

Where in Hong Kong? The Restaurateur's Problem



Introduction

Hong Kong, a super city of 7.5 million inhabitants. Despite having a total area of 2,755 km², Hong Kong's living area only amounts to 1,108 km², with the remainder of the territory being deemed as uninhabitable due to its mountainous geography. As a result, Hong Kong is one of the most densely populated places in the world.

For anyone who has lived or even visited Hong Kong, this becomes very apparent. Narrow roads overflowing with towering skyscrapers and eateries, it comes as no surprise that there are over 15,000 restaurants in Hong Kong catering to all cuisines and budget types. However, as it goes, not all neighbourhoods are created equal, and Hong Kong is testament to that. Whilst some neighbourhoods may fare an abundance in the choice of available eateries, others not so much; and this is just one of the few things a restaurateur must take into consideration when planning to open a new restaurant.

The other major factor is undoubtedly rent. Hong Kong is known to have some of the highest commercial and residential rental rates in the world. According to [NavigateHK](#), the average residential rental in Chung Hom Kok is HK\$ 95,000 per month, whereas the average monthly rental rate in Tsuen Wan is one of the cheapest in Hong Kong at HK\$ 5,500 per month.

Given the above, we can quickly see how much of a location game it is for restaurateurs and their restaurants. A restaurant in the neighbourhood such as The Peak may attract some of Hong Kong's wealthiest; however, you'd have to ensure there is enough business to sustain the high rent prices. Conversely, a restaurant in Sham Shui Po may have far lower operational costs; however, as one of the most restaurant heavy neighbourhoods in Hong Kong, competition would undoubtedly be tough.

This investigation seeks to explore how K-Means clustering can be applied to determine the best neighbourhood in Hong Kong to open a given type of restaurant. We will primarily be focusing on using geospatial data through the FourSquare API in Python and make use of available socio-economic data to help construct an informed judgement.

The Data

As mentioned, we require two types of data for this investigation:

Geospatial

- A list of neighbourhoods in Hong Kong
- Coordinates of those neighbourhoods
- Restaurants and restaurant types in the neighbourhoods of interest

