

A Review of Accessible Education Resources in Taipei City

Prepared by To-Hong WU

1. INTRODUCTION

Education is a global issue that concerns all family across different countries and time span. Parents always put the investment in children's education in the top of the list. It is so important that almost every country has the compulsory education policy which is usually between 9 to 12 years, throughout primary, middle and high schools. In order to carry out this compulsory obligation of their citizens, governments use the *school districts* to assign children to schools within each of the districts. However, considering the unevenly distributed socio-economic status of communities, there are usually good school district which parents are more inclined to send their kids to.

1.1 The Problem

There are lots of studies regarding children's education, from recognition to development, to the extent how the quality of education would effect the achievement in the future.

This study focuses on the education resources that can be adequately accessed by the neighborhoods. This study uses neighborhoods in Taipei City, Taiwan as a case study to examine if a gap exists among neighborhoods, regarding the accessibility to education resources.

1.2 Where to leverage Foursquare Places API

This study uses Foursquare Places API to collect the education resources (venues) that are accessible within a distance of each neighborhood.

1.3 Targeted Audience

The Government

Schools and libraries are important infrastructures that governments set up a budget for. However, it is difficult to measure the return on this investment, that is, the social benefit (in contrary to financial benefits) of schools and libraries, or where the government should put more resources. This study provides an example of good reference about where needs more attention from the government.

The parents

Many family would select a good spot to live in, and accessibility to education resources is always one of the most important feature when making such decision.

1.4 Quick facts of Taipei City, Taiwan

Taipei locates in East Asia and is the political and social-economic capital of Taiwan, where the Taiwan Government resides. Taipei City has a land area of 271.8 square kilometers (about the same size as Orlando, Florida), population of 2.6 million (close to the figure of Chicago), and a population density almost the same as New York City. The GDP (PPP) of Taipei City is estimated to be \$300 billion, slightly higher than Atlanta.

Chinese Mandarin is the official language, which is why in the rest of this study you will see Chinese characters, which this study will try to convert into English for better reading.

The old town of Taipei City was built in the end of 19th century, locating at southwestern side of current Taipei City. It is also where the Taipei Main Station locates. In this study you will see the old town is where most education resources are at.

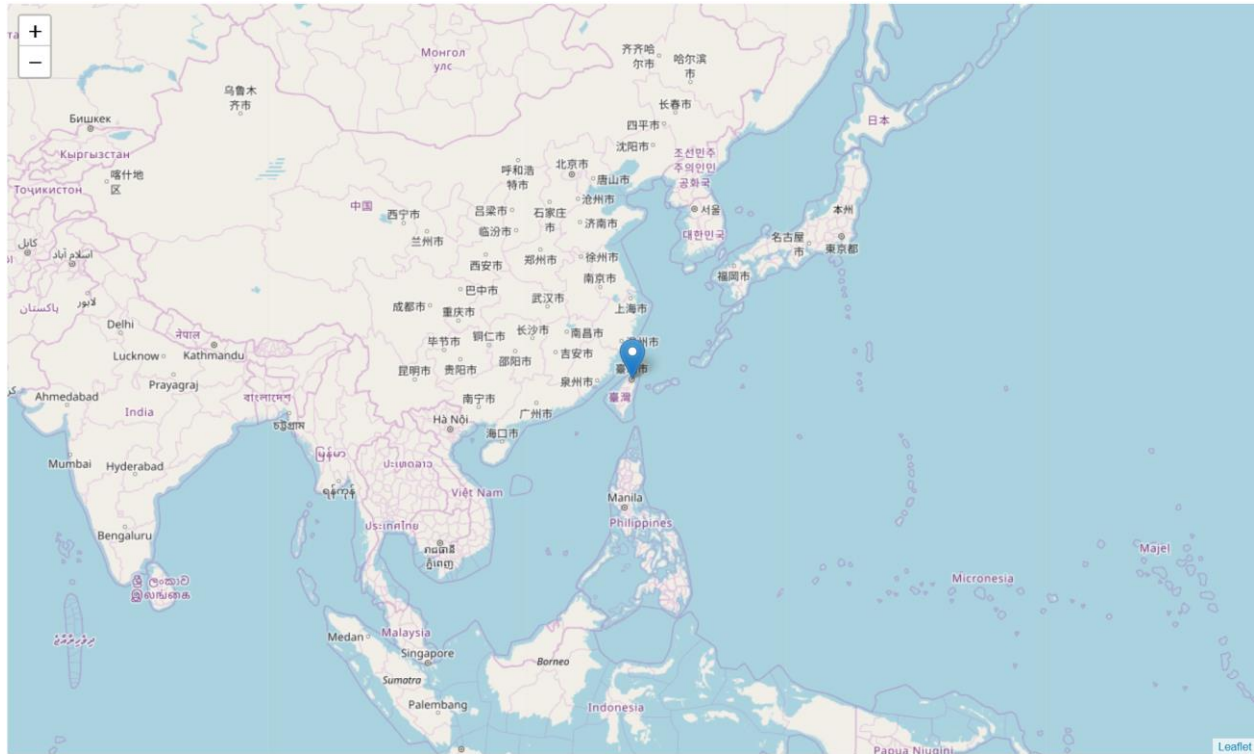


Figure 1. The location of Taipei City

2. METHODOLOGY

This study will collect the data of accessible education resources of each neighborhood in Taipei. By using K-means clustering, this study will divide the neighborhoods into 5 clusters and identify the cluster(s) which has a relatively weak accessibility to education.

The study will be carried out as follows:

2.1 Study of accessible education resources

This section will directly use K-means clustering on accessible education resources of each neighborhoods. The result will be a geometrical distribution of the clusters.

2.2 Study of accessible education resources per capita

The allocation of resources should take the population who will receive them into consideration. A more denser neighborhood should enjoy more resources. Following previous section, this section will calculate the accessible education resources per capita and remake the clustering model. The result should show where the per capita resources are distributed.

2.3 Study of accessible education resources per capita, taking socio-economic status into consideration

Some wealthier communities don't necessarily locate in the central part of the city. However this doesn't mean that kids from that community would have less access to education. Given that this study is aiming at providing a reference for policy making in regards of the allocation of education resources, the socio-economically weaker neighborhoods should be taken into account and given a priority in the resources.

In this section both socio-economic status and education resources will be used as variables to run the K-means clustering model, in order to identify the neighborhoods in need of the most help, that is, where the government budget could have the best return of social benefit.

2.4 Operational Definition of terms

Term	Definition
Education Resources	Venues acquired on Foursquare Places API. The categories of the venues limit to School, Library and Museum. Given that the compulsory education in Taipei is 9 years, this study will only consider the subcategories of Middle School, Elementary School and Preschool.
Accessible Education Resources	The education resources within a distance of 1,500 meters from the geometry center of a neighborhood.
Accessible Education Resources per capita	The accessible education resources divided by the number of population who are under 15 years old, due to the 9 years of compulsory education.
Socio-economic status	The median value of the household income of each neighborhood

3. Data Acquisition and Cleaning

3.1 GeoJSON of Taipei Neighborhoods and Their Geometric Centers

3.1.1 GeoJSON of Taipei Neighborhoods

This study needs a map containing all neighborhoods' coordinates to work on. Unfortunately such ready-for-use information couldn't be found during the study. However, there are a few GeoJSON and Shapely file of Taipei City. After searching and comparing different accuracy rate of GeoJSON files, this study uses the [GeoJSON file of Taipei](#) contributed by [littlebtc](#), as the infrastructure which this study will be constructed upon. Using GeoJSON instead of the coordinates also brings the benefits of better visualization effect on the map.

