

Bubble tea shop and venue data analysis in Taipei City

Justin Wu

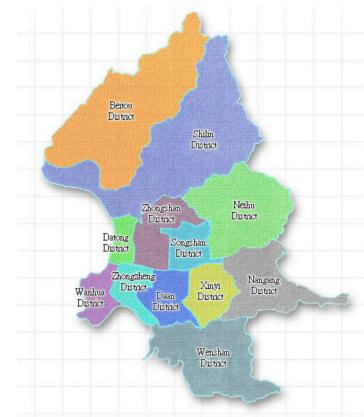
August 21, 2021

A. Introduction

- Business problem :

I live in Taipei City, which is the capital of Taiwan. In Taipei, there are 12 districts and the total area is around 270 km², which is also a home to an estimated population of 2.6 million. Taipei is a modern, energetic city and full of different cultures, so I choose Taipei as my topic of city to understand more detailed through this project.

- Bubble tea is a very popular drink in Taiwan. This project will recommend which district in Taipei is a smart choice for opening and managing a bubble tea shop. I would offer an analysis of the distribution of bubble tea shops in each district and do the clustering analysis with 5 features representing young people and crowd who are the bubble tea lovers. The discussion will also include the similarity between these 12 districts using Foursquare venue data to compare the districts.



B. Data description

Below are the data I used in this analysis :

- Taipei City Government website (<https://english.gov.taipei>) to obtain the name of districts.

- Use python library goopy.geocoders to get coordinates of the center of the 12 districts.
- Use Foursquare explore API to get the venue list for the 12 districts.
- Use Foursquare search API to get the information of the bubble tea shops in Taipei City.
- Use Foursquare search API to get the information of the 5 features(College, school, hotel, bus stop and metro) in Taipei City.

C. Methodology

- 12 districts in Taipei City :

From the Taipei City Government website, the name of the 12 districts in Taipei City are obtained, which are 1. Wanhua 2. Datong 3. Zhongzheng 4. Daan 5. Xinyi 6. Zhongshan 7. Songshan 8. Beitou 9. Shilin 10. Neihu 11. Nangang 12. Wenshan.

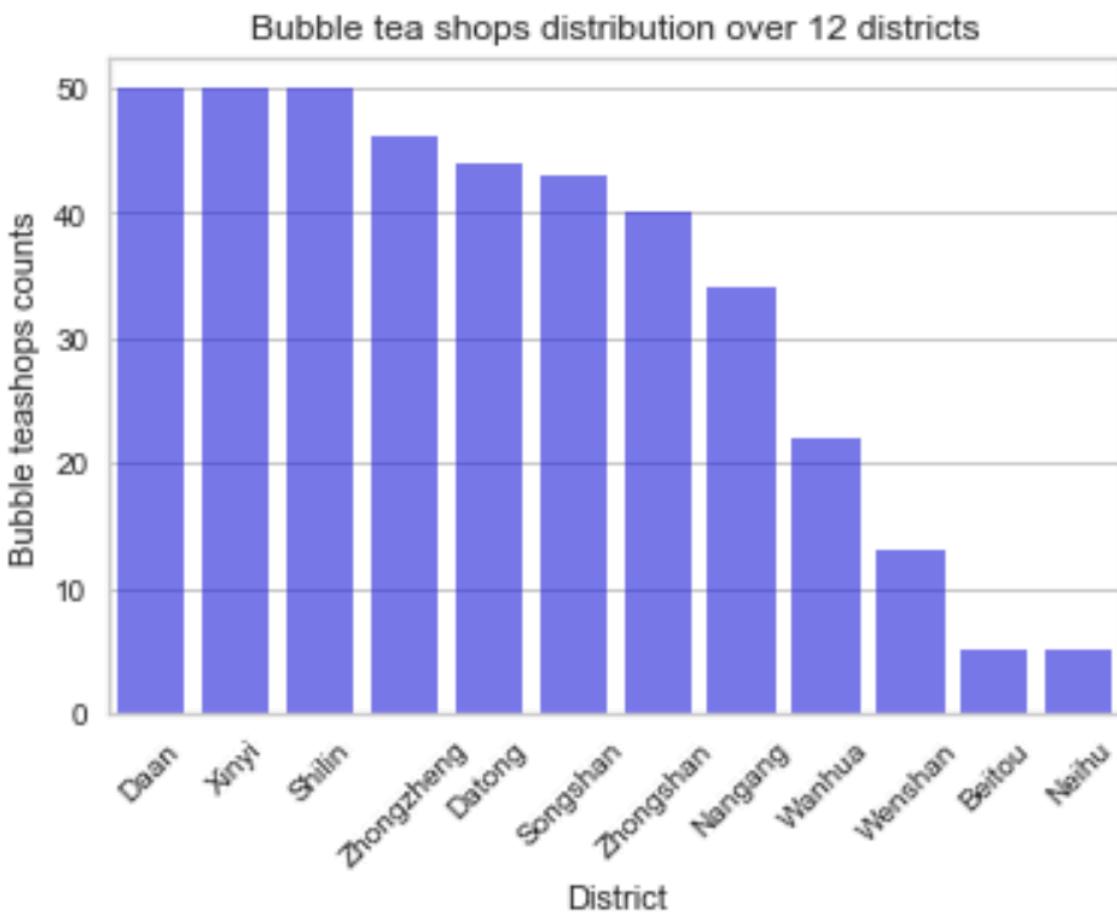
With these names, I am able to use Nominatim function to get the coordinate of the districts centers.

