**Capstone Project: The Battle of Neighbourhoods**

**1. Introduction**

**1.1 Description of the Problem**

There are many practical questions that require the comparison across city neighborhoods. For example, a job seeker with transferable skills may wish to focus his/her search on a single neighborhood with jobs that best match his/her qualifications, rather than dispersing his/her search efforts across multiple neighborhoods. Likewise, a restaurant looking to expand its locations might perhaps select neighborhoods it wishes to expand into before considering particular sites or neighborhoods. Additionally, many within-city computations might be aided by modelling a neighborhood’s relationship to other neighborhoods. For example, a person buying or renting a home in a new city might want to be able to compare the neighborhoods of the city.

**1.2 Objective**

Taipei and New Taipei City are two major cities in Taiwan. Both cities have been centers of attention for residential, job employment, tourism, education, shopping and sport activities. Both municipalities located in the north of Taiwan with New Taipei City surrounds Taipei.

Taipei is the capital city of Taiwan. Most part of New Taipei City are mountain area.

The aim of this project is to segment these two cities' neighborhoods based on data collected from Foursquare about venue categories in totally 41 neighborhoods across the cities. Using segmentation and clustering, I hope I can determine:

1. The similarity or dissimilarity between neighborhoods.
2. Classification of a neighborhood inside a city whether it is residential, tourism places, or others.

**2. Data**

This project will rely on public data from websites and Foursquare.

In order to segment the neighborhoods and explore them, I will need a dataset that contains the 2 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

A Wikipedia page: <https://en.wikipedia.org/wiki/List_of_townships,_county-administered_cities_and_districts_of_Taiwan> exists that has a list of townships, county-administered cities and districts of Taiwan. The Nominatim library can be used to get the latitude and longitude values of all neighborhoods.

### **3. Methodology**

#### 3.1 Data preparation

I scraped the Wikipedia page and wrangled the data, cleaned it, and then read it into a pandas dataframe so that it is in a structured format that consist of two columns of 2 Boroughs and 41 Neighborhoods of both cities.

I then use the Nominatim library to get the latitude and longitude values of all neighborhoods.

