Battle of the Neighborhoods – Singapore- Best Neighborhood for Living

-Coursera IBM Data Science Capstone Submission-

Hao Z.

Apr. 9, 2020

Table of Contents

1.0 Introduction

1.1 Background Information:

Singapore is a city-state country located off of the southern tip of the Malaysian peninsula. With clean streets free of littering and vandalism enforced by laws, Singapore has been constantly ranked as one of the top cities to live in around the world. Singapore not only acts as the economic hub of Southeast Asia, but it also has offers great educational opportunities.

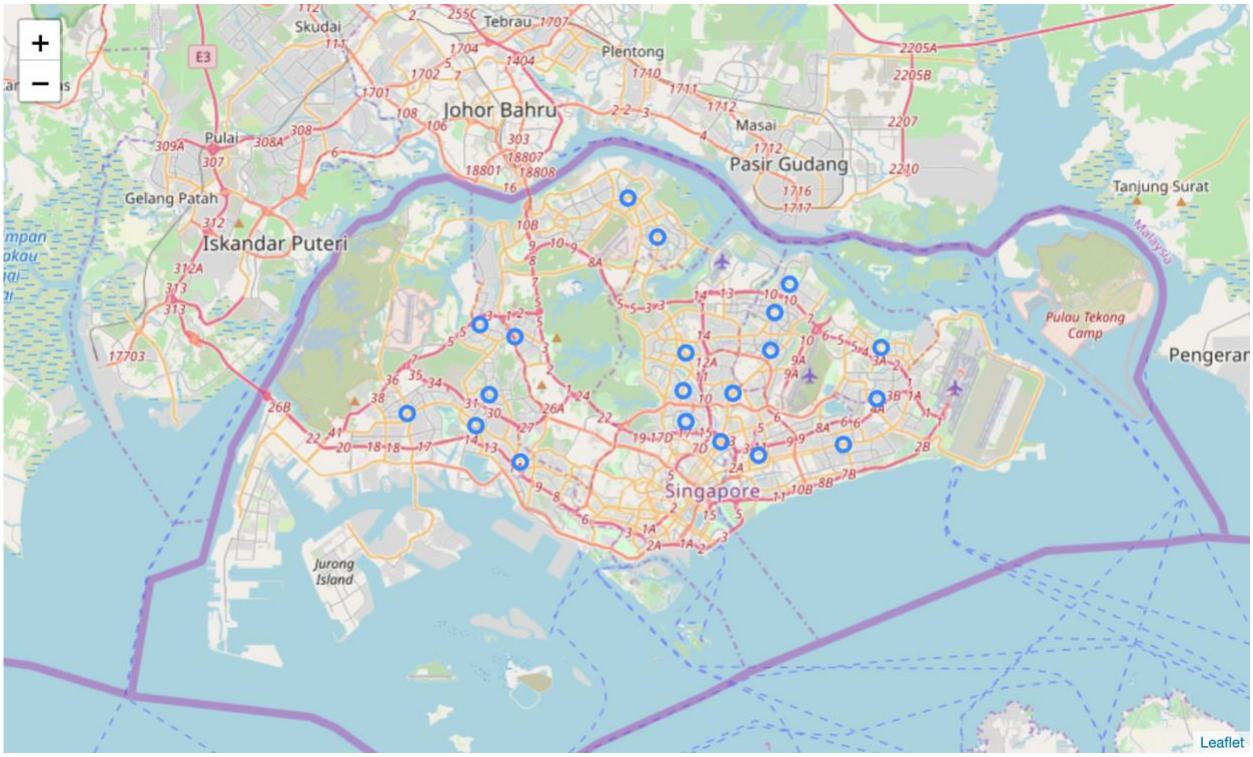

Figure 1: Map of Singapore visualized in Folium

There is a total of 27 core neighborhoods in Singapore, namely the following:

- Bukit Batok
- Bedok
- Bishan
- Bukit Merah
- Bukit Panjang
- Bukit Timah
- Choa Chu Kang
- Clementi
- Central Area
- Geylang
- Hougang
- Jurong East

- Jurong West
- Kallang/Whampoa
- Marine Parade
- Punggol
- Pasir Ris
- Queenstown
- Sembawang
- Serangoon
- Sengkang
- Tampines
- Tengah
- Toa Payoh
- Woodlands
- Yishun

Each of the above neighborhood represents a different geographical region in Singapore, and within each neighborhood, there are many public services and businesses that facilitate the daily lives of local Singaporeans.

1.2 Problem Statement

The goal of this project is to analyze which neighborhood in Singapore is the most suitable for a middle-income family. The neighborhoods will be evaluated based on the criteria listed below:

- Average household income and housing price of neighborhoods,
- School distribution,
- Public transit access,
- Access to grocery stores/markets,
- Access to restaurants and hawker centres, and
- Access to community and entertainment facilities.

The neighborhoods will be evaluated against each other and the mean frequency or value. Then, a recommendation will be made, using a weighted matrix.

2.0 Data

2.1 Neighborhood Data and Coordinates

The latitude and longitude of each neighborhood is retrieved using Geopy. This information is useful to retrieve venue data in each neighborhood from Foursquare.

