

THE BATTLE OF NEIGHBORHOODS IN HONG-KONG – Week 1

Where to open a French Pastry Shop in Hong-Kong?



A. Introduction & Business Problem

A. 1. Background

Hong Kong is one of the most cosmopolitan cities in Asia and also one of the most densely populated territories in the world.

But beneath the canopy of this concrete jungle lies some of the city's most exciting bars, restaurants and modern-art galleries.

Offering every kind of culinary treats from around the world, from cheap eats and street food to Michelin-starred restaurants, Hong Kong has a well-deserved reputation as a culinary paradise.

A. 2. Business Problem

As a French expat living in this urban jungle, I love being able to treat myself to a nice little sweet treat at every corner, but I realize how much I was missing it in Hong Kong.

Even locals are quite keen on buying cakes. You even find cake shops in the MTR stations. The thing is, the sweet offered in Hong Kong isn't really exciting. Most of the cakes are just basic sponge cakes, very often industrially baked.

As for French pastries, it seems you can only find them in luxury hotels whereas in France, yummy pastries are to be found at every street corner.

I saw it as a potential opportunity for me to launch my own company to bring something new and fresh to the French pastry scene in Hong Kong.



Now, the big question is **where to set up this new French pastry shop?**

A. 3. Target audience

- Entrepreneurs or pastry chefs considering opening a French pastry, bakery or dessert shops in HK
- Expats looking for a career transition and to seize food-related business opportunities in the city
- Tourists interested in knowing where are located the different types of desserts shops in HK, as well as the most popular ones.

B. Data

B.1 Data Source

Main Data is from 3 sources:

1. **List of districts and neighborhoods in Hong Kong** from Wikipedia
(https://en.wikipedia.org/wiki/List_of_places_in_Hong_Kong)

2. **Coordinate's data** from <https://www.maps.ie/coordinates.html>.

3. **Foursquare location data** to explore and compare neighborhoods in HK, get the most popular venues by neighborhoods, and understand the pastry scene in HK (what are the different types of dessert shops in HK and where are they located? / what are the existing French pastry shops and where are they located?)

B.2 Data Cleaning

Hong Kong consists of Hong Kong Island, the Kowloon Peninsula, the New Territories, the Outlying Islands (Lantau Island, and over 200 other islands). ***This project will focus only on Hong Kong Island and Kowloon*** data.

In fact, the inhabited part of Hong Kong is mainly: Kowloon, which is more traditional, and Hong Kong Island, which is the glitzier side of town.

After scraping and cleaning the Wikipedia and coordinates data, I created a csv file '**neighborhoods hongkong.csv**' combining the neighborhood's data of interest, along with their latitude and longitude.



Note: for a more comprehensive study, we would have needed data on the rental price of commercial and retail spaces in HK by neighborhoods. Unfortunately, those data are not easily found online or are not available in English.

C. Methodology

To solve the problem I defined above:

- We will start by some exploratory data analysis with: Pandas library, Visualization technique like Folium and Foursquare API location data to explore neighborhoods and most popular venues in Hong-Kong, and visualize these information on a map
- Then, we will use machine learning method like the K-Mean algorithm to cluster the neighborhoods according to their similarities
- Finally, combining the results from all the above analysis, we will discuss and recommend a proper location to open our French Pastry Shop in last sections of this report.

C. 1. Data pre-processing: loading and reading data into a Pandas dataframe

In order to segment the neighborhoods in Hong-Kong and explore them, we will need the dataset that contains the districts and the neighborhoods that exist in each districts, as well as the the latitude and logitude coordinates of each neighborhood.

Let's load my master data 'neighborhoods hongkong.csv' file and read it in a Pandas dataframe *df_hk*.

df_hk has the following components:

	District	Neighborhood	Latitude	Longitude
0	Central & Western	Central District	22.281322	114.160258
1	Central & Western	Mid-Levels	22.282405	114.145809
2	Central & Western	The Peak	22.272003	114.152417
3	Central & Western	Sai Wan	22.285838	114.134023
4	Central & Western	Sheung Wan	22.286870	114.150267

