

Another Convenience Store in Taipei?

by Matthew Holland

Introduction

As the title suggests, convenience stores are quite ubiquitous in Taipei. Part of the reason for this is that gas stations are mainly just for pumping gas and not much else. Convenience stores in Taipei are hubs for consumers. You can pay bills, mail and receive packages, buy train tickets, sit and have a cup of coffee, make photocopies, buy food and drink, etc.. They are open 24 hours a day and can be found on every corner of the city.

Having said all this, there are always opportunities for businesses to open new convenience stores in neighborhoods that are lacking convenience stores or in neighborhoods where the convenience stores are getting old.

In this project, I will work with companies that run convenience stores to find the optimum district (or districts) in which to open a convenience store in Taipei. We will focus on districts that have increasing populations (based on Foursquare data), do not have convenience stores in their most frequent locations, and have high population density.

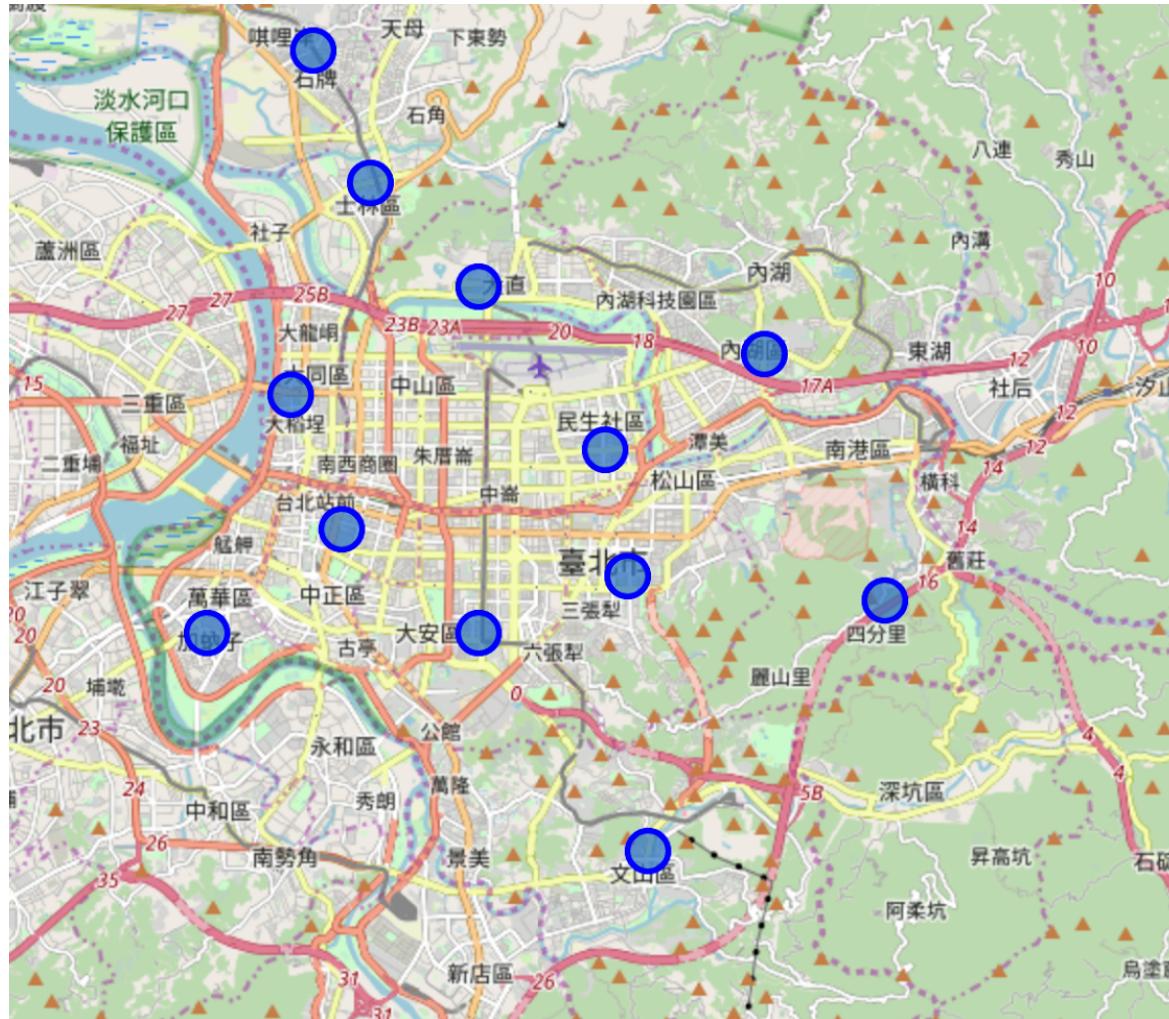
Data

I constructed a dataframe using statistical data from the Taipei City government. That information can be found in a PDF file [here](#). I pulled data for population, population density (person/sq. km), % population change (since 2008), and area (sq. km) for each of the twelve districts that make up Taipei City. Latitude and longitude were found by doing a Google search.

