Restaurants in Hong Kong

19/03/2019
CAPSTONE

KING WAI SIU

BACKGROUND

Hong Kong is a financial center, a special administrative region and with a population of over 8 million people. It is a crowded city and congested. And undoubtedly there are many choices for accommodation. However, a majority of land in Hong Kong are nature, only less than 10% is residential or so-called city area. Perhaps you've never seen so many mountains in your life and you decided to reserve some time for mountains. There are three main region, HK island, New Territories and Kowloon.

Hong Kong is a food paradise due to its multicultural characteristics. Hong Kong is already really crowded, that means there are restaurants everywhere, perhaps too much and too confusing. Yet you want to let the statistics decide for you.

How to choose a restaurant in Hong Kong

PROBLEM BACKGROUND

Hong Kong is a financial center, a special administrative region and with a population of over 8 million people. It is a crowded city and congested. And undoubtedly there are many choices for accommodation. However, most of the land in Hong Kong are nature, only less than 10% is residential or so-called city area. Perhaps you've never seen so many mountains in your life and you decided to reserve some time for mountains. There are three main region, HK island, New Territories and Kowloon.

Hong Kong is a food paradise due to its multicultural characteristics. Hong Kong is already really crowded, that means there are restaurants everywhere, perhaps too much and too confusing. Yet you want to let the statistics decide for you.

PROBLEM DESCRIPTION

Suppose I travel and keep changing places very frequently. This is very hectic and plus I get to experience very different types of environment, of which I do not have much knowledge about. In such situation, food can be an important factor for decided how you rate your trips and plus also recommending it to the people. Food can also attract people around to world to try it out if it were to be the best. In such scenarios, we need to find the right place, at reasonable cost, to serve us the best possible way. So, there are few questions that must be addressed, such as:

How many types of foods are available in the restaurant? which is the nearest to me with good rating? How many "similar" restaurants are available nearby? Do the "similar" restaurants cost more? if so, what specialty do that have?

To address such question, XXYZ company's manager decides to allocate this project to me not just to find out solutions to the questions but also build a system that can help in recommending new places based on their rankings compared to the previously visited by me.

Expectations from this recommender system is to get answer for the questions, and in such a way that it uncovers all the perspective of managing recommendations. It is sighted to show:

What types of restaurants are present in an area? Where are the similar restaurant present based on a preference to food? How do different restaurants rank with respect to my preferences?

TARGET AUDIENCE

Target audiences for this project does not limit to a person who keeps travelling but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food. People who rarely use restaurants would prefer the most rated restaurants nearby them and all this could be easily handled by our recommender system. So, target for this project is basically everyone who is exploring different places or similar places.

SUCCESS RATE

With restaurants evolving, new food categories emerge, hybrid food starts to be more popular, we need a system that could help us access vast number of food varieties. It is impossible for a person to ask each one about their visit to a place and not everyone remembers everything. On the other hand, Computers are good at remembering things, and with Machine learning to its peak, it high time technology will by our personal guidance and help us personally based on our likes and dislikes. So, people would care about this project as their personal assistance and success rate could certainly increase with time.

DATA REQUIREMENTS

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyze such large dataset and produce analytics to find a solution.

Suppose I want to find a restaurant, then logically, I need 3 things:

1. Its geographical coordinates (latitude and longitude) to find out where exactly it is located.
2. Population of the neighborhood where the restaurant is located.
3. Average income of neighborhood to know how much the restaurant is worth.

Let's take a closer look at each of these:

1. To access location of a restaurant, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively.

2. Population of a neighborhood is very important factor in determining a restaurant's growth and number of customers who turn up to eat. Logically, the more the population of a neighborhood, the more people will be interested to walk openly into a restaurant and less the population, a smaller number of people frequently visit a restaurant. Also, if more people visit, better the restaurant is rated because it is accessed by different people with different taste. Hence is very important factor.

3. Income of a neighborhood is also very important factor as population was. Income is directly proportional to richness of a neighborhood. If people in a neighborhood earns more than an average income, then it is very much possible that they will spend more however not always true with very less probability. So, a restaurant assessment is proportional to income of a neighborhood.

DATA COLLECTION

Unluckily, no one seemed to care about the precise district and towns in Hong Kong to create a csv for them. I had to go to the Wikipedia, and extra few different tables, remove elements to create a decent town csv.

