Clustering of neighborhoods in Taipei city by using Foursquare data

I. Introduction/Business Problem

Taipei, located in Northern Taiwan, is the capital and a special municipality of Taiwan. The city proper is home to approximately 2,7 million people, which makes a population density of nearly 10,000 people per square kilometer[i].

Taiwan is my favorite country for spending my vacations due to a number of reasons: its subtropical climate, friendly and hospitable people, plenty of tourist attractions of various types, well-connected public transports, and irresistible Taiwanese cuisine. Being the political, economic, educational and cultural center of Taiwan, Taipei attracts millions of overseas visitors each year, making it the 15th most visited city globally, and the most of any city in the Chinese-speaking world[ii].

However, despite the large amount of travel guides and recommendations, it is not that easy for people who are new to the city to makes decisions on the best places to stay for their first visits. Depending on different purposes and duration, visitors may have various preferences and requirements, for instance, some needs easy access to major transportation centers for convenience, some prefers to stay in a quiet place far from the city center, some wants to be close to the commercial centers, while others prefer to stay in an area where one gets most authentic experience of local life etc. After I have been to Taipei many times, my friends and relatives have for several times asked me for advice regarding visiting the different parts of Taipei city.

To accommodate this need, I decided to create a map of Taipei that visualize the characteristics of different clusters of neighborhoods in Taipei city, which provides new visitors a quick overview of distribution of different areas in Taipei city.

II. Data Description

Considering the above stated problem, I will use the following data to create the target map.

- List of the administrative districts of Taipei City with postal codes – scraped from this Wiki page. The data is in the second table from this page.
- Use geopy to get location coordinates of Taipei city and its administrative districts listed above.

	Postal code	Neighborhood	Latitude	Longitude
0	100	Zhongzheng District	25.032361	121.518267
1	103	Datong District	25.065986	121.515514
2	104	Zhongshan District	25.064361	121.533468
3	105	Songshan District	25.049885	121.577272
4	106	Daan District	25.026515	121.534395
5	108	Wanhua District	25.031933	121.499332
6	110	Xinyi District	25.033345	121.566896
7	111	Shilin District	25.094118	121.524788
8	112	Beitou District	25.131931	121.498593
9	114	Neihu District	25.069664	121.588998
10	115	Nangang District	25.054578	121.606600
11	116	Wenshan District	24.989786	121.570458

Table 1: Administrative Districts of Taipei city

- Use Foursquare API to explore popular venues in each Taipei district. I'll be querying the top 200 venues in each district within a radius of 1000m radius as it is a reasonable walking distance. The result will be analyzed to get the top venue categories for each district, which then will be used to cluster the districts.

III. Methodology

GitHub repository is used as database for this study.

I used Python web scraping tool **Beautiful Soup** and geocoding tool **geopy** to create my master data (Table 1), which contains Postal code, Neighborhood (administrative district), Latitude and Longitude of Taipei city.

Then I used **Foursquare explore API** to query the top 200 venues within a radius of 1000 meters for each neighborhood. Below is a sample request:

```
GET
https://api.foursquare.com/v2/venues/explore?client_id={{}}&client_secret={{}}
&ll=25.032361,121.518267&v=20210123&radius=1000&limit=200
```

The response result is in JSON format and contains data of the top venues returned by Foursquare for the given location coordinates and radius. Sample response looks like the following:

```
{'meta': {'code': 200, 'requestId': '6012046af3028a465b29b662'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Zhōngzhèng Qū',
  'headerFullLocation': 'Zhōngzhèng Qū, Taipei',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 116,
  'suggestedBounds': {'ne': {'lat': 25.041361109000007,
    'lng': 121.52818148278706},
   'sw': {'lat': 25.02336109099999, 'lng': 121.50835251721293}},
  'groups': [{'type': 'Recommended Places',
   'name': 'recommended',
   'items': [{'reasons': {'count': 0,
      'items': [{'summary': 'This spot is popular',
        'type': 'general',
        'reasonName': 'globalInteractionReason'}]},
    'venue': {'id': '4b8e4566f964a520891d33e3',
     'name': 'Kinfen Braised Pork Rice (金峰魯肉飯)',
     'location': {'address': '羅斯福路一段10號',
      'lat': 25.03219410314086,
```

