

Applied Data Science Capstone

Final Report

Singapore vs Hong Kong – A Tale of Two Cities

Mohammad Owais KERNEY
4-6-2020

Table of Contents

1. Introduction.....	2
1.1 Background.....	2
1.2 Problem	2
1.3 Project Aim	2
2. Data Acquisition, Cleaning and Relevance.....	3
2.1 Data Source	3
2.2 Data Cleaning and Feature Selection	3
2.3 Data Relevance	3
3. Methodology	4
3.1 Exploratory data analysis.....	4
3.1.1 Observing and Verifying Venue/Facility Data for Hong Kong from Foursquare	4
3.1.2 Observing and Verifying Venue/Facility Data for Singapore from Foursquare	4
3.1.3 Observing the number of venue categories of data obtained from Foursquare	4
3.2 Data Transformation	4
3.2.1 Performing OneHot encoding	4
3.2.2 Grouping the data and finding top venues.....	5
3.3 Clustering	5
4. Results.....	7
1. Discussion	8
5. Conclusion and Future Direction	9
References	10

1. Introduction

1.1 Background

Hong Kong and Singapore are the forerunners in the race to become 'Asia's World City' and the competition is fierce. The cities often find themselves trying to win bids to host major sports and entertainment events and give each other a tough fight. Media articles are filled with tales of this rivalry and several factors are used to compare the two cities. Lifestyle advantages, cost of residence and quality of life for expats are often some of the decisive ones [1]

It is important to highlight how similar these cities are despite the competition and rivalry. The similarities in culture, economic structure and situation of living make it hard to believe these cities are in fact part of separate states. Both economies depend heavily on the financial services industry and fintech and are employed as the regional headquarters for several foreign companies operating in Asia [2]

Due to this similitude, migration between Hong Kong and Singapore is a very common occurrence. Locals and expats looking for better job opportunities in the region have been known to move between these cities, often several times in their career. Be it better schooling systems, improving economic situation or changing immigration policies, several dynamic factors can lead to a family deciding to move between these cities. In mid-2019, migration from Hong Kong to Singapore peaked due to the political turmoil in Hong Kong and the widespread and sometimes violent protests in the city. Migration consultants were employed by Hong Kong families to arrange their movement as smoothly as possible [3]

1.2 Problem

When migrating from one city to another, people look at several factors in order to find a suitable district to live in. A big part of this search is a comparison of their current neighbourhood/district with others in the new city, to find similarities and settle in a place which has similar amenities and environment. For example, a family with school going children will want to move to a district with shorter distances to schools, family friendly housing estates and recreational facilities.

This research is a lot tougher than it seems. Simple search engine queries can contain incomplete or outdated information. Most blogs/articles focus on a general inter-city comparison which does not help much when it comes to comparing neighbourhoods with a precise level of detail. Comparison of neighbourhoods then ends up requiring several physical visits to the new city which results in additional overhead costs and stress.

1.3 Project Aim

This project aims to evaluate the similarity between Hong Kong and Singapore by analysing how several factors make districts in these two cities similar. For families and expats thinking about switching their city or for migration consultants looking to provide recommendations to

clients, this project will provide a detailed statistical analysis and recommend which district would be suitable to live in.

The project aims to make life easier for the migrating families and/or the consultants helping them rationally locate the right district to migrate to so they don't end up having a tough or elongated adjustment period.

In practical terms, the project aims to carry out a cluster analysis of districts in the two cities and group them in terms of similitude. This way, the migrating party can locate a district which is similar to the district they currently live in and have a smoother transition.

2. Data Acquisition, Cleaning and Relevance

2.1 Data Source

At first, data related to the districts and their location in both cities was obtained from Wikipedia and [GeoHack](#).

Then [Foursquare](#) API was then used to get venue data for all the districts in the two cities. The 'search' endpoint was used to get every single amenity/venue/facility in the respective districts.

