COURSERA CAPSTONE PROJECT: THE BATTLE OF NEIGHBORHOODS

Analysis of Singapore's Towns – A Relocation Guide

A. Introduction

Singapore is one of the preferred destinations to do business in Asia, with its economy ranked as the world's most competitive economy, based on the latest 2019 World Economic Forum Global Competitiveness Report. Favorable factors contributing to Singapore's economic performance include an open economy, strong labor-employer relations, diverse cosmopolitan workforce, as well as government stability and responsiveness to change.

Singapore also offers the best quality of life in Asia, based on Mercer's 2019 Quality of Living Survey, which considers factors like political stability, healthcare, education, crime, recreation and transport. People have been attracted to this cosmopolitan island state, due to its vibrant economy, low personal income taxes, cultural diversity and high quality of living. Presently, the immigrant population in Singapore number 2.16 million, and makes up ~40% of the total population of ~5.7 million people.

However, cost of living is a concern, with Singapore being rated as the world's most expensive city by the Economist Intelligence Unit's 2018 Worldwide Cost of Living Report. Furthermore, Singapore has a large population size for its size, with ~8,000 people per km^2. This makes Singapore 230 times denser than the United States, and more than 2,500 times denser than Australia.

As such, the goal of this analysis is to identify the most livable neighborhoods in Singapore for individuals looking to relocate to Singapore and those considering moving within Singapore. For the purpose of this exercise, we will define the most livable neighborhoods as having: (i) an affordable median rental price, (ii) a tolerable population density, and (iii) a balanced mix of amenities in the neighborhood.

B. Data Description

This analysis will require the use of the following data sources:

Singapore Median Rent by Town and Flat Type

Data on Singapore towns and corresponding median rental prices by town and flat type will be retrieved from Data.gov.sg (https://data.gov.sg), the government's one-stop access portal to publicly available datasets. Since Median Rent by Town and Flat Type data covers information from April 1, 2005 to December 31, 2019 on a quarterly basis, we will be using 2019-Q4 data for this analysis, as this is the most recent dataset. To simplify the analysis, the average rental price for each town will be determined by the median rental price for 4-room flat types in that town, as it is available as a benchmark across almost all towns.

Singapore Population Density by Town

Data on Singapore's population density by town will be obtained by scraping data from the Wikipedia page on 'Planning Areas of Singapore' (https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore), which contains data on town name, region, area (km2), population and density (/km2). Population density (people per km2) is a measure of the degree of 'crowding' of the town and is calculated by dividing the town's population by total area of town.

Singapore Town Location Data

Singapore's geospatial data will be retrieved from Data.gov.sg (https://data.gov.sg). Master Plan 2019 Planning Area Boundary (No Sea) data provides indicative polygons of planning area boundary, and this GeoJSON data on Singapore's planning areas will enable visualization on maps. In parallel, geographic coordinates of town centers will be retrieved using Google Maps, with coordinates of MRT stations being used as the center for all towns for the purpose of this analysis.

Singapore Venue Information from Foursquare API

Foursquare API (https://foursquare.com/) will be used to explore the neighborhoods of each town. Using Foursquare API, we will understand the various venues in each neighborhood, to assess if there are a balanced mix of amenities and to determine the most common venue categories. Ten key venue categories were selected for determining whether the neighborhood had a balanced mix of amenities: Coffee Shop, Food Court, Fast Food Restaurant, Café, Shopping Mall, Supermarket, Clothing Store, Bookstore, Convenience Store, Gym.

C. Methodology

Context & Location Orientation

To provide geographical context on the various planning areas in Singapore, a reference map for the planning areas in Singapore *(Figure 1)* was obtained from the Wikipedia page on 'Planning Areas of Singapore' (https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore).

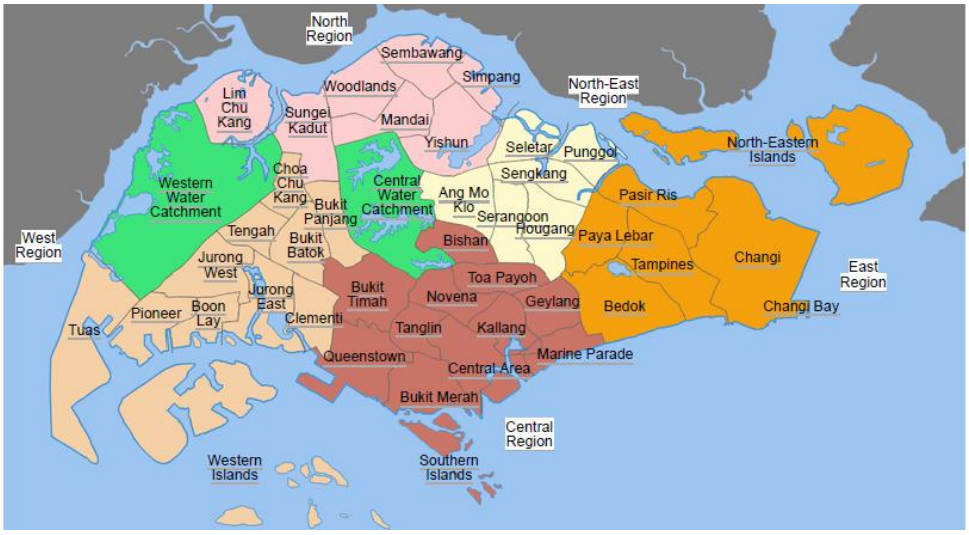

Figure 1: Reference Map for Planning Areas in Singapore

Exploratory Analysis

1. Median Rent

Of the 25 towns in Singapore, the Top 5 towns with the highest median rental prices are Queenstown, Outram, Bukit Merah, Kallang and Clementi with prices above SGD 2,400/month *(Figure 2)*. Many of these high rental towns are within, or near the Central Region. Proximity to the Central Region's attractions and commercial areas could be a possible reason for the higher rental, but in assessing livable neighborhoods in Singapore, this is unlikely to be a key consideration based on how we have defined the problem.

On the contrary, the Bottom 5 towns with the lowest median rental prices are Bukit Panjang, Woodlands, Yishun, Sembawang and Choa Chu Kang with prices least than or equal to SGD 1,800/month *(Figure 2)*. Many of these low rental towns are within, or near to the North Region.

