Title: Lowest Housing Sale Price Among Similiar Boroughs

1. Introduction/Business Problem

Singapore is a city-state with one of the most expensive housing market in the world. About 80% of Singapore residents stay in public housing , managed by the Housing Development Board (HDB). In year 2018, the median sale price of HDB flat is S$408,000. For someone looking to buy a new house in Singapore, this would mean careful research and planning is a must.
In this project, the objective to is help potential house owner make informed decision on the best Borough to purchase a HDB flat in, among other similar Boroughs. Best borough in this case would be the lowest priced Borough among the other similiar Boroughs.

2. Data

2.1 Definition:

1. Similiar Borough : Boroughs with simliar venues. As they have similiar amenities , this means the living conditions and lifestyles would be similar.
2. Lowest Housing Sale Price : Average price of all 2019 sale prices for each Borough

2.2 Data description:

1. Public housing flat resale price in 2019 from https://data.gov.sg/dataset/resale-flat-prices.
2. This raw data to be clean and average sale price computed for each Borough
3. Use geopy library to get the latitude and longitude values of each Borough
4. Forsquare API to get the most common venues of each Borough

3. Methodology

3.1 Raw Data:

The Public housing flat resale price in 2019 is downloaded from https://data.gov.sg/dataset/resale-flat-prices in json format.
It is transformed into Pandas DataFrame , and reduced features from 13 to 6 as follows:

```
The dataframe has 26 boroughs and 10396 records
```

	Borough	Street_Name	Resale_Price	Lease_Commence_Year	Floor_Area_Sqm	Flat_Type
0	PUNGGOL	PUNGGOL CTRL	470000	2014	95	4 ROOM
1	PUNGGOL	PUNGGOL WALK	590000	2015	113	5 ROOM
2	PUNGGOL	PUNGGOL WALK	555000	2015	113	5 ROOM
3	ANG MO KIO	ANG MO KIO AVE 1	270000	1981	68	3 ROOM
4	ANG MO KIO	ANG MO KIO AVE 1	295000	1976	73	3 ROOM

3.2 Data Processing:

The housing sale price is spread across different Flat_Type, therefore the average sale price of each Borough is computed and created under new dataframe:

	Borough	Resale_Price
0	ANG MO KIO	382347.010776
1	BEDOK	399831.014572
2	BISHAN	617004.078818
3	BUKIT BATOK	380307.621053
4	BUKIT MERAH	561061.817967

In order to retrieve the latitude and longitude of each borough in the dataframe, a geolocator function has to be created , which loop through each borough in the dataframe to get the location:

```
#define geolocator function:

def get_location(address):

    geolocator = Nominatim(user_agent="tt_explorer")
    location = geolocator.geocode(address)
    latitude_sg = location.latitude
    longitude_sg = location.longitude
    #print(address, latitude_sg, longitude_sg)
    return latitude_sg, longitude_sg
```

The objective of this project is to present the lowest price borough among similiar boroughs, where the similiarity between boroughs is defined as boroughs with simliar venues. As they have similiar amenities , this means the living conditions and lifestyles would be similar.
In order to get venues around each boroughs, Foursquare Explore API is used to retrieve update to 100 venues within 500metres of each borough:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ANG MO KIO	1.369842	103.846609	Old Chang Kee	1.369094	103.848389	Snack Place
1	ANG MO KIO	1.369842	103.846609	Subway	1.369136	103.847612	Sandwich Place
2	ANG MO KIO	1.369842	103.846609	Bun Master	1.369242	103.849031	Bakery
3	ANG MO KIO	1.369842	103.846609	MOS Burger	1.369170	103.847831	Burger Joint
4	ANG MO KIO	1.369842	103.846609	PLAYe	1.369109	103.848225	Hobby Shop

A total of 157 unique venue categories, among 26 borough are returned by Foursquare API.

