Living in Singapore

Introduction

Singapore's rising cost of living results from exorbitant housing prices. Singapore, being a small country, is densely populated. Moreover, as Singapore becomes more developed as one of the major trade hubs, this busy city is becoming a very highly skilled one. With greater purchasing power, the locals are looking towards a better lifestyle, striving for material items, with housing as one of the choices. The change in lifestyle, the lack of the availability of land and the relatively large population size push the houses prices up in Singapore. While married couples have to option to apply for Build-to-Order flats which are subsidized by the government, this option is sadly not available for singles, who can only purchase resale flats. Purchasing a cheap flat in Singapore is not an easy thing. In this project, we will look at possible options of housing area for single Singaporeans and compare the prices, ultimately making some recommendations.

Data Description

The data we need are mainly the housing prices in Singapore, and the coordinates of the towns in Singapore.

- We obtained the data on prices of resale flats from https://data.gov.sg/.
    - Some assumptions made: The individuals will look for the cheapest option, which are 2 room flats (for practical purposes in terms of the size of the house as well).
- For the coordinates of the towns in Singapore, we obtained the dataset of the coordinates of train stations in Singapore from Kaggle. As the name of the train stations usually correspond to the towns the stations are in, we will regard this dataset to be suitable.

We will then compare the amenities around every town using Foursquare location data before we make a recommendation on the suitability of the housing area.

Data Preparation

We upload the dataset with the coordinates of the form:

	town	lat	lng
0	JURONG EAST	1.333207	103.742308
1	BUKIT BATOK	1.349069	103.749596
2	BUKIT GOMBAK	1.359043	103.751863
3	CHOA CHU KANG	1.385417	103.744316
4	YEW TEE	1.397383	103.747523

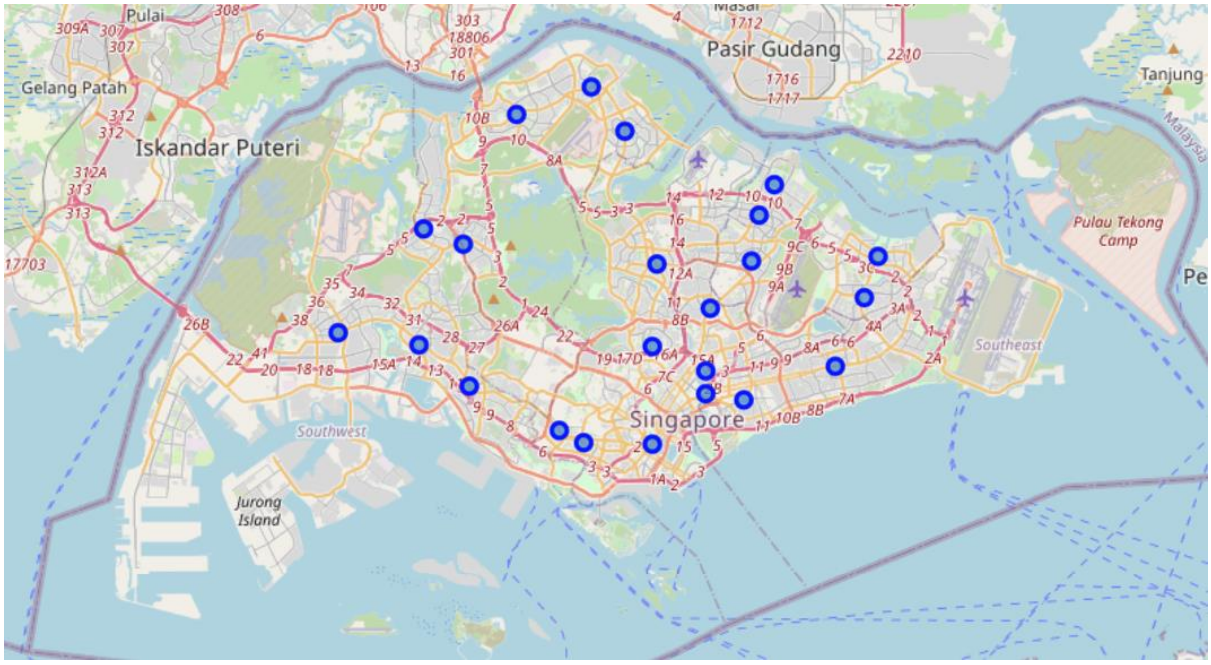
We upload the dataset with the housing prices of the form:

	town	resale_price
0	ANG MO KIO	213859.154930
1	BEDOK	219469.014085
2	BOON LAY	220723.525424
3	BUKIT PANJANG	233255.307692
4	CHOA CHU KANG	230084.848485