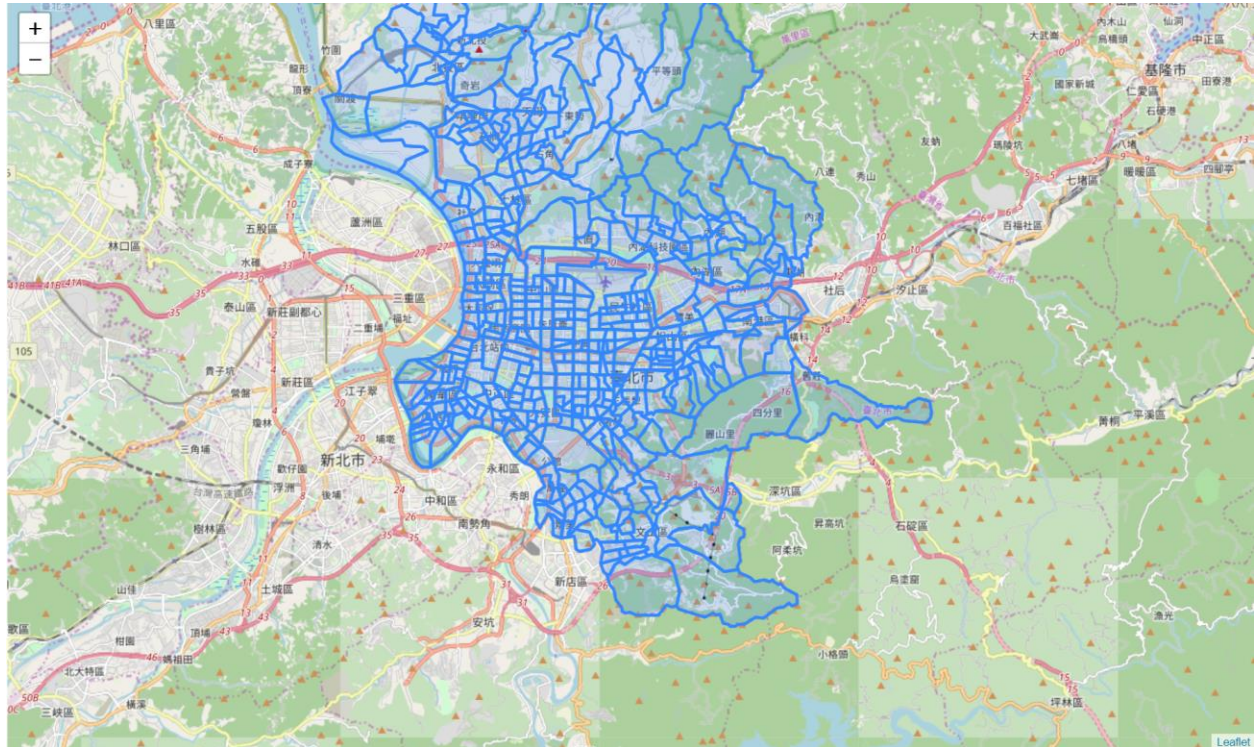


Figure 2. GeoJSON map view of Taipei City

3.1.2 Calculate the geometric centers of neighborhoods

The next task is to calculate the geometric center of each neighborhood polygon within the GeoJSON file, which I will use as the reference point to explore venues on Foursquare.

Though there are a few libraries working on different calculation of polygons, it seems that they are not created for GeoJSON. Thanks to [brandonxiang](#) who created a super useful project [geojson-python-utils](#) that solves the problem. With the centroid function, I can check these neighborhoods on the map.

It appears that the Chinese characters cannot *popup* properly on the folium map, and not too many discussion covers the encoding issues like this. Thanks to [Yafei's Blog](#) who solves this issue by bypassing it with html popups.

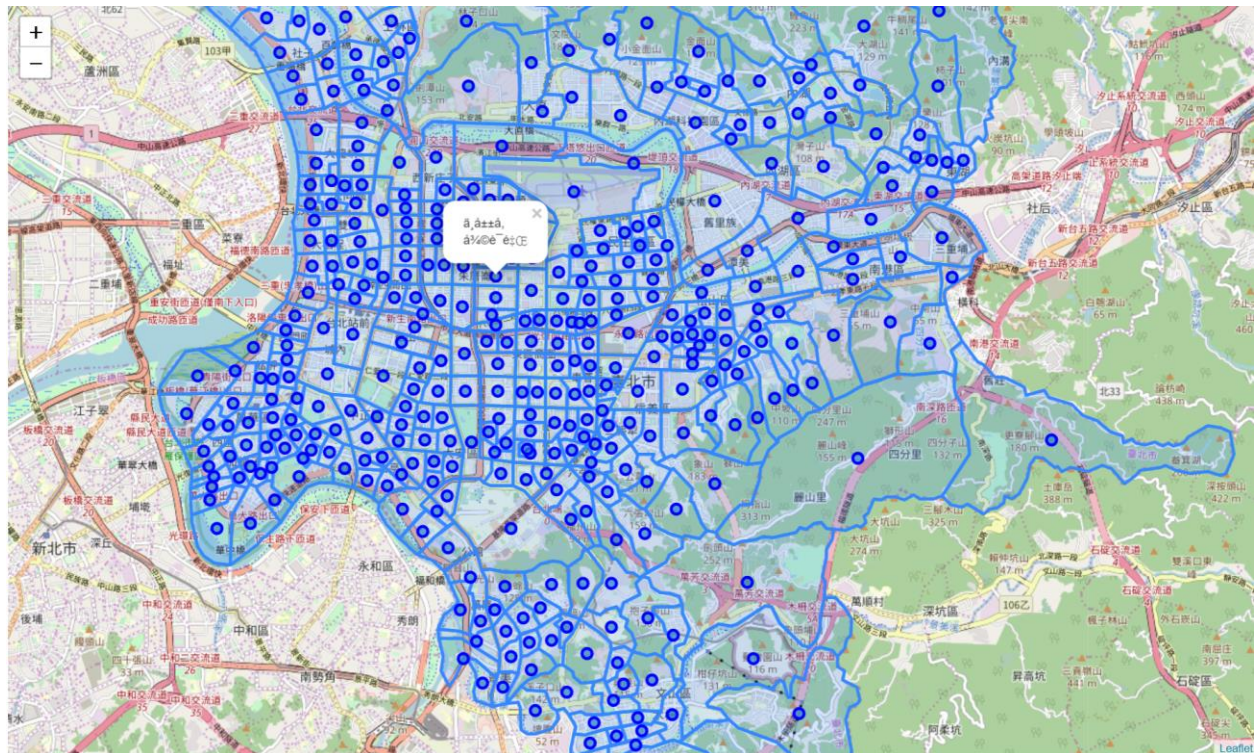


Figure 3. GeoJSON map with the geometric centers of neighborhoods with Chinese character issues

