Comparing Manhattan and Wan Chai, the top two richest places in cities with the most billionaires in the world, in terms of venues.

REPORT

Bationo Amos

September 24, 2020

1. Introduction

1.1. Background

The choice of a place to live might depend on factors such as safety, school, job, house cost, life cost, venues etc. In this project, the interest will be on last factor, venues.

More than 26% of the world's 2,095 billionaires live in just 10 cities on three continents. In 2020, the two first cities with the most billionaires are New York City, USA (**92**) and Hong Kong, China (**71**) (Forbes).

Manhattan in New York (USA) and Wan Chai in Hong Kong (China) are different in many ways. Indeed, they are located on two different continents separated by about 10,000 km. Added to this is the culture which is just as different and specific to each of them. From all these differences may result different type of neighborhoods.

1.1. Problem

What Manhattan and Wan Chai have in common is that they both are the richest place of the richest town of their respective country. Knowing all that, One might wonder what kinds of venues are in these two places? How similar are they in terms of venues? These are the question this project will attempt to answer.

1.2. Interest

This project will provide useful information to people who are willing to live in one of these two cities by helping to make a choice that fit them. It will also allow people who want to create a business (restaurant, shop ,etc) to identify the competitors and have an idea on which kinds of business are common in the two wealthiest place of the world's two first cities where billionaires decide to live.

2. Data acquisition and cleaning

2.1. Data sources

The data that will be used in this project will come from mainly four sources:

- wikipedia for Hong Kong's district list
- thttps://cocl.us/new_york_dataset for the list of New York neighborhoods
- Open Street Map geocoging library for python: Nominatim to transform districts and neighborhoods addresses into geographic coordinates (latitude and longitude)
- Foursquare to get venues list (not the complete list) in each districts and neighborhoods
- GADM to get the polygons of the district of Wan Chai

These sources have been chosen because of the data availability and their costs. Indeed, expect Foursquare, all the others sources are free. Fortunately, despite of some limitations the regular calls using the free API key of Foursquare provide quite enough data for this project.

For Foursquare, the response of the queries will be in JSON(JavaScript Object Notation) format combining *dictionaries* and *lists*. For each response object stored in a variable called *results*, venues list will be accessible through **results['response']['groups'][0]['items'].** For each venue results['response']['groups'][0]['items']**['venue']** , the features that will selected are

- **Venue name**: ['name']
- **Venue latitude**: ['location']['lat']
- **Venue longitude**: ['location']['lng']
- **Venue category**: ['categories'][0]['name']

2.2. Data cleaning

After cleaning, the data about New York neighborhoods will be similar to the one in the following figure

Figure 1: New York neighborhoods

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

and data about Hong Kong's districts provided by Wikipedia will look like *Figure 2*.

Figure 2: Wan Chai districts

	District	Chinese	Population[6]	Area(km2)	Density(/km2)	Region
0	Central and Western	中西區	244600	12.44	19983.92	Hong Kong Island
1	Eastern	東區	574500	18.56	31217.67	Hong Kong Island
2	Southern	南區	269200	38.85	6962.68	Hong Kong Island
3	Wan Chai	灣仔區	150900	9.83	15300.10	Hong Kong Island
4	Sham Shui Po	深水埗區	390600	9.35	41529.41	Kowloon

As we can see there are no predefined neighborhoods for the districts of Hong Kong, therefore I will generate some custom neighborhoods using the polygon districts got from GADM, specifically district of Wan Chai, and Nominatim library in order to set a name to each new neighborhood.

After cleaning polygon data got from GADM the table of polygons will looks like the following one.

Figure 3: Polygons of Hong Kong districts

	Name	geometry_type	geometry_coordinates
0	Central and Western	MultiPolygon	[[[[114.16697692871094, 22.281528472900447], [...
1	Eastern	MultiPolygon	[[[[114.17791748046886, 22.283470153808594], [...
2	Islands	MultiPolygon	[[[[113.93291473388672, 22.324028015136832], [...
3	Kowloon City	MultiPolygon	[[[[114.20491790771484, 22.334751129150447], [...
4	Kwai Tsing	MultiPolygon	[[[[114.14892578125011, 22.380218505859432], [...

Where the field **Name** contains all the district of Hong Kong.

3. Methodology

In this project, will focus on the similarity of Manhattan and Wan Chai by analyzing their neighborhoods. Note that the neighborhoods of these cities will not been merged, they will remains distinct for the study. Note that since there are not predefined neighborhoods for Wan Chain, we have generated some personalized neighborhoods using the bounds of the district polygon.

Assuming we have required data (venues and their location of each neighborhood and their location for both places), the first step will be to extract the venue category for each neighborhood and transformed into the columns of a data frame then used further as features for clustering models. For each neighborhoods the first 10th common venues will be gathered with the mean of the frequency of their occurrence and saved into a data frame.

In the second step, **kmeans** cluster models will built with different number of cluster for each the two cities, then the best number of clusters will be chosen based on **silhouette scores**.

Finally, for each place, each cluster will be examined and labeled based on their most common components and comparison will be made between the clusters of different place to highlight the similarities and differences.