East Rail Line [edit]

Livery	Name	Photo	Connections	District	Opened	Code	Notes
	Lo Wu		Line 1 (Shenzhen Metro)	North	14 October 1949	LOW	[a]
	Lok Ma Chau		Line 4 (Shenzhen Metro)	Yuen Long	15 August 2007	LMC	[b]
	Sheung Shui			North	16 May 1930	SHS	
	Fanling			North	1 October 1910	FAN	
	Tai Wo			Tai Po	9 May 1989	TWO	
	Tai Po Market			Tai Po	7 April 1983	TAP	[c]

I had to remove those 'formerly' elements.

|	District	Station	Latitude	Longitude
0	Central and Western	Hong Kong	22.279328	114.162813
1	Central and Western	Admiralty	22.278381	114.165013
2	Central and Western	Central	22.350627	114.184916
3	Central and Western	Kennedy Town	22.281363	114.127832
4	Central and Western	HKU	22.283976	114.135507
...	...	...	...	...
90	Yuen Long	Kam Sheung Road	22.434648	114.064956
91	Yuen Long	Yuen Long	22.446021	114.034944
92	Yuen Long	Long Ping	22.447670	114.025449
93	Yuen Long	Tin Shui Wai	22.459611	114.002238
94	Yuen Long	Lok Ma Chau	22.515505	114.066628

95 rows × 4 columns

I got most of the locations using geopy.