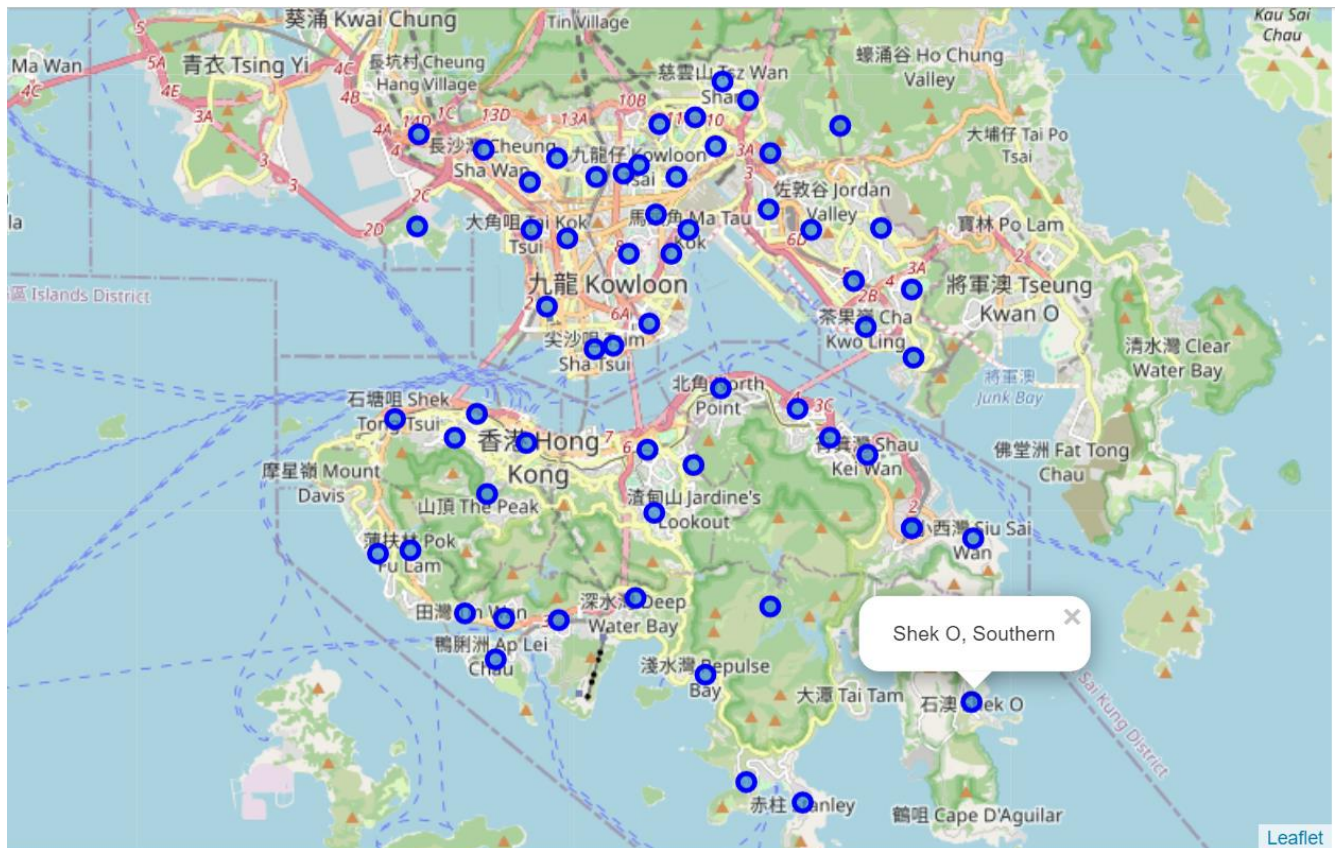
In total, the *df_hk* dataframe has 9 districts and 60 neighborhoods.

C. 2. Visualize the geographic data using Folium

We use:

- **Geopy** library to get the latitude and longitude values of Hong Kong
- **Python Folium** library to create a map of HK with neighborhoods superimposed on top. I used latitude and longitude values above to get the visual as below.

By clicking on the blue marker, you will see the names of the neighborhoods and districts.



C. 3. Exploring and Segmenting the Neighborhoods using Foursquare's API

I utilized the Foursquare API to explore the neighborhoods and segment them. I set the limit to **100 venues** within a radius of **500 meters** for each neighborhood from their given latitude and longitude information.

a. Exploring the first neighborhood in our dataframe

Let's get from Foursquare the top 100 venues that are in Central District within a radius of 500 meters.

	name	categories	lat	lng
0	Mandarin Grill + Bar (文華扒房 + 酒吧)	Steakhouse	22.281928	114.159408
1	Mandarin Oriental Hong Kong (香港文華東方酒店)	Hotel	22.281857	114.159382
2	Mott 32 (卅二公館)	Dim Sum Restaurant	22.280286	114.159080
3	8½ Otto e Mezzo Bombana	Italian Restaurant	22.281726	114.158767
4	The Mandarin Cake Shop	Bakery	22.281959	114.159416

77 venues were returned by Foursquare for Central District.