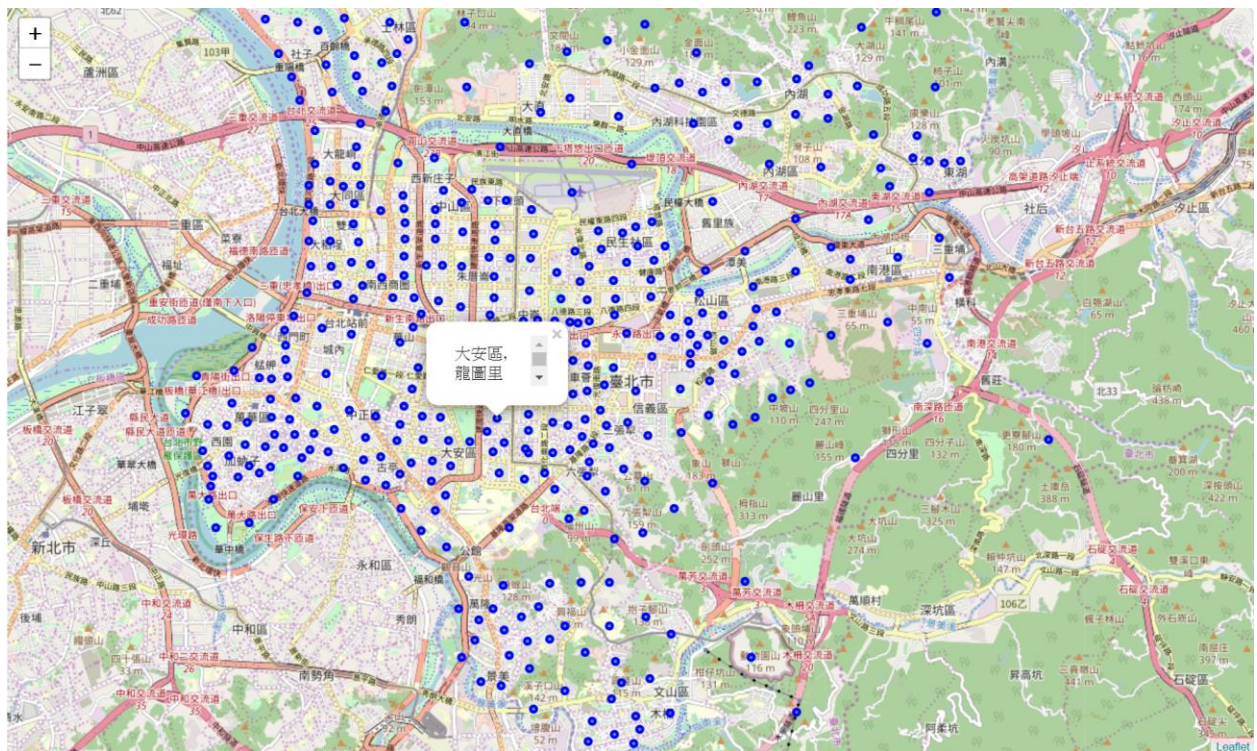


Figure 4. GeoJSON map with the geometric centers of neighborhoods (Chinese character issues fixed)

3.2 Taipei Household Income and Population under 15

3.2.1 Data cleaning

The statistics data is collected from the government website and stored on [github](#) for use.

The household income data is read from an excel file. Some of the titles of columns are renamed for future use, where "dist" and "neigh" represent district and neighborhood respectively. The population data was separated by ages and sex. A new column "under_15" is created to store the sum of the population. The two dataframes are then merged with the GeoJSON dataframe which contains the coordinates of all neighborhoods.

Table 1 shows the required information for later study, where dist_neigh, under_15 and income_median represent the neighborhood, population under 15, and the median value of household income.

When comparing with the original GeoJSON data where there are 456 neighborhoods, the new dataframe contains only 451. The five missing neighborhoods could be because of the lack of the socio-economic data, change of administration neighborhoods or any other reasons, which is beyond the scope of this study.

Table 1. List of neighborhoods, coordinates and its socio-economic data

	dist_neigh	Latitude	Longitude	under_15	income_median
0	松山區莊敬里	25.070698	121.563780	753	688
1	松山區東榮里	25.060350	121.558050	1476	919
2	松山區三民里	25.060024	121.562244	1045	800
3	松山區新益里	25.061795	121.567122	621	714
4	松山區富錦里	25.060958	121.564352	806	749
...
446	北投區關渡里	25.117710	121.467772	1679	641
447	北投區泉源里	25.153236	121.520774	247	584
448	北投區湖山里	25.149328	121.539538	188	614
449	北投區大屯里	25.160564	121.505817	166	570
450	北投區湖田里	25.179281	121.547548	101	589

451 rows × 5 columns

3.2.2 A quick look at the socio-economic data

Quickly checking the distribution of the population and household income of the neighborhoods, both figures show right-skewed distribution, meaning that they have longer tails to the right. In such case median instead of mean value could usually be a better choice to represent the group.

Below figures show how the neighborhoods distributed geometrically. When visualizing the data on map, it's hard not to notice that how similar these two figures are. Is it possible that household income and population under 15 are correlated for an unknown reason? Does wealthier family tend to have more kids than others? These topics are certainly deserved a more careful look in the future.

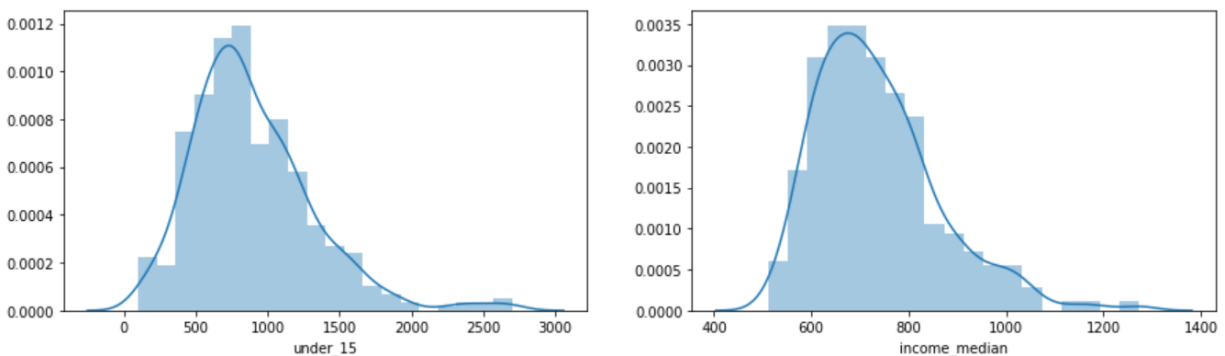


Figure 5. Distribution of neighborhoods for population under 15 and household income median