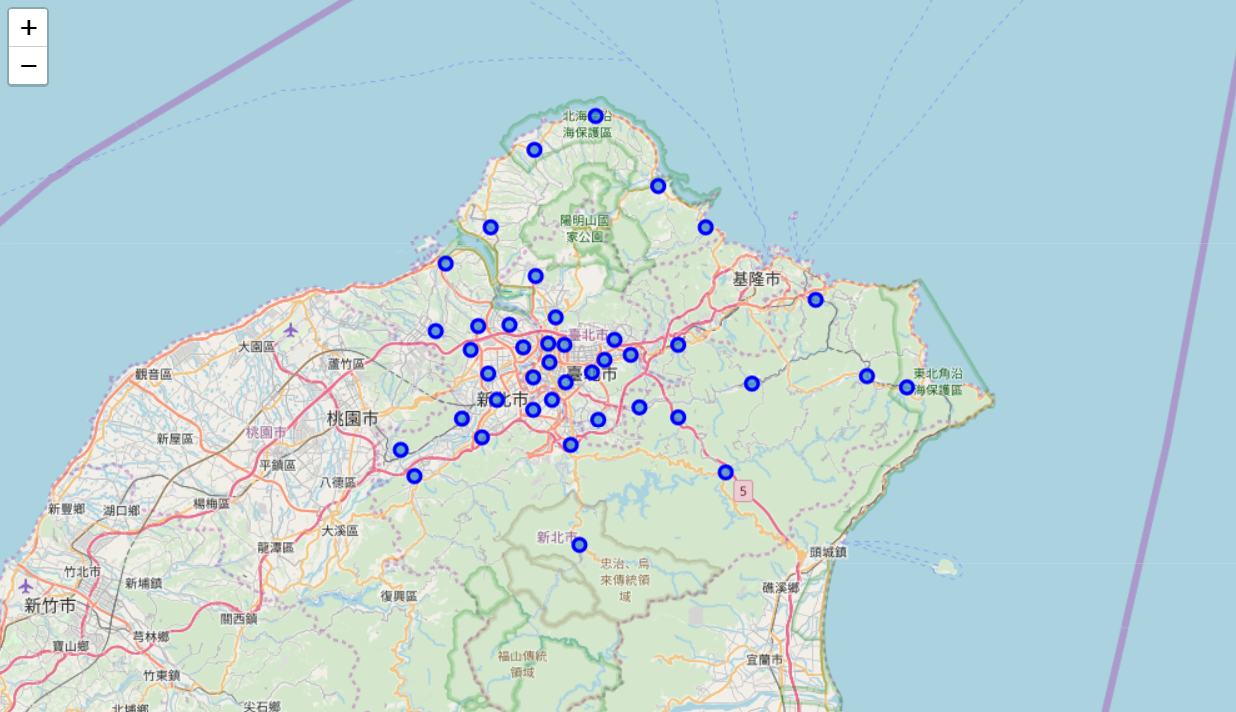


Fig. 1. Map of Taipei and New Taipei City. The markers are the neighborhoods in my analysis.

**3.2 Data acquisition**

The Foursquare API was used to search for nearby venues of each neighborhood in radius of 2000 meters. Only the venues name and venues category (i.e. café, restaurant, school, etc.) are extracted.

After obtaining all the venues, the total number of venues in each category is counted for each neighborhood. To clarify, I also builded a table with top 10 most frequent venue categories for each neighborhood.

**3.3 Data Analysis**

According to the total number of venues in each category, I applied k-means clustering method to compare neighborhoods among cities.

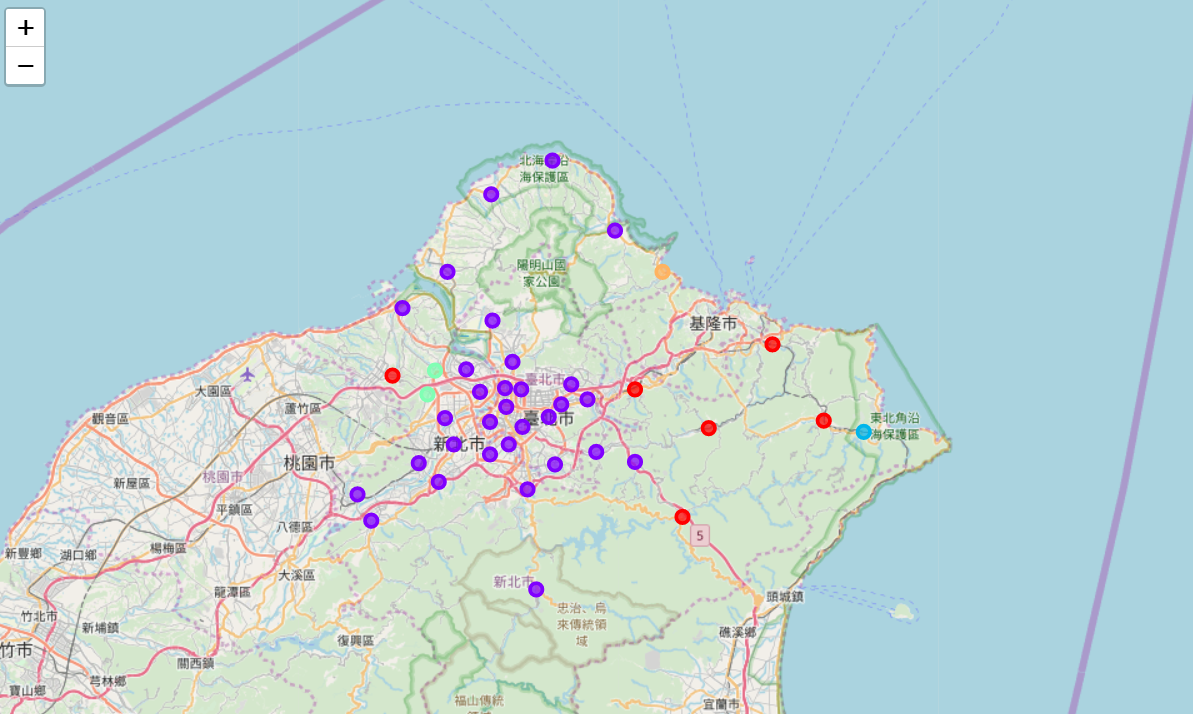


Fig. 2. k-means clustering visualization of all the neighborhoods in Taipei and New Taipei City.

### **4. Results**

**4.1 Cluster 0**

Cluster 0 contains the mountain neighborhoods of New Taipei City with historic sites, train stations, convenience stores, and restaurants.

