

Hong Kong Food Clusters analysis

Law Kwan Ping

May 2019

Hong Kong Food Clusters analysis	1
1. Introduction	1
2. Data acquisition and cleaning	3
Adjustment	4
3. Methodology	4
4. Exploratory Data Analysis	4
5. Clustering	6
6. Price levels analysis	10
7. Conclusions	11
Reference:	11

1. Introduction

Modern Hong Kong has a predominantly service-based economy,^[2] and restaurant businesses serve as a main economic contributor. With the third-densest population per square meters in the world and serving a population of 7 million,^[3] Hong Kong is host to a restaurant industry with intense competition. Due to its small geographical size, Hong Kong contains a high number of restaurants per unit area. It showing us a very favorable factor to perform a restaurant's geographic analysis.

Searching for a dinner place is always a daily question in many people's mind. Some people may find their ideas from food guides or magazine recommendations. However, there is not much information about how restaurant spread over hong kong. It would be interesting to understand what food choices do we have in the areas, what kinds of restaurants are more likely comes together and how the different restaurant categories form the districts. Besides, price is also an important factor to be considered. There are full of smart consumers in the city, price and the service have to be matched. Finding out a reasonable price for each cuisine is also a piece of helpful information. This project will analyze the geographic distribution of restaurants from a data science perspective and we will analyze the cuisine's clusters and the price levels around hong kong. Hopefully, It would be a good figure for consumers to make their choices.

In the nutshell, this project helps people understand the hong kong restaurant divergence in cuisine styles and the price level versus their geographic location.

2. Data acquisition and cleaning

In the sense of geographic analysis, we need an exhaustive dataset about hong kong restaurants. Foursquare API provides us a comprehensive way to explore hong kong restaurants. We can fit our location (including latitude and longitude) to the API and get back a description of the restaurants near us. In considering the analysis coverage, we decide to target the center part of Hong Kong and neglect those suburban areas, because it would increase the significance of the analysis.

To start exploring hong kong urban area, it would be great to travel the city with MTR. We pick the some most popular stations which well cover the Hong Kong island and Kowloon area as the starting points.

Kwun Tong Line, Island Line, part of Tsuen Wan Line and some popular stations were included.



Assuming that we travel along with these stations and search all the restaurants near the MTR stations, then we summarize all the restaurant data and consolidate to a list. It would provide us a good amount of data to start our analysis.

Adjustment

However, some MTR stations are not recognizable to the Foursquare site or the place has more than one result having a similar name in the world. It would cause an unsatisfactory result. We change the station's name to nearest location name which is recognizable by the server.

Besides, we also interest in analyzing price level deviation, Foursquare also provides us the price tier data.

3. Methodology

After data acquisition and cleaning, we grouped the data points of each station and found out how many restaurant categories does each station have. The categories frequencies are not ready for training the model, we applied one hot encoding to convert categories frequencies to numbers of dummy columns which stores the frequency of each category.

In this reason we want to discover any cluster exist in the city, we used unsupervised learning K-means algorithm to cluster the restaurant. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

We used python folium library to visualize geographic details of Hong Kong restaurants. Latitude and longitude values are the keys to visualize the distribution.

4.Exploratory Data Analysis

First of all, we got some basic statistic from the data fetching from Foursquare database. There are total 1749 data points get back though querying the explore type API for 35 locations.