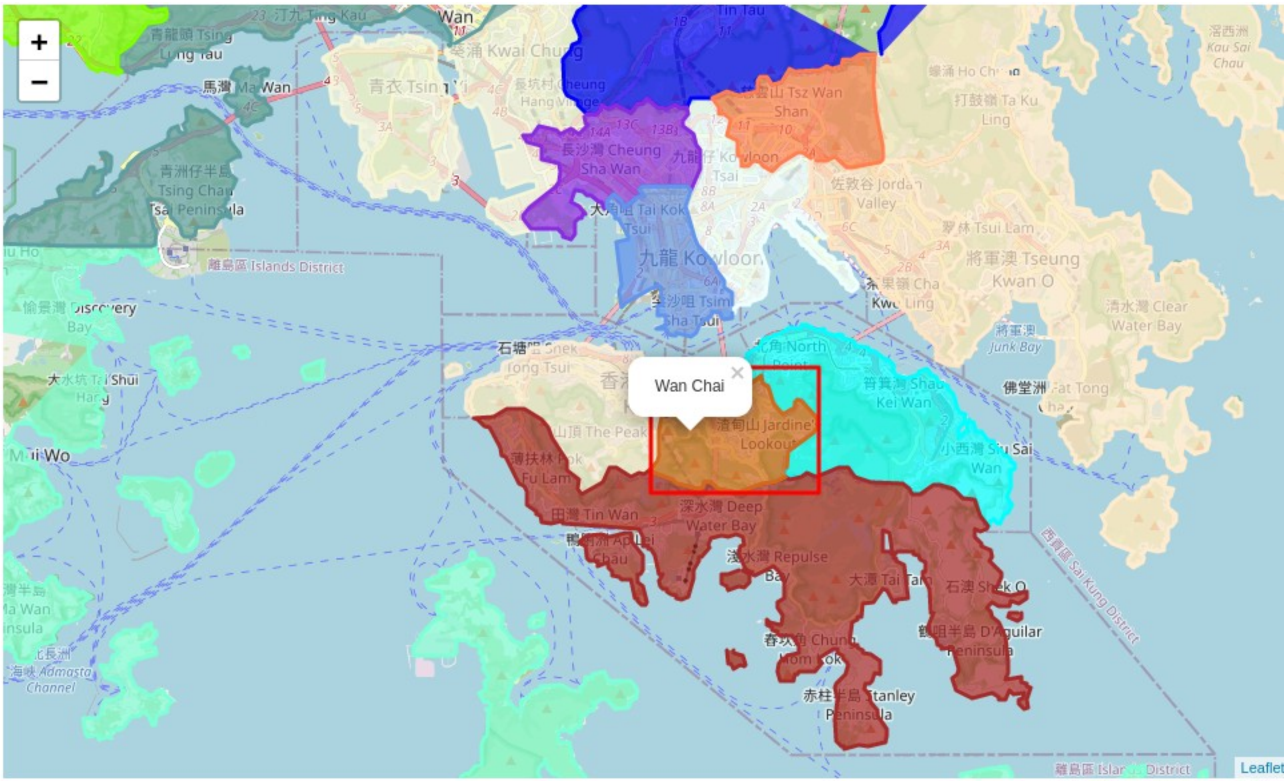
4. Analysis

4.1. Wan Chai custom neighborhoods

In order to create our own neighborhoods for the district of Wan Chai, I will use its polygon provided by GADM.

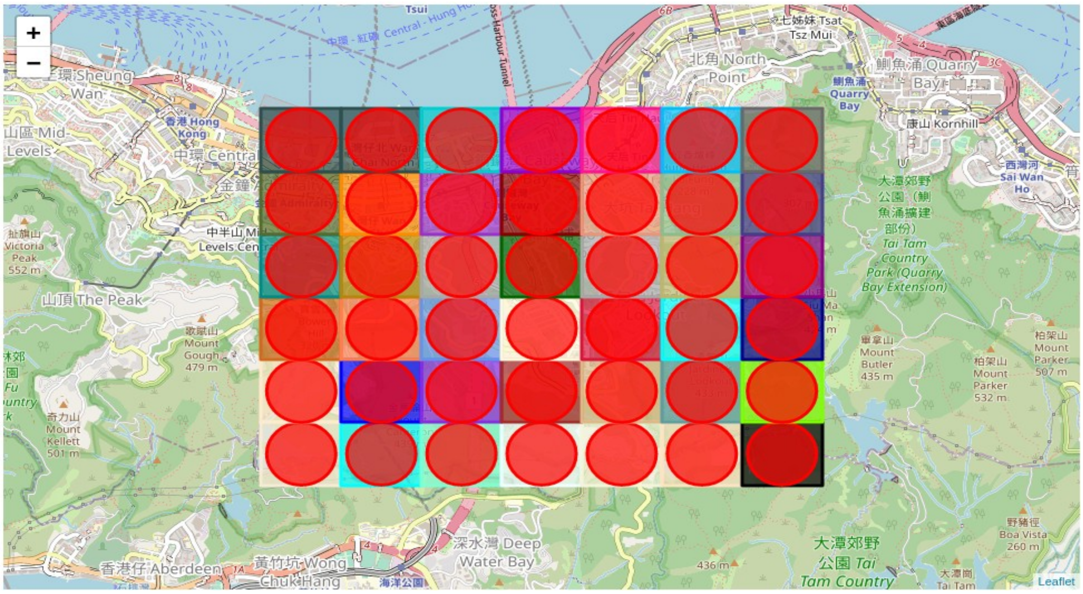
First of all I will draw a rectangle around the Wan Chai. That rectangle should be smallest as possible but also big enough to contain the district of Wan Chai. To do that I will merely draw the rectangle using the bounds of the Wan Chai polygon. That can be visualized in the following figure.

