Expanding Dessert Business To One Or More Districts in Hong Kong

Priscilla Yip

1 June 2019

1. Introduction

i. Background

Hong Kong is named Gourmet Paradise not only because of the variety of cuisines, but also numerous good restaurants in this little area. The culinary business in Hong Kong is intensely competitive. To run a business successfully, location is of utmost importance. Among ground level location, shopping mall and upper level shop location, ground level location is usually preferred due to high footfall. However, this option is also the most expensive. That said, many business owners are still willing to invest the lion's share on rents in exchange for higher business volume.

ii. Problem Description

Mr A, a culinary business owner, opened a dessert shop in Mong Kok (in Yau Tsim Mong District) two years back. Since then, the dessert shop has developed a good reputation and starts making profits lately. Mr A would like to leverage on the momentum to open shops under the same brand in other areas. He believes that the location of his first shop is the secret to his success, therefore, he wants to open shops in areas with similar surrounding environment.

iii. Surrounding Environment

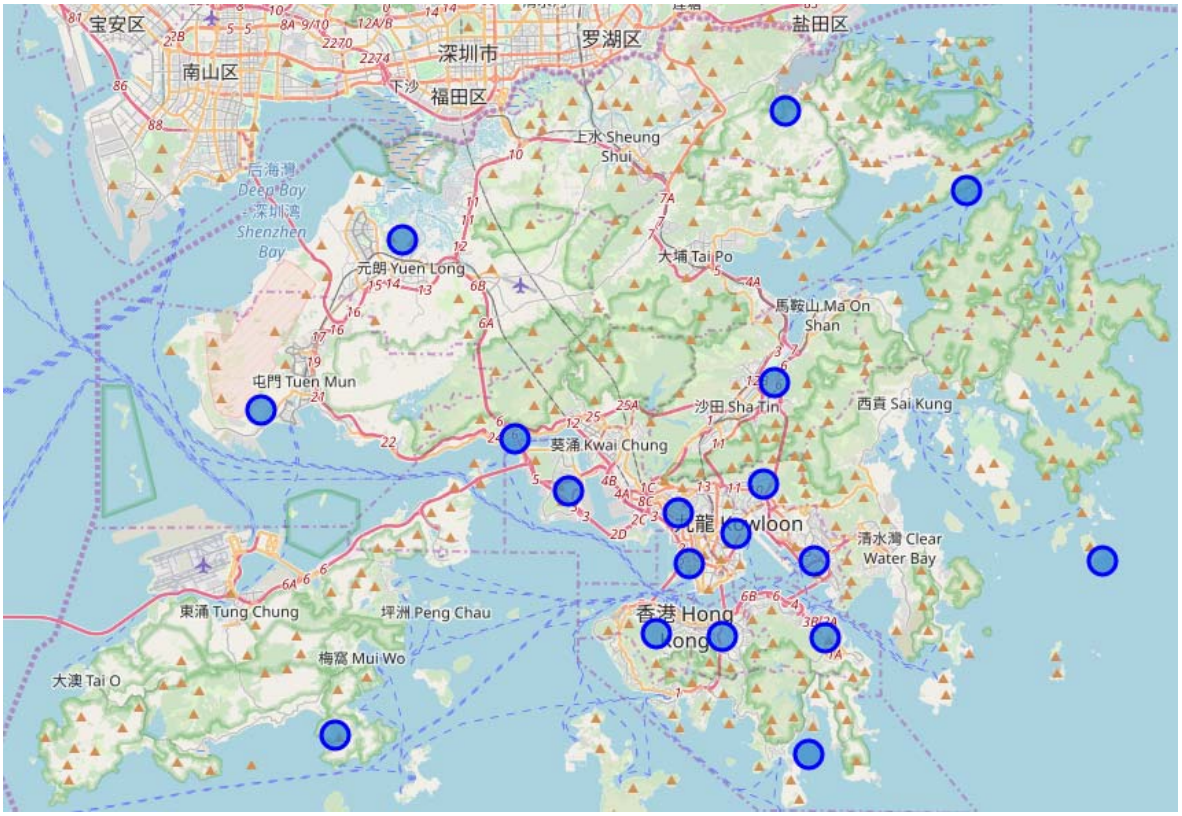
Desserts are appealing to Hongkongers, as well as tourists. However, people normally savor desserts after a lunch or dinner. Hence, the opening hours of these dessert shops are usually short. So as to compensate for the short opening period, it is very critical for these dessert shops to be opened in ground level areas, adjacent to other restaurants, shopping malls or other attractions with high footfall. Mr A is puzzling how to select the best locations based on these criteria.

