The Battle of Neighborhoods

A Look Around Universities in Taiwan

Brian L. Chen, Feb., 2021 - Source Code • Interactive Notebook • Blogpost • GitHub

1. Introduction

1.1 Background

Taiwan is known for its outstanding performance throughout this pandemic, and students from all over the world urge to know more about the unique island. Taiwan strictly performs its regulation and carries out a 14-day quarantine for people who are allowed to enter the island, including Taiwanese citizens.

Besides the restrictions, Taiwan is also known for its high population density (ranked 17th worldwide in 2019 according to statista), beautiful mountains, and convenience.

As a student currently studying in Taiwan, providing an insightful analysis to the local universities in terms of convenience would be interesting for future students when it comes to choosing which university to go to. Though it may be somewhat absurd only choosing universities based on their convenience, it is still something worth noticing, especially for those who haven't actually been to a specific university before.

Moreover, as a gym-goer and a sports fanatic, picking an university surrounded by sports facilities is always considerable.

1.2 Interest

The project is then aiming (1) **to examine the neighborhood of the top ten universities in Taiwan, exploring how it is like to live around their neighborhoods**, and (2) **to inspect the sports facilities around the top ten universities in Taiwan.**

Though at the discussion section, we only choose the top ten universities to inspect, the whole clustering and identification are made for universities listed from Ministry of Education, Taiwan.

The aim is to help future students choose their universities in terms on what their neighborhoods have to offer and what they would like to experience. The second part of this project could also help students playing sports gain a better perspective on which school to go to based on their preferences.

Figure 1. Taiwan is a beautiful place with beautiful people, and the project aims to provide insights regarding the universities on the island. Please do not add extra political perspective towards this project.

2. Data Acquisition and Cleaning

2.1 Data Sources

2.1.1 List of Accredited Taiwan Universities

The data of listed universities is from the Ministry of Education, which can be found here.

Since longitude and latitude for each school is not listed, **HERE API** is used to fetch the data needed to be further processed using Foursquare.

The final attributes include: *Serial Number, ID, Public/Private, System, Name, Eng. Name, Postal Code, Latitude, Longitude, Address, City, County, Neighborhood,* and *County (zh)*.

2.1.2 Geojson Data of Taiwan

The **geojson data** of Taiwan is obtained to plot the choropleth map. The data helps define the boundaries for each county in Taiwan. It is provided by **g0v**. Since the data hasn't been updated for nearly 6 years, some content from the dataset needs to be corrected.

I'm using this particular geojson data. As an example of the correction, 桃園縣 (Taoyuan County) needs to be corrected to 桃園市 (Taoyuan City).

2.1.3 Information of Venues Around Universities

Foursquare API is used to obtain the final list of venues to be further examined. Also note that Foursquare API uses special "category id" for each category. Since I am collecting a series of fitness/gym sites, scraping (instead of connecting through its API the venue category page works better for me.

The following information are collected: *University (zh), University (eng), University Latitude, University Longitude, Venue, Venue Latitude, Venue Longitude,* and *Venue Category*.

2.1.4 Results

Last but not least, there would be a total of **five** datasets to make analysis from:

1. Venues within 1,500m for each university on the list - to examine the convenience of each university
2. Sports venues within 1,000m and 2,000m
3. Gym venues only within 1,000m and 2,000m

For sports related venues, I am interested in the effects of two different distances, with 1,000 meters and 2,000 meters respectively. The reason behind this decision is because of 1 kilometer is the distance you can go on foot easily, without losing your desire/passion in exercising while 2 kilometer is a good distance for biking, which takes less than 10 minutes.

From where I study, I could bike for 10 minutes and not getting bothered. It is a good warm-up exercise as well. So with that in mind, let's just assume that the result for 1,000-meter clustering would be further interpreted using a pedestrian's perspective, while the other can be seen from a biker's perspective.

2.2 Data Processing

Please visit the `Data Processing` notebook for fully detailed data processing steps. For determining the convenience level of each university, the data is obtained by collecting all possible venues provided by Foursquare API. For determining sports and gym accessibility, we would run through each university in the list again, but only collecting the selected categories.

I perform a bit of data cleansing. After attaining and combining the data from the three main sources, the following result is to be seen.

Uni Venues (1,500 m):

	University (zh)	University (eng)	University Latitude	University Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	國立政治大學	National Chengchi Universerity	24.98741	121.57651	滇味廚房	24.987366	121.577474	Chinese Restaurant
1	國立政治大學	National Chengchi Universerity	24.98741	121.57651	道南館	24.988596	121.573602	Coffee Shop
2	國立政治大學	National Chengchi Universerity	24.98741	121.57651	JASONS Market Place	24.988345	121.576373	Supermarket
3	國立政治大學	National Chengchi Universerity	24.98741	121.57651	咖啡大亨磨坊 DICTY CAFÉ	24.987489	121.576419	Café
4	國立政治大學	National Chengchi Universerity	24.98741	121.57651	Starbucks (星巴克)	24.987716	121.575899	Coffee Shop

Uni Sports Venues (1,000 m):

	University (zh)	University (eng)	University Latitude	University Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	國立政治大學	National Chengchi Universerity	24.98741	121.57651	Roman Square (羅馬廣場)	24.986799	121.574540	Plaza
1	國立政治大學	National Chengchi Universerity	24.98741	121.57651	道南右岸河濱公園	24.995167	121.572406	Park
2	國立政治大學	National Chengchi Universerity	24.98741	121.57651	政大堤外球場	24.984681	121.572953	Baseball Field
3	國立政治大學	National Chengchi Universerity	24.98741	121.57651	政大網球場(山上) Tennis Courts	24.981970	121.578353	College Tennis Court
4	國立清華大學	National Tsing Hua University	24.79635	120.99686	清華大學新體育館 NTHU Stadium	24.793868	120.992318	College Stadium

Uni Sports Venues (2,000 m):

	University (zh)	University (eng)	University Latitude	University Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	國立政治大學	National Chengchi Universerity	24.98741	121.57651	臺北市文山運動中心 Taipei Wenshan Sports Center	24.996909	121.559737	Gym
1	國立政治大學	National Chengchi Universerity	24.98741	121.57651	Roman Square (羅馬廣場)	24.986799	121.574540	Plaza
2	國立政治大學	National Chengchi Universerity	24.98741	121.57651	道南右岸河濱公園	24.995167	121.572406	Park
3	國立政治大學	National Chengchi Universerity	24.98741	121.57651	萬芳公園 Wanfang Park	25.000369	121.567780	Park
4	國立政治大學	National Chengchi Universerity	24.98741	121.57651	政大堤外球場	24.984681	121.572953	Baseball Field

Uni Gym Venues (1,000 m):

	University (zh)	University (eng)	University Latitude	University Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	國立清華大學	National Tsing Hua University	24.79635	120.99686	清華大學新體育館 NTHU Stadium	24.793868	120.992318	College Stadium
1	國立清華大學	National Tsing Hua University	24.79635	120.99686	NCTU Gym	24.789599	120.995678	College Gym
2	國立清華大學	National Tsing Hua University	24.79635	120.99686	NCTU Swimming Pool	24.789435	120.995600	Gym Pool
3	國立臺灣大學	NATIONAL TAIWAN UNIVERSITY	25.01697	121.53369	World Gym	25.014470	121.534600	Gym / Fitness Center
4	國立臺灣大學	NATIONAL TAIWAN UNIVERSITY	25.01697	121.53369	台灣大學舊體育館	25.019733	121.535821	College Gym

Uni Gym Venues (2,000 m):

	University (zh)	University (eng)	University Latitude	University Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	國立政治大學	National Chengchi Universerity	24.98741	121.57651	臺北市文山運動中心 Taipei Wenshan Sports Center	24.996909	121.559737	Gym
1	國立清華大學	National Tsing Hua University	24.79635	120.99686	清華大學新體育館 NTHU Stadium	24.793868	120.992318	College Stadium
2	國立清華大學	National Tsing Hua University	24.79635	120.99686	NCTU Gym	24.789599	120.995678	College Gym
3	國立清華大學	National Tsing Hua University	24.79635	120.99686	NCTU Swimming Pool	24.789435	120.995600	Gym Pool
4	國立臺灣大學	NATIONAL TAIWAN UNIVERSITY	25.01697	121.53369	Taipei DaAn Sports Center (臺北市大安運動中心)	25.020867	121.545980	Gym / Fitness Center

Figure 2. DataFrames attained after using Foursquare API

3. Methodology

3.1 Exploratory Data Analysis

We will first look at the number of universities in each county/city.