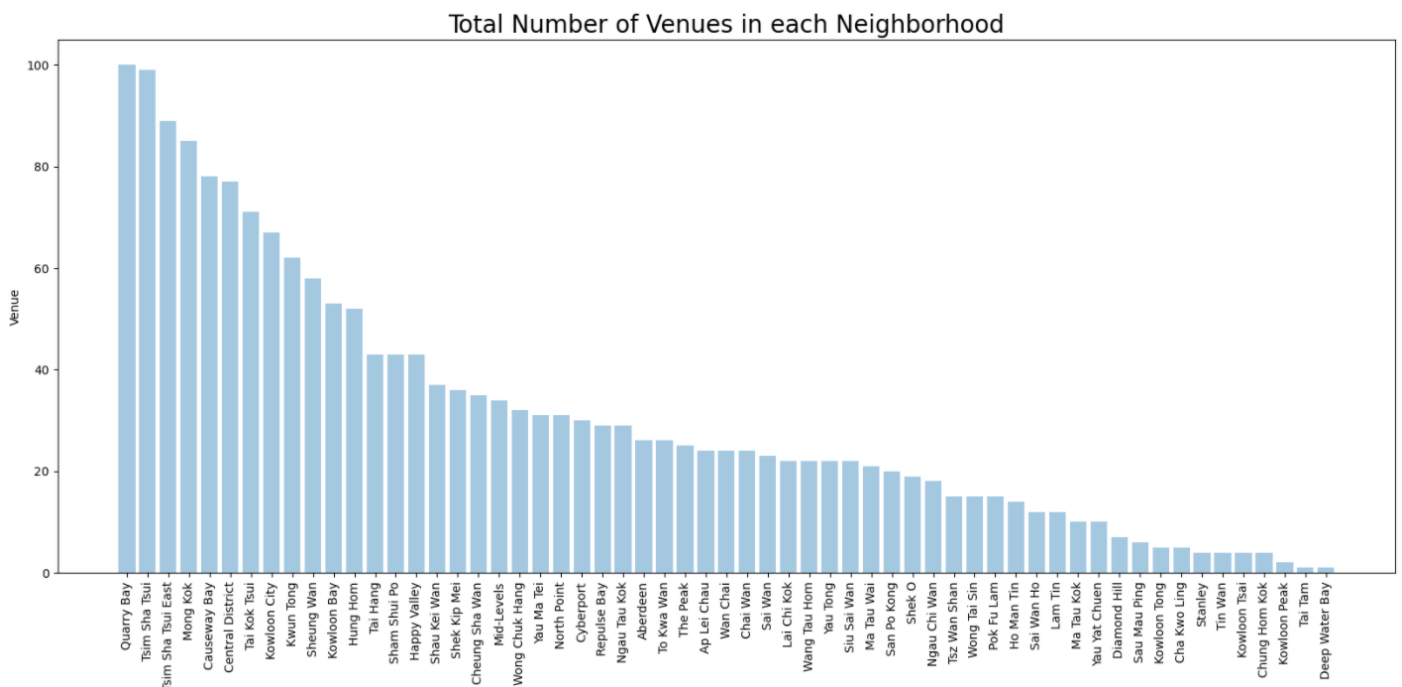
b. Creating a function to repeat the same process to all the neighborhoods in HK

1827 venues were returned by Foursquare for HK.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central District	22.281322	114.160258	Mandarin Grill + Bar (文華扒房 + 酒吧)	22.281928	114.159408	Steakhouse
1	Central District	22.281322	114.160258	Mandarin Oriental Hong Kong (香港文華東方酒店)	22.281857	114.159382	Hotel
2	Central District	22.281322	114.160258	Mott 32 (卅二公館)	22.280286	114.159080	Dim Sum Restaurant
3	Central District	22.281322	114.160258	8½ Otto e Mezzo Bombana	22.281726	114.158767	Italian Restaurant
4	Central District	22.281322	114.160258	The Mandarin Cake Shop	22.281959	114.159416	Bakery

Let's check how many venues were returned for each neighborhood in descending order and visualize the data in a bar chart.

	Neighborhood	Count
0	Quarry Bay	100
1	Tsim Sha Tsui	99
2	Tsim Sha Tsui East	89
3	Mong Kok	85
4	Causeway Bay	78



The 1827 venues fall into one of the 221 unique venues categories (Steakhouse, Hotel, Dim Sum Restaurant, Bakery, Dessert Shop, Pastry...)

c. Refining the analysis and explore the pastry scene

Now, let's explore and understand the pastry scene in HK: we slice the dataframe 'hk_venues' containing the most popular venues in HK and create a new dataframe 'df_pastry' containing only the venue falling into one of this Venue Category: bakery, pastry, dessert and cake shops (defined as the pastry scene).

