

Data acquisition and cleaning

The followings are data sources that I use for this project:

1st Data:

Distribution of Estimated Population & Population Density (person/km²) by Sector and Community from [Dubai Statistics Center website](#)

2nd Data:

- Since using **geocoder** library to retrieve a list of geographical coordinate doesn't get me the result I want, I had to use data from web-scraping using **beautifulsoup** library to get the coordinate for each community of Dubai city from Wikipedia page [List of communities in Dubai](#).
- Then I combined the two previous datasets and dropped the communities that stayed without coordinate.

3rd Data:

Finally, I use the coordinates from the previous combined dataset and foursquare credentials through its API and retrieve the popular venues along with their details and filter the result to get only the following venues that I use in clustering:

1. College and University.
2. Shopping mall.
3. Historical site.
4. Coffee shop.
5. Park.
6. Bank
7. Gym.

