Finding Optimal Locations at Taipei & New Taipei City to Open Convenience Store

Sherryl Santoso

March 27, 2020

1. Introduction

1.1 Background & Problem

Convenience store has been the most happening business in Taiwan. How its located determine the opportunities to generate customer. Number of Existing stores play important role whether business able to get optimal profit or not. Moreover, number of population of people in specific area and its area size also determine strategic location to open this Convenience store. Therefore, it is advantageous for stakeholder to accurately predict which districts location that could generate optimal profit.

1.2 Objective

Stakeholder would highly interest in optimal locations to open the convenience store.

2. Data acquisition and cleaning

2.1 Data Sources

Based on definition of our problem, factors that will influence our decision are:

- Number of Population in each district
- Area Size for each district
- Sample Number of existing convenience store

Taipei and New Taipei City Districts lists, population and area size can be found in Wikipedia here.

Following data sources will be needed to extract/generate the required information:

- Taipei and New Taipei City Districts lists, population and area size can be found in Wikipedia here.
- Approximate latitude and longitude for each district extracted using geolocator
- Number of existing of convenience store are extracted using Foursquare API

2.2 Data Cleaning

We removed districts that belong to mountain, beach or rural area.

2.2 Data Engineering

Note that for number of existing convenience store, we only extract from radius 1km as sample of numbers. Next, we sum all the sample number of existing convenience store for each districts as one of our variable.

Finally, we normalize our three variables {Population, area size and sample number of convenience store} to remove its 'metrics'.

Below for Distribution range of each variable at before and after normalization:

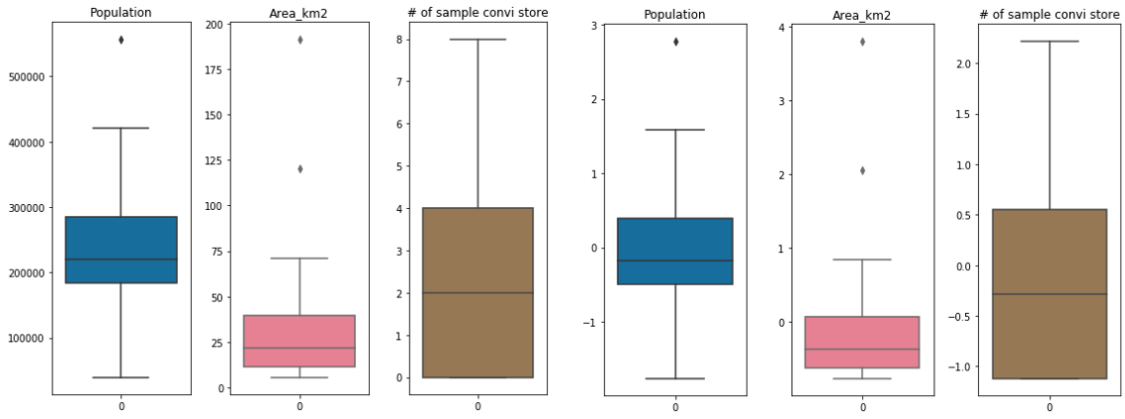

Figure 1 Before(left) VS After(Right) Normalization

3. Clustering Models

First, we will use k-means cluster as our clustering model. Then, use elbow method to determine best number of K (cluster) and we found that k=4 as our best number of clusters.

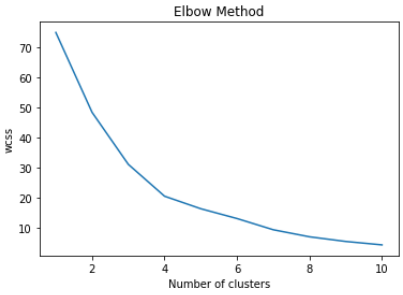

Figure 2 elbow method result

Next, we cluster all district using K-means cluster as below:

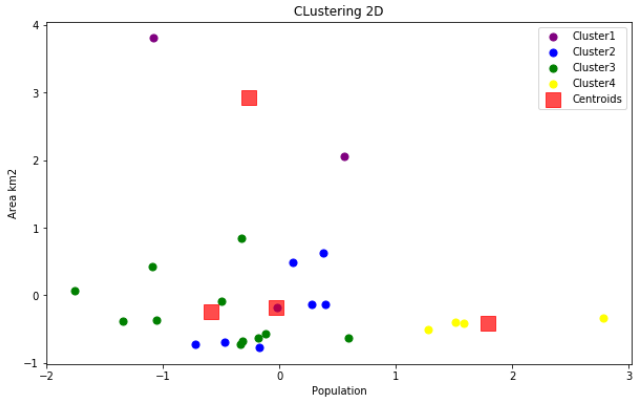

Figure 3 Clustering result in 2D

From above figure, we can see our districts divided into four different groups.

