

# Comparative Analysis between Singapore and Hong Kong

Chan Kok Siang

26 May 2020

## 1 Introduction

### 1.1 Background

Singapore and Hong Kong have often been associated with one another due to their similarities and statuses as major financial and shipping hubs. Their economic success over the years attracted top talents to settle and work in the cities but also resulted in high standards of living - both cities are currently ranked among the most expensive cities in the world.



Figure 1: Prominent skylines of Singapore (left) and Hong Kong (right)



	Singapore	Hong Kong
Flag		
Official Languages	English, Malay, Mandarin, Tamil	Mandarin, English
Ethnic Groups	74.3% Chinese Singaporeans, 13.3% Malay Singaporeans, 9.1% Indian Singaporeans, 3.3% Others	92.0% Han Chinese, 2.5% Filipino, 2.1% Indonesian, 0.8% White, 0.5% Indian, 0.3% Nepalese, 1.6% Others
Area	725.1 km <sup>2</sup>	2,755 km <sup>2</sup>
Population (2018)	Total: 5,638,700 Density: 7,804/km <sup>2</sup>	Total: 7,482,500 Density: 6,777/km <sup>2</sup>
GDP (PPP) (2019)	Total: \$589.187 billion Per capita: \$103,717	Total: \$490.880 billion Per capita: \$64,928
Time Zone	UTC+8	UTC+8
Geographic Coordinates	Latitude: 1.3, Longitude: 103.8	Latitude: 22.3, Longitude: 114.2

Table 1: General Information of Singapore and Hong Kong (source: Wikipedia)

Business travellers and leisure seekers often compare the neighbourhoods, shopping belts and central business districts (CBDs) between the two cities. Depending on the individual's real-world experience, travel inclinations, hobbies and interests, etc., the resulting comparisons can vary widely.

*“As both are overcrowded urban enclaves with no natural resources, they have to continue to pursue pro-growth economic policies, and their people have to constantly work hard and remain highly entrepreneurial in order to “earn” their economic growth.”*

Extracted quote from a book titled “Singapore and Hong Kong: Comparative Perspectives on the Occasion of the 20th Anniversary of the Handover” by Institute of Advanced Studies, Nanyang Technological University, Singapore.

*“The beauty of Hong Kong for me is that one could witness a dilapidated 4 storied building which looks to be on the verge of collapsing if one were to even sneeze in its proximity, right next to a sprawling skyscraper and there would be nothing incongruous about that set-up. It just blends into the Hong Kong feel. Singapore on the other hand has its homogenous looking HDBs well laid out maintaining a presentable image which further epitomizes the orderly nature of the city.”*

Extracted quote from an online post by Akshobh Giridharadas, Journalist, Business Desk at Channel NewsAsia.

## **1.2 Problem**

The analysis done at the individual's level can be subjective and influenced by the individual's personal experience, opinions and length of stay in the city. It can also be limited in scope since it is atypical for an individual to travel to all areas in both cities, hence rendering a comprehensive analysis to be almost impossible.

This report aims to provide an objective and data-driven comparative analysis of both cities, by exploring the different districts within each city, finding clusters within each city and determining similarities/ uniqueness between the two cities.

## **1.3 Interested Parties**

Readers who are interested to gain another perspective on how the two cities are similar (or dissimilar) may find the contents useful. This may include people who are planning to visit the cities in the near future, either for business or leisure purposes, or are finding districts with certain amenities and facilities to stay in for extended periods of time.

## 2 Data

### 2.1 Data Sources

The following data sources are used for the analysis in this report.

Postal district	Postal sector (1st 2 digits of 6-digit postal codes)	General location
01	01, 02, 03, 04, 05, 06	Raffles Place, Cecil, Marina, People's Park
02	07, 08	Anson, Tanjong Pagar
03	14, 15, 16	<a href="#">Bukit Merah</a> , <a href="#">Queenstown</a> , <a href="#">Tiong Bahru</a>
04	09, 10	Telok Blangah, Harbourfront
05	11, 12, 13	Pasir Panjang, Hong Leong Garden, Clementi New Town
06	17	High Street, Beach Road (part)
07	18, 19	Middle Road, Golden Mile
08	20, 21	<a href="#">Little India</a> , <a href="#">Farrer Park</a> , <a href="#">Jalan Besar</a> , <a href="#">Lavender</a>
09	22, 23	Orchard, Cairnhill, River Valley
10	24, 25, 26, 27	Ardmore, Bukit Timah, Holland Road, Tanglin
11	28, 29, 30	Watten Estate, Novena, Thomson
12	31, 32, 33	Balestier, Toa Payoh, Serangoon
13	34, 35, 36, 37	Macpherson, Braddell
14	38, 39, 40, 41	Geylang, Eunos
15	42, 43, 44, 45	Katong, Joo Chiat, Amber Road
16	46, 47, 48	Bedok, Upper East Coast, Eastwood, Kew Drive
17	49, 50, 81	Loyang, Changi
18	51, 52	Simei, Tampines, Pasir Ris
19	53, 54, 55, 82	Serangoon Garden, Hougang, Punggol
20	56, 57	Bishan, Ang Mo Kio
21	58, 59	Upper Bukit Timah, Clementi Park, Ulu Pandan
22	60, 61, 62, 63, 64	Jurong, Tuas
23	65, 66, 67, 68	Hillview, Dairy Farm, Bukit Panjang, Choa Chu Kang
24	69, 70, 71	Lim Chu Kang, Tengah
25	72, 73	Kranji, Woodgrove, Woodlands
26	77, 78	Upper Thomson, Springleaf
27	75, 76	Yishun, Sembawang
28	79, 80	Seletar

