

Picking A Singaporean Neighborhood to Live In

Mao Tian Tan

September 2021

Contents

1	Introduction	1
2	Data	1
3	Methodology	2
3.1	Exploratory Data Analysis	2
3.1.1	Neighborhoods of Singapore	2
3.1.2	Venues	2
3.1.3	Frequency of Venues	3
3.1.4	Neighborhoods with the most gyms	4
3.1.5	Neighborhoods with the most malls	4
3.1.6	Neighborhoods with the most coffee shops	6
3.2	Machine Learning Methods	6
4	Results	8
5	Discussion	8
6	Conclusion	10

1 Introduction

Singapore is a modern city with many districts, each containing many venues like restaurants or shopping centers. For someone wanting to rent or buy an apartment in Singapore, the choice of housing location can be rather daunting. The goal of this data science project is to explore the kinds of venues that are located within each district. For example, a health conscious tenant might want to know which district has the most gyms and the most salad places? A fan of the movies might be interested to know which district has the most movie theaters. In this data science project, we aim to answer some of these questions and try to narrow down the choices for any given tenant by grouping the neighborhoods into clusters using the k-means clustering algorithm.

2 Data

A list of neighborhoods of Singapore can be found in:

https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore

The list of neighborhoods are scraped from this wikipedia page, and the latitudes and longitudes of these neighborhoods are obtained from the geopy package. Venues with a 500 meter radius around these latitude and longitudes are obtained from the Foursquare API and put into a dataframe. It turns out that expanding the search radius from 500 meters to 2 kilometers does not have much impact on the most frequent venue categories for each neighborhood. This is because in each neighborhood, a lot of venues are located near the Mass Rapid Transport station (the Singaporean equivalent of the subway), while the rest of the neighborhoods are mostly residential areas. Therefore, the search radius used in this project will be fixed at 500 meters.

For each of these venues, the Foursquare API gives the following information

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue Name
- Venue Latitude
- Venue Longitude
- Venue Category

With this dataframe of neighborhoods, the neighborhoods can be ranked by the number of venues for a given venue category. For example, the neighborhoods can be ranked by the number of gyms. This information would be interesting to an prospective tenant who is a gym nut, for example.

3 Methodology

We begin this section by describing the exploratory data analysis performed on the data set.

3.1 Exploratory Data Analysis

Before solving the main problem of ranking neighborhoods by venue categories, let us explore the data obtained from Foursquare to get a sense of the neighborhoods and the kinds of venues contained in them.

3.1.1 Neighborhoods of Singapore

There are a total of 55 planning areas in Singapore which we take to be our neighborhoods. The neighborhoods are arranged in alphabetical order so that the first five neighborhoods in our dataframe are "Ang Mo Kio", "Bedok", "Bishan", "Boon Lay" and "Bukit Batok".

A map of Singapore along with these fifty five neighborhoods are shown in figure 1. The neighborhoods are spread out throughout the country although there is a higher density in the central part of Singapore. There are a few neighborhoods in the periphery of the country that are located far away from the urban areas. It is expected that these neighborhoods will be drastically different from the rest since they will contain atypical venues. One of these neighborhoods corresponds to the rightmost marker on the map and is called "North-Eastern Islands". The bottom two markers in 1, from left to right, are the "Western Islands" and "Southern Islands" neighborhoods respectively, and are also likely to be very different from the neighborhoods on the main island.

As evident from the map figure 1, there is a higher density of neighborhoods in the central region of Singapore. This area includes neighborhoods like "Singapore River" and "Orchard". The central business district of Singapore is located in this central part of Singapore and hence contains many offices. There are also lots of shopping malls in this area.

3.1.2 Venues

A call to the Foursquare API with the "explore" query with a search radius of 500 meters yields a dataframe containing 1647 venues. The first piece of information that can be extracted from this dataframe is the number of venues in each neighborhood. There are a few insights that can be gleaned from this data. First, the neighborhoods with the most venues found are "Orchard" and "Singapore River" with a 100 venues each. As mentioned in the previously, these two neighborhoods are located in the central region. However, there are other neighborhoods with large numbers of venues as well outside of this central region, like "Tampines", a neighborhood that is located in the east with 86 venues, as well as "Jurong East", a neighborhood in the west with 72 venues. These neighborhoods serve as regional hubs so that Singaporeans living in that area do not have to travel all the way downtown to the central region to have access to a wide range of