Bigger cities like Taipei City, New Taipei City, Taichung City, Tainan City, and Kaohsiung City all have more than 15 universities in the area, especially in the downtown region. In this stage, one would know that going to cities are definitely a wiser choice if he/she wants to have more "university neighbors". Going to big cities will never get yourself bored.

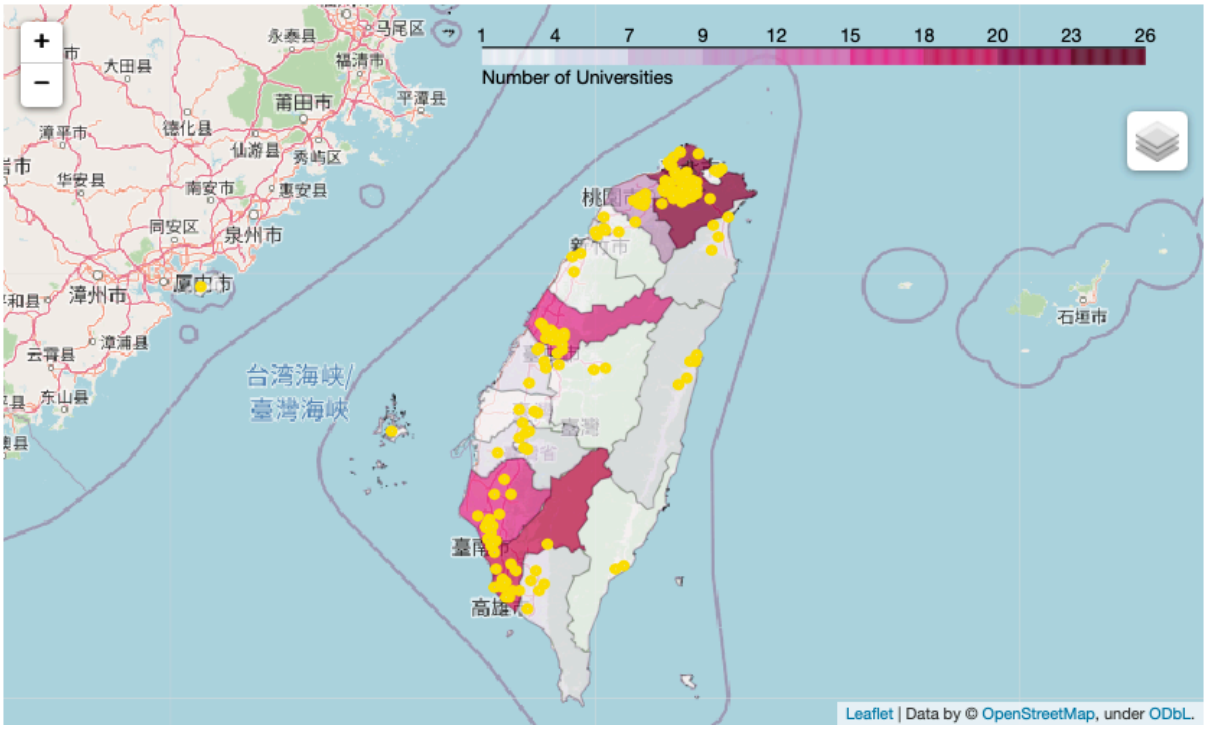

Figure 3. The choropleth map shows clear result of the number of universities in each county.

	University (zh)	University (eng)	Accessories Store	Airport	Airport Terminal	American Restaurant	Antique Shop	Arcade	Art Gallery	Art Museum	Arts & Crafts Store
0	國立政治大學	National Chengchi Universerity	0	0	0	0	0	0	0	0	0
1	國立政治大學	National Chengchi Universerity	0	0	0	0	0	0	0	0	0
2	國立政治大學	National Chengchi Universerity	0	0	0	0	0	0	0	0	0
3	國立政治大學	National Chengchi Universerity	0	0	0	0	0	0	0	0	0
4	國立政治大學	National Chengchi Universerity	0	0	0	0	0	0	0	0	0

Figure 4. The result of one-hot encoding.

3.2 Machine Learning

To analyze the data, we'd use a Machine Learning algorithm called **clustering** by the K-mean method. The method is applied to all five dataset; the steps for the method are listed in 3.2.1 section only.

3.2.1 Clustering University by Convenience

	University (zh)	Accessories Store	American Restaurant	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bar
0	一貫道崇德學院	0.0	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	0.000000	0.0
1	世新大學	0.0	0.011364	0.0	0.000000	0.022727	0.0	0.0	0.0	0.011364	0.0
2	中信金融管理學院	0.0	0.000000	0.0	0.000000	0.000000	0.0	0.0	0.0	1.000000	0.0
3	中原大學	0.0	0.000000	0.0	0.058824	0.000000	0.0	0.0	0.0	0.058824	0.0
4	中國文化大學	0.0	0.024390	0.0	0.000000	0.048780	0.0	0.0	0.0	0.024390	0.0

*Figure 5. The result of grouping universities by **mean***

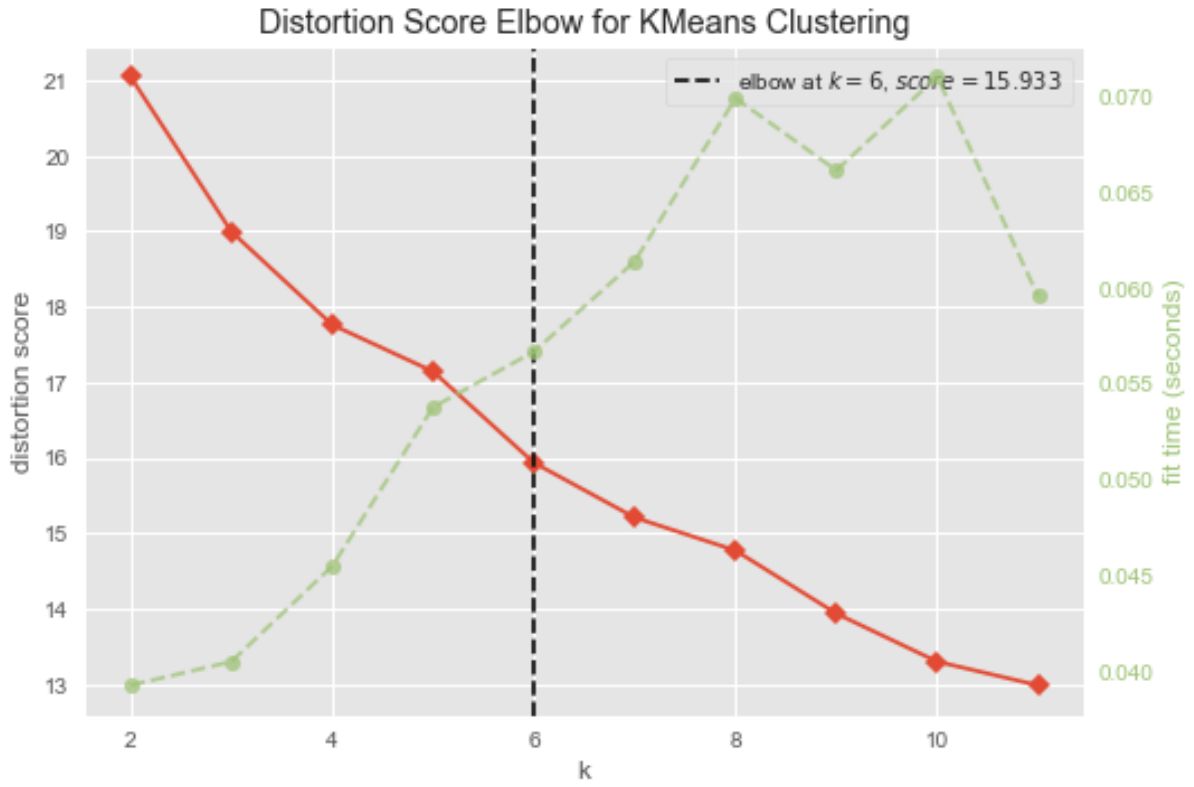

Figure 6. Determining the elbow for KMeans Clustering

First, we would perform the **one-hot encoding** technique to transform categorical data into numerical data for Machine Learning algorithm.

Then we choose a series of categories that might be more irrelevant to determine the “convenience” level. The categories include stores, restaurants, parks, clubs, etc. The full list can be seen at the source code of this project.

Then we now grouped the rows above by University by taking the **mean** of the frequency of occurrence of each Venue Category.

As a result, some universities may be left out the original list since Foursquare might not have enough data points or information regarding some locations.

Clustering Universities: Now we use Yellowbrick, an open source, pure Python project that extends the scikit-learn API with visual analysis and diagnostic tools, to help us determine the elbow for KMeans Clustering. In following sections, I will repeat the same method for finding the elbow for KMenas Clustering and thus won't specifically go through the details again.

