

The Battle of Neighborhoods

The best location to open a veterinary clinic in Singapore

Wang Zejia
2020/11/16

1. Introduction

1.1 Background

Singapore is one of the most densely populated and most affluent countries in the world. With the improvement in standard of living, demand for pets is also on the rise and is at an all-time high in 2020¹. As a pet owner myself, this poses questions on the availability and accessibility of veterinary clinics in the neighborhoods. On the other hand, potential veterinarians might be interested to find out if there is an opportunity to join the competition in the sector, and if yes, which is the best location to open one in Singapore.

1.2 Problem

This project aims to find out the most suitable location for a potential veterinarian to open a new clinic in Singapore. There are a few factors to consider when determining the best location:

1. Number of existing veterinary clinics in the neighborhoods
2. Accessibility of existing veterinary clinics

2. Data

Data used in the analysis are collected from multiple data sources and later aggregated.

2.1 Neighborhood data of Singapore

“Neighborhood” in this project is defined as the train station in Singapore, which there are 157 stations as of date. The location data of train station is obtained from Kaggle² and contains information on the 157 stations:

1. station_name: name of the train station
2. type: an indication of whether it is an MRT or LRT station
3. lat: the latitude where the station is located
4. lng: the longitude where the station is located

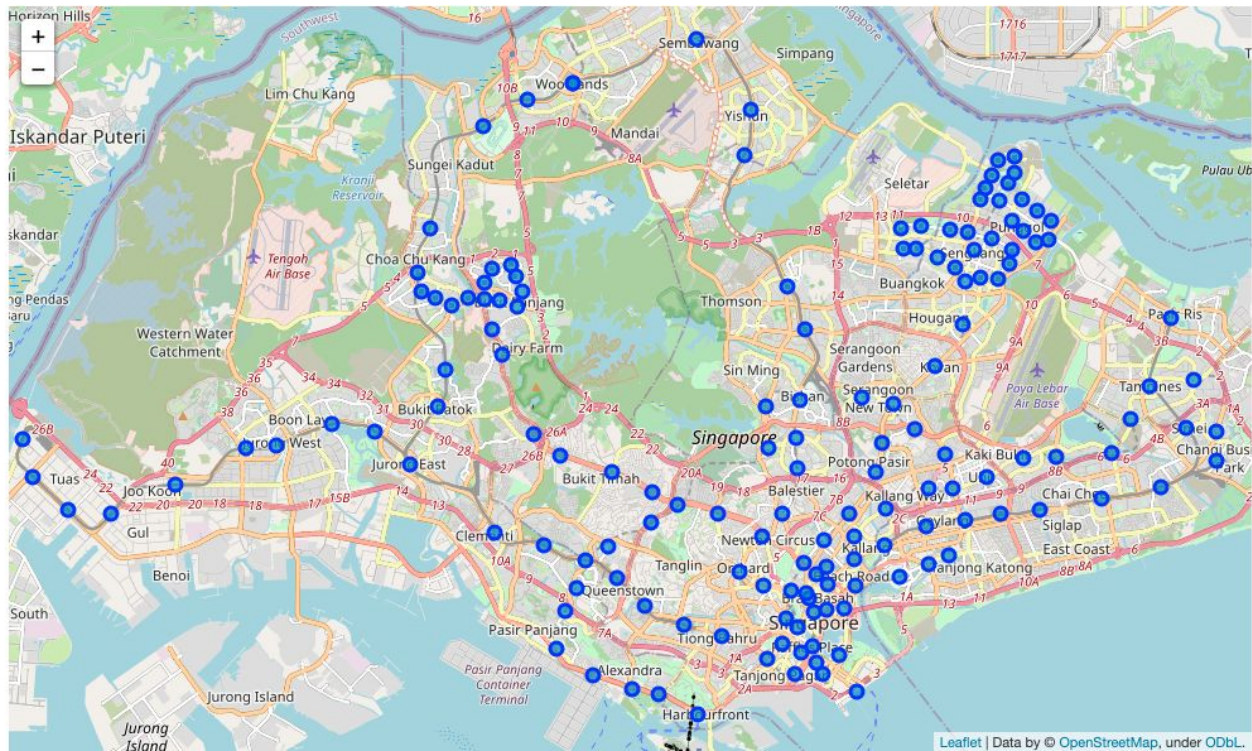
Sample of the dataset is shown below:

	station_name	type	lat	lng
0	Jurong East	MRT	1.333207	103.742308
1	Bukit Batok	MRT	1.349069	103.749596
2	Bukit Gombak	MRT	1.359043	103.751863
3	Choa Chu Kang	MRT	1.385417	103.744316
4	Yew Tee	MRT	1.397383	103.747523