We now perform a left join of the datasets based on the 'town' column:

	town	resale_price	lat	lng
0	ANG MO KIO	213859.154930	1.370025	103.849588
1	BEDOK	219469.014085	1.324043	103.930205
2	BOON LAY	220723.525424	1.338620	103.705817
3	BUKIT PANJANG	233255.307692	1.378340	103.762452
4	CHOA CHU KANG	230084.848485	1.385417	103.744316

The dataframe has 23 towns. Let us have a look at the Singapore map labelled with the towns.



We explore the first town listed in our dataframe, Ang Mo Kio. 40 venues were returned by Foursquare. The table below is an example.

	name	categories	lat	lng
0	Old Chang Kee	Snack Place	1.369094	103.848389
1	FairPrice Xtra	Supermarket	1.369279	103.848886
2	MOS Burger	Burger Joint	1.369170	103.847831
3	A&W	Fast Food Restaurant	1.369541	103.849043
4	Face Ban Mian 非板面 (Ang Mo Kio)	Noodle House	1.372031	103.847504

We now consider the other towns together as well in the below table. There are 172 unique venue categories.

	Town Lat	Town Lat	Venue	Venue Lat	Venue Lng	Venue Category
Town						
ANG MO KIO	40	40	40	40	40	40
BEDOK	58	58	58	58	58	58
BOON LAY	74	74	74	74	74	74
BUKIT PANJANG	48	48	48	48	48	48
CHOA CHU KANG	20	20	20	20	20	20
CLARKE QUAY	100	100	100	100	100	100
CLEMENTI	63	63	63	63	63	63

We perform one hot encoding to allow categorical data to be more expressive.

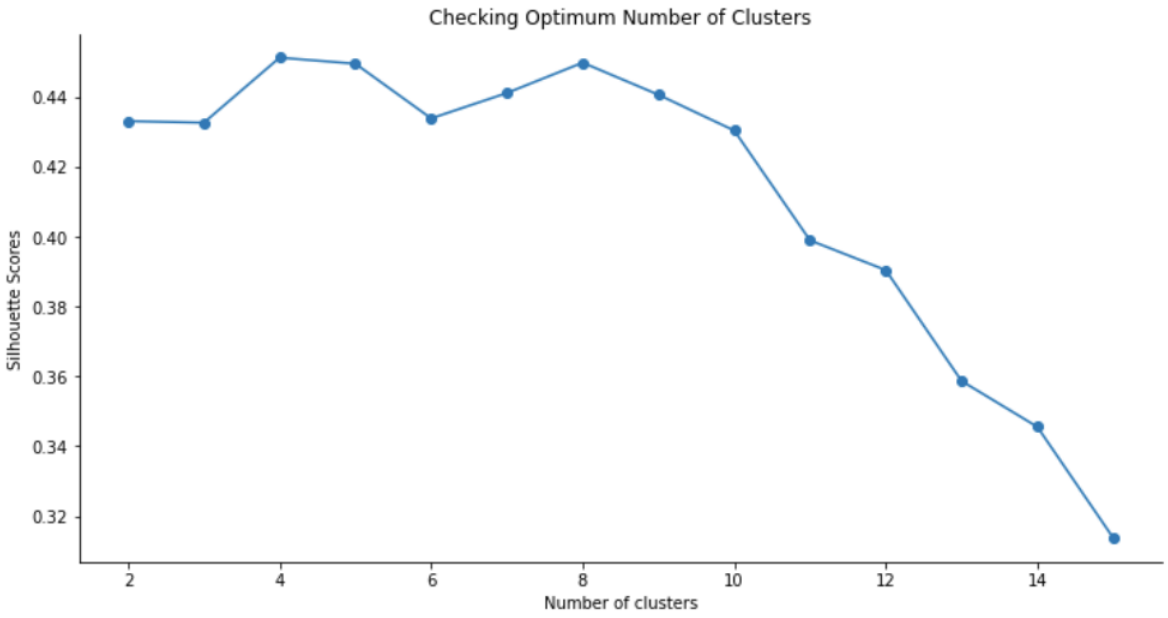
	Town	ATM	Accessories Store	American Restaurant	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Bar	Bas
0	ANG MO KIO	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.025000	0.000000	0.000000	0.025000	0.025	0.000000	0.0
1	BEDOK	0.000000	0.000000	0.017241	0.000000	0.00	0.000000	0.034483	0.000000	0.000000	0.017241	0.000	0.000000	0.0
2	BOON LAY	0.000000	0.000000	0.013514	0.000000	0.00	0.000000	0.094595	0.000000	0.013514	0.000000	0.000	0.000000	0.0
3	BUKIT PANJANG	0.000000	0.000000	0.020833	0.000000	0.00	0.000000	0.083333	0.000000	0.020833	0.020833	0.000	0.000000	0.0
4	CHOA CHU KANG	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.050000	0.000000	0.000000	0.050000	0.000	0.000000	0.0
5	CLARKE QUAY	0.000000	0.000000	0.000000	0.000000	0.02	0.000000	0.010000	0.000000	0.020000	0.010000	0.000	0.070000	0.0
6	CLEMENTI	0.000000	0.000000	0.015873	0.000000	0.00	0.015873	0.047619	0.000000	0.015873	0.015873	0.000	0.000000	0.0
7	DAKOTA	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.093750	0.000000	0.062500	0.000000	0.000	0.031250	0.0
8	GEYLANG	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000	0.000000	0.0