The result of the third step, i.e. the step we obtain the mean of the frequency of occurrence, is then to be put into a model from Yellowbrick, which outputs the optimal K value for KMean clustering. After that, we set the K value for the KMeans library we use to cluster the data provided by *sckit-learn*.

As we can see from *Figure 6*, the Elbow is at K = 6. Similar objects are then grouped into the same cluster automatically by Python.

3.2.2 Clustering University by Sports Facilities

We will then perform the same technique as before to see results for sports related venues.

Specific sports-related facilities are selected for this part of study. The full list can also be viewed in the source code of the project.

The K value is 4 for 1,000-meter distance and 7 for 2,000-meter distance.

3.2.3 Clustering University by Gym Facilities

We will then perform the same technique as before to see results for gym related business. The reason for using two different distances is exactly the same as above. Since gym workout has recently been a uprising fitness trend in Taiwan with more than 600 gyms opened in 2019.

In this section, we would merge "Gym" and "Gym / Fitness Center" into one category, "Gym / Fitness Center" since they are mostly the same. One example is that if you search for "World Gym" in Taipei, Taiwan, you will see some being categorized as "Gym" while others being "Gym / Fitness Center". There might be some other cases, but since clustering based on Gym is considered more specific, I will only manage to deal with this section.

The K value for both distances is 4.

3.3 Data Analysis

We'd then analyze the clusters obtained from the above steps.

The following functions are used to help generate insights:

1. `pretty_print`, `get_class_rules`, and `cluster_report` are modified from this article on Towards Data Science; they serve as an easy way to visualize and determine the compositions of each cluster.

   `cluster_report` will generate a DataFrame containing the rule produced by DecisionTree Machine Learning Technique.

	class_name	instance_count	rule_list
4	0	4	[1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant <= 0.32500000298023224) and (Café > 0.2916666716337204)
1	1	37	[1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House <= 0.2023809589445591) and (Convenience Store <= 0.1715686321258545) and (Fast Food Restaurant > 0.0342857139185071) [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House <= 0.2023809589445591) and (Convenience Store > 0.1715686321258545)
3	2	5	[1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant > 0.32500000298023224)
5	3	2	[1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant <= 0.32500000298023224) and (Café <= 0.2916666716337204) and (Trail > 0.3500000014901161)
2	4	9	[1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House > 0.2023809589445591) [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store > 0.45000000298023224)

Figure 7. Output of the cluster_report function

2. The `return_most_common_venues` function, originally given as a function to return top n venues listed in the DataFrame, was modified to find out the top n venues by percentage and will plot a pie chart for each cluster.

3. The `analysis_report` function will help pass the variables to `return_most_common_venues` function and will return a cluster report as a DataFrame and a folium map instance.

3.3.1 Clustering Universities by Convenience

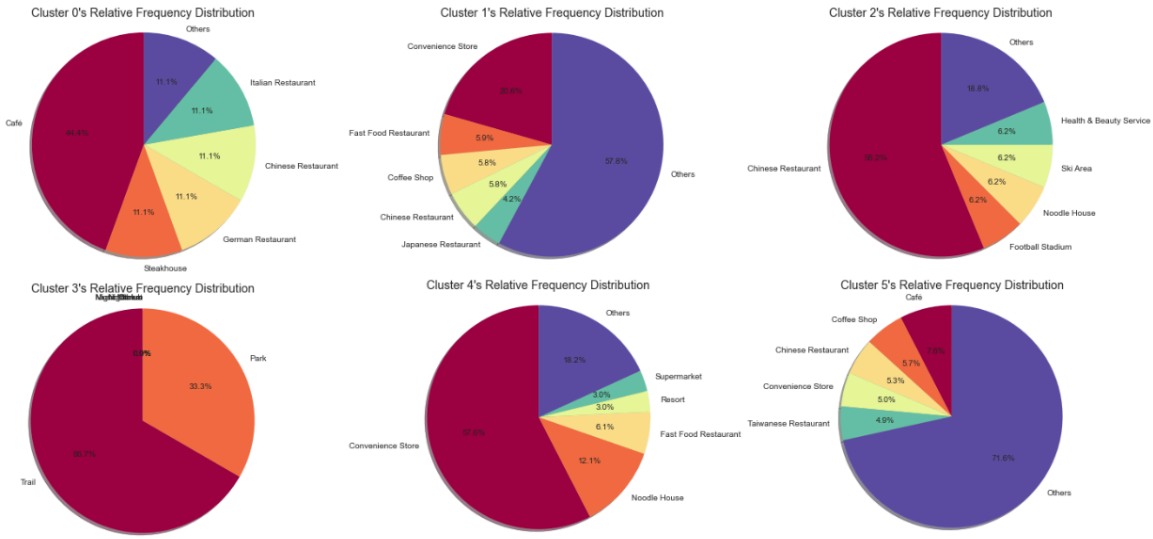

Figure 8. Clusters' Relative Frequency Distribution in Pie Charts

Cluster 0 (Red):

The report shows that Cluster 0 has a rule of:

> [1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant <= 0.32500000298023224) and (Café > 0.2916666716337204)

It means that with the probability of 1, an instance will have a convenience store taking up to less than 13% of all the venues within the 1,500m radius.

Cluster 0 also has a relatively high number of Café proportion.

Cluster 1 (Purple):

> [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House <= 0.2023809589445591) and (Convenience Store <= 0.1715686321258545) and (Fast Food Restaurant > 0.0342857139185071)

> [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House <= 0.2023809589445591) and (Convenience Store > 0.1715686321258545)

Cluster 1 has more to offer when it comes to venue categories. They are closer to the heart of the city and thus more venues can be observed. The top three

venues observed in this cluster are "Convenience Store", "Coffee Shop", and "Fast Food Restaurant".

Cluster 2 (Dodge Blue):

> [1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant > 0.32500000298023224)

There are a relatively high proportion of Chinese Restaurant in Cluster 2.

Cluster 3 (Light Blue):

> [1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant <= 0.32500000298023224) and (Café <= 0.2916666716337204) and (Trail > 0.3500000014901161)

Universities under Cluster 3 are high in trails and parks. And we can see that only *2* universities are listed.

Cluster 4 (Light Green):

> [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House > 0.2023809589445591)
>
> [1.0] (Convenience Store > 0.1342342346906662) and (Convenience Store > 0.45000000298023224)

Cluster 4 is mostly Convenience Store. As Convenience Store being the main venue for this cluster, we can see that for the most cases, the universities in this cluster couldn't be found in major cities' center areas.

Cluster 5 (Orange):

> [1.0] (Convenience Store <= 0.1342342346906662) and (Chinese Restaurant <= 0.32500000298023224) and (Café <= 0.2916666716337204) and (Trail <= 0.3500000014901161)
>
> [0.7142857142857143] (Convenience Store > 0.1342342346906662) and (Convenience Store <= 0.45000000298023224) and (Noodle House <= 0.2023809589445591) and (Convenience Store <= 0.1715686321258545) and (Fast Food Restaurant <= 0.0342857139185071)

There are 102 universities in this cluster. The venue categories in this cluster are more diverse with the top five venues only taken up to around a quarter of all others. This reflects the fact that Taiwan is well-known for its high

convenience around the neighborhoods. It does not change when it comes to universities.