Figure 2: Singapore Postal Districts (source: Wikipedia)

District	Chinese	Population <sup>[when?] [6]</sup>	Area (km <sup>2</sup> )	Density (/km <sup>2</sup> )	Region
Central and Western	中西區	244,600	12.44	19,983.92	Hong Kong Island
Eastern	東區	574,500	18.56	31,217.67	Hong Kong Island
Southern	南區	269,200	38.85	6,962.68	Hong Kong Island
Wan Chai	灣仔區	150,900	9.83	15,300.10	Hong Kong Island
Sham Shui Po	深水埗區	390,600	9.35	41,529.41	Kowloon
Kowloon City	九龍城區	405,400	10.02	40,194.70	Kowloon
Kwun Tong	觀塘區	641,100	11.27	56,779.05	Kowloon
Wong Tai Sin	黃大仙區	426,200	9.30	45,645.16	Kowloon
Yau Tsim Mong	油尖旺區	318,100	6.99	44,864.09	Kowloon
Islands	離島區	146,900	175.12	825.14	New Territories
Kwai Tsing	葵青區	507,100	23.34	21,503.86	New Territories
North	北區	310,800	136.61	2,220.19	New Territories
Sai Kung	西貢區	448,600	129.65	3,460.08	New Territories
Sha Tin	沙田區	648,200	68.71	9,433.85	New Territories
Tai Po	大埔區	307,100	136.15	2,220.35	New Territories
Tsuen Wan	荃灣區	303,600	61.71	4,887.38	New Territories
Tuen Mun	屯門區	495,900	82.89	5,889.38	New Territories
Yuen Long	元朗區	607,200	138.46	4,297.99	New Territories

Figure 3: Hong Kong Districts (source: Wikipedia)

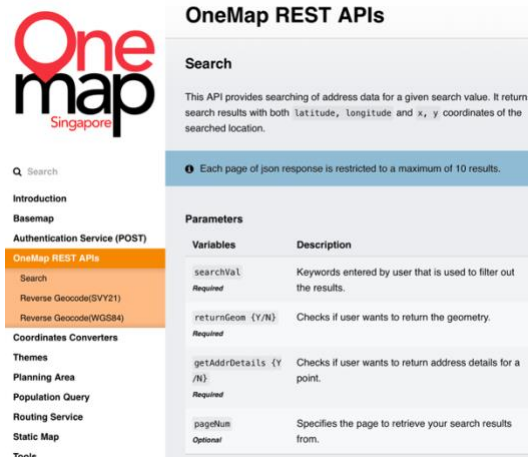


Figure 4: OneMap REST API on Search (source: OneMap)

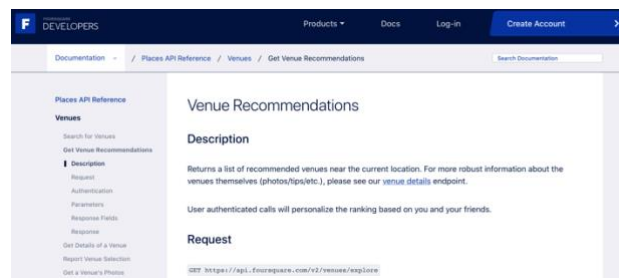


Figure 5: Foursquare Places API on Explore (source: Foursquare)

## 2.2 Data Cleaning

For Singapore districts, the information provided by Wikipedia does not have corresponding information on geographic coordinates. The general locations listed for each district are used to determine the centroid for that district by invoking the OneMap REST API on Explore and averaging their coordinates. It is assumed that simple averaging is acceptable since the general locations are close to each other within the district and hence the Earth can be treated as being relatively flat (i.e. negligible error when using planar coordinates, though it is more accurate to use complex calculation methods considering Earth's spherical surface).

