Exploring towns in Singapore using Foursquare API

Liza Chua

Table of Contents

I. Introduction

In Singapore, vast majority of the dwelling units are spread across 23 towns and 3 estates. Each town or estate is designed such that it can be self-contained with living, education, healthcare, and recreational needs. When someone is looking for a place to stay in Singapore, the question will be which town/estate is the best to live in?

To determine the ideal town or estate to stay, the available amenities within the town or estate is one of the common factors to consider.

Therefore, in this project, the objective is to investigate in the amenities within each town and estate. The goal is to give an insight of the most common amenities that can be found in each town and estate, as well as which town and estates are similar in terms of the available amenities.

The findings might be interesting to the target audiences as follows:

1. An individual/family looking for a place to stay in Singapore
2. Property agents/potential customers looking to buy/sell their property
3. Property developers doing market study for potential development areas

II. Data Sources

The data that is required for this project is as follows:

1. Singapore town and estates names and total area

   The list of Singapore town and estates is retrieved from Wikipedia page: https://en.wikipedia.org/wiki/New_towns_of_Singapore

2. Geographical location of each town and estates

   Data location of each town and estates is retrieved using GeoPy's Nominatim API based on the names of the town and estate.

3. Amenities in each town and estates

   The Foursquare API will be used to explore the common amenities that can be found within the radius of each town and estate. The explore function will be used to obtain the top 100 venues within the defined radius of each town and estate.

III.Methodology

1. Web-scraping

The list of Singapore towns and estates is obtained through web-scraping from Wikipedia using the BeautifulSoup package. The data field of interest in this project are the names of the towns and estates, as well as the total area, and only these relevant data fields are extracted. The extracted data are then presented in a data frame for exploration. There are a total of 23 towns and 3 estates in Singapore and the top 5 rows of the data frame is as shown in Fig. 1.

	Town/Estate	Area
0	Ang Mo Kio	6.38
1	Bedok	9.37
2	Bishan	6.90
3	Bukit Batok	7.85
4	Bukit Merah	8.58

Figure 1: Top 5 rows of data frame

2. Geocoding

With the list of the towns and estates, geocoding is performed to obtain the geographical coordinates of each town and estate. The coordinates of the town/estate are added to the data frame as shown in Fig. 2. Folium is then used to visualize the markers of each town and estate on the OpenStreetMap according to the geographical coordinates as shown in Fig. 3. A visual check on the location of the markers indicate that the geographical locations data are acceptable before moving to the next step.

	Town/Estate	Area	Latitude	Longitude
0	Ang Mo Kio	6.38	1.37161	103.84546
1	Bedok	9.37	1.32425	103.95297
2	Bishan	6.90	1.35079	103.8511
3	Bukit Batok	7.85	1.34952	103.75277
4	Bukit Merah	8.58	1.28306	103.81669