Table 1: Coordinates of Singapore Neighborhoods

	Areas	Coordinates	Area_Latitude	Area_Longitude
0	Ang Mo Kio	1.3700733, 103.8495157	1.3700733	103.8495157
1	Bukit Batok	1.3490572, 103.7495906	1.3490572	103.7495906
2	Bedok	1.3239765, 103.930216	1.3239765	103.930216
3	Bishan	1.3514521, 103.8482496	1.3514521	103.8482496
4	Bukit Merah	4.5592879, 101.0255816	4.5592879	101.0255816
5	Bukit Panjang	1.378629, 103.7621358	1.378629	103.7621358
6	Choa Chu Kang	1.3847493, 103.7445341	1.3847493	103.7445341
7	Clementi	1.3151003, 103.7652311	1.3151003	103.7652311
8	Central Area	36.5070827, -79.7447575	36.5070827	-79.7447575
9	Geylang	1.3181862, 103.8870563	1.3181862	103.8870563
10	Hougang	1.3719043, 103.892725	1.3719043	103.892725
11	Jurong East	1.333115, 103.7422968	1.333115	103.7422968
12	Jurong West	1.3396365, 103.7073387	1.3396365	103.7073387
13	Kallang/Whampoa	1.3251102, 103.8671678	1.3251102	103.8671678
14	Punggol	1.4052585, 103.9023302	1.4052585	103.9023302
15	Pasir Ris	1.3730307, 103.949255	1.3730307	103.949255
16	Queenstown	-45.0321923, 168.661	-45.0321923	168.661
17	Sembawang	1.4490928, 103.8200555	1.4490928	103.8200555
18	Serangoon	1.3497629, 103.873721	1.3497629	103.873721
19	Sengkang	1.3913304, 103.8952941	1.3913304	103.8952941
20	Tampines	1.3468925999999999, 103.94734087817068	1.3468925999999999	103.94734087817068
21	Toa Payoh	1.3353906, 103.8497414	1.3353906	103.8497414
22	Woodlands	30.1734194, -95.504686	30.1734194	-95.504686
23	Yishun	1.4293839, 103.8350282	1.4293839	103.8350282

2.2 Average household income:

Average income in different neighborhoods of Singapore:

The information acquired from Asia One shows the median annual salary recommended to purchase a 4BR HDB apartment in each neighborhood. HDB is a type of public housing for Singaporeans, sponsored by the Housing & Development Board of Singapore.

Average monthly household income based on the types of dwelling:

The Singaporean government has published information regarding the average monthly household income of different dwelling types (i.e. 2 BR HDB, 3 BR HDB, 4 BR HDB, condominiums, etc.). This value can be used as a benchmark for analysis.

2.3 Average housing price:

Average housing price in different neighborhoods of Singapore:

Asia One's article from above also contains information showing the average housing price for a 4BR HDB apartment in 24 neighborhoods of Singapore, as well as the recommended down payment that is required. Only the average housing price will be used.

HDB pricing data in Singapore of all neighborhoods, published by the Government of Singapore, can be accessed here. The link contains information regarding all HDBs in Singapore. This information can be used to determine the number of HDB buildings in each neighborhood.

The table below is a summarized data of family annual income and resale price for a four-bedroom HDB apartment across all neighborhoods in Singapore.

Table 2: Average Annual Income and Resale Price for 4 Bedroom Apartments in Singapore

	Areas	Income (SGD)	4BD Resale (k SGD)
0	Ang Mo Kio	61909	460.0
1	Bukit Batok	51142	380.0
2	Bedok	54911	408.0
3	Bishan	78059	580.0
4	Bukit Merah	89163	662.5
5	Bukit Panjang	49527	368.0
6	Choa Chu Kang	44413	330.0
7	Clementi	75031	557.5
8	Central Area	117089	870.0
9	Geylang	64601	480.0
10	Hougang	50066	372.0
11	Jurong East	54843	407.5
12	Jurong West	49124	365.0
13	Kallang/Whampoa	80751	600.0
14	Punggol	59419	441.5
15	Pasir Ris	54628	405.9
16	Queenstown	99324	738.0
17	Sembawang	45073	334.9
18	Serangoon	55180	410.0
19	Sengkang	55180	410.0
20	Tampines	58006	431.0
21	Toa Payoh	78059	580.0
22	Woodlands	44413	330.0
23	Yishun	45759	340.0

It is worth noting that a four-bedroom apartment has been used as a result of missing data from either income or resale value of other apartment types, thus making it difficult to directly correlate the variables.

2.4 School Distribution:

The Ministry of Education of Singapore has published a datasheet, summarizing all the school data in Singapore across various neighborhoods. This can be used to learn about the accessibility to educational institution in each neighborhood of Singapore.

Table 3: Neighborhood by School Count in Singapore

	Areas	Count
0	ANG MO KIO	14
1	BEDOK	22
2	BISHAN	10
3	BUKIT BATOK	13
4	BUKIT MERAH	11
5	BUKIT PANJANG	13
6	CENTRAL AREA	7
7	CHOA CHU KANG	15
8	CLEMENTI	10
9	GEYLANG	9
10	HOUGANG	17
11	JURONG EAST	6
12	JURONG WEST	22
13	KALLANG/WHAMPOA	5
14	PASIR RIS	11
15	PUNGGOL	13
16	QUEENSTOWN	9
17	SEMBAWANG	7
18	SENGKANG	18
19	SERANGOON	10
20	TAMPINES	20
21	TOA PAYOH	14
22	WOODLANDS	23
23	YISHUN	19

2.5 Public Transit Accessibility:

The number of bus stops, MRT stations and LRT stations is a good indicator to the connectivity of each neighborhood. Access to transit can improve family mobility and sustainability. Data from Foursquare will be used, for the count of MRT and LRT stations.

