

Background

Buying a home is on every come-of-age Singaporean's agenda. Most buyers have certain preferences in mind when looking for a home, such as families with kids will look for areas closer to schools).

Singapore is a small country and is divided into several planning areas. The assumption is that not all planning areas are structured equal.

This project attempts to classify and explore planning areas for someone who's on the hunt for a new home. At the end of the project, it is hoped that differences and similarities between planning areas can be identified to help in house hunting.

Data

Singapore Data (<https://docs.onemap.sg/>)

OneMap provides a set of extensive APIs to retrieve different data for API.

OneMap will primarily be used to retrieve the following details;

- Planning Areas
- Population Data

Foursquare API (<https://developer.foursquare.com/>)

Foursquare API will be used extract facilities and venues of individual planning areas.

Methodology

Data Collection

1. Planning areas are retrieved from OneMap.
2. Coordinates of each planning areas are retrieved using GeoData ArcGis service and stored locally.
3. Population data is also retrieved from OneMap for each planning area. Population data is not used part of clustering. However, it is later used to analyze if there are any relationships between population and venues.
4. Retrieve venues for each planning area from Foursquare.

All these data retrieved from respective services are stored locally for convenience and future use purposes.

Data Wrangling

1. Planning areas with 0 (zero) population are excluded from the clustering. Since the primary aim is analyze areas for house buying, areas without any residents are not useful.
2. Venues with "Neighborhood" venue category are excluded from the clustering. Upon analyzing the data, these venues typically are wards/towns in an area. As a small country, these venues might not have impacts on the home buying decisions.
3. Venues with "Building" venue category are excluded from the clustering. These do not bear any meaning in the clustering as most venues are considered buildings.

Clustering

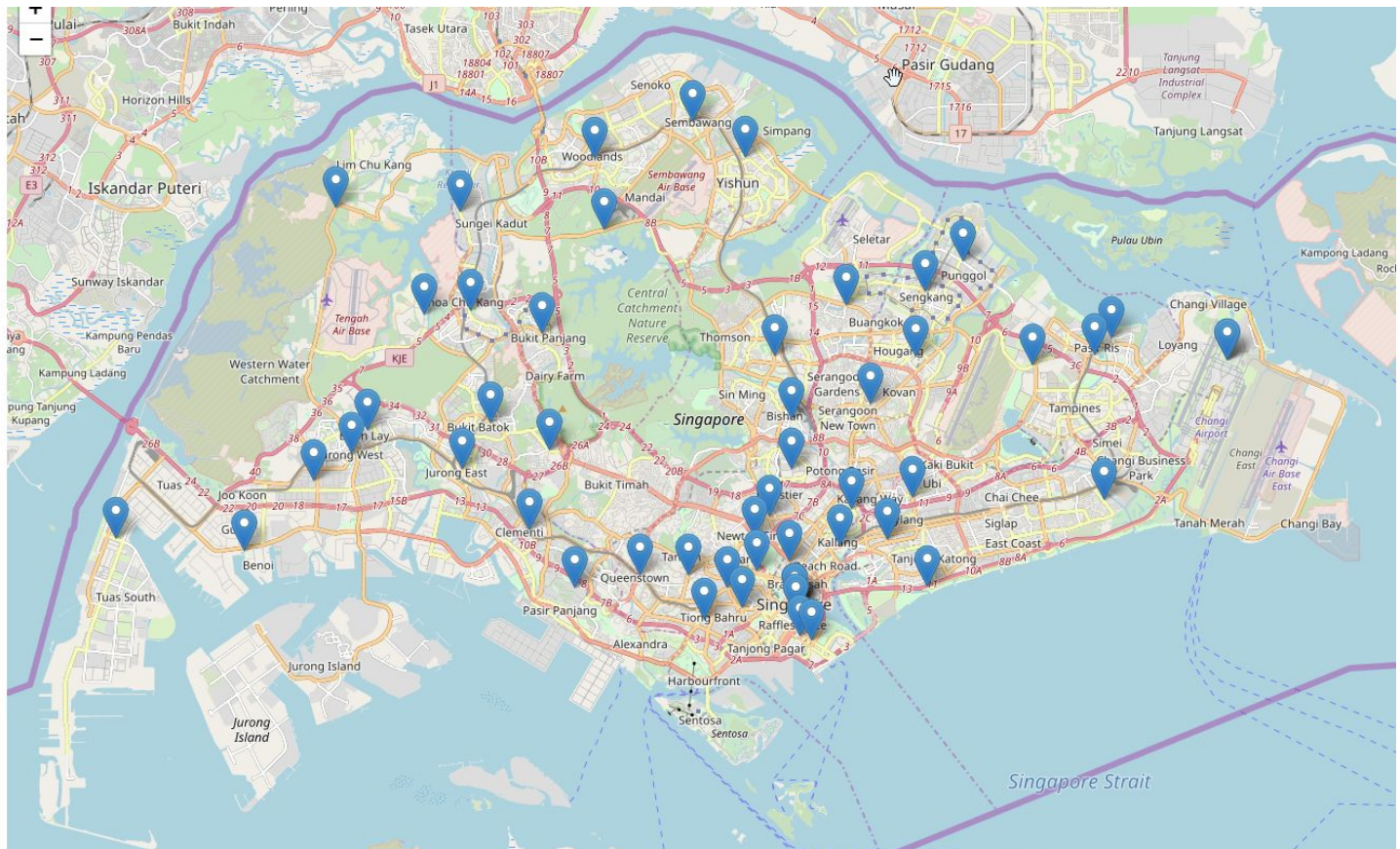
KMeans clustering technique is used to cluster planning areas based on venue categories retrieved from Foursquare.