Figure 2: Data frame with geographical coordinates of town

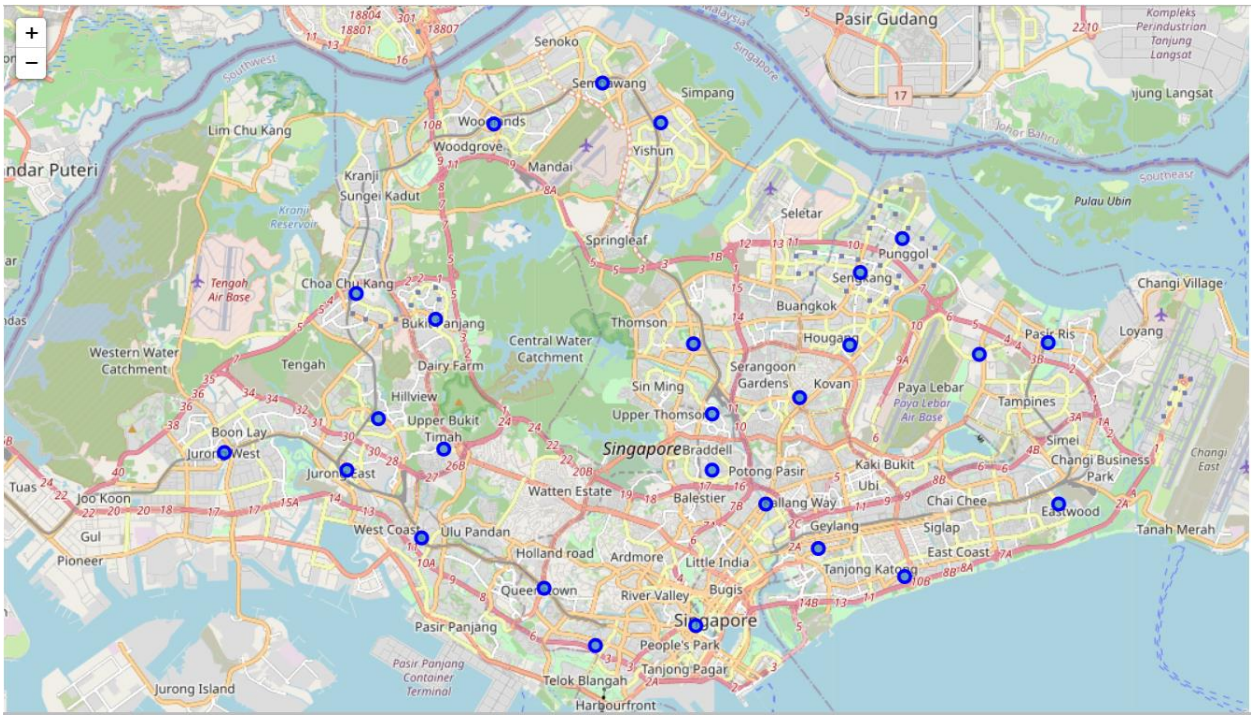

Figure 3: OpenStreetMap with geographical markers of each town/estate

3. RESTful API

Foursquare API explore function is then used to retrieve the information on the venues within each town and estate. The key information required for the request is geographical coordinates of each town, radius and the maximum number of venues to retrieve. The geographical coordinates are provided from data obtained through geocoding and the default maximum number of venues that can be retrieved is 100 for the free license from Foursquare.

The area of the town/estates varies and in order to capture data that is most representative of the whole area of the town/estates, the radius for each town/estate is individually defined according to the total area of the town/estate, assuming that area of each town/estate is circular. The calculated radius are added to the data frame as shown in Fig. 4.

	Town/Estate	Area	Latitude	Longitude	Radius (m)
0	Ang Mo Kio	6.38	1.37161	103.84546	1425.067393
1	Bedok	9.37	1.32425	103.95297	1727.010027
2	Bishan	6.90	1.35079	103.8511	1482.004796
3	Bukit Batok	7.85	1.34952	103.75277	1580.737994
4	Bukit Merah	8.58	1.28306	103.81669	1652.60365

Figure 4: Radius of each town/estate

4. Feature Engineering

The venues obtained are then grouped by the towns/estate and one-hot encoding is used to convert the categorical variables to numerical values. The mean of frequency of occurrence of each category is calculated to find the top 10 venues in each town/estate. An example of the top 10 most common venues of first 5 rows of the towns are shown in Fig. 5.

	Town/Estate	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	Food Court	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Park	Supermarket	Noodle House	Snack Place	Bubble Tea Shop	Sandwich Place
1	Bedok	Food Court	Café	Chinese Restaurant	Coffee Shop	Noodle House	Malay Restaurant	Sporting Goods Shop	Japanese Restaurant	Seafood Restaurant	Dessert Shop
2	Bishan	Food Court	Chinese Restaurant	Coffee Shop	Seafood Restaurant	Café	Asian Restaurant	Fast Food Restaurant	Supermarket	Thai Restaurant	Bubble Tea Shop
3	Bukit Batok	Food Court	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Sandwich Place	Supermarket	Convenience Store	Indian Restaurant	Grocery Store	Italian Restaurant
4	Bukit Merah	Chinese Restaurant	Food Court	Café	Coffee Shop	Park	Bakery	Asian Restaurant	Ice Cream Shop	Japanese Restaurant	Sandwich Place

Figure 5: Top 10 most common venue in each town/estate

5. Unsupervised learning

Unsupervised learning is implemented to determine the similarities among the towns/estates in terms of the available amenities. K-means clustering is chosen as the model and the optimal number of clusters is determined by using a combination of the elbow method and silhouette score.

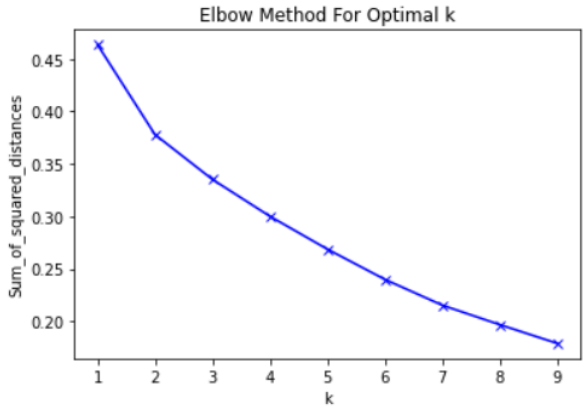

Figure 6: Elbow method

Figure 7: Silhouette score

The elbow method shows the elbow point at k=2, which coincides with the highest silhouette score at k=2. Therefore, the number of clusters chosen is 2. The results and discussions are presented in the next section.