We calculate the top venues in each town as shown:

	Town	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	10th Most Common Venue
0	ANG MO KIO	Coffee Shop	Dessert Shop	Food Court	Japanese Restaurant	Bubble Tea Shop	Supermarket	Snack Place	Asian Restaurant	Fried Chicken Joint	Noodle House
1	BEDOK	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Food Court	Chinese Restaurant	Dessert Shop	Noodle House	Fast Food Restaurant	Supermarket
2	BOON LAY	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Dessert Shop	Chinese Restaurant	Coffee Shop	Indian Restaurant	Café	Gym / Fitness Center	Karaoke Bar
3	BUKIT PANJANG	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Sushi Restaurant	Shopping Mall	Korean Restaurant	Supermarket	Indonesian Restaurant	Bus Station	Café
4	CHOA CHU KANG	Fast Food Restaurant	Coffee Shop	Food Court	Bakery	Café	Bookstore	Supermarket	Sandwich Place	Food Truck	Thai Restaurant

Methodology

We will perform K-means clustering on the towns. To determine how many clusters to use, we use the Silhouette method.

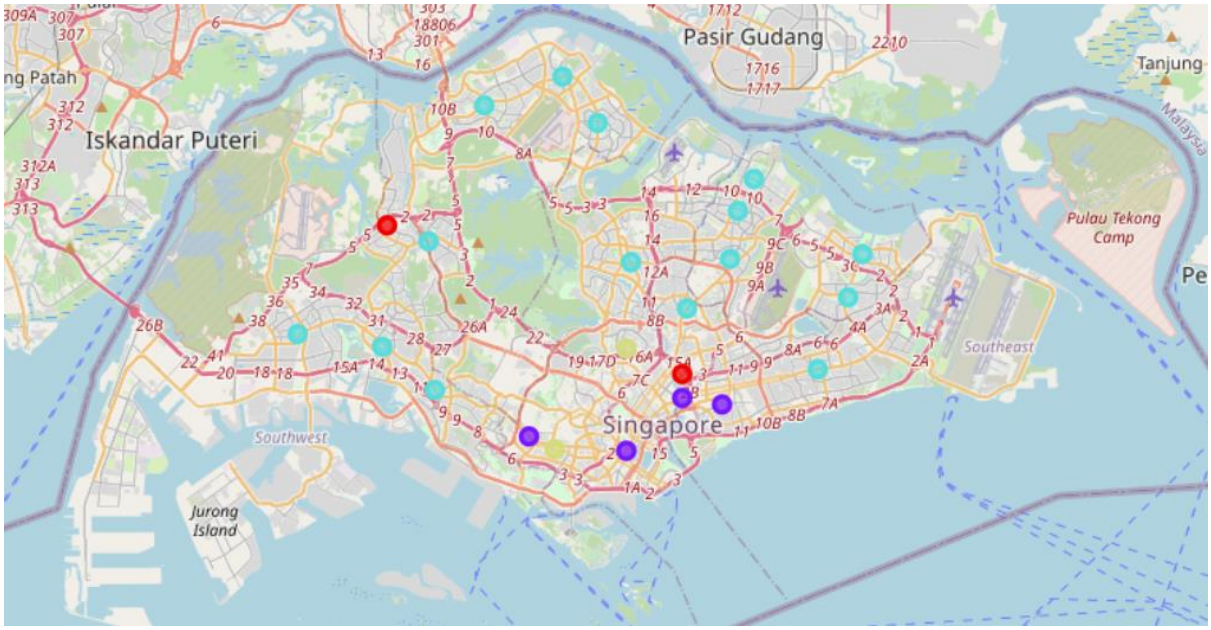


We see from the above graph that the optimum number of clusters is 4, having the highest Silhouette score.

We then perform K-means clustering with the final table merged below.

	town	resale_price	lat	lng	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	
0	ANG MO KIO	213859.154930	1.370025	103.849588	2	Coffee Shop	Dessert Shop	Food Court	Japanese Restaurant	Bubble Tea Shop	Supermarket	Snack Place	As Re
1	BEDOK	219469.014085	1.324043	103.930205	2	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Food Court	Chinese Restaurant	Dessert Shop	No Ho