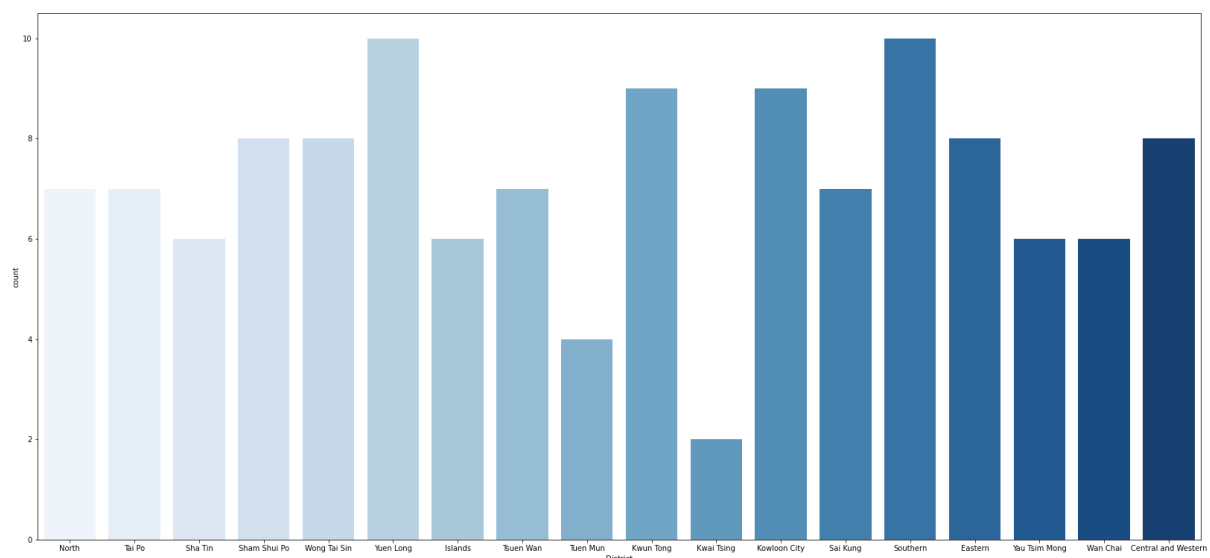
Socio-Economic Data

- Neighbourhood rent prices in Hong Kong

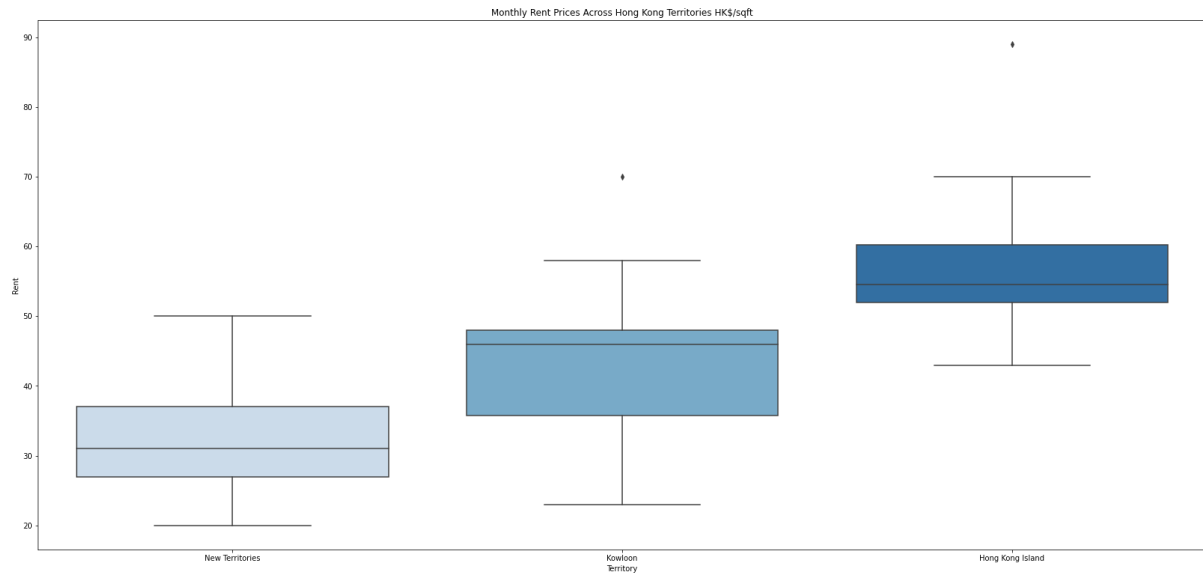
We can obtain data on restaurants and restaurant types within Hong Kong's neighbourhoods through the FourSquare API. However, in order to define the neighbourhoods and their respective geospatial indicators, we will need to obtain this data through the internet and either parse this information or manually create a working dataset. Luckily, data on Hong Kong's neighbourhoods are readily available on the internet and have been put to use in other studies. A simple Google search returns us a [dataset](#) which includes 'Neighbourhoods', 'Districts', 'Longitude', 'Latitude' and 'Rent per square foot'.

We are then left with a dataset consisting of 128 neighbourhoods across 18 districts:

	Territory	District	Neighbourhood	Latitude	Longitude	Rent (HK\$ per sqft)
0	Hong Kong Island	Central and Western	Kennedy Town	22.282435	114.128417	63
1	Hong Kong Island	Central and Western	Shek Tong Tsui	22.287735	114.134599	62
126	New Territories	Islands	Mui Wo	22.264545	114.001306	27
127	New Territories	Islands	Lamma Island	22.209441	114.149767	26



Whilst the obtained dataset is fit for purpose for our intended investigation, we could do with improving the data for visualisation purposes. In order to demonstrate the varying degree of rent prices across Hong Kong, it would be better to represent this as a boxplot covering all three primary territories of Hong Kong as opposed to all 18 districts. Labelling the neighbourhoods/districts under their respective territory, we find that that rent prices are generally higher in Hong Kong Islands, and the lowest in the New Territories.



