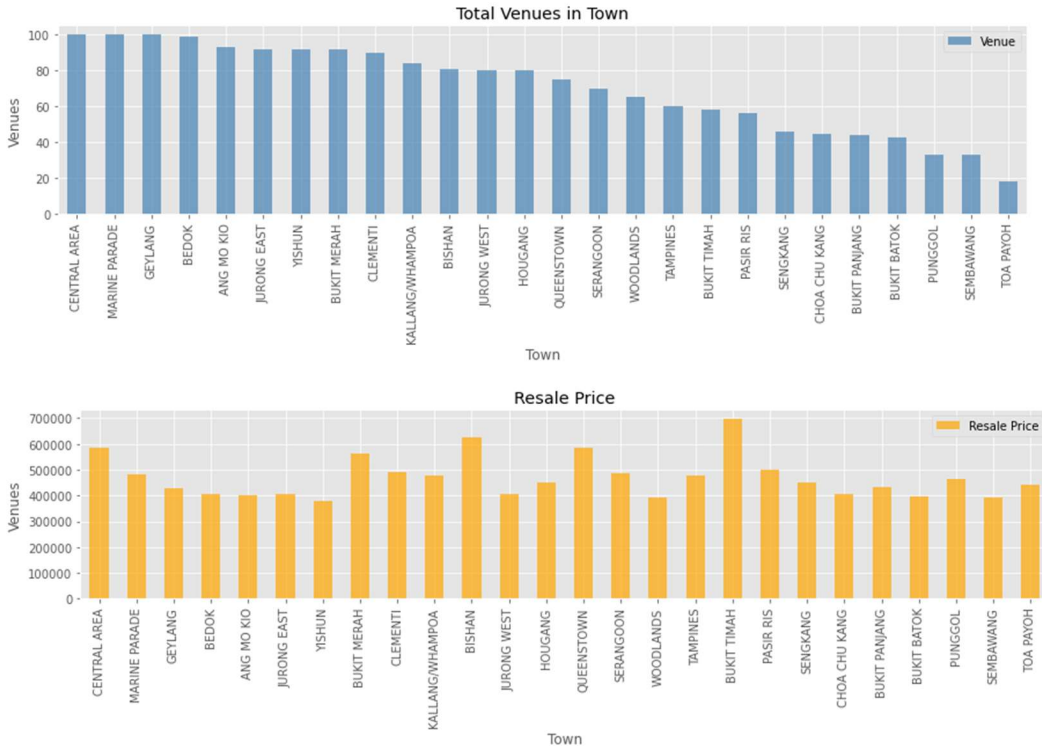
1. Correlation between HDB Resale Price vs Total Venues

The first analysis is to visualize the correlation between HDB Resale Price vs Total Venues.

From the collected venues data, then I will summarize the total venues for each town, as describe in following steps:

	Town	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Resale Price	Latitude	Longitude
0	CENTRAL AREA	100	100	100	100	100	100	585125.84	1.2789	103.8536
1	MARINE PARADE	100	100	100	100	100	100	480472.48	1.3020	103.8971
2	GEYLANG	100	100	100	100	100	100	429776.44	1.3201	103.8918
3	BEDOK	99	99	99	99	99	99	405873.45	1.3236	103.9273
4	ANG MO KIO	93	93	93	93	93	93	400148.74	1.3691	103.8454
5	JURONG EAST	92	92	92	92	92	92	407133.80	1.3329	103.7436
6	YISHUN	92	92	92	92	92	92	379274.21	1.4304	103.8354
7	BUKIT MERAH	92	92	92	92	92	92	562416.43	1.2819	103.8239
8	CLEMENTI	90	90	90	90	90	90	493022.24	1.3162	103.7649
9	KALLANG/WHAMPOA	84	84	84	84	84	84	475455.08	1.3245	103.8572
10	BISHAN	81	81	81	81	81	81	627302.21	1.3526	103.8352
11	JURONG WEST	80	80	80	80	80	80	407372.67	1.3404	103.7090
12	HOUGANG	80	80	80	80	80	80	449869.59	1.3612	103.8863
13	QUEENSTOWN	75	75	75	75	75	75	584474.49	1.2942	103.7861
14	SERANGOON	70	70	70	70	70	70	487274.65	1.3554	103.8679
15	WOODLANDS	65	65	65	65	65	65	392932.61	1.4382	103.7890
16	TAMPINES	60	60	60	60	60	60	478879.13	1.3496	103.9568
17	BUKIT TIMAH	58	58	58	58	58	58	696836.38	1.3294	103.8021
18	PASIR RIS	56	56	56	56	56	56	499348.52	1.3721	103.9474
19	SENGKANG	46	46	46	46	46	46	450835.38	1.3868	103.8914
20	CHOA CHU KANG	45	45	45	45	45	45	404028.30	1.3840	103.7470
21	BUKIT PANJANG	44	44	44	44	44	44	432441.11	1.3774	103.7719
22	BUKIT BATOK	43	43	43	43	43	43	397548.97	1.3590	103.7637
23	PUNGGOL	33	33	33	33	33	33	462420.83	1.3984	103.9072
24	SEMBAWANG	33	33	33	33	33	33	392777.98	1.4491	103.8185
25	TOA PAYOH	18	18	18	18	18	18	442313.77	1.3343	103.8563