3.3 Data Analysis:

Useful information to provide the potential house owner with is the top 5 venue for each borough. This information is computed by grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each category, from the Foursquare API data:

Top 5 Venues for each borough

```
----ANG MO KIO----
                  venue  freq
0           Coffee Shop  0.11
1            Food Court  0.06
2  Fast Food Restaurant  0.06
3       Bubble Tea Shop  0.06
4           Supermarket  0.04
```

```
----BEDOK----
                  venue  freq
0           Coffee Shop  0.07
1  Japanese Restaurant  0.05
2            Food Court  0.05
3        Sandwich Place  0.05
4     Asian Restaurant  0.05
```

```
----BISHAN----
                 venue  freq
0         Coffee Shop  0.10
1     Bubble Tea Shop  0.07
2          Food Court  0.07
3                Café  0.05
4  Chinese Restaurant  0.05
```

```
----BUKIT BATOK----
                 venue  freq
0         Coffee Shop  0.19
1  Chinese Restaurant  0.11
2          Food Court  0.07
3  Frozen Yogurt Shop  0.04
4         Bus Station  0.04
```

```
----BUKIT MERAH----
                 venue  freq
0    Asian Restaurant  0.13
1                Café  0.10
2  Chinese Restaurant  0.08
3         Coffee Shop  0.05
4           Bookstore  0.05
```

```
----BUKIT PANJANG----
                  venue  freq
0           Food Court  0.25
1                 Park  0.25
2  American Restaurant  0.12
3   Miscellaneous Shop  0.12
4               Market  0.12
```

```
----BUKIT TIMAH----
             venue  freq
0            Trail  0.58
1         Mountain  0.17
2        Rest Area  0.17
3  Scenic Lookout  0.08
4      Yoga Studio  0.00
```

```
----CENTRAL AREA----
                 venue  freq
0       Cocktail Bar  0.05
1        Event Space  0.04
2  French Restaurant  0.04
3       Concert Hall  0.04
4              Hotel  0.04
```

```
----CHOA CHU KANG----
                  venue  freq
0  Fast Food Restaurant  0.25
1           Coffee Shop  0.08
2            Food Court  0.08
3        Sandwich Place  0.08
4       Thai Restaurant  0.08
```

```
----CLEMENTI----
                  venue  freq
0           Food Court  0.08
1    Asian Restaurant  0.06
2  Japanese Restaurant  0.06
3         Noodle House  0.05
4          Coffee Shop  0.05
```

```
----GEYLANG----
                         venue  freq
0             Chinese Restaurant  0.19
1                   Noodle House  0.11
2                     Food Court  0.11
3  Vegetarian / Vegan Restaurant  0.08
4               Asian Restaurant  0.08
```

```
----HOUGANG----
          venue  freq
0   Food Court  0.12
1          Pool  0.04
2          Park  0.04
3     Bus Stop  0.04
4  Bus Station  0.04
```

```
----JURONG EAST----
                 venue  freq
0   Chinese Restaurant  0.08
1  Japanese Restaurant  0.07
2          Coffee Shop  0.06
3           Food Court  0.06
4      Bubble Tea Shop  0.04
```

```
----JURONG WEST----
                  venue  freq
0   Japanese Restaurant  0.13
1  Fast Food Restaurant  0.11
2      Asian Restaurant  0.10
3           Dessert Shop  0.06
4    Chinese Restaurant  0.06
```

