

The Battle of Neighborhoods: Tokyo versus Ho Chi Minh

Phat Nguyen

April 20, 2020

A. Introduction

A.1 Description & Discussion of the Background

Located in South-East Asia, Vietnam is a developing country with a large, young population of **95 million people**. The country also has an interesting S-shape, which can be seen on Google Maps. Ho Chi Minh city lies at the bottom of the S, and it is the financial centre and the most populous city of Vietnam. With its fast-growing economy and energetic labor force, Ho Chi Minh city has become an attractive location for foreign investment.

Famous for its very high cost of living, Tokyo, capital of Japan on the other hand, is one of the most populous and economically advanced metropolises in the world. The city attracts millions of tourists every year, as well as workers and students who seek for better career in a developed country. There are many advantages of living in Tokyo, such as convenient public transportation, superior education or a relatively safe environment.

As a business student who has experienced life in both cities, there are certain aspects that I both love and hate about the lifestyle in each city. However, I would not have the confidence to make an opinion about the differences of Ho Chi Minh and Tokyo because of the lack of relevant information. Using the knowledge that I obtained from the IBM Data Science Professional Certificate courses, I decided to explore my two favorite cities on a deeper level, compare and contrast the result, by using data and statistics.

B. Data Description

- A data of all Third-level Administrative Divisions of the Ho Chi Minh and their respective populations can be found on the Ho Chi Minh government homepage. <http://www.pso.hochiminhcity.gov.vn/web/guest/niengiamthongke-nam2010-dansovalaodong>
- From the Japanese government statistics website, I extracted municipality data of Tokyo with selected item of Total population and got the data of all neighborhoods in Tokyo along with their population. <https://www.e-stat.go.jp/en/regional-statistics/ssdsview/municipality>
- I cleaned the two data sets and got two dataframes with each dataframe contains two columns of Neighborhood and Population.
- I then used **Google Geocoding API** to get coordinates of all neighborhoods in Tokyo and Ho Chi Minh.
- I used **Foursquare API** to get the most common venues of each neighborhood.

C. Methodology

After the data cleansing process and having utilized Google Geocoding API, I created two master data sets respective to each city. This is how they look like:

	Neighborhood	Population	Latitude	Longitude
0	Adachi	670122	35.775017	139.804413
1	Bunkyo	219724	35.707869	139.752437
2	Chiyoda	58406	35.694031	139.753772
3	Taito	198073	35.712574	139.780204
4	Shinagawa	386855	35.609197	139.730336

Tokyo

	Neighborhood	Population	Latitude	Longitude
0	District 1	187435	10.775659	106.700424
1	District 2	140621	10.787273	106.749810
2	District 3	188945	10.784370	106.684409
3	District 4	183261	10.757826	106.701297
4	District 5	174154	10.754028	106.663375

Ho Chi Minh

Using the Foursquare API as the main search machine to explore each city, I set the limit as **100 venue** and the radius **500 meter** for each API request. Here is a part of the merged data that was returned from Foursquare API.

