**Comparative Analysis between Singapore and Hong Kong**

Chan Kok Siang

26 May 2020

1. Introduction
   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.

A sunset over a body of water with a city in the background

Description automatically generatedA view of a city at night

Description automatically generated

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

|  |  |  |
| --- | --- | --- |
|  | Singapore | Hong Kong |
| **Flag** | A picture containing drawing, food  Description automatically generated | A picture containing drawing  Description automatically generated |
| **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 km2 | 2,755 km2 |
| **Population (2018)** | Total: 5,638,700  Density: 7,804/km2 | Total: 7,482,500  Density: 6,777/km2 |
| **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. 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. 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.

1. Data
   1. Data Sources

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

A screenshot of a cell phone

Description automatically generated

Figure 2: Singapore Postal Districts (source: Wikipedia)

A screenshot of a cell phone

Description automatically generated

Figure 3: Hong Kong Districts (source: Wikipedia)

A screenshot of a cell phone

Description automatically generated

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

A screenshot of a social media post

Description automatically generated

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

* 1. 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.

* 1. 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.

1. Methodology
   1. Data Analysis

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

* + 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.

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

Figure 6: Sample Dataframes Created from Source HTML Pages

* + 1. 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.

A screenshot of a cell phone

Description automatically generated

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.

A screenshot of a cell phone

Description automatically generated

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

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

A screenshot of a cell phone

Description automatically generated

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.

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

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

Coordinates are then appended to the original dataframe.

A screenshot of a cell phone

Description automatically generated

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

* + 1. 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

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

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

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

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

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

A picture containing text, map

Description automatically generated A close up of a map

Description automatically generated

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**

A screenshot of a cell phone

Description automatically generated

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

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



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)**

A close up of text on a white background

Description automatically generated

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)**

A close up of text on a white background

Description automatically generated

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**

A close up of a piece of paper

Description automatically generated

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

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



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

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



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

* 1. 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 ks used to indicate visually the optimal k.

A close up of a map

Description automatically generated A close up of a map

Description automatically generated

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

1. 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

1. 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.

1. 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.