Top 5 Venues for each borough

```
----KALLANG/WHAMPOA----
                venue  freq
0  Chinese Restaurant  0.20
1         Coffee Shop  0.15
2  Dim Sum Restaurant  0.10
3            Bus Line  0.10
4                Park  0.05
```

```
----MARINE PARADE----
                 venue  freq
0                Hotel  0.11
1            Multiplex  0.08
2       Massage Studio  0.08
3  Japanese Restaurant  0.08
4   Chinese Restaurant  0.05
```

```
----PASIR RIS----
                venue  freq
0         Food Court  0.06
1    Asian Restaurant  0.06
2     Thai Restaurant  0.06
3  Chinese Restaurant  0.06
4         Coffee Shop  0.06
```

```
----PUNGGOL----
                venue  freq
0         Bus Station  0.33
1  Convenience Store  0.17
2  Chinese Restaurant  0.17
3            Bus Stop  0.17
4         High School  0.17
```

```
----QUEENSTOWN----
                    venue  freq
0       Chinese Restaurant  0.19
1             Noodle House  0.12
2               Food Court  0.12
3  Health & Beauty Service  0.06
4             Dessert Shop  0.06
```

```
----SEMBAWANG----
                  venue  freq
0           Coffee Shop  0.18
1  Fast Food Restaurant  0.12
2      Asian Restaurant  0.12
3               Bistro  0.06
4          Bus Station  0.06
```

```
----SENGKANG----
                  venue  freq
0  Fast Food Restaurant  0.10
1           Food Court  0.07
2           Restaurant  0.07
3          Coffee Shop  0.07
4      Asian Restaurant  0.07
```

```
----SERANGOON----
         venue  freq
0   Playground  0.25
1  Comedy Club  0.25
2     Bus Line  0.25
3         Café  0.25
4  Yoga Studio  0.00
```

```
----TAMPINES----
                venue  freq
0  Japanese Restaurant  0.06
1         Coffee Shop  0.05
2              Bakery  0.05
3                Café  0.05
4      Bubble Tea Shop  0.04
```

```
----TOA PAYOH----
                venue  freq
0        Noodle House  0.11
1         Coffee Shop  0.11
2  Chinese Restaurant  0.09
3          Food Court  0.09
4         Snack Place  0.09
```

```
----WOODLANDS----
                venue  freq
0  Japanese Restaurant  0.08
1         Pizza Place  0.06
2                Café  0.06
3   Chinese Restaurant  0.06
4           Food Court  0.04
```

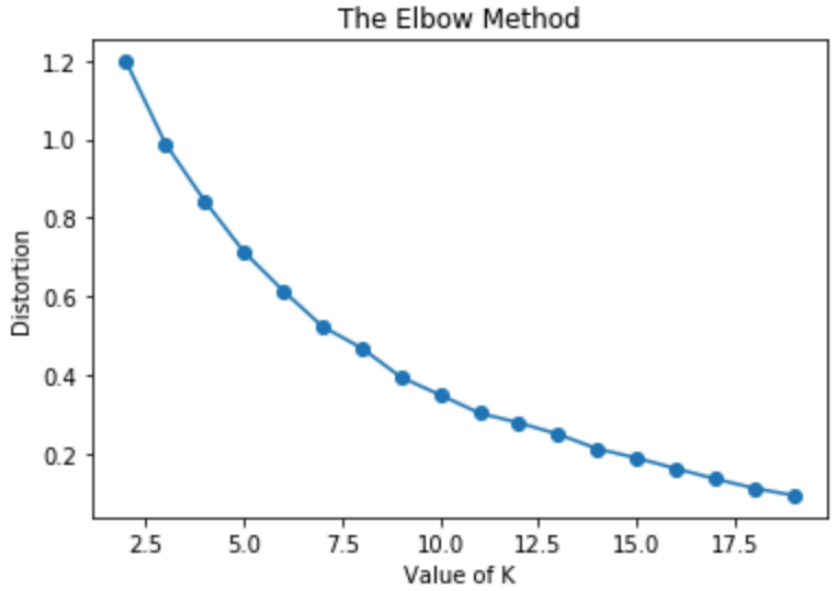
```
----YISHUN----
                venue  freq
0           Coffee Shop  0.11
1            Food Court  0.09
2    Chinese Restaurant  0.07
3          Noodle House  0.04
4  Fast Food Restaurant  0.04
```

Top 10 venues for each neighborhood can also be computed and provided to potential house owners for details. This is done using dataframe with sorted venue for each borough:

	Neighborhood	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	Food Court	Bubble Tea Shop	Fast Food Restaurant	Sushi Restaurant	Supermarket	Dessert Shop	Sandwich Place	Japanese Restaurant	Malay Restaurant
1	BEDOK	Coffee Shop	Food Court	Asian Restaurant	Japanese Restaurant	Sandwich Place	Chinese Restaurant	Fast Food Restaurant	Supermarket	Sushi Restaurant	Café
2	BISHAN	Coffee Shop	Bubble Tea Shop	Food Court	Cosmetics Shop	Café	Japanese Restaurant	Chinese Restaurant	Ice Cream Shop	Supermarket	Electronics Store
3	BUKIT BATOK	Coffee Shop	Chinese Restaurant	Food Court	Shopping Mall	Frozen Yogurt Shop	Department Store	Multiplex	Fast Food Restaurant	Café	Mobile Phone Shop
4	BUKIT MERAH	Asian Restaurant	Café	Chinese Restaurant	Noodle House	Bookstore	Coffee Shop	Yoga Studio	Bus Station	Boutique	Breakfast Spot

3.4 Data Mining:

K-means clustering will be used to group boroughs with similiar venues together before the minimum sale price among these groups can be determine.
K-means clustering is one of the most popular unsupervised machine learning algorithmn well suited for this use-case as the objective is to group similiar objects together.
One parameter to decide in K-means clustering is the number of clusters to use. For this purpose, the Elbow Curve is used. From the graph below, 8 clusters would be optimal number as distortion decreases slowly from K= 8 onwards.



The Cluster Number of each Neighorhood (Borough) is obtained after performing K-means clustering and the cluster label being added to the dataframe:

	Neighborhood	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	Cluster_Number
0	ANG MO KIO	Coffee Shop	Bubble Tea Shop	Food Court	Sushi Restaurant	Fast Food Restaurant	Snack Place	Supermarket	Dessert Shop	Sandwich Place	Japanese Restaurant	2
1	BEDOK	Coffee Shop	Food Court	Asian Restaurant	Sandwich Place	Japanese Restaurant	Café	Sushi Restaurant	Fast Food Restaurant	Supermarket	Chinese Restaurant	2
2	BISHAN	Coffee Shop	Food Court	Bubble Tea Shop	Supermarket	Chinese Restaurant	Café	Japanese Restaurant	Ice Cream Shop	Cosmetics Shop	Pet Store	2
3	BUKIT BATOK	Coffee Shop	Food Court	Chinese Restaurant	Fast Food Restaurant	Pharmacy	Department Store	Sandwich Place	Toy / Game Store	Asian Restaurant	Bus Station	0
4	BUKIT MERAH	Asian Restaurant	Café	Chinese Restaurant	Coffee Shop	Noodle House	Bookstore	Food Court	Flea Market	Japanese Restaurant	Hainan Restaurant	2

Now that we have Cluster Number information of each Neighborhood, we need to merge it together with Sale Price. This is achieved by merging the main dataframe with average priced dataframe created in section 3.2 above:

	Neighborhood	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	Cluster_Number	Resale_Price
0	ANG MO KIO	Coffee Shop	Bubble Tea Shop	Food Court	Sushi Restaurant	Fast Food Restaurant	Snack Place	Supermarket	Dessert Shop	Sandwich Place	Japanese Restaurant	2	382347.010776
1	BEDOK	Coffee Shop	Food Court	Asian Restaurant	Sandwich Place	Japanese Restaurant	Café	Sushi Restaurant	Fast Food Restaurant	Supermarket	Chinese Restaurant	2	399831.014572
2	BISHAN	Coffee Shop	Food Court	Bubble Tea Shop	Supermarket	Chinese Restaurant	Café	Japanese Restaurant	Ice Cream Shop	Cosmetics Shop	Pet Store	2	617004.078818
3	BUKIT BATOK	Coffee Shop	Food Court	Chinese Restaurant	Fast Food Restaurant	Pharmacy	Department Store	Sandwich Place	Toy / Game Store	Asian Restaurant	Bus Station	0	380307.621053
4	BUKIT MERAH	Asian Restaurant	Café	Chinese Restaurant	Coffee Shop	Noodle House	Bookstore	Food Court	Flea Market	Japanese Restaurant	Hainan Restaurant	2	561061.817967