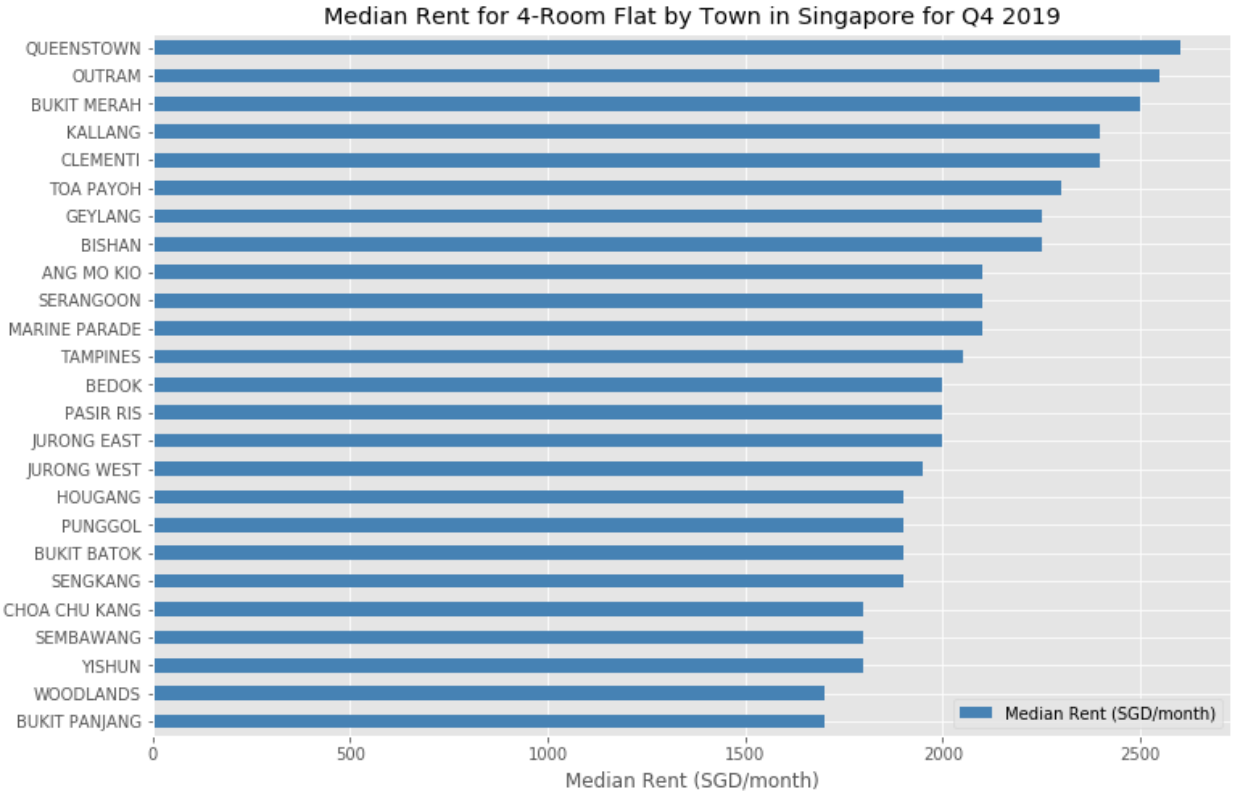

Figure 2: Median Rent for 4-Room Flat Type by Town in Singapore for Q4 2019

The distribution of median rental prices tends to be more right-skewed, with majority of towns (17 of 25 towns; 64% of towns) having median rental prices of SGD 2,100/month or below *(Figure 3)*.

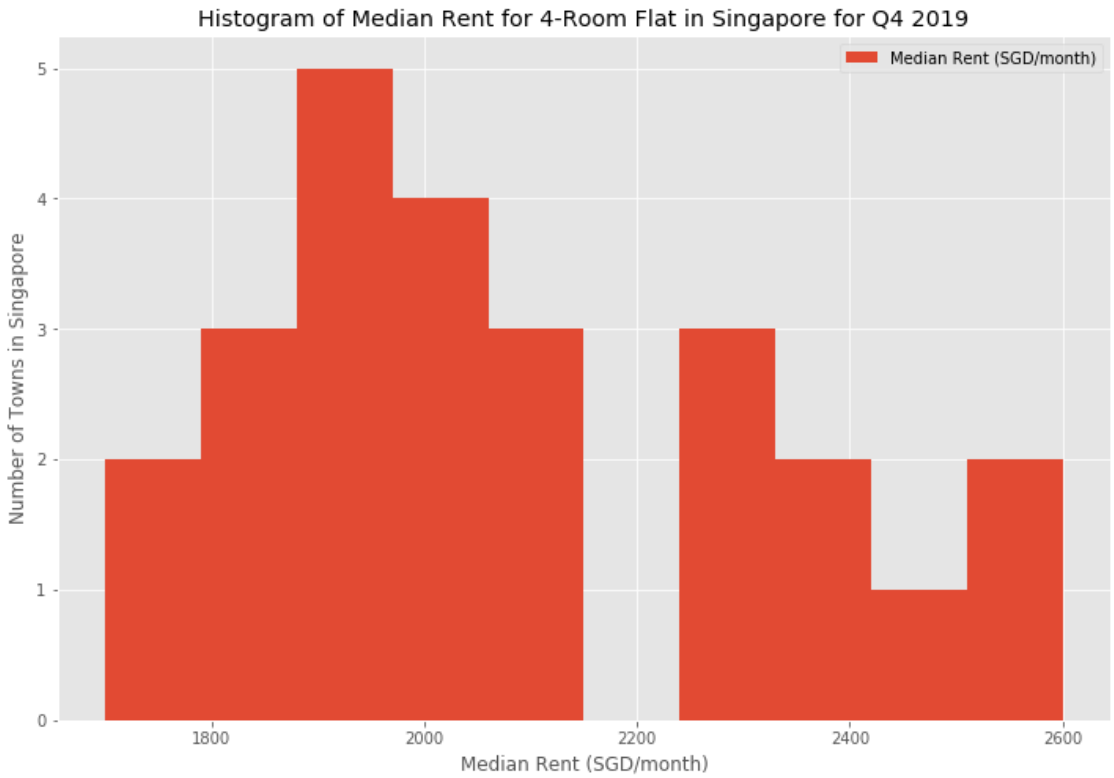

Figure 3: Histogram of Median Rent for 4-Room Flat Type by Town in Singapore for Q4 2019