However, as the retrieval of bus stops, MRT and LRT stations from Foursquare has been limited, public transit accessibility will not be used in the final evaluation of the neighborhood.

It is important to note that Singapore tops the global ranking in public transportation, according to the Straight Times.

2.6 Venue Data Retrieval using Foursquare API

As the neighborhoods in Singapore are around 5km by 5km in terms of size, a radius of 2000m is used to retrieve venue information around the neighborhood coordinates provided in Table 1 above. For each neighborhood, the limit is set to 100 results. The frequency of each category for each neighborhood is then assessed.

2.7 Restaurants/Hawker Centres:

Whether it is for convenience or for special occasions, having restaurants and hawker centres outside the doorstep can bring convenience to family living. Foursquare will be employed to retrieve dining options.

The top ten venue categories have been studied for all neighborhoods. As restaurants typically fall into a large number of categories depending on the cuisine, the following restaurant types are used, as they occurred most frequently in the frequency table:

- Food Court
- Cafe
- Chinese Restaurant
- Fast Food Restaurant
- Indian Restaurant
- Japanese Restaurant
- Asian Restaurant
- Coffee Shop

By the summation of frequency of each category in the neighborhood, we arrive at a total frequency of the restaurant category for each neighborhood:

Table 4: Frequency of Restaurants (partial collection) by Neighborhood

	Neighborhood	Total Restaurants
1	Bedok	0.450000
10	Hougang	0.420000
23	Yishun	0.420000
3	Bukit Batok	0.410000
17	Sembawang	0.400000
0	Ang Mo Kio	0.390000
21	Toa Payoh	0.360000
13	Kallang/Whampoa	0.360000
11	Jurong East	0.358696
12	Jurong West	0.350515
7	Choa Chu Kang	0.338028
20	Tampines	0.330000
2	Bishan	0.320000
19	Serangoon	0.320000
18	Sengkang	0.310000
8	Clementi	0.300000
9	Geylang	0.300000
15	Punggol	0.250000
5	Bukit Panjang	0.246914
4	Bukit Merah	0.200000
14	Pasir Ris	0.195876
16	Queenstown	0.119565
22	Woodlands	0.085106
6	Central Area	0.026316

From the table above, it can be seen that Bedok has the highest frequency for restaurants of the named categories.

2.8 Grocery Stores/Markets/Living Essentials:

Having grocery stores and supermarkets within walkable distance is important to the daily living of families. Foursquare will be used to retrieve grocery store and supermarket information within each neighborhood using the venues function. The total count per neighborhood will be used.

Foursquare API data retrieval process from Section 2.7 is repeated for Grocery Stores.

2.9 Leisure and Entertainment Facilities:

Foursquare API will be used to retrieve information within each neighborhood with regards to entertainment and community services, such as parks, movie theatres and gyms.

Foursquare API data retrieval process from Section 2.7 is repeated for Leisure and Entertainment Facilities.

The table below summarizes the data retrieved using Foursquare API for Restaurants, Living Essentials and Leisure Facilities.

Table 4: Summary of Total Frequency of Venue Categories in Singapore from Foursquare API

	Neighborhood	Total Living	Total Restaurants	Total Transpo	Total Leisure
0	Bukit Merah	0.200000	0.200000	0.000000	0.000000
1	Woodlands	0.106383	0.085106	0.000000	0.063830
2	Central Area	0.078947	0.026316	0.000000	0.026316
3	Choa Chu Kang	0.070423	0.338028	0.014085	0.014085
4	Bukit Panjang	0.061728	0.246914	0.024691	0.012346
5	Punggol	0.060000	0.250000	0.030000	0.030000
6	Jurong East	0.054348	0.358696	0.000000	0.021739
7	Pasir Ris	0.051546	0.195876	0.030928	0.020619
8	Sengkang	0.050000	0.310000	0.000000	0.030000
9	Jurong West	0.041237	0.350515	0.010309	0.020619
10	Bishan	0.040000	0.320000	0.000000	0.010000
11	Toa Payoh	0.040000	0.360000	0.000000	0.030000
12	Serangoon	0.040000	0.320000	0.010000	0.000000
13	Hougang	0.040000	0.420000	0.000000	0.010000
14	Kallang/Whampoa	0.040000	0.360000	0.000000	0.020000
15	Ang Mo Kio	0.030000	0.390000	0.000000	0.020000
16	Bedok	0.030000	0.450000	0.000000	0.000000
17	Clementi	0.020000	0.300000	0.000000	0.010000
18	Bukit Batok	0.020000	0.410000	0.000000	0.010000
19	Yishun	0.020000	0.420000	0.000000	0.020000
20	Sembawang	0.016667	0.400000	0.016667	0.000000
21	Geylang	0.010000	0.300000	0.000000	0.020000
22	Tampines	0.010000	0.330000	0.000000	0.040000
23	Queenstown	0.000000	0.119565	0.000000	0.021739

3.0 Methodology

The data collected in Section 2 above will be analyzed by the following methods:

Housing Price vs. Income: the relationship between housing price and income will be observed by regression plotting.

Housing Price plot: this bar plot is essential in visualizing the average pricing of apartments in each neighborhood compared with the mean pricing of all neighborhoods.

Housing Price vs. Number of Schools: the relationship between the housing of the neighborhoods will be studied using a regression plot.
Neighborhood Convenience: bar plots will be used to visualize the frequencies of restaurants, living essentials and leisure venues in each neighborhood, and will be plotted with the average frequencies of all neighborhoods.