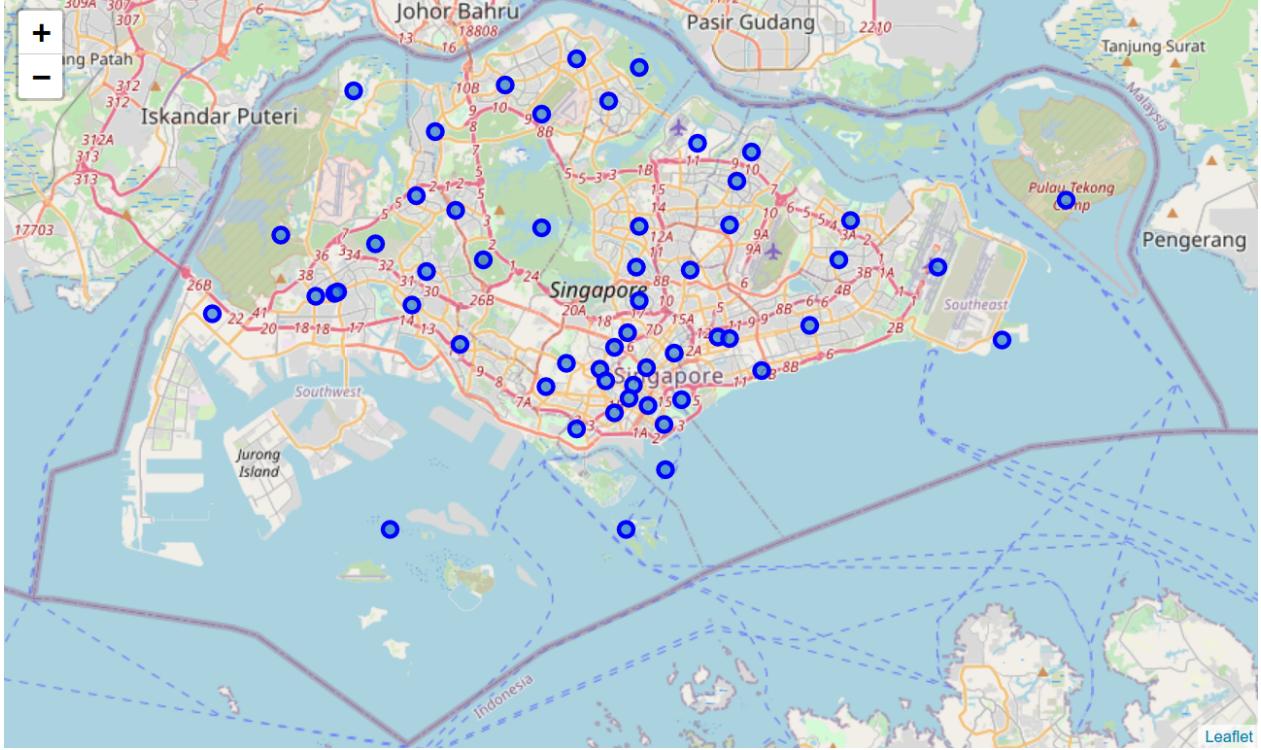


Figure 1: A map of Singapore along with the territorial boundary shown as a grey line. Parts of its northern neighbor Malaysia can also be seen in this picture. Plots of the fifty five Singaporean neighborhoods that are analyzed in this project are denoted with the blue markers.

Rank	Venue Category	Number of venues
1	Chinese Restaurant	84
2	Coffee Shop	81
3	Food Court	71
4	Japanese Restaurant	70
5	Asian Restaurant	59

Table 1: Top five most common venue categories

venue categories. There are also many other neighborhoods scattered throughout the country with comparable numbers of venues.

There are 4 "neighborhoods" that do not have any venues showing up on the foursquare API, like the aforementioned "North-Eastern Islands", "Western Islands" and "Southern Islands". Several other neighborhoods like "Mandai", "Simpang", "Straits View" and "Western Water Catchment" only have 1 venue. These neighborhoods with too little venues will not be relevant in our analysis. The low number of venues can be attributed to the fact that these areas are located away from urban areas and hence are not of any interest to a potential tenant looking for a place to stay in Singapore.