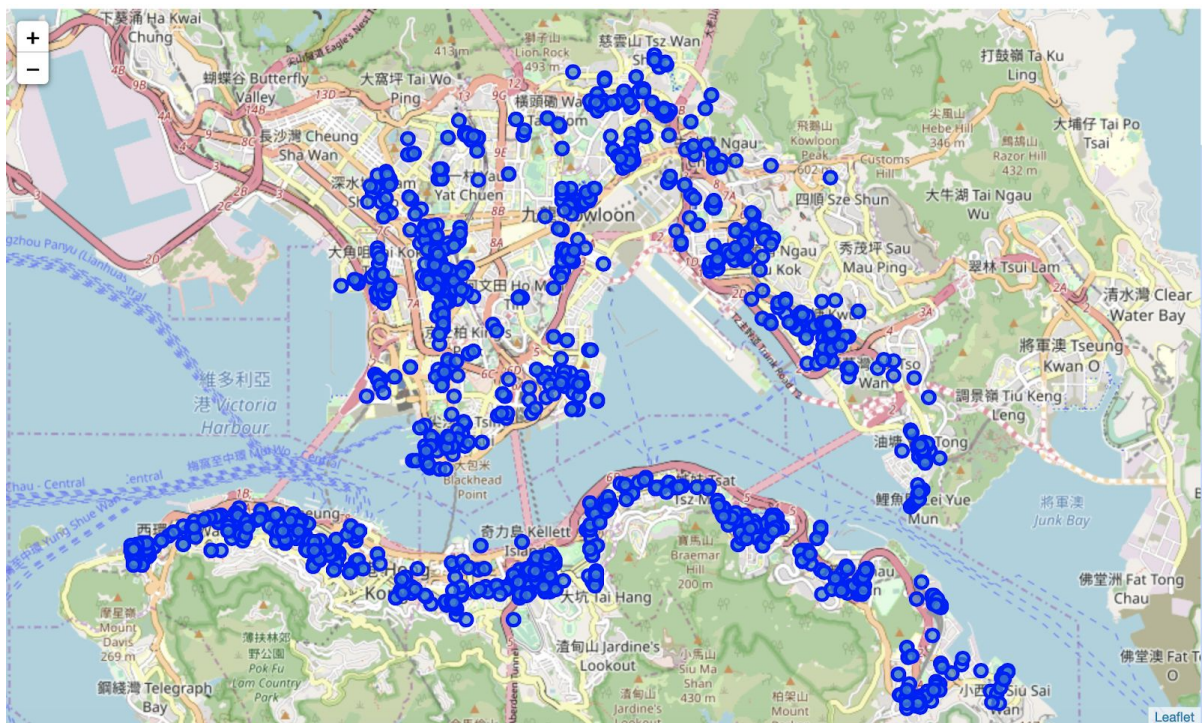
```
Out[11]:
```

	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.formattedAddress
0	Burger Joint	False	58e4d9d2d0bb3e769e4bd7c2	Shop FC01, G/F, TBG Mall, 3 Ngau Tau Kok Rd	HK	Kowloon Bay	香港	NaN	[Shop FC01, G/F, TBG Mall, 3 Ngau Tau Kok Rd, 香港]
1	Fast Food Restaurant	False	4bf89615508c0f471cad3e31	Unit 246-253, G/F, Amoy Plaza 2, 77 Ngau Tau K...	HK	牛头角	香港	NaN	[Unit 246-253, G/F, Amoy Plaza 2, 77 Ngau Tau ...]
2	Fast Food Restaurant	False	4ca7d7a1b0b8236a11bfade6	Shop P17, Telford Plaza, 33 Wai Yip St	HK	Kowloon Bay	香港	NaN	[Shop P17, Telford Plaza, 33 Wai Yip St, 香港]
3	Sandwich Place	False	542e0ccd498ec439e70a3d9e	港鐵彩虹站	HK	NaN	香港	NaN	[港鐵彩虹站, 香港]
4	Vegetarian / Vegan Restaurant	False	4d0f5554b765224b8716e432	Unit 242-245, G/F, Amoy Plaza 2	HK	牛头角	香港	77 Ngau Tau Kok Rd	[Unit 242-245, G/F, Amoy Plaza 2 (77 Ngau Tau ...]

