

2. Data acquisition and cleaning

2.1 Data sources

The data of latitude and longitude of districts of Shanghai and Beijing can be found at <https://zh.wikipedia.org/wiki/Template:上海行政区划地图> and <https://zh.wikipedia.org/wiki/北京市行政区划>, respectively. I found 19 districts for Beijing originally, including Beijing downtown, Dongcheng, Xicheng, Chongwen, Xuanwu, Chaoyang, Fengtai, Shijingshan, Haidian, Mentougou, Fanshang, Tongzhou, Shunyi, Changping, Daxing, Huairou, Pinggu, Miyun and Yanqing. I found 18 districts for Shanghai originally, including Huangpu, Xuhui, Changning, Jingan, Putuo, Zhabei, Hongkou, Yangpu, Minhang, Baoshan, Jiading, Pudong, Jinshan, Songjiang, Qingpu, Nanhui, Chongming and Fengxian. Later I will check those district on Foursquare to see if they have enough venues information and decide whether to keep them in the future clustering analysis or not.

2.2 Data cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were no missing values of the original data. However, there are still one problem with the datasets. I checked the venues around each district on Foursquare website <https://foursquare.com>, a location technology platform dedicated to improving how people move through the real world. I get venues information of a specific longitude and latitude by getting access to Foursquare database and defining a radius around that location. However, not every district has venues information within its range of location. This will have influence on the later cluster analysis. So at last, I reluctantly decided to delete those districts.

2.3 Area of Interest selection

Below is the information of dropped districts due to the lack of venues information from Foursquare website. Finally, we are going to keep 13 and 8 districts for Shanghai and Beijing, respectively, to do the following venues analysis and tentative clustering.

Cities	Kept Districts	Dropeed Districts	Reason for dropping
Shanghai	13, Shanghai (downtown), Huangpu, Xuhui, Changning, Jingan, Putuo, Zhabei, Hongkou, Yangpu, Minhang, Jiading, Songjiang, Nanahui	5, Baoshan, Jinshan, Qingpu, Chongming, Fengxian	Lack of venues information from Foursquare website
Beijing	8, Beijing (downtown), Dongcheng, Xicheng, Chongwen, Xuanwu, Chaoyang, Haidian, Changping	11, Fengtai, Shijingshan, Mentougou, Fangshan, Tongzhou, Shunyi, Daxing, Huairou, Pinggu, Miyun, Yanqing	