3.1.3 Frequency of Venues

Another interesting piece of information to analyze is the most common kinds of venues there are in Singapore as well as the most commons ones that are available in each neighborhood. From the data frame, the top five most common venue categories are The top five most common venues that were found by the Foursquare API are shown in table 1. All these venues are in the food and beverage industry. Singapore has a vibrant food culture with many food and beverage businesses and this is evident from the data. Table 1 also shows that East Asian cuisine is the most common kind that is available in Singapore.

The types and frequencies of different venues vary from one neighborhood to another. For example, in

Rank	Venue Category
1	Hotel
2	Japanese Restaurant
3	Nightclub
4	Thai Restaurant
5	Italian Restaurant

Table 2: Top five most common venue categories at the Singapore River.

Rank	Venue Category
1	Café
2	Bakery
3	Coffee Shop
4	Sushi Restaurant
5	Shopping Mall

Table 3: Top five most common venue categories at Tampines.

the centrally located neighborhood of "Singapore River", the top five most common venues are shown in table 2.

The Singapore River is a place for tourists and locals alike to go for entertainment and is not really a residential district, hence there are more restaurants and nightclubs in this area. Contrast this with a regional hub in the east, Tampines, with its top five most common venue categories shown in table 3.

Tampines is a residential district, so there are far fewer hotels or nightclubs here. Instead, it is more common to find cafes, bakeries and coffee shops here.

3.1.4 Neighborhoods with the most gyms

In this section, we discuss some useful results that we can obtain from our analysis. Firstly, let us look at the neighborhoods with the most gyms. This information would be useful to a prospective tenant who wishes to live in an area with a large number of gyms to choose from. The neighborhoods with more than one gym are shown in table 4.

The neighborhood with the most gyms is Tampines. The next two neighborhoods, Singapore River and Downtown Core, are not residential areas. Therefore, for a client who wishes to live in a place with more gyms, Tampines would be the top recommendation, although other neighborhoods like Ang Mo Kio, Pioneer and Clementi are not too far behind. A plot of these neighborhoods is shown in figure 2.

3.1.5 Neighborhoods with the most malls

This analysis can also be performed for a client who wishes to live in a place with many shopping malls. The top eight neighborhoods with the most shopping malls are shown in table 5.

The two neighborhoods with the most shopping malls, Paya Lebar and Orchard, are centrally located, with the former being primarily a residential district while the latter is not. There are other neighborhoods that are not too far behind as well, like Tampines and Jurong East. A map indicating these neighborhoods is shown in figure 4.

Rank	Neighborhood	Number
1	Tampines	4
2	Singapore River	3
3	Downtown Core	3
4	Ang Mo Kio	2
5	Pioneer	2
6	Clementi	2

Table 4: The Singaporean neighborhoods with the most gyms.