Figure 3: District of Wan Chai



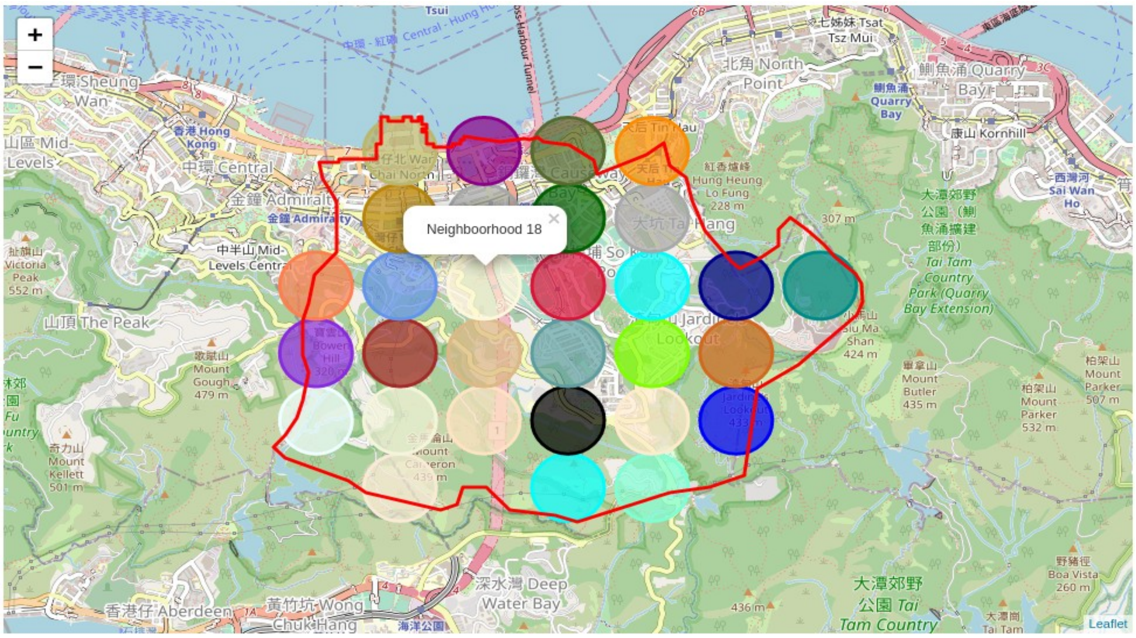
The second step will be to subdivide rectangle into other small ones. I arbitrarily chose 6x7 rectangles, which are 42 equal rectangles in total. Next, I draw a circle in each small rectangle. Each circle should be big enough but none of them should cross the boundaries of any small rectangles. The result can seen on the following map.

Figure 4: Subdivisions of Wan Chai district



The final step of creating neighborhoods of Wan Chai was to remove all circles those center is outside the district of Wan Chai then the remaining circles (30 circles) have been considered as neighborhoods. The result is in the figure below.