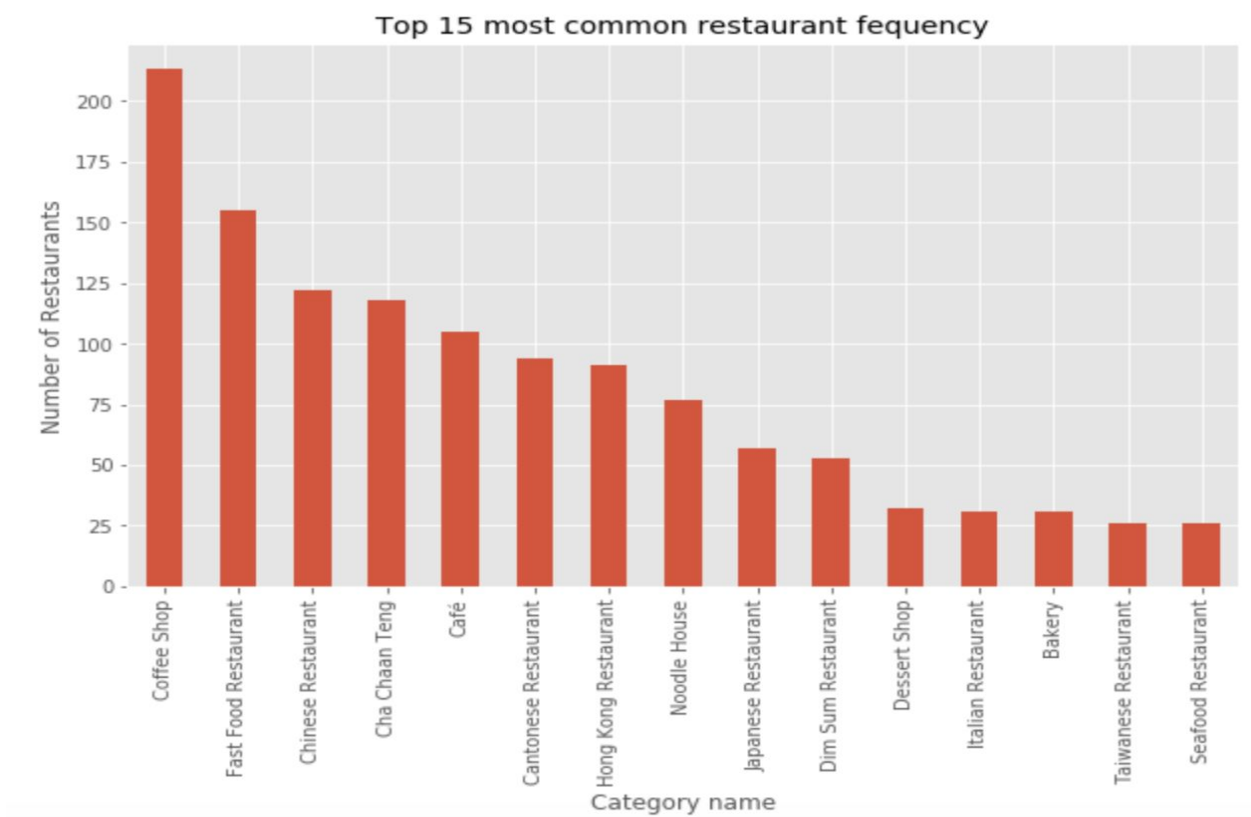
```
In [12]: hk_restaurant.shape
```

```
Out[12]: (1749, 19)
```

The below map visualizes the restaurant location retrived from Foursquare API.



In summary of categories, there are 91 unique categories were returned by Foursquare, then I created a bar chart which shows the distribution of the top 15 categories.



After the transformation of the location data, we can summarize the top 10 restaurant categories for each location and get a quick idea about them. The Top 10 common categories for each location are listing below.