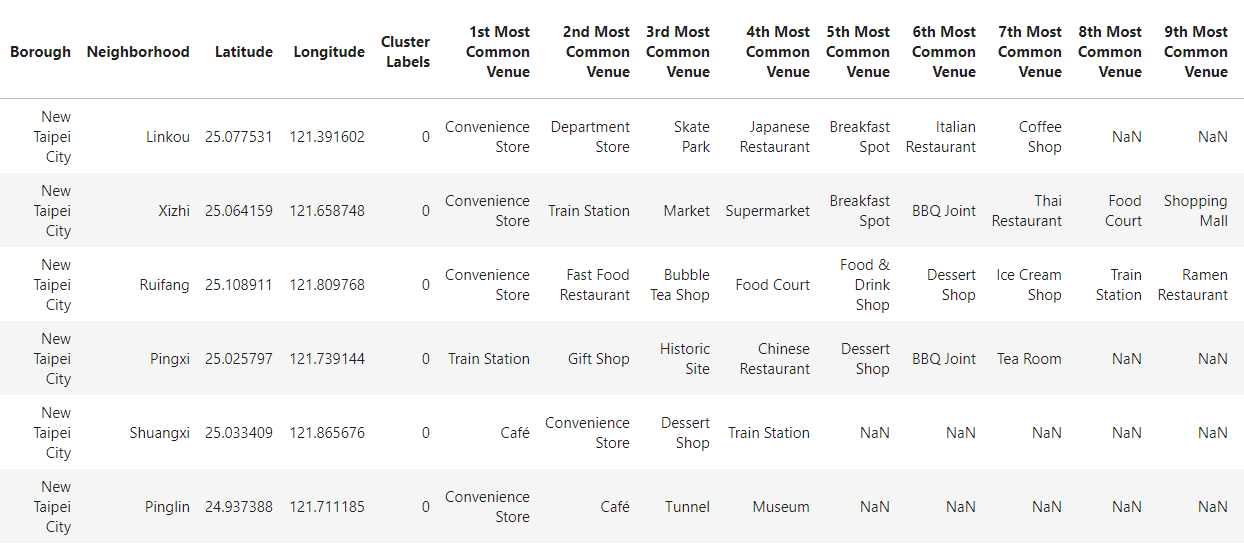


Table 1. List of neighborhoods in Cluster 0.