Methodology

Now that we have our working dataset, we can look to use the geospatial elements of our data to extract information regarding 'venues' within the vicinity of Hong Kong's neighbourhoods. Our arguments include setting the radius to 500m (which we would consider a reasonable argument given the approximate sizes of Hong Kong's neighbourhoods), and limiting the number of venues returned to 300 per neighbourhood, as we expect the vast majority of restaurants, if not all, to be captured within the first 300 results returned. The FourSquare API request returned 3063 venues across 267 unique categories, with categories ranging from hotels to waterfalls.

As our investigation primarily focuses on the catering industry of Hong Kong, we are only interested in restaurants. Grouping all restaurant categories together, we find the FourSquare API has returned a total of **1356** restaurants. For the purpose of this investigation, we are going to investigate Asian restaurants within Hong Kong and aim to determine the best neighbourhoods to open a new Pan-Asian restaurant, taking total restaurants, proportion of Asian restaurants and rent prices per square foot within Hong Kong's neighbourhoods into consideration. In order to achieve this, we will also have to group together Asian restaurants. Focusing purely on Chinese, Taiwanese, Japanese and Korean cuisines to define 'Asian restaurant', we find the number of Asian restaurants returned by the FourSquare API to total **847**. We can then calculate the proportion of Asian restaurants given the total number of restaurants for each neighbourhood.

Subsequently, are left with a final dataset consisting of all our variables of interest:

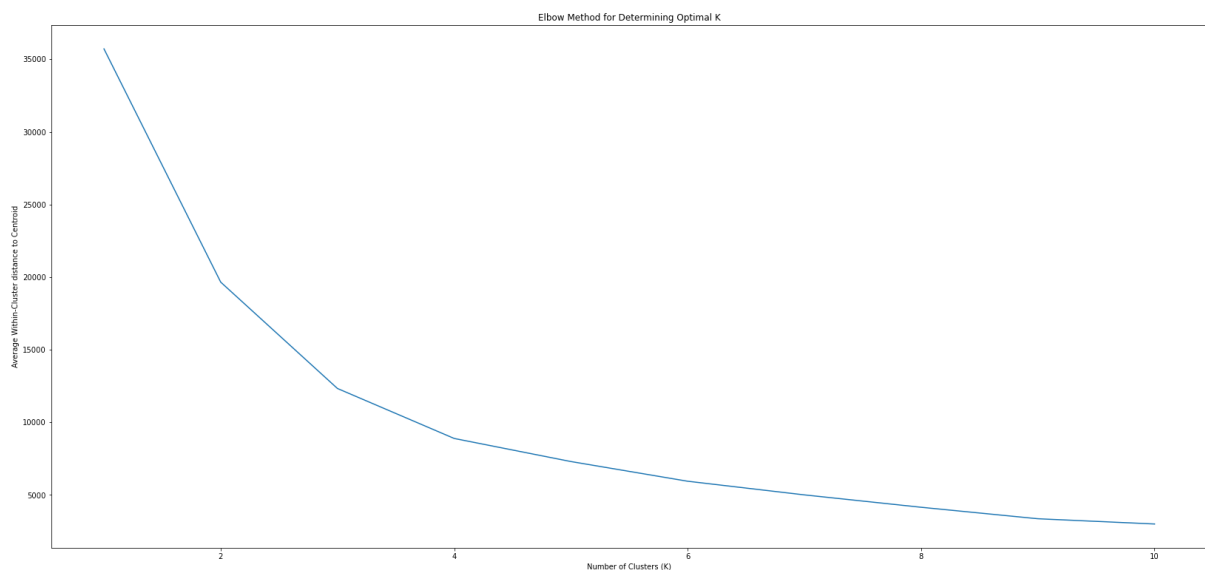
	Territory	District	Neighbourhood	Latitude	Longitude	Rent	Total Restaurants	Asian Restaurants	%Asian
0	New Territories	North	Luk Keng	22.518626	114.215191	20.0	1	1	1.00
1	New Territories	Tai Po	Kei Ling Ha	22.297180	114.172205	20.0	33	21	0.66
2	New Territories	North	Wu Kau Tang	22.506676	114.243548	20.0	0	0	0.00
125	Hong Kong Island	Central and Western	Peak	22.272003	114.152417	70.0	9	3	0.38
126	Kowloon	Yau Tsim Mong	Tsim Sha Tsui	22.297180	114.172205	70.0	32	21	0.66
127	Hong Kong Island	Central and Western	Admiralty	22.284242	114.158833	89.0	30	17	0.57

Now that we have our final dataset ready, we can compute our K-Means model.