IV. Results and Discussions

The towns/estates in Singapore can be grouped in 2 main clusters as visualized on the OpenStreetMap in Fig. 8.

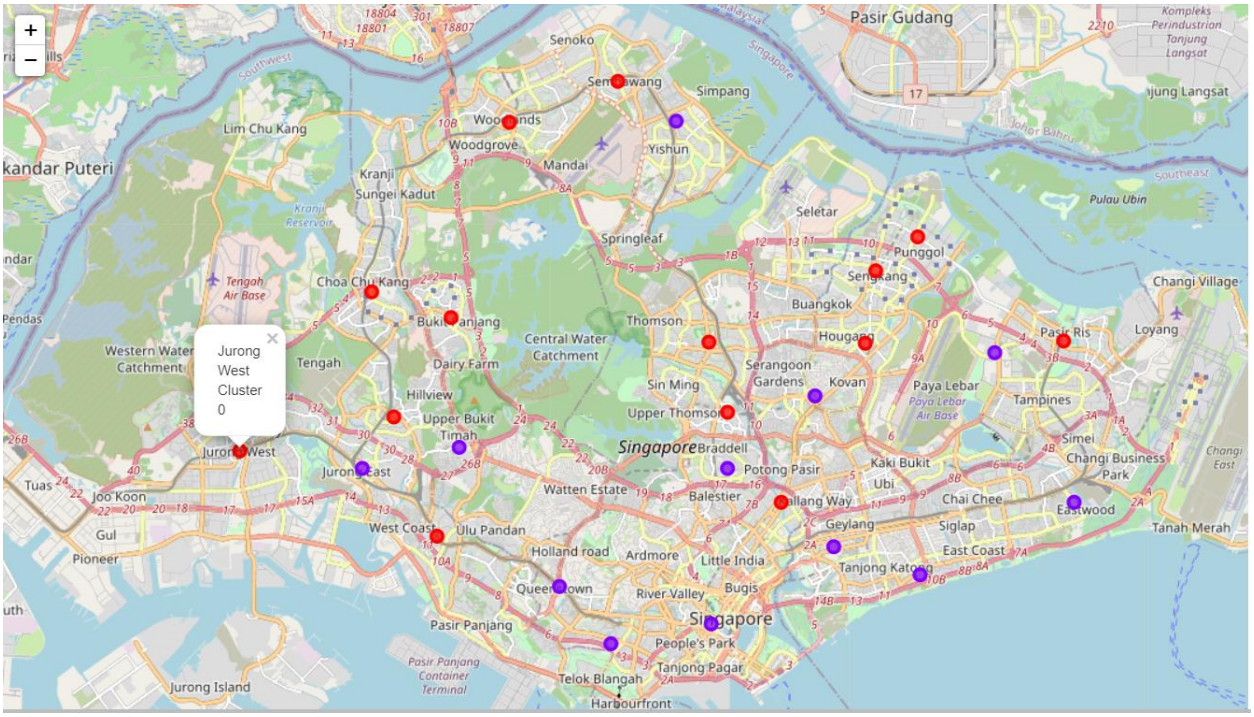

Figure 8: Clusters of town/estates

Based on the results above, it is seen that the clusters are segmented based on certain type of venue that are most common in that each neighborhood. Some observations can be made from each cluster's characteristics.

In Cluster 1, the most common venues are coffee shops, fast food restaurant and food court. There is availability of some restaurants and cafes which are less common. In addition to food venues, there are also availability of supermarkets and sports amenities.

In Cluster 2, the most common venues are Chinese/Asian restaurants and cafes. There is also availability of wide varieties of restaurants such as Indian, Japanese, Italian. Other than food venues, sports, recreational amenities and supermarkets are less common in this cluster and not within the top 10 most common venues for most of the town/estates in this cluster.

V.Conclusion

In this project, the main goal is to give an insight of the most common amenities that can be found in each town and estate in Singapore, as well as the similarities between the town/estates in terms of the available amenities.

The data sources were identified to obtain the list of town/estates in Singapore, area, and geographical coordinates. These information are used to enquire 100 most common venues in each town/estate using Foursquare API. K-means clustering is then used to perform clustering of the towns/estate to two main clusters.

An analysis of the results shows that cluster 1 has cheaper dining places, as well as availability of supermarkets and sports amenities. Cluster 2 presents more costly dining places where the most common venues are restaurants and cafes, but with a wide variety of cuisines choices. These findings can be a good starting point for an individual/family to decide on a suitable town/estate to stay in Singapore.

Future work can investigate in incorporating other factors such as the availability of transport facilities, schools, population density and housing prices to this study.