Here are all the subway stations in Hong Kong, well, the stations are literally a town on its own.
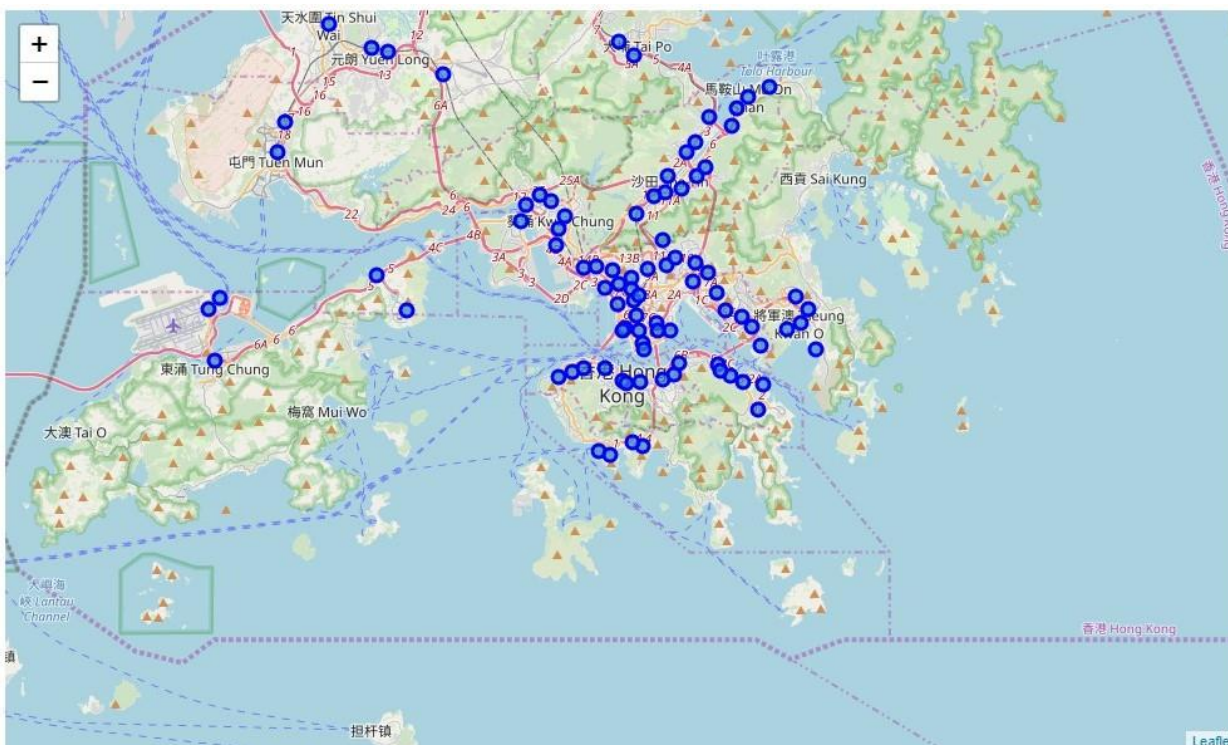
Colored version:
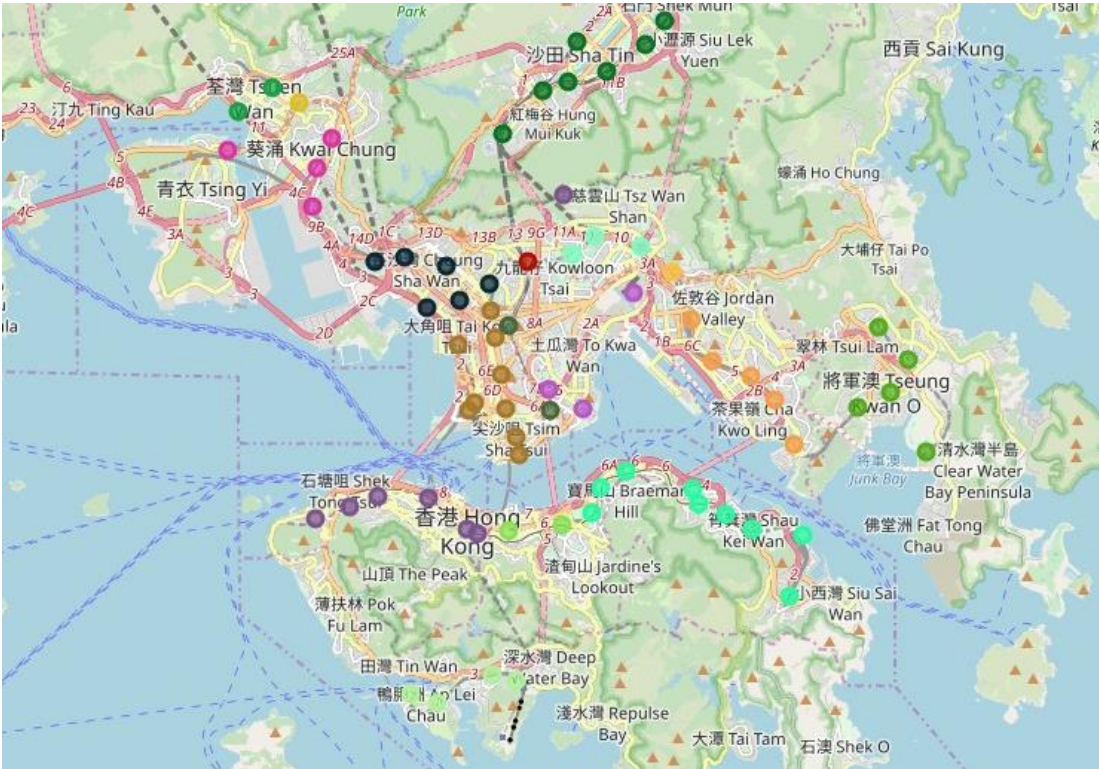
EXPLORE NEIGHBORHOOD

I added the suffix after each station name to ensure the result returned are the station. For example, there is a station named exactly same as Hong Kong, if I don't add Hong Kong Station, the results will return all the restaurants thus giving an inaccurate result.

	Station	District	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Hong Kong Station	Central and Western	22.279328	114.162813	Ruth's Chris Steak House (茹絲葵牛排餐廳)	22.279188	114.163833	Steakhouse
1	Hong Kong Station	Central and Western	22.279328	114.162813	Lobster Bar and Grill	22.277831	114.164887	Seafood Restaurant
2	Hong Kong Station	Central and Western	22.279328	114.162813	Howard's Gourmet (好酒好蔡)	22.281469	114.161404	Chinese Restaurant
3	Hong Kong Station	Central and Western	22.279328	114.162813	Mott 32 (卅二公館)	22.280696	114.159380	Dim Sum Restaurant
4	Hong Kong Station	Central and Western	22.279328	114.162813	Lawry's The Prime Rib	22.281377	114.162287	Steakhouse
5	Hong Kong Station	Central and Western	22.279328	114.162813	Café Gray Deluxe	22.277506	114.166320	Café
6	Hong Kong Station	Central and Western	22.279328	114.162813	Summer Palace (夏宮)	22.277120	114.164802	Cantonese Restaurant
7	Hong Kong Station	Central and Western	22.279328	114.162813	Waterside Terrace	22.277438	114.163969	Thai Restaurant
8	Hong Kong Station	Central and Western	22.279328	114.162813	TWG Tea	22.278135	114.165884	Café
9	Hong Kong Station	Central and Western	22.279328	114.162813	Tien Yi Chinese Restaurant (天一酒家)	22.277990	114.165901	Cantonese Restaurant