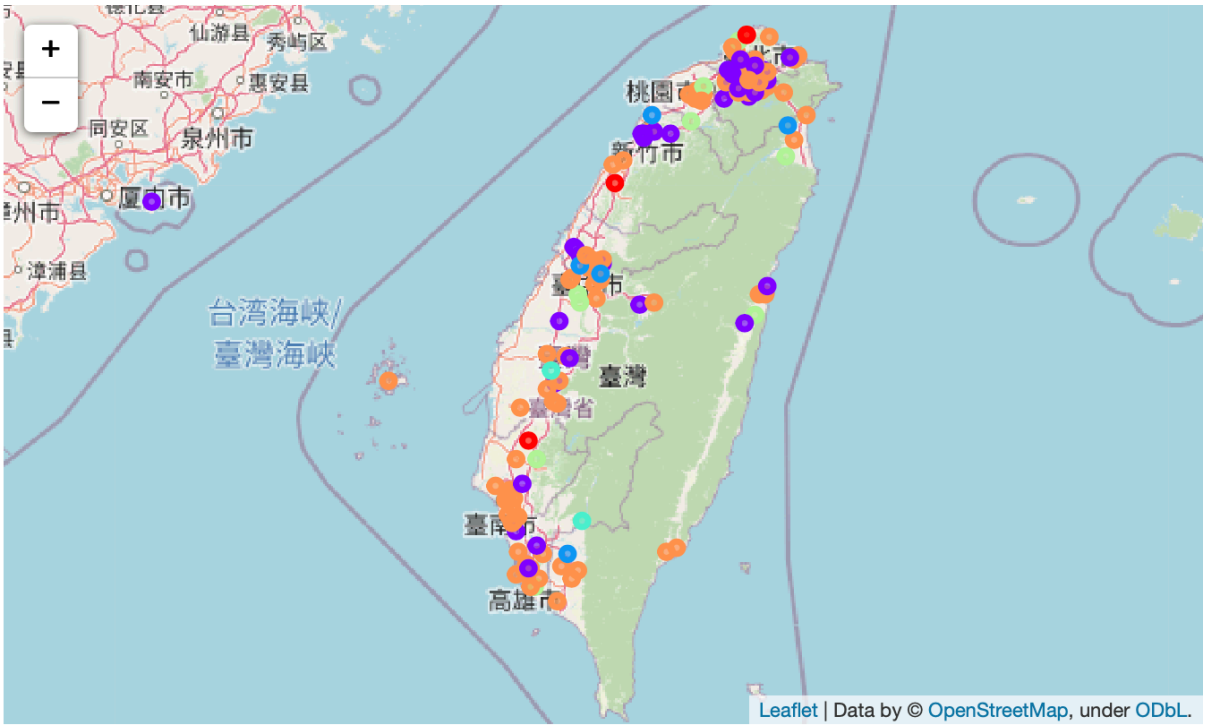

Figure 9. Map with 6 clusters

3.3.2 Clustering University by Sports Facilities

1,000-meter Radius Cluster Analysis (4)

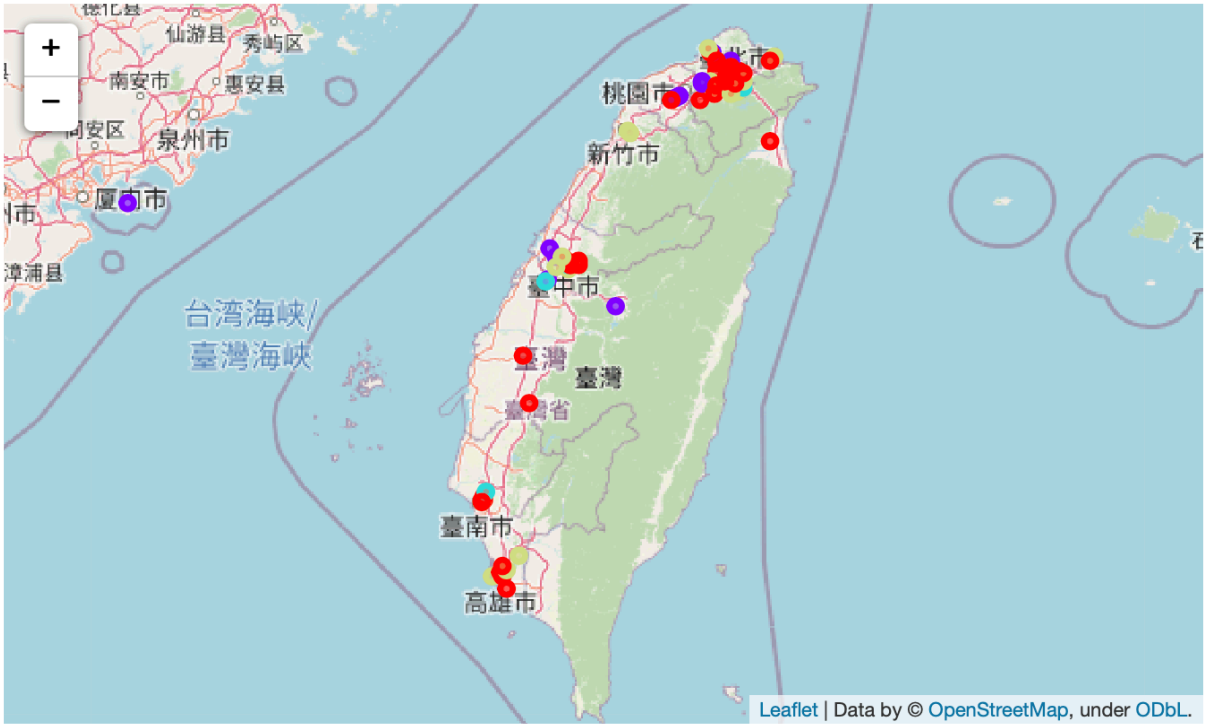

Figure 10. Map with 4 Clusters

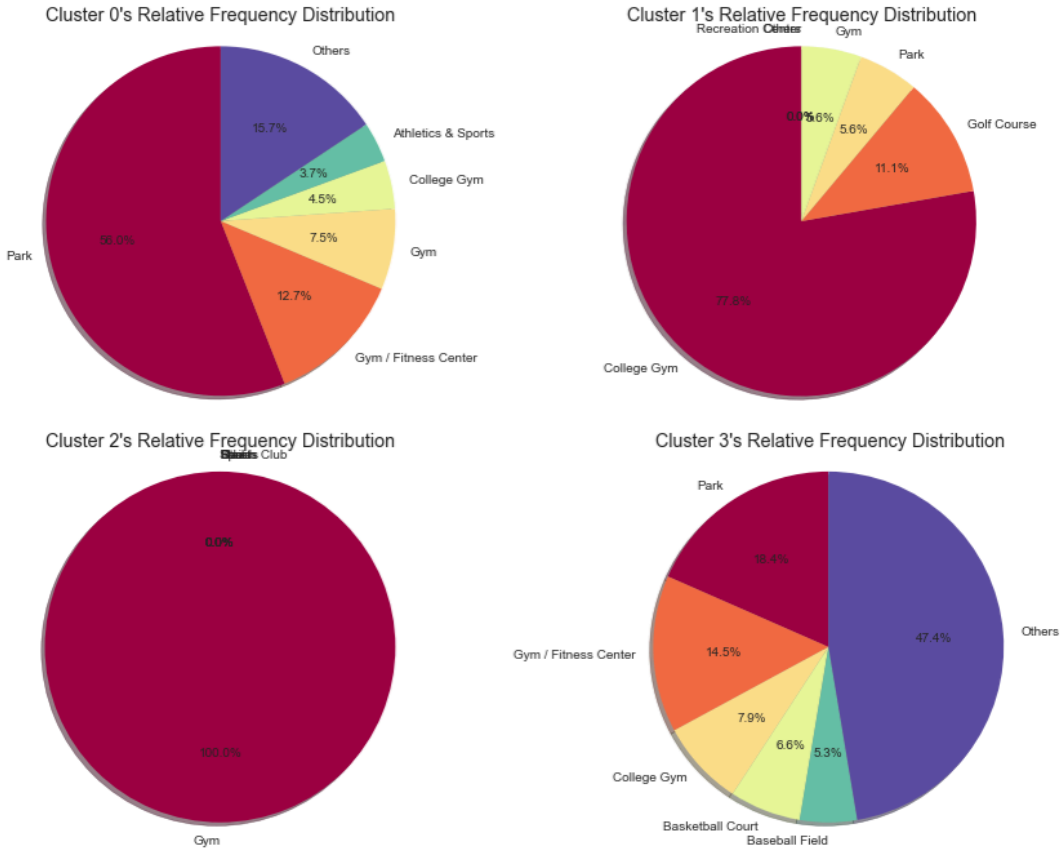

Figure 11. Clusters' Relative Frequency Distribution in Pie Charts

So what are the universities to go for convenient access to sports venues?

Cluster 0 (Red):

> [1.0] (Park > 0.3095238208770752)

Cluster 0 is the most commonly seen cluster among all with 45 instances. Park takes up to more than 30% for universities under this cluster. Other venues like Gym, Gym / Fitness Center (categorized as Gym as well), and College Gym take up to 25% of all sports venues.

Cluster 1 (Purple):

> [1.0] (Park <= 0.3095238208770752) and (College Gym > 0.4166666716337204)

Schools in Cluster 1 are equipped with at least one College Gym.

Cluster 2 (Light Blue):

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.4166666716337204) and (Gym > 0.7142857164144516)

Gym is the main source of venues under Cluster 2. Only four schools are in this cluster. Though, from the map down below, we can see that one can experience other sports facilities if they are willing to walk or drive further away from school. It is just that the closest place to go to on foot seems to be gyms only.

Cluster 3 (Olive Green):

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.4166666716337204) and (Gym <= 0.7142857164144516)

Cluster 3 has a more diverse compositions in general. Parks are expected to be less than other clusters.