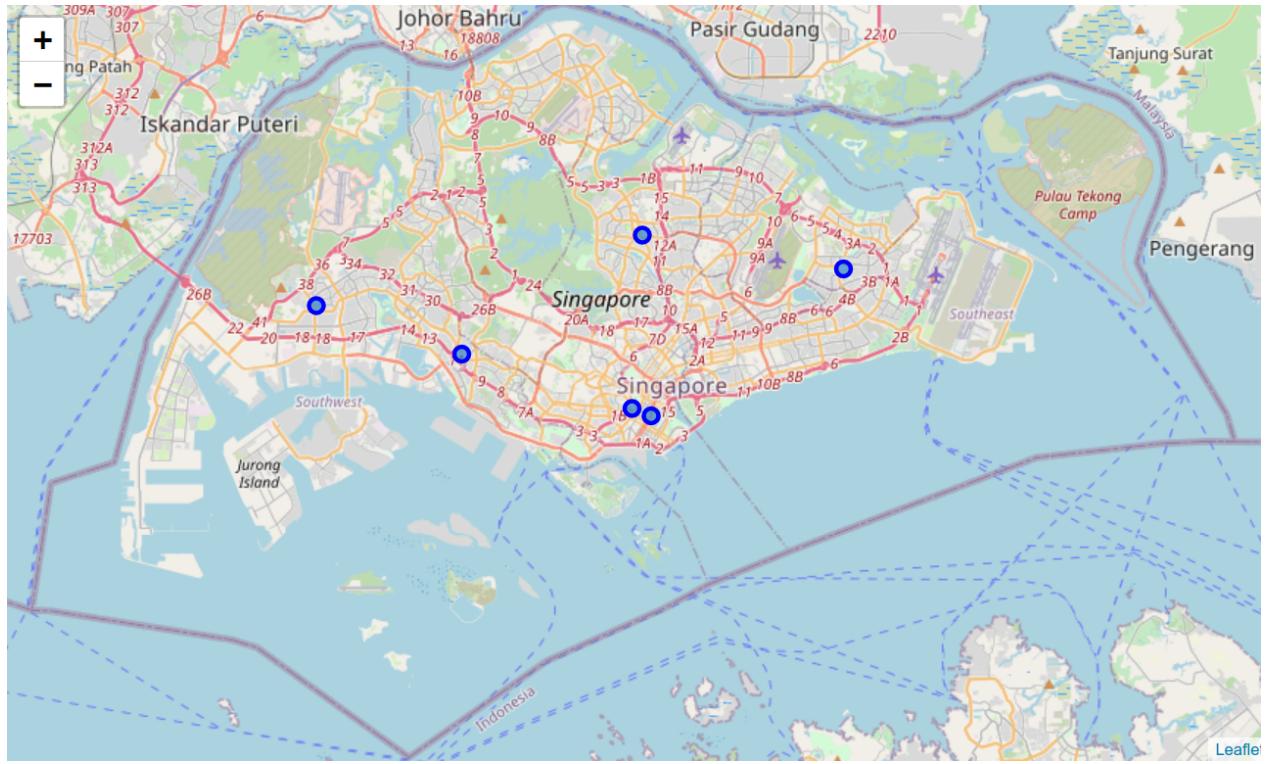


Figure 2: A map of Singapore with the neighborhoods containing the most gyms (table 4) depicted by the blue markers.

Rank	Neighborhood	Number
1	Paya Lebar	6
2	Orchard	4
3	Tampines	3
4	Jurong East	3
5	Sembawang	2
6	Bukit Panjang	2
7	Woodlands	2
8	Clementi	2

Table 5: The Singaporean neighborhoods with the most shopping malls.