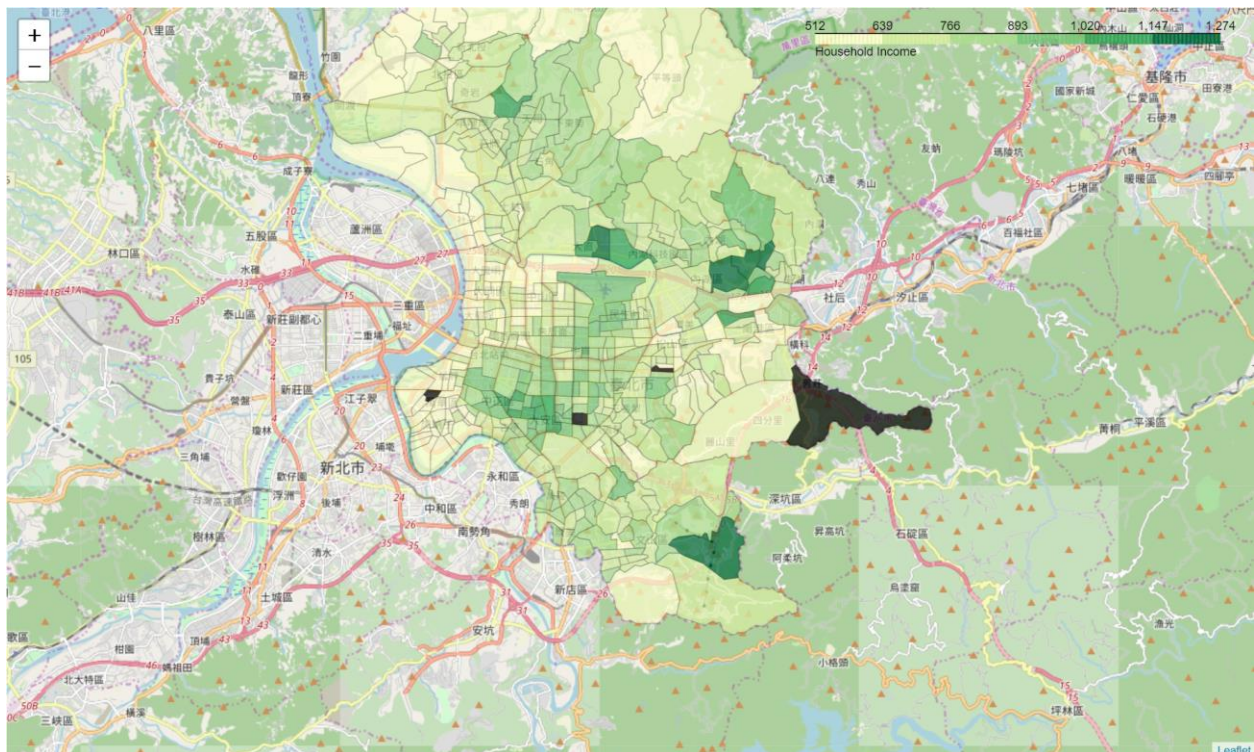


Figure 6. Neighborhoods distribution based on household income

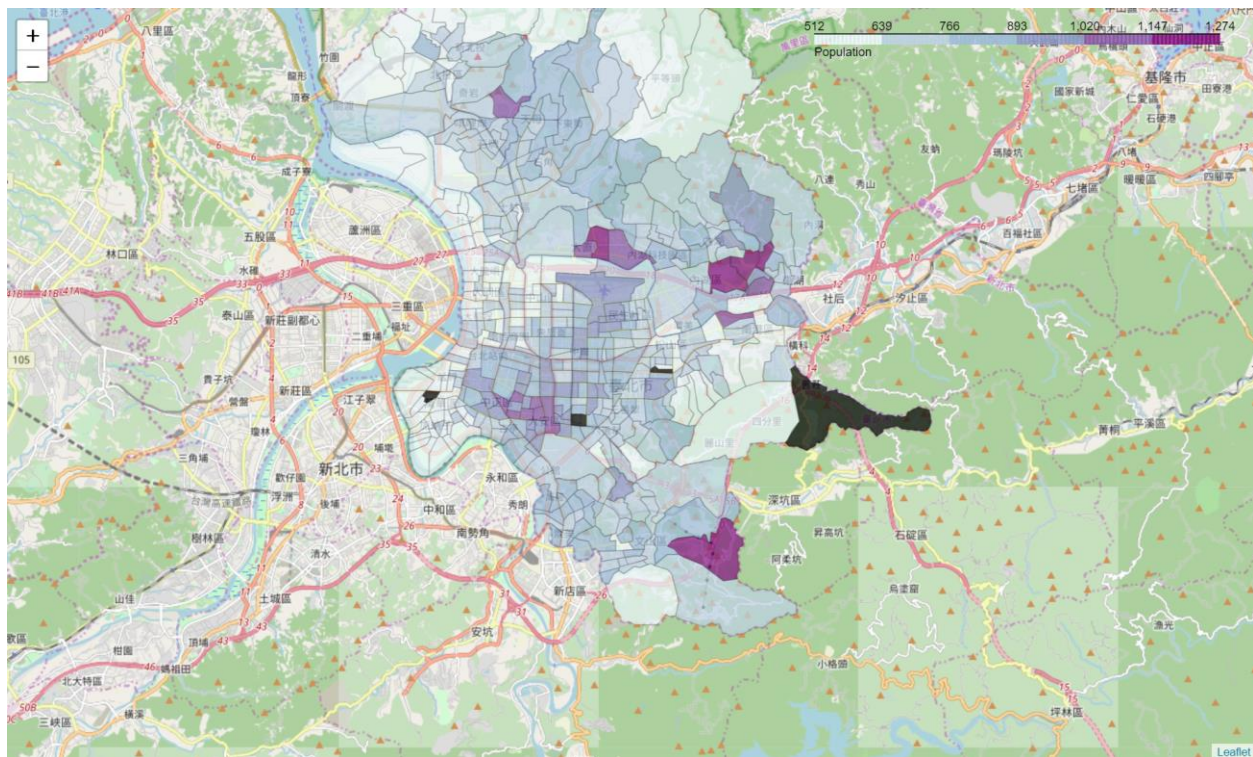


Figure 7. Neighborhoods distribution based on under-15 population

3.3 Education Resources by Foursquare

3.3.1 Data cleaning

To get to know what categories of venues we can find on Foursquare, first this study pulls out the entire categories list provided by Foursquare. However, it's not very easy to understand even if we convert the list to a dataframe because of its hierarchical structure.

I turned out to directly scrap it on Foursquare webpage of [Venues Categories](#), and find "Museum" subcategory under "Arts & Entertainment", and "School" and "Library" under "Professional & Other Places". These category IDs are then used to get the list of accessible education resources of each neighborhood, where more than 9,000 venues are successfully retrieved.

However the results still show a wide range of categories where many of them should not be identified as education resources (Figure 8). This study narrows down the category "School" to only include School, Elementary School, Middle School and Preschool, and "Museum" to all related kinds of museums. The list is significantly cut down by 2000 to 7225 venues.

Table 2 List of Education Resources on Foursquare

	categoryName	categoryId
0	School	4bf58dd8d48988d13b941735
1	Library	4bf58dd8d48988d12f941735
2	Museum	4bf58dd8d48988d181941735

```

array(['Community College', 'Library', 'School', 'Elementary School',
      'Middle School', 'Museum', 'Memorial Site', 'High School',
      'Comedy Club', 'Nursery School', 'Art Museum', 'Public Art',
      'Café', 'Music School', 'Factory', 'Preschool', 'History Museum',
      'College Academic Building', 'Indoor Play Area',
      'Performing Arts Venue', 'Trade School', 'Medical School',
      'Monument / Landmark', 'Language School', 'College Arts Building',
      'Music Venue', 'Temple', 'Park', 'Government Building',
      'Private School', 'University', 'Church', 'Swim School',
      'Historic Site', 'Office', 'Music Store', 'Science Museum',
      'Coffee Shop', 'Driving School', 'Planetarium'], dtype=object)

```

Figure 8. List of categories of venues retrieved from Foursquare

Table 3. Refined list of venues

	Neighbourhood	venue_id	venue_name	venue_latitude	venue_longitude	venue_category_name
2	松山區莊敬里	4eb253b0f5b94483884b11fa	Happy Marian	25.058909	121.557636	School
4	松山區莊敬里	4cd258364aa9f04dcdde56cb	臺北市立三民國民小學 Taipei Municipal SanMin Elementary ...	25.064342	121.565906	Elementary School
5	松山區莊敬里	4c81bbbe74d7b60cbe1d7ad8	台北市立民權國民小學 Taipei Municipal MinQuan Elementary...	25.062611	121.563132	Elementary School
6	松山區莊敬里	4c46fd491ddec928e9f19b32	臺北市立濱江國民中學 Taipei Municipal BinJiang Junior Hi...	25.079964	121.559820	Middle School
7	松山區莊敬里	4cdd91967e2e236af2b3791b	臺北市立民生國民中學 Taipei Municipal MinSheng Junior Hi...	25.059093	121.566732	Middle School

3.3.2 A quick look at the accessible education resources

Checking on how the neighborhoods access the education resources, one can easily find that the top 5 neighborhoods which have the most accessible resources are all in the old town, the southwestern part of Taipei City, while the last five are located in outer area.