	District	lat	lng
0	Wanhua	25.031933	121.499332
1	Datong	25.065986	121.515514
2	Zhongzheng	25.032361	121.518267
3	Daan	25.026515	121.534395
4	Xinyi	25.033345	121.566896
5	Zhongshan	25.064361	121.533468
6	Songshan	25.049885	121.577272
7	Beitou	25.131931	121.498593
8	Shilin	25.091840	121.524207
9	Neihu	25.069664	121.588998
10	Nangang	25.054578	121.606600
11	Wenshan	24.989786	121.570458

- The current distribution of bubble tea shops in the 12 districts :

I am interested in the amount of bubble tea shops in each district since this information would give the potential districts with low density of bubble tea shops currently.

I use Foursquare API to search for the category of bubble tea shops in the circle region of radius of 1 km for each district and make the plot for the distribution over 12 districts and we can find the districts with less than 30 bubble teashops are Wanhua, Wenshan, Beitou and Neihu.

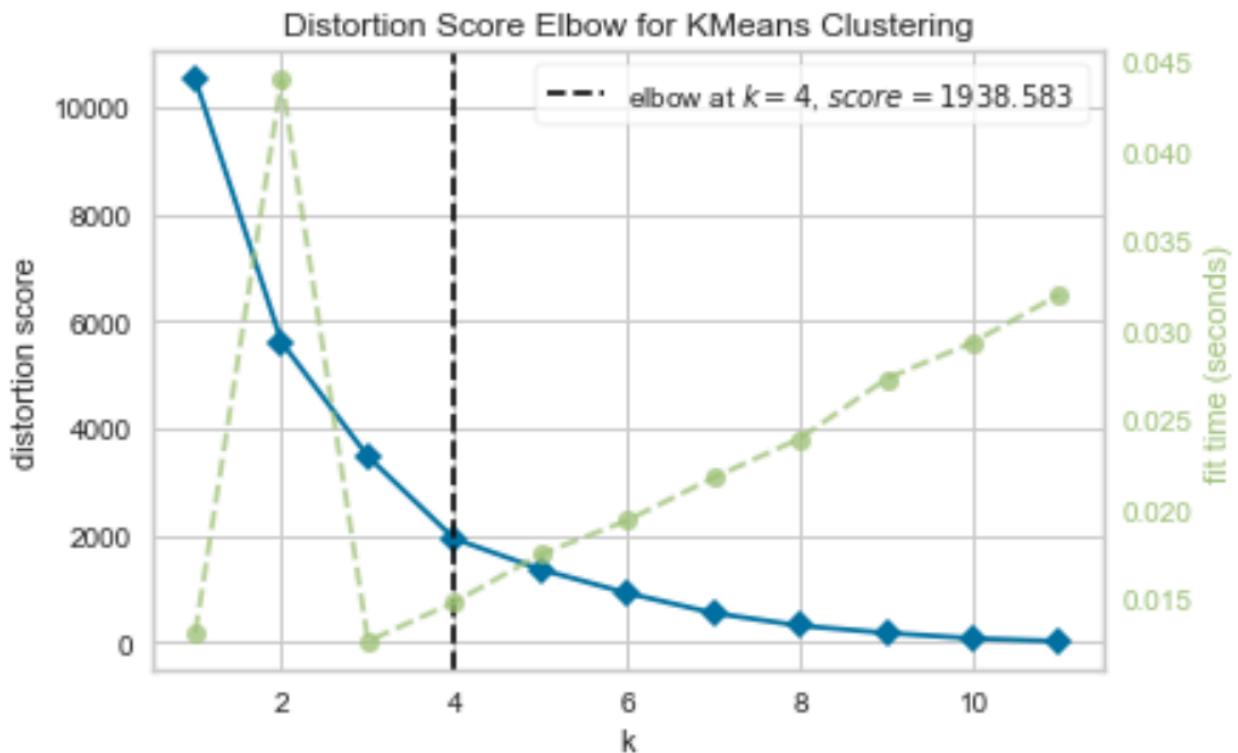


- The clustering analysis with 5 features representing young people and crowd:

Bubble tea is a very popular drink in Taipei, especially in young group and I also think the more convenient you approach to the tea shops the more often you would buy the drinks. Therefore I select

5 features representing young groups and crowd to execute clustering analysis. The 5 features are colleges, schools, bus stops, metro stations and hotels and their information of amount and location in each district are derived through Foursquare API searching by each category ID.

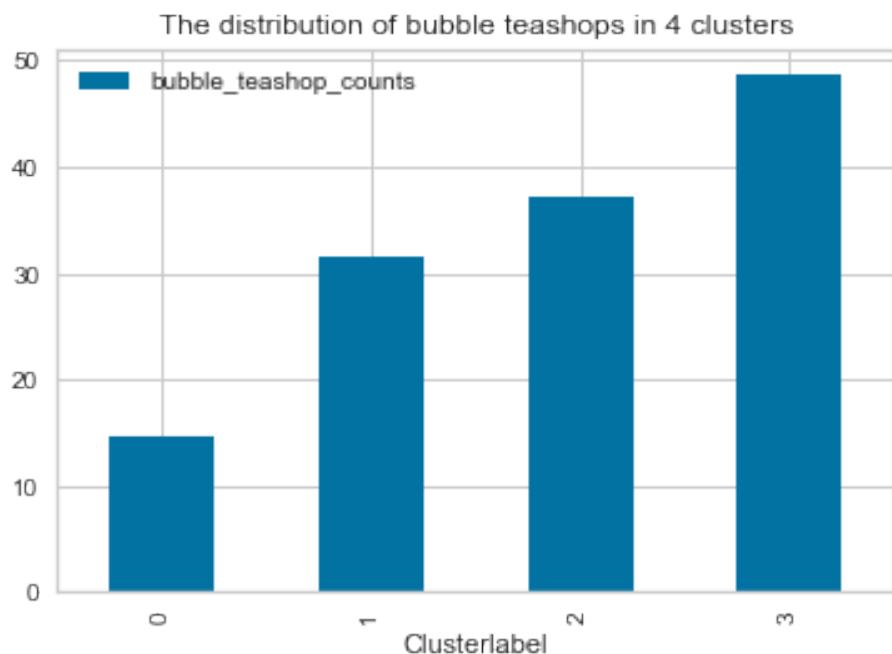
I use Kmeans clustering method to classify these 12 districts and the number of cluster is 4, which is decided using elbow method.