¹<https://www.businesstimes.com.sg/brunch/paws-perous-business-the-booming-pets-trade-thats-also-feeding-an-illicit-market>

² <https://www.kaggle.com/yxlee245/singapore-train-station-coordinates>

The following graph shows the location of each station on the Singapore map.



2.2 Venue data of veterinary clinics

Given the coordinates of each train station, the venue data of existing veterinary clinics are obtained via the Foursquare API³. The following information are collected for all clinics within 2km of each station:

1. vet_id: ID of the venue in Foursquare
2. vet_name: Name of the clinic
3. vet_lat: Latitude of the clinic
4. vet_lng: Longitude of the clinic
5. vet_distance: Distance in meter of the clinic from the train station

The dataset contains 559 rows and a sample is shown below:

	station_name	station_lat	station_lng	vet_id	vet_name	vet_lat	vet_lng	vet_distance
0	Jurong East	1.333207	103.742308	4d89579b7139b1f7a335b4d4	Island Veterinary	1.339389	103.734316	1124
1	Jurong East	1.333207	103.742308	54549ac2498e22fa72234c5b	The Joyous Vet	1.326232	103.724836	2093
2	Jurong East	1.333207	103.742308	4de1f041ae60e7f3abe43f19	Singapore Veterinary Animal Clinic	1.343995	103.730574	1774
3	Bukit Batok	1.349069	103.749596	51091730e4b026ea03ad83fc	My Family Vet	1.349688	103.759894	1148
4	Bukit Batok	1.349069	103.749596	4d89579b7139b1f7a335b4d4	Island Veterinary	1.339389	103.734316	2013

³ <https://developer.foursquare.com/docs/api-reference/venues/search/>

2.3 Data processing

Descriptive statistics of each train station is obtained from the venue dataset, and later combined with the neighborhood dataset to give the following table (157 rows, 8 columns):

1. station_name: Name of the train station
2. station_lat: Latitude of the station
3. station_lng: Longitude of the station
4. vet_count: Number of veterinary clinics within 2km of the station
5. vet_mean_distance: Mean distance (in meter) of veterinary clinics from the station
6. vet_min_distance: Minimum distance (in meter) of veterinary clinics from the station
7. vet_median_distance: Median distance (in meter) of veterinary clinics from the station
8. vet_max_distance: Maximum distance (in meter) of veterinary clinics from the station

	station_name	station_lat	station_lng	vet_count	vet_mean_distance	vet_min_distance	vet_median_distance	vet_max_distance
0	Jurong East	1.333207	103.742308	3.0	1663.666667	1124.0	1774.0	2093.0
1	Bukit Batok	1.349069	103.749596	5.0	1492.600000	972.0	1148.0	2190.0
2	Bukit Gombak	1.359043	103.751863	3.0	925.666667	56.0	1349.0	1372.0
3	Choa Chu Kang	1.385417	103.744316	1.0	1025.000000	1025.0	1025.0	1025.0
4	Yew Tee	1.397383	103.747523	0.0	-2000.000000	-2000.0	-2000.0	-2000.0

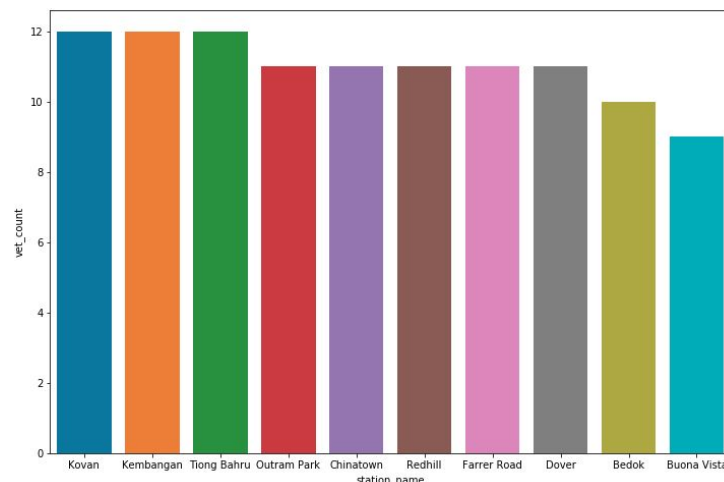
For stations without any existing veterinary clinics (For example, Yew Tee station in row 5), vet_count is set as 0 and distance metrics are set as -2000.