4. Analysis & Results

From figure 4, we can see that each cluster has different range of Population size, area size and sample number of existing Convenience store. Each cluster has characteristic as below:

- Cluster 1: lower mid of population, large area, lower mid number of existing convenience store
- Cluster 2: upper mid of population, upper mid area, big number of existing convenience store
- Cluster 3: small population, lower mid area, low number of existing convenience store
- Cluster 4: large population, small area, upper mid number of existing convenience store

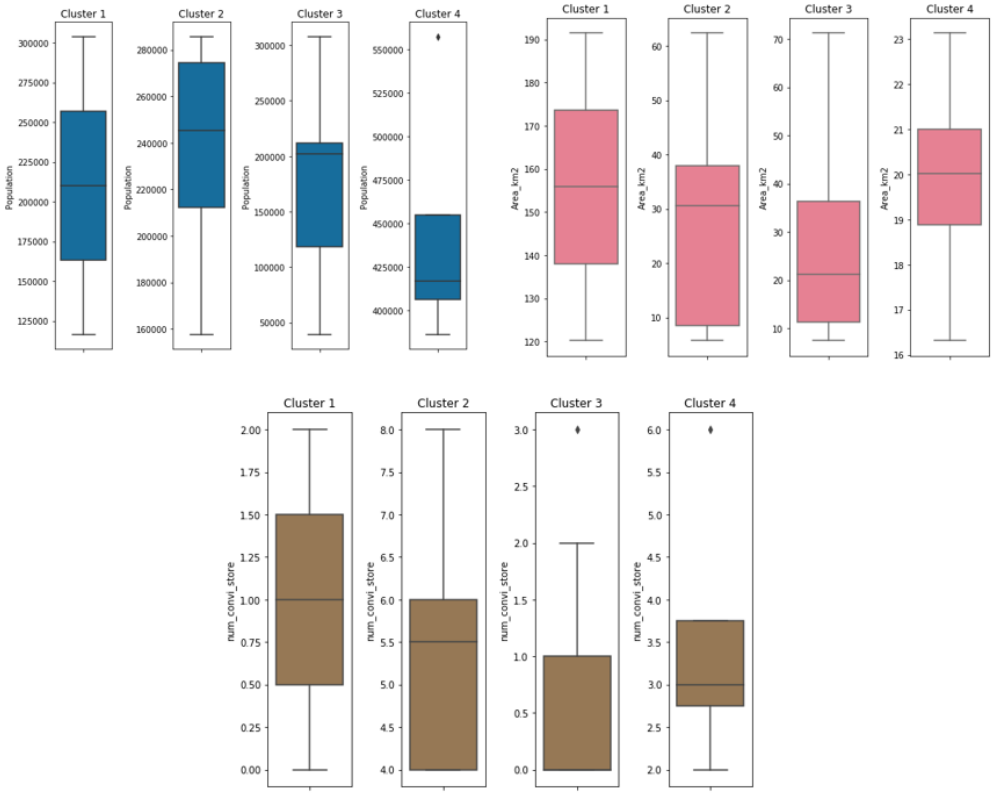

Figure 4 Population(top left), Area(top right) and Sample number Convenience Store (bottom) range of each cluster

Next, we will apply scoring method to get the best cluster. That case, we will give score 1-4 from low to big for area and population. While for number of convenience store, we will give opposite as 1-4 from large to small number of available convenience store.

Table 1 Score Table of clusters

Cluster	Population	Area	# of store	Total
1	2	4	2	8
2	3	3	4	10
3	1	2	1	4
4	4	1	3	8

From table 1, we can derive that districts belong to cluster 2 are the best districts to open new convenience store. Districts that belong to cluster 2 are Yonghe District, Tucheng District, Zhongzheng District, Wanhua District, Wenshan District, Neihu District, Shilin District, Beitou District.

5. Conclusion

Purpose of this project was to identify Optimal Districts from Taipei or New Taipei City to open new Convinience store. Based on Population numbers, Area Size and extracting Number of existng of convience store from foursquare, we applied normalization to removed 'metrics' units of thos criterias. Next, Clustering of those Criterias was then performed in order to create candidate of interest (containing best potential locations) starting points for final exploration by stakeholders.

Our proposed best districs are: Yonghe District, Tucheng District, Zhongzheng District, Wanhua District, Wenshan District, Neihu District, Shilin District, Beitou District.

Final decision on optimal districts location will be made by stakeholders based on specific characteristics in every recommended zone, taking into consideration additional factors like attractiveness of each location (other area facilities), levels of noise / proximity to major roads, real estate availability, etc.

6. Future Directions

For future research, we could add variables such as real estate availability, levels of noise / proximity to major roads and attractiveness of each location to get much accurate and narrower group of districts.