In order to easily retrieve the lowest sale price Neighbourhood within each cluster, the dataframe is sorted by Cluster Number, follow by Sale Price:

	Neighborhood	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	Cluster_Number	Resale_Price
19	SEMBAWANG	Coffee Shop	Asian Restaurant	Wings Joint	Chinese Restaurant	Bus Station	Bistro	Fast Food Restaurant	Café	Supermarket	Shopping Mall	0	365279.450980
8	CHOA CHU KANG	Fast Food Restaurant	Playground	Asian Restaurant	Coffee Shop	Shoe Store	Sandwich Place	Thai Restaurant	Food Court	Cosmetics Shop	Park	0	378231.267123
3	BUKIT BATOK	Coffee Shop	Food Court	Chinese Restaurant	Fast Food Restaurant	Pharmacy	Department Store	Sandwich Place	Toy / Game Store	Asian Restaurant	Bus Station	0	380307.621053
5	BUKIT PANJANG	Food Court	Park	Market	Noodle House	Miscellaneous Shop	Salon / Barbershop	Electronics Store	Food & Drink Shop	Food	Flea Market	1	429277.918854
24	WOODLANDS	Japanese Restaurant	Chinese Restaurant	Café	Coffee Shop	Asian Restaurant	Fast Food Restaurant	Clothing Store	Basketball Court	Shopping Mall	Electronics Store	2	367538.507446

4. Results

4.1 Result Visualization

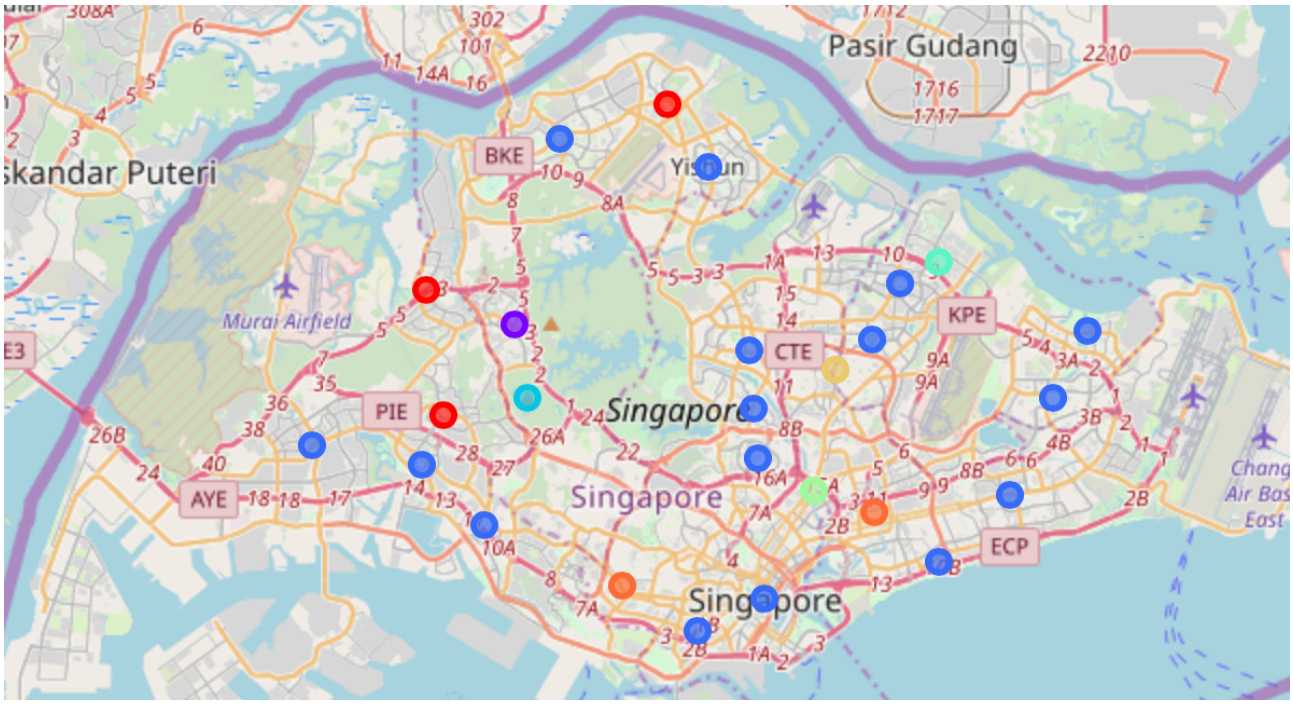
To visualize the clusters in a map, the latitude and longitude information of each Neighbourhood will be needed to be merge with the main dataframe.
To achieve this, a copy of dataframe with Neighbourhood and latitude and longtitude information is created:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude
0	ANG MO KIO	1.369842	103.846609
48	BEDOK	1.323976	103.930216
105	BISHAN	1.351452	103.848250
144	BUKIT BATOK	1.349057	103.749591
168	BUKIT MERAH	1.280628	103.830591