2. Data Description

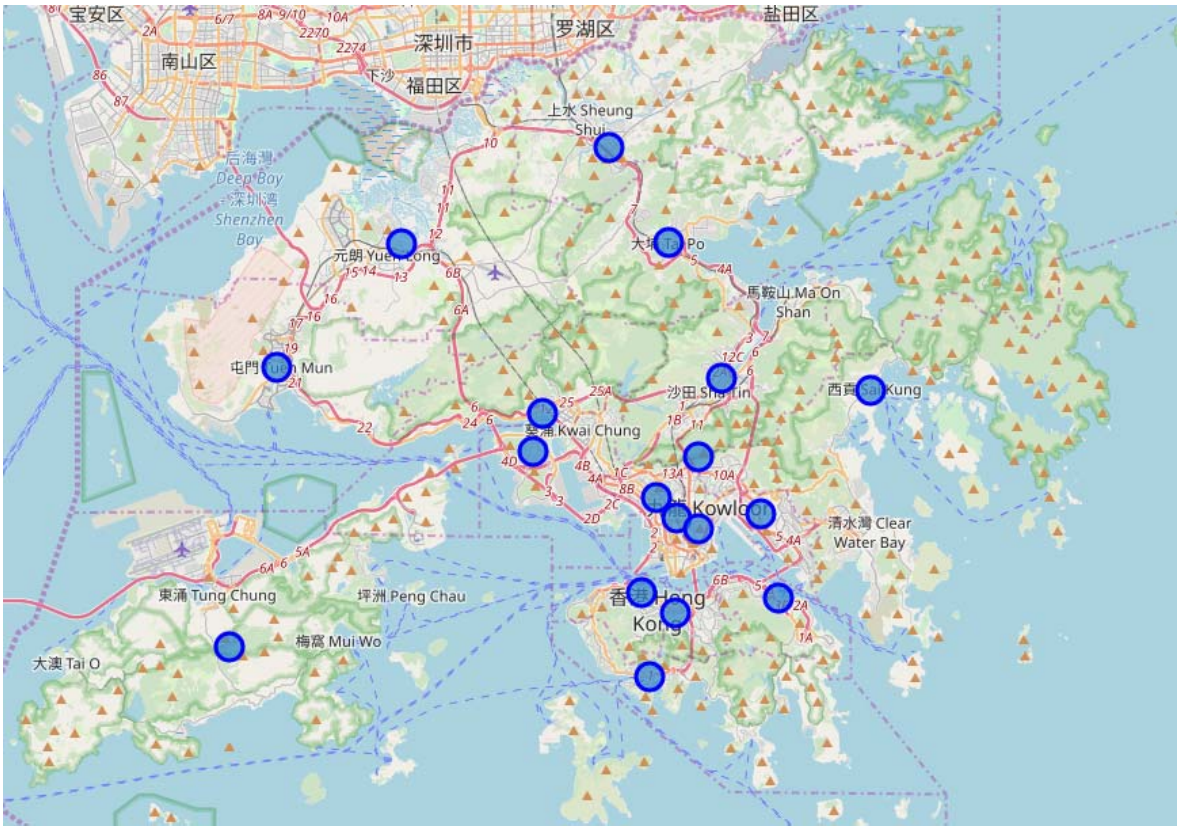
Hong Kong is so tiny that there is no borough. The whole city is divided into 18 cities, with individual area ranging from 9 km^2 to 175 km^2. Average footfall by district is a useful feature for my modeling. Given this data is unavailable, I took population density (population divided by area) as a proxy of footfall. 2017 population data is extracted from the Census and Statistics Department, whereas district areas are sourced from Wikipedia. Average rental by district is good to have but also not available. I did not attempt to take samples from the Internet to estimate average rentals because rentals vary tremendously within the same district. Random sampling may create bias.

I tried using Nominatim API to get the coordinates of these 18 cities. However, based on my knowledge of the city, I found that the output is not very accurate. Thus, I searched their coordinates through the Internet.