K-Means Clustering

K-Means is an unsupervised clustering machine learning algorithm, which ultimately groups objects and observations that share similarities into clusters.

Using the elbow method to determine the optimal number of K clusters for our investigation, we decide to compute the K-Means algorithm using 6 clusters (although an argument could be made for 4 clusters looking at the elbow plot):



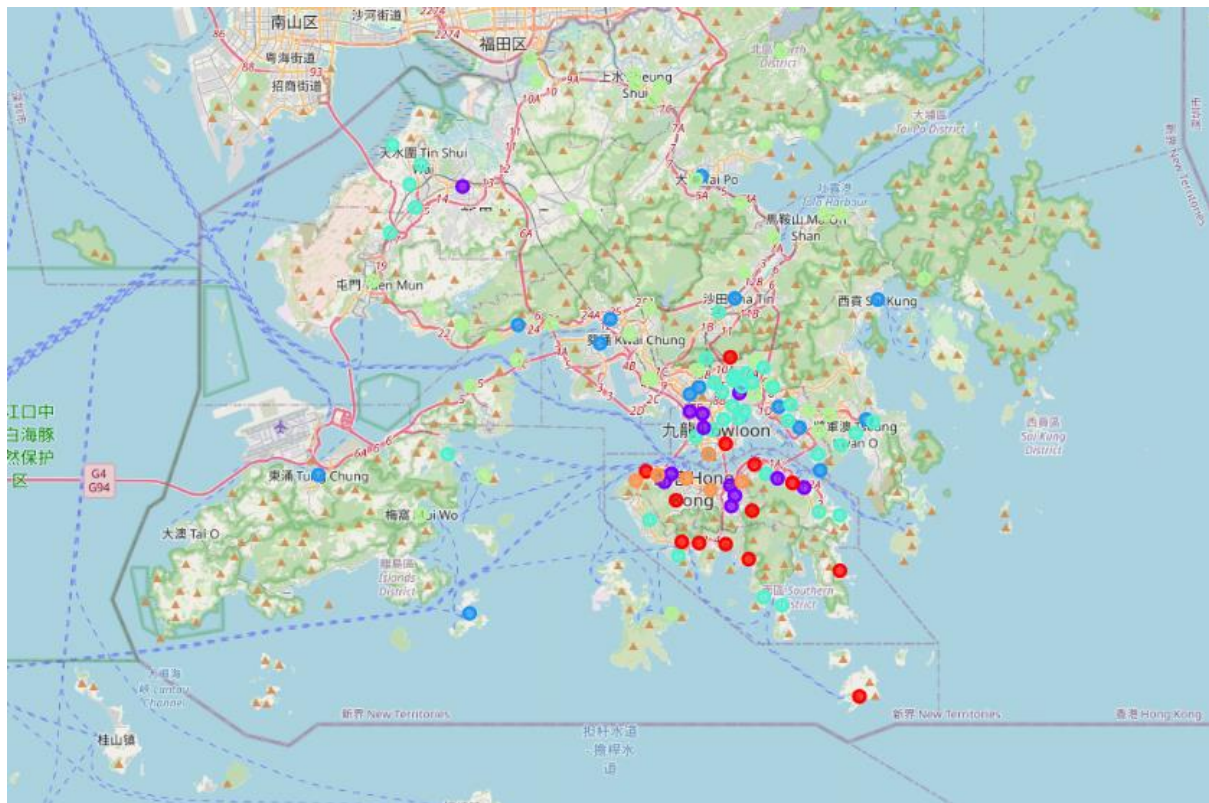
Results

Having used the K-Means algorithm to divide the data into 6 clusters, we find definitive features amongst each cluster. In summary, the clusters display the following properties:

City	Total Neighbourhoods	Total Restaurants	Total Asian Restaurants	Restaurants per Neighbourhood	%Asian	Average Rent (HK\$)
Hong Kong	128	1356	847	10.59	62.46%	41

KEY:		Below HK avg.	Above HK avg.			
Cluster	Total Neighbourhoods	Total Restaurants	Total Asian Restaurants	Restaurants per Neighbourhood	%Asian	Average Rent (HK\$)
1	12	306	197	25.50	64.38%	53.08
2	15	318	207	21.20	65.09%	33.20
3	36	276	176	7.67	63.77%	45.06
4	42	146	89	3.48	60.96%	28.74
5	6	228	134	38.00	58.77%	67.83
6	13	82	44	6.31	53.66%	58.85

Visually, we can superimpose our clusters over a map of Hong Kong using Python's Folium library:



Where

- Cluster 1 - Purple
- Cluster 2 - Blue
- Cluster 3 - Teal
- Cluster 4 - Green
- Cluster 5 - Orange
- Cluster 6 - Red

Discussion

Whilst there isn't much disparity in the proportion of Asian restaurants to the total number of restaurants amongst the clusters (-8.8% to +2.63%), we can turn our attention to rent prices and the total number of restaurants within a neighbourhood for a better indicator of where to open a new Pan-Asian restaurant. Cluster 4 may seem like an appealing option at first, but when we get to understand the dynamics of Hong Kong, we realise it may not be an attractive option after all.

As mentioned previously, Hong Kong is home to over 15,000 restaurants; however, the majority of those restaurants are densely situated within the Southern Kowloon and Northern Hong Kong Island regions. Those areas act as Hong Kong's financial, commercial, and tourist hubs, and therefore it is no surprise that the spread of restaurants. Taking this into consideration, we can understand why Cluster 4 has its limitations. It also comes as no surprise that the majority of neighbourhoods within Cluster 4 are located in the New Territories, where the population density of Hong Kong is also the lowest, when compared to Kowloon and Hong Kong Island.

Cluster 2 and Cluster 5 present interesting options. Cluster 2 has an average of 21.2 restaurants per neighbourhood, whilst the average rent within Cluster 2 is almost HK\$10 lower than the Hong Kong average. However, Cluster 2 seems to also contain the highest saturation of Asian restaurants. Cluster 5 on the other hand seems to house the major catering hubs of Hong Kong, with an average of 38 restaurants per neighbourhood. Of those restaurants, only 58.77% are accounted for by Asian restaurants; however, the average rent in Cluster 5 is the highest amongst all Clusters at HK\$67.83. A deeper dive into Cluster 5 shows the lowest rent within the cluster is Tin Hau (HK\$ 56.00 a month per square foot); however, 70% of Tin Hau's restaurants are accounted for by Asian restaurants. A deep dive into Cluster 2 can identify Sai Kung as a potential neighbourhood to open a new Pan-Asian restaurant. Rent in Sai Kung is very closely aligned to the Hong Kong average (HK\$42.00 a month per square foot), and the neighbourhood houses a reasonable number of restaurants (23) suggesting it to somewhat of catering and entertainment hub, of which only 3 are actually Asian.

Conclusion

Whilst we can draw some insights from our data and investigation, it would be unwise to draw a conclusive answer from the data we have at hand. The clusters themselves do well to depict the various neighbourhoods of Hong Kong when taking rent prices, total number of restaurants and the proportion of Asian restaurants into consideration; however, we will ultimately need to take a lot more factors into consideration if when it comes to this question. Yes, our current investigation can provide a good foundation, but we may also need to take into consideration the population of each neighbourhood, visitor appeal (Central London being a key example where the population may not be great, but restaurants can effectively sustain themselves off tourism), wealth of a neighbourhood amongst other factors to effectively come out on top in this location game.