Private Property Prices and Foursquare Data Clustering Analysis

Sng Jun Qiang

10 Jun 2019

Introduction: Business Problem

Private property prices had been souring recently in Singapore. Demands often form correlation with transportations, amenities, schools etc. Other than the previous mentioned obvious factors that caused the souring prices of private property, this analysis would like to identify if factors such as restaurants, cafes, bars, shopping malls or places of visit related to the prices of private property in Singapore.

Analysis will involve extracting data from open source and Foursquare to analyze the relationships between private property prices in each neighborhood districts and their venues nearby, within 1km, which will provide information for buyers who are considering purchasing private property for personal residential needs or investment purposes.

Description of the Data

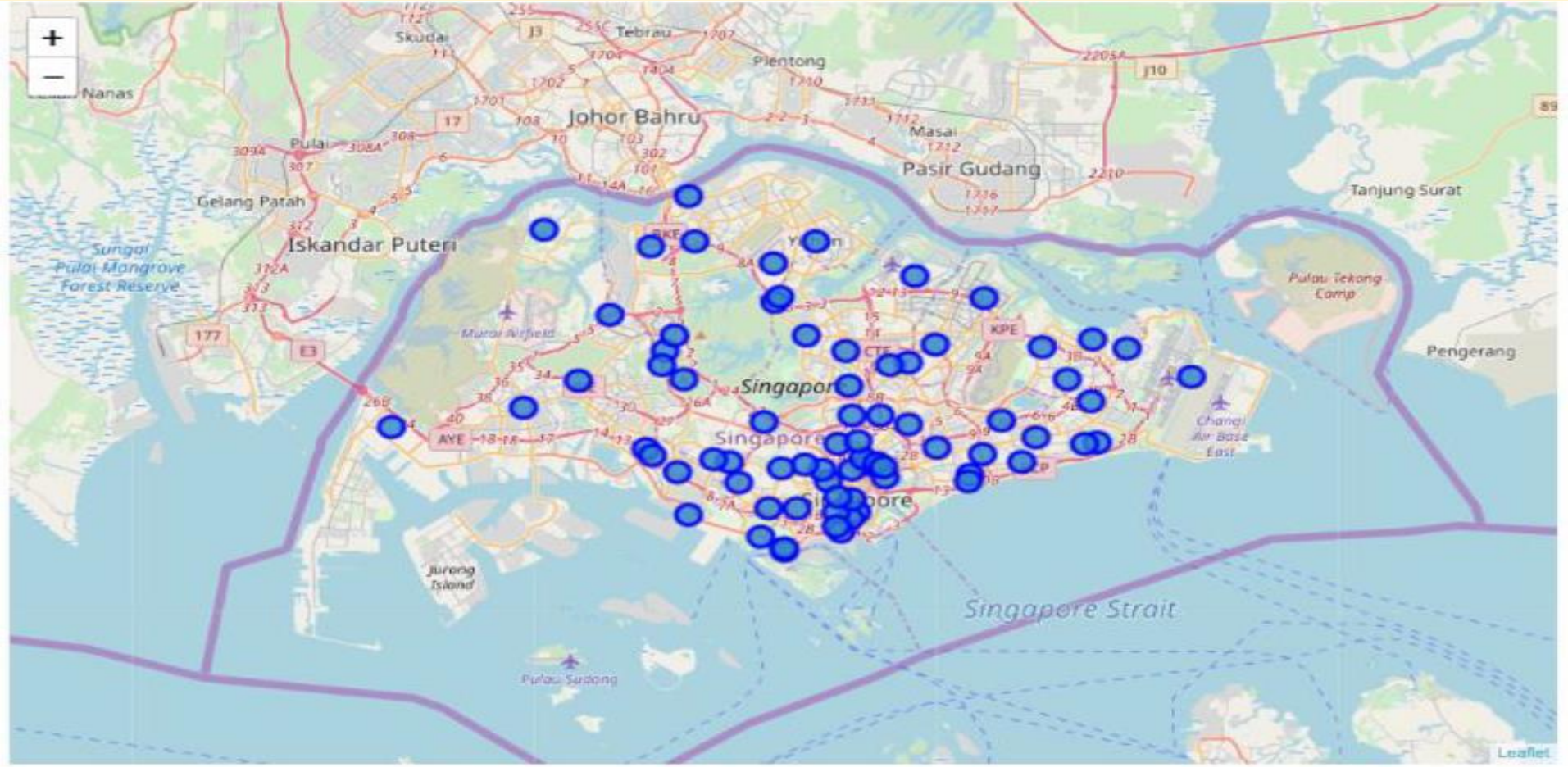
- Neighborhood list and postal district code were retrieved from Wikipedia. Thereafter, the latitude and longitude will be retrieved using geopy and save as a data frame named “sg_latlng”.
- Prices per square feet of private property for 2017 and 2018 will be extracted from Urban Redevelopment Authority (URA) as csv and merge with data frame “sg_latlng” to create data frame “sg_latlng_psf”.
- A function will be created to extract venues nearby from Foursquare to create data frame “sg_venues” and combined with data frame “sg_latlng_psf” to form data frame “sg_venues_psf” for analysis.
- Top venues of each neighbourhoods were retrieved by introducing dummy variables and computing the highest frequency of top venues creating data frame “neighborhoods_venues_sorted”.
- Finally, both data frame “sg_latlng_psf” and “neighborhoods_venues_sorted” were merged to create final data frame “sg_latlng_psf_venues” as shown in the next slide.