Coordinates generated by Nominatim API (inaccurate)



Coordinates obtained from Internet (more accurate)



My data is quite clean, thus no need for data wrangling. The full list of districts for illustration:

	District	Latitude	Longitude	Female	Male	Total Population	0-14	15-24	45-64	25-44	65+	Median age	Area (km2)	radius	Density
0	Central and Western	22.286660	114.15497	133900	107700	241600	23918	25368	76104	77554	38656	43	12.44	1.990423	19421.221865
1	Wan Chai	22.277020	114.17232	101100	78300	179400	17761	15608	57408	59561	29242	44	9.83	1.769343	18250.254323
2	Eastern	22.284110	114.22414	301600	244800	546400	63382	51908	167198	173755	90156	43	18.56	2.431219	29439.655172
3	Southern	22.247250	114.15884	145400	118500	263900	30612	25598	82073	84184	41432	43	38.85	3.517472	6792.792793
4	Yau Tsim Mong	22.321380	114.17260	179800	153800	333600	37363	36362	102749	108086	49373	42	6.99	1.492017	47725.321888
5	Sham Shui Po	22.330983	114.16224	215600	185000	400600	49274	40461	122984	125788	61692	42	9.35	1.725603	42844.919786
6	Kowloon City	22.316670	114.18333	225800	186100	411900	47780	41602	127277	132220	62609	43	10.02	1.786360	41107.784431
7	Wong Tai Sin	22.350000	114.18333	224800	195800	420600	44584	45004	119871	137116	74026	45	9.30	1.720983	45225.806452
8	Kwun Tong	22.323300	114.21540	355300	308800	664100	79028	67738	195245	208527	113561	43	11.27	1.894511	58926.353150
9	Kwai Tsing	22.352880	114.10004	270400	237200	507600	58882	53806	152280	159894	82739	43	23.34	2.726375	21748.071979
10	Tsuen Wan	22.370660	114.10479	169000	144600	313600	36064	32301	97216	102234	46099	43	61.71	4.433155	5081.834387
11	Tuen Mun	22.392110	113.97011	256200	224300	480500	54777	50933	142709	160007	71595	43	82.89	5.137907	5796.839184
12	Yuen Long	22.450000	114.03333	334800	290200	625000	75625	68750	196875	188750	94375	41	138.46	6.640447	4513.939044
13	North	22.494710	114.13812	165900	146900	312800	39413	32531	94778	98219	47858	42	136.61	6.595936	2289.729888
14	Tai Po	22.450070	114.16877	164200	139400	303600	35218	30360	92902	100492	44326	43	136.15	6.584821	2229.893500
15	Sha Tin	22.387150	114.19534	367800	309800	677600	78602	71148	196504	220898	110449	44	68.71	4.677838	9861.737738
16	Sai Kung	22.381430	114.27052	252600	211200	463800	52873	47771	143778	147488	71889	42	129.65	6.425715	3577.323563
17	Islands	22.261140	113.94608	88600	71700	160300	18114	17312	53220	47449	24045	41	175.12	7.467978	915.372316

In the end, I used Foursquare API to get the most common venues of each of the 18 districts. I found that maximum number of venues returned by Foursquare API is limited

to 100 only no matter how I altered the “LIMIT” parameter. When radius is 500, only 563 venues are returned. In Yuen Long and Tuen Mun District, there are only a handful of venues returned. These figures are too low for meaningful modeling. Therefore, I used the district area to derive a radius for each district. These radii are then plugged into the API request URL. As a result, the Foursquare API returns 1,616 venues, falling into 215 unique categories.

The following shows a statistics of the venues returned by district:

```
District
Central and Western    100
Eastern                 99
Islands                100
Kowloon City           100
Kwai Tsing              78
Kwun Tong              100
North                  100
Sai Kung               100
Sha Tin                100
Sham Shui Po           100
Southern               100
Tai Po                  64
Tsuen Wan               69
Tuen Mun                60
Wan Chai               100
Wong Tai Sin            46
Yau Tsim Mong          100
Yuen Long              100
Name: Venue, dtype: int64
```

The first 5 rows of the venue list are as follows:

	District	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central and Western	22.286660	114.15497	Four Seasons Hotel Hong Kong (香港四季酒店)	22.286554	114.156929	Hotel
1	Central and Western	22.286660	114.15497	The Spa at Four Seasons	22.286279	114.157623	Spa
2	Central and Western	22.286660	114.15497	VEA Restaurant and Bar	22.284890	114.152953	French Restaurant
3	Central and Western	22.286660	114.15497	Coco Espresso	22.285117	114.152466	Coffee Shop
4	Central and Western	22.286660	114.15497	HK Brewcraft	22.283352	114.154738	Beer Store