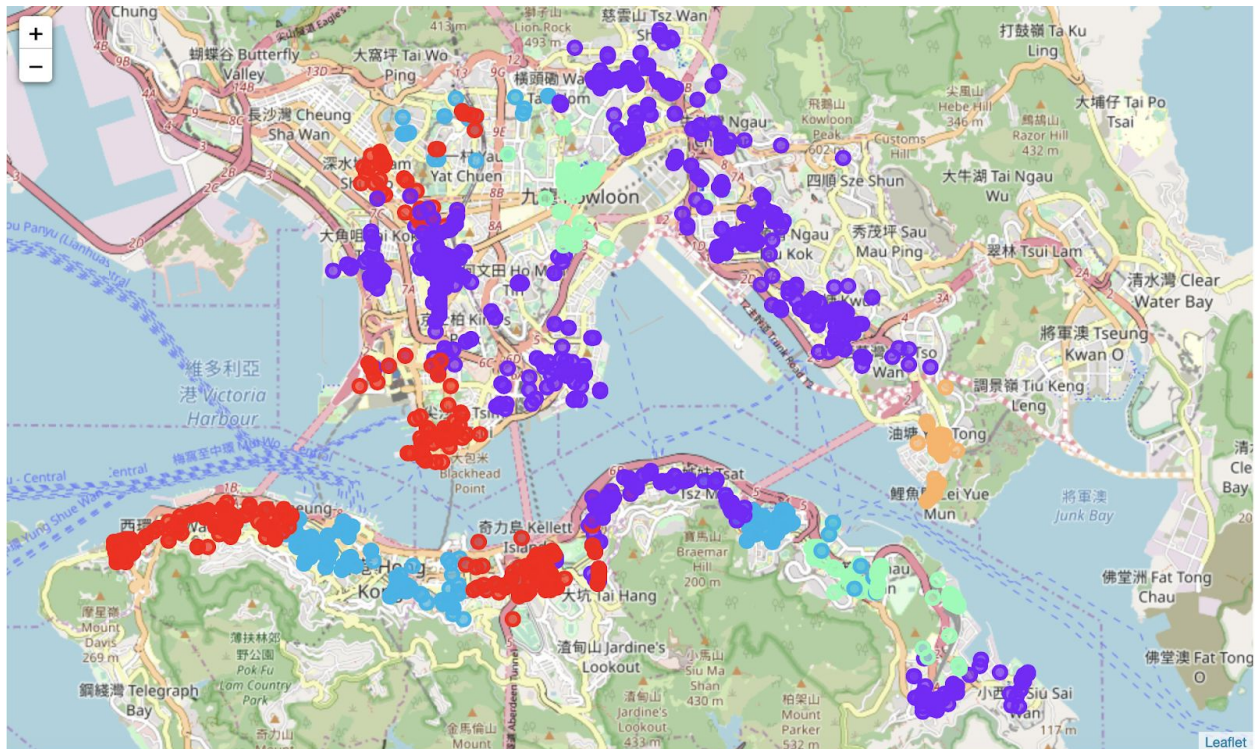
	Near	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category	6th Most Common Category	7th Most Common Category	8th Most Common Category	9th Most Common Category	10th Most Common Category
0	Admiralty	Café	Coffee Shop	Dim Sum Restaurant	Fast Food Restaurant	Italian Restaurant	Noodle House	Sandwich Place	Burger Joint	Chinese Restaurant	Cantonese Restaurant
1	CENTRAL,HK	Coffee Shop	Café	Dim Sum Restaurant	Fast Food Restaurant	Noodle House	Italian Restaurant	Burger Joint	Sandwich Place	Chinese Restaurant	Cha Chaan Teng
2	CHUK YUEN	Chinese Restaurant	Coffee Shop	Cha Chaan Teng	Cantonese Restaurant	Pizza Place	Dim Sum Restaurant	Japanese Restaurant	Hong Kong Restaurant	Café	Sushi Restaurant
3	Causeway Bay	Coffee Shop	Cha Chaan Teng	Hong Kong Restaurant	Noodle House	Chinese Restaurant	Cantonese Restaurant	Sandwich Place	Fast Food Restaurant	Restaurant	Ramen Restaurant
4	Chai Wan	Fast Food Restaurant	Hong Kong Restaurant	Chinese Restaurant	Cha Chaan Teng	Coffee Shop	Restaurant	Noodle House	Cafeteria	Taiwanese Restaurant	Cantonese Restaurant

We could see that Cafe and coffee Shop are the most common restaurants in Adminarity and Central.