Here is what that dataframe looks like.

	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude
0	Beitou	254138	4473	1.756142	56.82	25.1152	121.5150
1	Daan	308722	27173	-1.633275	11.36	25.0262	121.5427
2	Datong	127086	22368	1.951818	5.68	25.0627	121.5113
3	Nangang	120897	5535	6.356007	21.84	25.0312	121.6112
4	Neihu	286834	9083	7.505772	31.58	25.0689	121.5909
5	Shilin	285017	4570	-0.366350	62.37	25.0950	121.5246
6	Songshan	205219	22096	2.321785	9.29	25.0542	121.5639
7	Wanhua	188225	21263	-1.122079	8.85	25.0263	121.4970
8	Wenshan	273040	8665	4.325632	31.51	24.9929	121.5713
9	Xinyi	221606	19773	-2.706239	11.21	25.0348	121.5677
10	Zhongshan	228285	16685	4.315462	13.68	25.0792	121.5427
11	Zhongzheng	158583	20847	-0.473211	7.61	25.0421	121.5199

Here is a map of Taipei City with the center of each of the twelve districts noted with a blue circle.



Next, I accessed data from [Foursquare](#) to find out more about venues in each of the twelve districts. I accessed the name of the venue, the coordinates of the venue, the category of each venue, and the distance of that venue from the center of the district. The first ten rows of that data from the Beitou District are below.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
0	25.1152	121.515	水龜伯古早味	25.116794	121.515918	Dessert Shop	200
1	25.1152	121.515	石牌夜市 Shipai Nightmarket	25.116622	121.516702	Night Market	233
2	25.1152	121.515	東方泰國小館	25.114670	121.515385	Thai Restaurant	70
3	25.1152	121.515	蕭記大餛飩	25.116001	121.517358	Chinese Restaurant	253
4	25.1152	121.515	台北市北投運動中心 Taipei Beitou Sports Center	25.116769	121.509748	Athletics & Sports	557
5	25.1152	121.515	露特西亞 Lutetia	25.114342	121.527354	Café	1248
6	25.1152	121.515	慶熹宮韓國料理	25.115802	121.518145	Korean Restaurant	323
7	25.1152	121.515	宋江餡餅粥	25.118556	121.526267	Chinese Restaurant	1195
8	25.1152	121.515	一品山西刀削麵之家	25.118909	121.528256	Chinese Restaurant	1398
9	25.1152	121.515	瓦城泰國料理 Thai Town Cuisine	25.118716	121.523570	Thai Restaurant	948

Methodology

In this project, we will work to detect districts in Taipei that have a low density of convenience stores and a high population density. If possible, these areas will also have an increasing population. The search radius for each district will be 2000m around the district center.

Above, we collected the necessary data from the Taipei City government and from Foursquare. This data was organized into a dataframe.

Next, we will create clusters within the districts using k-means clustering. We hope to identify a cluster with an increasing population, high population density, and a district with convenience stores outside of their top ten most popular venues.

Exploratory Data Analysis

To begin the exploratory data analysis, I used one hot encoding to make the data more readable for machine learning. After cleaning the data and finding which types of venues were the most popular in each district, I organized the data into the data frame below.