MOST POPULAR TYPE OF RESTAURANTS

Top 15 of the most restaurant type in HK
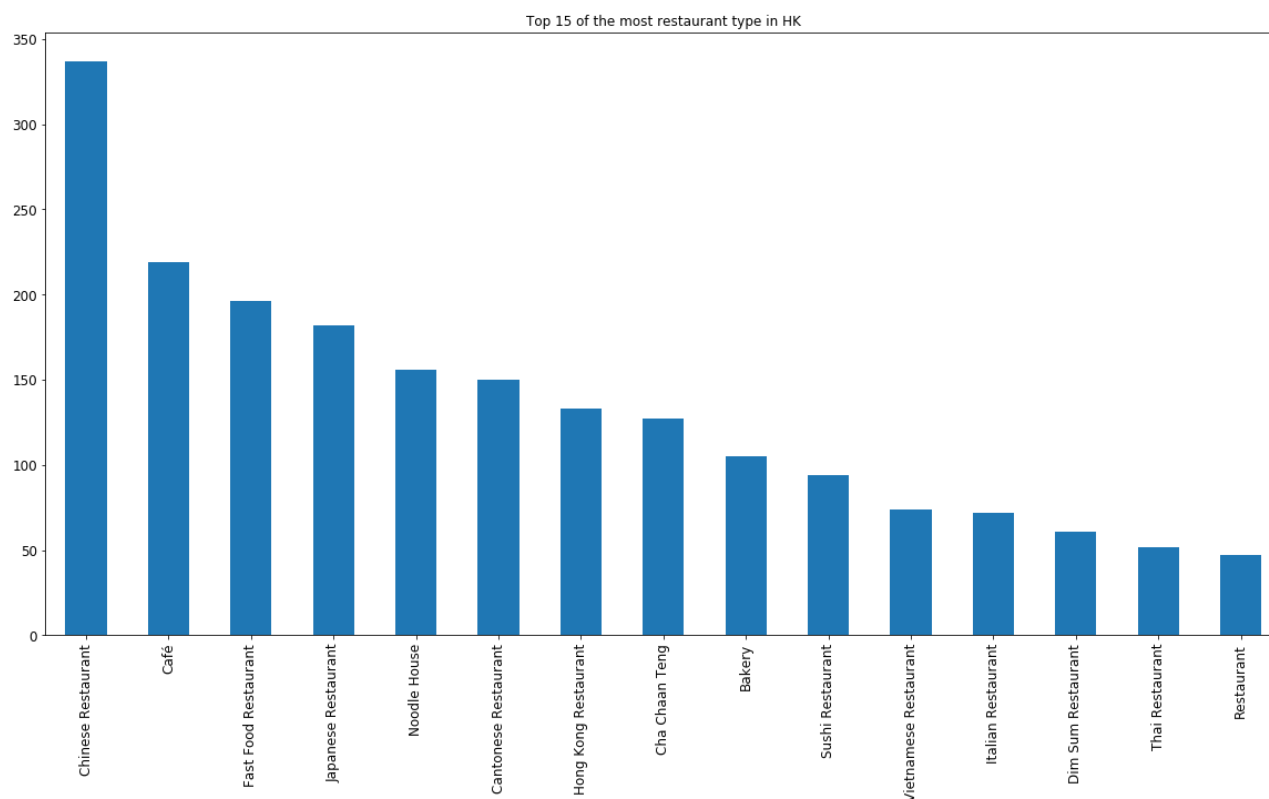
Traditional Chinese restaurants outnumber others by a significant portion, some of those type care repeated.