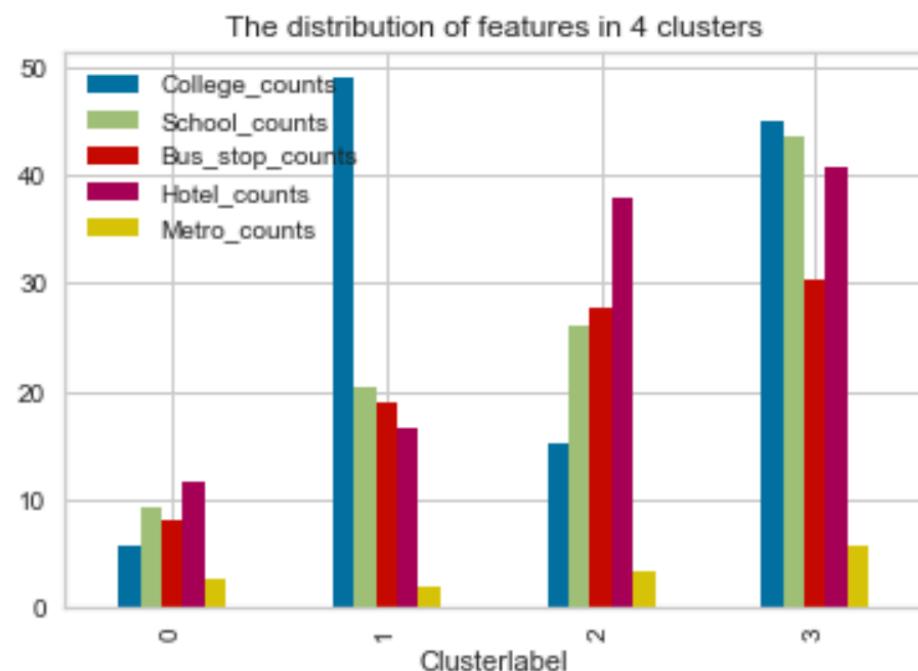
Clusterlabel	District	lat	lng	bubble_teashop_counts	College_counts	School_counts	Bus_stop_counts	Hotel_counts	Metro_counts
0	2	Wanhua	25.031933	121.499332	22	6	29	28	28
1	2	Datong	25.065986	121.515514	44	27	28	26	47
2	3	Zhongzheng	25.032361	121.518267	46	46	44	36	26
3	3	Daan	25.026515	121.534395	50	49	49	31	50
4	3	Xinyi	25.033345	121.566896	50	40	38	24	46
5	2	Zhongshan	25.064361	121.533468	40	17	18	24	47
6	2	Songshan	25.049885	121.577272	43	11	29	33	30
7	0	Beitou	25.131931	121.498593	5	7	9	7	24