From the data above, I created bar chart for total venues and bar chart for housing resale price for each town.

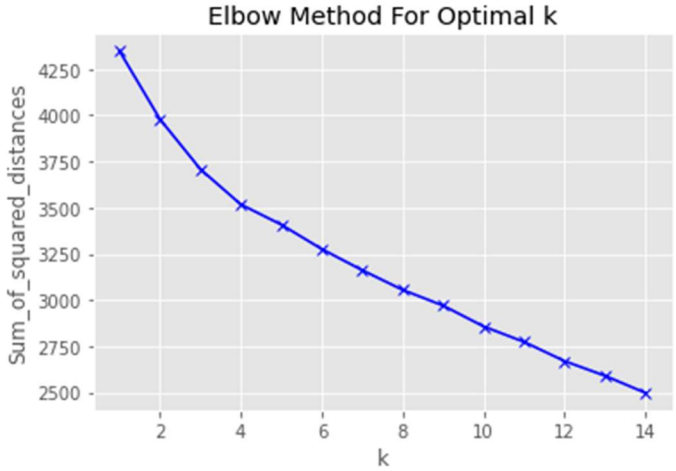


Hence from the bar chart above, it is shown that housing resale price is not depend on the total venues.

2. Finding optimal k and clustering approach

In order to find the optimal k, the best method is using Elbow Method.

This process is described as below steps.