The same method is used to get the top 200 venues for all the neighborhoods of Taipei. The data is then merged with coordinates data into one table.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Zhongzheng District	25.032361	121.518267	Kinfen Braised Pork Rice (金峰魯肉飯)	25.032194	121.518534	Taiwanese Restaurant
1	Zhongzheng District	25.032361	121.518267	National Theater (國家戲劇院)	25.035197	121.518188	Theater
2	Zhongzheng District	25.032361	121.518267	虎記商行	25.031744	121.519284	Café
3	Zhongzheng District	25.032361	121.518267	樂田麵包屋 Gakuden Boulangerie	25.032757	121.517534	Bakery
4	Zhongzheng District	25.032361	121.518267	Chiang Kai-Shek Memorial Hall (中正紀念堂)	25.034555	121.521835	Monument / Landmark

From the results, I can see that a total of 886 venues were found by Foursquare for the 12 neighborhoods. Number of venues returned for each neighborhood is shown as follows:
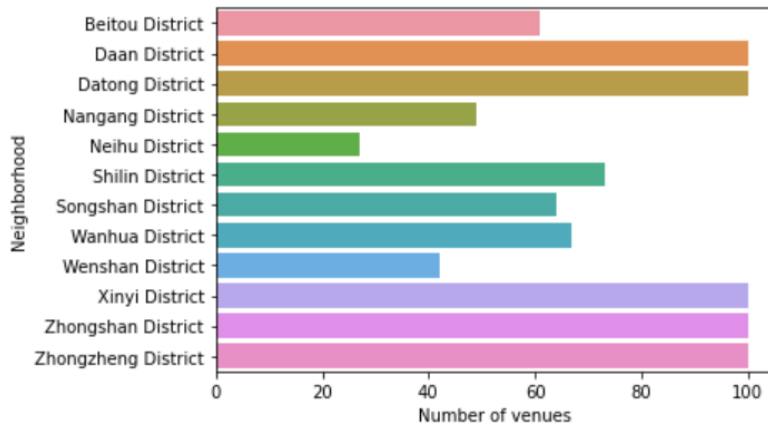
Figure 1: Number of venues for each neighborhood

From the graph above, we can see that districts of Daan, Datong, Shilin, Xinyi, Zhongshan and Zhongzheng get 100 recommended venues each, while Neihu district only gets under 30 venues. However, these returned venues do not cover all possible results for each of the neighborhoods, as the inquiry is based on a given combination of latitude and longitude for each neighborhood. Changing coordinates combination for the same neighborhood will impact on the results.

I applied one-hot encoding technique to indicate the existence/non-existence of each venue category for each neighborhood, and then I created another table to show the top 10 venue categories for each neighborhood, which has its first rows as follows:

	Neighborhood	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	Beitou District	Hotel	Noodle House	Convenience Store	Coffee Shop	Chinese Restaurant	Park	Asian Restaurant	Café	Metro Station	Taiwanese Restaurant
1	Daan District	Café	Taiwanese Restaurant	Noodle House	Tea Room	Chinese Restaurant	Ice Cream Shop	Dessert Shop	Bakery	Vegetarian / Vegan Restaurant	Dim Sum Restaurant
2	Datong District	Taiwanese Restaurant	Convenience Store	Dessert Shop	Café	Noodle House	Coffee Shop	Hotpot Restaurant	Hotel	Chinese Restaurant	Market
3	Nangang District	Coffee Shop	Convenience Store	Japanese Restaurant	Thai Restaurant	Hotpot Restaurant	Hotel	Shopping Mall	Train Station	Café	Fast Food Restaurant
4	Neihu District	Convenience Store	Chinese Restaurant	Japanese Restaurant	Asian Restaurant	Coffee Shop	Italian Restaurant	French Restaurant	Sporting Goods Shop	Duty-free Shop	Korean Restaurant

*Table 2: Top 10 venue categories for each neighborhood*

From Table 2, we can observe a common characteristic across all neighborhoods: a large portion of the top venue categories are different types of food/beverage catering, such as restaurants of special cuisines, cafes, tea houses, and ice-cream shops. Therefore, no matter where you are in Taipei city, there should be no difficulty of finding something delicious to feed yourself.

In order to differentiate the neighborhoods to cater for different needs of visitors who are new to Taipei city, I planned to use unsupervised learning **algorithm K-means** to cluster the neighborhoods.