2,000-meter Radius Cluster Analysis (7)

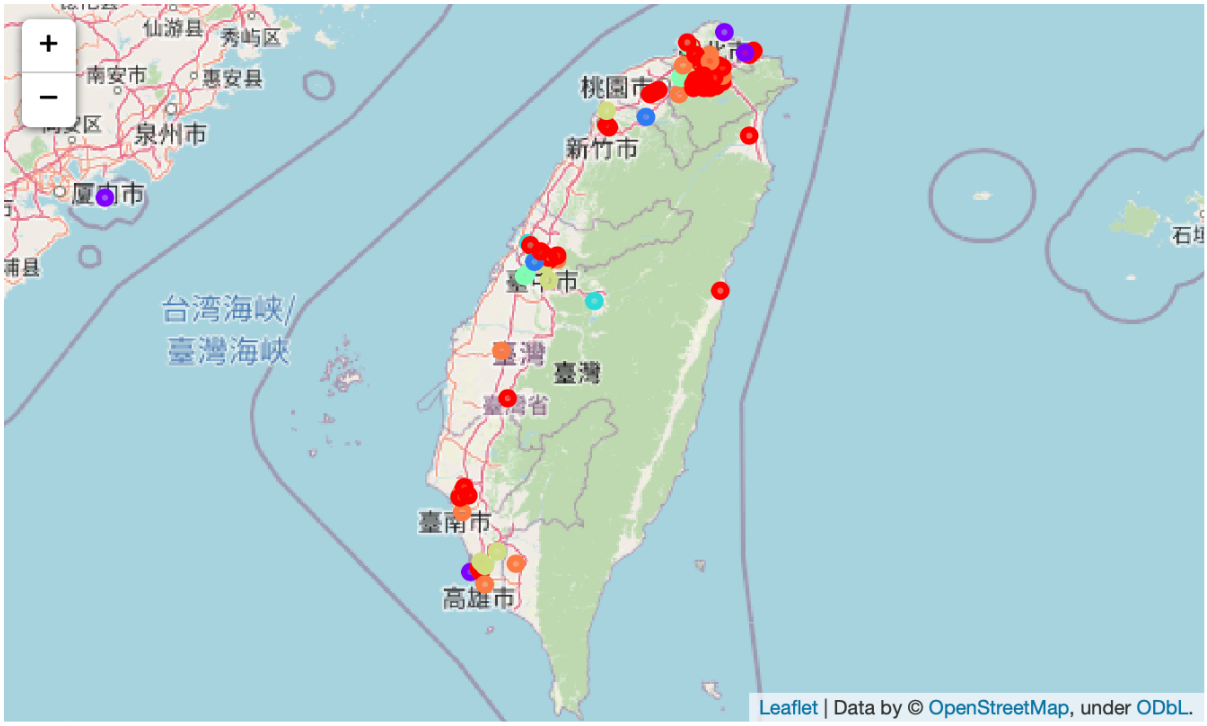

Figure 12. Map with 7 Clusters

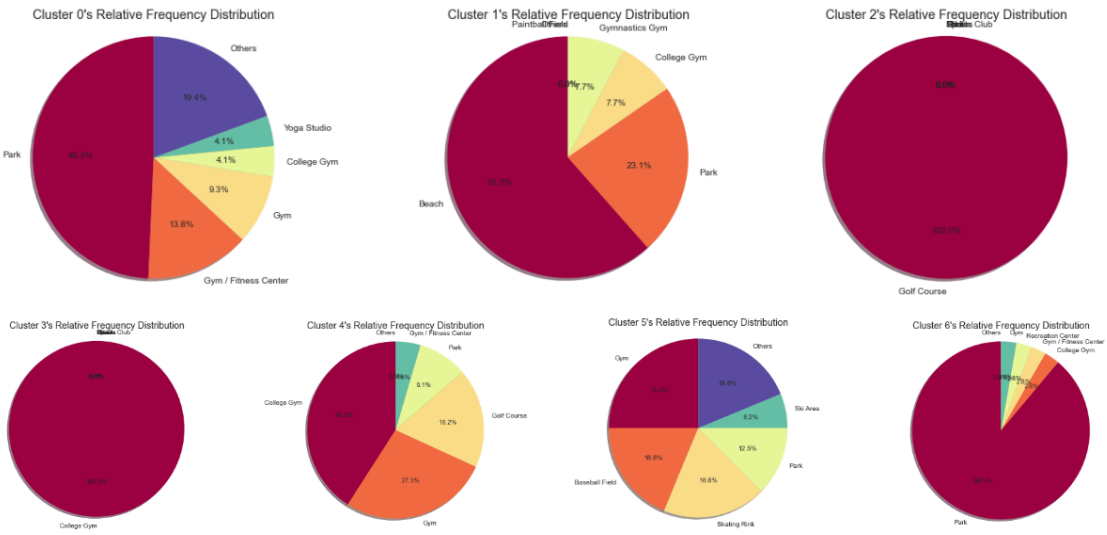

Figure 13. Clusters' Relative Frequency Distribution in Pie Charts

Cluster 0 (Red):

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.2916666716337204) and (Beach <= 0.3392857164144516) and (Golf Course <= 0.125) and (Other Great Outdoors > 0.0714285746216774)

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.2916666716337204) and (Beach <= 0.3392857164144516) and (Golf Course > 0.125) and (Bike Trail > 0.125)
>
> [1.0] (Park > 0.3095238208770752) and (Park <= 0.7128205299377441)

Park is the main category in Cluster 0. 61 universities are in this cluster. With gym-related taking up to a quarter of the whole, park taking up to almost 50%, the universities are more likely to be in the city or areas with more people.

Cluster 1 (Purple):

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.2916666716337204) and (Beach > 0.3392857164144516)

There are 4 universities in Cluster 5 with all of them having beaches to go to within 2,000-meter distance. All 4 universities are located near the coast line with one in Kinmen County, an island-based county near Mainland China.

Cluster 2 (Dodge Blue):

> [1.0] (Park <= 0.3095238208770752) and (College Gym <= 0.2916666716337204) and (Beach <= 0.3392857164144516) and (Golf Course > 0.125) and (Bike Trail <= 0.125)

The only sports venue in this cluster is Golf Course with only two universities. One being Ling Tung University and the other being Hsin Sheng Junior College of Medical Care and Management.

We can see that Ling Tung University is in Cluster 2 of the 1,000-meter group. The rule for that group is " (Park <= 0.35087719559669495) and (College Gym <= 0.4166666716337204) and (Gym <= 0.7142857164144516) ", thus with Goal Course taking 100% in its whole venue categories, we can expect the other three categories listed being none.

Hsin Sheng Junior College is not listed in the previous group.

Cluster 3 (Light Blue):

> [1.0] (Park <= 0.3095238208770752) and (College Gym > 0.2916666716337204) and (College Gym > 0.75)

College Gym is the only category in Cluster 3.

Cluster 4 (Tiffany Green):

> [1.0] (Park <= 0.3095238208770752) and (College Gym > 0.2916666716337204) and (College Gym <= 0.75)

Gym and Golf Course are the two main categories in Cluster 4. With 8 universities are in the cluster, we are expected to see less parks.

Cluster 5 (Olive Green):

> [0.8888888888888888] (Park <= 0.3095238208770752) and (College Gym <= 0.2916666716337204) and (Beach <= 0.3392857164144516) and (Golf Course <= 0.125) and (Other Great Outdoors <= 0.0714285746216774)

With 8 universities, Cluster 6 presents diverse options of sports facilities. The top 5 categories are: Gym, College Gym, Skating Rank, Park, Ski Area. As can be seen from the map below, those universities are mainly located in Taichung and Kaohsiung both with a number of 3.

Cluster 6 (Orange):

> [1.0] (Park > 0.3095238208770752) and (Park > 0.7128205299377441)

In Cluster 6, we are expected to see more parks with a proportion of more than 71%.