The pastry scene is comprised of 65 shops.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Category
0	Central District	22.281322	114.160258	The Mandarin Cake Shop	22.281959	114.159416	Bakery
1	Central District	22.281322	114.160258	Urban Bakery Works	22.281240	114.158146	Bakery
2	Mid-Levels	22.282405	114.145809	Sweeties' Dessert 糖痴豆 (糖痴豆)	22.284321	114.142419	Dessert Shop
3	The Peak	22.272003	114.152417	Tai Cheong Bakery (泰昌餅家)	22.270565	114.149801	Bakery
4	The Peak	22.272003	114.152417	The Sweet Dynasty (糖朝)	22.270996	114.149834	Dessert Shop

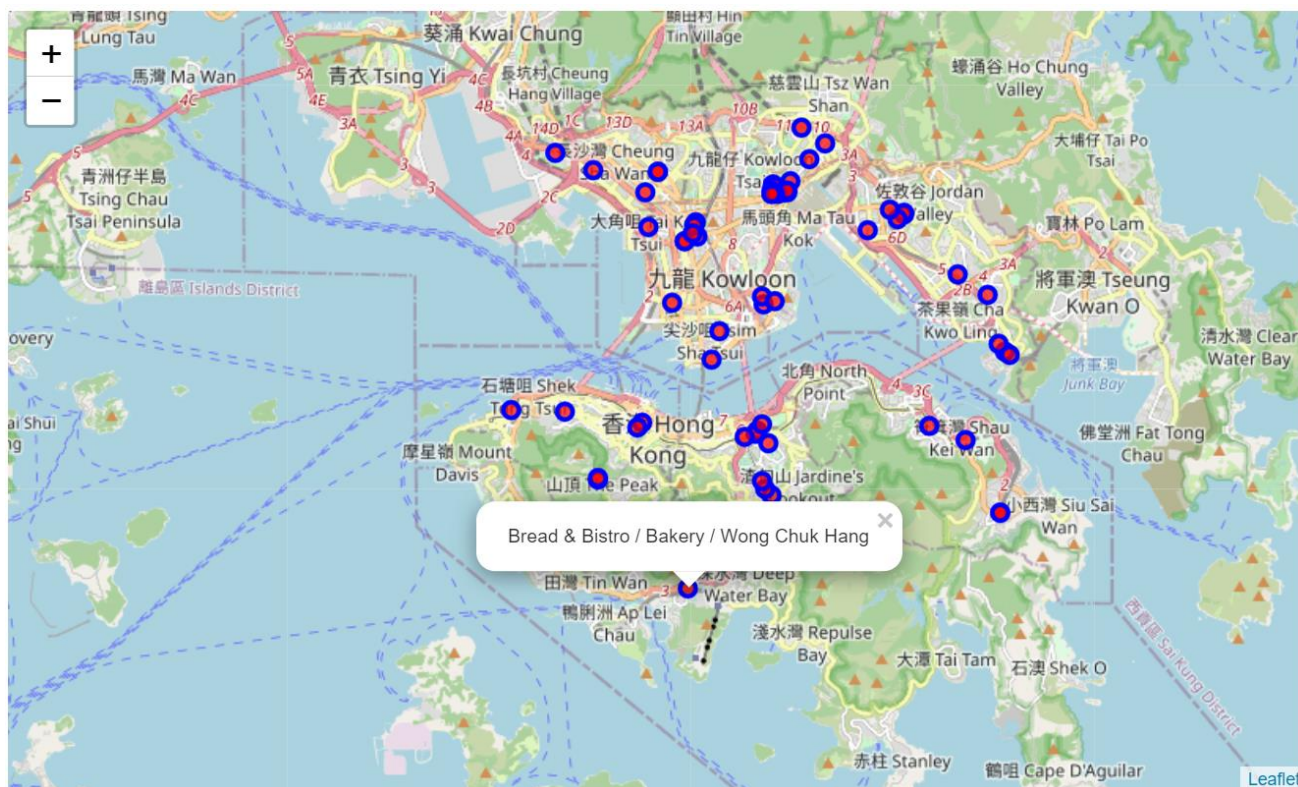
Let's create a new dataframe with only the French pastry shops:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Happy Valley	22.268464	114.186029	Maison Eric Kayser	22.270354	114.185429	Bakery
1	Kowloon City	22.331013	114.190366	Ceres Boulangerie et Patisserie	22.328188	114.187598	Bakery
2	Ma Tau Wai	22.324115	114.186379	Ceres Boulangerie et Patisserie	22.328188	114.187598	Bakery
3	Mong Kok	22.319685	114.168397	Passion by Gérard Dubois	22.318521	114.168517	Bakery
4	Tsim Sha Tsui	22.298872	114.174118	Maison Eric Kayser	22.296739	114.173747	Coffee Shop
5	Tsim Sha Tsui	22.298872	114.174118	Gontran Cherrier Artisan Boulanger Paris	22.294884	114.174504	Bakery

Note: 'Ceres Boulangerie et Patisserie' appears twice but there is only 1 shop (same venue latitude and longitude, but the shop have been referenced twice because of the neighborhood name: Ma Tau Wai is simply the Chinese name of the Kowloon City neighborhood)