5. Clustering

We run K-Means clustering to cluster the locations into 5 groups.

The below map visualizes the restaurant clusters result by K-Means algorithm.



When we examine above map we can label each cluster as follows:

- **Blue** Cluster 0 : Office Area with more western cuisine
- **Purple** Cluster 1: Fairly distributed Food Area
- **Red** Cluster 2 : Working or Shopping Area
- **Light Blue** Cluster 3: Asian Food Cluster
- **Yellow** Cluster 4 : Seafood intensive Area

The below list showing the detail ingredients of each cluster:

<p>---- Cluster 0: ----</p> <p>resturant freq</p> <p>0 Coffee Shop 0.17</p> <p>1 Café 0.15</p> <p>2 Fast Food Restaurant 0.06</p> <p>3 Dim Sum Restaurant 0.06</p> <p>4 Noodle House 0.05</p> <p>5 Italian Restaurant 0.05</p> <p>6 Cantonese Restaurant 0.04</p> <p>7 Sandwich Place 0.04</p> <p>8 Hong Kong Restaurant 0.04</p> <p>9 Chinese Restaurant 0.04</p> <p>10 Cha Chaan Teng 0.04</p> <p>11 Burger Joint 0.03</p> <p>12 Japanese Restaurant 0.02</p> <p>13 Restaurant 0.02</p> <p>14 Bakery 0.02</p> <p>15 American Restaurant 0.02</p> <p>16 Middle Eastern Restaurant 0.01</p> <p>17 Australian Restaurant 0.01</p> <p>18 Mexican Restaurant 0.01</p> <p>19 Asian Restaurant 0.01</p> <p>20 Pizza Place 0.01</p> <p>21 Ice Cream Shop 0.01</p> <p>22 Juice Bar 0.01</p> <p>23 Shanghai Restaurant 0.01</p> <p>24 Seafood Restaurant 0.01</p> <p>25 Taiwanese Restaurant 0.01</p> <p>26 Sushi Restaurant 0.01</p> <p>27 Soup Place 0.00</p> <p>28 Modern European Restaurant 0.00</p> <p>29 Jiangsu Restaurant 0.00</p>	<p>---- Cluster 1: ----</p> <p>resturant freq</p> <p>0 Fast Food Restaurant 0.11</p> <p>1 Chinese Restaurant 0.10</p> <p>2 Cha Chaan Teng 0.09</p> <p>3 Coffee Shop 0.07</p> <p>4 Noodle House 0.06</p> <p>5 Hong Kong Restaurant 0.05</p> <p>6 Cantonese Restaurant 0.05</p> <p>7 Café 0.05</p> <p>8 Japanese Restaurant 0.03</p> <p>9 Dim Sum Restaurant 0.03</p> <p>10 Thai Restaurant 0.03</p> <p>11 Sushi Restaurant 0.02</p> <p>12 Restaurant 0.02</p> <p>13 Taiwanese Restaurant 0.02</p> <p>14 Bakery 0.02</p> <p>15 Dessert Shop 0.02</p> <p>16 Seafood Restaurant 0.01</p> <p>17 Vietnamese Restaurant 0.01</p> <p>18 Hunan Restaurant 0.01</p> <p>19 Snack Place 0.01</p> <p>20 Shanghai Restaurant 0.01</p> <p>21 Hotpot Restaurant 0.01</p> <p>22 Italian Restaurant 0.01</p> <p>23 Asian Restaurant 0.01</p> <p>24 Sandwich Place 0.01</p> <p>25 Pizza Place 0.01</p> <p>26 Food Court 0.01</p> <p>27 Bubble Tea Shop 0.01</p> <p>28 Szechuan Restaurant 0.01</p> <p>29 Tea Room 0.01</p>
<p>---- Cluster 2: ----</p> <p>resturant freq</p> <p>0 Coffee Shop 0.20</p> <p>1 Fast Food Restaurant 0.06</p> <p>2 Chinese Restaurant 0.06</p> <p>3 Cha Chaan Teng 0.06</p> <p>4 Noodle House 0.04</p> <p>5 Cantonese Restaurant 0.04</p> <p>6 Café 0.04</p> <p>7 Hong Kong Restaurant 0.04</p> <p>8 Dim Sum Restaurant 0.03</p> <p>9 Dessert Shop 0.03</p> <p>10 Malay Restaurant 0.02</p> <p>11 French Restaurant 0.02</p>	<p>---- Cluster 3: ----</p> <p>resturant freq</p> <p>0 Coffee Shop 0.13</p> <p>1 Fast Food Restaurant 0.10</p> <p>2 Cantonese Restaurant 0.08</p> <p>3 Hong Kong Restaurant 0.08</p> <p>4 Japanese Restaurant 0.06</p> <p>5 Cha Chaan Teng 0.05</p> <p>6 Café 0.05</p> <p>7 Chinese Restaurant 0.04</p> <p>8 Vietnamese Restaurant 0.03</p> <p>9 Noodle House 0.02</p> <p>10 Dim Sum Restaurant 0.02</p> <p>11 Italian Restaurant 0.02</p>