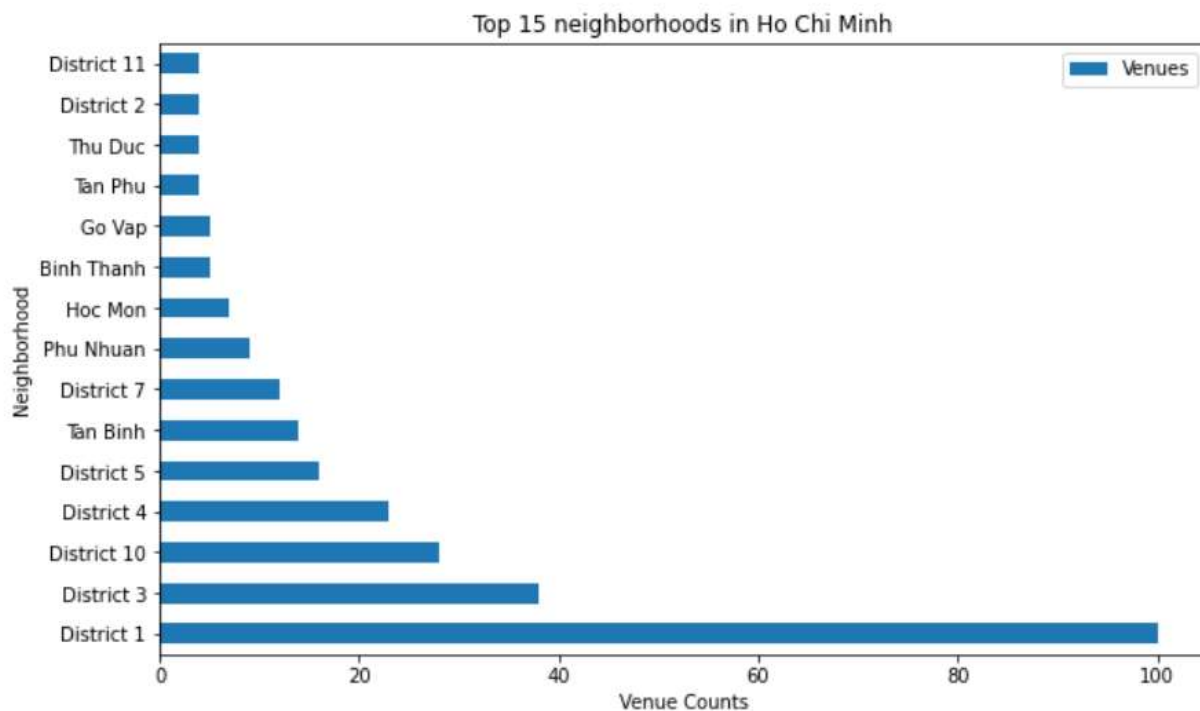
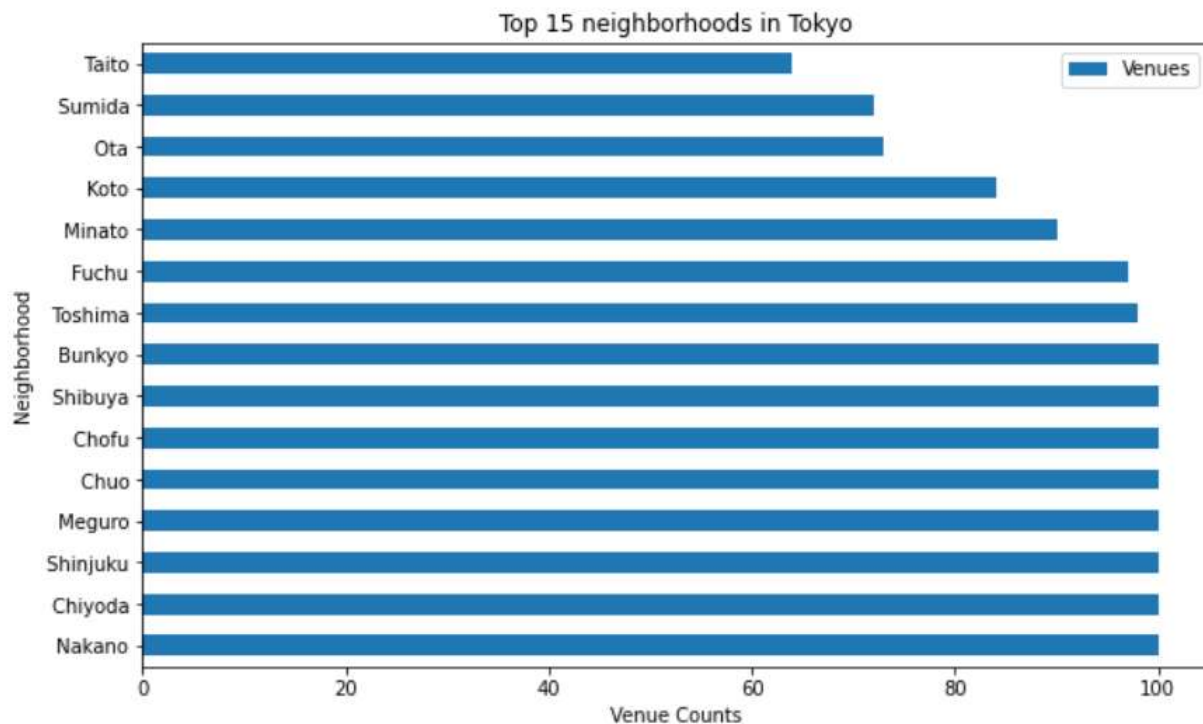
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adachi	35.775017	139.804413	MEGA Don Quijote (MEGAドン・キホーテ 環七梅島店)	35.778288	139.804967	Discount Store
1	Adachi	35.775017	139.804413	Nitori (ニトリ 環七梅島店)	35.778259	139.803108	Furniture / Home Store
2	Adachi	35.775017	139.804413	Ikinari Steak (いきなりステーキ)	35.777730	139.802890	Steakhouse
3	Adachi	35.775017	139.804413	セブンイレブン 足立梅島2丁目店	35.777031	139.803106	Convenience Store
4	Adachi	35.775017	139.804413	レストラン メヒコ 足立区役所14階	35.774582	139.803969	Restaurant