The above dataframe is then merged with the main dataframe to create the master dataframe to be used for visualization in map:

	Neighborhood	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	Cluster_Number	Resale_Price	Neighborhood Latitude	Neighborhood Longitude
0	SEMBAWANG	Coffee Shop	Asian Restaurant	Wings Joint	Chinese Restaurant	Bus Station	Bistro	Fast Food Restaurant	Café	Supermarket	Shopping Mall	0	365279.450980	1.448065	103.820760
1	CHOA CHU KANG	Fast Food Restaurant	Playground	Asian Restaurant	Coffee Shop	Shoe Store	Sandwich Place	Thai Restaurant	Food Court	Cosmetics Shop	Park	0	378231.267123	1.389260	103.743728
2	BUKIT BATOK	Coffee Shop	Food Court	Chinese Restaurant	Fast Food Restaurant	Pharmacy	Department Store	Sandwich Place	Toy / Game Store	Asian Restaurant	Bus Station	0	380307.621053	1.349057	103.749591
3	BUKIT PANJANG	Food Court	Park	Market	Noodle House	Miscellaneous Shop	Salon / Barbershop	Electronics Store	Food & Drink Shop	Food	Flea Market	1	429277.918854	1.377921	103.771866
4	WOODLANDS	Japanese Restaurant	Chinese Restaurant	Café	Coffee Shop	Asian Restaurant	Fast Food Restaurant	Clothing Store	Basketball Court	Shopping Mall	Electronics Store	2	367538.507446	1.436897	103.786216

To create the map visualization, Folium map rendering library is used, and the result is shown as follows:



4.1 Detail Result

Besides map visualization, the straightforward answer to provide potential house buyer is the information on the lowest-priced borough for each cluster retrieved by filtering the first borough of each cluster in the master dataframe. This will get the lowest-priced borough as the master dataframe is already sort by cluster number, follow by sale price in section 3.3 above.

For each cluster, the lowest-priced borough and its venues (amenities) is displayed for potential buyer to know the type of living conditions and lifestyles of each cluster.