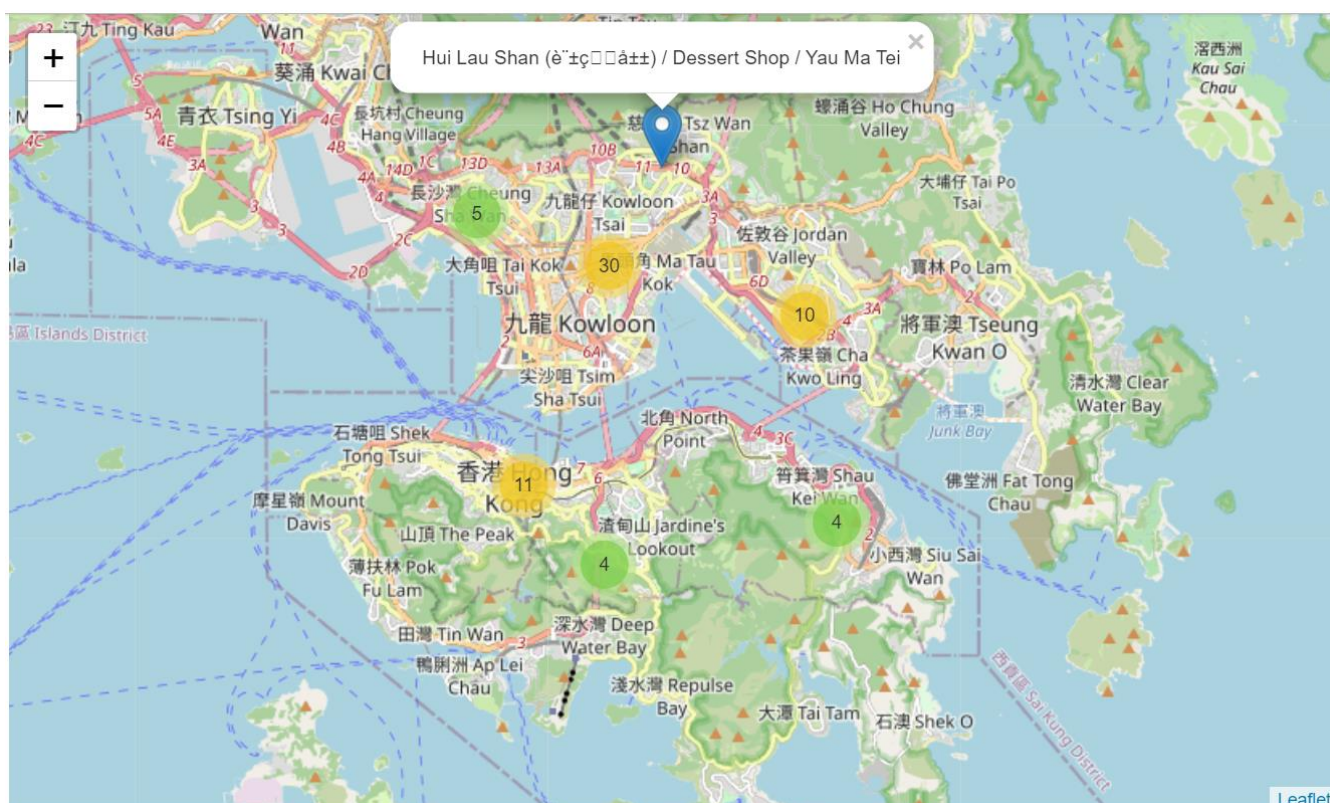
HK has 65 pastry shops, of which 5 are French. Now that we reduced the data a little bit, let's visualize where all these shops are located.

By clicking on the red marker, you will see the name of shops, its category and the neighborhood.



