

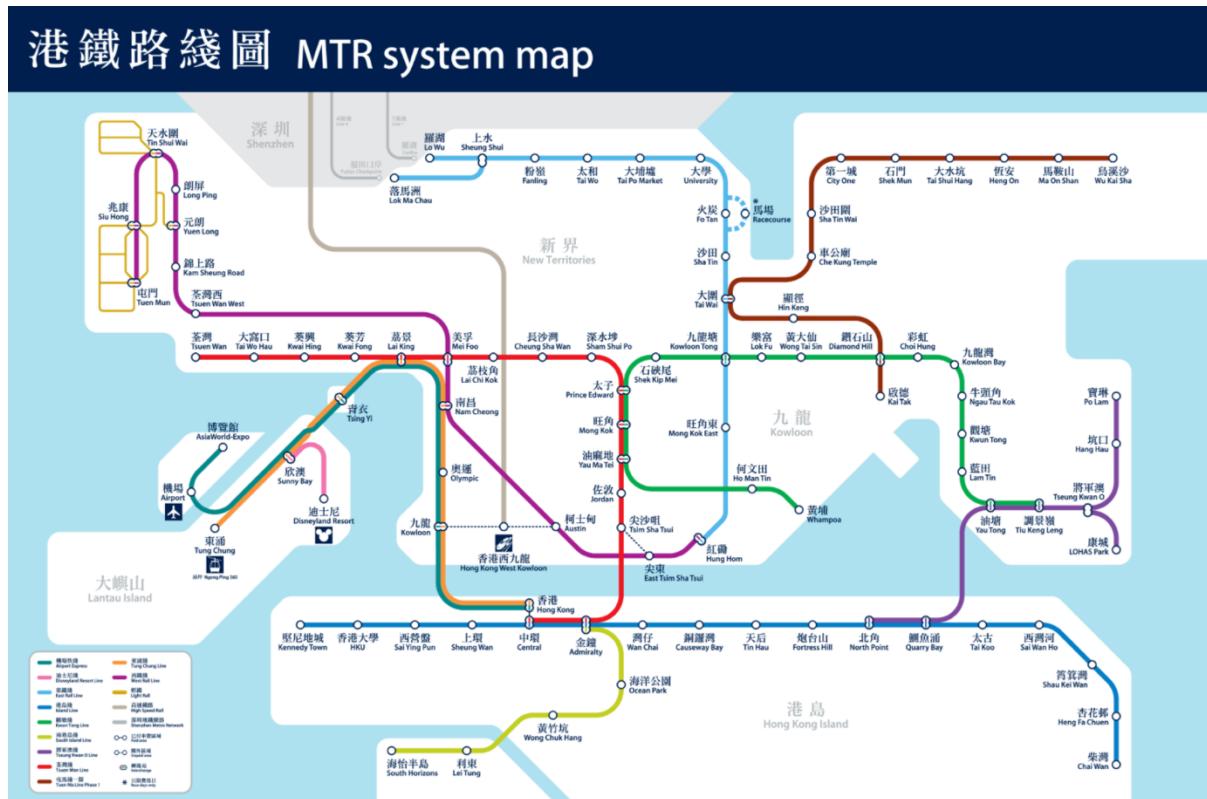
A Study of Neighbour of MTR Stations in Hong Kong

1. Introduction

Hong Kong is one of the major financial cities in the world, a former British colony and now a Special Administrative Region of the People's Republic of China. It has over 7.5 millions residents and is one of the most densely populated places in the world.

From report of the Hong Kong government over 90 percent of daily trips are made on public transport. The major public transport network is the Mass Transit Railway (MTR), and it has an average weekday patronage of about 5.6 million passengers. Over the years it has developed into a 11 lines of heavy rail plus 10 lines of light rail commuter network. Most parts of the city are covered by heavy rail while only a small area in the north-west is covered by additional light rail.

As a resident of the city I would like to explore the neighbours around MTR stations in the city, and see how different or alike they are. From which we will try to analyze what local business is popular among the neighbourhoods, and evaluate the business prospects for opening a local business.