To determine the best K value for clustering, I used **Elbow method** to find the best K. Using silhouette metric, i.e. the mean ratio of intra-cluster and nearest-cluster distance does not slow an elbow curve, the knee point detection algorithm calculates a likely optimal K value of 6 (Graph 2), however the score is quite low, which indicates all of these clustering configurations are not appropriate. This could mean that there might not be clusters in my dataset. I also plotted the inertia, i.e. the sum of squared distances of data points to their closest cluster center, then it gives an elbow curve with optimum K=2 (Graph 3).
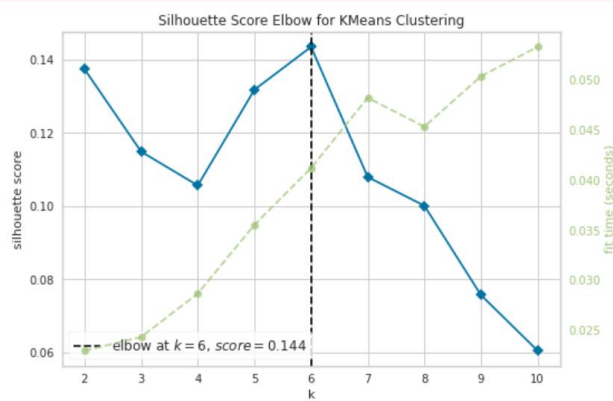
Graph 2: Elbow Method using Silhouette
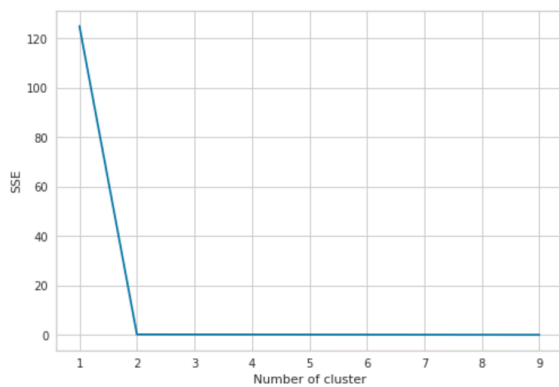
Graph 3: Elbow Method using SSE (Inertia)

The above analysis seems show that my dataset is not very clustered, meaning that the 12 different neighborhoods of Taipei city do not show significant dissimilarities among themselves. Actually, this conclusion confirms the previous observation that the top venues for all 12 neighborhoods seem to be dominated by food and beverage caterings.

However, I decided to make an attempt to cluster my dataset into 2 respectively 6 clusters and see if the resulting clustering make sense.

IV. Results

Table 3 shows the merged table with cluster labels for each neighborhood when clustered into 2 clusters.

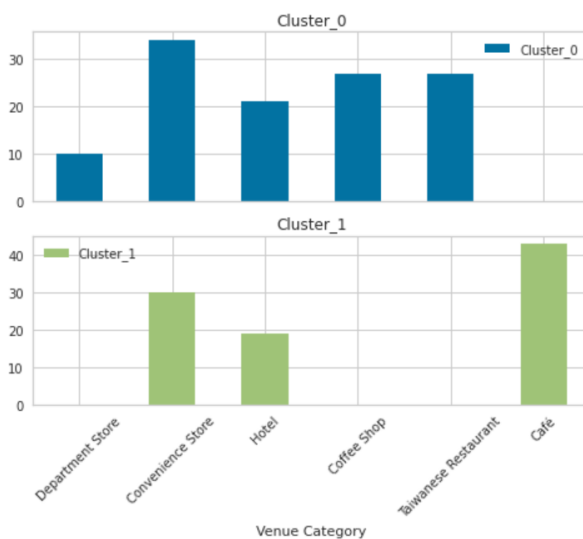|	Postal code	Neighborhood	Latitude	Longitude	Cluster Labels_2	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	100	Zhongzheng District	25.032361	121.518267	1	Café	Noodle House	Convenience Store	Coffee Shop	Japanese Restaurant	Breakfast Spot	Dumpling Restaurant	History Museum	Hotpot Restaurant
1	103	Datong District	25.065986	121.515514	0	Taiwanese Restaurant	Convenience Store	Dessert Shop	Chinese Restaurant	Coffee Shop	Café	Asian Restaurant	Park	Hotel
2	104	Zhongshan District	25.064361	121.533468	1	Hotel	Taiwanese Restaurant	Convenience Store	Chinese Restaurant	Café	Hotpot Restaurant	Japanese Restaurant	BBQ Joint	Asian Restaurant
3	105	Songshan District	25.049885	121.577272	1	Convenience Store	Chinese Restaurant	Breakfast Spot	Taiwanese Restaurant	Park	Café	Asian Restaurant	Coffee Shop	Seafood Restaurant
4	106	Daan District	25.026515	121.534395	1	Café	Noodle House	Coffee Shop	Taiwanese Restaurant	Tea Room	Ice Cream Shop	Chinese Restaurant	Dim Sum Restaurant	Dessert Shop