3.3.3 Clustering Universities by Gym Facilities

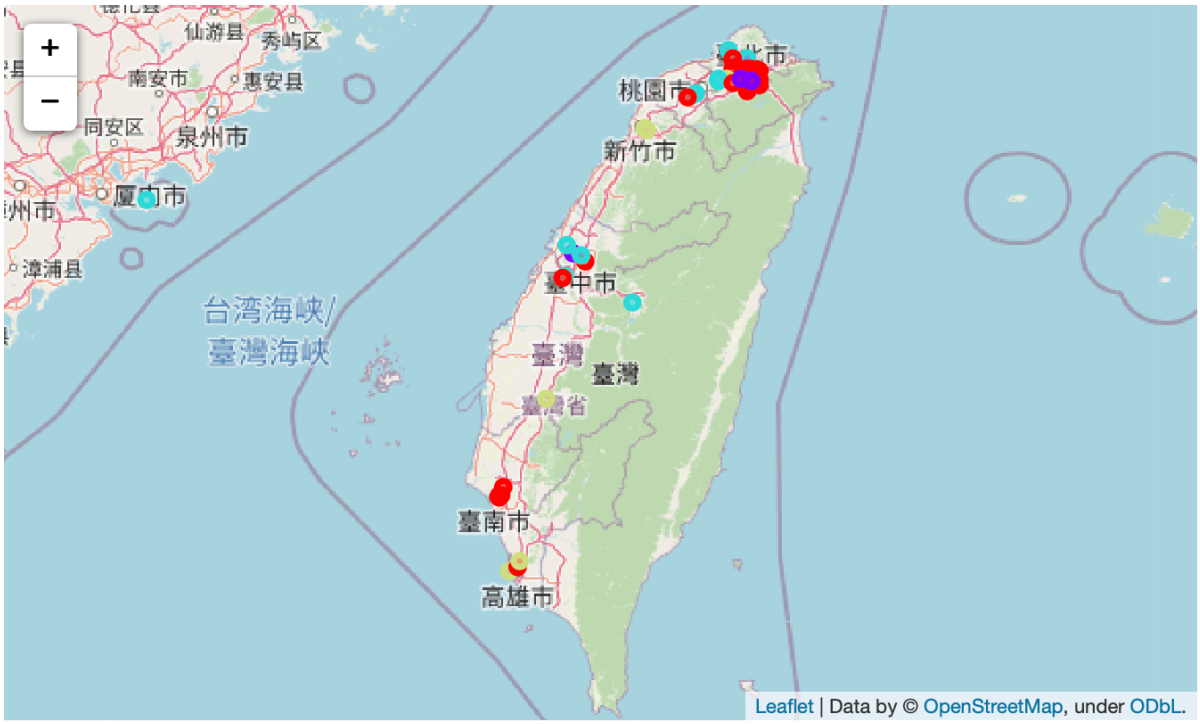

Figure 14. Map with 4 Clusters

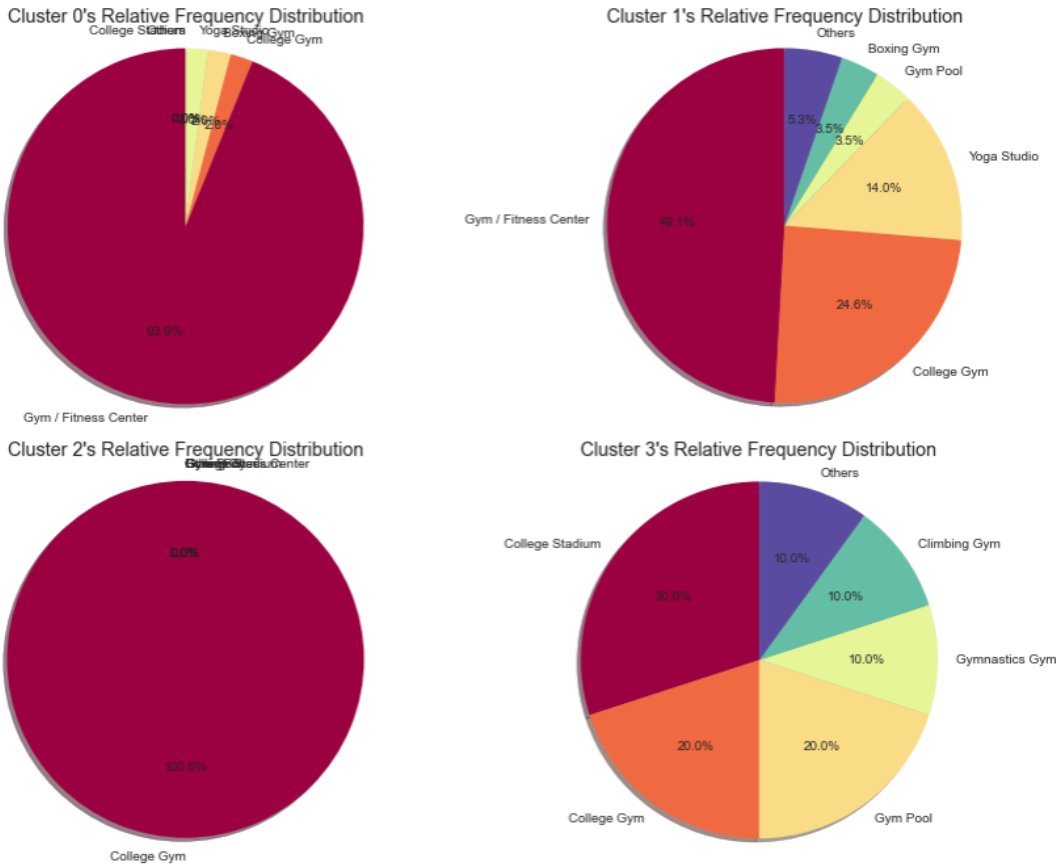

Figure 15. Clusters' Relative Frequency Distribution in Pie Charts

1,000-meter Radius Cluster Analysis (4)

Cluster 0 (Red):

> [1.0] (Gym / Fitness Center > 0.6333333551883698)

Gym and Fitness Center is the most common venue in the cluster. 24 universities are in the cluster.

Cluster 1 (Purple):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (College Gym <= 0.75) and (Gym / Fitness Center > 0.1666666716337204)

12 universities are in the cluster. Gym selections are expected to be more in the universities under Cluster 1. Students are expected to experience Gym facilities with a less-than-1,000-meter walk. The universities are mostly located in Taipei City, the central hub of Taiwan.

Cluster 2 (Light Blue):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (College Gym > 0.75)

12 universities are again to be seen in the cluster. College Gym are there to be used in those universities while you can expect nothing else from it.

Cluster 3 (Olive Green):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (College Gym <= 0.75) and (Gym / Fitness Center <= 0.1666666716337204)

Universities in Cluster 3 show more variety than others. Gym and Fitness Center are at the lowest percentage among all.

2,000-meter Radius Cluster Analysis (4)