12 Bubble Tea Shop 0.02 13 Snack Place 0.02 14 Italian Restaurant 0.02 15 Shanghai Restaurant 0.02 16 Japanese Restaurant 0.02 17 Bakery 0.02 18 Gastropub 0.02 19 Dumpling Restaurant 0.02 20 English Restaurant 0.01 21 Ramen Restaurant 0.01 22 Seafood Restaurant 0.01 23 Pizza Place 0.01 24 Restaurant 0.01 25 Sushi Restaurant 0.01 26 Sandwich Place 0.01 27 Taiwanese Restaurant 0.01 28 Mexican Restaurant 0.01 29 Burger Joint 0.01	12 Taiwanese Restaurant 0.02 13 Snack Place 0.01 14 Shanghai Restaurant 0.01 15 Seafood Restaurant 0.01 16 Sandwich Place 0.01 17 Fried Chicken Joint 0.01 18 American Restaurant 0.01 19 Salad Place 0.01 20 Steakhouse 0.01 21 Restaurant 0.01 22 Asian Restaurant 0.01 23 Hotpot Restaurant 0.01 24 Pizza Place 0.01 25 Ice Cream Shop 0.01 26 New American Restaurant 0.01 27 Jiangsu Restaurant 0.01 28 Food Court 0.01 29 Korean Restaurant 0.01
---- Cluster 4: ---- resturant freq 0 Seafood Restaurant 0.27 1 Fast Food Restaurant 0.12 2 Chinese Restaurant 0.08 3 Café 0.06 4 Bakery 0.04 5 Cha Chaa Teng 0.04 6 Japanese Restaurant 0.04 7 Cantonese Restaurant 0.02 8 Hakka Restaurant 0.02 9 Donburi Restaurant 0.02 10 Hong Kong Restaurant 0.02 11 Dim Sum Restaurant 0.02 12 Dessert Shop 0.02 13 Sushi Restaurant 0.02 14 Fish Market 0.02 15 Modern European Restaurant 0.02 16 Snack Place 0.02 17 Bubble Tea Shop 0.02 18 Food Court 0.02 19 Bistro 0.02 20 Thai Restaurant 0.02 21 Korean Restaurant 0.02 22 Juice Bar 0.02 23 Pizza Place 0.02 24 Ramen Restaurant 0.00 25 Restaurant 0.00	