Determining Optimal k Value for KMeans

Elbow method is used to determine the number of optimal k value for clustering.

Clustering & Analysis

Planning Areas in Singapore



Each marker represents the planning areas in Singapore.

Top 10 Venue Categories

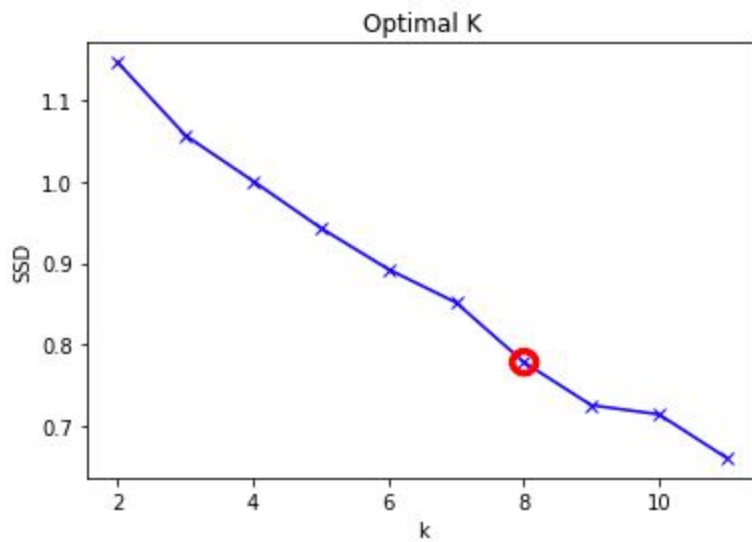
Venue Category	Count
Office	126
Chinese Restaurant	105
Residential Building (Apartment / Condo)	95
Café	95
Coffee Shop	88
Asian Restaurant	83
Bakery	67
Bus Station	66
Salon / Barbershop	62
Food Court	59

As one of the top finance hub in Asia, it is not surprising that the most common venue is Office in Singapore. Other venue categories are of no surprise as Singapore is known for its foods and cafes hopping is one of the popular weekend activities.

As having one of the best transport systems in the world, it is also not surprising to see Bus Station being the one of the top 10 venues.

(Note: Due to the limitation of free Foursquare API. Analysis results here may not accurately represent the actual scene. Nevertheless, let's see what we have here.)