Final Evaluation Matrix:

Table 5: Evaluation Matrix for Neighborhoods

Category	Housing Price	School	Restaurants	Living Essentials	Leisure
Weight (%)	35%	15%	10%	25%	15%

The matrix has been developed with the following considerations:

- Housing price is the single most important factor that determines where the families can live, as a family's affordability cannot change overnight.
- Being able to live near grocery stores, pharmacies and convenience stores is important to ensure that the families can access the resources that they need every day.
- There are a number of both independent and public schools available in Singapore, therefore it offers each family some options to choose school within the neighborhood or further away.
- Spending time with family and access to leisure and entertainment is crucial to the new generation, for work-life balance.
- Restaurant preferences are the most difficult to predict, since it varies from one person to another. The assigned score is to reflect the available options nearby, but families can choose their favorite joints further away.

4.0 Analysis

4.1 Housing Price vs. Income:

Income determines the affordability of an individual or a family, therefore the income is plotted along the x-axis as the independent variable, while housing price is plotted along the y-axis as the dependent variable to income.

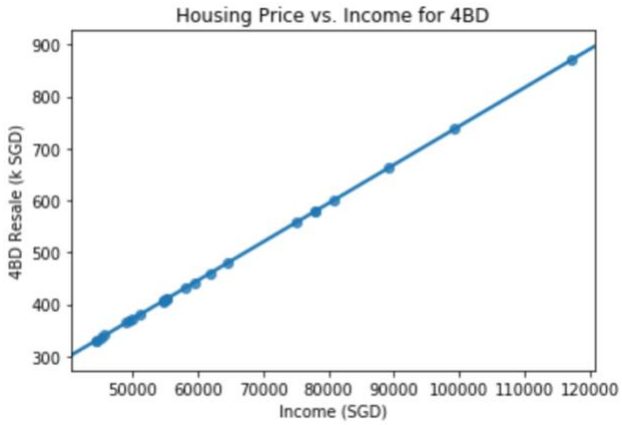

Figure 2: Housing Price vs. Income for 4 Bedroom Apartment in Singapore

As illustrated by the figure above, housing price is linearly dependent of the income of families. As the income of families increase, the resale values of the apartments also increase. This is suggestive that higher income families tend to be clustered in neighborhoods with higher housing values, while the lower income families are clustered together with limited housing affordability.

4.2 Housing Price plot:

Housing price in each neighborhood is compared to the mean housing price of all neighborhoods living in a 4BD HDB apartment complex.

This will be meaningful in determining the need for a middle-class family. Too low or too high in terms of the housing price will hurt the family financially, therefore the neighborhoods with close-to-average housing price in Singapore will be looked at more closely.

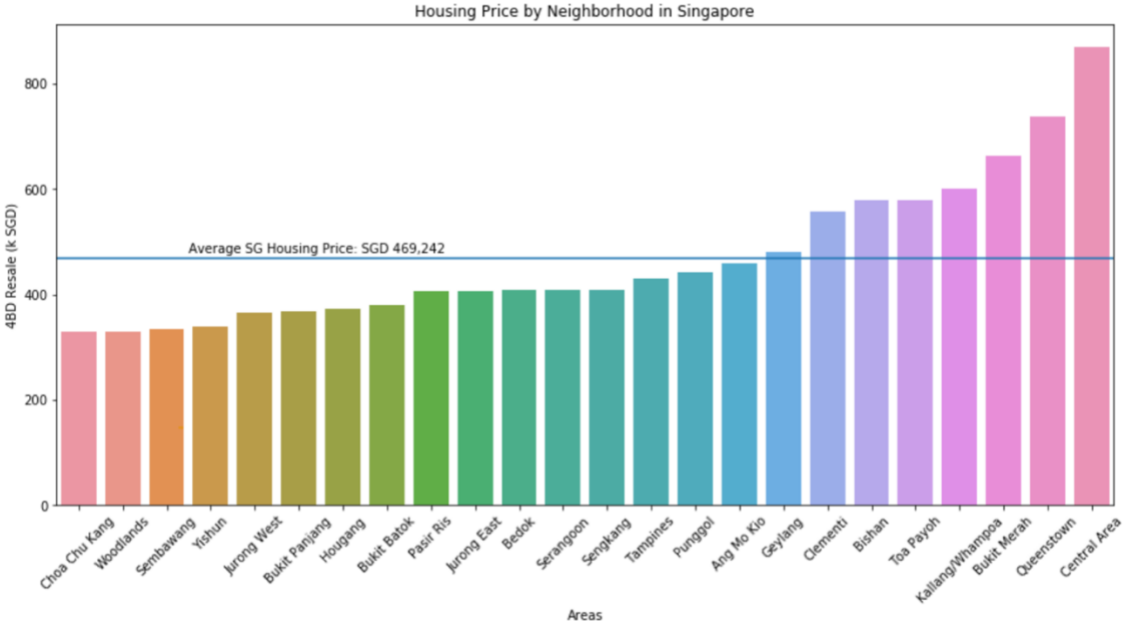

Figure 3: Housing Price in Neighborhoods of Singapore and Mean Housing Price for 4BD HDB

A third of the neighborhoods in Singapore have apartments above the average housing price in Singapore of SGD 469,242. In the Central Area in particular, the housing prices are roughly 40% higher than the average.

It is recommended that middle income families to choose neighborhoods around average Singapore housing prices, to avoid large financial burdens.

4.3 Housing Price vs. Number of Schools

The number of schools can be used as a good indicator of housing pricing. The total number of schools in each neighborhood will be plotted against the housing price, to gain a general understanding of the trend and effect of number schools on the price.