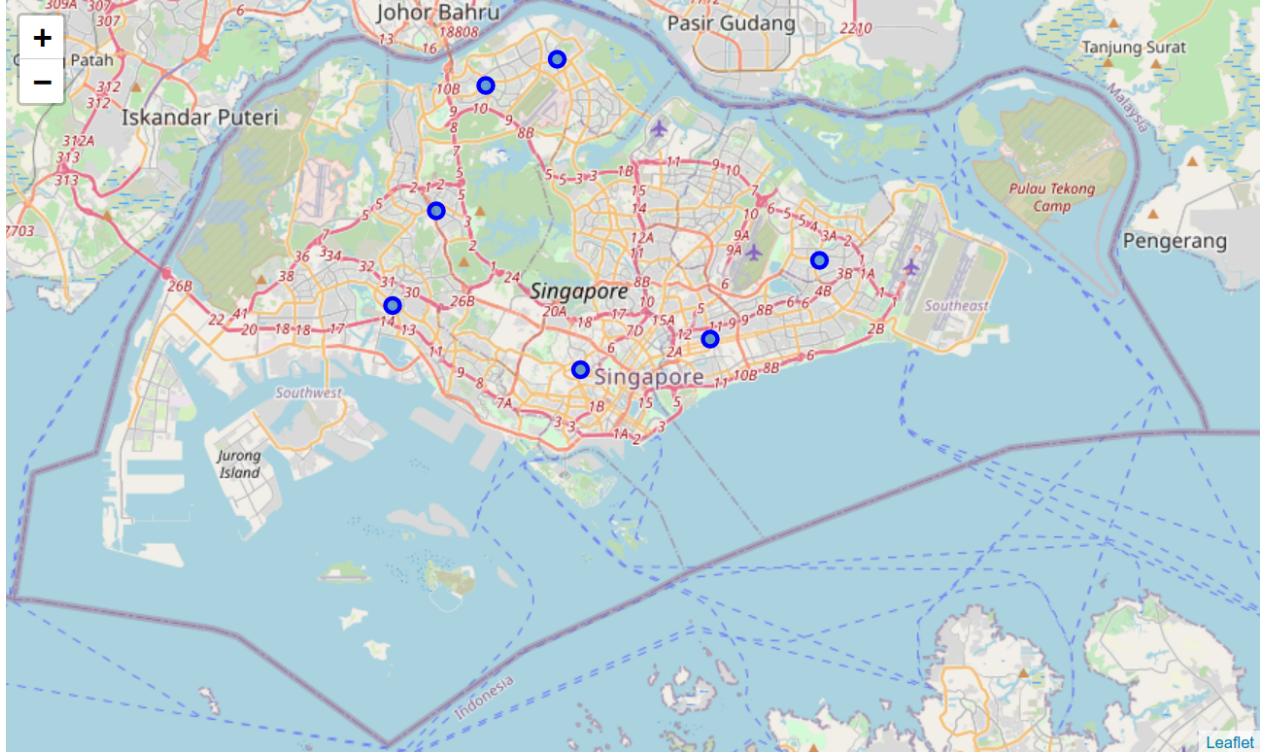


Figure 3: A map of Singapore with the neighborhoods containing the most shopping malls (table 5) depicted by the blue markers.

3.1.6 Neighborhoods with the most coffee shops

Let us conclude our exploratory data analysis by looking at the neighborhoods with the greatest number of coffee shops. The neighborhoods with more than 3 coffee shops are shown in table 6. Unlike gyms, there are many neighborhoods with multiple coffee shops. This is because coffee shops are the second most common venue in Singapore as shown in table 1.

The location of these neighborhoods are shown in figure 4. They are dispersed throughout the country. As any Singaporean or tourist could attest, these coffee shops can be found all over the country.

3.2 Machine Learning Methods

In this project, a k-means cluster algorithm is employed to group similar neighborhoods together. The grouping will help any potential tenants to divide up the large number of neighborhoods into a few group with distinct qualities. Such a clustering will be invaluable in choosing a neighborhood to live in.

We will use the KMeans function that comes readily available in the Sci-Kit Learn package. The random

Rank	Neighborhood	Number
1	Novena	6
2	Jurong East	6
3	Clementi	6
4	Bedok	4
5	Bishan	4
6	Bukit Batok	4
7	Ang Mo Kio	4
8	Toa Payoh	4
9	Tampines	4

Table 6: The Singaporean neighborhoods with the most coffee shops.