However, there are 5 neighborhoods missing due to that they have zero access to education. These neighborhoods have been brought back by setting the numbers to zero, for later analysis.

Table 4. Numbers of accessible education resources of each neighborhood

	dist_neigh	Latitude	Longitude	under_15	income_median	Library	Museum	School	edu_Total
0	松山區莊敬里	25.070698	121.563780	753	688	2.0	0.0	6.0	8.0
1	松山區東榮里	25.060350	121.558050	1476	919	4.0	1.0	11.0	16.0
2	松山區三民里	25.060024	121.562244	1045	800	3.0	0.0	9.0	12.0
3	松山區新益里	25.061795	121.567122	621	714	1.0	0.0	7.0	8.0
4	松山區富錦里	25.060958	121.564352	806	749	3.0	0.0	9.0	12.0
...
446	北投區關渡里	25.117710	121.467772	1679	641	0.0	0.0	5.0	5.0
447	北投區泉源里	25.153236	121.520774	247	584	0.0	0.0	1.0	1.0
448	北投區湖山里	25.149328	121.539538	188	614	0.0	0.0	2.0	2.0
449	北投區大屯里	25.160564	121.505817	166	570	0.0	0.0	0.0	0.0
450	北投區湖田里	25.179281	121.547548	101	589	0.0	0.0	1.0	1.0

451 rows × 9 columns

3.3.3 Check the unique venues and where they locate

There are more than 7,000 venues in the list, where most of them are duplicated simply because of the area covered by the radius of 1,500 meters. The search results are refined by removing duplicates and find 333 unique venues, of which 310 are located within only 204 neighborhoods in Taipei City.

It's not surprising that more than half neighborhoods have no education resources given that the land area of neighborhoods could be very small inside the old town.

Table 5. List of where unique venues locate

v_cat	venue_neigh	Library	Museum	School	Total
29	中正區黎明里	2	3	5	10
19	中正區建國里	2	3	1	6
17	中正區南門里	0	3	2	5
113	士林區臨溪里	1	3	1	5
106	士林區福佳里	2	2	0	4
...
89	南港區舊莊里	1	0	0	1
90	南港區萬福里	0	0	1	1
91	南港區西新里	0	0	1	1
11	中山區永安里	0	0	1	1
102	士林區永倫里	0	0	1	1

204 rows × 5 columns

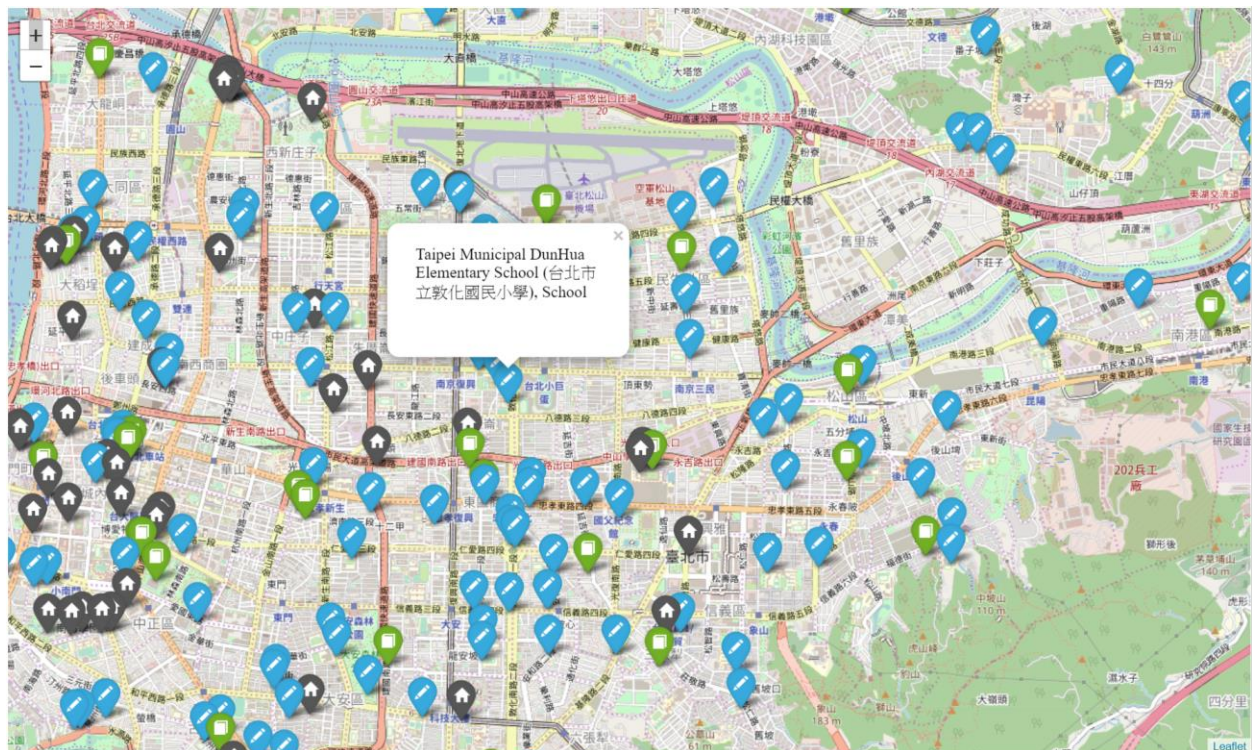


Figure 8. Map where the unique venues are located

4. RESULTS

4.1 Clusters of Accessible Education Resources

This study uses K-means clustering to run the numbers of accessible education resources for all neighborhoods. The K values is set as 5 due to the general practice for socio-economic studies such as Gini coefficient to analyze the top and bottom 20%.

Table 6 and 7 shows the results of clustering. It seems that the number of each cluster doesn't exist some kind of ranking. That is, Cluster 1 is not necessarily better or worse than Cluster 0. While Cluster 1 has the most school, Cluster 2 should be the best one given that it has all kinds of resources. Meanwhile Cluster 4 has the least resources among the clusters.

Taking a careful consideration of mean value of each variable, this study reassign the cluster in an ascending order where Cluster 4 is the best one. By doing this helps the visualization on map where the audience will find the result is much easier to understand. One can easily see that the neighborhoods with the best accessibility to education resources are concentrated in southwestern part of Taipei City, which is the old town.