TOTAL RESTAURANTS IN DISTRICTS

District	count_col
Eastern	487
Yau Tsim Mong	478
Central and Western	338
Sha Tin	243
Kwun Tong	167
Sham Shui Po	158
Wan Chai	135
Sai Kung	128
Kowloon City	103
Yau Tsim Mong/Kowloon City	101
Islands	99
Kwai Tsing	83
Tsuen Wan	76
Yuen Long	76
Southern	70
Tuen Mun	52
Wong Tai Sin	51
North	38
Tai Po	35
Kwai Tsing/Tsuen Wan	22
Sham Shui Po/Kowloon City	22
Wong Tai Sin/Kwun Tong	4

Eastern district seems to have the most restaurants results. Let's further analyze the eastern district.

TOP RESTAURANT TYPES IN EASTERN DISTRICT


Top 20 restaurant type in Eastern District
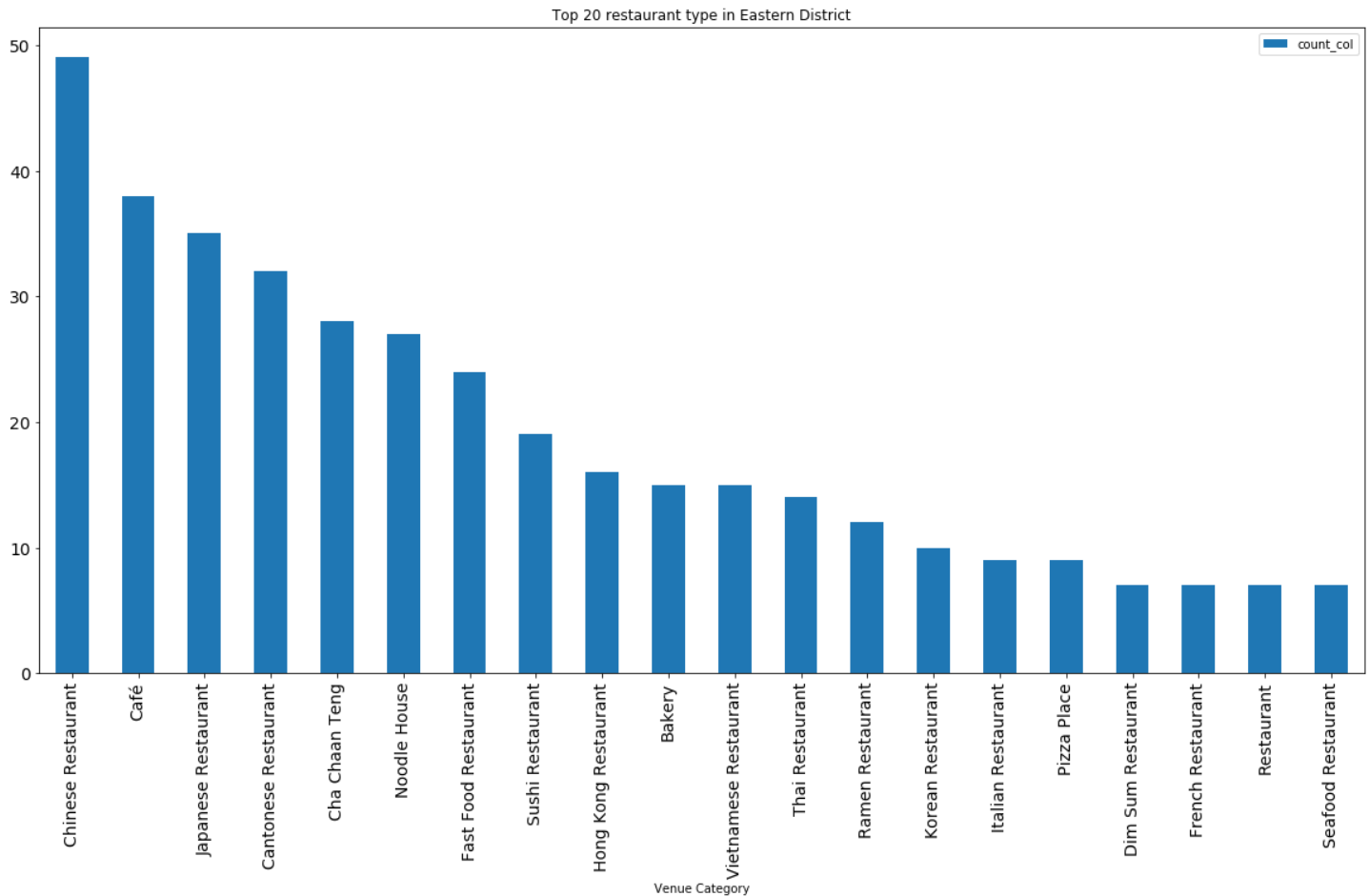
The top 10 shows a great decrease in trend. After pizza place, it seems to level off. Fast food restaurant is not that popular here as the eastern district is also an affluent neighborhood.