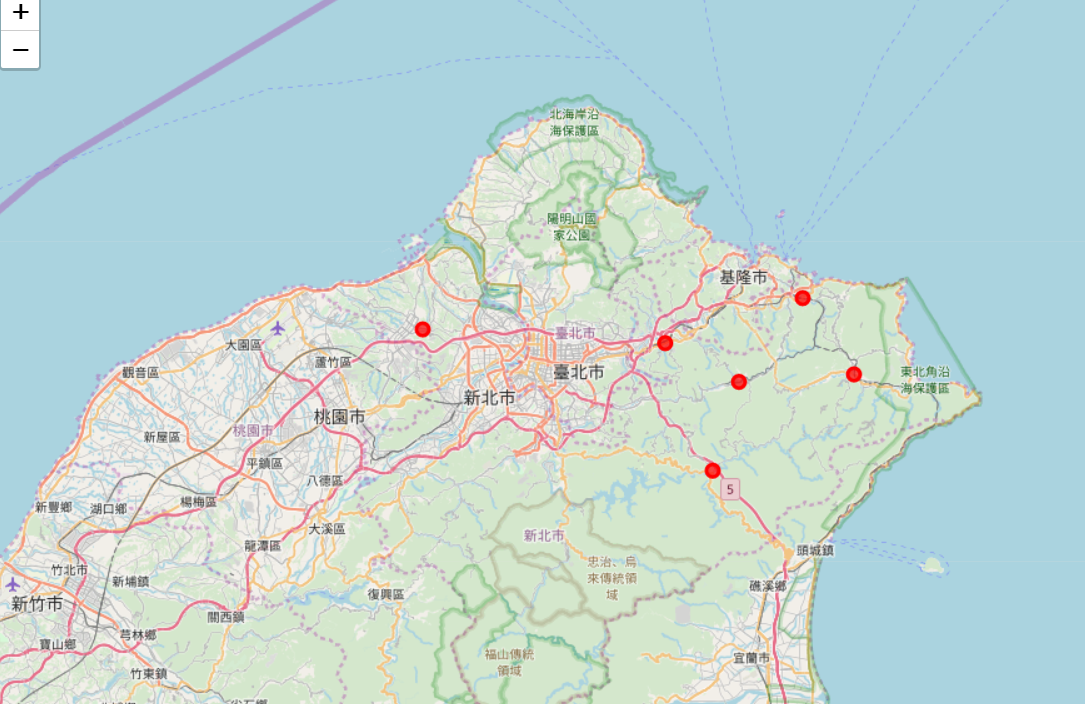


Fig. 3 Neighborhoods in Cluster 0.

**4.2 Cluster 1**

Cluster 1 has the heart of both cities. All Taipei neighborhoods and most New Taipei City neighborhoods close to Taipei are included. All these neighborhoods are similar to each other to have access to a variety of venues, including restaurants, hotels, convenience stores and department stores.







Table 2. List of neighborhoods in Cluster 1.

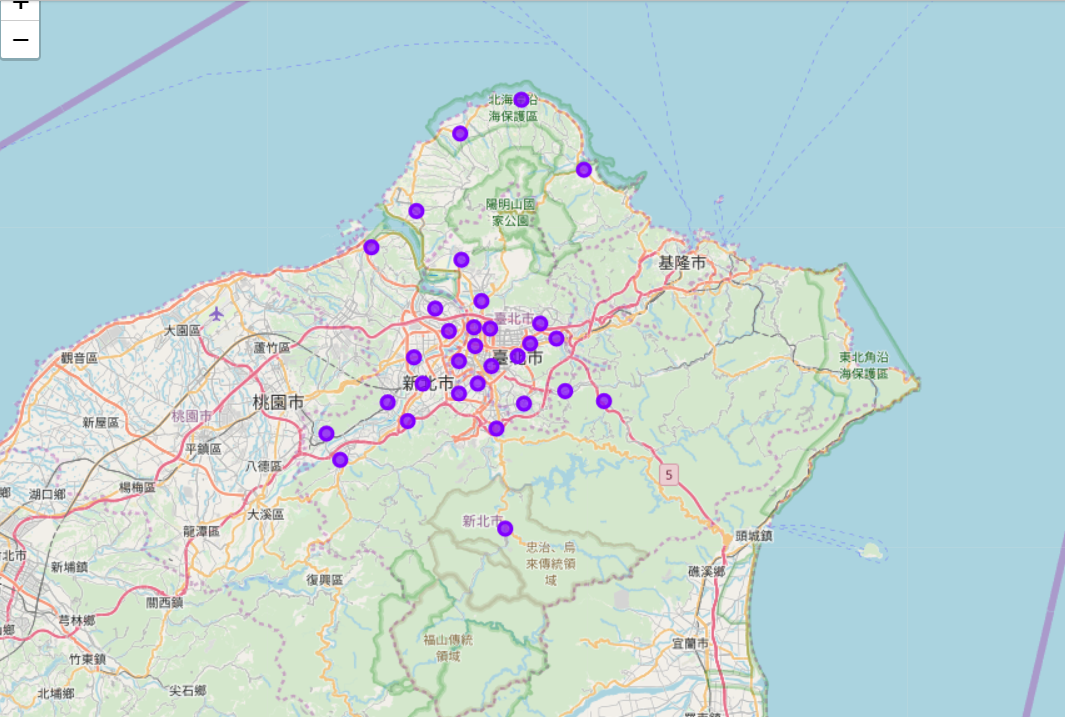
****

Fig. 4 Neighborhoods in Cluster 1.

**4.3 Cluster 2**

Cluster 2 contains only one far outside neighborhood of New Taipei City. The Foursquare only returned two categories (train station and campground) of venues for it.

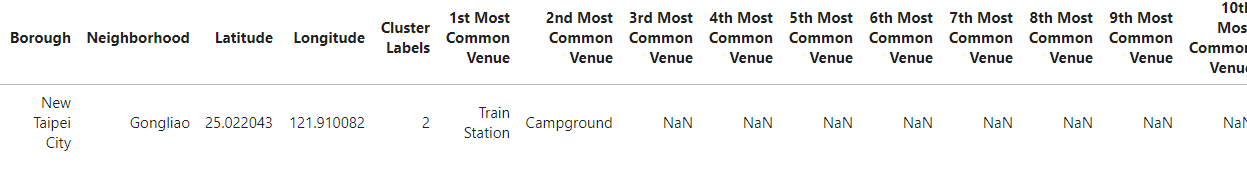
****

Table 3. List of neighborhoods in Cluster 2.