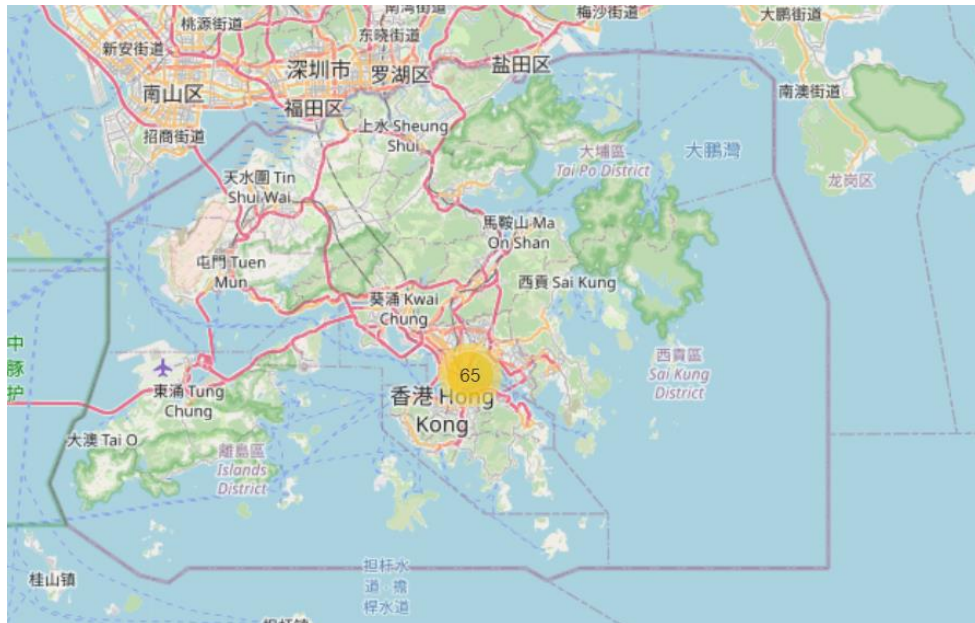
If we find the map to be so congested with all these circle markers, the remedy is to group the markers into different clusters. Each cluster is then represented by the number of shops in each neighborhood. These clusters can be thought of as pockets of HK which we can then analyze separately.

The 65 shops are now grouped in 7 clusters (6 clusters + 1).



Notice how when we zoom out all the way, all markers are grouped into one cluster 'the global cluster' of 65 markers, which is the total number of shops in our df_pastry dataframe.

Once you start zooming in, the 'global cluster' will start breaking up into smaller clusters (see above). Zooming in all the way will result in individual markers.



Now that we understand the big picture, it is useful to:

- **Distinguish the shops according to their category.**

We know from our df_pastry dataframe that the shops are categorized mainly in either 'Bakery' or 'Dessert Shop'.

Note: according to the Foursquare classification, our French pastry shop will fall more into the "Bakery" category than the "Dessert Shop" category. In fact, the former is comprised of existing French Bakery (see above) and the latter is mainly comprised of locals' dessert or bakery shops.

The way to do that in Folium is to create a 'feature group' with its own features and style and then add it to the map.

Bakery are marked in **Orange**, the **Dessert Shops** in **Green** and others category in Yellow.

- **See where the French pastry shops are located:** marked with the **blue** markers.

C. 4. Analyze Each Neighborhood

Previously, we extracted the total numbers of venues for each neighborhood. Now, let's get deeper into our analysis and understand the composition of each neighborhood. So that, we can cluster these neighborhood according to their similarities.

Top 10 venues for each neighborhood:

	Neighborhood	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	Aberdeen	Boat or Ferry	Fast Food Restaurant	Sushi Restaurant	Taiwanese Restaurant	Athletics & Sports	Park	Supermarket	Bus Station	Market	Shopping Mall
1	Ap Lei Chau	Fast Food Restaurant	Chinese Restaurant	Shopping Mall	Seafood Restaurant	Park	Pet Store	Bus Station	Mountain	Café	Furniture / Home Store
2	Causeway Bay	Sushi Restaurant	Chinese Restaurant	Coffee Shop	Japanese Restaurant	Clothing Store	Hotel	Dessert Shop	Noodle House	Pizza Place	Seafood Restaurant
3	Central District	Steakhouse	Social Club	Hotel	Gym / Fitness Center	Italian Restaurant	Japanese Restaurant	Lounge	Cocktail Bar	Shopping Mall	Café
4	Cha Kwo Ling	Hong Kong Restaurant	Convenience Store	Fast Food Restaurant	Shopping Mall	Soccer Field	Dessert Shop	Electronics Store	French Restaurant	Food Court	Flea Market
5	Chai Wan	Chinese Restaurant	Fast Food Restaurant	Coffee Shop	Bus Stop	Multiplex	Cha Chaan Teng	Market	Park	Athletics & Sports	Seafood Restaurant