Table 3 Merged table with cluster labels (clustered into 2 clusters)

A visualization of the 2 clusters of neighborhoods of Taipei on a map is shown in Map 1 below.
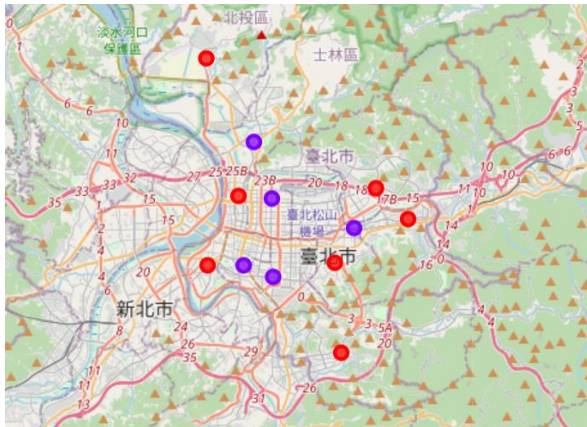
Map 1: Taipei districts clustered into 2 clusters

Then I did some data processing to get the number of the "1st Most Common Venue" in these 2 clusters, which is shown in Graph 4 below.


Graph 4: Number of 1st most common venues in each cluster

Combine the above graph and Map 1, we can label these 2 clusters as follows.

Cluster 0: Outer Ring Multifunctional Life Social Circles

Cluster 1: Inner Ring Café & Accommodation Areas

I also clustered the neighborhoods into 6 clusters, which are visualized on Map 2 below. Details of the neighborhoods in each cluster are shown in Table 4-9.
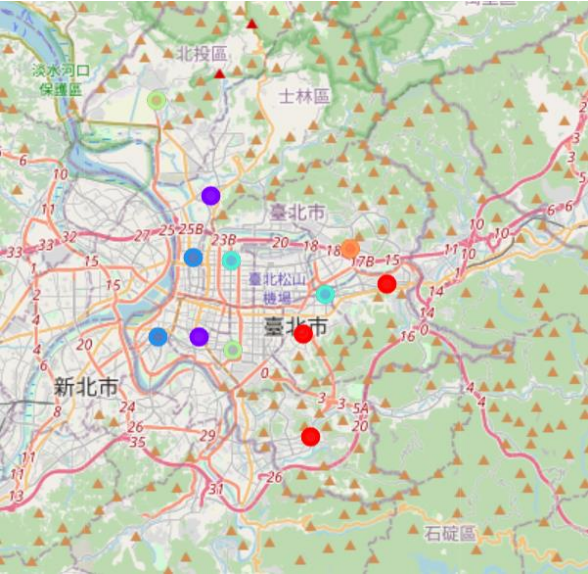
Map 2: Taipei districts clustered into 6 clusters

	Neighborhood	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	Xinyi District	Department Store	Hotel	Chinese Restaurant	Bar	Electronics Store	Café	Japanese Restaurant	Lounge	Coffee Shop	Noodle House
10	Nangang District	Convenience Store	Coffee Shop	Thai Restaurant	Japanese Restaurant	Fast Food Restaurant	Chinese Restaurant	Café	Shopping Mall	Breakfast Spot	Noodle House
11	Wenshan District	Coffee Shop	Convenience Store	Japanese Restaurant	Bus Station	Café	Pool	Cable Car	Chinese Restaurant	Sandwich Place	Sake Bar

Table 4 Cluster 0

	Neighborhood	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	Zhongzheng District	Café	Noodle House	Convenience Store	Coffee Shop	Japanese Restaurant	Breakfast Spot	Dumpling Restaurant	History Museum	Hotpot Restaurant	Taiwanese Restaurant
7	Shilin District	Café	Convenience Store	Hotel	Food Court	Bubble Tea Shop	Noodle House	Taiwanese Restaurant	Ice Cream Shop	Food Truck	Japanese Restaurant