For Hong Kong districts, the coordinates for each district are obtained from its own Wikipedia's page.

Separate CSV files are created to store the contents and coordinates of the districts.

## 2.3 Feature Selection

The Foursquare Places API on Search returns a list of venues in a district based on its coordinates, with details such as id, name, location and categories. One hot encoding is applied to "categories" and the results are used as features to find clusters within each city.

### 3 Methodology

#### 3.1 Data Analysis

The key steps are: (i) Dataframe Creation, (ii) Coordinates Addition, and (iii) District Clustering.

##### 3.1.1 Dataframe Creation

The Beautiful Soup python library is used to extract data from the source html pages. Some data cleaning and column renaming are done to facilitate subsequent processing.

SGDistrict		GeneralLocation	HKDistrict		Region
0	1	Raffles Place, Cecil, Marina Square, People's ...	0	Central and Western	Hong Kong Island
1	2	Anson, Tanjong Pagar	1	Eastern	Hong Kong Island
2	3	Bukit Merah, Queenstown, Tiong Bahru	2	Southern	Hong Kong Island
3	4	Telok Blangah, Harbourfront	3	Wan Chai	Hong Kong Island
4	5	Pasir Panjang, West Coast Terrace, Clementi	4	Sham Shui Po	Kowloon

Figure 6: Sample Dataframes Created from Source HTML Pages

##### 3.1.2 Coordinates Addition

For each district within Singapore, the coordinates for its general locations are found via OneMap API on Search. Multiple locations may be returned via the API. Coordinates of the first returned location are used since returned locations are assumed to be near each other.

	GeneralLocation	SearchedLocation	SearchedLatitude	SearchedLongitude
0	Raffles Place, Cecil, Marina Square, People's ...	RAFFLES PLACE	1.283933	103.851463
1	Raffles Place, Cecil, Marina Square, People's ...	158 CECIL STREET SINGAPORE 069545	1.279789	103.848043
2	Raffles Place, Cecil, Marina Square, People's ...	MARINA SQUARE	1.291040	103.857624
3	Raffles Place, Cecil, Marina Square, People's ...	PEOPLE'S PARK CENTRE	1.285758	103.843930
4	Anson, Tanjong Pagar	79 ANSON ROAD SINGAPORE 079906	1.274098	103.845669

Figure 7: Sample Coordinates of First Returned Locations for Singapore's Districts

The centroid for each district is determined by calculating the means of coordinates for the district's general locations.

	SearchedLatitude	SearchedLongitude
GeneralLocation		
Raffles Place, Cecil, Marina Square, People's Park	1.285130	103.850265
Anson, Tanjong Pagar	1.274526	103.843923
Bukit Merah, Queenstown, Tiong Bahru	1.285461	103.820705
Telok Blangah, Harbourfront	1.269926	103.813697
Pasir Panjang, West Coast Terrace, Clementi	1.310088	103.767941

Figure 8: Sample Coordinates of Centroids for Singapore's Districts

Coordinates of the centroids are then appended to the original dataframe.

SGDistrict	GeneralLocation	SearchedLatitude	SearchedLongitude
0	1 Raffles Place, Cecil, Marina Square, People's ...	1.285130	103.850265
1	2 Anson, Tanjong Pagar	1.274526	103.843923
2	3 Bukit Merah, Queenstown, Tiong Bahru	1.285461	103.820705
3	4 Telok Blangah, Harbourfront	1.269926	103.813697
4	5 Pasir Panjang, West Coast Terrace, Clementi	1.310088	103.767941

Figure 9: Sample Merged Dataframe with Coordinates of Singapore's Districts

For districts within Hong Kong, their coordinates are found on the individual district's Wikipedia page. The Beautiful Soup python library is used to extract both the URLs of the individual pages and coordinates.

HKDistrict	URL	Latitude	Longitude
0 Islands	<a href="https://en.wikipedia.org/wiki/Islands_District">https://en.wikipedia.org/wiki/Islands_District</a>		
1 Kwai Tsing	<a href="https://en.wikipedia.org/wiki/Kwai_Tsing_District">https://en.wikipedia.org/wiki/Kwai_Tsing_District</a>		
2 North	<a href="https://en.wikipedia.org/wiki/North_District_(...">https://en.wikipedia.org/wiki/North_District_(...</a>		
3 Sai Kung	<a href="https://en.wikipedia.org/wiki/Sai_Kung_District">https://en.wikipedia.org/wiki/Sai_Kung_District</a>		
4 Sha Tin	<a href="https://en.wikipedia.org/wiki/Shan_Tin_District">https://en.wikipedia.org/wiki/Shan_Tin_District</a>		

HKDistrict	Latitude	Longitude
Islands	22.26114	113.94608
Kwai Tsing	22.35488	114.08401
North	22.49471	114.13812
Sai Kung	22.38143	114.27052
Sha Tin	22.38715	114.19534

Figure 10: Sample Extracted URLs and Coordinates of Hong Kong's Districts

Coordinates are then appended to the original dataframe.

