Define Land Types and Social Activities

Based on Foursquare Spatial Data

The urban structure in high-density cities are usually very difficult to identify. When considering the gap between physical built environment and socio-economic activities, the challenge is getting even bigger.

This project proposed a method to define land type on the basis of street block. We'll able to describe what the spatial and social structure of Kennedy Town in Hong Kong is.

Data

▶ (1) block data

I firstly scraped road data (in .shp format) from OpenStreetMap. Then preprocessed it by ArcGIS Map to block polygons. (The calculation processes the techniques applied are out of the scope of this course, so I will not include the details here.) Because block is a relatively homogeneous land piece surrounded by streets, it is representative to define land type.

Based on the block polygon data,

(i) A GeoJSON file was created for further choropleth map production;

(ii) Centroids' latitudes and longitudes data of blocks were also calculated via ArcGIS Map.

▶ (2) Foursquare location data

I utilized Foursquare API to explore the blocks and segment them. Each block has scraped top 100 venues within a radius of 500 meters.
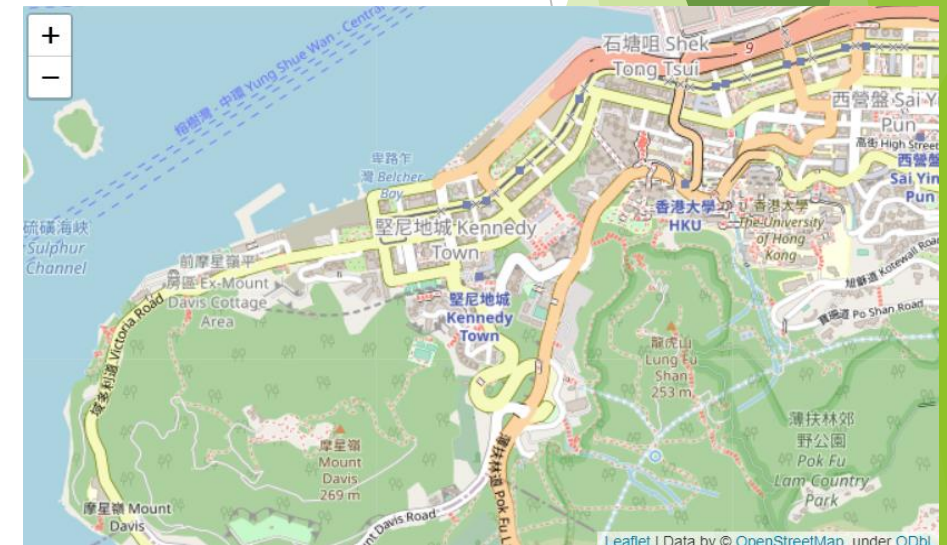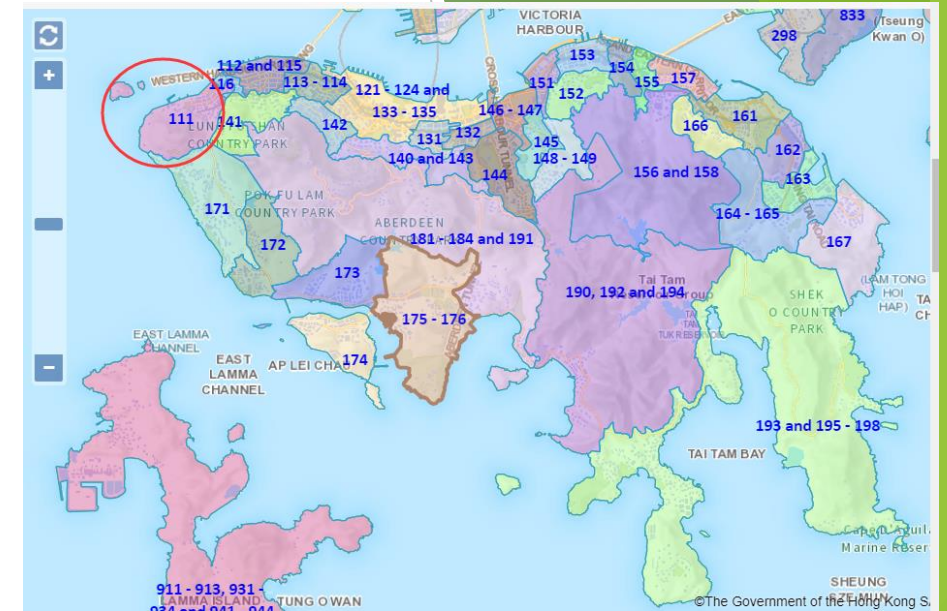
Methodology

▶ (1) site description

I chose Kennedy Town in Hong Kong as the study site here. Kennedy Town locates at the western end of Sai Wan on Hong Kong Island in Hong Kong.