Columns 4-8 will be used as features to cluster neighborhoods in subsequent analysis.

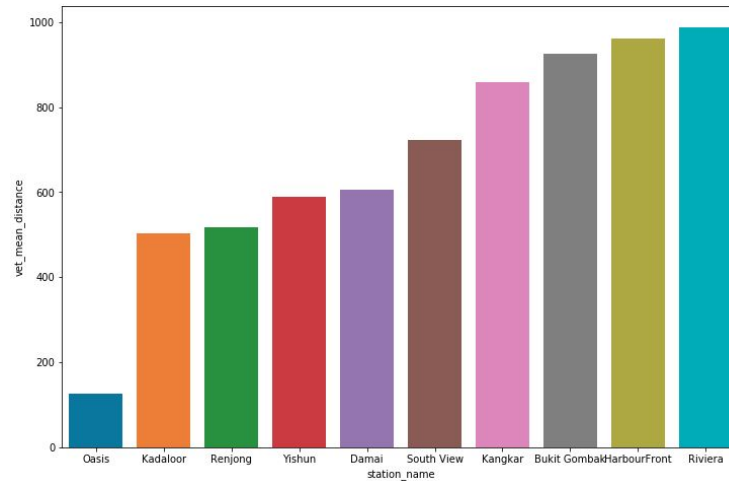
3. Methodology

3.1 Exploratory Data Analysis

Exploratory data analysis is performed to understand the distribution of veterinary clinics in each neighborhood. The chart below shows the top 10 stations with the most number of clinics nearby: Kovan, Kembangan, Tiong Bahru top the chart with 12 clinics in the neighborhood.



In terms of accessibility, Oasis station has veterinary clinics nearest to the station, at an average of less than 200 meters, followed by Kadaloor, Renjong and Yishun.



3.2 Clustering

K-Means algorithm is used to cluster the neighborhoods according to their characteristics pertaining to veterinary clinics nearby. Number of clusters is set at 5, and the features used in the clustering are: vet_count, vet_mean_distance, vet_min_distance, vet_median_distance, vet_max_distance. Cluster labels are appended to the original dataset to obtain the following:

	cluster_label	station_name	station_lat	station_lng	vet_count	vet_mean_distance	vet_min_distance	vet_median_distance	vet_max_distance
0	3	Jurong East	1.333207	103.742308	3.0	1663.666667	1124.0	1774.0	2093.0
1	3	Bukit Batok	1.349069	103.749596	5.0	1492.600000	972.0	1148.0	2190.0
2	0	Bukit Gombak	1.359043	103.751863	3.0	925.666667	56.0	1349.0	1372.0
3	4	Choa Chu Kang	1.385417	103.744316	1.0	1025.000000	1025.0	1025.0	1025.0
4	1	Yew Tee	1.397383	103.747523	0.0	-2000.000000	-2000.0	-2000.0	-2000.0