Figure 4: Housing Price vs. Total Number of Schools in Neighborhoods of Singapore

The figure above shows that as the available educational institution in neighborhoods decrease, the resale value of the apartments increases. The two variables are inversely proportional.

This could be due to that only four-bedroom apartments are studied, and that larger apartments are situated in less-dense neighborhoods. As the density of neighborhoods are low, the need for educational institution is low as well. This could also mean that more expensive four-bedroom apartments are situated on the outskirts of the city-state.

The figure below shows the count of schools per neighborhood as well as the average number of schools in Singapore per neighborhood.

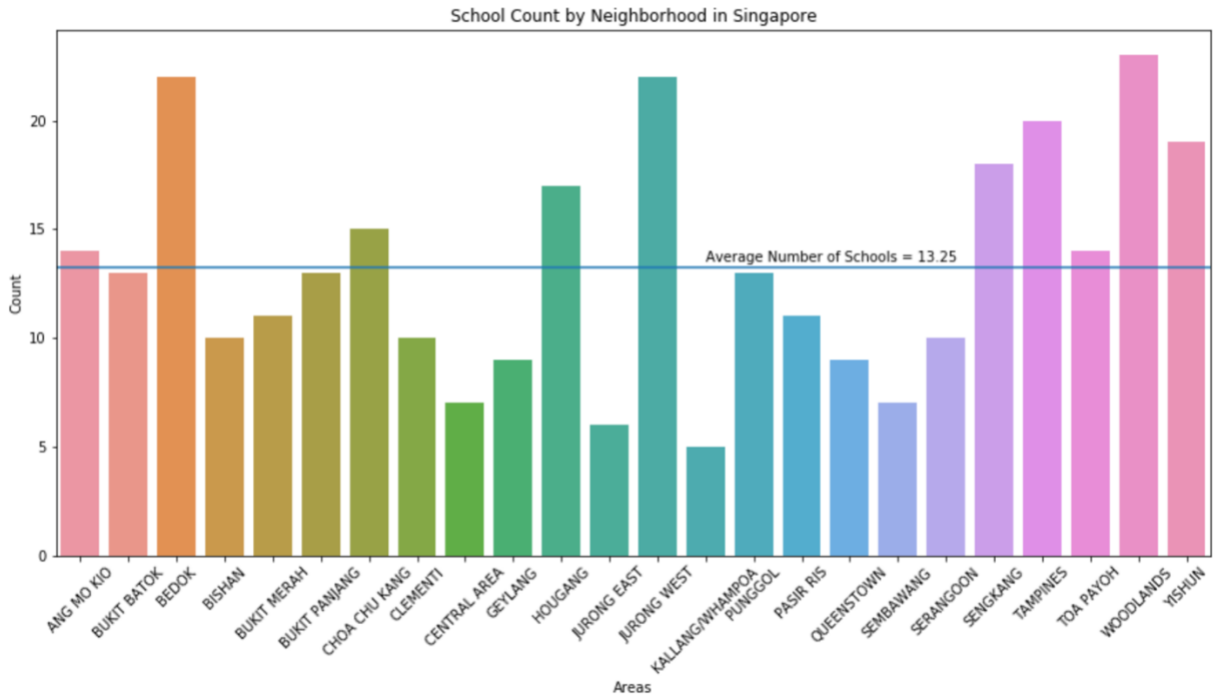

Figure 5: School Count by Neighborhood and Average in Singapore

4.4 Neighborhood Convenience

Restaurants:

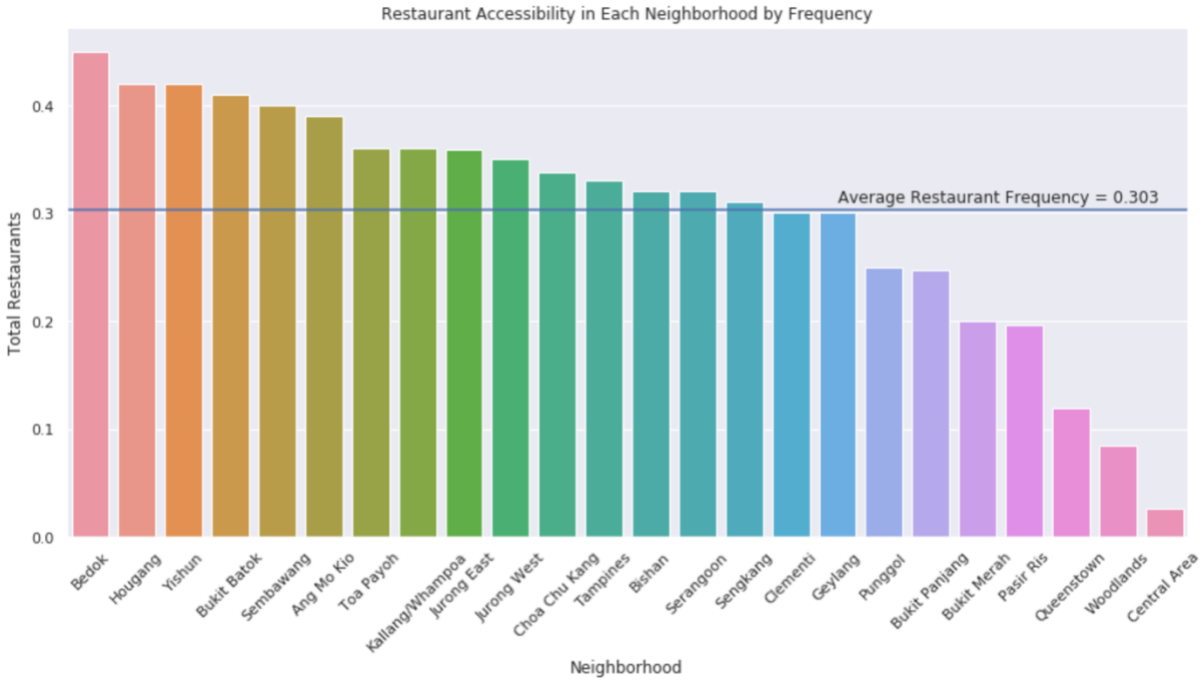

Figure 6: Restaurant Accessibility in Neighborhoods of Singapore

60% of the areas in Singapore are around or above the average restaurant frequency, meaning that they have more options available than average neighborhoods. However, the frequency of the restaurants is only calculated based on the selected eight types of restaurants outlined in Section 2.7. In particular, in areas such as the Central Area, higher frequency restaurants may include high-end dining venues, which are not included in the calculation as they would not be frequently visited by average families.

Living Essentials:

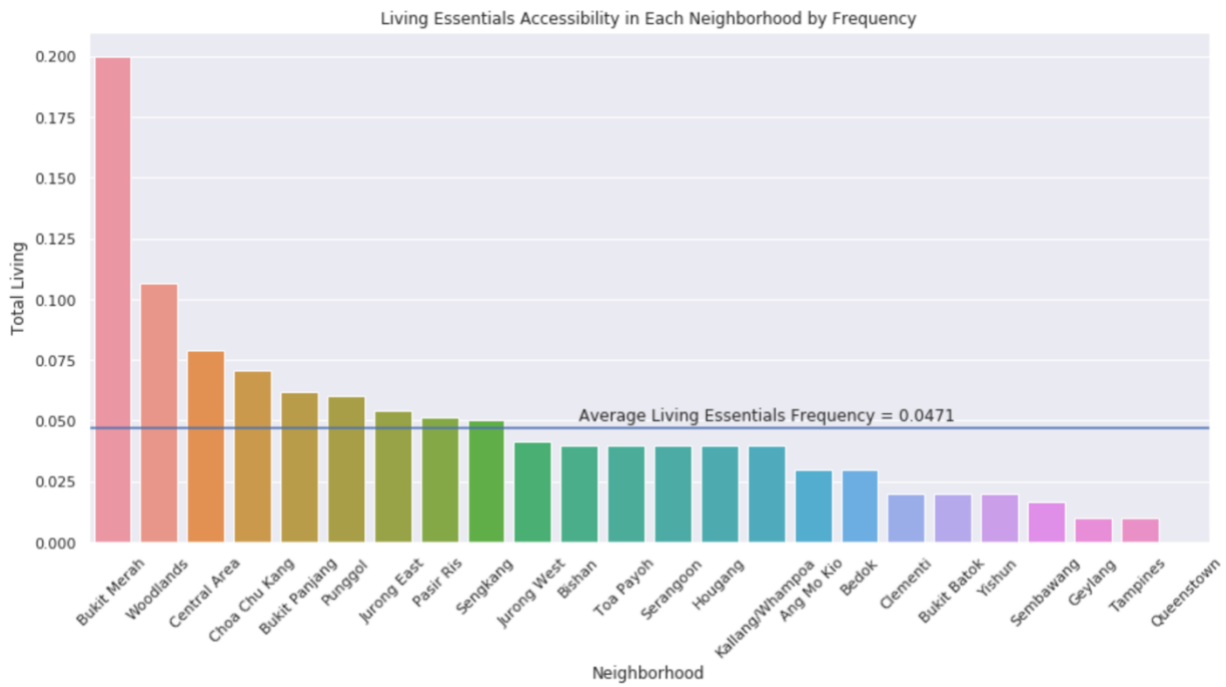

Figure 7: Stores for Essential Items Accessibility in Neighborhoods of Singapore

Singapore neighborhoods show a fairly steady and equal access to stores with living essentials, with a few outliers on both ends.

Leisure and Entertainment Facilities:

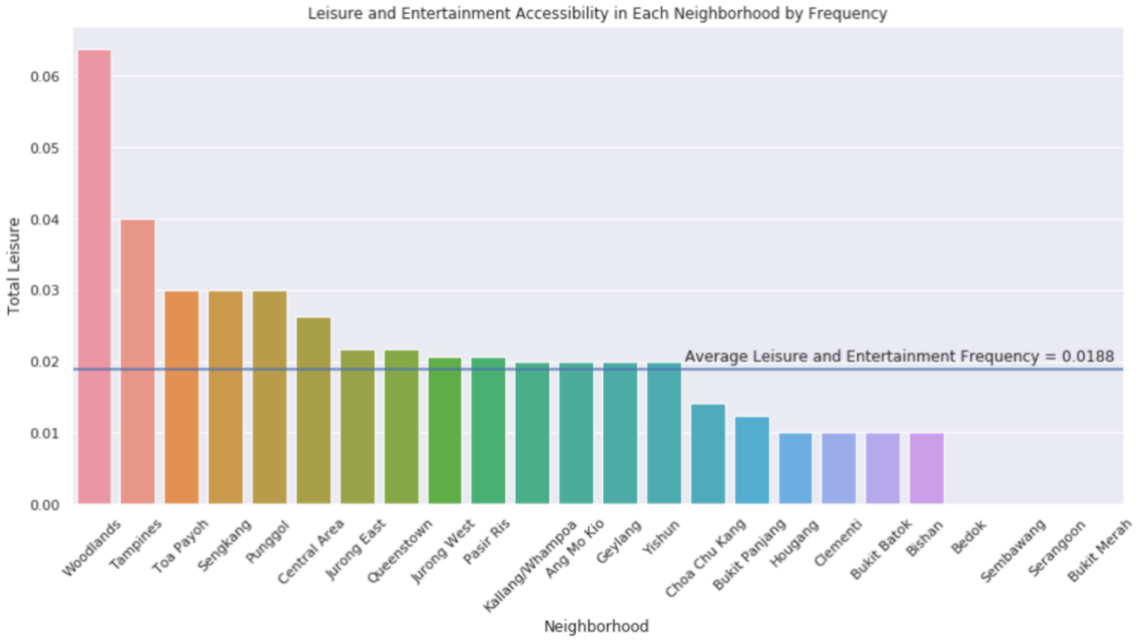

Figure 8: Stores for Essential Items Accessibility in Neighborhoods of Singapore

The above figure shows that Singapore neighborhoods have fairly equal access to leisure, entertainment and community facilities.

It is worth noting that there are missing values for Bedok, Sembawang, Serangoon and Bukit Merah. This could be due to the categorical keyword selection when querying data or that the venues are not properly logged in Foursquare.

4.5 Weighted Matrix

The evaluation matrix will be applied at this final stage to determine the best neighborhood for a middle-class family.

Normalization of Housing and School Data

Housing price and school data have been normalized to compare the neighborhoods with the mean. The range of normalized values lie between -1 and 1, depending on its values compared with the mean. As a result, it is preferred to have normalized values closer to 0, which means that the housing price and school distribution do not deviate significantly from the mean. This is crucial to ensure that the housing price is affordable.

The normalized data is further processed by taking the absolute values of normalized values.

Restaurants/Living Essentials/Leisure

The frequency data for restaurants, access to essential stores and leisure facilities is to be used when the evaluation factors are applied.

The frequency and normalized values of each category are multiplied by the factors as indicated in Table 5 above, or summarized below:

- Housing: 35%
- Schools: 15%
- Restaurants: 10%
- Living Essentials: 25%
- Leisure: 15%

The neighborhoods are then arranged in ascending order based on the sum of the evaluation factors. The best neighborhood is chosen with the minimum score. This is due to that absolute normalized values have been used for housing price and school count. Minimum score ensures that the chosen neighborhood has a close-to-average housing price and school count.

The final calculated table is shown below.

Table 6: Ranking of Neighborhoods based on Descending Order of Final Evaluated Weight

Neighborhood	Living	Restaurants	Leisure	Norm_School_Count	Norm_Housing_Price	Evaluated Weight
ANG MO KIO	0.030000	0.380000	0.020000	0.041667	0.017114	0.060740
PUNGGOL	0.070000	0.250000	0.030000	0.013889	0.051373	0.067064
GEYLANG	0.010000	0.300000	0.020000	0.236111	0.019923	0.077890
PASIR RIS	0.054348	0.217391	0.021739	0.125000	0.117299	0.098392
BUKIT BATOK	0.020000	0.410000	0.010000	0.013889	0.165262	0.107425
BUKIT PANJANG	0.060241	0.228916	0.012048	0.013889	0.187485	0.107462
SERANGOON	0.040000	0.300000	0.020000	0.180556	0.109707	0.108481
CLEMENTI	0.020000	0.290000	0.010000	0.180556	0.163441	0.119788
SENGKANG	0.050000	0.270000	0.030000	0.263889	0.109707	0.121981
TAMPINES	0.010000	0.330000	0.040000	0.375000	0.070818	0.122536
TOA PAYOH	0.040000	0.350000	0.030000	0.041667	0.205108	0.127538
BISHAN	0.040000	0.320000	0.010000	0.180556	0.205108	0.142371
JURONG EAST	0.033333	0.355556	0.022222	0.402778	0.114336	0.147657
HOUGANG	0.040000	0.420000	0.010000	0.208333	0.180077	0.147777

CHOA CHU KANG	0.078947	0.315789	0.013158	0.097222	0.257855	0.158122
BEDOK	0.030000	0.460000	0.000000	0.486111	0.113410	0.166110
YISHUN	0.020000	0.410000	0.020000	0.319444	0.239336	0.180684
SEMBAWANG	0.034483	0.431034	0.000000	0.347222	0.248781	0.190881
JURONG WEST	0.050000	0.350000	0.020000	0.486111	0.193040	0.190981
KALLANG/WHAMPOA	0.040000	0.350000	0.010000	0.458333	0.242145	0.200001
BUKIT MERAH	0.200000	0.200000	0.000000	0.125000	0.357886	0.214010
WOODLANDS	0.108696	0.086957	0.108696	0.541667	0.257855	0.223673
QUEENSTOWN	0.000000	0.130435	0.021739	0.236111	0.497701	0.225916
CENTRAL AREA	0.081081	0.027027	0.027027	0.347222	0.742145	0.338861

The neighborhood of Ang Mo Kio has the lowest score, indicating that it is the best neighborhood for middle-class families. This is reflected by its housing price, that it is very close to the average. Ang Mo Kio also has great access to restaurants.

On the other hand, Central Area is the least suitable for a middle-class family, primarily as a result of its high housing price. Such high housing price significantly deviates away from the average housing price across Singapore, making it a financial burden for middle-class families.