Table 6. Result of clusters assigned to each neighborhood

	Cluster1	Library	Museum	School
0	4	2.0	0.0	6.0
1	0	4.0	1.0	11.0
2	0	3.0	0.0	9.0
3	4	1.0	0.0	7.0
4	0	3.0	0.0	9.0
...
446	4	0.0	0.0	5.0
447	4	0.0	0.0	1.0
448	4	0.0	0.0	2.0
449	4	0.0	0.0	0.0
450	4	0.0	0.0	1.0

451 rows × 4 columns

Table 7. Mean value of each cluster (before reassigning)

	Library		Museum		School	
	mean	count	mean	count	mean	count
Cluster1						
0	3.098214	112	1.187500	112	10.544643	112
1	4.575342	73	4.232877	73	18.657534	73
2	5.400000	40	13.525000	40	15.775000	40
3	1.872340	47	7.680851	47	11.106383	47
4	1.273743	179	1.016760	179	4.413408	179

Table 8. Mean value of each cluster (after reassigning)

	Library		Museum		School	
	mean	count	mean	count	mean	count
Cluster1_reassigned						
0.0	1.273743	179	1.016760	179	4.413408	179
1.0	3.098214	112	1.187500	112	10.544643	112
2.0	1.872340	47	7.680851	47	11.106383	47
3.0	4.575342	73	4.232877	73	18.657534	73
4.0	5.400000	40	13.525000	40	15.775000	40

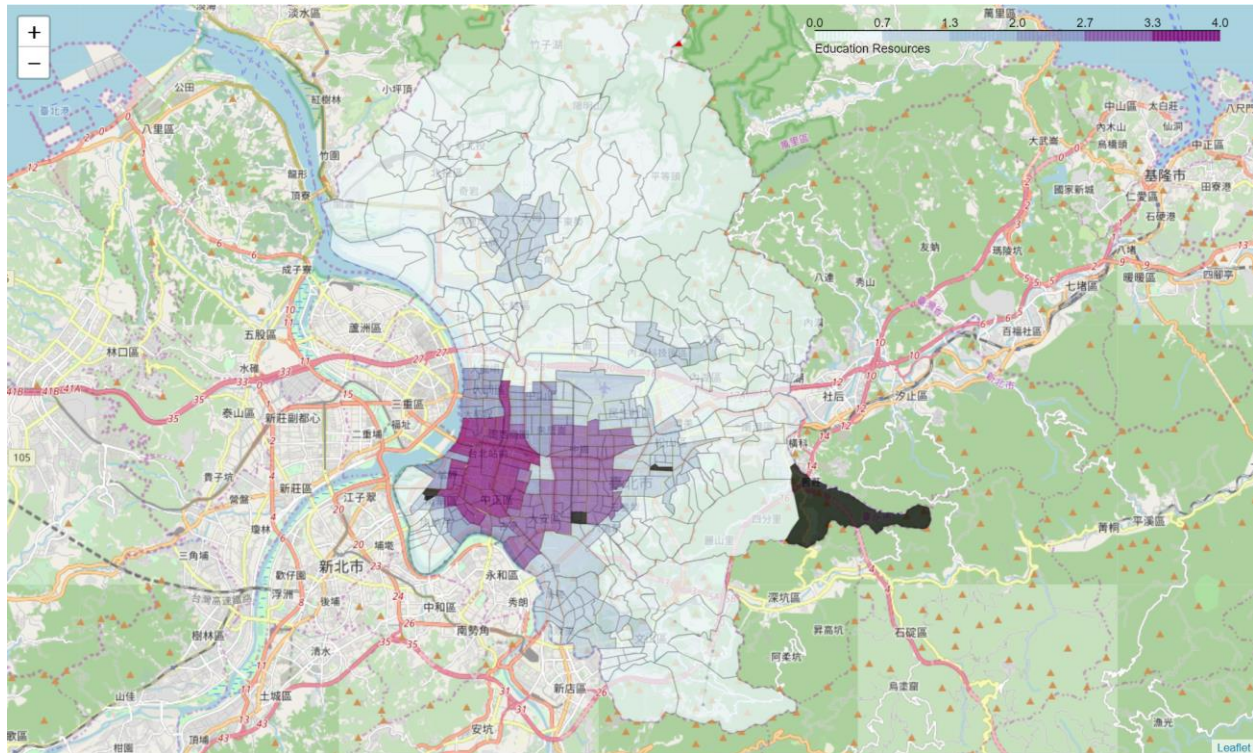


Figure 9. Clusters of Accessible Education Resources

4.2 Clusters of Accessible Education Resources Per Capita Under 15

Previous section shows an unevenly distribution of education resources in Taipei City. However, one should understand that more resources would be allocated to highly populated area to achieve higher social benefits. This section further analyses the per-capita education resources and make new clusters based on them.

Table 9 and 10 shows the accessible education resources per thousand capita under 15 and the cluster number.

In Table 10 finds that Cluster 0 and 4 are closely tied to each other in terms of total per capita resources. While Cluster 0 has far more schools (per capita), Cluster 4 has good numbers of both schools and museums. It is hard to say that which education resources, School or Museum, should have higher weight to provide a new ranking.

This study reassigns the clusters by the total numbers of resources. Table 11 and Figure 10 show the results of distribution. One can see that while the original Cluster 0 (with the most School) are distributed in discrete area, original Cluster 4 still tends to concentrate in the old town area where lots of museum locate.

Table 9. Result of clusters assigned to each neighborhood

	Cluster2	Library	Museum	School
0	1	2.656042	0.000000	7.968127
1	1	2.710027	0.677507	7.452575
2	1	2.870813	0.000000	8.612440
3	3	1.610306	0.000000	11.272142
4	3	3.722084	0.000000	11.166253
...
446	1	0.000000	0.000000	2.977963
447	1	0.000000	0.000000	4.048583
448	3	0.000000	0.000000	10.638298
449	1	0.000000	0.000000	0.000000
450	1	0.000000	0.000000	9.900990

451 rows × 4 columns

Table 10. Mean value of each cluster (before reassigning)

	Library		Museum		School	
	mean	count	mean	count	mean	count
Cluster2						
0	9.637153	19	11.405141	19	42.941080	19
1	1.368081	184	1.095917	184	4.944486	184
2	5.444704	85	8.123020	85	22.108947	85
3	3.520382	135	3.085905	135	12.703711	135
4	8.803661	28	26.184080	28	28.357054	28

Table 11. Mean value of each cluster (after reassigning)

	Library		Museum		School	
	mean	count	mean	count	mean	count
Cluster2_reassigned						
0.0	1.368081	184	1.095917	184	4.944486	184
1.0	3.520382	135	3.085905	135	12.703711	135
2.0	5.444704	85	8.123020	85	22.108947	85
3.0	8.803661	28	26.184080	28	28.357054	28
4.0	9.637153	19	11.405141	19	42.941080	19