Tokyo Venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	District 1	10.775659	106.700424	Pasteur Street Brewing Company	10.775220	106.700894	Brewery
1	District 1	10.775659	106.700424	Liberty Central Saigon Citypoint Hotel	10.774758	106.700795	Hotel
2	District 1	10.775659	106.700424	B3 - Steakhouse & Craft Beer	10.775190	106.702492	Steakhouse
3	District 1	10.775659	106.700424	The Old Compass Cafe	10.774816	106.700685	Café
4	District 1	10.775659	106.700424	O Lé	10.774772	106.699524	Spanish Restaurant

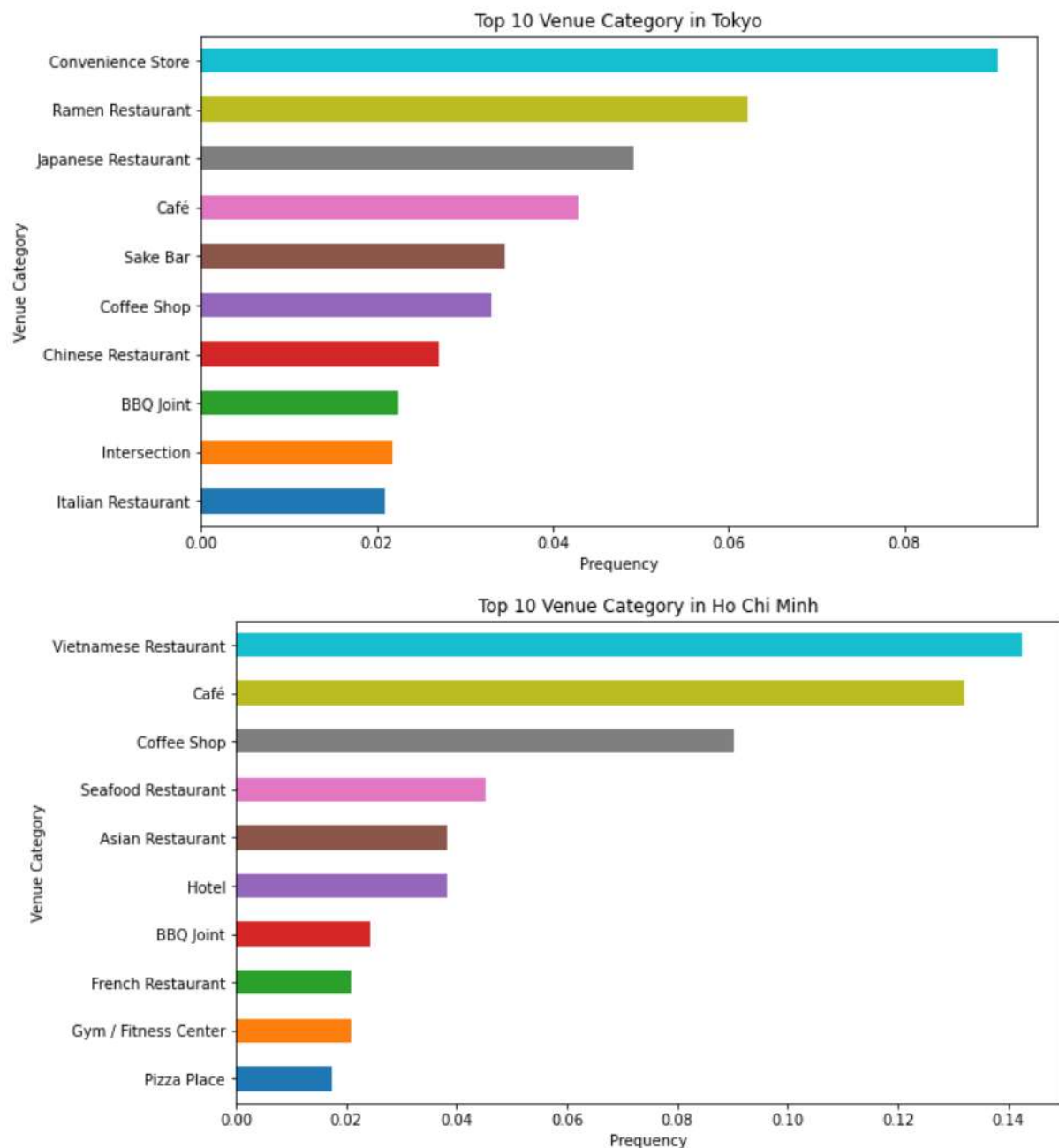
Ho Chi Minh venues

Now, to see how much more advanced the Tokyo's economy compared to Ho Chi Minh's, I used **horizontal bar charts** to visualize the top 15th busiest neighborhoods in each city based on their total venues:



We can clearly see the difference between the two cities' **economy size** by looking at those two charts. Tokyo has eight neighborhoods that reach the limit of 100 venues, while Ho Chi Minh only has one which is District 1. We can have a rough estimation that Tokyo's economy is **at least** 8 times larger in size than Ho Chi Minh's.

Next, to understand the difference in **characteristics of venues** in each city, I listed out ten most common venues in each city and used horizontal bar charts again for better visualization of the data.



We can see that some of the unique keywords for each city would be:

- **Tokyo:** “convenient store”, “ramen restaurant”, “Japanese restaurant”, “sake bar”, “café/coffee shop”, “Chinese restaurant”
- **Ho Chi Minh:** “Vietnamese restaurant”, “Café/Coffee shop”, “Seafood”, “Hotel”, “