2.2 Data Cleaning and Feature Selection

It was difficult to find latitude and longitude data for the two cities' districts and hence several data sources were used to verify the information collected.

In addition, since some of Singapore's districts do not have official names or identities and are just referred to as a group of popular neighbourhoods in the district, just one of the neighbourhoods was selected from these districts to get the latitude and longitude.

For each venue/facility found in the districts, we had the following features:

1. Name
2. Latitude
3. Longitude
4. Category (clinic, school, restaurant etc)
5. City

Only Name and Place Type are relevant data features for our analysis while the rest of the features are used for identification purposes only.

2.3 Data Relevance

The focus on venues and facilities in each district will help us determine which districts are similar in terms of the number and type of facilities found. This will help a migrating facility

find a suitable district in terms of the amenities they need and will lead to a smoother migration for them.

3. Methodology

3.1 Exploratory data analysis

3.1.1 Observing and Verifying Venue/Facility Data for Hong Kong from Foursquare

The following dataframe was extracted:

	Name	Category	Latitude	Longitude	City	District	District Latitude	District Longitude
0	Immigration Department Travel Documents Issuin...	Government Building	22.286864	114.155123	Hong Kong	Central and Western	22.28666	114.15497
1	Harbour Building (海港政府大樓)	Government Building	22.286592	114.154997	Hong Kong	Central and Western	22.28666	114.15497
2	Marine Department 海事處	Government Building	22.286571	114.154943	Hong Kong	Central and Western	22.28666	114.15497
3	Infinitus Plaza (無限極廣場)	Shopping Mall	22.286463	114.153689	Hong Kong	Central and Western	22.28666	114.15497
4	Fitness First Platinum	Gym / Fitness Center	22.285881	114.154117	Hong Kong	Central and Western	22.28666	114.15497

The total venues were 15,824.

3.1.2 Observing and Verifying Venue/Facility Data for Singapore from Foursquare

The following dataframe was extracted:

	Name	Category	Latitude	Longitude	City	District	District Latitude	District Longitude
0	Arcade Money Changers	Currency Exchange	1.283997	103.851197	Singapore	District 01 - Raffles Place, Marina, Cecil	1.28372	103.851239
1	Clifford Centre	Office	1.283816	103.852219	Singapore	District 01 - Raffles Place, Marina, Cecil	1.28372	103.851239
2	City Money Changers	Financial or Legal Service	1.283475	103.850161	Singapore	District 01 - Raffles Place, Marina, Cecil	1.28372	103.851239
3	Ocean Financial Centre	Building	1.283260	103.851521	Singapore	District 01 - Raffles Place, Marina, Cecil	1.28372	103.851239
4	In Touch Physio	Medical Center	1.283503	103.851163	Singapore	District 01 - Raffles Place, Marina, Cecil	1.28372	103.851239

The total venues were 26,800.

3.1.3 Observing the number of venue categories of data obtained from Foursquare

There were 304 and 386 different categories in Hong Kong and Singapore data respectively.

3.2 Data Transformation

3.2.1 Performing OneHot encoding

OneHot encoding was performed to transform the dataset with Categories set as columns. It looked like the following:

	District	ATM	Accessories Store	Acupuncturist	Adult Education Center	Advertising Agency	African Restaurant	Airport	Airport Gate	Airport Lounge	...	Waterfront	Well	Wine Bar	Wine Shop	Winery	Wings Joint
0	Central and Western	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
1	Central and Western	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
2	Central and Western	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3	Central and Western	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
4	Central and Western	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

3.2.2 Grouping the data and finding top venues

The data was then grouped by District and the top ten venues were found for each district as shown below:

	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	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space	Lawyer	Dentist's Office	Gym / Fitness Center	Coffee Shop
1	District 01 - Raffles Place, Marina, Cecil	Office	Bakery	Bank	Building	Café	Salad Place	Health & Beauty Service	Medical Center	Japanese Restaurant	Asian Restaurant
2	District 02 - Tanjong Pagar, Chinatown	Japanese Restaurant	Office	Café	Coffee Shop	Building	Sandwich Place	Asian Restaurant	Chinese Restaurant	Cosmetics Shop	Park
3	District 03 - Tiong Bahru, Alexandra, Queenstown	College Classroom	Building	Coffee Shop	Chinese Restaurant	Residential Building (Apartment / Condo)	Asian Restaurant	Food Court	Dentist's Office	Medical Center	Nail Salon
4	District 04 - Mount Faber, Telok Blangah, Harb...	Clothing Store	Chinese Restaurant	Office	Cosmetics Shop	Jewelry Store	Food Court	Boutique	Café	Sporting Goods Shop	Housing Development

3.3 Clustering

We have chosen to cluster the districts using k-means clustering and we have chosen to create 4 clusters.

K-Means clustering was chosen as the machine learning approach because it is simple and less expensive in terms of time complexity. It reduces the space into disjoint smaller sub-spaces, hence saving resources.

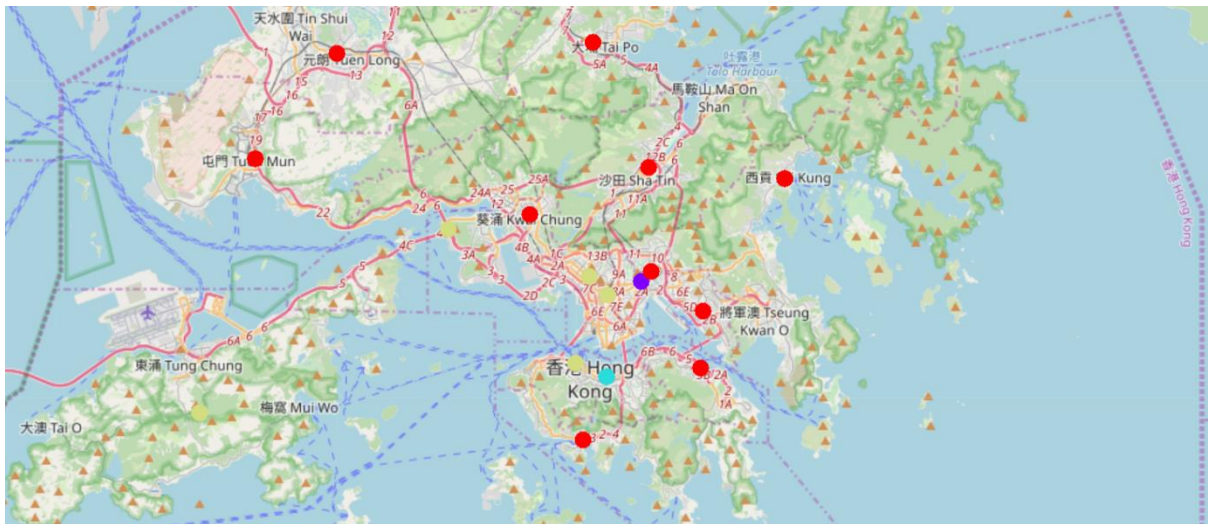
As for the value of k, we experimented with k=3,4,5,6,7,8,9 and 10 but we repeatedly found that after k=4, the new clusters formed were having just one district each or no district at all which indicated that k=4 was the optimum value.

The result of the clustering looked like the following:

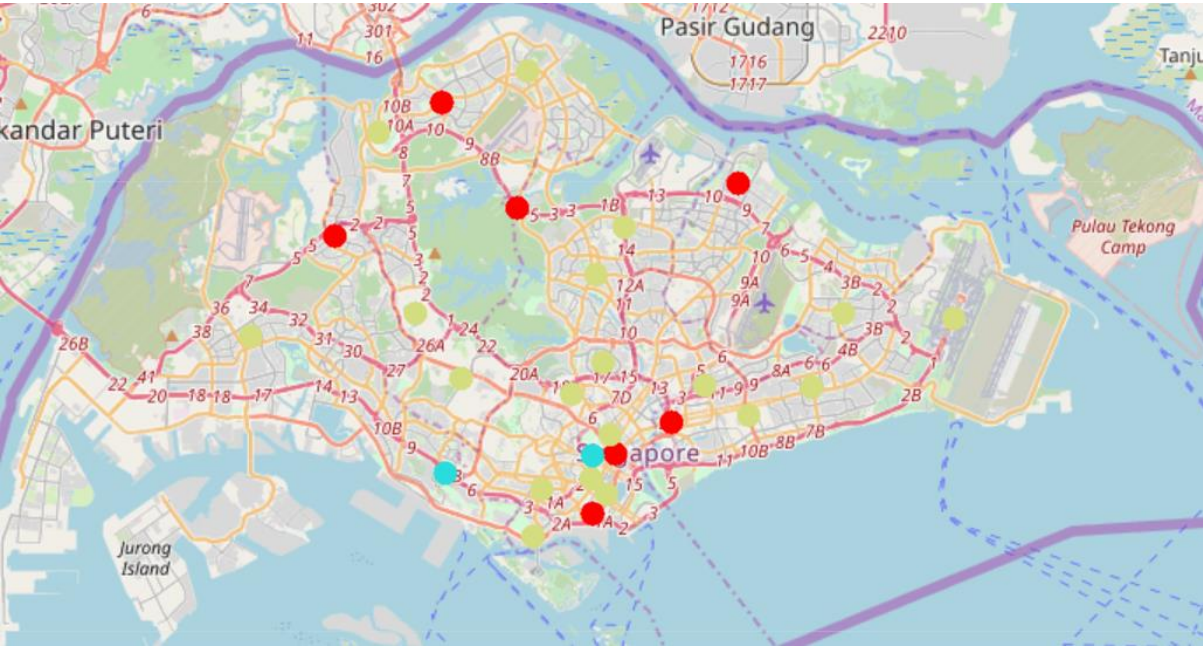
	Name	Category	Latitude	Longitude	City	District	District Latitude	District Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Immigration Department Travel Documents Issuin...	Government Building	22.286864	114.155123	Hong Kong	Central and Western	22.28666	114.15497	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space
1	Harbour Building (海 港政府大樓)	Government Building	22.286592	114.154997	Hong Kong	Central and Western	22.28666	114.15497	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space
2	Marine Department 海事處	Government Building	22.286571	114.154943	Hong Kong	Central and Western	22.28666	114.15497	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space
3	Infinitus Plaza (無限 極廣場)	Shopping Mall	22.286463	114.153689	Hong Kong	Central and Western	22.28666	114.15497	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space
4	Fitness First Platinum	Gym / Fitness Center	22.285881	114.154117	Hong Kong	Central and Western	22.28666	114.15497	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space

The results, visualized using Folium, are as follows:

Hong Kong



Singapore



4. Results

The two largest clusters had the following number of districts:

- Cluster 1: 18
- Cluster 4: 24

The clusters can be seen below:

Cluster 1

	Category	District	District Latitude	District Longitude	Cluster Labels	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
0	Courthouse	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop
1	Housing Development	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop
2	Bus Station	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop
3	Police Station	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop
4	Park	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop
5	Grocery Store	Eastern	22.284110	114.224140	0	Chinese Restaurant	Housing Development	Hong Kong Restaurant	Café	Convenience Store	Park	Japanese Restaurant	Bakery	Coffee Shop