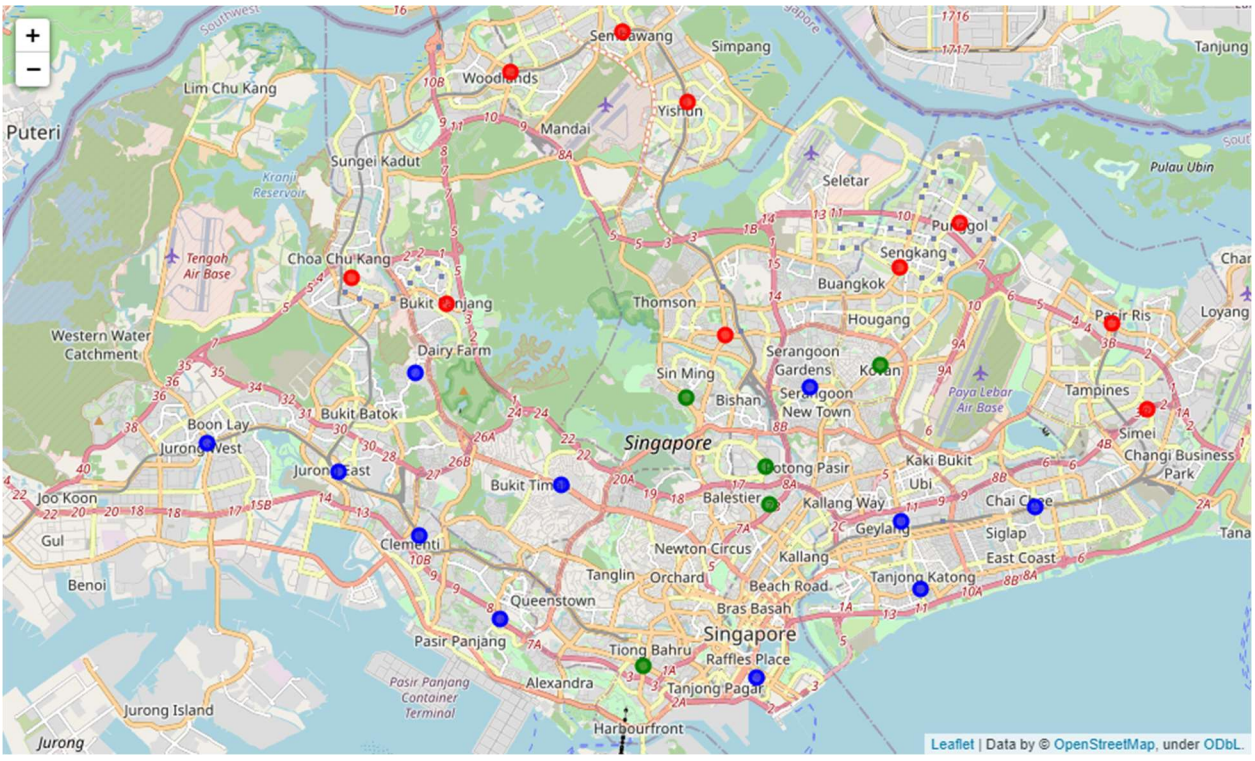


Even though elbow method is not so obviously see the elbow in the graph, but I can estimate **the best k cluster is 3.**

Now, I will generate clustering based on the k cluster above.

	Town	Resale Price	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	...	15th Most Common Venue	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Resale Price_right	Latitude_right	Longitude_right
0	BISHAN	627302.21	1.3526	103.8352	0	Chinese Restaurant	Café	Thai Restaurant	Bakery	Coffee Shop	...	Dessert Shop	81	81	81	81	81	81	627302.21	1.3526	103.8352
1	BUKIT MERAH	562416.43	1.2819	103.8239	0	Chinese Restaurant	Food Court	Coffee Shop	Noodle House	Café	...	Japanese Restaurant	92	92	92	92	92	92	562416.43	1.2819	103.8239
2	TOA PAYOH	442313.77	1.3343	103.8563	0	Chinese Restaurant	Coffee Shop	Pool	Trail	Snack Place	...	Filipino Restaurant	18	18	18	18	18	18	442313.77	1.3343	103.8563
3	HOUGANG	449869.59	1.3612	103.8863	0	Chinese Restaurant	Bus Station	Coffee Shop	Noodle House	Food Court	...	Building	80	80	80	80	80	80	449869.59	1.3612	103.8863
4	KALLANG/WHAMPOA	475455.08	1.3245	103.8572	0	Chinese Restaurant	Noodle House	Asian Restaurant	Food Court	Dessert Shop	...	Hostel	84	84	84	84	84	84	475455.08	1.3245	103.8572
5	BUKIT TIMAH	696836.38	1.3294	103.8021	1	Bakery	Italian Restaurant	Pizza Place	Café	Hotel	...	Playground	58	58	58	58	58	58	696836.38	1.3294	103.8021
6	GEYLANG	429776.44	1.3201	103.8918	1	Food Court	Chinese Restaurant	Asian Restaurant	Noodle House	Fast Food Restaurant	...	Coffee Shop	100	100	100	100	100	100	429776.44	1.3201	103.8918
7	BEDOK	405873.45	1.3236	103.9273	1	Food Court	Chinese Restaurant	Coffee Shop	Noodle House	Supermarket	...	Pizza Place	99	99	99	99	99	99	405873.45	1.3236	103.9273
8	JURONG WEST	407372.67	1.3404	103.7090	1	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Chinese Restaurant	Food Court	...	Sandwich Place	80	80	80	80	80	80	407372.67	1.3404	103.7090
9	BUKIT BATOK	397548.97	1.3590	103.7637	1	Italian Restaurant	Café	Historic Site	Shopping Mall	Bus Station	...	Bus Stop	43	43	43	43	43	43	397548.97	1.3590	103.7637
10	SERANGOON	487274.65	1.3554	103.8679	1	Chinese Restaurant	Café	Asian Restaurant	Coffee Shop	Noodle House	...	Shopping Mall	70	70	70	70	70	70	487274.65	1.3554	103.8679
11	CLEMENTI	493022.24	1.3162	103.7649	1	Food Court	Chinese Restaurant	Coffee Shop	Indian Restaurant	Bakery	...	Chinese Breakfast Place	90	90	90	90	90	90	493022.24	1.3162	103.7649
12	QUEENSTOWN	584474.49	1.2942	103.7861	1	Coffee Shop	Café	Asian Restaurant	Sandwich Place	Indian Restaurant	...	Snack Place	75	75	75	75	75	75	584474.49	1.2942	103.7861
13	CENTRAL AREA	585125.84	1.2789	103.8536	1	Hotel	Café	Japanese Restaurant	Gym / Fitness Center	Waterfront	...	Mexican Restaurant	100	100	100	100	100	100	585125.84	1.2789	103.8536
14	MARINE PARADE	480472.48	1.3020	103.8971	1	Chinese Restaurant	Asian Restaurant	Café	Noodle House	Japanese Restaurant	...	Massage Studio	100	100	100	100	100	100	480472.48	1.3020	103.8971
15	JURONG EAST	407133.80	1.3329	103.7436	1	Coffee Shop	Café	Japanese Restaurant	Food Court	Chinese Restaurant	...	Sushi Restaurant	92	92	92	92	92	92	407133.80	1.3329	103.7436
16	ANG MO KIO	400148.74	1.3691	103.8454	2	Chinese Restaurant	Food Court	Coffee Shop	Fast Food Restaurant	Japanese Restaurant	...	Bubble Tea Shop	93	93	93	93	93	93	400148.74	1.3691	103.8454
17	CHOA CHU KANG	404028.30	1.3840	103.7470	2	Fast Food Restaurant	Coffee Shop	Food Court	Café	Bookstore	...	Sandwich Place	45	45	45	45	45	45	404028.30	1.3840	103.7470
18	WOODLANDS	392932.61	1.4382	103.7890	2	Food Court	Fast Food Restaurant	Japanese Restaurant	Coffee Shop	Supermarket	...	Pizza Place	65	65	65	65	65	65	392932.61	1.4382	103.7890
19	SENGKANG	450835.38	1.3868	103.8914	2	Food Court	Fast Food Restaurant	Coffee Shop	Bus Line	Chinese Restaurant	...	Food Stand	46	46	46	46	46	46	450835.38	1.3868	103.8914
20	SEMBAWANG	392777.98	1.4491	103.8185	2	Coffee Shop	Bus Station	Chinese Restaurant	Park	Food Court	...	Convenience Store	33	33	33	33	33	33	392777.98	1.4491	103.8185