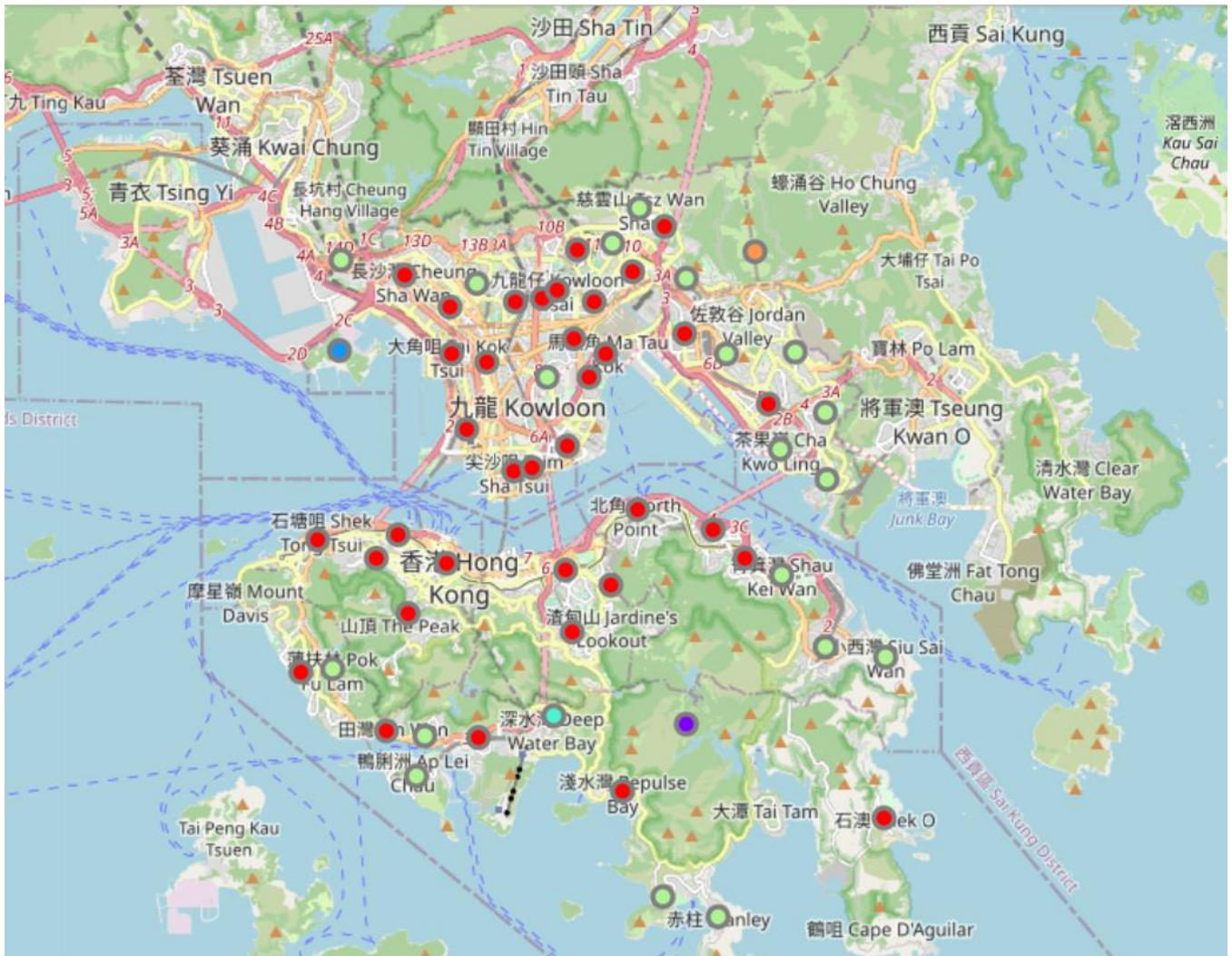
C. 5. Clustering using K-mean algorithm

As we can see from the above table, we have some common venue categories between neighborhoods. For this reason, I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning. I will run K-Means to cluster the neighborhoods into 6 clusters (labelled from n°0 to 5 by the algorithm in the below table, but from 1 to 6 in the below map and the rest of the report).

I choose 6 clusters given that we have 60 neighborhoods and that we want get the finest possible analysis of the neighborhoods of HK and understand what characterized them the most.

	Neighborhood	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	Aberdeen	Athletics & Sports	Ramen Restaurant	Park	Bubble Tea Shop	Fast Food Restaurant	Garden	Noodle House	Supermarket	Boat or Ferry	Sushi Restaurant
1	Ap Lei Chau	Furniture / Home Store	Fast Food Restaurant	Clothing Store	Chinese Restaurant	Shopping Mall	Coffee Shop	Garden	Café	Paper / Office Supplies Store	Park
2	Causeway Bay	Chinese Restaurant	Sushi Restaurant	Coffee Shop	Japanese Restaurant	Cha Chaan Teng	Hotel	Dessert Shop	Clothing Store	Noodle House	Gift Shop
3	Central District	Steakhouse	Coffee Shop	Italian Restaurant	Japanese Restaurant	Gym / Fitness Center	Lounge	French Restaurant	Social Club	Hotel	Sushi Restaurant
4	Cha Kwo Ling	Bus Station	Shopping Mall	Dumpling Restaurant	Hong Kong Restaurant	Fast Food Restaurant	Soccer Field	Zhejiang Restaurant	Farmers Market	Furniture / Home Store	Fruit & Vegetable Store
5	Chai Wan	Chinese Restaurant	Coffee Shop	Cha Chaan Teng	Tea Room	Shopping Mall	Grocery Store	Fast Food Restaurant	Bakery	Seafood Restaurant	Athletics & Sports
6	Cheung Sha Wan	Chinese Restaurant	Sandwich Place	Japanese Restaurant	Coffee Shop	Seafood Restaurant	Fast Food Restaurant	Cha Chaan Teng	English Restaurant	Restaurant	Shopping Mall
7	Chung Hom Kok	Hill	Park	Beach	Zhejiang Restaurant	Garden	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
8	Cyberport	Coffee Shop	Bus Stop	Japanese Restaurant	Cantonese Restaurant	Gym	Irish Pub	Buffet	Sports Club	Multiplex	Café
9	Deep Water Bay	Campground	Zhejiang Restaurant	Deli / Bodega	Gastropub	Garden	Furniture / Home Store	Fruit & Vegetable Store	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
10	Diamond Hill	Chinese Restaurant	Imported Food Shop	Basketball Court	Cha Chaan Teng	Burger Joint	Vietnamese Restaurant	Bus Station	Park	Food Court	Flea Market