2. Population Density

Of the 55 planning areas / towns in Singapore, the Top 5 towns with the highest population density are Choa Chu Kang and Sengkang by a clear margin with population density of 30k people/km^2 and 23k people/km^2, followed by Woodlands, Jurong West and Punggol *(Figure 4)*. Coincidentally, Choa Chu Kang and Sengkang are also amongst the lowest median rental towns in Singapore. High population density is likely a predictor for lower median rental prices. The distribution of population tends to be highly right-skewed, with a large majority of towns have less than 15k people/km^2 *(Figure 5)*.

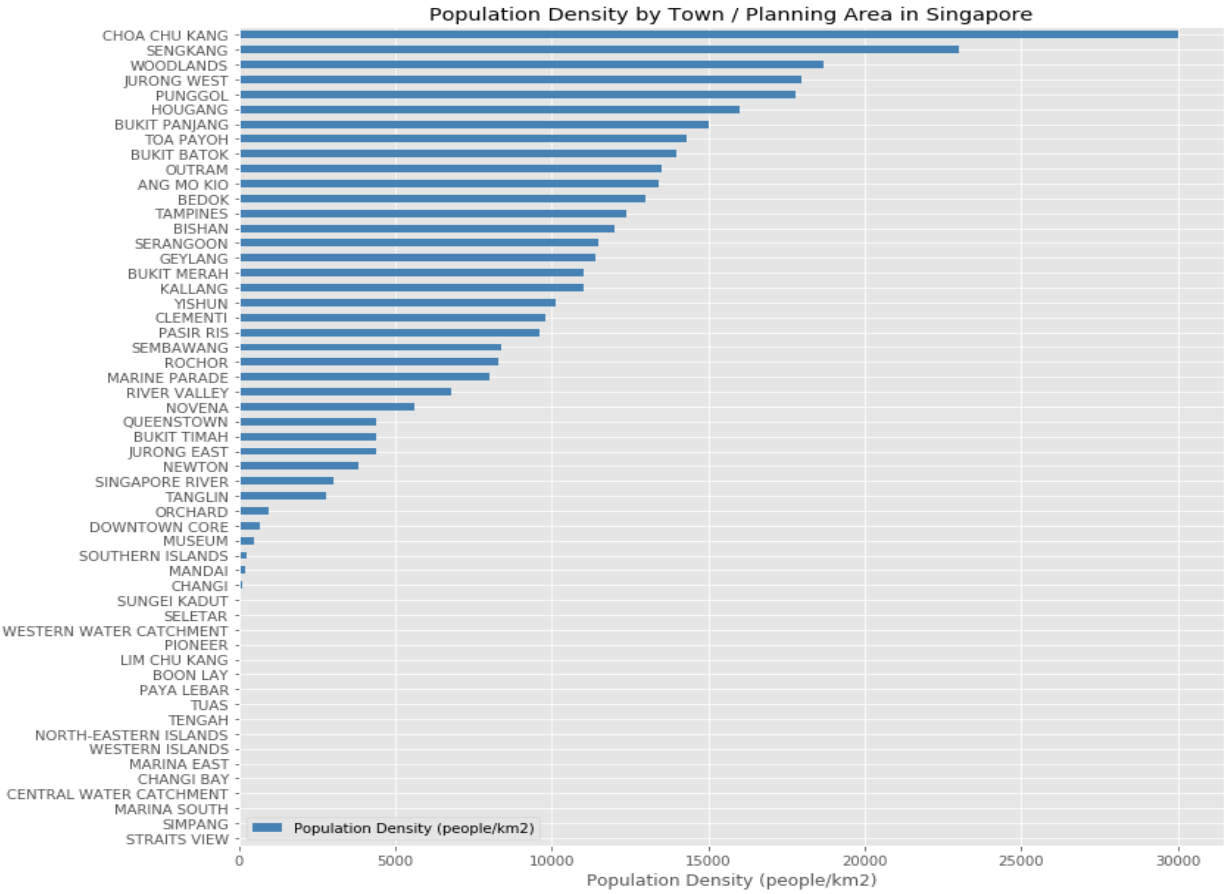

Figure 4: Population Density by Town / Planning Area in Singapore

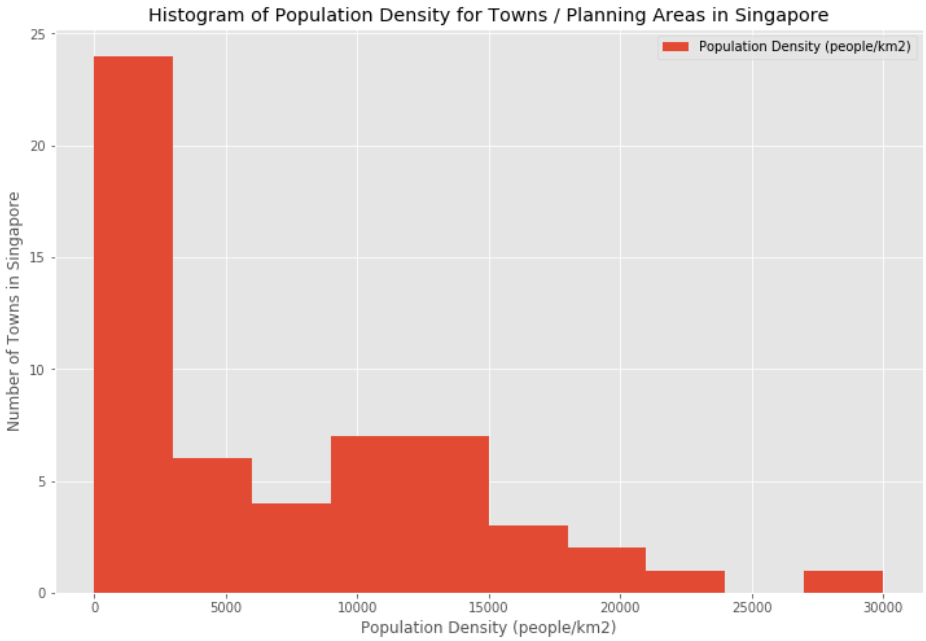

Figure 5: Histogram of Population Density by Town / Planning Area in Singapore

3. Relationship Between Median Rent and Population Density

There is a possibility that high population density is a predictor for lower median rental prices, as we have seen from the examples of Choa Chu Kang and Sengkang. However, it appears that population density only provides some predictive value for median rent, with a Pearson correlation coefficient (r) of -0.450, and this relationship tends to be more evident for higher population densities *(Figure 6).* This reveals that other factors likely contribute to median rental prices, possibly location and proximity to key amenities, attractions, commercial areas, etc.