2	BOON LAY	220723.525424	1.338620	103.705817	2	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Dessert Shop	Chinese Restaurant	Coffee Shop	Indian Restaurant	Ca
3	BUKIT PANJANG	233255.307692	1.378340	103.762452	2	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Sushi Restaurant	Shopping Mall	Korean Restaurant	Supermarket	In Re
4	CHOA CHU KANG	230084.848485	1.385417	103.744316	0	Fast Food Restaurant	Coffee Shop	Food Court	Bakery	Café	Bookstore	Supermarket	Sa Pl

The results are as follows.



Cluster 0:

	resale_price	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	10th Most Common Venue
4	230084.848485	Fast Food Restaurant	Coffee Shop	Food Court	Bakery	Café	Bookstore	Supermarket	Sandwich Place	Food Truck	Thai Restaurant
8	201506.097561	Coffee Shop	Indian Restaurant	Café	Food Court	Supermarket	Noodle House	Park	Dog Run	Donut Shop	Dumpling Restaurant

Cluster 1:

	resale_price	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	10th Most Common Venue
5	243653.846154	Bar	Nightclub	Hotel	Japanese Restaurant	Food Court	Noodle House	Café	Mexican Restaurant	Lounge	Italian Restaurant
7	216000.000000	Noodle House	Dessert Shop	Asian Restaurant	Chinese Restaurant	Seafood Restaurant	Snack Place	BBQ Joint	Coffee Shop	Food Court	Dim Sum Restaurant
11	225843.478261	Hostel	Food Court	Snack Place	BBQ Joint	Thai Restaurant	Historic Site	Noodle House	Chinese Restaurant	Rock Club	Seafood Restaurant
14	250556.338028	Food Court	Chinese Restaurant	Café	Noodle House	Indian Restaurant	Seafood Restaurant	Thai Restaurant	Stadium	Italian Restaurant	Train Station