Data Frame

	Postal District	Neighborhood	Latitude	Longitude	2017 psf	2018 psf	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
0	1	Raffles Place	1.283595	103.851568	2333.80	2540.88	Hotel	Cocktail Bar	Korean Restaurant	Yoga Studio	Japanese Restaurant	Italian Restaurant	Café	Coffee Shop
1	1	Cecil	1.280805	103.848668	2333.80	2540.88	Hotel	Korean Restaurant	Gym / Fitness Center	Chinese Restaurant	Cocktail Bar	Restaurant	Café	Bar
2	1	Marina	1.263270	103.820293	2333.80	2540.88	Theme Park Ride / Attraction	Clothing Store	Fast Food Restaurant	Coffee Shop	Chinese Restaurant	Spa	Multiplex	Theme Park
3	1	People's Park	1.284139	103.842683	2333.80	2540.88	Hotel	Japanese Restaurant	Cocktail Bar	Café	Chinese Restaurant	Spanish Restaurant	Dessert Shop	Wine Bar
4	2	Anson	1.273845	103.844512	2366.87	3362.38	Japanese Restaurant	Coffee Shop	Café	Ramen Restaurant	Hotel	Bakery	Food Court	Italian Restaurant

Data frame "sg_latlng_psf_venues".

Visualization of Neighborhoods in Singapore



Visualization of Neighbourhoods in Singapore.