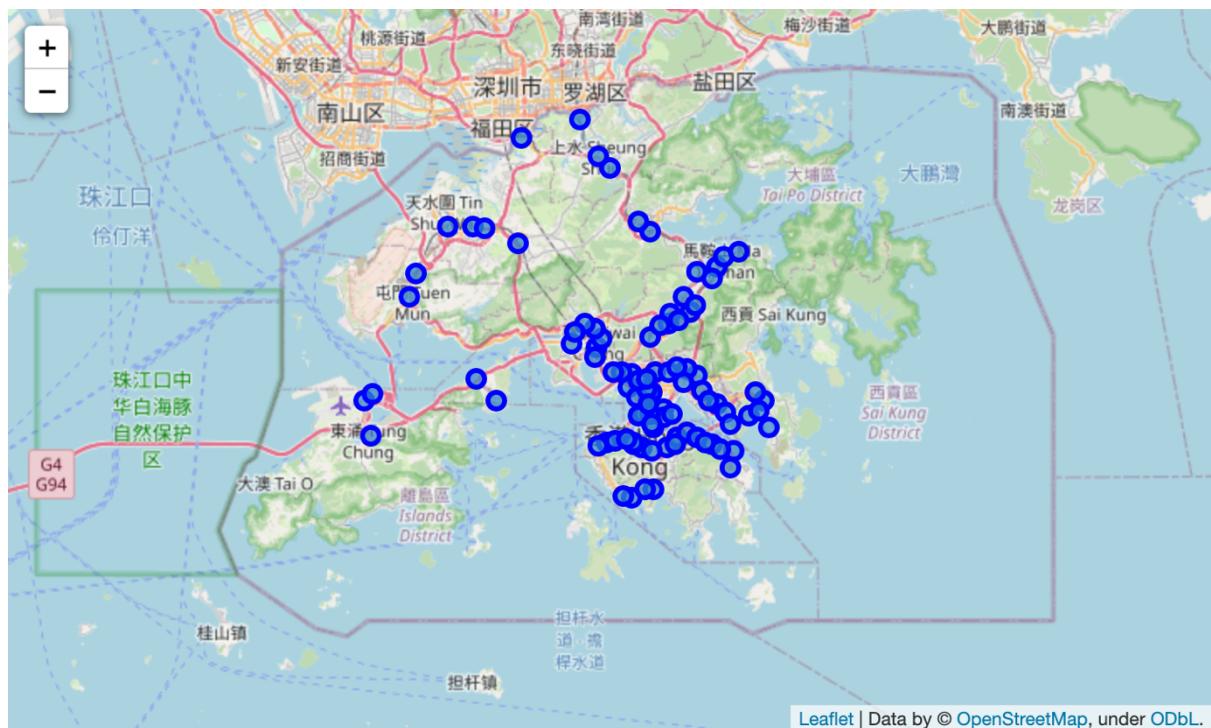
2. Data Acquisition

From DATA.GOV.HK on the Hong Kong Government website, we can obtain different published data. This includes MTR routes, fares and barrier-free facilities data set.

I first obtained data downloaded from data.gov.hk website to retrieve the MTR network information. There are in total 94 stations and 11 lines. I used geopy to load coordinates of all stations, and create a table of coordinates of all MTR stations. From the results I discovered quite a few coordinates out of positions and geopy even failed to retrieve Whampoa station coordinates. Then I verified the coordinates of each station with coordinates on Google Map, and finalized a table of appropriate coordinates for all stations as below.

	Station	Latitude	Longitude
0	Admiralty	22.279643	114.164590
1	Airport	22.316087	113.936478
2	AsiaWorld-Expo	22.321251	113.942971
3	Austin	22.305491	114.166135
4	Causeway Bay	22.280621	114.185087
...
89	Wong Tai Sin	22.341654	114.193859
90	Wu Kai Sha	22.429171	114.243983
91	Yau Ma Tei	22.313617	114.170540
92	Yau Tong	22.298172	114.236941
93	Yuen Long	22.446469	114.034680

Here is the visualization of MTR stations upon a map of Hong Kong.