The above venue list is not good for modeling. I transformed it by the one hot encoding method into a DataFrame with frequency of each venue category in individual column. The following tables illustrate the top 10 most common venue categories in first three districts.

```
----Central and Western----
                  venue  freq
0           Coffee Shop  0.06
1                   Bar  0.04
2                 Hotel  0.04
3  Gym / Fitness Center  0.04
4          Yoga Studio  0.04
5   Japanese Restaurant  0.04
6    Italian Restaurant  0.03
7     French Restaurant  0.03
8        Ice Cream Shop  0.03
9                 Café  0.03
```

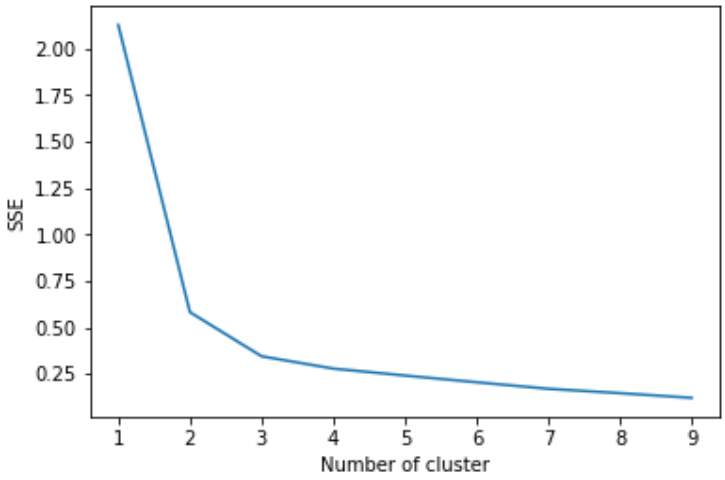
```
----Eastern----
                  venue  freq
0                  Park  0.06
1    Chinese Restaurant  0.05
2      Department Store  0.04
3    Seafood Restaurant  0.04
4   Japanese Restaurant  0.04
5  Fast Food Restaurant  0.03
6              Mountain  0.03
7           Coffee Shop  0.03
8          Noodle House  0.03
9         Shopping Mall  0.03
```

```
----Islands----
                  venue  freq
0       Airport Lounge  0.09
1                Beach  0.06
2          Coffee Shop  0.05
3               Bakery  0.05
4   Seafood Restaurant  0.03
5            Gift Shop  0.03
6                Hotel  0.03
7         Dessert Shop  0.02
8   Chinese Restaurant  0.02
9  Japanese Restaurant  0.02
```

3. Methodology

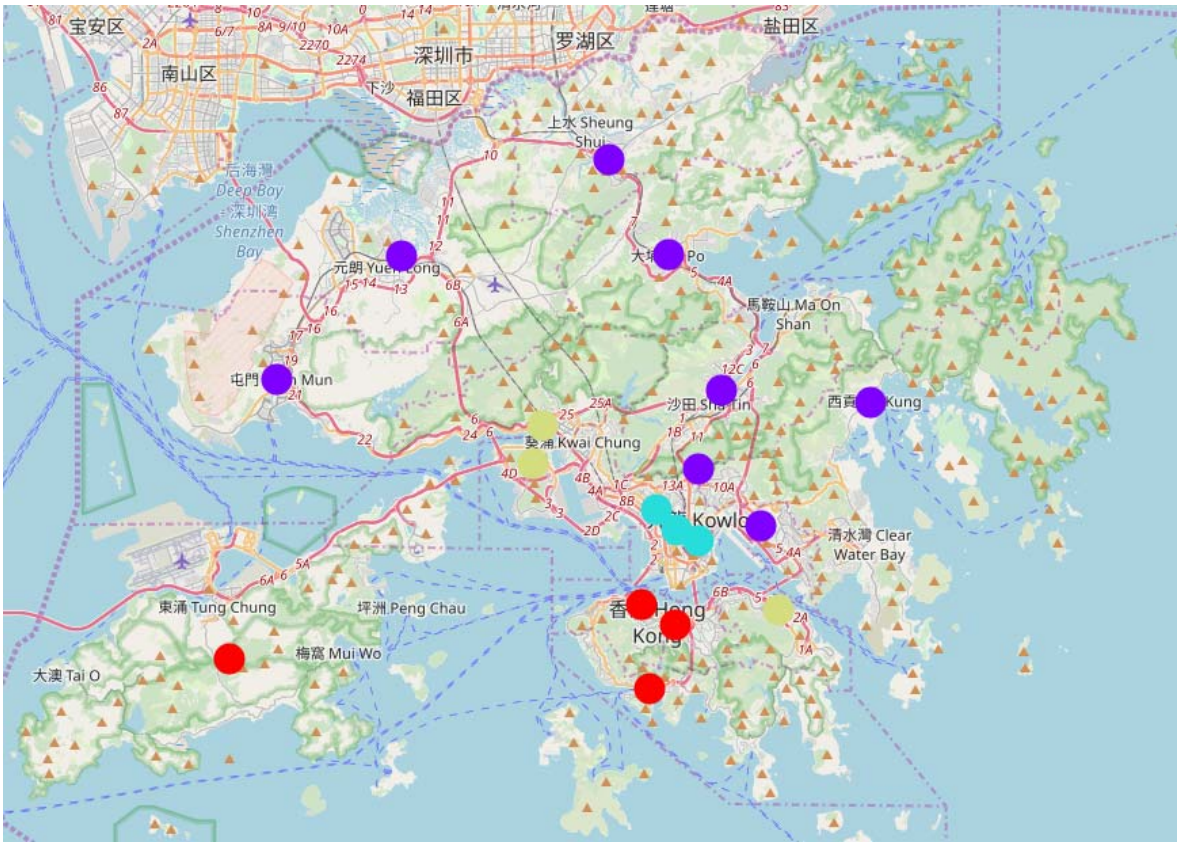
To identify districts having similar surrounding environment as Yau Tsim Mong, I use K-means model to segment the 18 districts into clusters, and then select districts that are in the same cluster as Yau Tsim Mong. The occurrence frequency of venue categories and population density are used as feature input to run the K-means model. The population density is normalized for modeling.

