

Clustering the Districts in Hong Kong and Giving Recommendations

1. Introduction

Scenario:

I am a data scientist residing in Hong Kong. One day, a friend of mine asked me, where is the most suitable place for her to open her own restaurant.

Business Problem:

The challenge is to find a suitable district to open a new restaurant, where the market is not saturated yet.

2. Data Processing

2.1. Data description

The following data is required to answer the issues of the problem:

- List of districts and regions in HK
- Venues for each district (that can be clustered)
- Coordinates for the districts and venues (for mapping purpose)

The data will be used as follows:

- Use Foursquare and geopy data to map top 10 venues for all HK districts and clustered in groups (as per Course LAB)
- Data will be searched in open data sources if available, such as Wikipedia.

2.2. Data cleaning

The district data is scrapped from the Wikipedia using BeautifulSoup. We dropped some unnecessary columns, getting the district names and the corresponding regions in the end.

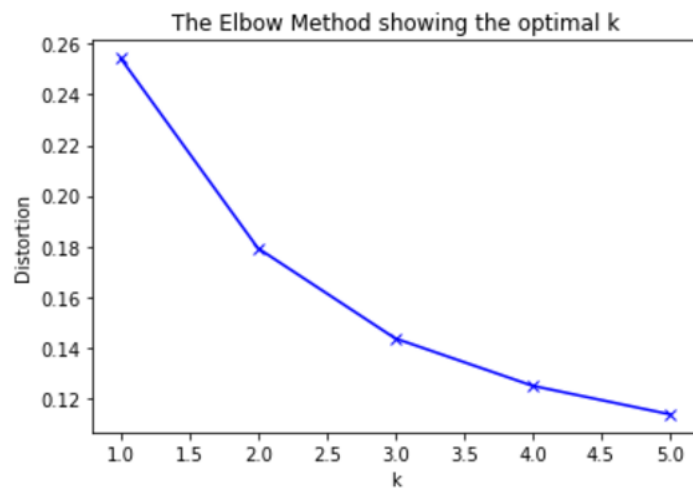
Using geopy, we get all the venues in each district. After one-hot coding, we calculate the frequency of each venue, and sort the top 10 venues.

	Neighbourhood	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	Adelaide, King, Richmond	Coffee Shop	Café	Thai Restaurant	Bar	Steakhouse	Gym	Restaurant	American Restaurant	Breakfast Spot	Hotel
1	Agincourt	Lounge	Breakfast Spot	Skating Rink	Chinese Restaurant	Sandwich Place	Eastern European Restaurant	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, S...	Park	Playground	Yoga Studio	Eastern European Restaurant	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
3	Albion Gardens, Beaumont Heights, Humbergate,...	Grocery Store	Fast Food Restaurant	Pizza Place	Sandwich Place	Coffee Shop	Beer Store	Pharmacy	Fried Chicken Joint	Empanada Restaurant	Electronics Store
4	Alderwood, Long Branch	Pizza Place	Gym	Pool	Skating Rink	Pharmacy	Pub	Coffee Shop	Sandwich Place	Diner	Discount Store