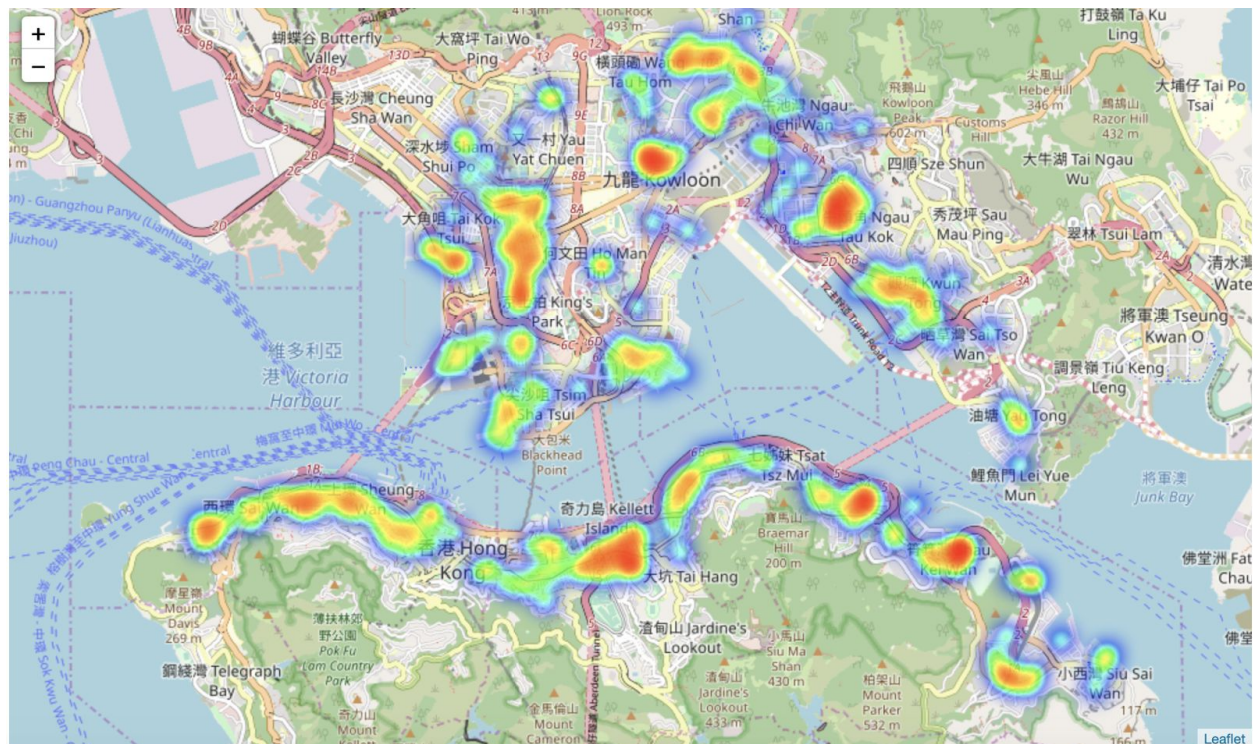
26	New American Restaurant	0.00
27	Noodle House	0.00
28	Sandwich Place	0.00
29	Middle Eastern Restaurant	0.00

6. Price levels analysis

We also conducted a price tier analysis on the location data. In order to exclude some meaningless data, we apply data punching to the above restaurant data. Some restaurant types have standard pricing or usually cheap which will not contribute much to that area of the restaurant. Therefore we will ignore those data.

There is a total of 945 data points left for the price level analysis. Price Tier level has the range between 0 to 4, from least price 0 to most price 4. The average price tier is 1.78 for all the data points.

The below Heat map visualizes the restaurant price distribution.



Expensive restaurant presenting in deep red color. It happens in the center of Mong Kok, Kowloon, Causeway Bay and the seaside area of Hong Kong Island. In these areas, it would be more likely to have to pay more for your dinner.

7. Conclusions

As we mentioned before, this report aiming to discover the restaurant cluster, we can easily see the outcome of the clustering is very obvious and can be easily distinguished. And it helps us find out those restaurant patterns. Also, the price tier analysis helps us understand the price distribution of the restaurants. Hope it would be a helpful figure for consumers to reference to.

Reference:

- Hong Kong census. "[Census labour data pdf](#)." *Labor*. Retrieved on 14 March 2007.
- HK Census. "[HK Census](#)." *Statistical Table of population*. Retrieved on 16 March 2007.
- https://en.wikipedia.org/wiki/List_of_MTR_stations
- [3] [Forsquare API](#)
- [5] [Google Map](#)