To determine the optimal number of clusters, I run the K-means model within a for loop from a count of 1 to 10. Within-cluster sum of squared errors, or SSE, is used to evaluate the performance of K-means. SSE declines as the count increases but stabilizes when number of clusters reaches 7. The lower the SEE, the higher the performance. When there are too few clusters, the results are not too useful. Hence, I decided to adopt 4 clusters for my K-means model.



4. Results

A label list with value ranging from 0 to 3 is returned by the K-means model. Each label represents the cluster a district is assigned to. This label list is then appended to the DataFrame having 18 districts as row, and coordinates of these districts, their population density and occurrence frequency of their venue categories as column. The distribution and clustering of the districts are then visualized by using the python folium library. Each cluster is indicated by one color.



There are 4 clusters in total. In each cluster, all districts have similar frequencies and types of venue categories, as well as population density. This implies that these districts have similar surrounding environment and footfall. The biggest cluster has 8 districts whilst the smallest one has 3.

By examining each cluster, we can determine the discriminating venue categories and density that distinguish each cluster:

Cluster 0	Cluster 1	Cluster 2	Cluster 3
Districts close to airport, or where tourists like to stay and visit	Districts with a variety of restaurants and coffee shops	High density districts with a lot of Asian cuisines, coffee shops and shopping malls	Districts with many parks, Asian cuisines and close to mountains

Cluster 0 (Red)

	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	Density
0	Central and Western	Coffee Shop	Bar	Gym / Fitness Center	Hotel	Yoga Studio	Japanese Restaurant	Café	Cantonese Restaurant	Italian Restaurant	Indian Restaurant	19421.221865
1	Wan Chai	Hotel	Café	Italian Restaurant	Hotel Bar	Bar	Coffee Shop	Steakhouse	Chinese Restaurant	Szechuan Restaurant	Cantonese Restaurant	18250.254323
3	Southern	Theme Park Ride / Attraction	Hotel	Café	Fast Food Restaurant	Coffee Shop	Aquarium	Furniture / Home Store	Clothing Store	Shopping Mall	Scenic Lookout	6792.792793
17	Islands	Airport Lounge	Coffee Shop	Bakery	Beach	Trail	Hotel	Gift Shop	Mountain	Seafood Restaurant	Market	915.372316

Cluster 1 (Purple)