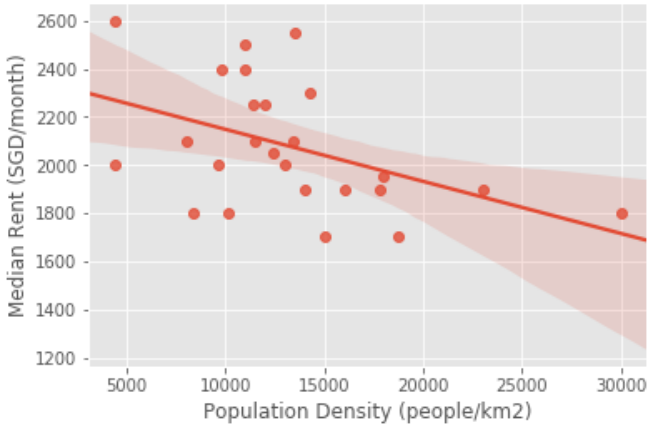

Figure 6: Scatter Plot for Median Rent vs. Population Density

4. Visualization of Towns on Map of Singapore

Using Folium, a visualization of a map of Singapore with town centers superimposed on top was created to facilitate geolocation analysis *(Figure 7)*.

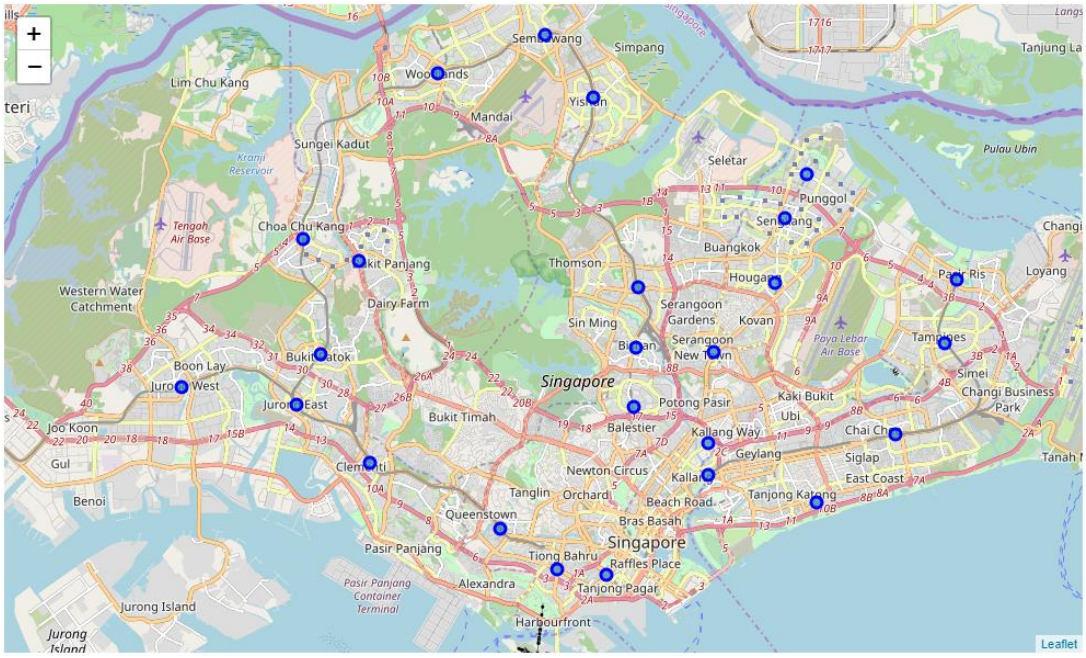

Figure 7: Map of Singapore with Towns

To enable the identification of high population density areas visually, a choropleth map was generated *(Figure 8)*. As we can see, Choa Chu Kang has the highest population density at 30k people/km^2, and is shaded dark red on the choropleth map. Other high population density towns, such as Sengkang, Woodlands and Jurong West can also be readily identified, with dark orange shading.