Kennedy Town (KT) in this project refers the area of Tertiary Planning Unit 111. Tertiary Planning Unit is a geographic reference system demarcated by the Planning Department for the Territory.

Geopy library was used to get the latitude and longitude values of KT.

Methodology

▶ (2) Data preprocessing

There are total 47 blocks in KT. The sizes of blocks vary a lot.

I firstly imported the data from csv and collated it into a pandas DataFrame.

After this, choropleth was used to generate a leaflet map indicating all the blocks' boundaries and their centroids.

Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions
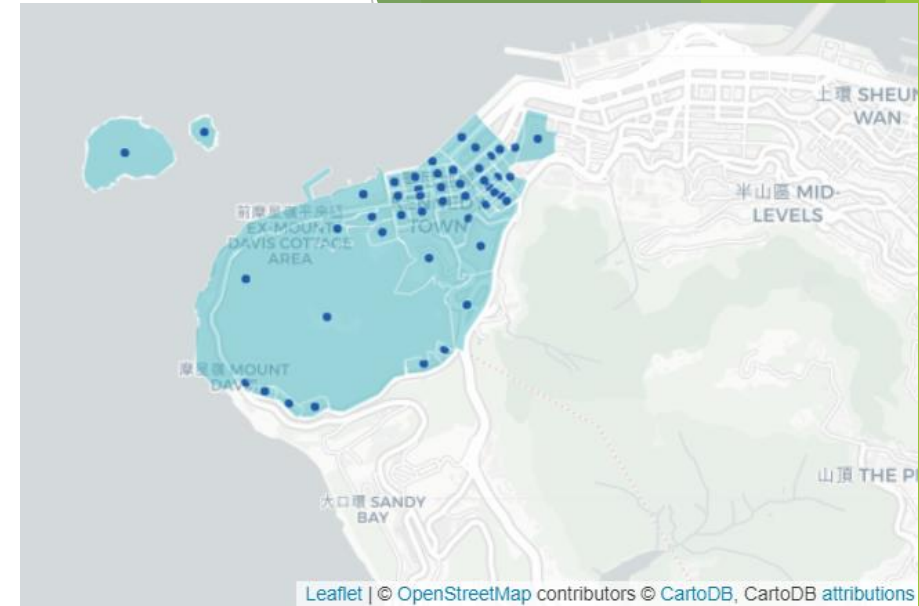
	FID	lat	lng	Count
0	0	22.272960	114.122259	1
1	1	22.273106	114.121010	1
2	2	22.273676	114.119808	1
3	3	22.274046	114.118835	1
4	4	22.274937	114.127660	1

Methodology

▶ (3) Import Foursquare spatial data

Foursquare version was set to be '20180605' with limit 100 to extract the top 100 venues around each block within a radius of 500 meters. There're total 2782 venues in 77 unique categories were found and extracted. All the records were grouped by blocks. Within each block, the frequency of occurrence for each category was calculated.
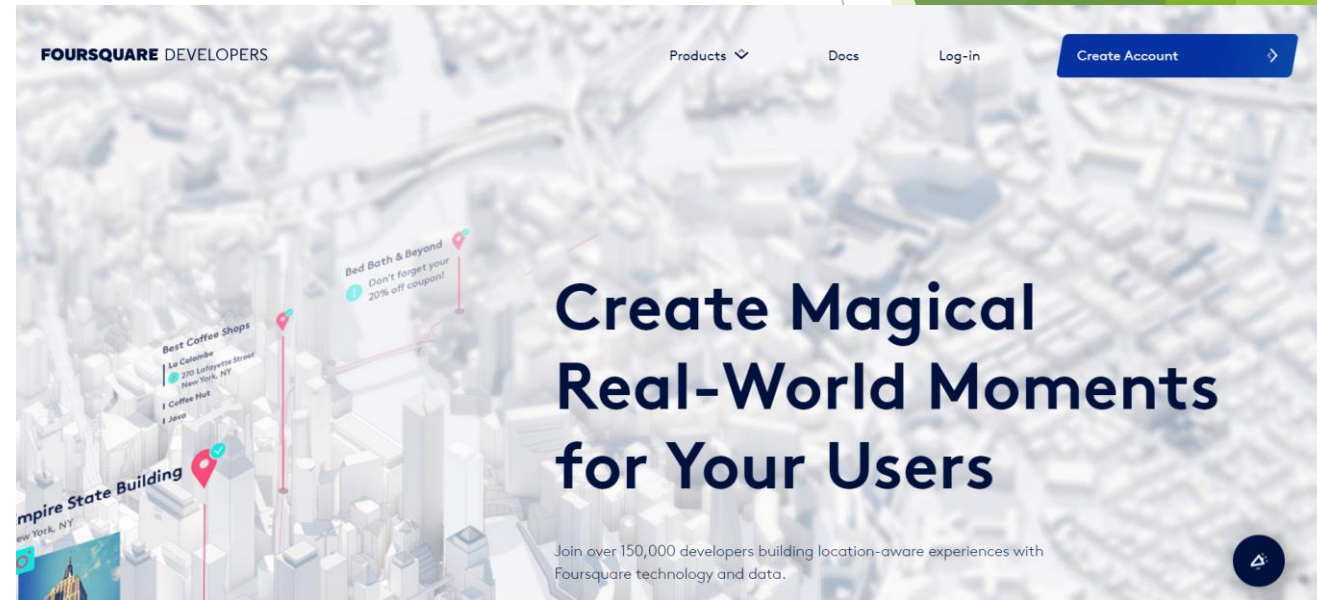
Methodology

