

Coffee and Bread in Taipei City

Introduction

I just finished my master degree in the middle of Taiwan. However, there are few jobs related to data science in where I'm live in, so I decide to move to the capital city of Taiwan, Taipei City. Living anywhere is fine to me, but the only one of my live rule is that I must have coffee or bread in my breakfast everyday. My classmates told me that there are enough coffee shop and bakery in Zhongzheng District in Taipei City, which totally can satisfy my habit. Thus, to find neighborhoods that similar to Zhongzheng District, I am going to use venues data related to coffee shop and bakery to cluster the district(neighborhood) in Taipei City and New Taipei City.

The target audience of this project is the coffee and bread lovers in Taipei City, and I think this problem can catch the eyes of whom are going to move to Taipei and also a coffee or bread lovers.

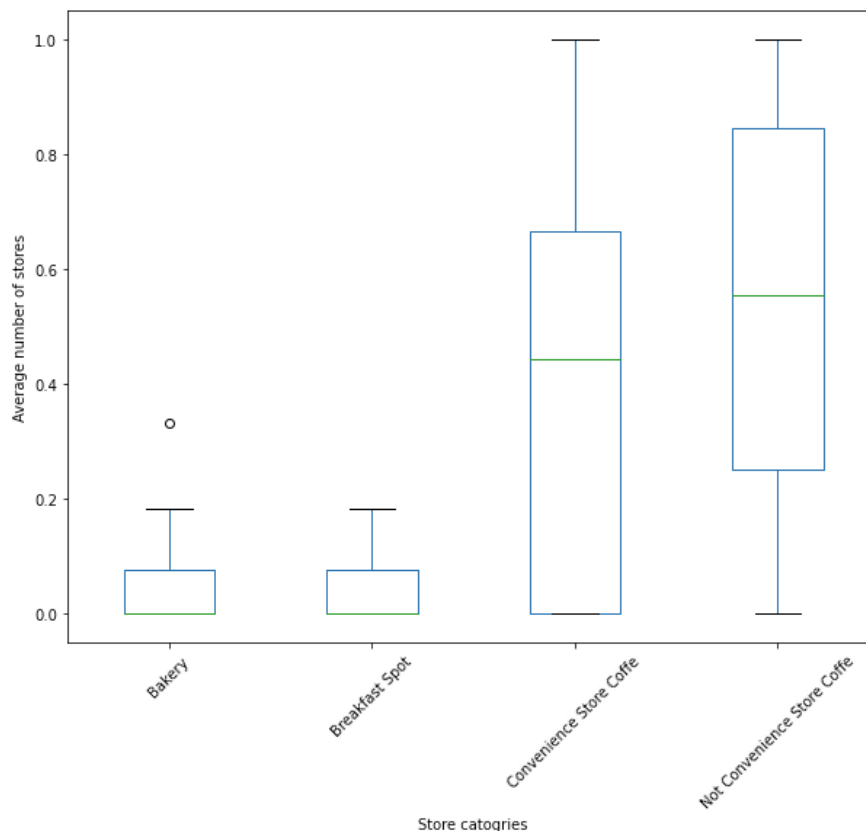
Data Description

Postal code data of Taipei city is scripted from wiki website, and I am already download a CSV data file about latitude and longitude from website. Besides, I will use foursquare api to get venues data of each neighborhood, and keep the venues related to coffee and bakery. By the way, you can get coffee and bakery in convenience store in Taiwan, so the convenience stores are also included.

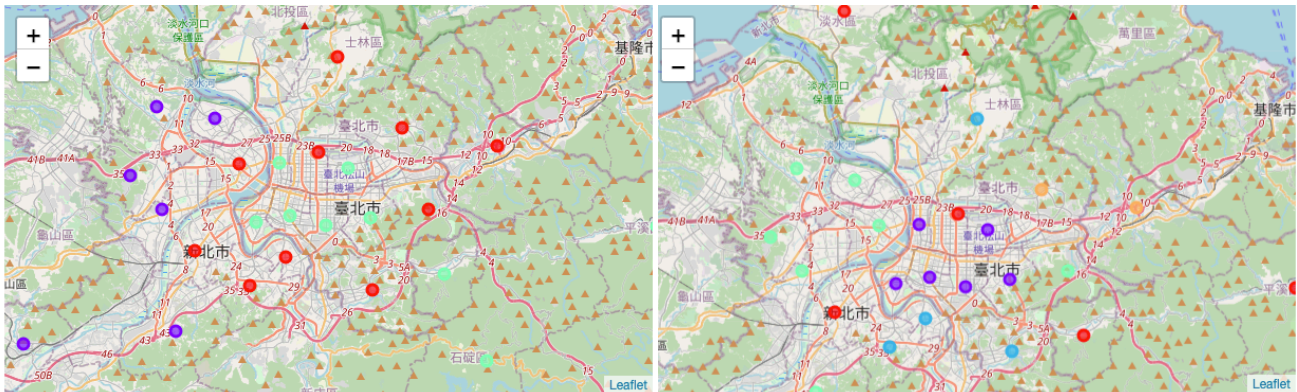
After merging the venues data and the location data, one-hot encoding method is applied on the venues data. By grouping the one-hot encoded data on neighborhoods, we can obtain proportion of stores related to coffee and bread in each neighborhoods. At this moment, I find that number of neighborhoods decrease. Because there is lack of venues data in Taiwan on the Foursquare website, some neighborhoods are dismiss from the current data.