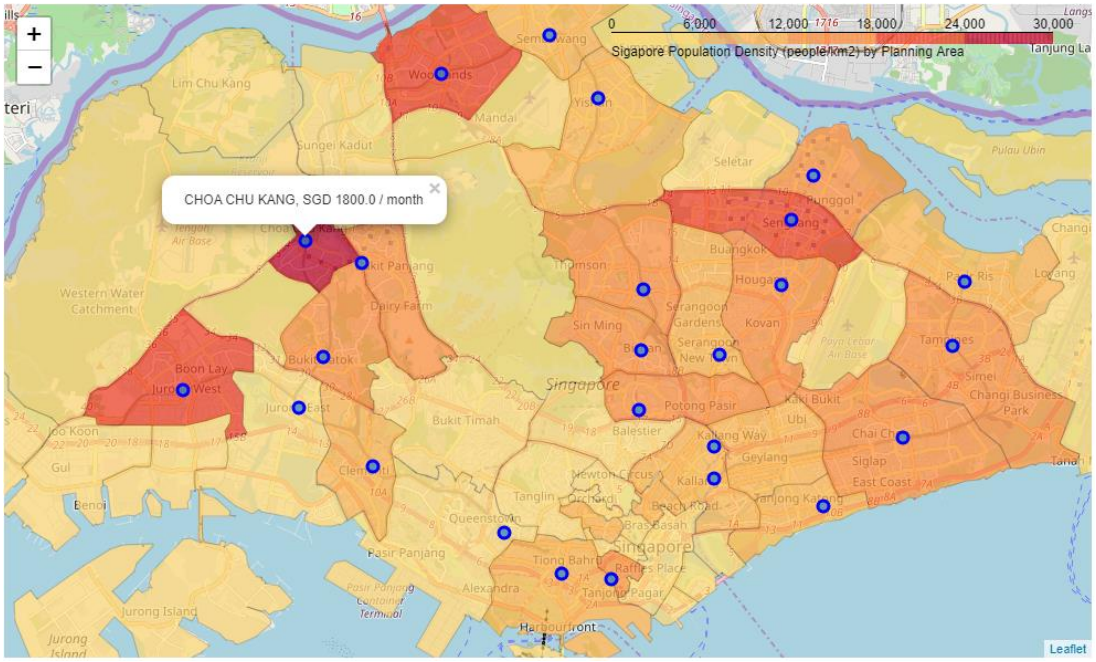

Figure 8: Choropleth Map of Singapore based on Population Density

5. Assessment of Balanced Mix of Amenities

Using Foursquare API, the closest 100 venues within a 500m radius from each town center across all towns were queried *(Figure 9)*. Based on this query, 163 unique venue categories were identified.

	Town	Town Latitude	Town Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ANG MO KIO	1.370017	103.84945	FairPrice Xtra	1.369279	103.848886	Supermarket
1	ANG MO KIO	1.370017	103.84945	Old Chang Kee	1.369094	103.848389	Snack Place
2	ANG MO KIO	1.370017	103.84945	Subway	1.369136	103.847612	Sandwich Place
3	ANG MO KIO	1.370017	103.84945	MOS Burger	1.369170	103.847831	Burger Joint
4	ANG MO KIO	1.370017	103.84945	NTUC FairPrice	1.371507	103.847082	Supermarket

Figure 9: Dataframe Showing Venue Details by Town based on Foursquare API Query

To enable the identification of what would best constitute a balanced mix of amenities, it was critical to examine the composition and frequency for the Top 40 venue categories across all towns *(Figure 10)*.

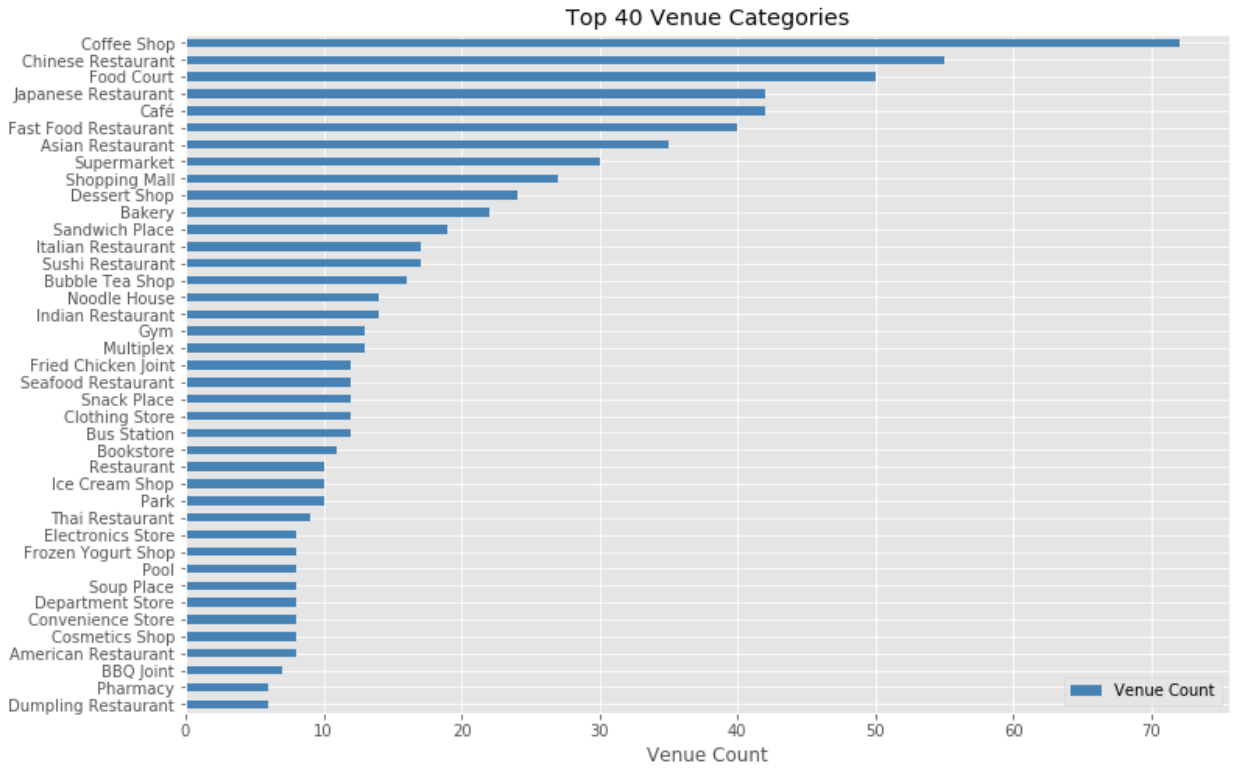

Figure 10: Top 40 Venue Categories Across all Towns in Singapore

Unsurprisingly, a largely majority (21 of 30; 70% of Top 30) of these Top 30 venue categories were related to food. However, food alone will not provide a balanced mix of amenities. Accordingly, ten key venue categories were selected from the Top 30, to assess if there are a balanced mix of amenities in each town *(Figure 11)*. These key venue categories had the minimally acceptable food options *(i.e. Coffee Shop, Food Court, Fast Food Restaurant, Café)*, essential retail options *(i.e. Shopping Mall, Supermarket, Clothing Store, Bookstore, Convenience Store)*, and a fitness option *(i.e. Gym)*.