We can see that there are some patterns in those keywords that will help us distinguish the lifestyle in each city. For a deeper understanding, I used unsupervised learning **K-means algorithm** to cluster the neighborhoods and study how venues are grouped in each city. K-Means algorithm is one of the most common and easy to understand Machine Learning clustering algorithm.

As choosing too large K would make it harder to analyze both data sets at the same time, I chose a small $K = 3$ clusters for each city. Using the data that I got from Foursquare API, I created a tablet with top 10 most common venues of each neighborhood. Next, I merged the table with the cluster labels, and this is how they look like:

	Neighborhood	Population	Latitude	Longitude	Cluster Labels	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	
0	Adachi	670122	35.775017	139.804413	0	Convenience Store	Bus Stop	Park	Intersection	Grocery Store	Kids Store	Bakery	Diner	R
1	Akiruno	80954	35.728590	139.293466	0	Ramen Restaurant	Convenience Store	Pharmacy	Chinese Restaurant	Steakhouse	Sushi Restaurant	Thrift / Vintage Store	Intersection	E
2	Akishima	111539	35.705910	139.354016	0	Convenience Store	Supermarket	Ramen Restaurant	Zoo Exhibit	French Restaurant	Fishing Store	Flower Shop	Food	
3	Aogashima	178	32.456994	139.765208	1	Campground	Sauna / Steam Room	Scenic Lookout	Historic Site	Park	Food & Drink Shop	Fish Market	Fishing Store	
4	Arakawa	212264	35.735934	139.783476	0	Chinese Restaurant	Convenience Store	BBQ Joint	Intersection	Discount Store	Memorial Site	Park	Grocery Store	