Lowest Housing Sale Price Borough in Cluster 1

Neighborhood	SEMBAWANG
1st Most Common Venue	Coffee Shop
2nd Most Common Venue	Asian Restaurant
3rd Most Common Venue	Wings Joint
4th Most Common Venue	Chinese Restaurant
5th Most Common Venue	Bus Station
6th Most Common Venue	Bistro
7th Most Common Venue	Fast Food Restaurant
8th Most Common Venue	Café
9th Most Common Venue	Supermarket
10th Most Common Venue	Shopping Mall
Cluster_Number	0
Resale_Price	365279
Neighborhood Latitude	1.44806
Neighborhood Longitude	103.821

Lowest Housing Sale Price Borough in Cluster 2

Neighborhood	BUKIT PANJANG
1st Most Common Venue	Food Court
2nd Most Common Venue	Park
3rd Most Common Venue	Market
4th Most Common Venue	Noodle House
5th Most Common Venue	Miscellaneous Shop
6th Most Common Venue	Salon / Barbershop
7th Most Common Venue	Electronics Store
8th Most Common Venue	Food & Drink Shop
9th Most Common Venue	Food
10th Most Common Venue	Flea Market
Cluster_Number	1
Resale_Price	429278
Neighborhood Latitude	1.37792
Neighborhood Longitude	103.772

Lowest Housing Sale Price Borough in Cluster 3

Neighborhood	WOODLANDS
1st Most Common Venue	Japanese Restaurant
2nd Most Common Venue	Chinese Restaurant
3rd Most Common Venue	Café
4th Most Common Venue	Coffee Shop
5th Most Common Venue	Asian Restaurant
6th Most Common Venue	Fast Food Restaurant
7th Most Common Venue	Clothing Store
8th Most Common Venue	Basketball Court
9th Most Common Venue	Shopping Mall
10th Most Common Venue	Electronics Store
Cluster_Number	2
Resale_Price	367539
Neighborhood Latitude	1.4369
Neighborhood Longitude	103.786

Lowest Housing Sale Price Borough in Cluster 4

Neighborhood	BUKIT TIMAH
1st Most Common Venue	Trail
2nd Most Common Venue	Rest Area
3rd Most Common Venue	Mountain
4th Most Common Venue	Scenic Lookout
5th Most Common Venue	Dim Sum Restaurant
6th Most Common Venue	Flea Market
7th Most Common Venue	Fast Food Restaurant
8th Most Common Venue	Event Space
9th Most Common Venue	Electronics Store
10th Most Common Venue	Dumpling Restaurant
Cluster_Number	3
Resale_Price	713304
Neighborhood Latitude	1.35469
Neighborhood Longitude	103.776

Lowest Housing Sale Price Borough in Cluster 5

Neighborhood	PUNGGOL
1st Most Common Venue	Food Court
2nd Most Common Venue	Movie Theater
3rd Most Common Venue	Chinese Restaurant
4th Most Common Venue	High School
5th Most Common Venue	Bus Station
6th Most Common Venue	Electronics Store
7th Most Common Venue	Food & Drink Shop
8th Most Common Venue	Food
9th Most Common Venue	Flea Market
10th Most Common Venue	Fast Food Restaurant
Cluster_Number	4
Resale_Price	463184
Neighborhood Latitude	1.39803
Neighborhood Longitude	103.907

Lowest Housing Sale Price Borough in Cluster 6

Neighborhood	KALLANG/WHAMPOA
1st Most Common Venue	Chinese Restaurant
2nd Most Common Venue	Coffee Shop
3rd Most Common Venue	Pet Store
4th Most Common Venue	Restaurant
5th Most Common Venue	Food Court
6th Most Common Venue	Hotel Pool
7th Most Common Venue	Juice Bar
8th Most Common Venue	Dance Studio
9th Most Common Venue	Gym / Fitness Center
10th Most Common Venue	Bus Line
Cluster_Number	5
Resale_Price	494525
Neighborhood Latitude	1.32511
Neighborhood Longitude	103.867