8	1	Shilin	25.091840	121.524207	50	48	21	22	31
9	0	Neihu	25.069664	121.588998	5	5	12	6	2
10	0	Nangang	25.054578	121.606600	34	5	7	11	9
11	1	Wenshan	24.989786	121.570458	13	50	20	16	2

D. Results

There are 4 clusters for these 12 districts. To see how different these 4 clusters are, I plot the distribution of bubble tea shops and 5 features over 4 clusters.



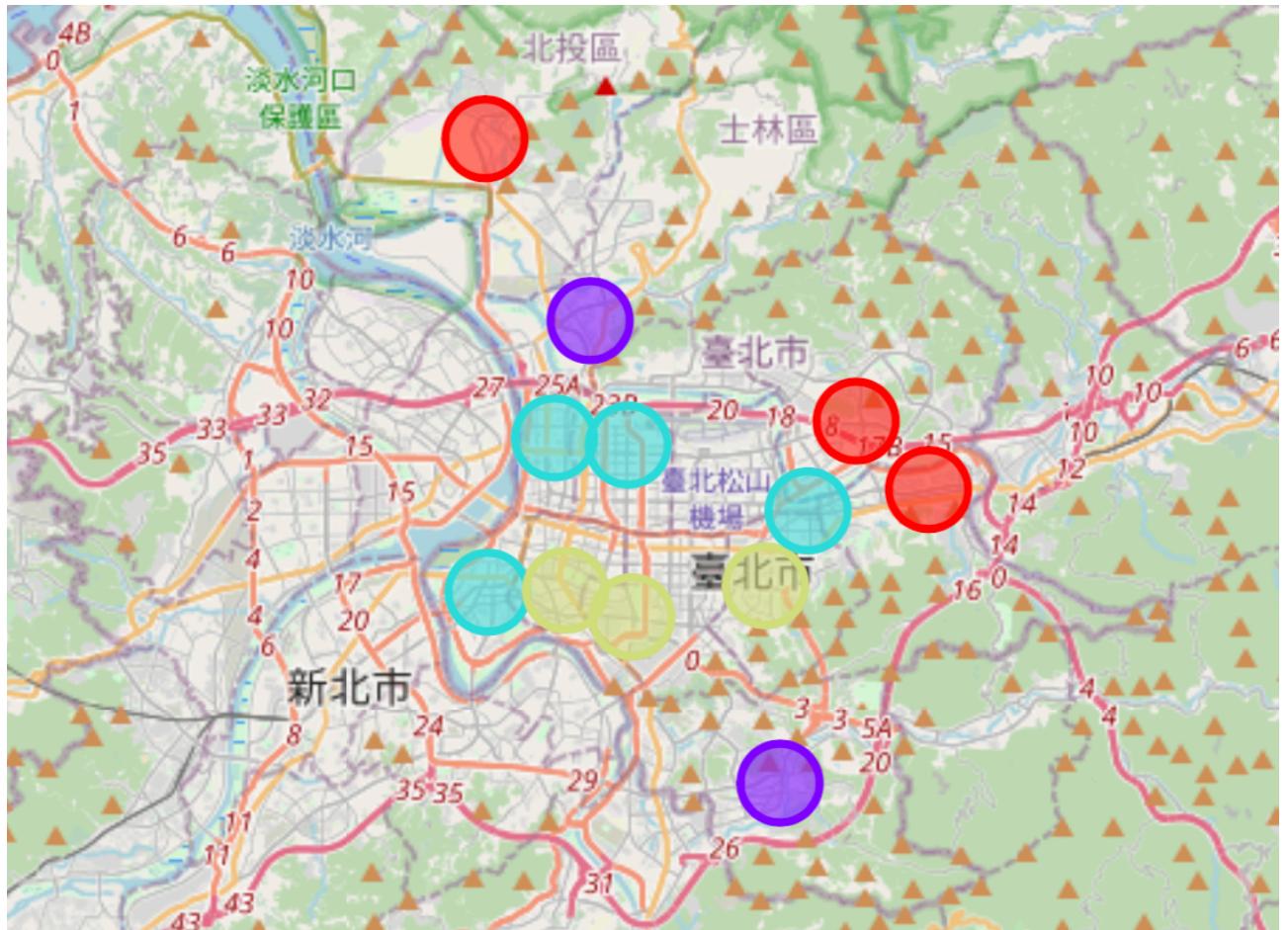
We can find the amount of bubble tea of cluster 0 is the fewest while the cluster 3 is the most. We can also find the distribution of 5