Tokyo Merged

	Neighborhood	Population	Latitude	Longitude	Cluster Labels	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 Common Venue	
0	District 1	187435	10.775659	106.700424	1	Coffee Shop	Hotel	Vietnamese Restaurant	Café	French Restaurant	Asian Restaurant	Bar	Burger Joint	Mu	
1	District 2	140621	10.787273	106.749810	1	Restaurant	Vietnamese Restaurant	Snack Place	Whisky Bar	Flower Shop	Dessert Shop	Dim Sum Restaurant	Dog Run	Ei Euro Reste	
2	District 3	188945	10.784370	106.684409	1	Vietnamese Restaurant	Coffee Shop	Café	French Restaurant	Asian Restaurant	Seafood Restaurant	Rock Club	Juice Bar		
3	District 4	183261	10.757826	106.701297	1	Seafood Restaurant	Café	Vietnamese Restaurant	Gym / Fitness Center	Snack Place	Mexican Restaurant	Clothing Store	Pizza Place	Playg	
4	District 5	174154	10.754028	106.663375	1	Vietnamese Restaurant	Chinese Restaurant	Dim Sum Restaurant	Café	Asian Restaurant	BBQ Joint	Multiplex	Sporting Goods Shop	F C	

Ho Chi Minh merged

Now to set a proper label for each cluster, maybe we can count the 1st most common venue of each cluster.

- Tokyo:

-----Cluster 1-----

1st Most Common Venue	
Convenience Store	21
Bakery	1
Park	1
Bus Stop	1
Intersection	1
BBQ Joint	1
Drugstore	1
Chinese Restaurant	1
Ramen Restaurant	1
Café	1

-----Cluster 3-----

1st Most Common Venue	
Scenic Lookout	1

- Cluster 1: Convenient venue
- Cluster 2: Food, drinks and entertainment venue
- Cluster 3: Sightseeing venue

- Ho Chi Minh

- Cluster 1: Asian venue
- Cluster 2: Food, drinks venue
- Cluster 3: Flower venue

-----Cluster 2-----

1st Most Common Venue	
Café	5
Ramen Restaurant	5
Convenience Store	5
Japanese Restaurant	3
Intersection	2
Hotel	2
Bus Stop	1
Harbor / Marina	1
Soba Restaurant	1
Sushi Restaurant	1
Zoo Exhibit	1
Gift Shop	1
Plaza	1
Campground	1
Baseball Stadium	1