We utilized the Foursquare API with the coordinates data to explore neighbourhoods of each station within a range of radius 300 meters. It resulted in 3280 venues listed as below :

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Admiralty	22.279643	114.16459	Ruth's Chris Steak House (茹絲葵牛排餐廳)	22.279188	114.163833	Steakhouse
1	Admiralty	22.279643	114.16459	Tamar Park (添馬公園)	22.281361	114.165583	Park
2	Admiralty	22.279643	114.16459	Lane Crawford (連卡佛)	22.277843	114.165009	Furniture / Home Store
3	Admiralty	22.279643	114.16459	Pacific Place (太古廣場)	22.277696	114.165048	Shopping Mall
4	Admiralty	22.279643	114.16459	Great Food Hall	22.277923	114.164712	Supermarket

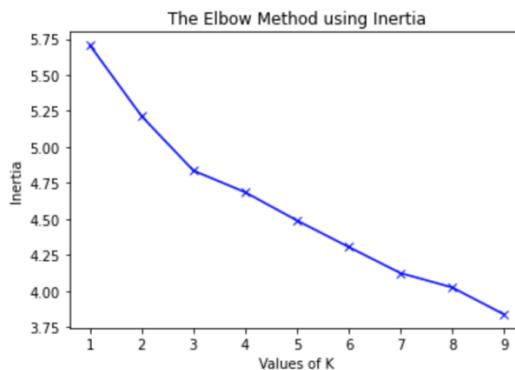
And grouping the data we found that there are 279 unique venue categories.

3. Methodology

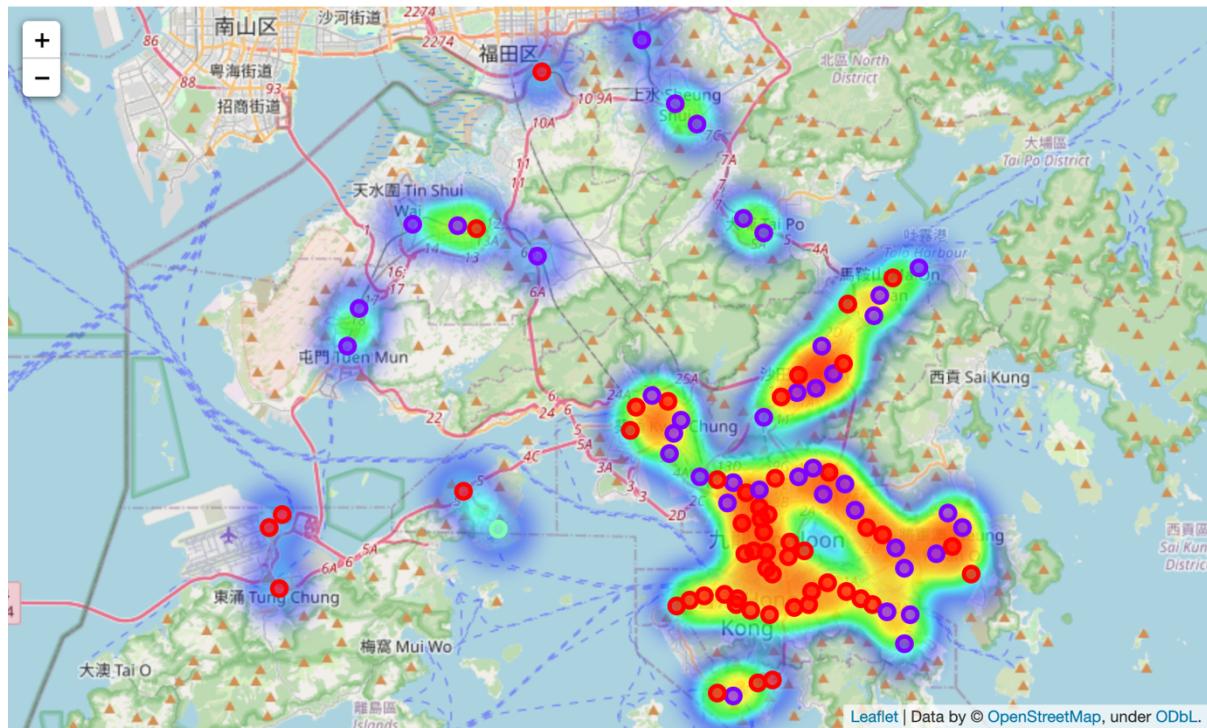
In order to determine the patterns of venues distribution in the neighbours, we will use methodology that can divide stations into different clusters without human supervision. Here all neighbourhood venues information for each station is collected, and we do not have any present label reference on stations.