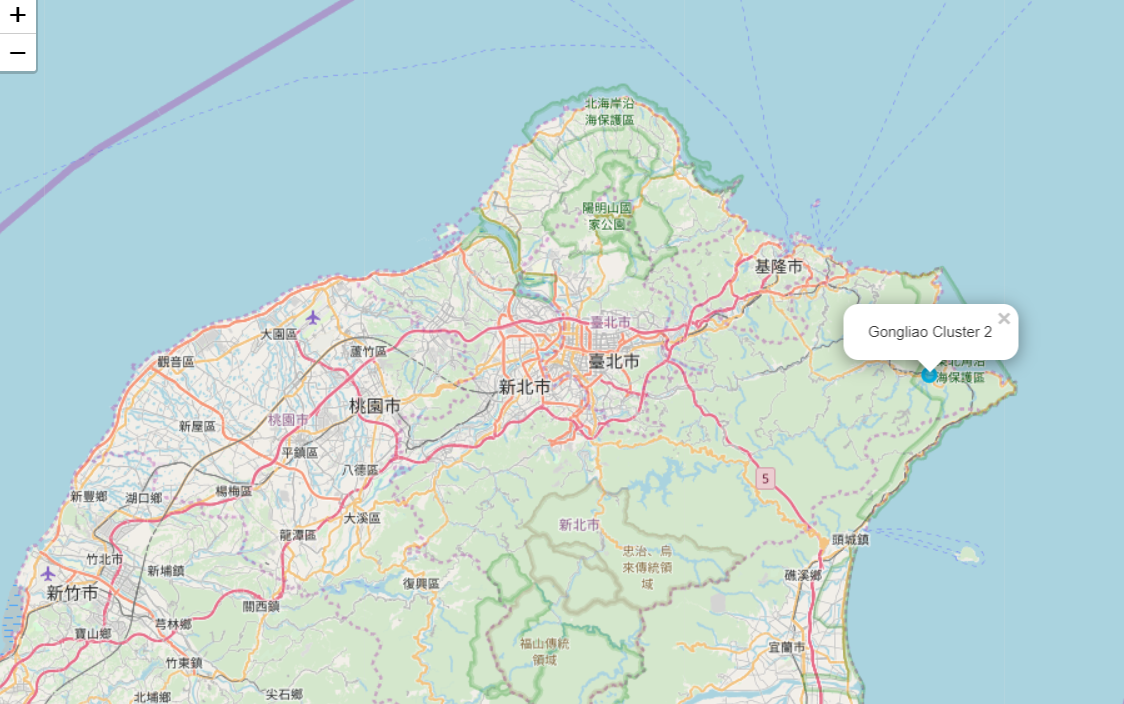
****

Fig. 5 Neighborhoods in Cluster 2.

**4.4 Cluster 3**

Cluster 3 includes two close to center mountain neighborhoods of New Taipei City. Both neighborhoods are similar to have top access to bakeries, convenience stores.

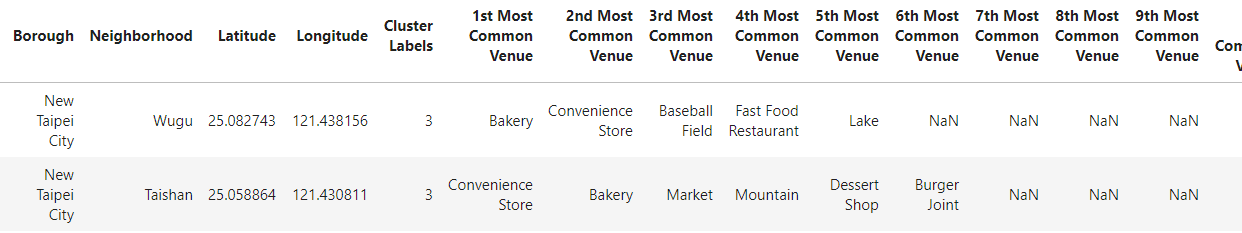
****

Table 4. List of neighborhoods in Cluster 3.