-----Cluster 1-----

1st Most Common Venue	
Café	2
Asian Restaurant	1

-----Cluster 3-----

1st Most Common Venue	
Flower Shop	1

-----Cluster 2-----

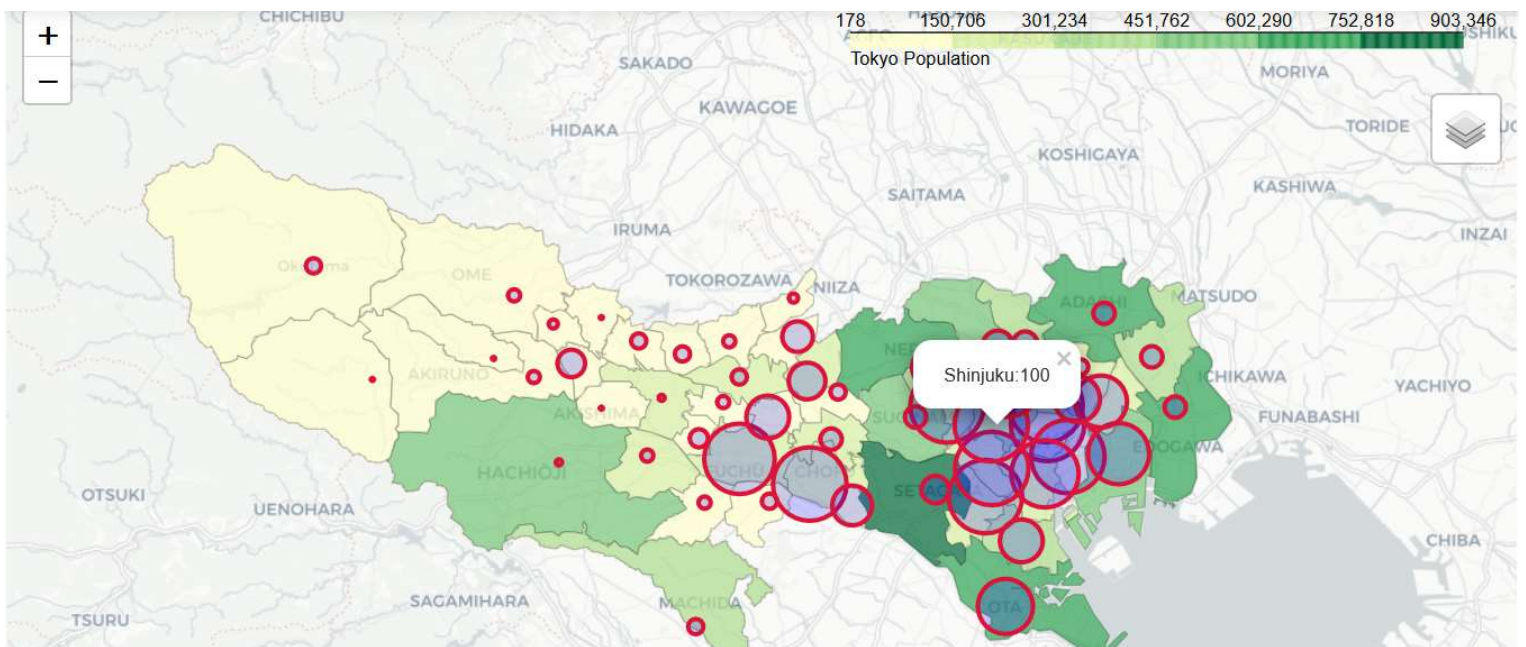
1st Most Common Venue	
Vietnamese Restaurant	8
Seafood Restaurant	3
Café	2
Coffee Shop	2
Restaurant	1
Pizza Place	1
Electronics Store	1