	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	Density
7	Wong Tai Sin	Café	Fast Food Restaurant	Chinese Restaurant	Dessert Shop	Trail	Park	Coffee Shop	Japanese Restaurant	Asian Restaurant	Shopping Mall	45225.806452
8	Kwun Tong	Chinese Restaurant	Coffee Shop	Shopping Mall	Café	Fast Food Restaurant	Hong Kong Restaurant	Multiplex	Sushi Restaurant	Park	Department Store	58926.353150
11	Tuen Mun	Seafood Restaurant	Shopping Mall	Fast Food Restaurant	Chinese Restaurant	Coffee Shop	Hong Kong Restaurant	Cantonese Restaurant	Train Station	Italian Restaurant	Park	5796.839184
12	Yuen Long	Chinese Restaurant	Fast Food Restaurant	Dessert Shop	Coffee Shop	Noodle House	Shopping Mall	Train Station	Seafood Restaurant	Pizza Place	Café	4513.939044
13	North	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Hotel	Shopping Mall	Café	Noodle House	Dessert Shop	Cha Chaan Teng	Burger Joint	2289.729888
14	Tai Po	Chinese Restaurant	Coffee Shop	Fast Food Restaurant	Noodle House	Café	Restaurant	Bakery	Trail	Plaza	Hong Kong Restaurant	2229.893500
15	Sha Tin	Chinese Restaurant	Café	Fast Food Restaurant	Coffee Shop	Train Station	Cantonese Restaurant	Shopping Mall	Park	Hotel	Dessert Shop	9861.737738
16	Sai Kung	Café	Fast Food Restaurant	Seafood Restaurant	Coffee Shop	Pub	Thai Restaurant	Campground	Chinese Restaurant	Train Station	Beach	3577.323563

Cluster 2 (Blue)

	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	Density
4	Yau Tsim Mong	Coffee Shop	Chinese Restaurant	Hotel	Dim Sum Restaurant	Shopping Mall	Snack Place	Malay Restaurant	Noodle House	Dumpling Restaurant	Cha Chaan Teng	47725.321888
5	Sham Shui Po	Noodle House	Chinese Restaurant	Dessert Shop	Dim Sum Restaurant	Coffee Shop	Dumpling Restaurant	Cha Chaan Teng	Shopping Mall	Japanese Restaurant	Cantonese Restaurant	42844.919786
6	Kowloon City	Dessert Shop	Chinese Restaurant	Hotel	Coffee Shop	Shopping Mall	Noodle House	Toy / Game Store	Beer Bar	Sporting Goods Shop	Snack Place	41107.784431

Cluster 3 (Green)

	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	Density
2	Eastern	Park	Chinese Restaurant	Japanese Restaurant	Department Store	Seafood Restaurant	Mountain	Noodle House	Dim Sum Restaurant	Cantonese Restaurant	Coffee Shop	29439.655172
9	Kwai Tsing	Chinese Restaurant	Park	Shopping Mall	Sushi Restaurant	Japanese Restaurant	Cha Chaan Teng	Market	Coffee Shop	Noodle House	Dessert Shop	21748.071979
10	Tsuen Wan	Chinese Restaurant	Coffee Shop	Park	Hotel	Japanese Restaurant	Dessert Shop	Noodle House	Cantonese Restaurant	Mountain	Cha Chaan Teng	5081.834387

5. Recommendation

Yau Tsim Mong District, where Mr A's existing dessert shop is located, is inside Cluster 2. In this cluster, there are Sham Shui Po District and Kowloon City District. Mr A is recommended to open dessert shops in either one of these two districts or both. Sham Shui Po District is preferred because Kowloon City already has a lot of competitors. Alternatively, he can grow his chain of shops in the existing district (Yau Tsim Mong District) given that the competition is less intense there. Rental is not considered in my model

because of data limitation and the fact that rentals can vary significantly within the same district. Mr A is recommended to firstly choose district(s) based on the modelling results, and then decide the final options by considering other factors like rentals.

6. Conclusion

In this study, I use clustering to help a dessert shop owner make a business decision. I located the 18 districts in Hong Kong and extracted venues in these districts using the Foursquare API. These districts were segmented into 4 clusters by K-means model based on the occurrence frequency of the venue categories and the normalized population density. Districts in the same cluster have similar surrounding environment. Based on the results, I recommend Mr A to open new dessert shops in districts in the same cluster as his existing dessert shop. Rental is not a feature in my model, but should be an important factor after deciding the district(s).

7. References

I. **Official Statistics**
https://www.censtatd.gov.hk/hkstat/sub/sp150_tc.jsp?productCode=FA100096

II. **Area by District**
https://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF%E8%A1%8C%E6%94%BF%E5%8D%80%E5%8A%83