In these circumstances I decided to use k-means clustering to process the data. I turned each record of venue against the venue categories to a one hot vector, then calculated the mean for each venue category for each station. We input the means (weights) of venue categories of each station in to the model, thus the model would divide stations into different clusters based on these weights.

After fitting the data to the k-means model I checked the inertias, according to the elbow method results showed the most appropriate number of clusters for the k-means model is 3.



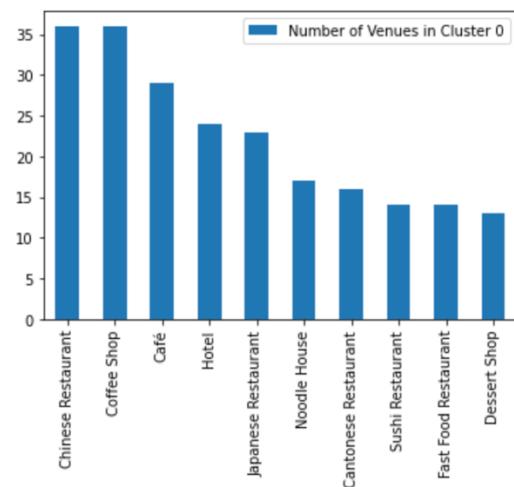
Applying k-means with a value of k equals to 3, the model returned labels of clusters for each stations and I plotted the results of all stations with color :



Red dots are stations in cluster 0, purple dots are stations in cluster 1 and the green dot is the only station in cluster 2. With the heat map one can also see the coverage of MTR stations network distributed in the city.

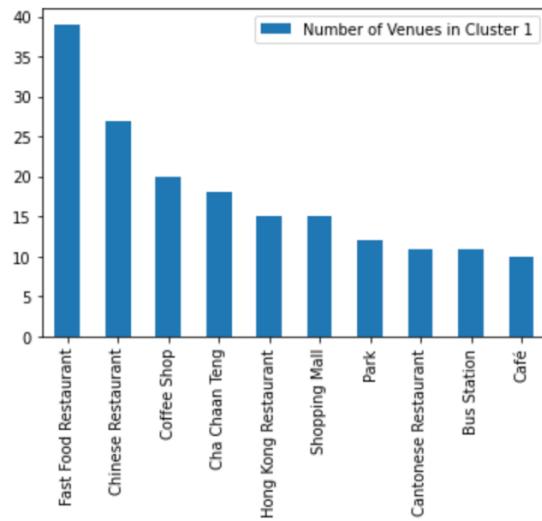
I listed out the top ten most common venue categories in each cluster and plotted them in bar charts below :

Number of Venues in Cluster 0	
Chinese Restaurant	36
Coffee Shop	36
Café	29
Hotel	24
Japanese Restaurant	23
Noodle House	17
Cantonese Restaurant	16
Sushi Restaurant	14
Fast Food Restaurant	14
Dessert Shop	13