Based on the data above, it is enough data to visualize the clustering in the map, as follows:



Finally in the map, it is shown 3 cluster for all town in Singapore based on venues data from Foursquare.

5. Results and Discussion

In the section Analysis shows that there are 3 towns has the highest amount of venues, they are CENTRAL AREA, MARINE PARADE, and GEYLANG. This is no doubt because those 3 towns are considered matured town, which means these towns has been developed earlier than other towns.

In the bar chart, we can see that all towns has uniformity in venues which mostly are food and restaurant. The amount of venues for each town has small gap between other town. There are no town that has very high venues compare with others. This means every town in Singapore is equally developed.

In the bar chart of housing resale price shows that the price does not depend on the total venues.

Based on Elbow Method which described in Analysis section, I can find the optimal k-cluster = 3.

The home buyer or business owner can focus on Cluster 2 which has lower housing price and has fewer variety of venues. The home buyer has more opportunities to buy their dream house. The business owner can invest other business sector which has lower competitor.

6. Conclusion

Based on analyst and result which the data given from Singapore HDB Resale Price and Foursquare data, I can conclude these following:

There is no proof that the amount of venue in certain location effected the housing resale price in Singapore

Housing resale price in Singapore basically depend upon the amenities, maturity of the town, and location near city area, while Foursquare doesn't have enough data to support those criterias.

Home buyer can target the house in Cluster 2 which is lower house price.

Business owner can invest other business sector in Cluster 2 which those towns has lower competitor compared with other towns.

7. References

[1] Public Housing in Singapore

[2] Resale Flat Prices