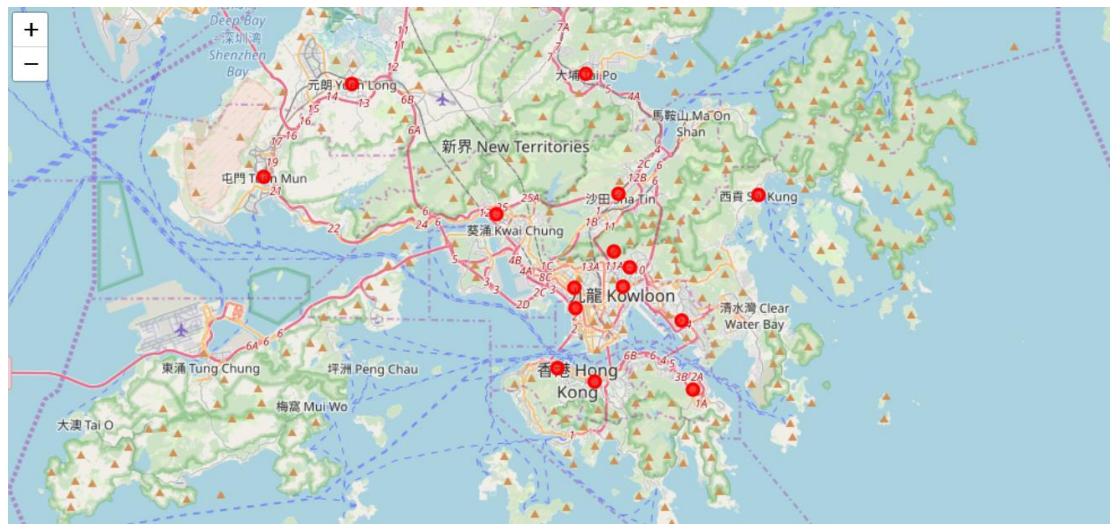
3. Clustering

We choose k-means as the clustering method. The k-means++ is used to speed up the clustering process. The clusters are formed based on the frequency of each venues in one district.

We also use the elbow method to determine the best K to choose as cluster number. It turns out that 2 is the best choice.



Clustering result



The final result is surprising though. Only one district is in the second cluster, while all others are in the first.

From the cluster details below, we can see that the first cluster contains most of the districts. Restaurants, coffee shop and shopping mall have the highest presence. In the second cluster, there is actually only one district, which is the Eastern. This may be because of the small number of districts and limited data volume. It seems that the Eastern district is a relatively wilder place, with less entertainment venues and more public facilities.

	Area	Cluster_Labels	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	12.44	0.0	Café	Japanese Restaurant	Coffee Shop	Cocktail Bar	French Restaurant	Indian Restaurant	Chinese Restaurant	Wine Bar	Bar	Italian Restaurant
3	9.83	0.0	Café	Coffee Shop	Italian Restaurant	Hotel	Hong Kong Restaurant	Thai Restaurant	Chinese Restaurant	Bar	Bakery	Cantonese Restaurant
4	9.35	0.0	Chinese Restaurant	Noodle House	Dessert Shop	Hong Kong Restaurant	Italian Restaurant	Snack Place	Dumpling Restaurant	Indonesian Restaurant	Café	Cha Chaan Teng
5	10.02	0.0	Thai Restaurant	Dessert Shop	Café	Chinese Restaurant	Fast Food Restaurant	Cha Chaan Teng	Coffee Shop	Halal Restaurant	Bakery	Noodle House
6	11.27	0.0	Chinese Restaurant	Café	Coffee Shop	Cha Chaan Teng	Hong Kong Restaurant	Japanese Restaurant	Sushi Restaurant	Fast Food Restaurant	Restaurant	Department Store
7	9.30	0.0	Coffee Shop	Fast Food Restaurant	Temple	Cantonese Restaurant	Park	Chinese Restaurant	Shopping Mall	Szechuan Restaurant	Astrologer	Pool
8	6.99	0.0	Coffee Shop	Japanese Restaurant	Noodle House	Café	Fast Food Restaurant	Clothing Store	Hotpot Restaurant	Vietnamese Restaurant	Shanghai Restaurant	Hotel
10	23.34	0.0	Scenic Lookout	Sushi Restaurant	Nature Preserve	Trail	Zoo	Fish & Chips Shop	French Restaurant	Food Truck	Food Court	Flea Market
11	136.61	0.0	Seafood Restaurant	Café	Thai Restaurant	Coffee Shop	Italian Restaurant	Chinese Restaurant	Burger Joint	Park	Pub	Pizza Place

Cluster 1

	Area	Cluster_Labels	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
1	18.56	1.0	Bus Stop	Zoo	Gastropub	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Food Truck	Food Court	Flea Market	Fish & Chips Shop

Cluster 2

4. Conclusion

Based on the analysis, I would recommend Eastern District as the best place to open a new restaurant.

Reference:

https://en.wikipedia.org/wiki/Districts_of_Hong_Kong