HKDistrict	Region	Latitude	Longitude
0 Central and Western	Hong Kong Island	22.28666	114.15497
1 Eastern	Hong Kong Island	22.28411	114.22414
2 Southern	Hong Kong Island	22.24725	114.15884
3 Wan Chai	Hong Kong Island	22.27968	114.17168
4 Sham Shui Po	Kowloon	22.33074	114.16220

Figure 11: Sample Merged Dataframe with Coordinates of Hong Kong's Districts

### 3.1.3 District Clustering

The following sections break the comparative analysis between Singapore and Hong Kong into several steps, namely:

(i) Exploring districts in both cities by retrieving information on nearby recommended venues from Foursquare

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	1 - Raffles Place, Cecil, Marina Square, People...	1.28513	103.850265	CITY Hot Pot Shabu shabu	1.284173	103.851585	Hotpot Restaurant
1	1 - Raffles Place, Cecil, Marina Square, People...	1.28513	103.850265	Ritual Gym	1.285965	103.848651	Gym
2	1 - Raffles Place, Cecil, Marina Square, People...	1.28513	103.850265	Sabaai Sabaai Traditional Thai Massage	1.286964	103.849512	Massage Studio
3	1 - Raffles Place, Cecil, Marina Square, People...	1.28513	103.850265	Virgin Active	1.284608	103.850815	Gym / Fitness Center
4	1 - Raffles Place, Cecil, Marina Square, People...	1.28513	103.850265	Amoy Hotel	1.283118	103.848539	Hotel

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central and Western	22.28666	114.15497	Four Seasons Hotel Hong Kong (香港四季酒店)	22.286554	114.156929	Hotel
1	Central and Western	22.28666	114.15497	Samsen (泰館)	22.285082	114.152286	Thai Restaurant
2	Central and Western	22.28666	114.15497	Godenya	22.284673	114.153195	Japanese Restaurant
3	Central and Western	22.28666	114.15497	IFC Mall Ferry Pier Walkway	22.287095	114.157675	Scenic Lookout
4	Central and Western	22.28666	114.15497	The Spa at Four Seasons	22.286279	114.157623	Spa

Figure 12: Sample Nearby Recommended Venues in Singapore (left) and Hong Kong (right)

(ii) Analysing each district by determining its most recommended venue categories

District	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	1 - Raffles Place, Cecil, Marina Square, People...	Hotel	Waterfront	Seafood Restaurant	Cocktail Bar	Event Space	Japanese Restaurant	Plaza	Concert Hall	Spanish Restaurant
1	2 - Anson, Tanjong Pagar	Japanese Restaurant	Coffee Shop	Hotel	Cafe	Food Court	Cocktail Bar	Bakery	Tapas Restaurant	Gym / Fitness Center
2	3 - Bukit Merah, Queenstown, Tiong Bahru	Chinese Restaurant	Cafe	Food Court	Park	Bakery	Coffee Shop	Noodle House	Hainan Restaurant	Indian Restaurant
3	4 - Telok Blangah, Harbourfront	Chinese Restaurant	Scenic Lookout	Trail	Cafe	Food Court	Multiplex	Park	Clothing Store	Pool
4	5 - Pear Peninsula, West Coast Terrace, Clementi	Coffee Shop	Food Court	Cafe	Bakery	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Park	Dessert Shop

District	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	Central and Western	Japanese Restaurant	French Restaurant	Cafe	Italian Restaurant	Hotel	Cocktail Bar	Coffee Shop	Gym / Fitness Center	Thai Restaurant
1	Eastern	Cafe	Japanese Restaurant	Chinese Restaurant	Coffee Shop	Park	Department Store	Noodle House	Seafood Restaurant	French Restaurant
2	Southern	Chinese Restaurant	Furniture / Home Store	Cafe	Park	Restaurant	Coffee Shop	Fast Food Restaurant	Vietnamese Restaurant	Shopping Mall
3	Wan Chai	Hotel	Cafe	Coffee Shop	Italian Restaurant	Chinese Restaurant	Cartoonish Restaurant	Burger Joint	Seafood Restaurant	Snack Place
4	Sham Shui Po	Noodle House	Chinese Restaurant	Dessert Shop	Hong Kong Restaurant	Cha Chan Teng	Cafe	Dim Sum Restaurant	Coffee Shop	Malay Restaurant