5.0 Discussions

As this is a preliminary report for the choice of neighborhoods for a middle-class family, a more in-depth investigation can be conducted. Ang Mo Kio has been deemed as the most suitable neighborhood for residence of a middle-class family in Singapore.

Ang Mo Kio has hit average scores with housing price, school count and access to leisure facilities. It also has above-average access to restaurants nearby in the specified categories. However, Ang Mo Kio's access to groceries stores, pharmacies and supermarkets are slightly lacking compared with the average accessibility of all neighborhoods.

While family income and housing price has a positive linear correlation in Singapore, the housing price is inversely proportional to the number of education institutions in each neighborhood, as explained previously in Section 4.3 of the report.

There is a number of factors that have not been considered in this research, such as transportation. As mentioned earlier in Section 2.5 of the report, there was limitations with transit data from Foursquare API. Since Singapore is renowned for its public transport system, this research was conducted based on the understanding that access to transit is equal across all neighborhoods. In future, other sources can be employed to improve the accuracy of the prediction.

6.0 Conclusion

An investigation has been conducted to determine which neighborhood is the most suitable for middle-class families in Singapore. Five criteria have been reviewed, namely housing price, educational access, access to restaurants, access to essentials stores and access to leisure facilities. Each of the above criteria has been assigned a weight to determine the final score of the neighborhoods.

Ang Mo Kio is deemed to be the most-suitable neighborhood for middle-class families, as a result of its average housing price and great access to resources required in daily lives.

Ang Mo Kio neighborhood is visualized in the map below.

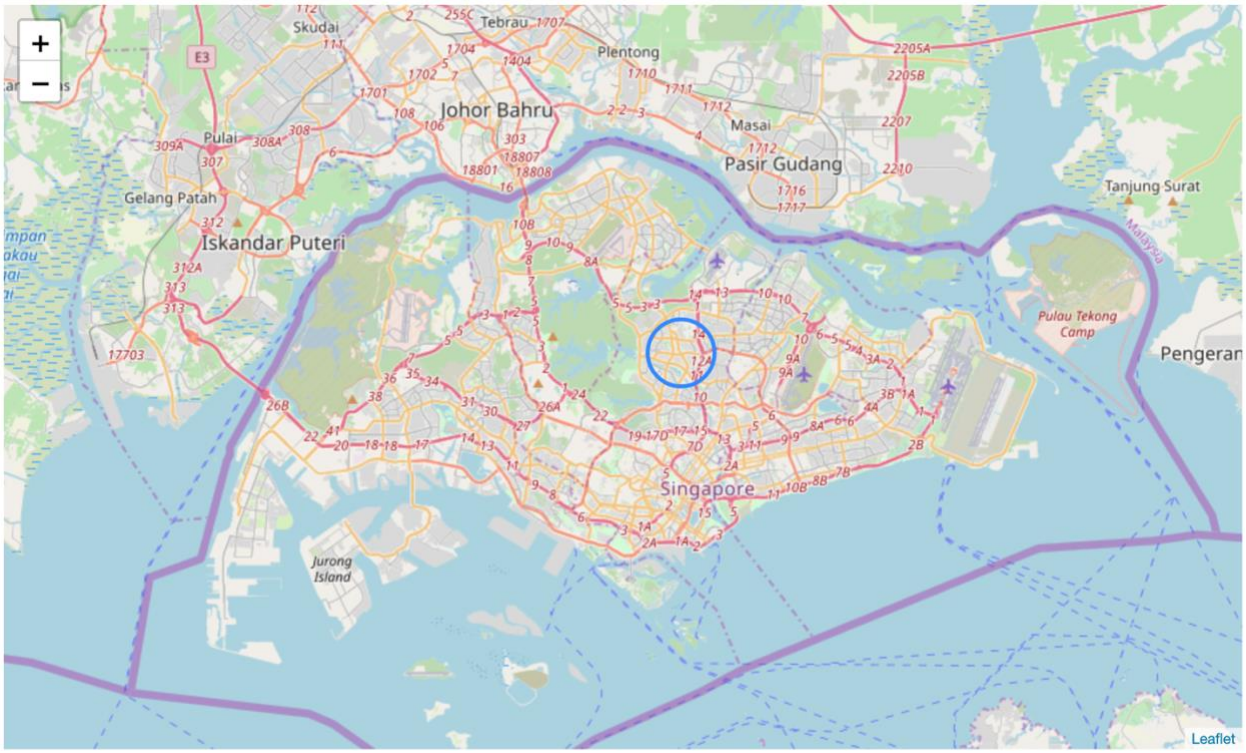

Figure 9: Ang Mo Kio Neighborhood in Singapore visualized in Folium

Reference List

1. https://www.asiaone.com/singapore/what-average-income-singaporeans-living-different-neighbourhoods
2. https://www.tablebuilder.singstat.gov.sg/publicfacing/createSpecialTable.action?refId=16779
3. https://data.gov.sg/dataset/hdb-property-information?view_id=82d8145d-daa6-460b-b269-9848c99ed32d&resource_id=482bfa14-2977-4035-9c61-c85f871daf4e
4. https://data.gov.sg/dataset/school-directory-and-information?view_id=ba7c477d-a077-4303-96a1-ac1d4f25b190&resource_id=ede26d32-01af-4228-b1ed-f05c45a1d8ee
5. https://www.straitstimes.com/singapore/transport/spore-public-transport-system-tops-global-list