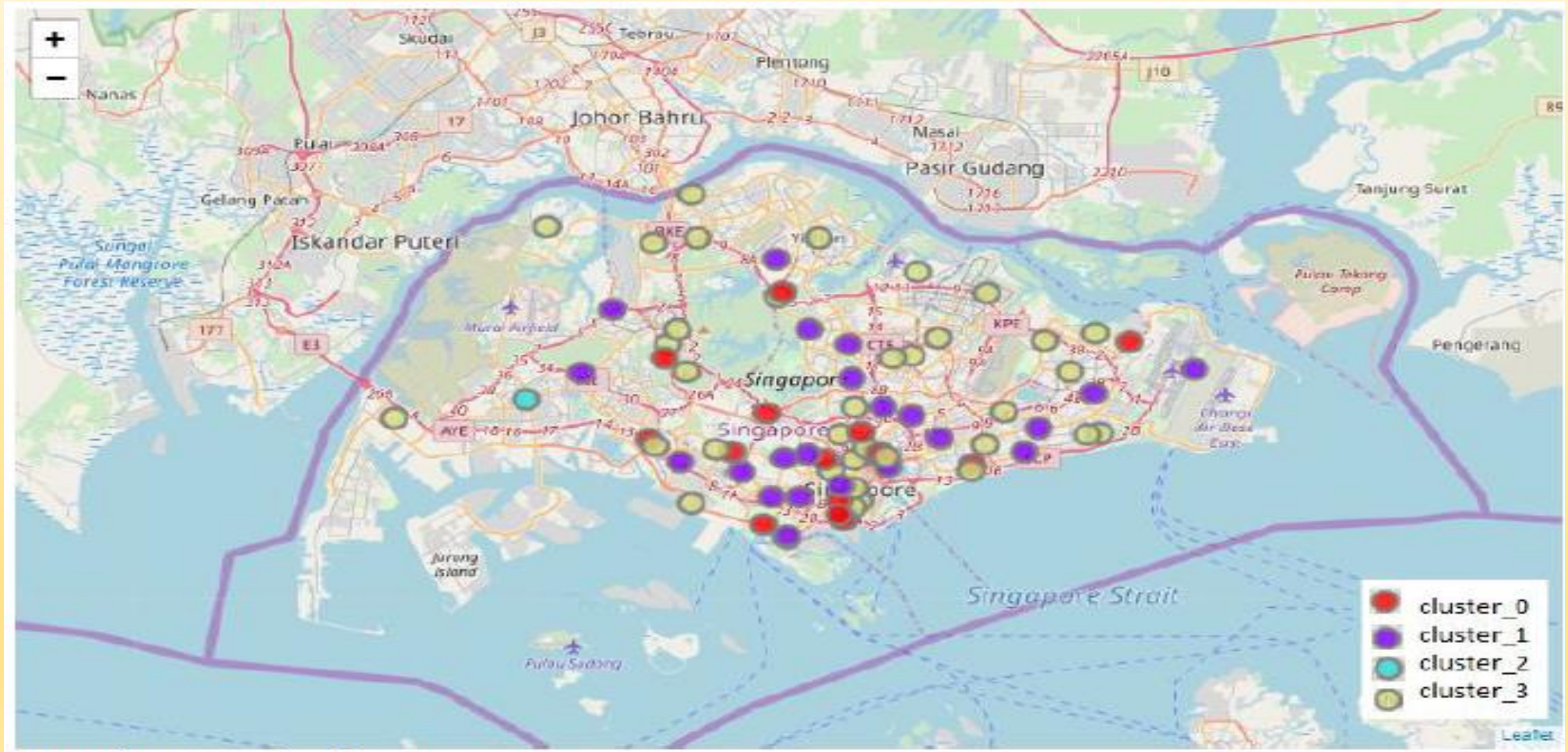
Methodology: K-Means Clustering

K-Means clustering method was used in the analysis with pre-set number of clusters as 4 and the clusters were grouped based on the below characteristics:

- 1) cluster_0: 13 neighborhood districts located mainly central with only 1 neighborhood district in the eastern part of Singapore. The average 2018 psf was at 1945.12 which was the highest among the 4 clusters. The common venues within 1 km of the districts in cluster_0 were generally Japanese Restaurant, Chinese Restaurant, Café and Coffee Shop.
- 2) cluster_1: 16 neighborhood districts located mainly from north to south, east and a couple in north west of Singapore. The average 2018 psf was at 1658.66 which was the second highest of the 4 clusters. The common venues within 1 km of the districts in cluster_1 were generally Chinese Restaurant, Food Court, Café, Convenience Store.
- 3) cluster_2: 1 neighborhood district located in the west of Singapore. The 2018 psf was at 1371.44 which was the lowest of the 4 clusters. The common venues within 1 km of the district in cluster_2 were generally Restaurants, Food Court, Playground and Karaoke Bar.
- 4) cluster_3: 21 neighborhood districts spread over Singapore with average 2018 psf of 1500.78. The common venues within 1 km of the districts in cluster_3 were generally Chinese Restaurant, Japanese Restaurant, Coffee Shop, Café, Bus Station and Supermarket.

Looking at the 4 clusters, it appears that restaurants and cafe are easily accessible within 1 km in almost all neighborhood districts in Singapore. There was no distinct separation of the clusters in terms of locations. However, in terms of average psf, there are 4 different levels. The unique cluster which consist only 1 neighborhood district will be cluster_2 and hold the lowest psf value.

Visualization of 4 clusters



Visualization of 4 clusters

Conclusion

To conclude the analysis, we revisit the business problem "Does souring prices of private property affected by factors such as restaurants, cafes, bars, shopping malls or places of visit in Singapore?". The data from Foursquare provides information of nearby restaurants, cafes, bars, shopping malls or places of visit in Singapore within the neighborhoods. The analysis does not show distinct relationship between venues nearby collected from Foursquare and private property prices. However, further analysis can be achieved by using other type of data such as geo locations of train stations, future development of districts etc.

End

Thank You