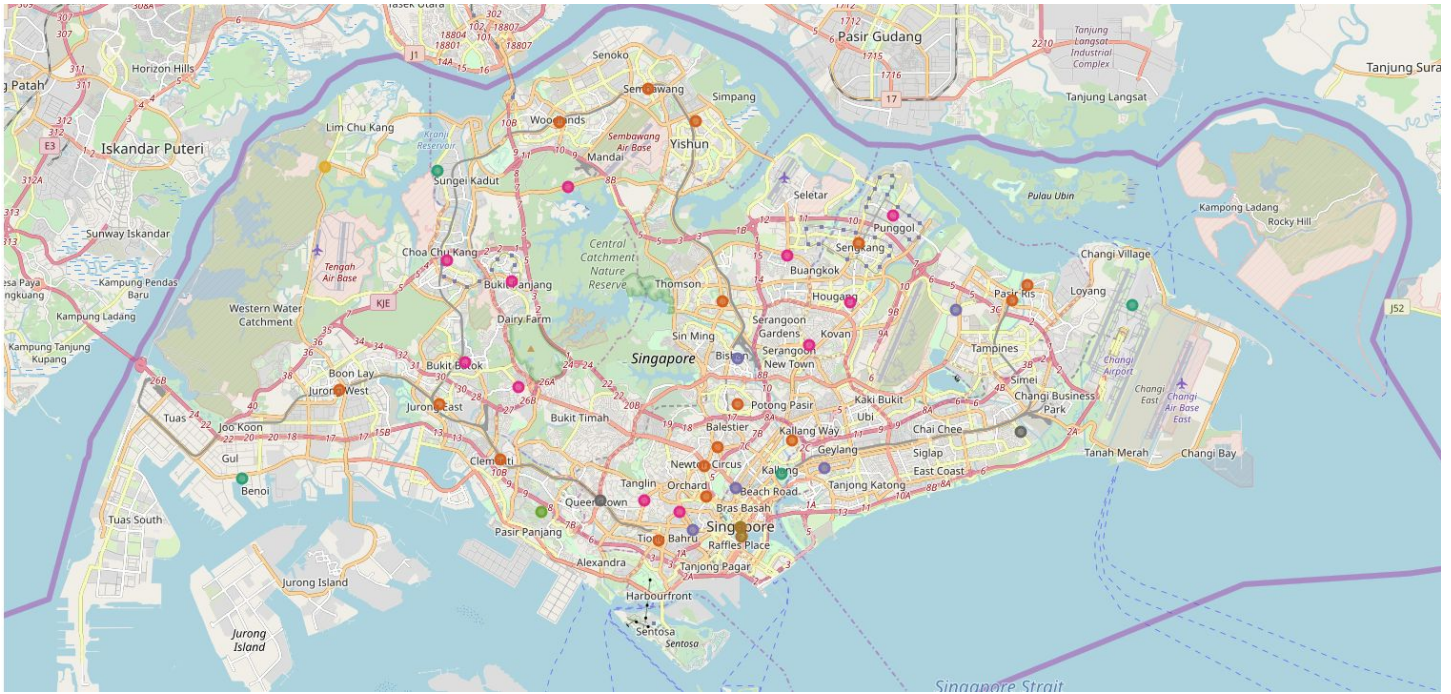
Finding Optimal k value for KMeans



Based on the elbow point, optimal k value is determined to be 8.

During multiple runs, it is also noted that each run gives different value for k . However, 8 is the most common value achieved from multiple runs.

Results



Roles of Population - Population for each planning area was retrieved to compare planning areas and see if similar planning areas in the same cluster have similar population and/or if there is any correlation between the venue categories and population. Population is not used as a feature for clustering, but added to the cluster information to give an overview.

As stated below, population does not seem to have correlation to the type of venues/facilities available in the planning area. However, it is important to note that the size of the planning areas may be different that affects the population.

Cluster 1 - Places for Schools

```
merged_df.loc[merged_df['Cluster Labels'] == 0, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
14	BEDOK	281300	College Classroom	Church	School	High School	Office	Art Gallery	Playground	Basketball Court	College Library	Asian Restaurant
50	QUEENSTOWN	46050	Church	College Classroom	Bus Station	School	Basketball Court	Medical Center	Other Great Outdoors	Convenience Store	Gym	Parking

These planning areas are saturated with schools and classroom facilities. These places could well be suitable for family with school kids and teens going to college.

Cluster 2 - Offices & Industrial Venues

```
merged_df.loc[merged_df['Cluster Labels'] == 1, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
2	PIONEER	90	Office	Coworking Space	Conference Room	Factory	Police Station	Cafeteria	Beach	Bus Station	Parking	Harbor / Marina
4	KALLANG	50250	Office	Bus Station	Furniture / Home Store	River	General Entertainment	Fast Food Restaurant	Metro Station	Electronics Store	Non-Profit	Design Studio
22	CHANGI	2080	Office	Airport Gate	Coworking Space	Fire Station	Airport	Asian Restaurant	Storage Facility	Bus Station	Airport Terminal	Hardware Store
27	SUNGEI KADUT	390	Office	Factory	Furniture / Home Store	Industrial Estate	Metro Station	Fire Station	Parking	Coffee Shop	Coworking Space	Auto Workshop

Although planning areas in this cluster may not be of interest to the home buyer, one could use this result to look for new office places and/or coworking spaces.