Cluster 2

	Category	District	District Latitude	District Longitude	Cluster Labels	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
0	Thai Restaurant	Kowloon City	22.3282	114.19155	1	Thai Restaurant	Dessert Shop	Chinese Restaurant	Cha Chaan Teng	Grocery Store	Hotpot Restaurant	Restaurant	Seafood Restaurant	BBQ Joint
1	Bus Stop	Kowloon City	22.3282	114.19155	1	Thai Restaurant	Dessert Shop	Chinese Restaurant	Cha Chaan Teng	Grocery Store	Hotpot Restaurant	Restaurant	Seafood Restaurant	BBQ Joint
2	Seafood Restaurant	Kowloon City	22.3282	114.19155	1	Thai Restaurant	Dessert Shop	Chinese Restaurant	Cha Chaan Teng	Grocery Store	Hotpot Restaurant	Restaurant	Seafood Restaurant	BBQ Joint
3	Thai Restaurant	Kowloon City	22.3282	114.19155	1	Thai Restaurant	Dessert Shop	Chinese Restaurant	Cha Chaan Teng	Grocery Store	Hotpot Restaurant	Restaurant	Seafood Restaurant	BBQ Joint

Cluster 3

	Category	District	District Latitude	District Longitude	Cluster Labels	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
0	Government Building	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café
1	Government Building	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café
2	Post Office	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café
3	None	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café
4	Hotel	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café
5	Café	Wan Chai	22.279680	114.171680	2	Office	Government Building	Building	Bar	Bus Stop	Automotive Shop	Auto Dealership	General Entertainment	Café

Cluster 4

	Category	District	District Latitude	District Longitude	Cluster Labels	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
0	Government Building	Central and Western	22.286660	114.154970	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space	Lawyer	Dentist's Office	Gym / Fitness Center
1	Government Building	Central and Western	22.286660	114.154970	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space	Lawyer	Dentist's Office	Gym / Fitness Center
2	Government Building	Central and Western	22.286660	114.154970	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space	Lawyer	Dentist's Office	Gym / Fitness Center
3	Shopping Mall	Central and Western	22.286660	114.154970	3	Building	Government Building	Doctor's Office	Office	Bus Stop	Coworking Space	Lawyer	Dentist's Office	Gym / Fitness

1. Discussion

From the cluster analysis, we found that most of the districts in Hong Kong and Singapore can be classified into two main categories which goes on to show the similarity between the two cities. The most popular cluster, Cluster 4, is comprised of Government buildings and amenities with services such as clinics and law firms in the vicinity. Less popular venues in this cluster include gyms and fitness centers.

The second most popular cluster, Cluster 1, consists mainly of public housing, parks and convenience stores and seems like a cluster that a family would consider residing in.

The k-means cluster analysis has identified similar and dissimilar districts in Singapore and Hong Kong and the final visualization map makes a good resource for families/expats moving between the cities to identify the type of neighbourhood they want to live in and the types of amenities they can expect in that area.

5. Conclusion and Future Direction

The k-means clustering algorithm in this project was fed location and venue data from Foursquare. It was effective in identifying similar neighbourhoods in the two cities. However, to improve the model, other factors such as population of the districts as well as crime rates would be an interesting input enhancement to consider as a future direction.

Moreover, the model could be narrowed down to neighbourhoods instead of districts to find more specific areas for residence and that model would be more suitable for migration consultants or real-estate agents serving people in the area.

References

[1] How the Hong Kong vs Singapore thing got started, and why Lion City doesn't indulge in such petty comparisons

South China Morning Post - Jason Wordie

<https://www.scmp.com/magazines/post-magazine/article/2096458/how-hong-kong-vs-singapore-thing-got-started-and-why-lion>

[2] Frenemy cities: a look at Hong Kong, Singapore's rivalry and partnership as Lion City marks key anniversary

South China Morning Post – Gary Cheung

<https://www.scmp.com/news/hong-kong/economy/article/1847695/hong-kong-and-singapores-transformations-story-both-rivalry>

[3] Emigrating Hongkongers seek stability elsewhere, heading to Australia, Canada, Singapore

South China Morning Post – Fiona Sun

<https://www.scmp.com/news/hong-kong/politics/article/3023166/port-storm-emigrating-hongkongers-look-for-stability-elsewhere>