- (4) K-means Clustering

The top 10 categories of venues in each block were used for k-means clustering. Before clustering, non-related columns like "Block ID" were dropped off as data preparation.

Result

I firstly set 3 cluster (K=3) to see the results. It turns out that 3 would be fine enough.

Among 47 blocks, 39 blocks are of cluster 0, 5 blocks are of cluster 1, 2 blocks are of cluster 2. 1 block in color grey represents a large country park, Mount Davis, which has no venue records scraped from Foursquare.
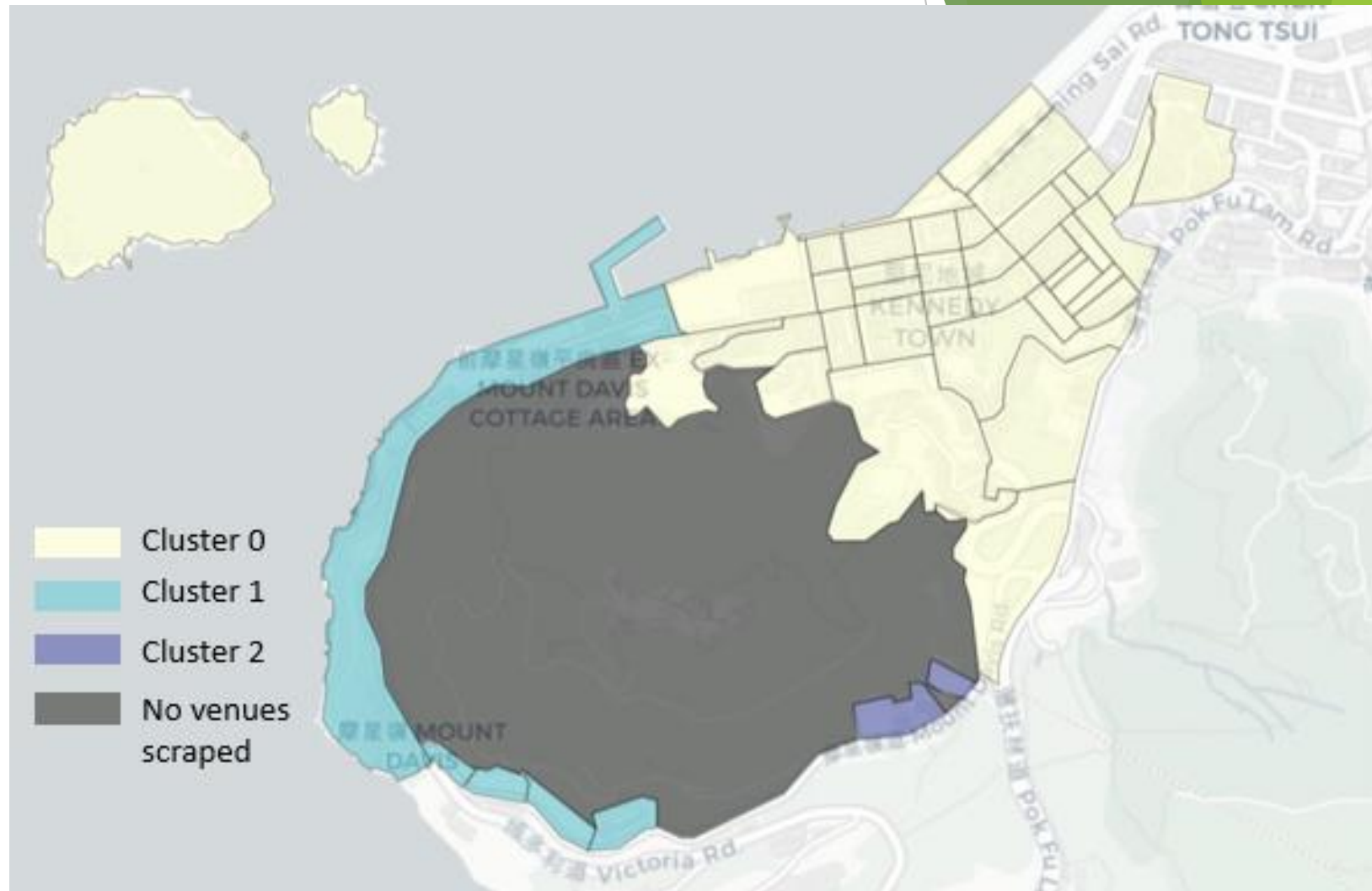
Result