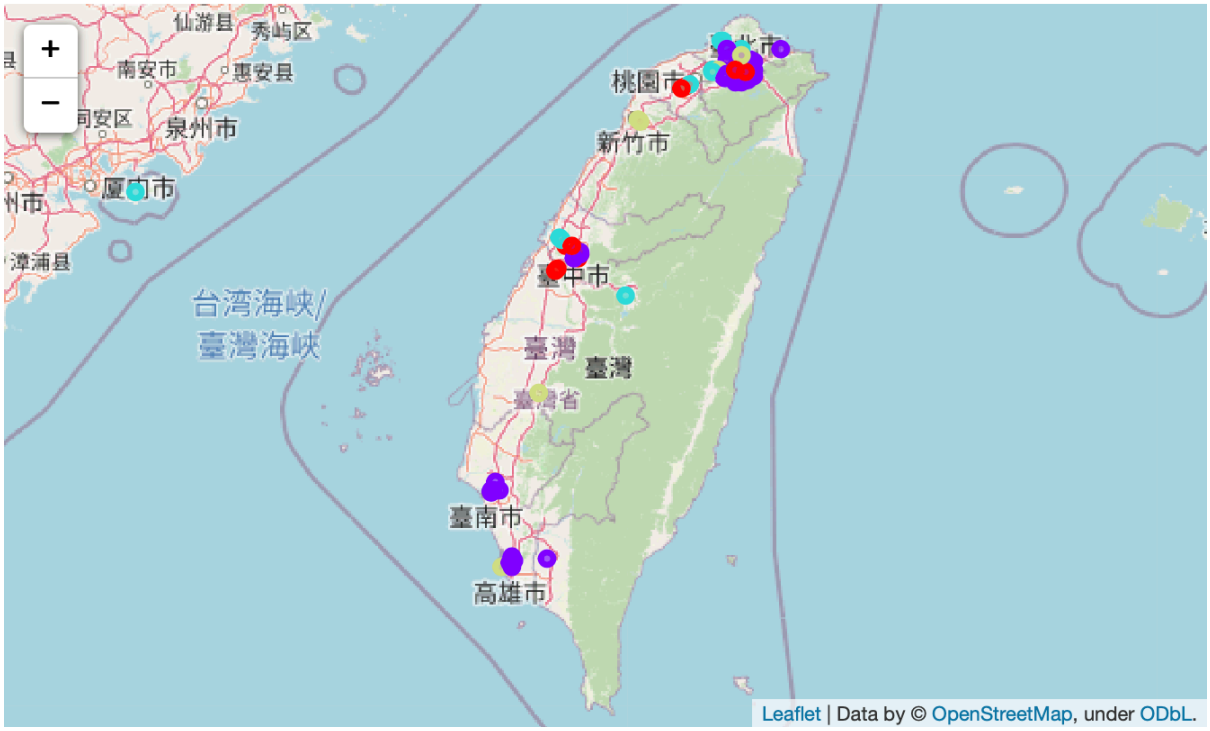

Figure 16. Map with 4 Clusters

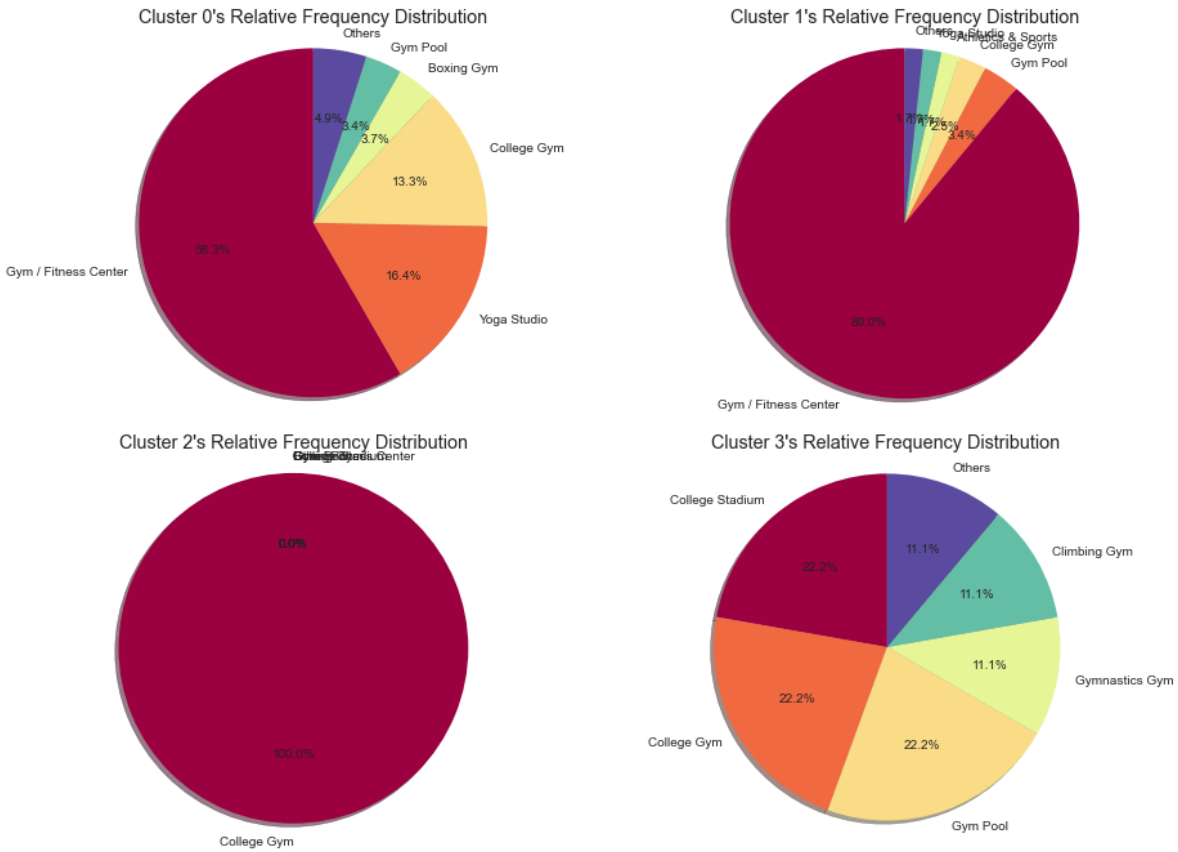

Figure 17. Clusters' Relative Frequency Distribution in Pie Charts

Cluster 0 (Red):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (Gym / Fitness Center > 0.1428571492433548)
>
> [1.0] (Gym / Fitness Center > 0.6333333551883698) and (College Gym <= 0.22500000149011612) and (Boxing Gym > 0.02500000037252903)
>
> [1.0] (Gym / Fitness Center > 0.6333333551883698) and (College Gym > 0.22500000149011612)

24 universities are in Cluster 0. Cluster 0 and Cluster 1 in the previous group are more alike according to the map and the pie charts generated. Gym / Fitness Center are the main source of gyms in these two categories and yoga studios are expected in the area. College Gyms are also provided.

Cluster 1 (Purple):

> [1.0] (Gym / Fitness Center > 0.6333333551883698) and (College Gym <= 0.22500000149011612) and (Boxing Gym <= 0.02500000037252903)

Cluster 1 and Cluster 0 in the previous group are more alike. Gym / Fitness Center are the most commonly seen source of gyms. 35 of the universities listed are in the cluster.

Cluster 2 (Light Blue):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (Gym / Fitness Center <= 0.1428571492433548) and (College Gym > 0.6666666716337204)

Universities in this cluster are similar to the ones in Cluster 2 in the previous group. Two universities are dropped from the previous group because of the expansion in radius, other than that, the rest is the same as above.

Cluster 3 (Olive Green):

> [1.0] (Gym / Fitness Center <= 0.6333333551883698) and (Gym / Fitness Center <= 0.1428571492433548) and (College Gym <= 0.6666666716337204)

In Cluster 3, universities have college gym and their roundabout have at least one gym that is not owned by the school.

4. Discussion

Since there are too many universities to consider, choosing a wide range of universities to examine is generally not a good idea. Therefore, I will be using the top ten results from QS world ranking to see the how the life is around these ten universities.

The universities chosen are:

Rank	University Name	Located City/ County
66	National Taiwan University (NTU)	Taipei City
168	National Tsing Hua University	Hsinchu City
234	National Cheng Kung University (NCKU)	Tainan City
240	National Chiao Tung University	Hsinchu City
267	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City
331	National Taiwan Normal University	Taipei City
387	Taipei Medical University (TMU)	Taipei City
416	National Sun Yat-sen University	Kaohsiung City
465	National Central University	Taoyuan City
488	National Taipei University of Technology	Taipei City

For the rest of the other 155 universities, please do feel free to play with the interactive notebook and find the university of your choice.

4.1 By Convenience

There are 6 clusters in this grouping result with Cluster 3 being the most common one.

In terms of life convenience, judging on the venue categories around the universities within a not-so-far walking distance of 1,500 meters, the clustering results of the top-10 universities are:

Cluster	University Name	County	Remarks
5 (Orange)	National Taiwan University (NTU)	Taipei City	Located in the capital city of Taiwan, NTU is definitely a school surrounded by stores and various shops.

5 (Orange)	National Tsing Hua University	Hsinchu City	Cluster 1 and Cluster 5 are very much alike, both have various venue categories to choose from. But for Cluster 1, Fast Food Restaurants are more to be seen. Since both NTHU and NCTU are right next to each other, we can expect the similar result. Since NTHU's location is closer to the main route in Hsinchu, there might be some deviation, and thus resulting in the difference between NTHU and NCTU.
5 (Orange)	National Cheng Kung University (NCKU)	Tainan City	As the highest ranked university in Tainan City, NCKU has a lot to offer when it comes to diversity of venues.
1 (Purple)	National Chiao Tung University	Hsinchu City	See NTHU.
5 (Orange)	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City	As the neighbor of NTU, NTUST shares some of venues with it. While some venues are on the other side of NTU, most NTUST students can get to those venues with extra walking time.
5 (Orange)	National Taiwan Normal University	Taipei City	NTNU is also in the neighborhood of NTU, so you could expect mostly the same lifestyle, shops, stores all around as NTU.
5 (Orange)	Taipei Medical University (TMU)	Taipei City	Near to the famous Taipei 101, TMU is expected to have similar convenience level as NTU.
5 (Orange)	National Sun Yat-sen University	Kaohsiung City	Sitting right next to coastline, NSYSU is surrounded by great views. As a tourist destination, different kinds of venues are expected.
5 (Orange)	National Central University	Taoyuan City	Being the top university in Taoyuan City, convenience is a must. Since it is not so far away from the downtown area, diverse venues can be found around campus.

5 (Orange)	National Taipei University of Technology	Taipei City	NTUT is located at a busy intersection in Taipei City. Since it is on the most commonly passed route in Taipei, convenience level is expected to be on or beyond the average as well.

4.2 By Sports Venues

4.2.1 Within 1,000m 🚶

Closer sports venues can be more motivating and appealing for students who long for doing sports. 1,000-meter range is not far for walking. The following table concludes what each university has to offer.