	Town	Coffee Shop	Food Court	Fast Food Restaurant	Café	Shopping Mall	Supermarket	Clothing Store	Bookstore	Convenience Store	Gym	Total Amenities
0	ANG MO KIO	1	1	1	0	1	1	0	0	1	1	7
1	BEDOK	1	1	1	1	1	1	1	1	0	0	8
2	BISHAN	1	1	0	1	1	1	0	1	0	1	7
3	BUKIT BATOK	1	1	1	1	1	0	0	0	1	0	6
4	BUKIT MERAH	1	0	0	0	0	0	0	0	1	0	2

Figure 11: Dataframe Showing Key Venue Categories to Assess Balanced Mix of Amenities in each Town

To better assess the degree to which each of these towns had a balanced mix of amenities, the count of key venue categories selected was plotted for each town *(Figure 12)*. Of the 25 towns in Singapore, only 18 towns (72%) have 6 or more key venue categories that qualify as a balanced mix of amenities.

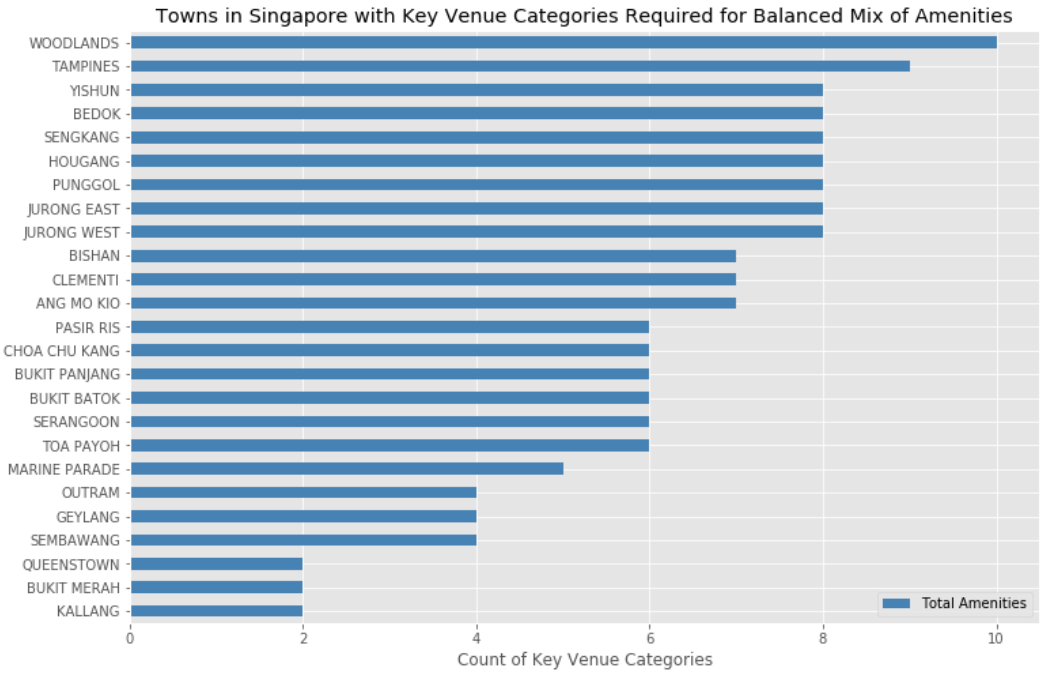

Figure 12: Towns in Singapore with Key Venue Categories Required for Balanced Mix of Amenities

The distribution of towns with key venue categories tends to be more right-skewed, with most towns (16 of 25; 64%) having between 6-8 key venue categories.

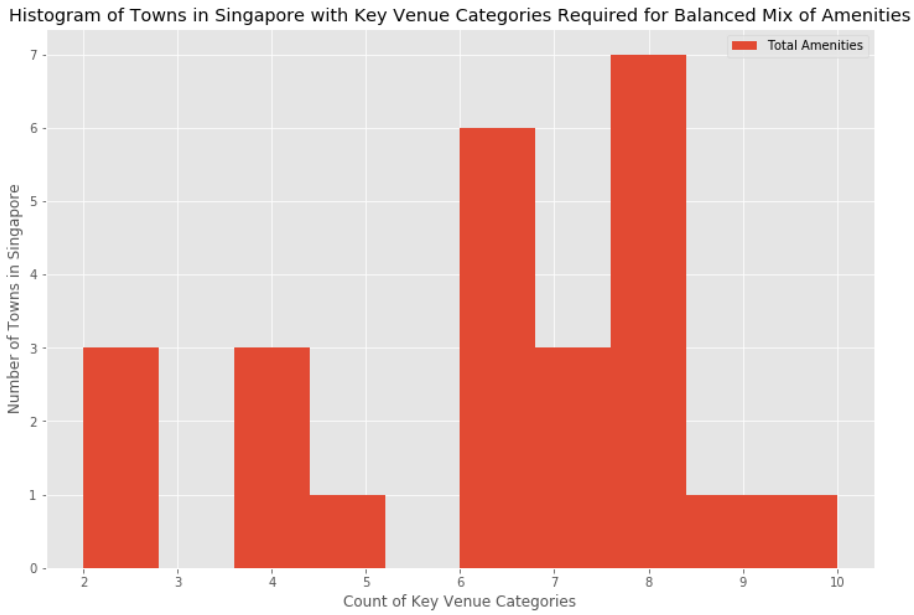

Figure 13: Histogram of Towns in Singapore with Key Venue Categories for Balanced Mix of Amenities

6. Relationship Between Median Rent, Population Density and Total Amenities

We have previously established that Population Density only provides some predictive value for Median Rent, and this relationship tends to be more evident for higher population densities. It is also important to assess the relationship of Total Amenities *(i.e. count of key venue categories required to assess balanced mix of amenities)* with Median Rent and Population Density, and their respective Pearson correlation coefficients were determined (Figure).

	Median Rent (SGD/month)	Population Density (people/km2)	Total Amenities
Median Rent (SGD/month)	1.000000	-0.450388	-0.653017
Population Density (people/km2)	-0.450388	1.000000	0.364211
Total Amenities	-0.653017	0.364211	1.000000

Figure 14: Correlation between Median Rent, Population Density and Amenities

It appears that Total Amenities is a good predictor for Median Rent, since the Pearson correlation coefficient is -0.653, with higher total amenities resulting in lower median rent. However, the scatter plot of Total Amenities and Median Rent shows that there is quite a wide spread of median rental prices across the range of total amenities *(Figure 15)*. The inverse relationship between Total Amenities and Median Rent may be more applicable to the extremes where total amenities may be a consequence of population density. This observation is further confirmed in the scatter plot of Population Density and Total Amenities, with Choa Chu Kang being a clear outlier with 30k people/km^2 *(Figure 16)*.