Lowest Housing Sale Price Borough in Cluster 7

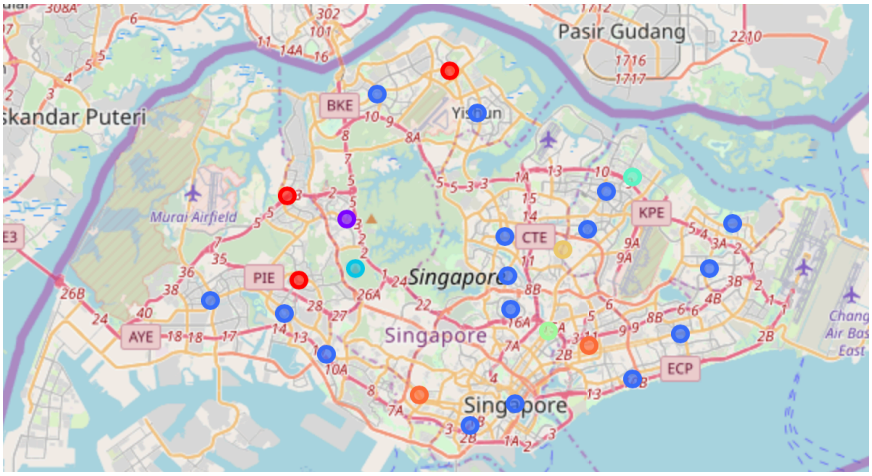
Neighborhood	SERANGOON
1st Most Common Venue	Playground
2nd Most Common Venue	Café
3rd Most Common Venue	Pet Store
4th Most Common Venue	Bus Line
5th Most Common Venue	Bus Station
6th Most Common Venue	Basketball Court
7th Most Common Venue	Electronics Store
8th Most Common Venue	Food
9th Most Common Venue	Flea Market
10th Most Common Venue	Fast Food Restaurant
Cluster_Number	6
Resale_Price	480534
Neighborhood Latitude	1.36324
Neighborhood Longitude	103.874

Lowest Housing Sale Price Borough in Cluster 8

Neighborhood	GEYLANG
1st Most Common Venue	Chinese Restaurant
2nd Most Common Venue	Food Court
3rd Most Common Venue	Noodle House
4th Most Common Venue	Vegetarian / Vegan Restaurant
5th Most Common Venue	Dim Sum Restaurant
6th Most Common Venue	Asian Restaurant
7th Most Common Venue	Grocery Store
8th Most Common Venue	Dessert Shop
9th Most Common Venue	Train Station
10th Most Common Venue	Seafood Restaurant
Cluster_Number	7
Resale_Price	429117
Neighborhood Latitude	1.31819
Neighborhood Longitude	103.887

5. Discussion

As Singapore is a tiny city-state, only 26 boroughs are being used for this project. With the small number of boroughs, K-means clustering returns an uneven distribution among the boroughs as can be seen from the visualization below as well as computed distribution of clusters below.




```
#check the distribution of clusters
unique, counts = np.unique(labels, return_counts=True)

print (np.asarray((unique, counts)).T)
```

```
[[ 0  3]
 [ 1  1]
 [ 2 16]
 [ 3  1]
 [ 4  1]
 [ 5  1]
 [ 6  1]
 [ 7  2]]
```

The distribution shows that cluster index 2 contains more than 60% of the boroughs, while 5 clusters have only 1 borough in them.
Such results may not give true picture of the right borough to choose for potential house buyer as the result is skewed.

Improvement recommendation would be not to use borough as clustering object due to the small size of Singapore which resulted in small number of boroughs for even distribution among clusters. Alternative to borough would be street name, which will increase the number of clustering objects for better distribution among clusters.

6. Conclusion

The objective of this project is to help potential house owner make informed decision on the lowest-priced Borough to purchase a HDB flat in, among other similar Boroughs. Similiarity among Boroughs is determined by the similiar venues.
The project resulted in 8 clusters from which potential buyers can choose from, which has been mapped to a Singapore map for better visiualization to buyer.
The detail list of lowest-priced borough of each cluster is also displayed.
Price may not be the only factor when potential buyer is looking for a home. Therefore, in the result section, the lowest-priced borough and its venues (amenities) is displayed for potential buyer to know the type of living conditions and lifestyles of each cluster.
With this additional information, potential buyer will be able to make more informed decision on where to buy the house.