Figure 5: Customs neighborhoods of Wan Chai district

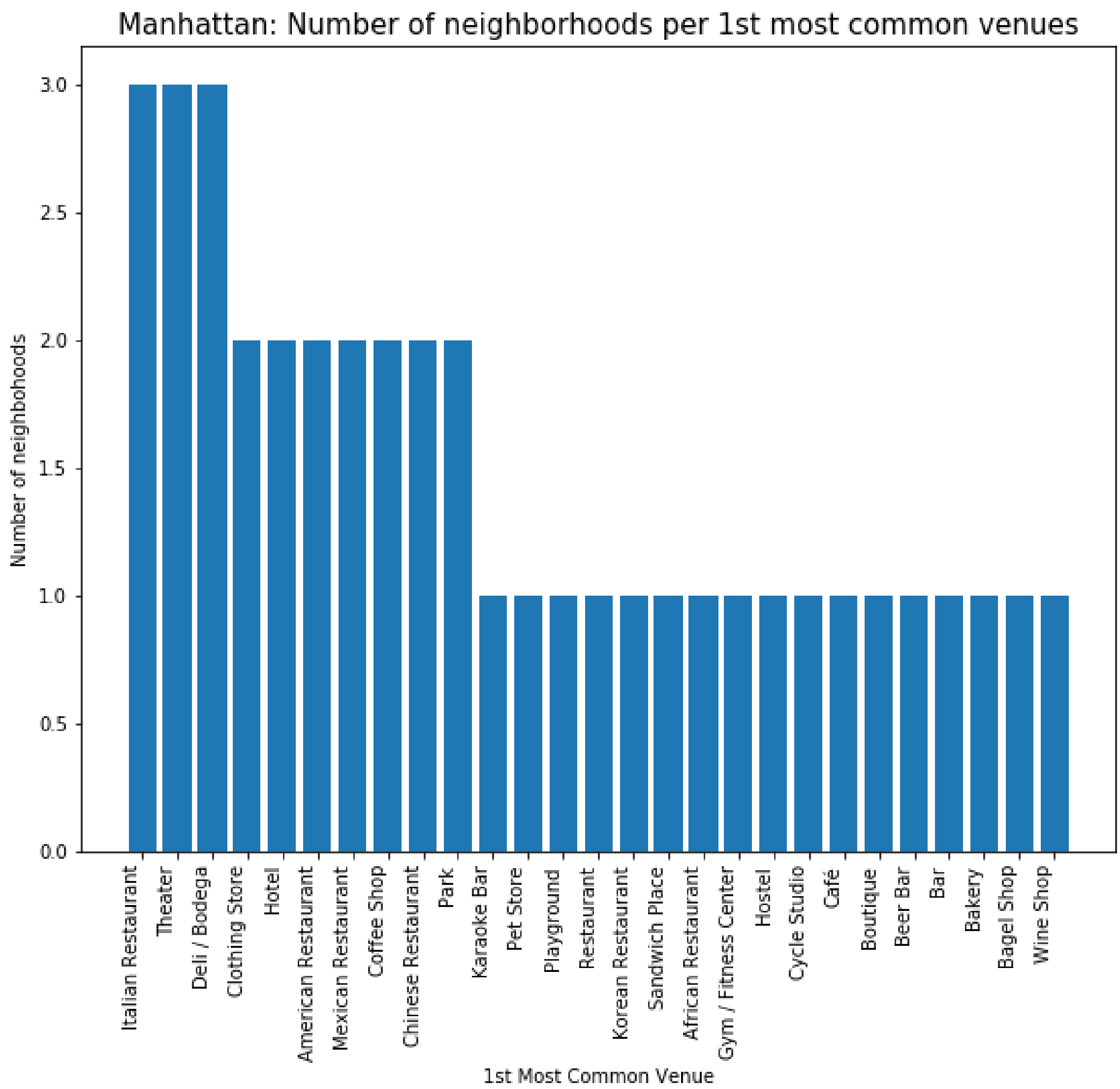


One remaining problem is the fact not all the circles are completely inside the boundaries of the district, some of them spread out outside the boundaries. In order to avoid considering venues that are located outside Wan Chai will just be ignored.

4.2. Analyze Manhattan Neighborhoods

After getting the first 10th most common venues of each neighborhoods of Manhattan it's possible to group the neighborhoods per ***1st most common venues*** and found out which "*1st most common venues"* are/is the most common venues among Manhattan neighborhoods. From the figure bellow we can see that most of the neighborhoods In Manhattan, **Italian Restaurant**, **deli / Bodega** and **theater** are the **1st Most Common Venue**.

<u>*Figure 6*</u>*: Manhattan: Number of neighborhoods per 1st most common venues*

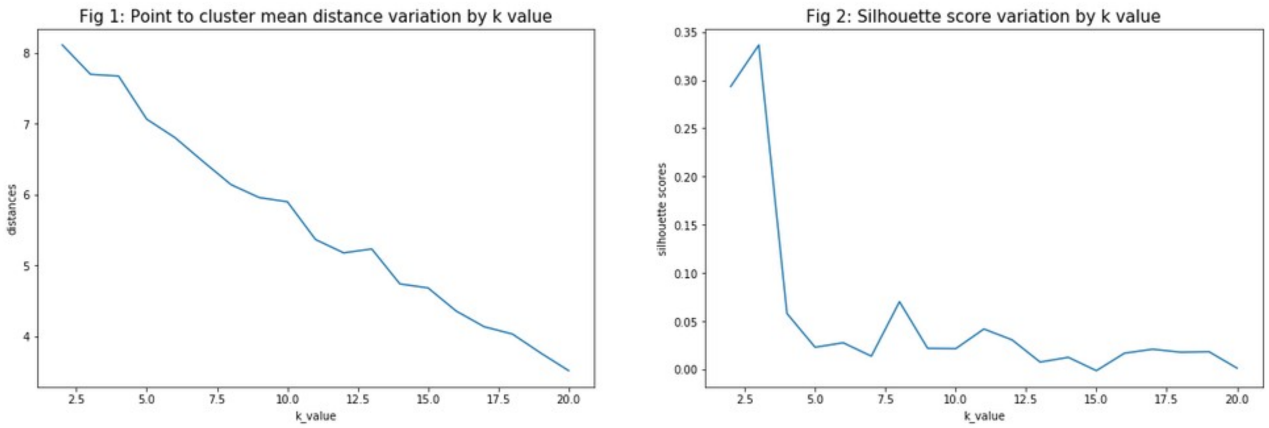


Foursquare has returned 267 venues categories for the 40 neighborhoods in Manhattan. Clustering using the kmeans algorithm first requires choosing the number of clusters. To do that, the model has been fitted iteratively while varying the value of clusters from 2 to 20 then the best number of clusters is the one that produces best fitting. With the **elbow** method, the best k is found where the minimum *mean distance within cluster* starts becoming sharply constant, like an *elbow* . But, we do not always have an enough clear and sharp *elbow* on the graph to make a decision with elbow method.

Unlike **elbow** method, **silhouette** method can often bring more clear information. Indeed, with this method the best k value is the first one that has the highest silhouette score. According to Wikipedia the silhouette value measures how similar an object (here a neighborhood) is to its own cluster (cohesion) compared to other clusters (separation).

From the figure bellow, the left graph the *mean distance within cluster* is not sharp enough to use the elbow method; but for silhouette variation on the right graph the maximum *silhouette score* value is clearly identifiable at *k* equals to 3 which the best number of clusters for Manhattan neighborhoods.