Figure 13: Sample Most Recommended Venues in Singapore (left) and Hong Kong (right)

(iii) Clustering districts by k-means clustering and visualising clusters in a Folium map

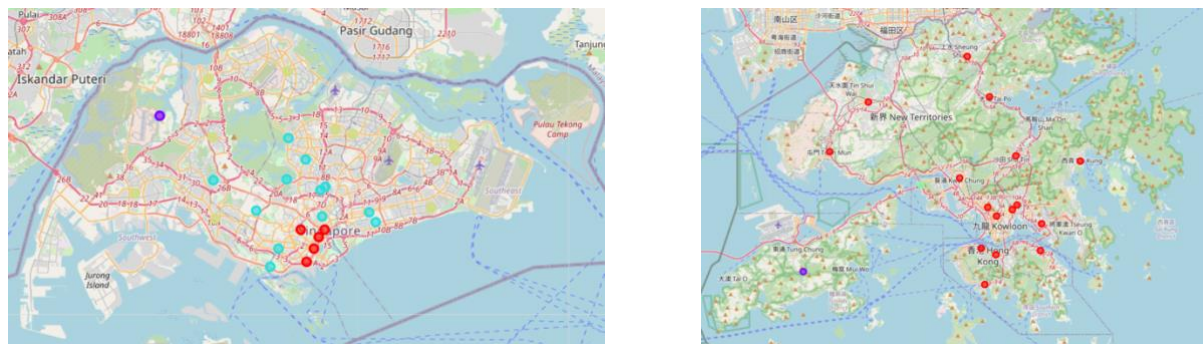


Figure 14: Clusters in Singapore (left) and Hong Kong (right)

(iv) Examining districts by evaluating commonalities between their most recommended venue categories and providing descriptive names to clusters

The clustering algorithm has determined that there are 4 clusters in Singapore. Coupled with personal knowledge of the districts and having a closer look at the data, districts within the same cluster do have similar venues and hence are similar.

### Cluster 1 – Cluster with Commercial Facilities and Multinational F&B Options

	District	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	1 - Raffles Place, Cecil, Marina Square, Peopl...	Hotel	Waterfront	Seafood Restaurant	Cocktail Bar	Event Space	Japanese Restaurant	Plaza	Concert Hall	Spanish Restaurant	Performing Arts Venue
1	2 - Anson, Tanjong Pagar	Japanese Restaurant	Coffee Shop	Hotel	Café	Food Court	Cocktail Bar	Bakery	Tapas Restaurant	Spanish Restaurant	Gym / Fitness Center
5	6 - High Street, Beach Road	Hotel	Shopping Mall	Cocktail Bar	Event Space	Italian Restaurant	Café	Japanese Restaurant	Coffee Shop	Bakery	Beer Bar
6	7 - Middle Road, Golden Mile	Hotel	Café	Thai Restaurant	Chinese Restaurant	Cocktail Bar	Indian Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Shopping Mall	Event Space
8	9 - Orchard, Cairnhill, River Valley	Hotel	Shopping Mall	Japanese Restaurant	Clothing Store	Bakery	Yoga Studio	Wine Bar	Café	Cosmetics Shop	Coffee Shop

Figure 15: Districts and Most Common Venues in Singapore's Cluster 1

### Cluster 2 – Cluster with Nature-Themed Facilities (outlier)

	District	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
23	24 - Lim Chu Kang, Tengah	Farm	Garden Center	Spa	Café	Farmers Market	Chinese Restaurant	Flower Shop	Zoo Exhibit	Food	Food Court

Figure 16: Districts and Most Common Venues in Singapore's Cluster 2



### Cluster 3 – Cluster with Residential Facilities and More Individual-Themed Restaurants (i.e. more matured estates, closer to CBD)