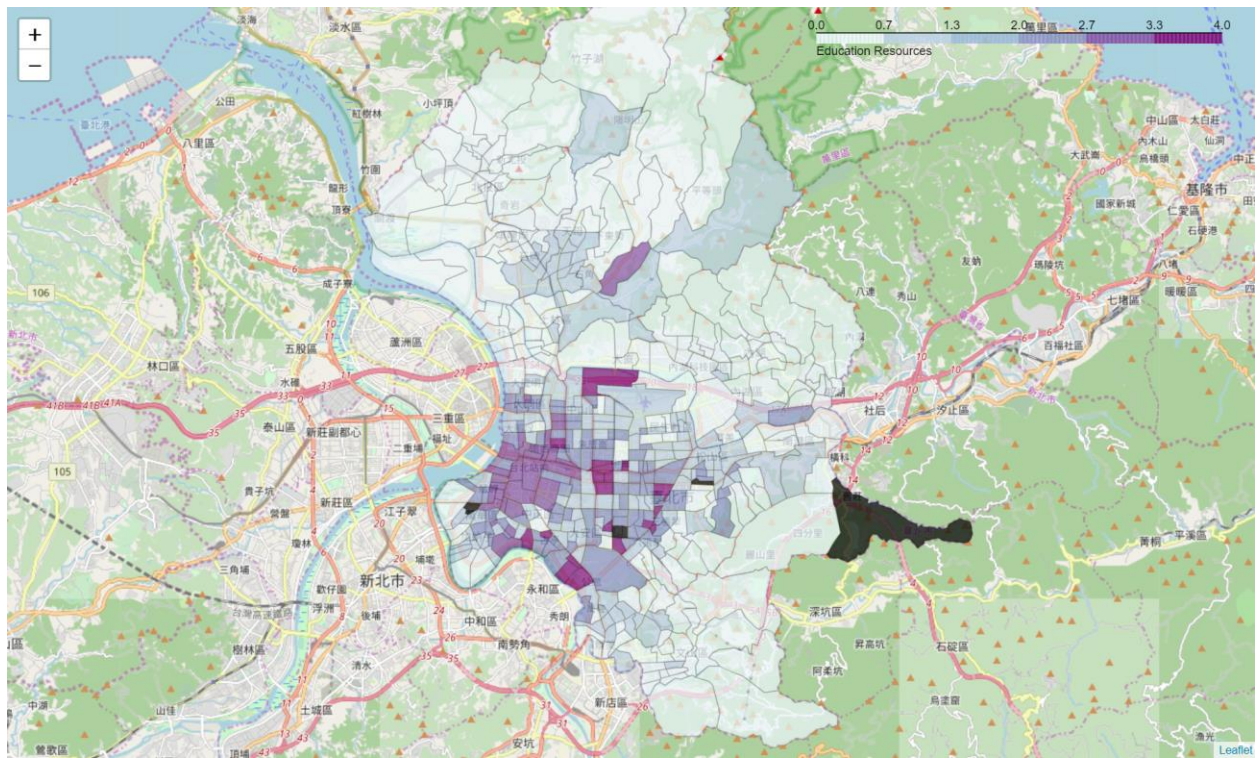


Figure 10. Clusters of Accessible Education Resources per capita under 15

4.3 Clusters of Accessible Education Resources Per Capita, Taking Socio-economic Status into Consideration

Finally, this study takes socio-economic data into consideration. In order to simplify the question regarding whether resources of different categories should be given different weights, this study uses the total accessible education resources per 1000 capita as one variable. The median of household income is the other variable (Table 12). The scale of two variables differs a lot so the data is normalized before running the K-means clustering model.

Table 13 shows the result of the clustering, which is hard to tell if a ranking exists for comparison. This study then reassigns the clusters based on the normalized sum of the two variables and creates new clusters (Table 14).

We can see from Table 14 about how income effects the clusters of distribution. The two clusters (reassigned Cluster 3 and 1) that have highest income only rank third and fourth place in the numbers of education resources, while the cluster with highest education resources takes the third place in household income.

It is notable that the reassigned Cluster 0 have both the lowest income and education resources.

Table 12. List of household income and education resources of each neighborhood

	income_median	edu_per_cap
0	688	10.624170
1	919	10.840108
2	800	11.483254
3	714	12.882448
4	749	14.888337
...
446	641	2.977963
447	584	4.048583
448	614	10.638298
449	570	0.000000
450	589	9.900990

Table 13. Mean value of each cluster (before reassigning)

	income_median		edu_per_cap	
	mean	count	mean	count
Cluster3				
0	720.977778	45	64.345839	45
1	634.036765	136	10.398817	136
2	978.844828	58	20.492272	58
3	683.285714	84	34.099251	84
4	772.796875	128	12.727933	128

Table 14. Mean value of each cluster (after reassigning)

	Cluster3		income_median		edu_per_cap	
	mean	count	mean	count	mean	count
Cluster3_reassigned						
0.0	1	136	634.036765	136	10.398817	136
1.0	4	128	772.796875	128	12.727933	128
2.0	3	84	683.285714	84	34.099251	84
3.0	2	58	978.844828	58	20.492272	58
4.0	0	45	720.977778	45	64.345839	45

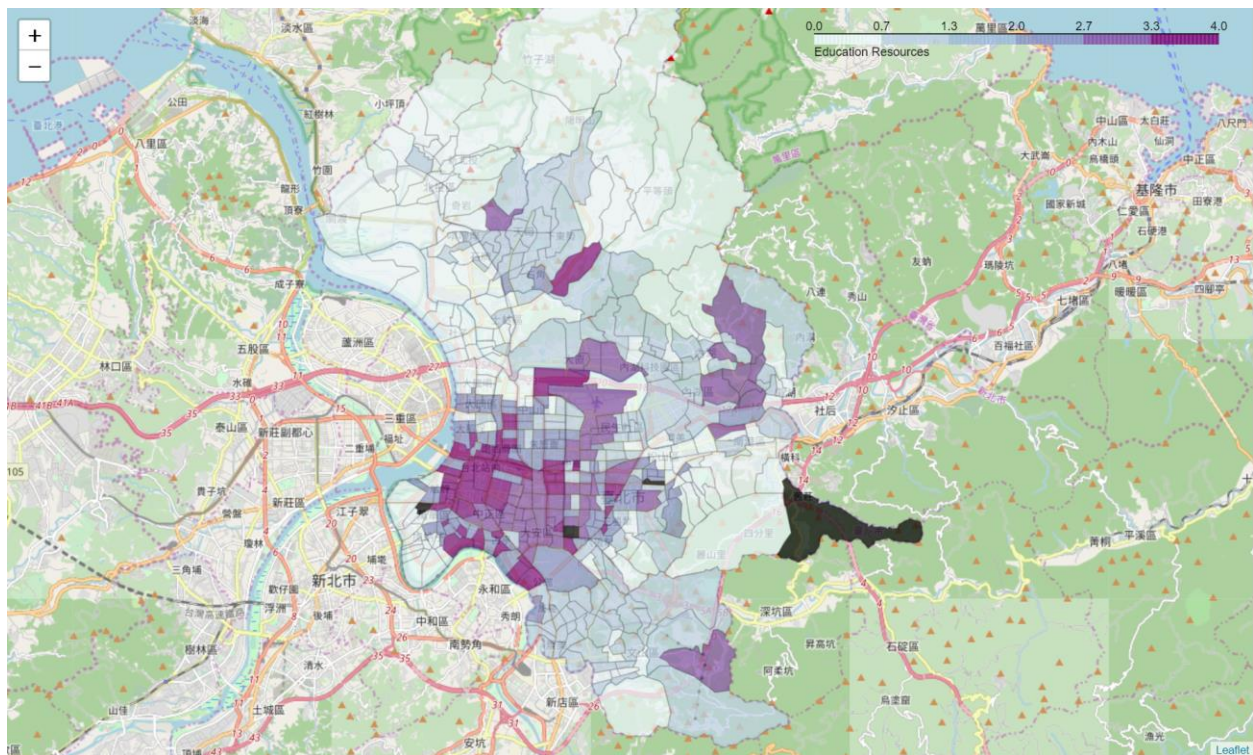


Figure 11. Clusters of Accessible Education Resources per capita under 15, taking household income into consideration