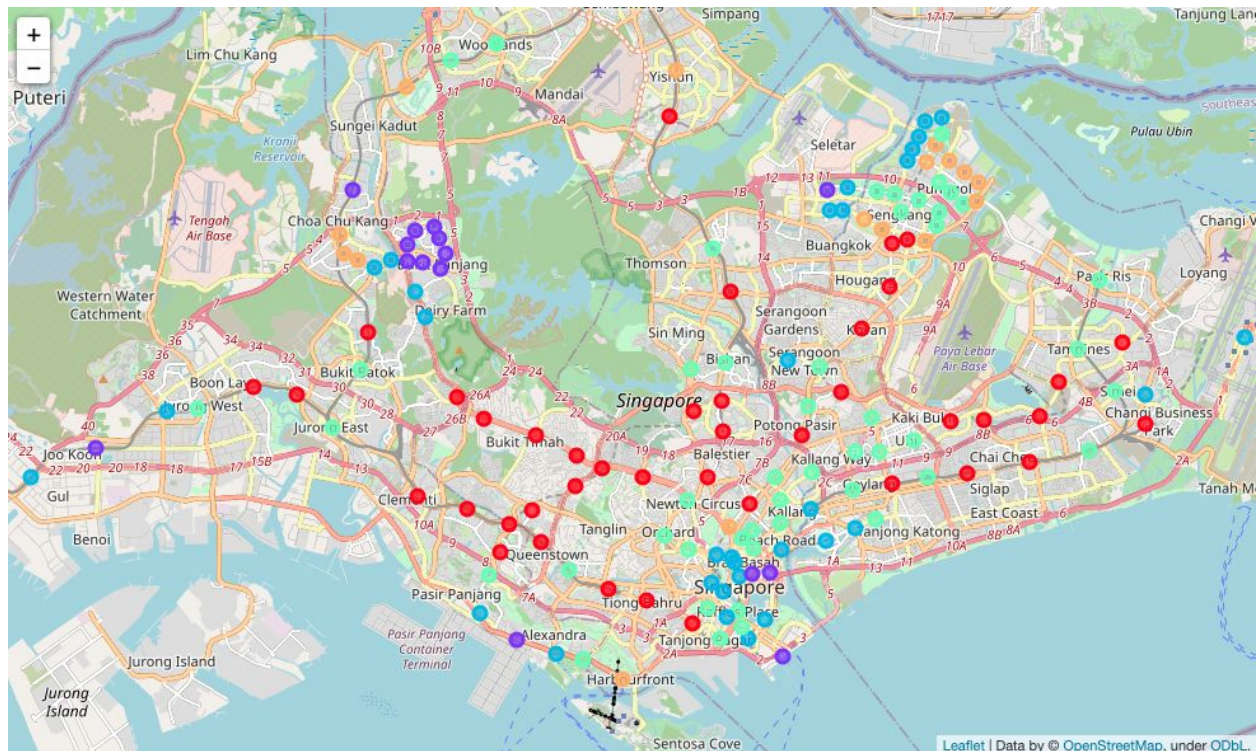
4. Results & Discussion

From the results, each cluster has very distinctive characteristics from each other:

- Cluster 0: Most veterinary clinics present (average count of 6.95) and easily accessible (average distance of 1.37km from the station)
- Cluster 1: No veterinary clinic present within 2km from the station (average count of 0)
- Cluster 2: Moderate number of veterinary clinics present (average count of 2.08) but least accessible (average distance of 1.98km from the station)
- Cluster 3: Moderate number of veterinary clinics present (average count of 3.63) and moderately accessible (average distance of 1.58km from the station)
- Cluster 4: Few veterinary clinics present (average count of 1.56) but most accessible (average distance of 859m from the station)

	vet_count	vet_mean_distance	vet_min_distance	vet_median_distance	vet_max_distance
cluster_label					
0	6.951220	1366.755788	478.487805	1399.134146	2154.780488
1	0.000000	-2000.000000	-2000.000000	-2000.000000	-2000.000000
2	2.088235	1979.222619	1836.852941	1992.470588	2089.705882
3	3.632653	1581.172811	1119.081633	1596.857143	1977.224490
4	1.562500	859.046875	841.625000	857.281250	884.687500

The map below shows the location of each station, and its corresponding cluster. (Cluster 0: red; cluster 1: purple; cluster 2: blue; cluster 3: green; cluster 4: orange)



Given the results, the recommendation on the next best location to open a new veterinary clinic is summarised below:

- The most desirable location to open a new veterinary clinic is in cluster 1, as there is currently no clinic nearby, which implies low competition and unmet demand
- Second best location is in cluster 2, where existing veterinary clinics are least accessible. A new clinic nearer to the station would have competitive advantage over existing ones
- The least desirable location is in cluster 0, where there are already sufficient veterinary clinics to meet the demand

5. Conclusion

In this project, K-Means clustering algorithm is used on location data to find out the best location to open a new veterinary clinic in Singapore. The results and findings would help potential veterinarians to decide where to start their business. The same methodology can be extended to similar topics or other locations, to provide valuable insights to entrepreneurs in different industries and regions.