Clusters map



Let's label the clusters according to their most common venues:

- Cluster 1 (red): Restaurants, Cafés and Coffee Shops
- Cluster 2 (purple): Park
- Cluster 3 (dark blue): Health Food Store
- Cluster 4 (turquoise): Campground
- Cluster 5 (light green): Chinese Restaurants and Fast-Food
- Cluster 6 (orange): Scenic Lookout

D. Results

Understanding the pastry scene using Foursquare API and visualize it using Folium

HK has 65 popular pastry shops in the broadest sense (Bakery|Pastry|Dessert|Cake), of which just 5 are French (mainly located on the Kowloon Peninsula):

- Maison Eric Kayser (Happy Valley, Hong-Kong Island)

- Maison Eric Kayser (Tsim Sha Tsui, Kowloon Peninsula)
- Passion by Gérard Dubois (Mongkok, Kowloon Peninsula)
- Ceres Boulangerie et Patisserie (Kowloon City = Ma Tau Wai, Kowloon Peninsula)
- Gontran Cherrier Artisan Boulanger Paris (Tsim Sha Tsui, Kowloon Peninsula)

K-Mean Clustering and most common venues:

Only the cluster 1 in red regrouping a numbers of Restaurants from diverse nationality (Japanese, Thai, French, Spanish, Chinese, Italian, Hong-Kong...), Cafés and Coffee Shops appears relevant to set up our new business. Note that this Cluster also represents the tourist area of HK, with a numbers of hotels, restaurants, supermarkets...

E. Discussion

From the above cluster map, we notice that the neighborhoods of our selected Cluster (n°1 in red) are distributed on both side of Hong-Kong: the Kowloon Peninsula at the top and Hong-Kong Island below.

Therefore, the question is now: once we have determined the seemingly ideal Cluster thanks to the different analysis we did (visualization of the geographical location and its neighborhoods, understanding the pastry scene, most common venues by neighborhoods and clustering), in which part of Hong-Kong to set up our business?

To answer this, as mentioned in the Data section, it would have been helpful to have at our disposal the rental price of commercial and retail spaces in HK by neighborhoods. Unfortunately, those data are not easily found online or are not available in English.

In fact, we could have superimposed the choropleth map of rental prices on top of the cluster map, which would have helped us to refine our study.

Notwithstanding the lack of these data, it is commonly known that rents are much higher in the Hong-Kong Island side, especially in the Central District, which might explain why most of the popular French pastry shops are rather located on the Kowloon Peninsula.

F. Conclusion

"Where is the proper location to open a French pastry shop?"

From all the above results of our analysis, in combination with some local knowledge of Hong-Kong:

- Obviously clusters 2, 3, 4 and 6 are excluded because these are mountain, park, beach, or port (as we can actually see on the map) or irrelevant (campground or health food store as most common venue).
- Cluster 5 represents mainly residential area where locals live, hence a majority of Chinese and Fast-Food restaurants.

- Cluster 1 is where commercial, office and touristic areas are concentrated. More specifically, the Hong-Kong Island part of the Cluster 1 regroups the neighborhoods where expats work and live, hence the significant numbers of cafés and international restaurants.

Therefore, Cluster 1 will be the ideal location given the types of venues and profile of the potential clients in this area. More specifically:

- The pastry shop will be on the Hong-Kong Island side if money is not a problem, and because expats and offices are concentrated on that part of HK. Besides, there is only 1 popular French pastry shop of that side (Maison Eric Kayser) vs. 4 on Kowloon.
- Otherwise, the Kowloon side still represents a good choice given that it is a very touristic area that concentrates the most numbers of venues by neighborhood and rents are cheaper.

Cluster 1 – The ideal neighborhoods to set up the French Pastry Shop