Table 5 Cluster 1

	Neighborhood	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
1	Datong District	Taiwanese Restaurant	Convenience Store	Dessert Shop	Chinese Restaurant	Coffee Shop	Café	Asian Restaurant	Park	Hotel	Hotpot Restaurant
5	Wanhua District	Taiwanese Restaurant	Convenience Store	Chinese Restaurant	Dessert Shop	Night Market	Coffee Shop	Asian Restaurant	Café	Hotel	Bakery

Table 6 Cluster 2

	Neighborhood	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
2	Zhongshan District	Hotel	Taiwanese Restaurant	Convenience Store	Chinese Restaurant	Café	Hotpot Restaurant	Japanese Restaurant	BBQ Joint	Asian Restaurant	Seafood Restaurant
3	Songshan District	Convenience Store	Chinese Restaurant	Breakfast Spot	Taiwanese Restaurant	Park	Café	Asian Restaurant	Coffee Shop	Seafood Restaurant	Hotel

Table 7 Cluster 3

	Neighborhood	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	Daan District	Café	Noodle House	Coffee Shop	Taiwanese Restaurant	Tea Room	Ice Cream Shop	Chinese Restaurant	Dim Sum Restaurant	Dessert Shop	Bakery
8	Beitou District	Hotel	Café	Convenience Store	Coffee Shop	Noodle House	Park	Asian Restaurant	Dessert Shop	Metro Station	Hot Spring

Table 8 Cluster 4

	Neighborhood	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
9	Neihu District	Convenience Store	Asian Restaurant	Taiwanese Restaurant	Hotel	Japanese Restaurant	Cafeteria	Burger Joint	Coffee Shop	Sporting Goods Shop	Golf Course

Table 9 Cluster 5

By examining the above clusters, I can see some common characteristics of the neighborhoods that are clustered into same clusters, however, the differences among the clusters do not seem to be big enough. For instance, the top 2 most common venues for both Zhongzheng and Daan Districts are the same, i.e. Café and Noodle House, and the rest of the top venues are very similar to each other.

Since there are only 12 districts in total, which in general do not seem to be significantly different from each other, splitting them into 6 clusters does not make a good clustering.

V.   Discussion

In this study, I used Foursquare data to explore the popular venues in each district of Taipei and then made attempts to cluster the neighborhoods in order to provide some help to new visitors to the city.

Only 12 coordinates were used for query Foursquare data, which could have put some limitations on the dataset. More coordinates can be used for future explorations.

When using K means algorithm to cluster the neighborhoods, the elbow method gave me different best K values, based on different metrics used. Therefore I made two attempts of clustering with potential best K values of 2 and 6. By analyzing the clustering results, I came to the conclusion that splitting the neighborhoods into 2 clusters is a better clustering.

One problem of this study lies on the dependency on Foursquare data, which is updated on daily basis. I noticed that my dataset changed every time I acquired the Foursquare data and this has caused the clustering results were changing from time to time, especially when the top venue categories of each neighborhood turned out to essentially quite similar.

Apart from this, I realized that I could have acquired different and more data from Foursquare to improve my dataset so that the dissimilarities among the neighborhoods are better identified to make a better clustering. I used Foursquare's explore endpoint to get my data, which is dominated by "food" venues at subcategory level. Potential improvements can be done by acquiring number of Foursquare's 10 main venue categories (i.e. Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Professional & Other Places, Residence, Shop & Service, Travel & Transport) in each neighborhood, or using trending endpoint to get the most popular venues at certain time periods, etc.

VI.   Conclusion

According to this study using Foursquare data, activity density differs among Taipei's neighborhoods, as the number of top venues within the defined same radius for each neighborhood show some differences (see figure 1), while the lifestyles and local environment do not show significant differences among the 12 neighborhoods of Taipei city.

However, the neighborhoods can be roughly split into 2 clusters, namely the "Outer Ring Multifunctional Life Social Circles" where large number of various life, accommodation, and social venues are available, as well as the "Inner Ring Café & Accommodation Areas" where convenience stores, hotels and cafés are most popular.

For you, visitors new to Taipei city, wherever you choose to stay in the city, you will be able to have an all-around experience of Taipei in a slightly different way.

[i] https://worldpopulationreview.com/countries/taiwan-population
[ii] https://en.wikipedia.org/wiki/Taipei#Tourism