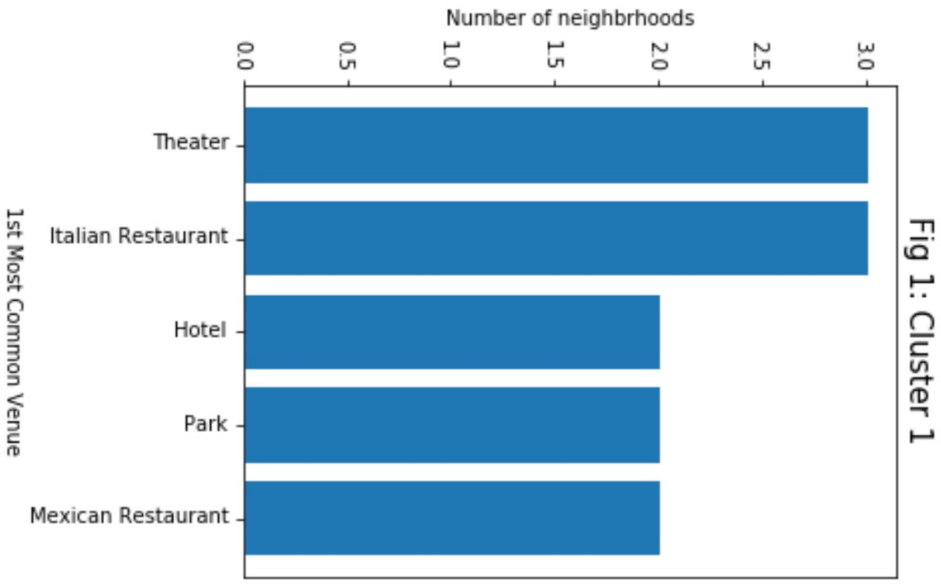
Figure 7: Manhattan: Mean distance within cluster and silhouette score per k value

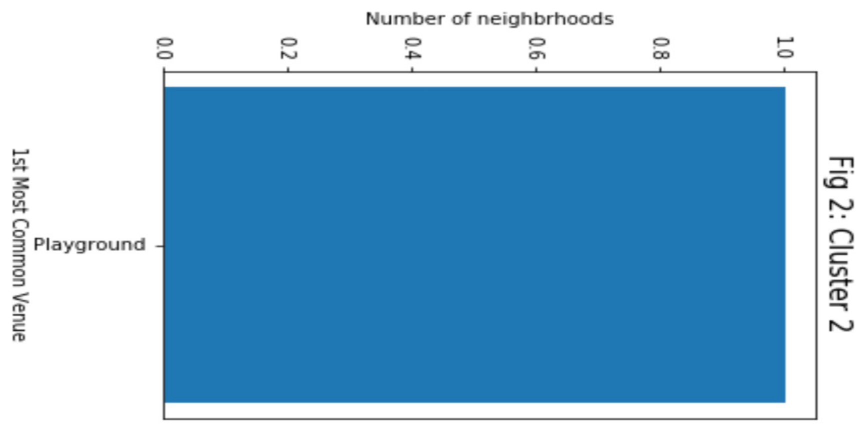


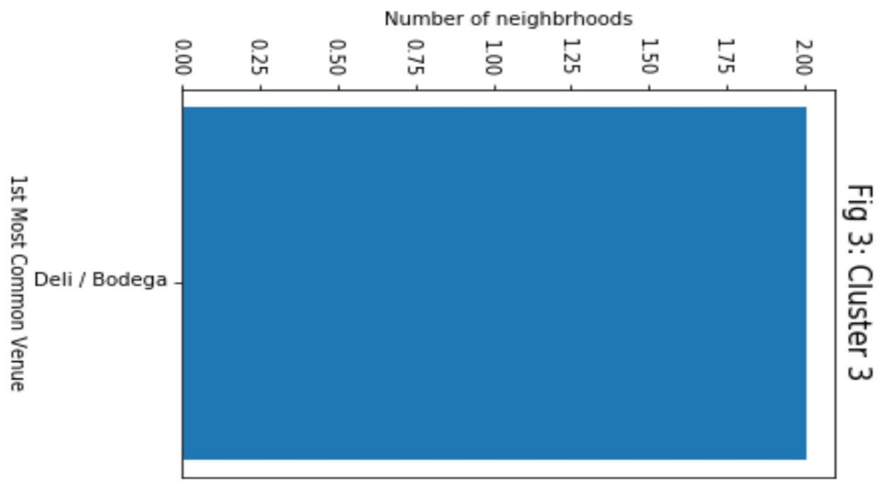
After clustering Manhattan neighborhoods into 3 clusters we can label those clusters based on their components (see Figure 8). Based their components, the following labels can be suggested for each clusters:

- **Cluster 1**: Restaurants and spare times
- **Cluster 2**: Entertainment
- **Cluster 3**: Shop

