Hong Kong is the city with the highest property price per m2 among the world, USD$28,570, and as a result, the citizens are rather to rent an apartment than purchasing one so that the moving rate increased. Therefore, there is a demand for information about different area to assist those to choose their home location, but, for now, there is not an efficient database to fulfil this demand. And so this research is going to cluster the 18 districts in Hong Kong by the combination of the facilities in each of them.

Data sources

To solve this problem, two types of data are needed. First, the list of districts in Hong Kong, which is available on Wikipedia, shows the districts name and was last updated in Nov 2018. Second, the facilities data from Foursquare API shows the facilities within the set radius around the centre of each district.

Data cleaning

The first data, district list, has 6 columns, which are district, Chinese, population, area, density, region. As the research is going to cluster the districts just by the facilities, only the district column is needed, while other columns are dropped from the data set. In order to find the location of each district centres, full addresses are needed, so I added ‘, Hong Kong, China’ after the name to use Geopy to find each longitude and latitude. Below table shows the first 5 rows of the table:

|  |  |  |  |
| --- | --- | --- | --- |
|  | District | latitude | longitude |
| **0** | Central and Western | 22.285239 | 114.150679 |
| **1** | Eastern | 22.279594 | 114.232420 |
| **2** | Southern | 22.247692 | 114.154762 |
| **3** | Wan Chai | 22.277465 | 114.173026 |
| **4** | Sham Shui Po | 22.330095 | 114.160940 |

Methodology

I used the foursquare API to find the venues within 1000m from each district centers by explore function.