****

Fig. 6 Neighborhoods in Cluster 3.

**4.5 Cluster 4**

Cluster 4 includes only one neighborhood of New Taipei City. It is close to the seashore and has access to seafood restaurants, beaches, harbor/marina, and resorts etc.

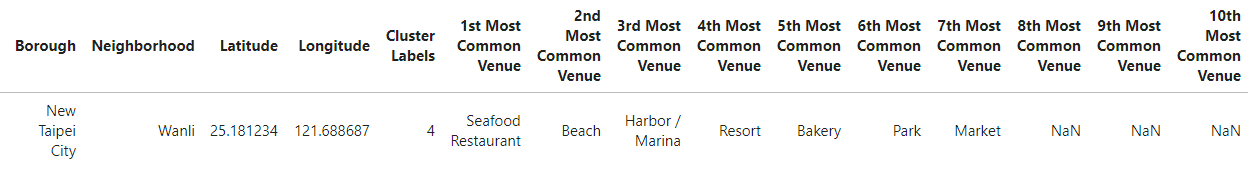


Table 5. List of neighborhoods in Cluster 4.

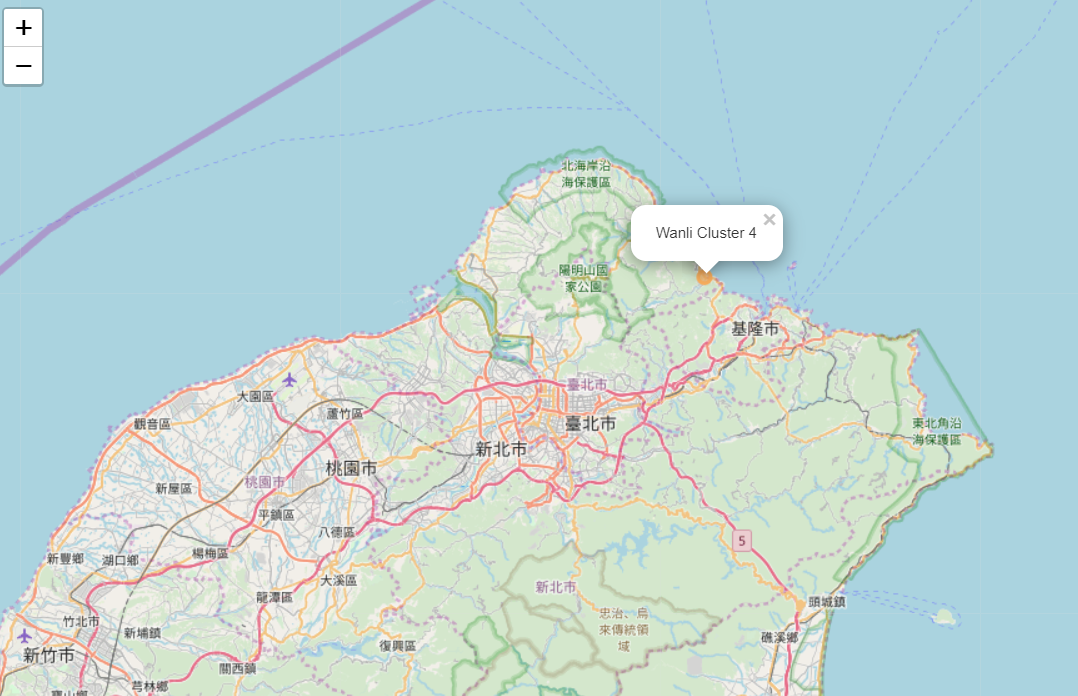


Fig. 7 Neighborhoods in Cluster 4.

### **5. Discussion**

Overall, the results are not surprising. All Taipei neighborhoods and the New Taipei City neighborhoods surrounding Taipei are much similar than other New Taipei City mountain neighborhoods.

For those blue-collar job seekers, neighborhoods in cluster 1 will be a good consideration. These neighborhoods located in city center, has a good choice of restaurants and supermarkets.

For a Japanese restaurant looking to expand its locations might perhaps select neighborhoods in cluster 0 because there are not many Japanese restaurants there in these districts.

**6. Conclusion**

I believe the limitation on the Foursquare API has its effect on the result of the neighborhood clustering. If more venues can be returned then the feature of each neighborhood will be highlighted and in return the effectiveness of the result will be improved.