	District	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	3 - Bukit Merah, Queenstown, Tiong Bahru	Chinese Restaurant	Café	Food Court	Park	Bakery	Coffee Shop	Noodle House	Hainan Restaurant	Indian Restaurant	Hotel
3	4 - Telok Blangah, Harbourfront	Chinese Restaurant	Scenic Lookout	Trail	Café	Food Court	Multiplex	Park	Clothing Store	Pool	Japanese Restaurant
7	8 - Little India, Farrer Park, Jalan Besar, La...	Indian Restaurant	Chinese Restaurant	Café	Bakery	Thai Restaurant	Coffee Shop	Hotel	Restaurant	Vietnamese Restaurant	Italian Restaurant
9	10 - Ardmore, Bukit Timah, Holland Road, Tanglin	Bakery	Chinese Restaurant	Café	Ice Cream Shop	Food Court	Bar	Supermarket	Asian Restaurant	Thai Restaurant	Noodle House
10	11 - Watten Estate, Novena, Thomson	Café	Chinese Restaurant	Flower Shop	Trail	Spa	Bakery	Thai Restaurant	Ice Cream Shop	Dessert Shop	Supermarket
11	12 - Balestier, Toa Payoh, Serangoon	Chinese Restaurant	Food Court	Hotel	Asian Restaurant	Noodle House	Dessert Shop	Coffee Shop	Bakery	Seafood Restaurant	Supermarket
12	13 - Macpherson, Braddell	Chinese Restaurant	Coffee Shop	Asian Restaurant	Bakery	Convenience Store	Noodle House	Snack Place	Market	Seafood Restaurant	Gym
13	14 - Geylang, Eunos	Chinese Restaurant	Asian Restaurant	Food Court	Noodle House	Vegetarian / Vegan Restaurant	Multiplex	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Ice Cream Shop
14	15 - Katong, Joo Chiat, Amber Road	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Italian Restaurant	Hotel	Coffee Shop	Noodle House	Bakery	Multiplex	Japanese Restaurant
19	20 - Bishan, Ang Mo Kio	Chinese Restaurant	Coffee Shop	Food Court	Café	Japanese Restaurant	Thai Restaurant	Ice Cream Shop	Fast Food Restaurant	Supermarket	Spa
20	21 - Upper Bukit Timah, Clementi Park, Ulu Pandan	Food Court	Chinese Restaurant	Korean Restaurant	Café	Thai Restaurant	Asian Restaurant	Shopping Mall	Coffee Shop	Bus Station	Hainan Restaurant
25	26 - Upper Thomson, Springleaf	Food Court	Chinese Restaurant	Bus Stop	Coffee Shop	Park	Seafood Restaurant	Asian Restaurant	Indian Restaurant	Shanghai Restaurant	Soup Place

Figure 17: Districts and Most Common Venues in Singapore's Cluster 3

### Cluster 4 – Cluster with Residential Facilities and More Commoditised F&B Options (less matured estates, further away from CBD)

	District	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
4	5 - Pasir Panjang, West Coast Terrace, Clementi	Coffee Shop	Food Court	Café	Bakery	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Park	Dessert Shop	Sandwich Place
15	16 - Bedok, Upper East Coast, Eastwood, Kew Drive	Food Court	Coffee Shop	Noodle House	Asian Restaurant	Chinese Restaurant	Malay Restaurant	Café	Dessert Shop	Beach	Playground
16	17 - Loyang, Changi	Coffee Shop	Supermarket	Café	Snack Place	Bus Station	Asian Restaurant	Thai Restaurant	Fast Food Restaurant	Chinese Restaurant	Resort
17	18 - Simei, Tampines, Pasir Ris	Coffee Shop	Bakery	Café	Gym	Supermarket	Thai Restaurant	Fried Chicken Joint	Japanese Restaurant	Gym / Fitness Center	Clothing Store
18	19 - Serangoon Garden, Hougang, Punggol	Coffee Shop	Food Court	Bus Station	Fast Food Restaurant	Asian Restaurant	Café	Gym	Basketball Court	Supermarket	Sandwich Place
21	22 - Jurong, Tuas	Japanese Restaurant	Asian Restaurant	Fast Food Restaurant	Café	Chinese Restaurant	Food Court	Coffee Shop	Dessert Shop	Exhibit	Shopping Mall
22	23 - Hillview, Dairy Farm, Bukit Panjang, Choa...	Coffee Shop	Food Court	Fast Food Restaurant	Supermarket	Park	Asian Restaurant	Shopping Mall	Sushi Restaurant	Noodle House	Bus Station
24	25 - Kranji, Woodgrove, Woodlands	Racetrack	Food Court	Grocery Store	Coffee Shop	Bakery	Gaming Cafe	Monument / Landmark	Food	Go Kart Track	Hainan Restaurant
26	27 - Yishun, Sembawang	Coffee Shop	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Fried Chicken Joint	Thai Restaurant	Café	Hainan Restaurant	Sushi Restaurant	Vegetarian / Vegan Restaurant
27	28 - Seletar Road	Food Court	Asian Restaurant	Coffee Shop	Fast Food Restaurant	Café	Supermarket	Dessert Shop	American Restaurant	Shopping Mall	Sandwich Place

Figure 18: Districts and Most Common Venues in Singapore's Cluster 4

The clustering algorithm has determined that there are 3 clusters in Hong Kong. By doing additional research and having a closer look at the data, most districts in Hong Kong have similar venues, except for 2 districts that are clearly different from the rest.