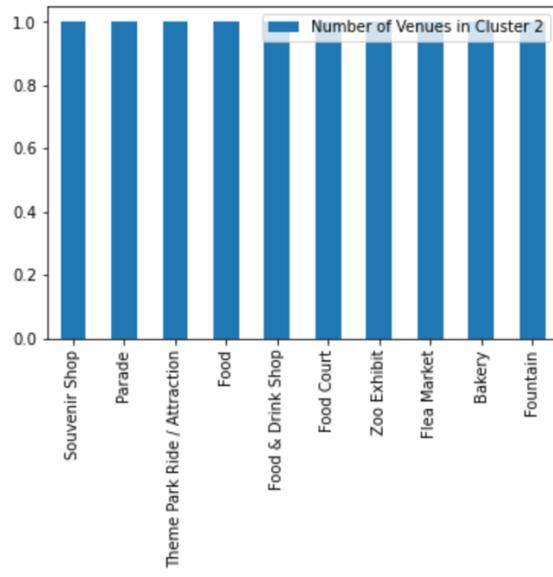
The top ten venue categories in Cluster 0 are Chinese Restaurant, Coffee Shop, Cafe, Hotel and Japanese Restaurant, Noodle House, Cantonese Restaurant, Fast Food Restaurant, Sushi Restaurant and Dessert Shop.

Number of Venues in Cluster 1	
Fast Food Restaurant	39
Chinese Restaurant	27
Coffee Shop	20
Cha Chaan Teng	18
Hong Kong Restaurant	15
Shopping Mall	15
Park	12
Cantonese Restaurant	11
Bus Station	11
Café	10



The top ten venues in Cluster 1 are Fast Food Restaurant, Chinese Restaurant, Coffee Shop, Cha Chaan Teng, Shopping Mall, Hong Kong Restaurant, Park, Bus Station, Cantonese Restaurant and Café.

Number of Venues in Cluster 2	
Souvenir Shop	1
Parade	1
Theme Park Ride / Attraction	1
Food	1
Food & Drink Shop	1
Food Court	1
Zoo Exhibit	1
Flea Market	1
Bakery	1
Fountain	1



The top ten venue categories in cluster 2 (Disneyland Resort) are Souvenir Shop, Parade, Theme Park Ride / Attraction, Food, Food & Drink Shop, Food Court, Zoo Exhibit, Flea Market, Bakery and Fountain.

4. Results and Discussion

In cluster 0, the stations in most busy commercial areas and around tourist attractions are included such as Admiralty, Central, Causeway Bay, Kwun Tong, Mong Kok, Quarry Bay and Tsim Sha Tsui. However some densely populated residential areas are also included such as North Point and Sai Wan Ho, etc. In cluster 1, most stations are in residential area but because of the densely populated environment in the city, there are also increasing numbers of commercial buildings in Kowloon Bay and Kwai Fong, etc. Cluster 2 only has station in Disneyland Resort, and opening business in Disneyland Resort is out of scope in this project.

In comparison of top 10 most common venue categories, the venue categories only exist in cluster 0 are Hotel, Japanese Restaurant, Noodle House, Sushi Restaurant and Dessert Shop.

- more commercial premises and tourist attractions lead to more hotels around these stations.
- Japanese restaurants (including Sushi restaurant) are usually pricey in the local market, so we see more of them in the neighbourhoods because we have more dense commercial areas and tourist attractions around.
- Noodle shop here includes different style of noodles here but they are usually for a working day quick meal.
- Dessert shops mainly serve dessert, especially Chinese style dessert soup or pudding, and usually sell with a higher unit price with regards to normal meal.

The venue categories only exist in cluster 1 are Cha Chaan Teng, Shopping Mall, Park and Bus Station.

- Cha Chaan Teng is the special type restaurant originated in Hong Kong that provides both local style cuisine and a mixture of Eastern and Western cuisines. Therefore more stations in the neighbourhoods of dense residential areas lead to more number of them. They serve food fulfilling most of the local residents' needs. For those interested to read more, please refer to : https://en.wikipedia.org/wiki/Cha_chaan_teng
- more stations in residential areas also lead to more local Shopping Mall, Park and Bus Station. These are typically what local residents need for their daily life.