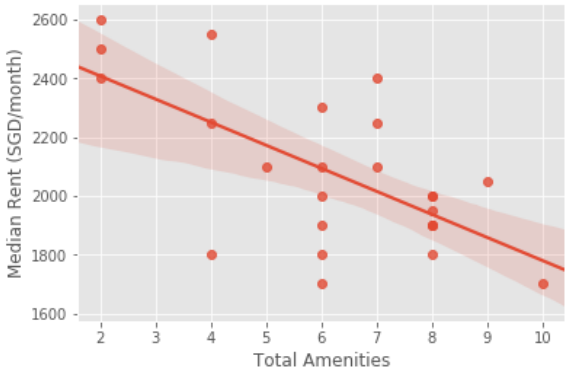

Figure 15: Scatter Plot for Median Rent vs. Total Amenities

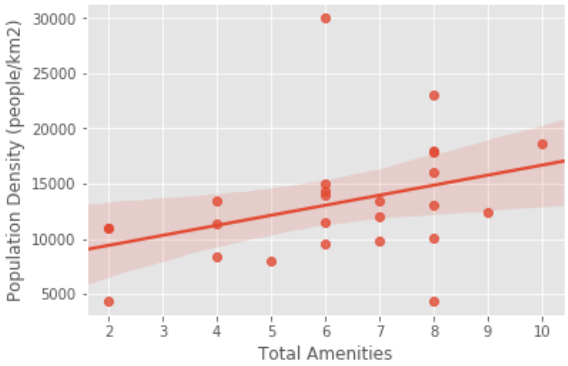

Figure 16: Scatter Plot for Population Density vs. Total Amenities

Clustering with K-Means

To better understand which towns are similar in terms of Median Rent, Population Density and Total Amenities, we will use K-Means Clustering to identify clusters of towns with similar characteristics, an unsupervised machine learning technique. Prior to clustering with K-Means, the optimal k for clustering was determined to be k=3, using the elbow method *(Figure 17)*.

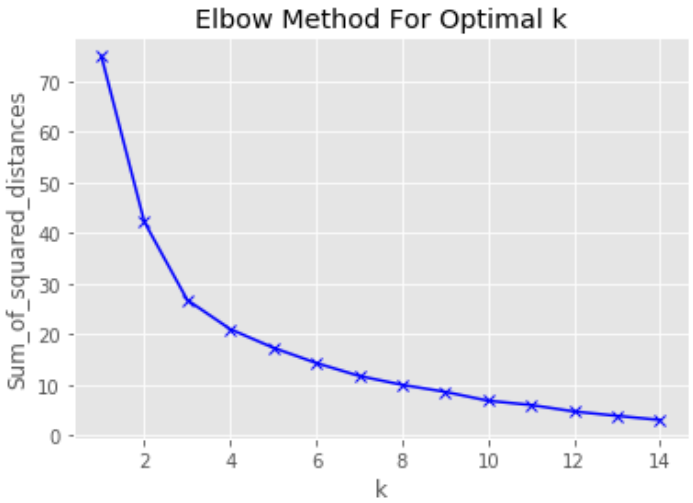

Figure 17: Elbow Method for Optimal k

Based on K-Means clustering, 3 clusters of towns were determined, based on Median Rent, Population Density and Total Amenities. The mean of values for these parameters for each of the 3 clusters was calculated to enable interpretation *(Figure 18)*. In addition, the descriptive statistics for each cluster were analyzed to enable better qualitative interpretation for the clusters. Finally, a choropleth map based on population density with town clusters superimposed on top was generated to enable visualization of the resultant clusters *(Figure 19)*.

Clus_km	Median Rent (SGD/month)	Population Density (people/km2)	Total Amenities
1	1850.000000	26500.000000	7.000000
2	2177.272727	9054.545455	4.909091
3	2025.000000	14841.666667	7.250000

Figure 18: Mean of Median Rent, Population Density and Total Amenities for Clusters

D. Results

Town Clusters based on Median Rent, Population Density and Total Amenities

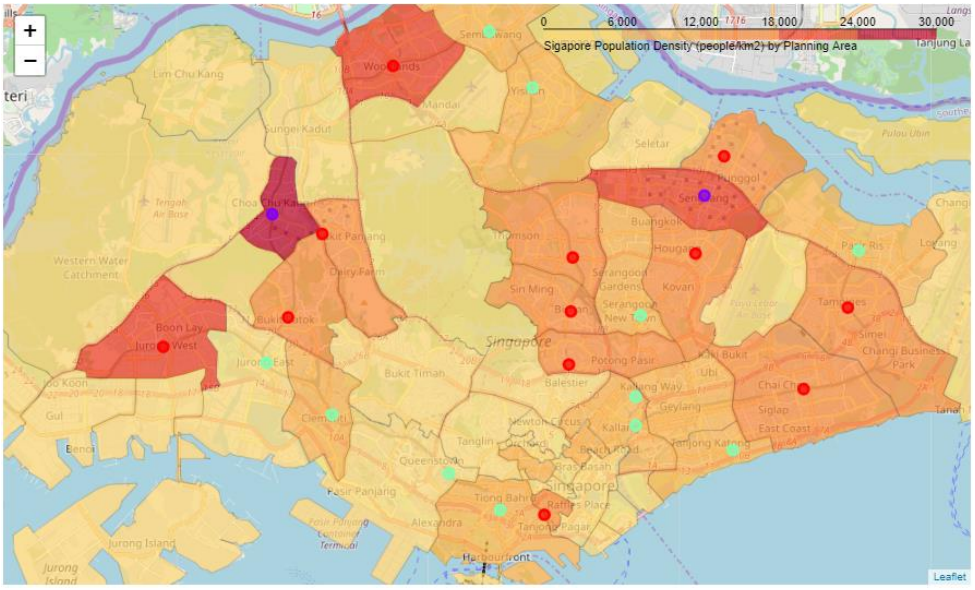

Figure 19: Choropleth Map of Singapore based on Population Density with Town Clusters