Cluster 3 - Foodies' place

```
merged_df.loc[merged_df['Cluster Labels'] == 2, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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	MARINE PARADE	25050	Noodle House	Asian Restaurant	Bakery	Chinese Restaurant	Residential Building (Apartment / Condo)	Snack Place	Coffee Shop	Dessert Shop	Café	BBQ Joint
3	JURONG WEST	266720	Japanese Restaurant	Bakery	Sporting Goods Shop	Asian Restaurant	Shoe Store	Fast Food Restaurant	Cosmetics Shop	Dessert Shop	Medical Center	Food Court
5	NOVENA	23700	Sporting Goods Shop	Office	Café	Outdoor Supply Store	Ice Cream Shop	Japanese Restaurant	Dance Studio	Thai Restaurant	Basketball Court	Juice Bar
6	PASIR RIS	74440	Bank	Dessert Shop	Bakery	Cosmetics Shop	Japanese Restaurant	Café	Sushi Restaurant	Food Court	Fast Food Restaurant	Convenience Store
8	JURONG EAST	40910	Japanese Restaurant	Cosmetics Shop	Dessert Shop	Pharmacy	Bakery	Café	Chinese Restaurant	Juice Bar	Korean Restaurant	Snack Place
13	ANG MO KIO	165710	Chinese Restaurant	Residential Building (Apartment / Condo)	Karaoke Bar	Noodle House	Medical Center	Dessert Shop	Doctor's Office	Fast Food Restaurant	Asian Restaurant	Steakhouse
18	BUKIT MERAH	73120	Chinese Restaurant	Coffee Shop	Food Court	Noodle House	Asian Restaurant	Seafood Restaurant	Café	Wings Joint	Bakery	Gym
24	SENGKANG	122280	Café	Mobile Phone Shop	Food Court	Japanese Restaurant	Salon / Barbershop	Snack Place	Noodle House	Chinese Restaurant	Pizza Place	Bakery
26	DOWNTOWN CORE	1250	Café	Resort	Pool	Dessert Shop	Parking	Diner	Chinese Restaurant	Coffee Shop	Food Court	Sports Bar
29	ORCHARD	530	Cosmetics Shop	Shopping Mall	Department Store	Shoe Store	Japanese Restaurant	Italian Restaurant	Coffee Shop	Chinese Restaurant	Juice Bar	Bar
32	TOA PAYOH	63070	Chinese Restaurant	Coffee Shop	Church	Café	Medical Center	Indian Restaurant	Pool Hall	Dentist's Office	Asian Restaurant	Noodle House
34	WOODLANDS	125810	Snack Place	Chinese Restaurant	Bakery	Café	Jewelry Store	Fast Food Restaurant	Bank	Thai Restaurant	Miscellaneous Shop	Coffee Shop
43	YISHUN	106170	Coffee Shop	Indian Restaurant	Bakery	Grocery Store	Market	Asian Restaurant	Restaurant	Medical Center	Parking	Bus Line
45	NEWTON	4030	Seafood Restaurant	Chinese Restaurant	BBQ Joint	Office	Noodle House	Coffee Shop	Juice Bar	Dessert Shop	Asian Restaurant	Bus Stop
47	CLEMENTI	48700	Salon / Barbershop	Chinese Restaurant	Pizza Place	Café	Bank	Bubble Tea Shop	Dessert Shop	Bakery	Dentist's Office	BBQ Joint
51	SEMBAWANG	44140	Salon / Barbershop	Bus Station	Chinese Restaurant	Electronics Store	Department Store	Convenience Store	Dessert Shop	Asian Restaurant	Malay Restaurant	Japanese Restaurant

If one is looking for a place with easy access to food places, he/she should consider planning area in this cluster.

Cluster 4 - Mixed use places

```
merged_df.loc[merged_df['Cluster Labels'] == 3, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
5	BISHAN	42900	Coffee Shop	Food Court	Residential Building (Apartment / Condo)	Playground	College Classroom	Seafood Restaurant	Fried Chicken Joint	Housing Development	Post Office	Other Great Outdoors
1	GEYLANG	55250	Hotel	Temple	Coffee Shop	Residential Building (Apartment / Condo)	Chinese Restaurant	Motel	Gas Station	Korean Restaurant	Strip Club	Food Court
3	TAMPINES	131330	Temple	Coffee Shop	Bus Station	Bus Stop	Café	Office	Asian Restaurant	Martial Arts Dojo	Athletics & Sports	Spiritual Center
0	OUTRAM	20030	Hotel	Nightclub	Hotel Bar	Buffet	Event Space	Pool	Meeting Room	Bus Station	Lounge	Office
4	ROCHOR	13450	Hotel	Bar	Chinese Restaurant	Asian Restaurant	Coffee Shop	Hostel	Turkish Restaurant	Boutique	Food Court	Korean Restaurant

Closer to entertainment venues, these areas could well be upscale residential areas.