Figure 8: Manhattan: Neighborhoods clusters components

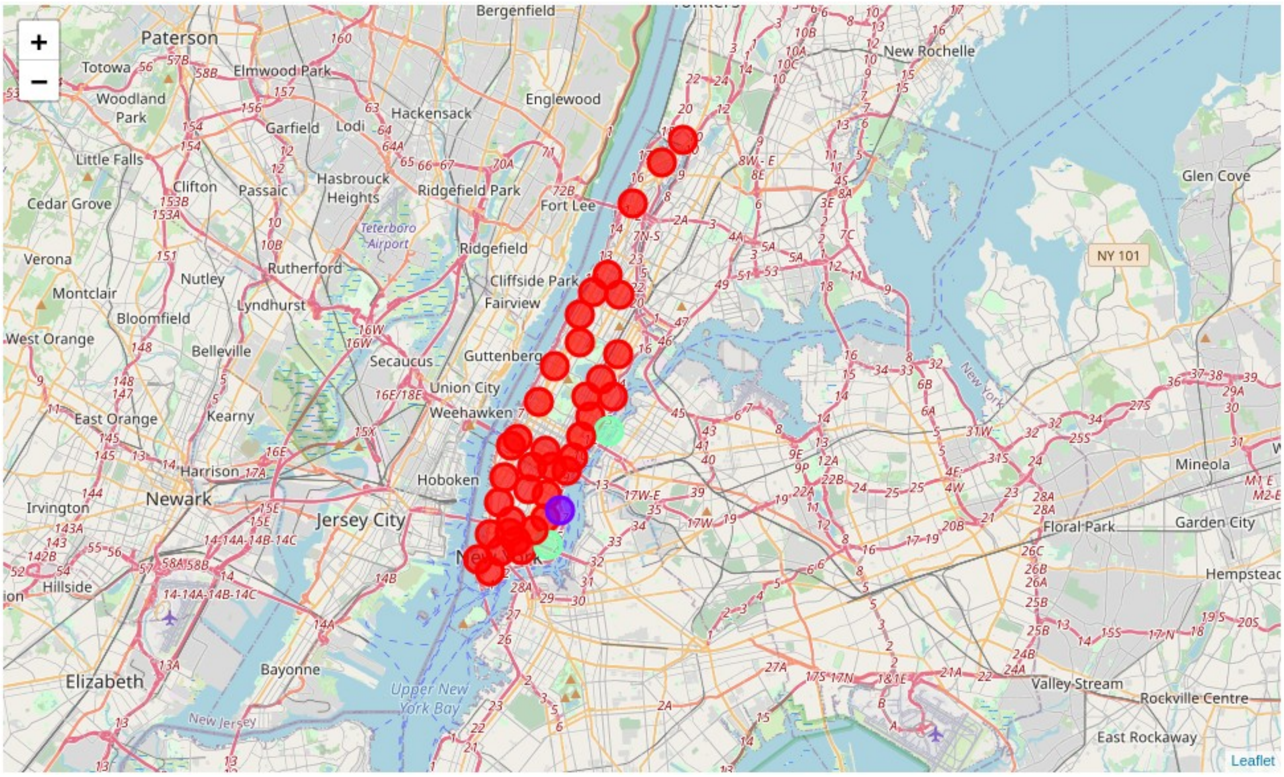






The following figure represents each neighborhood in circles and the color represents the cluster to which they belongs. We notice that *cluster 1* (in red) representing ***restaurants and spare times*** is the most common among the neighborhoods of Manhattan.

<u>Figure 9</u>: Manhattan: Clusters of Neighborhoods



4.3. Analyze Wan Chai Neighborhoods

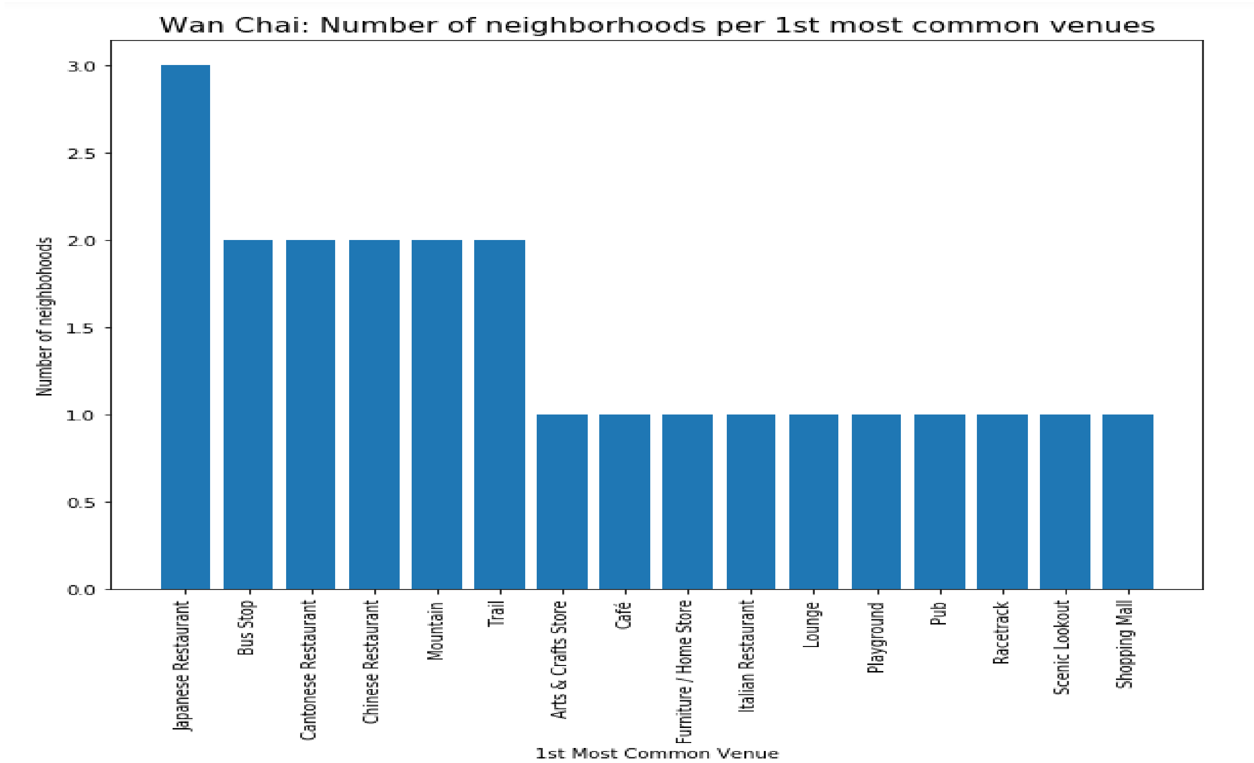
As for the district of Wan Chai, the same analyzes as for the Borough of Manhattan were carried out.