	Clus_km	1	2	3
Median Rent (SGD/month)	count	2.000000	11.000000	12.000000
	mean	1850.000000	2177.272727	2025.000000
	std	70.710678	273.279743	247.257687
	min	1800.000000	1800.000000	1700.000000
	25%	1825.000000	2000.000000	1900.000000
	50%	1850.000000	2100.000000	1975.000000
	75%	1875.000000	2400.000000	2137.500000
	max	1900.000000	2600.000000	2550.000000
Population Density (people/km2)	count	2.000000	11.000000	12.000000
	mean	26500.000000	9054.545455	14841.666667
	std	4949.747468	2566.462015	2282.127529
	min	23000.000000	4400.000000	12000.000000
	25%	24750.000000	8200.000000	13300.000000
	50%	26500.000000	9800.000000	14150.000000
	75%	28250.000000	11000.000000	16450.000000
	max	30000.000000	11500.000000	18700.000000
Total Amenities	count	2.000000	11.000000	12.000000
	mean	7.000000	4.909091	7.250000
	std	1.414214	2.300198	1.602555
	min	6.000000	2.000000	4.000000
	25%	6.500000	3.000000	6.000000
	50%	7.000000	5.000000	7.500000
	75%	7.500000	6.500000	8.000000
	max	8.000000	8.000000	10.000000

Figure 20: Table of Descriptive Statistics for Town Clusters

The following town clusters were generated using K-Means clustering:

- **Cluster 1 (Purple)**: High Population Density, Low Rent, Moderately Balanced Mix of Amenities
- **Cluster 2 (Green)**: Low Population Density, Variable Rent, Variable Mix of Amenities
- **Cluster 3 (Red)**: Moderate Population Density, Variable Rent, Most Balanced Mix of Amenities

To recap, the goal of this analysis is to identify the most livable neighborhoods in Singapore for individuals looking to relocate to Singapore and those considering moving within Singapore, with the most livable neighborhoods defined as having: (i) an affordable median rental price, (ii) a tolerable population density, and (iii) a balanced mix of amenities in the neighborhood.

Towns in Cluster 1 should be avoided, since despite the low rents (SGD 1,800-1,900/month), these towns have the highest population density (≥23k people/km^2), which falls outside tolerability thresholds. These towns tend to have moderately balanced mix of amenities (6-8 key venue categories). Cluster 1 only has 2 towns: Sengkang, Choa Chu Kang.

People with a preference for less crowded neighborhoods will most likely prefer to reside in Cluster 2 towns. Towns in Cluster 2 have the lowest population density (4.4-11.5k people/km^2), and variable mix of amenities (2-8 key venue categories). Median rents range between SGD 1,800-2,600/month, averaging SGD 2,177/month, so the choice of town will depend on one's budget and ability to afford. Cluster 2 has 11 towns: Yishun, Serangoon, Bukit Merah, Sembawang, Clementi, Geylang, Queenstown, Jurong East, Kallang, Marine Parade, Pasir Ris.

Towns in Cluster 3 are the most livable neighborhoods, as these have moderate population density (12.0-18.7k people/km^2) and the most balanced mix of amenities (4-10 key venue categories). Median rents range between SGD 1,700-2,550/month, averaging SGD 2,025/month, so likewise, the choice of town will depend on one's budget and ability to afford. Cluster 3 has 12 towns: Toa Payoh, Tampines, Punggol, Ang Mo Kio, Hougang, Outram, Bukit Panjang, Bukit Batok, Bishan, Bedok, Woodlands, Jurong West.

E. Discussion

Interpretation of Town Cluster Results

While 3 parameters – Median Rent, Population Density and Total Amenities – have been used in the clustering of towns, it appears that Population Density was the primary parameter used in this segmentation exercise. This is probably the best determinant for a livable neighborhood in Singapore besides median rent, as Singapore is an island city state, where space is likely to be a key constraint. Although Singapore has largely relied on high rise housing, the strain on surrounding amenities will likely remain for densely populated towns.

In addition, given that the small size of Singapore as a country, getting from one end of the island to another only takes less than an hour on car, so the proximity of amenities is likely to rank as the least

important criteria, given the relative proximity of towns to one another. Ultimately, the choice of a suitable town to reside in will depend on the individual's affordability and housing budget.

Limitations of K-Means Clustering

We used K-Means clustering to identify the town clusters for this purpose of this analysis, but this unsupervised algorithm has some inherent limitations. Namely, the town clusters may differ slightly in different runs of K-Means clustering, since the algorithm works by randomly placing k centroids and iteratively shifting the centroid position to create clusters with minimum error, or the densest clusters. To ensure better results, it may be helpful to run the K-Means clustering algorithm multiple times with different starting conditions to randomize the starting centroids and assess the consistency of clusters for the towns.

Suggestions for Future Work

This study determined the most livable neighborhoods in Singapore based on: (i) an affordable median rental price, (ii) a tolerable population density, and (iii) a balanced mix of amenities in the neighborhood. In the future, for more granular analysis, sub-clusters can be identified within each main cluster to further segment towns based on other criteria, including availability of specific venues *(e.g. restaurants, bars, parks)*, accessibility to top-rated venues, attractions, and workplaces, amongst others. These segments can then be mapped to different profiles of individuals looking to reside in Singapore.

F. Conclusion

In this study, the most livable neighborhoods in Singapore were identified based on: (i) an affordable median rental price, (ii) a tolerable population density, and (iii) a balanced mix of amenities in the neighborhood. K-Means clustering was performed on Singapore towns using Median Rent, Population Density and Total Amenities *(i.e. count of key venue categories required to assess balanced mix of amenities)*. Towns in Cluster 3 have been identified as the most livable neighborhood, followed by Cluster 2 towns with the lowest population density, and with Cluster 1 towns being excluded due to high population density. The results of this study will be useful for individuals looking to relocate to Singapore and those considering moving within Singapore, given Singapore's reputation as a preferred business destination in Asia, with its vibrant economy, low personal income taxes, cultural diversity and high quality of living.