features in each cluster is showing the similar trend, cluster 0 has less amount of features while cluster 4 has many. Cluster 1 and cluster 2 are comparable and the number of college is the most in 4 clusters.

Compare these two distributions in 4 clusters, we can convince our hypothesis that the regions where the more crowd and younger people living the more bubble teashops locate.

The map shows the 12 districts clustering in 4 groups.



A. Discussion

From the distribution of the bubble tea shops in 12 districts, we know there are 4 districts with less than 30 shops, which are Wanhua, Wenshan, Beitou and Neihu. Since Beitou and Neihu are classified as cluster 0 with the 5 features clustering method, I think people here may not like bubble tea.

Wenshan district would be a good choice to open bubble teashops with classified as cluster 1. There are many college in Cluster 1, which means there are so many young people who love buying and drinking bubble tea.

Wanhua is a nice region to open more bubble tea shops since it is classified as cluster 2. In cluster 2 group, we can find Datong, Zhongshan and Songshan have many bubble teashops comparing to Wanhua. If the 5 selected features can be the index measuring how people love bubble tea shops in the region, Wanhua is a district having equal level as Datong, Zhongshan and Songshan. Therefore I recommend the managers to think twice opening bubble teashops in Wanhua.

I also analyze the similarity between these 12 districts by exploring all kinds of venues with Foursquare API and clustering

	name	latitude	longitude	Clusterlabel
0	Wanhua	25.031933	121.499332	1
1	Datong	25.065986	121.515514	1
2	Zhongzheng	25.032361	121.518267	2
3	Daan	25.026515	121.534395	2
4	Xinyi	25.033345	121.566896	4
5	Zhongshan	25.064361	121.533468	1
6	Songshan	25.049885	121.577272	3
7	Beitou	25.131931	121.498593	0
8	Shilin	25.091840	121.524207	2
9	Neihu	25.069664	121.588998	3
10	Nangang	25.054578	121.606600	5
11	Wenshan	24.989786	121.570458	2

analysis. Clustering analysis is executed with Kmeans with the number of clusters is 6 decided by elbow method.

The clustering result of all kinds of venues is different from what I get by only features, but it may be useful for managers to do further analysis.



E. Conclusion

This project is to analyze the relationship between bubble tea shops and the 5 features representing crowd and young people group. Through the clustering analysis, we can obtain the information of how similar the 12 districts in Taipei are and furthermore compare the density of bubble tea shops in the districts to recommend Wenshan district and Wanhua District are the two regions where the manager of bubble teashops may find the business opportunity to attract more people to buy bubble tea.