Cluster 2:

	resale_price	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	10th Most Common Venue
0	213859.154930	Coffee Shop	Dessert Shop	Food Court	Japanese Restaurant	Bubble Tea Shop	Supermarket	Snack Place	Asian Restaurant	Fried Chicken Joint	Noodle House
1	219469.014085	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Food Court	Chinese Restaurant	Dessert Shop	Noodle House	Fast Food Restaurant	Supermarket
2	220723.525424	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Dessert Shop	Chinese Restaurant	Coffee Shop	Indian Restaurant	Café	Gym / Fitness Center	Karaoke Bar
3	233255.307692	Fast Food Restaurant	Asian Restaurant	Coffee Shop	Sushi Restaurant	Shopping Mall	Korean Restaurant	Supermarket	Indonesian Restaurant	Bus Station	Café
6	302217.391304	Coffee Shop	Food Court	Chinese Restaurant	Dessert Shop	Asian Restaurant	Chinese Breakfast Place	Fried Chicken Joint	Shopping Mall	Bus Station	Fast Food Restaurant
9	235102.857143	Fast Food Restaurant	Coffee Shop	Chinese Restaurant	Food Court	Supermarket	Café	Frozen Yogurt Shop	Shopping Mall	Soup Place	Pharmacy
10	240772.704545	Café	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Food Court	Shopping Mall	Pharmacy	Clothing Store	Bubble Tea Shop	Multiplex
12	263368.421053	Food Court	Diner	Sandwich Place	Fast Food Restaurant	Italian Restaurant	Coffee Shop	Pool	Recreation Center	Restaurant	Fruit & Vegetable Store
13	253139.742857	Japanese Restaurant	Bubble Tea Shop	Chinese Restaurant	Café	Bakery	Fast Food Restaurant	Supermarket	Sushi Restaurant	ATM	Light Rail Station
16	224294.800000	Coffee Shop	Asian Restaurant	Fast Food Restaurant	Chinese Restaurant	Shopping Mall	Bus Station	Sporting Goods Shop	Food	Supermarket	Bistro
17	248427.587097	Fast Food Restaurant	Coffee Shop	Sushi Restaurant	Food Court	Bakery	Bus Station	Sculpture Garden	Sandwich Place	Café	Candy Store
18	218608.500000	Coffee Shop	Chinese Restaurant	Korean Restaurant	Noodle House	Clothing Store	Ice Cream Shop	Asian Restaurant	Japanese Restaurant	Fast Food Restaurant	Multiplex
19	264170.000000	Bakery	Coffee Shop	Café	Gym	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Bubble Tea Shop	Fast Food Restaurant	Shopping Mall
21	221511.466667	Japanese Restaurant	Café	Indian Restaurant	Shopping Mall	Frozen Yogurt Shop	Food Court	Fast Food Restaurant	Electronics Store	Coffee Shop	Clothing Store
22	223834.506667	Food Court	Chinese Restaurant	Fried Chicken Joint	Coffee Shop	Supermarket	Fast Food Restaurant	Tea Room	Hainan Restaurant	Italian Restaurant	Shopping Mall

Cluster 3:

	resale_price	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	10th Most Common Venue
15	249547.293103	Chinese Restaurant	Asian Restaurant	Coffee Shop	Supermarket	Malay Restaurant	Gym	Hainan Restaurant	Park	Food Court	Train Station
20	204460.367816	Chinese Restaurant	Coffee Shop	Snack Place	Dessert Shop	Cosmetics Shop	Grocery Store	Monument / Landmark	Café	Frozen Yogurt Shop	Bubble Tea Shop

Results and Discussion

From the 4 clusters, we observe that cluster 2 consists of flats of relatively higher prices compared to the other clusters. Venue wise, all the clusters are populated with a variety of food options. Some venues like parks are only found in cluster 0 and cluster 3. Cluster 1 does not have any groceries or supermarkets while Cluster 0 does not have any train stations. Thus cluster 3 seems to be a suitable option for living, with transport, amenities like gym, park etc. while having lower prices. If flexibility in prices is allowed, cluster 2 is the better choice, with a better variety in venues. Based on our assumptions that the individual is looking for the cheapest options, cluster 3 is the better choice, with the basic amenities like transport, park, and supermarket.

Conclusion

We have analyzed the towns in Singapore to determine the optimal location for singles in Singapore to live in. This project can be further improved by considering other factors like the different types of flats (which we made an assumption), ethnic quota etc. Thank you for reading up to this point.