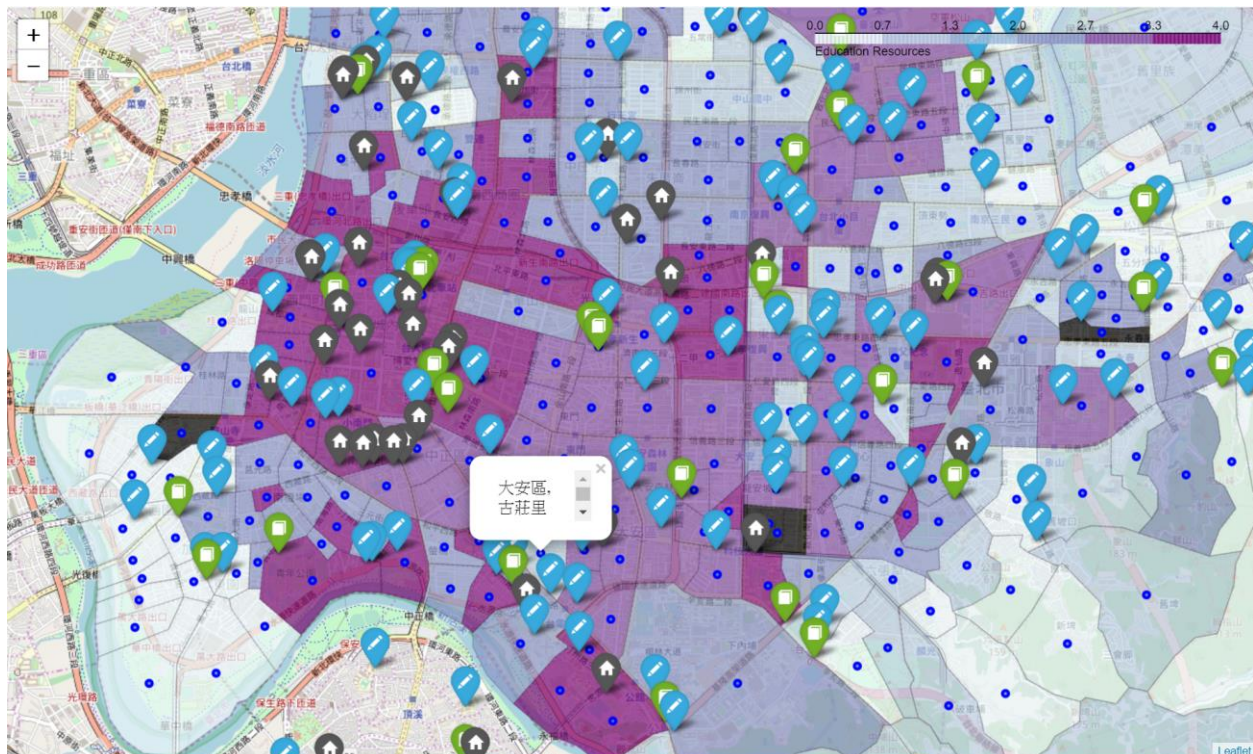


Figure 12. A closer look at how education resources are distributed among different clusters

5. DISCUSSION

5.1 Key Findings

The neighborhoods from the old town have more accessible education resources

As shown in Figure 9, the neighborhoods with higher accessibility to education resources concentrate in the southwestern part of Taipei City, which is the “old town” with the longest development history. It’s not surprising that most education resources are located in the oldest area of a city.

Except the core area, there are three areas that needs further attention. “Tianmu”, located in northern part of Taipei, is where Taipei American School and Taipei Japanese School are at. It is also the place where many foreigners live since after World War II. “Neihu” locates in the geometric central part of Taipei while “Nangang” in central east part. These three areas have both higher household income and under-15 population (see Figure 6 and 7), which have an impact on the analysis.

Some highly populated neighborhoods from the old town show weaker per capita accessibility

From Figure 10 we see the distribution of accessibility has shifted. Not every neighborhoods in the old town enjoy the same level of per capita accessibility. A bunch of high populated neighborhoods share a limited number, though still higher than other places, of resources, while some outer neighborhoods could benefit from a higher quality of education.

Meanwhile, please note that the top 2 cluster have almost the same amount of per capita resources. They are separated into to different clusters because of the strength in different type of education resources. Cluster 4 (deepest color) has significantly higher number of schools than other, and Cluster 3 has the most museums. As we can see neighborhoods in Cluster 4 are scattered in a wider area, those in Cluster 3 still concentrate in the old town.

Considering socio-economic status, some outer neighborhoods show greater accessibility

There are wealthier communities located in outer part of Taipei City. Though there may not exist too many education resources, however, one should take this situation into consideration especially when allocating the government budget for future investment. Certainly we’d like to see the money from the taxpayer can be used more wisely on projects that create the highest social benefit.

Table 14 shows that while Cluster 4 has the highest accessibility, its income only rank the 3rd place. In the contrary, the wealthiest Cluster (Cluster 3) also takes the 3rd place in accessible education resources. Shown in Figure 11, now we can see “Tianmu”, “Neihu” and “Nangang” stand out when we consider

both education and income as variables. These relatively wealthier areas may not demand resources as urgent as Cluster 0, which has both the lowest income and accessibility to education resources.

5.2 Limitation of the Study

The measurement of accessibility is difficult and arbitrary

This study use the distance between venues and the geometric centers of the neighborhood polygons to define the “accessible” education resources. This convenient approach has at least two limitation that need further attention in future studies:

It neglects the distribution of the population. For example, a neighborhood near the waterfront could have its geometric center in the middle of the river. While a big neighborhood could find no venues at all within the radius of 1,500 meters.

It doesn't consider the option of using transportation. There are great public transportation in Taipei City include MRT (the subway), a network of bus routes covering a great area, and even youBike (a bike sharing system). With the help of transportation, the accessibility should be re-define as a combined function of time and cost for using it. Much more resources could be reached by neighborhoods that besides a subway station, which can completely the clustering results.

Different weights should be assigned to education resources

This study doesn't take a careful look at the venues themselves but only the categories that Foursquare assigns to. This could create issues in different ways:

The weights between different categories: Should different categories have the same weight? Can a community library from 1,200 meters away have the same impact on an 8-year-old kid as does the elementary school from 100 meters?

The weights within the same category: Should National Palace Museum, which has a permanent collection of nearly 700,000 pieces of ancient Chinese imperial artifacts and artworks, share the same weight as an art gallery that displays private collections?

6. CONCLUSION

Starting from the care of education, this study aims at providing a solid support for the government for the allocation its resources. By giving a definition to the terms, collecting and cleaning the data, and applying the data into a machine learning model (K-means Clustering), this study successfully identifies

the relatively strong and weak neighborhoods in accessibility to education resources. A cluster of weakest neighborhoods in both accessible resources and household income will need further attention for the government's policy making. Furthermore, this study has its limitation which one should bear in mind during reading it. This limitation needs to be refined in future studies to provide more thorough findings.