Referring to *Figure 10* it appears that **Japanese Restaurant** is the most common ***first common venues*** among the Wan Chai neighborhoods. In the other hand, silhouette scores variation of *Figure 11* shows that 5 is best number of clusters for neighborhoods of Wan Chai.

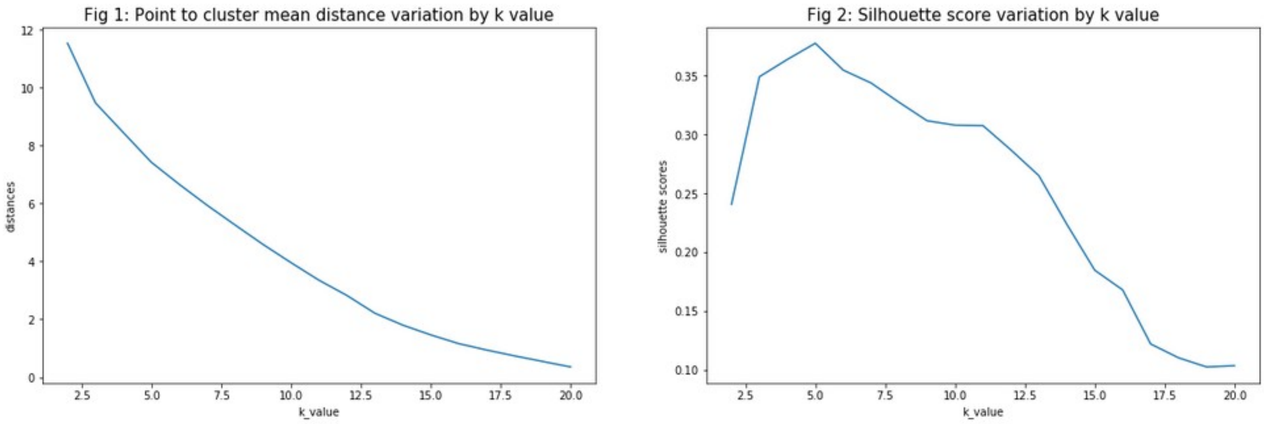
After clustering, the 5 clusters of neighborhoods was labeled as following:

- Cluster 1 (Japanese restaurant, Cantonese restaurant, Chinese restaurant, Arts and crafts store, bus stop): **Asian restaurants**, **transport** and **manual work**
- Cluster 2 (Scenic Lookout): **Observatory**
- Cluster 3 (Mountain): **Trip**
- Cluster 4 (Bus Stop): **Transport**
- Cluster 5 (Trails): **Trail**