Methodology

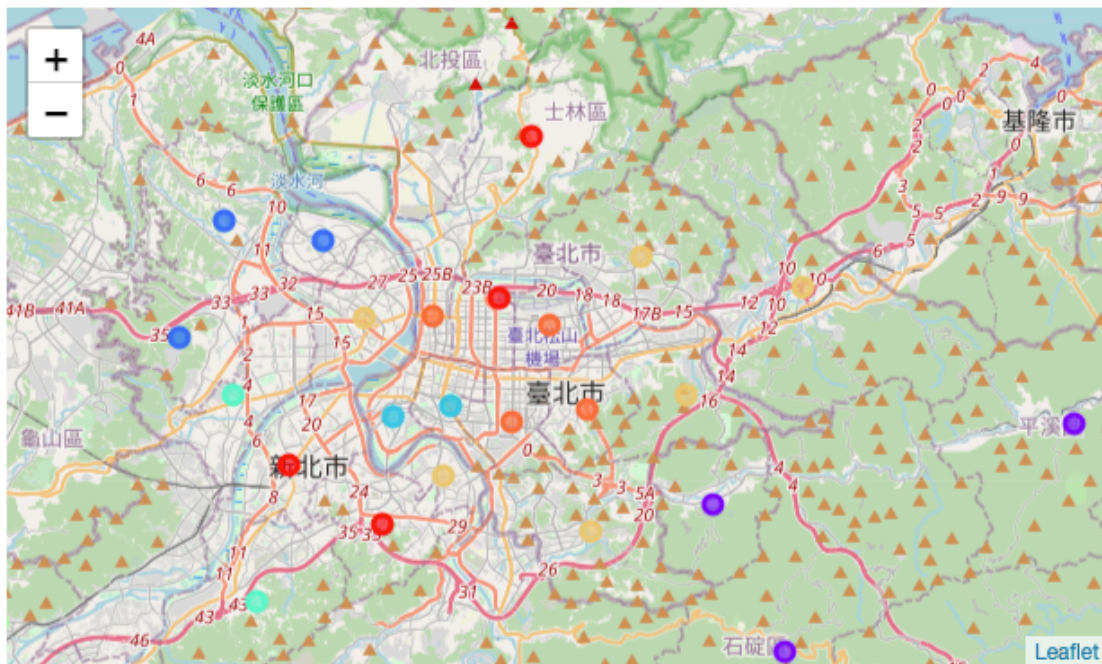
I also plot a box plot about the target venues data. The figure show that high variance on convenience store coffee and coffee shop coffee, so the average number of venues related to coffee is the main feature in this problem.



To cluster neighborhoods, I use K-means to analysis on the venues data. At first, number of K is setting to 3. From the map below, the cluster result only provides the information between downtown area and suburban area. In the second experiment, number of K is set to 5. Only the neighborhoods in the suburb area are assigned to different group obviously, but not for neighborhoods in the downtown area.



To divide the downtown area into more clusters, number of K increases to 8. In this try, not only neighborhoods in the suburban but also in the downtown are separated in the map, which is an acceptable clustering result.