Analysis from the data shows that opening a Japanese Restaurant, Dessert Shop or Cha Chaan Teng may suit in one cluster of neighbourhoods only. In the other cluster the lower number of such category already implicate a low demand of the service or product.

In contrast both clusters have significant numbers of Coffee Shop, Chinese Restaurant, Café, Cantonese Restaurant and Fast Food Restaurant in the neighbourhoods. Therefore in all neighbourhoods these categories of business have a large demand and can exist together at a same time.

There is one category in cluster 1 that appears to be overlapped with other venue categories. The "Hong Kong Restaurant" term seems ambiguous and let's inspect the first 50 of them :

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
204	Causeway Bay	22.280621	114.185087	Wing's Catering (榮式燒雞扒)	22.277385	114.186429	Hong Kong Restaurant
216	Causeway Bay	22.280621	114.185087	Kowloon Restaurant (九龍餐室)	22.279086	114.180639	Hong Kong Restaurant
331	Chai Wan	22.264783	114.237238	Yeh Lam Kwok (椰林閣)	22.263780	114.236988	Hong Kong Restaurant
344	Chai Wan	22.264783	114.237238	泉記金喜來海鮮飯店	22.265425	114.240492	Hong Kong Restaurant
356	Cheung Sha Wan	22.335578	114.156231	Sum's Cuisine & Restaurant (森記 餐室飯館)	22.336142	114.158479	Hong Kong Restaurant
369	Cheung Sha Wan	22.335578	114.156231	Redrice (Redrice 米飯主題餐廳)	22.337675	114.158110	Hong Kong Restaurant
382	Cheung Sha Wan	22.335578	114.156231	Rice Love (飯意主題餐廳)	22.334337	114.160764	Hong Kong Restaurant
401	City One	22.383017	114.203754	ItaMoMo Café (意八餐廳)	22.381653	114.205117	Hong Kong Restaurant
415	City One	22.383017	114.203754	叻哥茶餐廳	22.386703	114.204142	Hong Kong Restaurant
538	East Tsim Sha Tsui	22.295301	114.174600	Mui Kee Cookfood Stall (妹記大排檔)	22.299378	114.173436	Hong Kong Restaurant
574	Fortress Hill	22.287563	114.193315	協盛大排檔	22.285654	114.191834	Hong Kong Restaurant

From the list we can see they are actually Cha Chaan Teng or Cantonese Restaurant. In fact usually Cha Chaan Teng are positioned in the lower price range, branding themselves as a Hong Kong Restaurant may be more appropriate to their pricing strategies.

Coffee Shop and Café actually means the same type of business, and it indicates the demand for a cup of tea/coffee and quick meal is high too. A coffee shop requires much less space than a restaurant, so it costs less to operate one in comparison. And this is why in cluster 0 we have more Coffee Shop or Café, due to the expensive cost of rent in dense commercial areas.

5. Conclusion

The purpose of this project is to explore the neighbours around MTR stations in the city and evaluate the business prospects for opening a local business nearby. We extracted location data for each MTR stations, we used Foursquare API to prepare data of surrounding neighbourhoods, and then we used K-means clustering to build a model for further segmentation. After analysis of the collected data Coffee Shop, Chinese Restaurant, Café, Cantonese Restaurant and Fast Food Restaurant are the type of business in demand, and have better opportunities to sustain in the neighbourhood.

According to 2016 Population By-census (<https://www.bycensus2016.gov.hk/en/bc-mt.html>), 92% of the population of Hong Kong are ethnic Chinese. Although it did not categorize Han Chinese subgroups, it also explains the reason why this analysis results in Chinese Restaurant, especially Cantonese Restaurant, stands out among the venue categories. Indeed the Fast Food restaurants in the city serve a lot of Chinese dishes and Cantonese dishes as well.

Opening a local business will still need to consider a lot of factors such as investor's interest, wages level, competitors nearby, location from the MTR station, actual customer flow, etc. The analysis above should provide additional tools in the decision making during the process.