<u>Figure 10</u>: Wan Chai: Number of neighborhoods per 1st most common venues

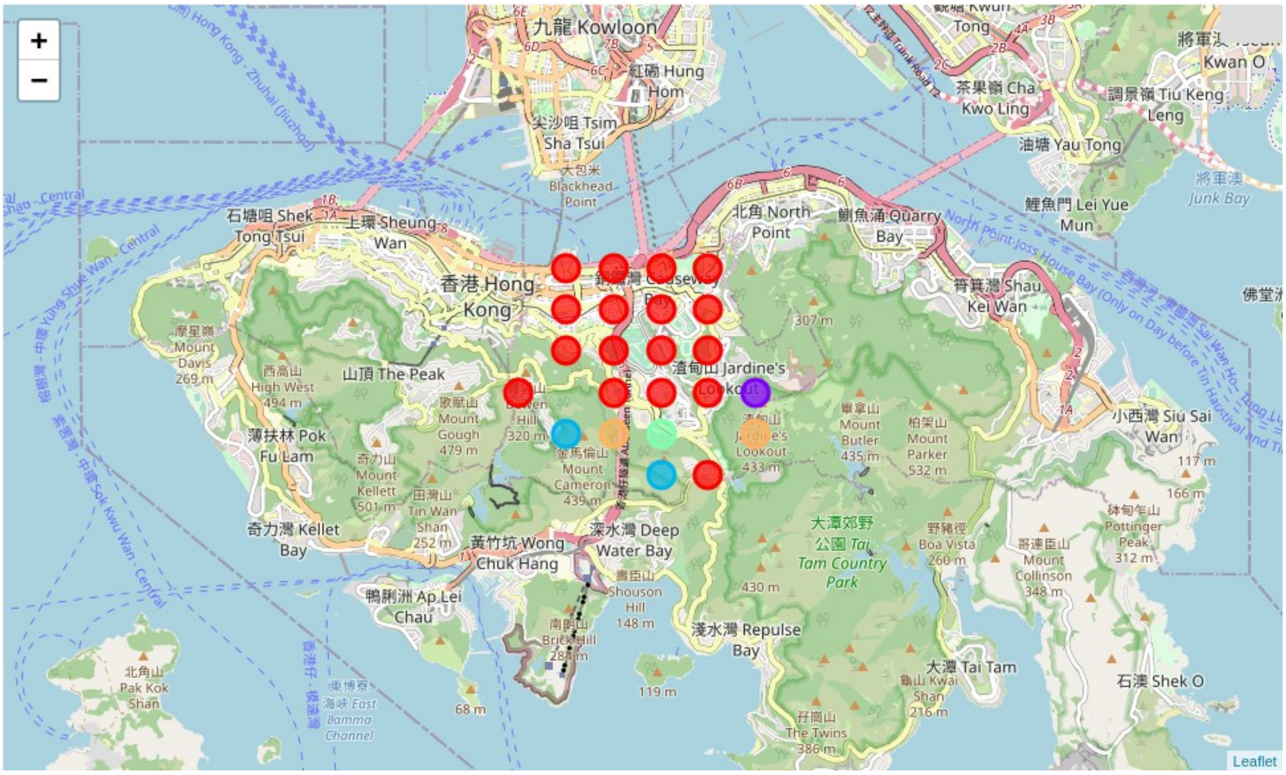


<u>Figure 11</u>: Wan Chai: Mean distance within cluster and silhouette score per k value



Finally, the clusters of Wan Chai neighborhoods can be visualized on the following map. Each color represents a specific cluster. From the map we can notice that the cluster of **Asian restaurants**, **transport** and **manual work** (cluster 1) is the most common among the neighborhoods of Wan Chai.

<u>Figure 12</u>: Wan Chai: Clusters of Neighborhoods



5. Results and Discussion

From this analysis, it has been shown that Manhattan has more venues (registered in Foursquare database) than Wan Chai. If in fact, Fourquare's users in Manhattan are more active than those in Wan Chai, this can explain the low number of venues in that Hong Kong district. But, others things may explain this high difference in venues number. For example Wan Chai is known to have a significant part of its area covered by vegetation reducing the space for building substructures creating more type of venues. The size of Wan Chai may also be a reason.

We can also notice that, despite the high number of neighborhoods in Manhattan (40) the best number of clusters found using silhouette scores was 3 while the on for Wan Chai was 5 although its had less than 30 neighborhoods while clustering. Meaning that the neighborhoods that we have created are more heterogeneous than those of Manhattan in terms of venues.

Looking at most common clusters of Wan Chai and Manhattan, it comes that restoration is the most common type of venues in these two city (all specification combined). But, a deeper look shows that in **Asian restaurants** are the most common type of venue (probably due to their culture), while in Manhattan **theaters** and I**talian restaurant** are the leading venues.

One more interesting fact about the most common clusters of these two cities is in structure of clusters. Indeed, in Manhattan, the most common cluster of neighborhoods has as type of venues **theater**, I**talian restaurant**, **hotel**, **park** and M**exican restaurant** which I have labeled to **restaurants** and **spare times**. Now, considering the most common cluster of neighborhood in Wan Chai, it is made of **Japanese restaurant**, **Cantonese restaurant**, **Chinese restaurant**, **arts and craft store** and **bus stop** which we have labeled as **Asian restaurants**, a bit **transport** and **manual work**. This choice of labeling may go along with thought that in this part of Asia people spend more time in working.

6. Conclusion

Purpose of this project was, to identify, beyond their cultures, a potential similarity of dissimilarity between the neighborhoods of Manhattan and Whan Chai. Except the fact these two places are the wealthiest of their respective cities (Manhattan in New York and Whan Chai in Hong Kong) and that those two cities count the highest number of billionaires in the world, the culture may generate a difference even in their neighborhoods.

It has been shown that despite a relative slightly difference in their respective most common cluster, on thing common to them is restoration. Providing and getting food remains the most common activities for people living in this two parts of the world.

Now, for a investor looking for what to do (create a venue as a business) the following two recommendations may useful. First if that person chooses to Manhattan for its business, an Asian restaurant could be a great choice since one, restoration is already doing well, and two, there are few meaning less competition. If that person chooses Wan Chai, Italian or American restaurants will be a good choice for the same reasons as for the choice of Manhattan. Finally, for people who are switching from one of these two places to the other, they will be able to find a neighborhood which will have some similarities with where tey came from so that the adaptation process will go smoothly.

Also, since there was no entertainment venues in the most common cluster of Wan Chai, it may be interesting to do another research to check whether the type venues around in a particular place may affect people working rhythm. If yes, how?

References

- World's Richest Cities: The Top 10 Cities Billionaires Call Home 2020, (2020, April 09). Retrieved September 05, 2020, from https://www.forbes.com/sites/giacomotognini/2020/04/07/worlds-richest-cities-the-top-10-cities-where-most-billionaires-call-home-2020/#d82ea6f3101a
- Districts of Hong Kong. (2020, September 22). Retrieved September 05, 2020, from https://en.wikipedia.org/wiki/Districts_of_Hong_Kong
- New York neighborhoods. Retrieved September 05, 2020, from thtps://cocl.us/new_york_dataset

- Open-source geocoding: Nominatim. Retrieved September 05, 2020, from https://nominatim.org/
- Foursquare api. Retrieved September 05, 2020, from https://developer.foursquare.com/docs/places-api/
- Countries geo data. Retrieved September 05, 2020, from https://gadm.org/download_country_v3.html