▶ (1) cluster 0: residence places

Cluster 0 covers the most part of blocks in KT. Based on the top 10 avenue categories in the table below, it's probably representing land types, or urban places, of residence. There're plenty of facilities for daily activities, e.g. restaurant, coffee shop, market, and so on.

BlockID	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
6	Chinese Restaurant	Vietnamese Restaurant	Sports Bar	Mountain	Market	Japanese Restaurant	Playground	French Restaurant	Coffee Shop	Australian Restaurant
7	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Café	Mexican Restaurant	Vietnamese Restaurant	Italian Restaurant	Fish & Chips Shop	Hong Kong Restaurant	Cha Chaan Teng
8	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Café	Mexican Restaurant	Vietnamese Restaurant	Fish & Chips Shop	French Restaurant	Italian Restaurant	Cha Chaan Teng
10	Japanese Restaurant	Chinese Restaurant	Café	Coffee Shop	Vietnamese Restaurant	Bakery	Hong Kong Restaurant	Cha Chaan Teng	Mexican Restaurant	Trail
11	Japanese Restaurant	Coffee Shop	French Restaurant	Mexican Restaurant	Bakery	Italian Restaurant	Cha Chaan Teng	Ice Cream Shop	Hotpot Restaurant	Hotel
12	Japanese Restaurant	Coffee Shop	Chinese Restaurant	Vietnamese Restaurant	Café	Mexican Restaurant	Hong Kong Restaurant	French Restaurant	Fish & Chips Shop	Cha Chaan Teng
13	Japanese Restaurant	Chinese Restaurant	Coffee Shop	Vietnamese Restaurant	Café	Mexican Restaurant	Hong Kong Restaurant	Italian Restaurant	Bakery	French Restaurant

Result

▶ (2) cluster 1: coastal areas for tourism and leisure

All 5 blocks are near marina with places for leisure and fun, e.g. historic site, dessert shop, etc.

	BlockID	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	0	Harbor / Marina	Hostel	Historic Site	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant
1	1	Harbor / Marina	Hostel	Historic Site	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant
2	2	Harbor / Marina	Hostel	Historic Site	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant
3	3	Harbor / Marina	Hostel	Historic Site	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant
30	31	Harbor / Marina	Hostel	Playground	Fish & Chips Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant

Result

▶ (3) cluster 2: public facilities for sports and physical activities

These 2 blocks are close to each other located near mountain foot. Both of them have playgrounds.

	BlockID	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
4	4	Mountain	Playground	Vietnamese Restaurant	Chinese Restaurant	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant	English Restaurant	Fast Food Restaurant
5	5	Mountain	Trail	Playground	Scenic Lookout	Fast Food Restaurant	Coffee Shop	Dessert Shop	Dim Sum Restaurant	Discount Store	Dumpling Restaurant

Discussion

Based on foursquare spatial venue data and k-means clustering, we can find similar blocks and grouped them into different clusters. Based on each cluster's characteristics, combined with real world experience, we are able to define each cluster's type and guess what social activities might be related to such cluster types.

Conclusion

▶ Social activities are highly related with physical built environment and facilities. Based on this, explore spatial data around a given area would be an efficient way to quick define the area type and social characteristics at very low cost.