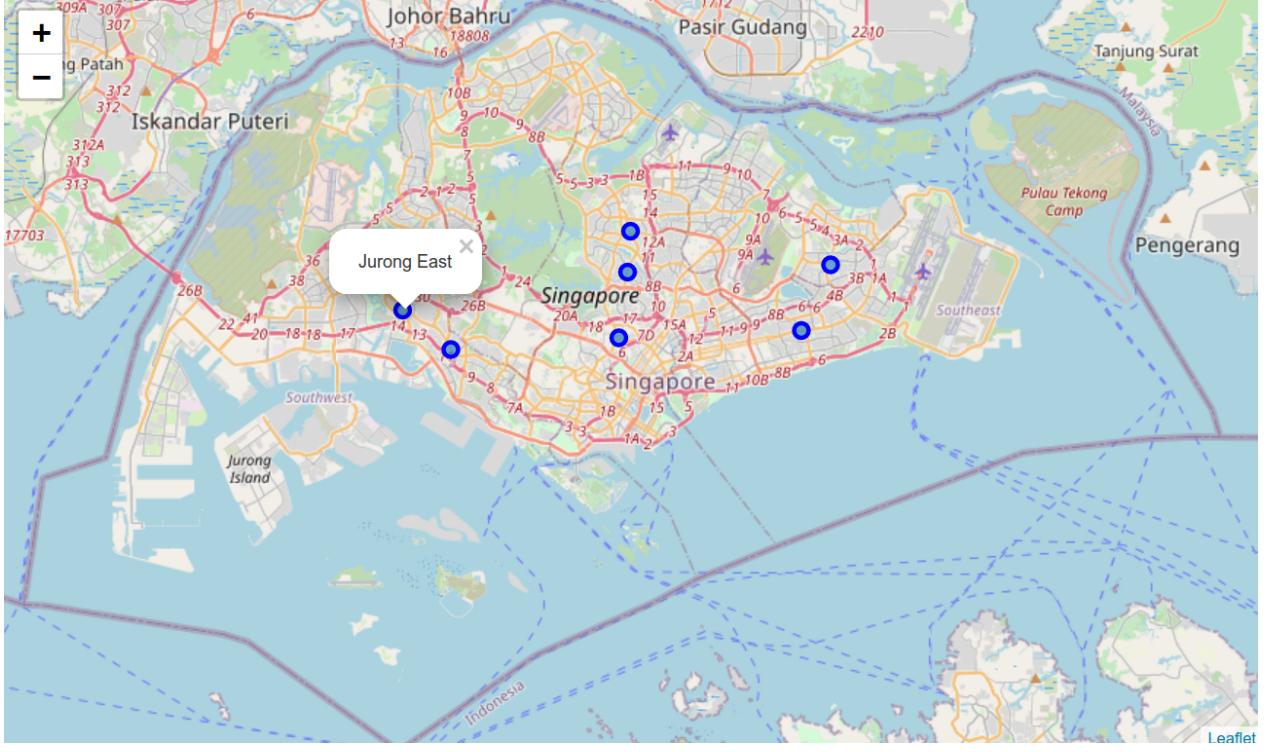


Figure 4: A map of Singapore with the neighborhoods containing the most shopping malls (table 6) depicted by the blue markers.

state is fixed at 0 so that the seed for the initial cluster positions are always fixed and the analysis is reproducible.

Naively running the k-means clustering will result in a single cluster with almost all the neighborhoods and a few outliers occupying their own individual clusters. Such a clustering is useless to any potential tenant. To overcome this, we can remove the outlier neighborhoods from our dataset. These outlier neighborhoods are Mandai, North-Eastern Islands, Simpang, Southern Islands, Straits View, Tengah, Western Islands, Western Water Catchment, Bukit Timah, Central Water Catchment, Marina East, Tuas, Changi, Changi Bay, Kallang, Lim Chu Kang, Marina South, Rochor, Seletar, Sungei Kadut, and Jurong West. Most of these neighborhoods are not residential areas hence will have very unique venue categories. For example, the Western Water Catchment and the Central Water Catchment are reservoirs designed to store water that is collected from the rain that is then used for consumption after processing. Tengah has an airbase and Seletar and Changi have airports.

After removing the outlier neighborhoods, the last potential subtlety to deal with is the choice of the number of clusters. If the number of clusters is too large, there will be several clusters that contain only single neighborhood. Such clusters provide no useful information. On the other hand, there should be enough clusters to divide up their neighborhoods so that each cluster will have a distinctive character. To determine an optimal number of clusters, the k-means clustering algorithm is performed with both 4 and 5 clusters. The number of neighborhoods in each cluster for both choices are shown in table 5. When there are only four clusters, the largest cluster 2 has more than twice the number of neighborhoods than the next largest cluster, 0. The largest cluster is also almost four times larger than the third largest cluster. On the other hand, when the number of clusters is fixed at 5, the largest cluster is almost as large as the second largest cluster and only twice as large as the third largest cluster. Therefore, setting the number of clusters to five results in a more even distribution of neighborhoods amongst the clusters. Using a larger number of clusters will result in a few clusters containing a single outlier neighborhood. For the rest of project, the number of clusters will be fixed at five.

Cluster	Number of Neighborhoods
2	19
0	8
3	5
1	2

Cluster	Number of Neighborhoods
1	13
0	10
4	6
2	4
3	1

Figure 5: The number of neighborhoods in each cluster when there are four (**left**) and five (**right**) clusters respectively.

Cluster	Neighborhoods
0	Boon Lay, Marine Parade, Museum, Orchard, Punggol, River Valley, Singapore River, Tampines, Tanglin, Woodlands
1	Ang Mo Kio, Bedok, Bishan, Bukit Merah, Bukit Panjang, Clementi, Downtown Core, Jurong East, Novena, Pasir Ris, Paya Lebar, Serangoon, Yishun
2	Geylang, Newton, Queenstown, Toa Payoh
3	Outram
4	Bukit Batok, Choa Chu Kang, Hougang, Pioneer, Sembawang, Sengkang

Table 7: The five clusters and their neighborhoods.

4 Results

After dropping the outlier neighborhoods, there are 34 neighborhoods left. The result of clustering them into five groups is shown in table 7. The location of the neighborhoods in these five clusters are shown in figure 6.

To understand why the neighborhoods are clustered this way, it helps to look at the top five most common venues that occur in each cluster. The top five most common clusters are shown in table 8. All venues that tie for the fifth spot are also shown in the table. The number of venues in each of these categories that are present in each of these clusters are shown in the parenthesis.