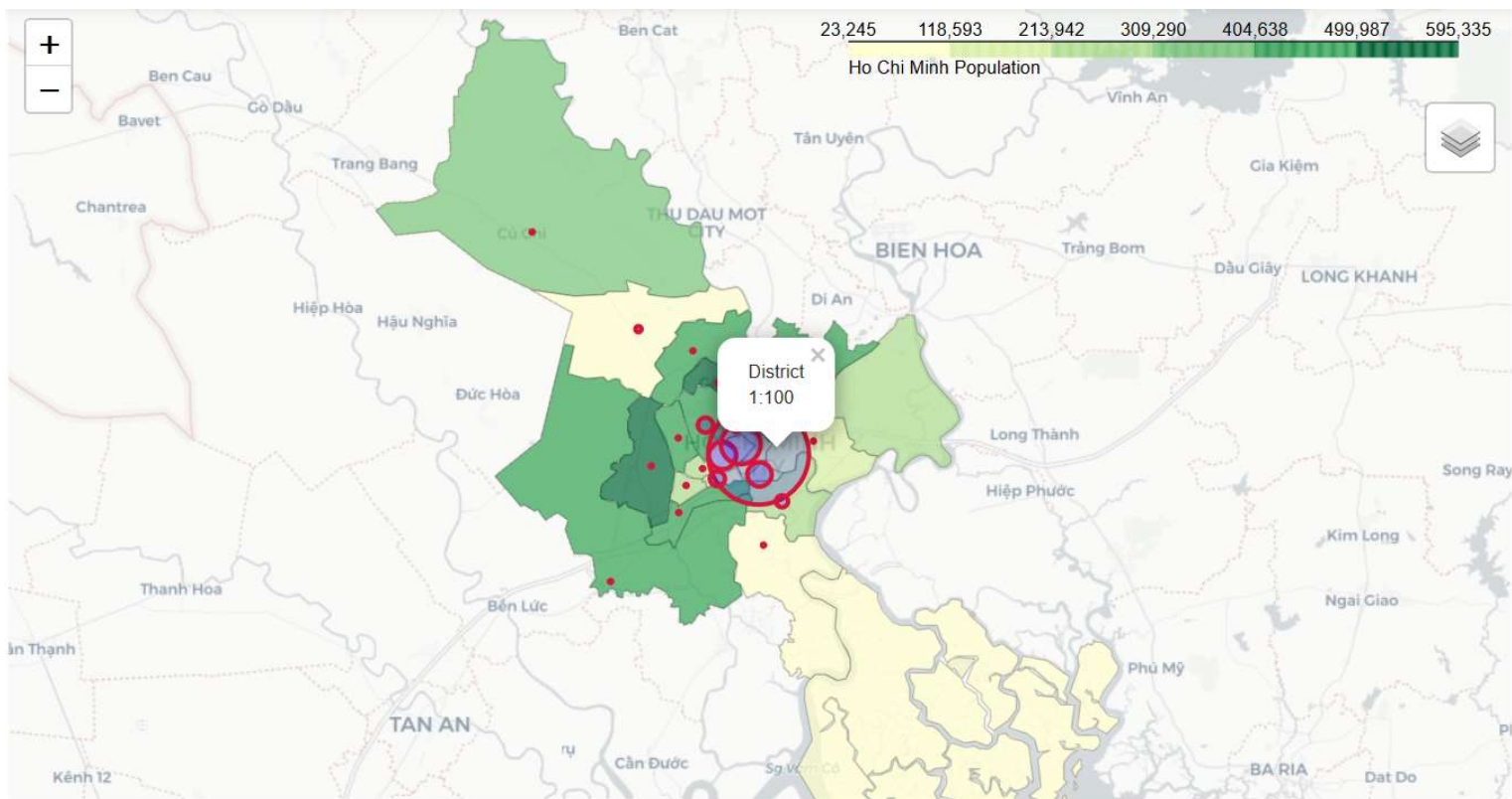
D. Result

The results of bar charts and K-means clustering clearly indicate the differences between Tokyo and Ho Chi Minh. However, there are also small similarities such as Café/Coffee shop being the more common venue or there are a lot of restaurants inside both cities.

As many people may not know how Ho Chi Minh looks like, I want to make a comparison of two cities using maps. I looked for GeoJson files of Tokyo and Ho Chi Minh using Google, and used **Folium** to create choropleth maps that will display information as below:

- Population distribution (as different color layers)
- Number of venues (as red circles)
- Neighborhood names





E. Discussion

Even though they are both Asian countries that share many similarities in culture, there seems to be a large difference in lifestyle of each city. Aside from the discussion in the Methodology section, here are some additional discussion:

- Convenient store seems to be a popular **business model** in Tokyo, while Vietnamese restaurant and café are more popular in Ho Chi Minh.
- Popular **foreign cuisines** are Chinese and Italian in Tokyo, while French cuisine and pizza are preferred in Ho Chi Minh.
- Access to **hotels and accommodation** is easier in Ho Chi Minh.
- People care about **fitness** more in Ho Chi Minh than Tokyo.

F. Conclusion

This study has provided me with much deeper understanding of the two cities. As I plan to work and do business in both Tokyo and Ho Chi Minh in the future, the results of this study would prove to be crucial.

I hope this study would provide useful information to those who want to do business or work that are related to both cities.