Cluster 5 - Ideal residential areas?

```
merged_df.loc[merged_df['Cluster Labels'] == 4, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
7	HOUGANG	223010	Bus Station	Parking	Salon / Barbershop	Student Center	Lounge	Field	Metro Station	Asian Restaurant	Residential Building (Apartment / Condo)	Moving Target
10	MANDAI	2090	Exhibit	Ice Cream Shop	Trail	Other Great Outdoors	Office	Electronics Store	Asian Restaurant	Meeting Room	Park	Coffee Shop
11	BUKIT PANJANG	71290	Residential Building (Apartment / Condo)	Parking	Bus Station	College Administrative Building	Office	Park	Housing Development	Government Building	Gym	Café
12	BUKIT TIMAH	41450	Café	Residential Building (Apartment / Condo)	Seafood Restaurant	Doctor's Office	Spa	Bus Line	Dentist's Office	Pet Store	Bakery	Bike Shop
17	BUKIT BATOK	73550	Doctor's Office	Residential Building (Apartment / Condo)	Furniture / Home Store	Spa	Parking	Bus Station	Bank	Salon / Barbershop	Bus Line	Bakery
19	CHOA CHU KANG	94070	Bus Station	Student Center	Karaoke Bar	Bakery	Food Truck	Other Nightlife	Gas Station	Jewelry Store	Nightclub	Salon / Barbershop
20	TANGLIN	11270	Office	Conference Room	Residential Building (Apartment / Condo)	Café	College Academic Building	Mexican Restaurant	Chinese Restaurant	Tech Startup	School	Bus Line
23	SELETAR	110	Residential Building (Apartment / Condo)	Bus Stop	Bus Station	Japanese Restaurant	Language School	Playground	Asian Restaurant	Bakery	Bank	Dessert Shop
25	SERANGOON	56710	Residential Building (Apartment / Condo)	Pub	Chinese Restaurant	Asian Restaurant	Church	Coffee Shop	Bar	Automotive Shop	Café	Gaming Cafe
49	PUNGGOL	161570	Residential Building (Apartment / Condo)	Housing Development	Food Court	Parking	Asian Restaurant	Chinese Restaurant	Multiplex	Professional & Other Places	Other Great Outdoors	Indian Restaurant
53	RIVER VALLEY	5330	Residential Building (Apartment / Condo)	Bar	Chinese Restaurant	Church	Pool	Office	Indian Restaurant	Food	Bakery	Fraternity House

With a good mix of residential places, food places and other amenities, these could well be ideal places to buy a home.

Cluster 6 - University grounds

```
merged_df.loc[merged_df['Cluster Labels'] == 5, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
44	MUSEUM	210	College Lab	College Science Building	College Classroom	College Quad	University	College Administrative Building	College Academic Building	Coffee Shop	Asian Restaurant	College Library

This area seems to be have been occupied by a university and/or universities and not so much of a residential area.

Cluster 7 - No-go Zone

```
merged_df.loc[merged_df['Cluster Labels'] == 6, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
9	LIM CHU KANG	50	Military Base	Bridge	Campground	Office	Auditorium	Farm	Salon / Barbershop	General Entertainment	Other Great Outdoors	Government Building

Mainly seems to have been occupied by military bases and government buildings, this is a no-go zone for home buyers.

Cluster 8 - Places for a day out

```
merged_df.loc[merged_df['Cluster Labels'] == 7, merged_df.columns[[0] + [1] + list(range(3, merged_df.shape[1]))]]
```

	Planning Area	population	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
36	SINGAPORE RIVER	1560	Hotel	Outdoor Sculpture	Office	Café	Event Space	Gym	Hotel Bar	Art Gallery	Lounge	Bridge
39	SOUTHERN ISLANDS	960	Event Space	Art Gallery	Salon / Barbershop	Art Museum	Bar	Café	Spa	Asian Restaurant	Restaurant	Racetrack
42	WESTERN WATER CATCHMENT	720	Event Space	Art Gallery	Salon / Barbershop	Art Museum	Bar	Café	Spa	Asian Restaurant	Restaurant	Racetrack

These areas have a lot more to offer for a day-out, but barely residential areas.

Conclusion

Based on the clustering results, most suitable residential areas, in my opinion, are as below;

- Cluster 5
Areas in this cluster have a good mix of amenities and places to eat out.
- Cluster 3
Foodies' heavens, these areas are a good choice when it comes to filling up your tummy.
- Cluster 4
Residential areas in this area could be in the higher price range due to their proximity to hotels and entertainment venues.

Limitations

Due to the fact that venues are extracted based on central coordinates of planning areas, some venues may not have been extracted. This may adversely impact cluster results.

Improvements

Limitations stated above may be overcome by using GeoJSON data that indicates the boundaries of each planning area. Based on the maximum radius, venues around each coordinate can be retrieved to further improve the cluster results.