	District	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	Convenience Store	Chinese Restaurant	Coffee Shop	Café	Thai Restaurant	Breakfast Spot	Dumpling Restaurant	American Restaurant	Supermarket	Market
1	Daan	Café	Hotel	Bakery	Chinese Restaurant	Hotpot Restaurant	Japanese Restaurant	Coffee Shop	Massage Studio	Dumpling Restaurant	Noodle House
2	Datong	Hotel	Taiwanese Restaurant	Dessert Shop	Japanese Restaurant	Asian Restaurant	Coffee Shop	Hotpot Restaurant	Chinese Restaurant	Beer Bar	Café
3	Nangang	Convenience Store	Supermarket	Gym / Fitness Center	Market	Park	Café	Exhibit	Food & Drink Shop	Flower Shop	Flea Market
4	Neihu	Convenience Store	Japanese Restaurant	Café	Coffee Shop	Supermarket	Bakery	Fast Food Restaurant	Chinese Restaurant	Breakfast Spot	Italian Restaurant
5	Shilin	Café	Breakfast Spot	Convenience Store	Japanese Restaurant	Taiwanese Restaurant	Ice Cream Shop	Pizza Place	Snack Place	Department Store	Market
6	Songshan	Café	Noodle House	Hotpot Restaurant	Hotel	Dumpling Restaurant	Japanese Restaurant	Bookstore	Chinese Restaurant	Stadium	Comfort Food Restaurant
7	Wanhua	Taiwanese Restaurant	Park	Café	Noodle House	Coffee Shop	Hotel	Convenience Store	Chinese Restaurant	Japanese Restaurant	Hostel
8	Wenshan	Convenience Store	Exhibit	Café	Bus Station	Coffee Shop	Zoo Exhibit	Zoo	Cable Car	Italian Restaurant	Japanese Restaurant
9	Xinyi	Department Store	Chinese Restaurant	Hotel	Coffee Shop	Japanese Restaurant	Gym / Fitness Center	Café	Food Court	Cocktail Bar	Dessert Shop
10	Zhongshan	Hotel	Convenience Store	Café	Park	Japanese Restaurant	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant	Steakhouse
11	Zhongzheng	Hotel	Café	Noodle House	Coffee Shop	Taiwanese Restaurant	Bakery	Japanese Restaurant	Chinese Restaurant	Hostel	Monument / Landmark

A visual analysis of the data shows that four of the twelve districts have “convenience store” as their most common venue.

Machine Learning

I ran k-means clustering on the data about Taipei to see if I could find any groups of districts that would be ideal places to open a convenience store.

Here is the array with k = 3.

```
array([0, 1, 1, 2, 0, 1, 1, 1, 0, 1], dtype=int32)
```

As you can see, k = 3 doesn't cluster the data into meaningful groups. The third group has a single district.

Here is the array with k = 4.

```
array([1, 2, 2, 0, 3, 1, 2, 2, 3, 1], dtype=int32)
```

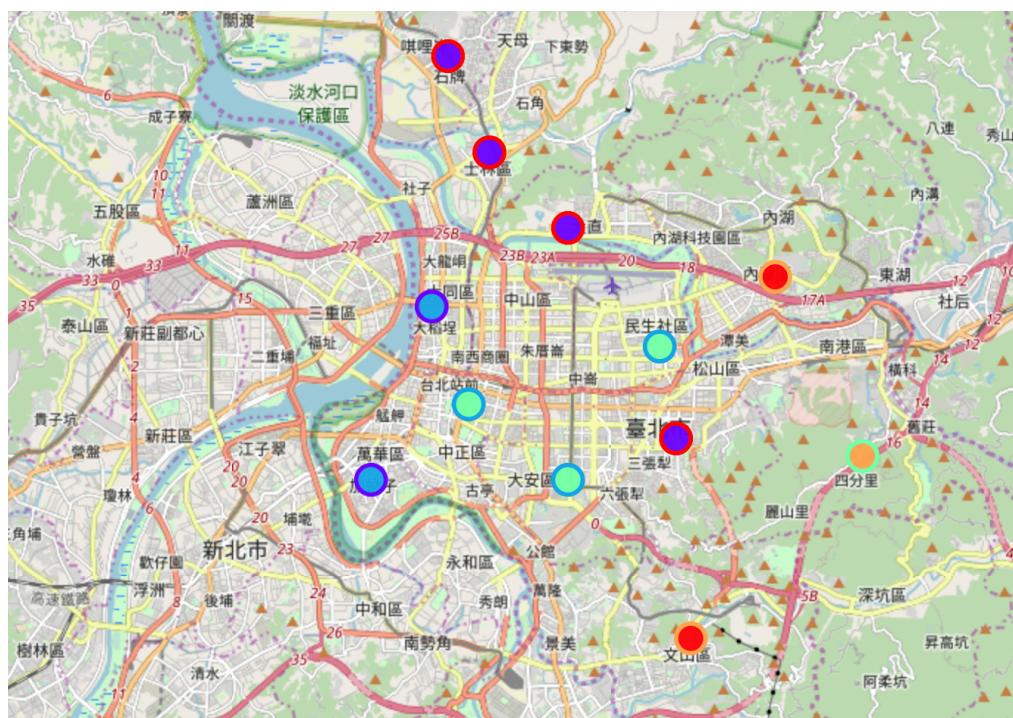
Getting better, but two of the clusters are still very small.