Cluster	University Name	County	Remarks
0 (Red)	National Taiwan University (NTU)	Taipei City	Park is the most commonly seen venue in Cluster 0, and Cluster 0 serves as the most popular cluster among all. As a school with more than 30,000 students athletes and students, there should be other types of sports venues around NTU.
3 (Olive Green)	National Tsing Hua University	Hsinchu City	NTHU and NCTU almost share the same space and area. It's no surprise that they both belong to Cluster 2 with less park ratio and more other sorts of sports venues.
0 (Red)	National Cheng Kung University (NCKU)	Tainan City	Parks are expected to be visited around NCKU. As seen from the map, there are plenty of green spaces around campus.
3 (Olive Green)	National Chiao Tung University	Hsinchu City	See NTHU.
3 (Olive Green)	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City	Since some parks are located on one side of NTU, NTUST, though being close to NTU, is expected to see less parks (but still have parks), and thus other sports venues are expected.
0 (Red)	National Taiwan Normal University	Taipei City	Parks can also be seen around NTNU.
0 (Red)	Taipei Medical University (TMU)	Taipei City	Near the center of Taipei, TMU still has some parks around.
3 (Olive Green)	National Sun Yat-sen University	Kaohsiung City	As a coastline university, NSYSU has balanced amount of sports resources for students to explore.

N/A	National Central University	Taoyuan City	There's not enough sports venues around NCU within 1,000-meter distance.
0 (Red)	National Taipei University of Technology	Taipei City	Close to the busy intersection, NTUT still has Parks with over 35% among all sports venues around.

Most universities are in **Cluster 0** (Red), which is of no surprise. Parks are popular sports venues for most people, but for students, there needs to be more other sorts of sports venues, especially for those larger universities. With that said, parks are still important assets and it seems that Taiwan is doing a decent job of greening the environment in cities.

4.2.2 Within 2,000m 🚴

Cluster	University Name	County	Remarks
0 (Red)	National Taiwan University (NTU)	Taipei City	With a 2,000-meter range, there are more sorts of venues around NTU as expected. Students can not only experience greeneries in the city, but also some gym resources.
0 (Red)	National Tsing Hua University	Hsinchu City	With a 2,000-meter ride, students can find themselves sharing the similar portion of sports facilities/resources around NTHU and NCTU.
0 (Red)	National Cheng Kung University (NCKU)	Tainan City	The compositions are similar to other universities (with results) in Taiwan. Besides greeneries, other sports facilites are there to be seen.
0 (Red)	National Chiao Tung University	Hsinchu City	
0 (Red)	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City	Located just next to NTU, this time, with 2,000-meter radius, we can expect more similar resources around both campus.
0 (Red)	National Taiwan Normal University	Taipei City	Close up to NTU, NTNU agin belongs to the same cluster as NTU.
0 (Red)	Taipei Medical University (TMU)	Taipei City	Parks, gym, fitness center are sure to be found around TMU's neighborhoods.
1 (Purple)	National Sun Yat-sen University	Kaohsiung City	Since it is special to have beaches near universities, NSYSU shares the same cluster with only other three universities. Surfing and other water sports can be more popular for students on campus.
N/A	National Central University	Taoyuan City	There are still not enough information regarding this place and its roundabout.

0 (Red)	National Taipei University of Technology	Taipei City	From the results above, it is not surprising that NTUT shares again the same cluster as other top universities in Taipei City.

For the very first time, both NTHU (Hsinchu City) and NCTU (Hsinchu City) belong to the same cluster as NTU (Taipei City). That shows with more walk or biking distance, students can enjoy similar resources. Biking can be regarded as a way to warm up regardless of sports. Moreover, most universities share the same cluster (**Cluster 0**, Red) together. That shows that the proportion of sports venues around these top universities in Taiwan are generally the same within 2,000-meter reach.

4.3 By Gym

Gym plays an important role in modern day daily source of exercise. Since it is considered more convenient and easy to get started, gym has become more popular these days. Generally speaking, College Gyms are cheaper than branded Gyms out of campus.

4.3.1 Within 1,000m 🚶

There are four clusters in this group. Though College Gyms are cheaper for sure, having more gyms around campus can guarantee a less crowded workout environment. I believe universities in Cluster 3 can have the students experience more events and activities, while universities in Cluster 1 can guarantee a less crowded gym.

Cluster	University Name	County	Remarks
1 (Purple)	National Taiwan University (NTU)	Taipei City	Students at NTU have more gym options, whether to go to school gym or gyms outside, which seems to be a must for a 30,000-student size university.
3 (Olive Green)	National Tsing Hua University	Hsinchu City	Students in both NTHU and NCTU are expected to experience various sorts of training gyms, including gym pools. Note that some facilities on one campus might not be shared with the other.
0 (Red)	National Cheng Kung University (NCKU)	Tainan City	Students might have to look into gyms outside. But in Southern Taiwan, gym membership fees are much cheaper than that of the Northern side, students have higher chance to workout at gyms outside.
3 (Olive Green)	National Chiao Tung University	Hsinchu City	See NTHU.

1 (Purple)	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City	Students at NTUST also have options to go to school gym or gyms outside. Less crowded gyms are expected.
1 (Purple)	National Taiwan Normal University	Taipei City	Near NTU, NTNU also has more options in terms of gyms.
0 (Red)	Taipei Medical University (TMU)	Taipei City	Gyms are seen, but might be a little be more expensive.
3 (Olive Green)	National Sun Yat-sen University	Kaohsiung City	Not only beaches as seen above, NSYSU also provides wide choices of gyms for students to go to.
N/A	National Central University	Taoyuan City	As seen in the sports venue section, NCU seeks no sports venues as per Foursquare API's fetch.
0 (Red)	National Taipei University of Technology	Taipei City	Gyms can be seen out of campus as it is located in a busy intersection in Taipei City. Same as TMU, those gyms can be more expensive.

4.3.2 Within 2,000m 🚴

Cluster	University Name	County
0 (Red)	National Taiwan University (NTU)	Taipei City
3 (Olive Green)	National Tsing Hua University	Hsinchu City
1 (Purple)	National Cheng Kung University (NCKU)	Tainan City
3 (Olive Green)	National Chiao Tung University	Hsinchu City
0 (Red)	National Taiwan University of Science and Technology (Taiwan Tech)	Taipei City
0 (Red)	National Taiwan Normal University	Taipei City
0 (Red)	Taipei Medical University (TMU)	Taipei City
3 (Olive Green)	National Sun Yat-sen University	Kaohsiung City
N/A	National Central University	Taoyuan City
0 (Red)	National Taipei University of Technology	Taipei City

Top universities in Taipei City are all in **Cluster 0**, which shows that with a slightly longer distance, students can have a less crowded workout experience. There are Boxing Gyms in their neighborhoods as well.

And again, NTHU, NCTU, and NSYSU share a different cluster than others. Students have a wide range of gym variations.

One thing worth noticing from the map is that universities in the suburban area of Taipei are mostly Cluster 1, which means that those universities have no College Gym records, but only with Gym / Fitness Center labels. If you are already on budget, might have to look into those schools more.

In terms of gyms, 1,000-meter range and 2,000-meter range seem to differ less. But in general, students can pick universities with more gym variations to have a less crowded workout experience; and students could pick ones in Cluster 3 (Olive Green) - 1,000-meter / Cluster 3 (Olive Green) - 2,000-meter if they need more training variations.

5. Conclusion

Though the clustering results can be referred to some extents, students should consider more and look into more details when it comes to choosing an university to go to. The clustering results are a source to help search for similar places with similar venue categories.

One thing about Foursquare API is that the results in Taiwan are not complete. For example, Fitness Factory, as the second largest gym brand in Taiwan has almost no results on Foursquare. In addition, due to pricing, I can only fetch up to 100 venues for one location. This may result in inaccurate clustering results since it is not reflecting the real-time status around the particular place. For above reasons, it is suggested that one visit the university of your heart before actually choosing it. If one can achieve more than a hundred venues from a single location, it might be better to use that results, extract sports venues from the entire sets, and analyze the proportion of them in the entire neighborhood.

Through Data Science, we should be able to achieve the results without visiting in reality. But since the data are missing somehow in the first place, there is still a lot of improvements about this project.

Further application of this project can also be applied to business problem like where to open up a new branch for gyms by looking into the neighborhoods with less gyms. I know that there are still many things to consider, but at least it is an easy start to approach the problem and can give a firsthand insights to the gym owners.

Overall, I learned a lot from the project itself and the entire specialization. Machine Learning techniques are also used besides traditional statistics method. Rooms are left for improvements, and I am open to all kinds of suggestions.