5 Discussion

We begin this section by trying to understand the result of the k-means clustering algorithm. Keep in mind that the irrelevant neighborhoods for any prospective tenant have been removed so the remaining neighborhoods do have viable residential spots. The first thing to note about the map in figure 6 is that cluster 0, which is depicted in red, is heavily concentrated in the central part of Singapore. This area contains the central business district of Singapore and also contains many places of interests to tourists. This explains why there are 21 hotels in this cluster as seen in table 8, making up the third largest venue category in cluster 1. What is interesting to note is that there are also neighborhoods in the periphery of the island

Cluster	Top Five Most Common Venues
0	Japanese Restaurant (41), Café (22), Hotel (21), Chinese Restaurant (20), Bakery (20)
1	Coffee Shop (43), Food Court (38), Chinese Restaurant (30), Asian Restaurant (24), Café (23)
2	Chinese Restaurant (21), Noodle House (14), Food Court (12), Seafood Restaurant (6), Asian Restaurant (5), Coffee Shop (5)
3	Spanish Restaurant (1), Seafood Restaurant (1), Nail Salon (1), Hotel (1), Grocery Store (1), Gastropub (1), Farmers Market (1), Dim Sum Restaurant (1), Coffee Shop (1), Café (1), Bakery (1)
4	Coffee Shop (13), Fast Food Restaurant (12), Bus Station (12), Chinese Restaurant (11), Food Court (11)

Table 8: The top five most common venues in each of these clusters with the number of occurrences listed in the parenthesis.

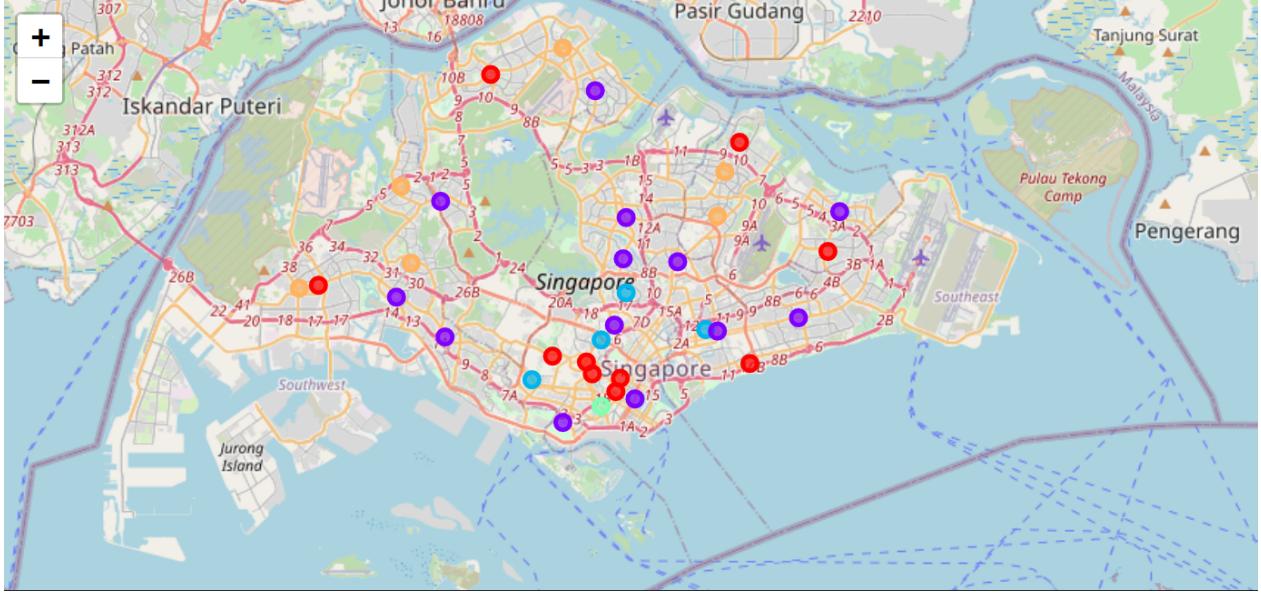


Figure 6: Map of Singapore along with the relevant neighborhoods grouped into clusters 0 (red), 1 (purple), 2(teal), 3 (green) and 4 (orange).

that are also lumped into this cluster. One reason is that some of these places serve as regional hubs so that Singaporeans living in the vicinity of these hubs do not have to travel all the way down to the central region. For example, the easternmost neighborhood in cluster 0 is Tampines which serves as a hub in the east, containing multiple shopping malls as well as a large sports hub. The cluster 0 neighborhood in the northeastern part of Singapore is Punggol, a place that has seen a lot of recent developments and is a nice place with new apartments for young couples to move to. That these neighborhoods are grouped by the k-means clustering algorithm with the centrally located hot spots makes perfect sense. It is also worth noting that the top three neighborhoods with the most gyms fall in this cluster as seen in table 4. We can think of cluster 0 as the "exciting neighborhoods".