### Cluster 1 – Cluster with Varied F&B Options and Commercial Facilities

	District	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	Central and Western	Japanese Restaurant	French Restaurant	Café	Italian Restaurant	Hotel	Cocktail Bar	Coffee Shop	Gym / Fitness Center	Thai Restaurant	Yoga Studio
1	Eastern	Café	Japanese Restaurant	Chinese Restaurant	Coffee Shop	Park	Department Store	Noodle House	Seafood Restaurant	French Restaurant	Shopping Mall
2	Southern	Chinese Restaurant	Furniture / Home Store	Café	Park	Restaurant	Thai Restaurant	Coffee Shop	Fast Food Restaurant	Vietnamese Restaurant	Trail
3	Wan Chai	Hotel	Café	Coffee Shop	Italian Restaurant	Chinese Restaurant	Cantonese Restaurant	Szechuan Restaurant	Burger Joint	Seafood Restaurant	Snack Place
4	Sham Shui Po	Noodle House	Chinese Restaurant	Dessert Shop	Hong Kong Restaurant	Cha Chaan Teng	Café	Dim Sum Restaurant	Coffee Shop	Malay Restaurant	Italian Restaurant
5	Kowloon City	Coffee Shop	Dessert Shop	Thai Restaurant	Chinese Restaurant	Café	Fast Food Restaurant	Hotel	Cha Chaan Teng	Noodle House	Asian Restaurant
6	Kwun Tong	Chinese Restaurant	Coffee Shop	Park	Cha Chaan Teng	Café	Hong Kong Restaurant	Hotel	Noodle House	Shopping Mall	Hakka Restaurant
7	Wong Tai Sin	Dessert Shop	Coffee Shop	Thai Restaurant	Chinese Restaurant	Fast Food Restaurant	Café	Cha Chaan Teng	Noodle House	Asian Restaurant	Cantonese Restaurant
8	Yau Tsim Mong	Coffee Shop	Noodle House	Cha Chaan Teng	Dim Sum Restaurant	Hotel	Hong Kong Restaurant	Beer Bar	Cantonese Restaurant	Dumpling Restaurant	Toy / Game Store
11	North	Fast Food Restaurant	Cantonese Restaurant	Chinese Restaurant	Coffee Shop	Noodle House	Burger Joint	Café	Shopping Mall	Metro Station	Cha Chaan Teng
12	Sai Kung	Seafood Restaurant	Café	Thai Restaurant	Harbor / Marina	Dessert Shop	Burger Joint	Chinese Restaurant	Pub	Coffee Shop	Pizza Place
13	Sha Tin	Chinese Restaurant	Fast Food Restaurant	Café	Cantonese Restaurant	Shopping Mall	Coffee Shop	Hong Kong Restaurant	Train Station	Hotel	Dessert Shop
14	Tai Po	Chinese Restaurant	Fast Food Restaurant	Shopping Mall	Cha Chaan Teng	Hong Kong Restaurant	Coffee Shop	Plaza	Bus Station	Athletics & Sports	Korean Restaurant
15	Tsuen Wan	Chinese Restaurant	Fast Food Restaurant	Bus Stop	Shopping Mall	Café	Hong Kong Restaurant	Park	Cantonese Restaurant	Coffee Shop	Cha Chaan Teng
16	Tuen Mun	Seafood Restaurant	Fast Food Restaurant	Coffee Shop	Light Rail Station	Shopping Mall	Pizza Place	Multiplex	Cantonese Restaurant	Chinese Restaurant	Park
17	Yuen Long	Chinese Restaurant	Noodle House	Fast Food Restaurant	Dessert Shop	Shopping Mall	Thai Restaurant	Coffee Shop	Market	Café	Light Rail Station

Figure 19: Districts and Most Common Venues in Hong Kong's Cluster 1

### Cluster 2 – Cluster with Nature-Themed Facilities (outlier)

	District	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	Islands	Trail	Mountain	Zoo	Dumpling Restaurant	Flea Market	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market

Figure 20: Districts and Most Common Venues in Hong Kong's Cluster 2

### Cluster 3 – Cluster with Suburban Facilities & Transport Hub (outlier)

	District	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
10	Kwai Tsing	Scenic Lookout	Bus Station	Café	Beach	Trail	Electronics Store	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant

Figure 21: Districts and Most Common Venues in Hong Kong's Cluster 3

### 3.2 Machine Learning Algorithm

K-means is used as the clustering algorithm which clusters data by trying to separate samples in  $n$  groups of equal variances, minimising the distortion/ inertia (i.e. measure of how internally coherent clusters are).

The elbow method is used to determine the optimal number of clusters (i.e.  $k$ ) to use for the clustering process. Distortions/ inertias are plotted against the different  $k$ s used to indicate visually the optimal  $k$ .