In order to further analyze the eastern district, one-hot encoding is used.

	Venue Category_American Restaurant	Venue Category_Asian Restaurant	Venue Category_Australian Restaurant	Venue Category_BBQ Joint	Venue Category_Bakery	Venue Category_Beijing Restaurant	Venue Category_Belgian Restaurant	Venue Category_Bistro
0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0

CLUSTERING

First, we need to find out the best k number to do the k-means clustering. An iterative test for inertia of the number of clusters is plotted.
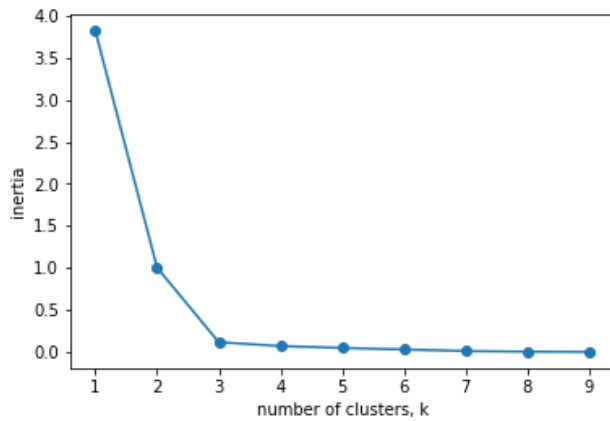
k number should be 3, after 3 the inertia level off
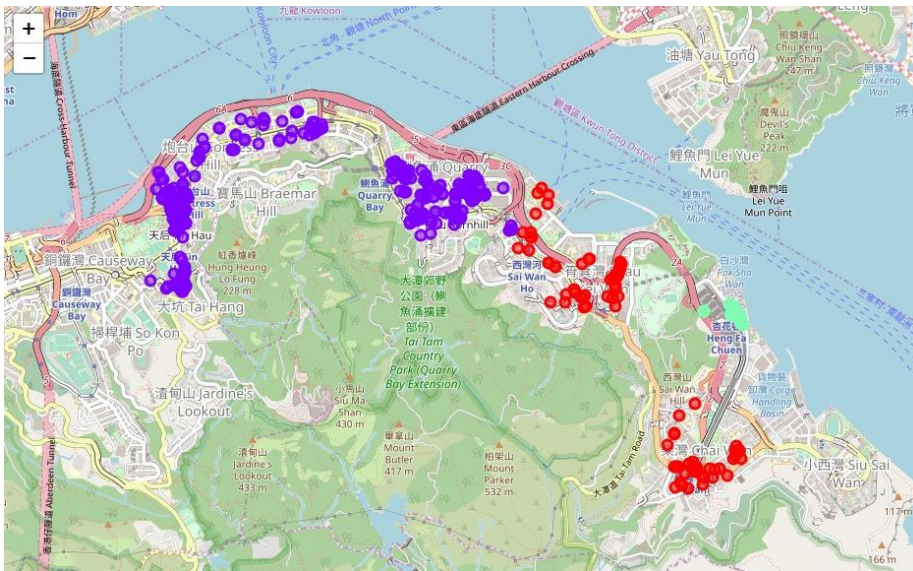
After clustering, the restaurants can be seen as the above clusters. The cluster containing the least number is actually the richest neighborhood, maybe that concludes their different restaurant preference.

CONCLUSION

It seems that nearby neighborhoods have similar eating preference, and somehow a small cluster is found because of their unique preference. Well, the rich people in Hong Kong can keep their eating habits.