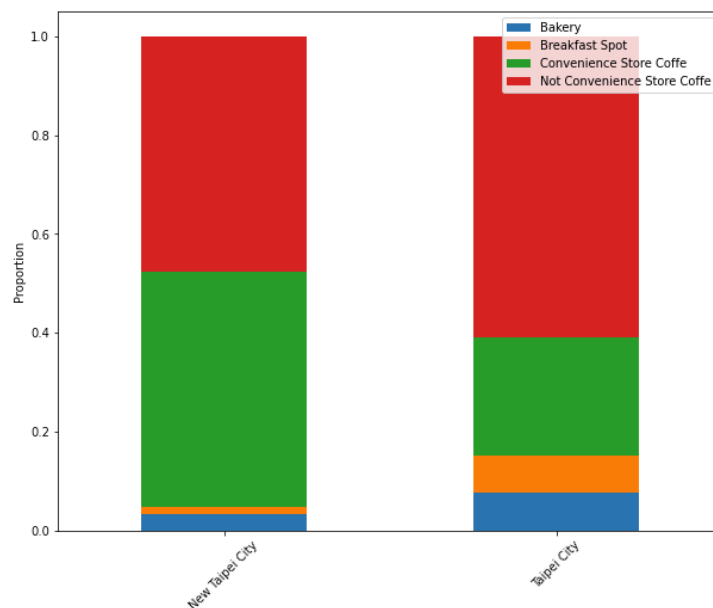
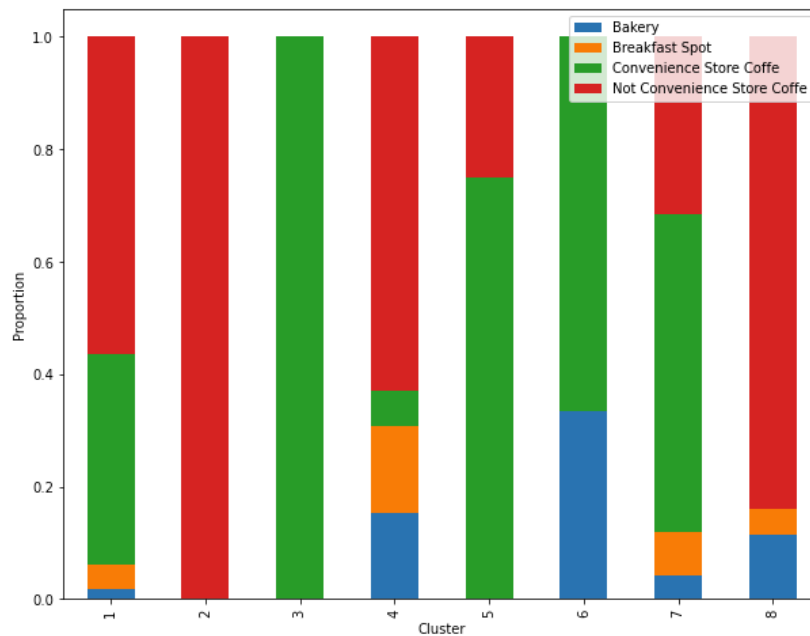
Results

From the map of 8 clusters, there is only one neighborhood similar to Zhongzheng District where recommend by my classmate, which is Wanhua District. The neighborhoods in each cluster are listed in the following table.

Cluster_No	Neighborhoods in English	Neighborhoods in Mandarin
1	Zhongshan District,Shilin District,Banqiao Dis...	中山區,士林區,板橋區,中和區
2	Shenkeng District,Shiding District,Ruifang Dis...	深坑區,石碇區,瑞芳區,平溪區,三峽區,淡水區

3	Jinshan District,Taishan District,Luzhou Distr...	金山區,泰山區,蘆洲區,五股區
4	Zhongzheng District,Wanhua District	中正區,萬華區
5	Tucheng District,Xinzhuang District	土城區,新莊區
6	Yingge District	鶯歌區
7	Neihu District,Nangang District,Wenshan Distri...	內湖區,南港區,文山區,汐止區,永和區,三重區
8	Datong District,Songshan District,Daan Distric...	大同區,松山區,大安區,信義區

Based on number of K equals to 8, the proportion of stores in each cluster is visualized into the bar-chart plot. Each bar-chart shows the proportion of venues related to bakeries, breakfast store, coffee and convenience store. Besides, the proportion of each store in different city is also visualized.



Discussion

Some observations are listed below.

1. If you really dislike cups of coffee from machines, you should not move in the neighborhoods of the 3 cluster.

2. There is no convenience store in the neighborhoods of the 3 and 8 cluster.

3. The habit of coffee lover must can be satisfied in the neighborhoods of the 1,2, 4 and 8 clusters, because the proportion of the coffee shops is more than 50% in these area.

It can be observed that the average number of coffee shop in Taipei City is more than in New Taipei City. On the other hand, there are more convenience store in New Taipei City. Therefore, if you want to enjoy the coffee in the after noon, there are more options for you in Taipei City.

Conclusion

Form this analysis, it can confirmed that the introduction from my classmate about Zhongzheng District is correct. Further, I find a similar neighborhood, Wanhua District, where exists different coffee shops and bakeries.

Reference

Postal code in Taiwan : https://en.wikipedia.org/wiki/Postal_codes_in_Taiwan

Latitude and longitude in Taiwan : <https://data.gov.tw/dataset/25489>

Foursquare : <https://developer.foursquare.com>