The next largest cluster is cluster 1, indicated by the purple markers in figure 6. These neighborhoods are scattered throughout Singapore with a fairly uniform density. These neighborhoods are basically the "regular neighborhoods". While they might not have as many special and exciting venues as cluster 1, they still have many things to offer and are decent places to live. The most common venues in cluster 1 are coffee shops and food courts, with 43 and 38 of them respectively. These are places where regular Singaporeans visit on a daily basis, hence there is a proliferation of these venues in these "regular neighborhoods".

The third largest cluster is cluster 2, consisting of only four neighborhoods. These neighborhoods are depicted with the teal markers in figure 6. One key feature of this cluster is that they are centrally located and yet the k-means clustering algorithm excluded them from cluster 0. While centrally located, these neighborhoods are not as new and gentrified as those in cluster 0. In fact, Queenstown and Toa Payoh are the first and second oldest satellite towns [1] while Geylang is notorious for being a red-light district [2]. The difference between cluster 0 and cluster 2 is evident from their top five most common venues shown in table 8. Only "Chinese Restaurant" appears in the top five for both of these clusters which is not entirely surprising since chinese restaurants are the most common venue in Singapore as seen in table 1 and therefore is likely to show up in almost every cluster. To summarize, cluster 2 should be thought of as the "centrally located old neighborhoods".

Cluster 3 is interesting as it only contains a single neighborhood, Outram. The reason is that the selection of venues in Outram is rather different from the rest as seen in table 8. It does not contain a single chinese restaurant or food court, two of the most common venues in Singapore as seen in table 1. In fact, the only venue in table 1 that it contains is a single coffee shop. Instead, Outram has a spanish restaurant, a gastropub as well as a farmers market. Such venues are rather unique in Singapore. The unique variety of venues in Outram explains why it was put into its own cluster.

Finally, there is cluster 4 which has 6 neighborhoods. Like cluster 2, these neighborhoods are also not gentrified. Their most common venue are coffee shops, just like cluster 2. However, they are located far away

from the central region of Singapore, unlike cluster 2. It is therefore appropriate to think of cluster 4 as the "ungentrified neighborhoods located away from the city center".

The grouping of neighborhoods has led to a clustering that would make sense to any native Singaporeans. Each cluster of neighborhoods have a rather distinctive flavor to them. For a prospective tenant who wishes to live in a gentrified location with lots of exciting venues, the neighborhoods in cluster 0 are a clear choice. A tenant who wishes to live in a regular neighborhood which likely comes with a more regular rent could pick one of the neighborhoods in cluster 1. The neighborhoods with less gentrification are in clusters 2 and 4, with cluster 2 being centrally located and cluster 4 being located far away from the central region. These neighborhoods maybe the most affordable option for tenants on a tighter budget.

6 Conclusion

The goal of this project is to provide useful information to prospective tenants who are trying to figure out which part of Singapore they might want to live in. Given the large number of neighborhoods, it is helpful to try to group the neighborhoods into clusters of similar neighborhoods. We begin by performing some exploratory data analysis. With the location data obtained from Foursquare, we created a dataframe containing information of 1676 venues along with their neighborhoods. With this dataset, we could obtain all kinds of information, like the neighborhoods with the most gyms for example.

After the preliminary data analysis, we performed k-means clustering on the neighborhoods based on the frequencies of the venue categories occurring in them. The k-means algorithm grouped the neighborhoods into five clusters which can be thought of as the "exciting neighborhoods", "regular neighborhoods", "centrally located old neighborhoods", "Outram" and "ungentrified neighborhoods located away from the city center". A prospective tenant can then narrow down his selection of neighborhoods based on his budget and preference of venues.

References

- [1] Toa payoh. https://en.wikipedia.org/wiki/Toa_Payoh. Accessed: 2021-09-05.
- [2] Geylang. <https://en.wikipedia.org/wiki/Geylang>. Accessed: 2021-09-05.