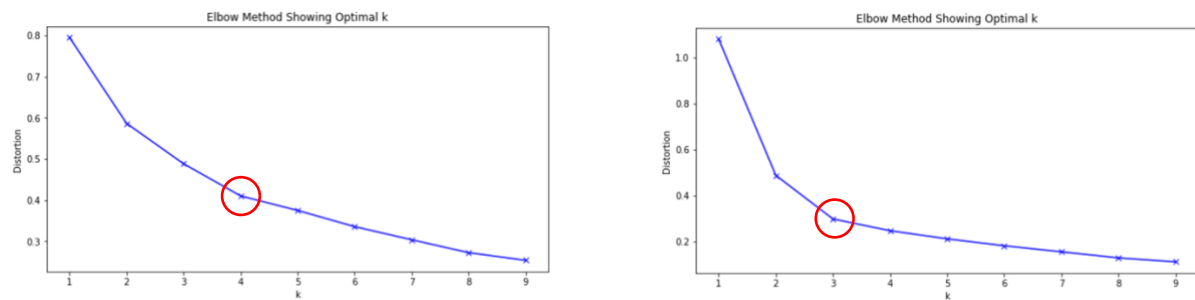


Figure 22: Optimal ks for Singapore (left) and Hong Kong (right)

## 4 Results

The results have highlighted the distinctness of each city. One probable hypothesis is that Singapore's districts, facilities within the districts and their compositions are developed based on their proximities to the CBD – you will find more facilities catered to a working crowd when the district is closer to the CBD and more facilities catered for daily amenities when the district is further away from the CBD. Maturity of the districts (or housing estates) will also determine the facility types located in the districts – more specialised F&B options are found in matured estates, probably due to the fact that certain ethnic groups and demographics have settled there over the years, thereby justifying for more specialised F&B options to cater to their food preferences.

On the other hand, most Hong Kong's districts do not have a clear differentiation between commercial and residential facilities, which suggests that the districts are typically mixed developments and geographically less spaced out. Compared to Singapore, Hong Kong's districts are more homogeneous.

The following table highlights similar clusters between Singapore and Hong Kong, based on the findings uncovered by the clustering algorithm and subsequent analysis.

Singapore	Hong Kong
Cluster 1 – Cluster with Commercial Facilities and Multinational F&B Options	Cluster 1 – Cluster with Varied F&B Options and Commercial Facilities  (Note: This cluster appears to be an amalgamation of Singapore's Clusters 1, 3 and 4.)
Cluster 3 – Cluster with Residential Facilities and More Individual-Themed Restaurants (i.e. more matured estates, closer to CBD)	
Cluster 4 – Cluster with Residential Facilities and More Commoditised F&B Options (less matured estates, further away from CBD)	
Cluster 2 – Cluster with Nature-Themed Facilities (outlier)	Cluster 2 – Cluster with Nature-Themed Facilities (outlier)
Nil	Cluster 3 – Cluster with Suburban Facilities & Transport Hub (outlier)

Table 2: Cluster Comparison between Singapore and Hong Kong

## 5 Discussion

Going through the above workflows has uncovered areas where changes can be made to improve the results of the comparative analysis. The experience gained will also be useful for conducting similar analysis based on machine learning algorithms.

The key observations and learnings are described in the following paragraphs:

- Data preparation, which included finding appropriate data sources, cleaning data, etc., took more time than expected. Try allocating more time for this task for future data science projects.
- Data retrieval from online resources was not always successful, dependent on the sites' reliability, slow and had access limits. If having the most updated data is not critical, try storing retrieved data in a persistent format (e.g. CSV files) for subsequent read operations. This is especially helpful during debugging/ testing when multiple read operations are required.
- Current analysis leveraged only recommended venues and their frequencies of occurrence as features during the clustering process. Inclusion of additional features can improve the clustering results, such as population density, number of residents, land area, demographics information, etc.
- A standard search radius was used for all districts to find recommended venues via the Foursquare API on Explore. Try using specific search radius for each district to increase the accuracy of the search results.

## 6 Conclusion

The comparative analysis has provided another perspective on how Singapore and Hong Kong are similar/ different, based on information gathered from public users. Machine learning has provided an efficient method to process large amount of data for analysis purposes – this automates many tasks and minimises the time taken to complete them, which then frees up valuable resources to focus on higher-value work (e.g. drawing conclusions based on patterns defined automatically by machine learning algorithm).

This concludes the report on “Comparative Analysis between Singapore and Hong Kong”. Contents have been developed as part of a learning and training exercise and hence should not be further distributed for non-educational purposes.