Here is the array with k = 5.

```
array([1, 3, 2, 4, 0, 1, 3, 2, 0, 1], dtype=int32)
```

Five clusters splits the data up into more distinct groups allowing us to examine them more closely.

Here is what the clusters look like on the map.



Here are the five clusters represented in data frames.

Cluster 0

	Population	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
4	286834	Neihu	286834	9083	7.505772	31.58	25.0689	121.5909	Convenience Store	Japanese Restaurant	Café	Coffee Shop	Supermarket	Bakery
8	273040	Wenshan	273040	8665	4.325632	31.51	24.9929	121.5713	Convenience Store	Exhibit	Café	Bus Station	Coffee Shop	Zoo Exhibit

Cluster 1

	Population	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	254138	Beitou	254138	4473	1.756142	56.82	25.1152	121.5150	Convenience Store	Chinese Restaurant	Coffee Shop	Café	Thai Restaurant
5	285017	Shilin	285017	4570	-0.366350	62.37	25.0950	121.5246	Café	Breakfast Spot	Convenience Store	Japanese Restaurant	Taiwanese Restaurant
9	221606	Xinyi	221606	19773	-2.706239	11.21	25.0348	121.5677	Department Store	Chinese Restaurant	Hotel	Coffee Shop	Japanese Restaurant
10	228285	Zhongshan	228285	16685	4.315462	13.68	25.0792	121.5427	Hotel	Convenience Store	Café	Park	Japanese Restaurant

Cluster 2

	Population	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
2	127086	Datong	127086	22368	1.951818	5.68	25.0627	121.5113	Hotel	Taiwanese Restaurant	Dessert Shop	Japanese Restaurant	Asian Restaurant	Coffee Shop
7	188225	Wanhua	188225	21263	-1.122079	8.85	25.0263	121.4970	Taiwanese Restaurant	Park	Café	Noodle House	Coffee Shop	Hotel

Cluster 3

	Population	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
1	308722	Daan	308722	27173	-1.633275	11.36	25.0262	121.5427	Café	Hotel	Bakery	Chinese Restaurant	Hotpot Restaurant	Japanese Restaurant
6	205219	Songshan	205219	22096	2.321785	9.29	25.0542	121.5639	Café	Noodle House	Hotpot Restaurant	Hotel	Dumpling Restaurant	Japanese Restaurant
11	158583	Zhongzheng	158583	20847	-0.473211	7.61	25.0421	121.5199	Hotel	Café	Noodle House	Coffee Shop	Taiwanese Restaurant	Bakery

Cluster 4

	Population	District	Population	Population Density	% Population Change	Area (sq. km)	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
3	120897	Nangang	120897	5535	6.356007	21.84	25.0312	121.6112	Convenience Store	Supermarket	Gym / Fitness Center	Market	Park	Café

Results

- **Cluster 0** contains large districts with low population density on the outskirts of Taipei City.
- **Cluster 1** contains districts that are mostly in the northern part of the city. The residents of these districts tend to be more affluent.
- **Cluster 2** contains districts on the western part of the city. These districts are in the oldest part of the city and they are densely populated.
- **Cluster 3** contains districts in the center of the city. These districts are smaller and more densely populated. The residents of these districts tend to be more affluent.
- **Cluster 4** is a single district, Nangang. It is one of the latest districts to be developed.

Discussion

A visual analysis of the clusters shows that cluster 3 has three districts with high population density. In addition, these districts do not have convenience stores in the top ten most common venues.

It would be my advice to target these areas for opening a new convenience store. Songshan District would be my top recommendation since it has an increasing population. My second recommendation would be Zhongzheng District and my third recommendation would be Daan District. Although these two districts have decreasing populations, they still share many characteristics with Songshan District.

Xinyi District would also be worth looking at as a possibility, although it doesn't fit into the cluster that the other three districts are in.

Conclusion

The purpose of this project was to identify districts in Taipei that would be good places to open a convenience store. By using population data from the Taipei City government, location data from Foursquare, and machine learning, we were able to find districts that would be ideal for opening a convenience store. This data will help to narrow down the area in which stakeholders will need to search in order to find a place to open a convenience store.