

IBM Data Science Professional Certificate
Capstone Project

Building sustainable cities

Comparing the world-leading cities
Singapore and New York

Jhoanna Chan-Punzalan
24 August 2020

Building sustainable cities

Comparing Singapore and New York City

Business Problem

Due to the rise of globalization, most of the world's population – as many as two out of three people by 2050¹ - are going to reside in cities. In order to sustain a healthy population and economic growth, these urban areas must adapt to maintain their attractiveness as a place to live, work and visit.

According to the United Nations, many governments will “face challenges in meeting the needs of their growing urban populations, including for housing, transportation, energy systems and other infrastructure, as well as for employment and basic services such as education and health care.” Providing a good quality of life for residents is therefore an urgent need for the world's cities.

According to Mercer, a consultancy firm which publishes the annual Quality of Living Ranking and works with city governments and planners on assessing factors to improve their rankings, it is “more necessary than ever for city leaders to have a clear understanding of how well they are doing to realize the goal of being competitive economically, socially, culturally and environmentally, for now and for the future.”²

Ranking 231 cities on 39 factors, the Mercer Quality of Living Ranking provides recommendations to companies on where to deploy mobile workers and set up new offices. The index shows that the best cities understand that the quality of living is an essential component of a city's attractiveness for businesses and mobile talent. In the 2019 survey, all US cities covered in the analysis fell in the rankings with the exception was New York (ranked 44), rising one place. In Asia, Singapore (ranked 25) has the highest quality of living followed by Tokyo.³

Two of the key performance metrics used by Mercer to analyze living conditions in a city are:

- Public services and transportation (electricity, water, public transportation, traffic congestion)
- Recreation (restaurants, theatres, cinemas, sports and leisure)

Targeting decisionmakers and stakeholders such as urban planners and city governments, I propose to use Location Based Social Networks (LBSN) such as Foursquare to determine the developments needed for a top-tier, future-oriented city with a fast-growing population. Due to the wide scope of factors which determine the attractiveness of a city, I shall limit my review to the data available on Foursquare, such as public transportation and recreation. The objective is to provide city planners and governments evidence-based guidelines so they can make informed decisions on how they should invest in and develop their cities based on forecasted population sizes and density. This study may also support them in locating sites – such as a new hospital or park - for future development.

I will use the industrialized cities of Singapore (ranked 25) and New York City (ranked 44) as benchmarks of what a model city should be due to their high ranking.

¹ “68% of the world population projected to live in urban areas by 2050, says UN”

16 May 2018, New York <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>

² “City Attractiveness Index”, Mercer <https://www.mercer.com/what-we-do/workforce-and-careers/city-attractiveness-index.html>

³ “Vienna Tops Mercer's 21st Quality of Living Ranking”, 13 March 2019 <https://www.mercer.com/newsroom/2019-quality-of-living-survey.html>

An Introduction to the Cities

1. Singapore

Singapore is an island city-state in Southeast Asia. With a total area of 725.7 km², the country is composed of one main island, 63 satellite islands and islets, and one outlying islet, the combined area of which has increased by 25% since the country's independence in 1965 as a result of extensive land reclamation projects. It has the second greatest population density in the world.⁴ The country is home to almost 5.7 million residents.⁵

There are four official languages of Singapore: English, Malay, Chinese, and Tamil; with English being the lingua franca. This reflects in its rich cultural diversity and extensive variety of cuisine and major festivals.

2. New York City

Described as the cultural, financial, and media capital of the world, New York City is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km²), New York City is also the most densely populated major city in the United States. It is also the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities.⁶

Situated on one of the world's largest natural harbors, New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island—were consolidated into a single city in 1898.

Since New York City is much larger than Singapore in terms of area size and population, the borough of Brooklyn was chosen to be the focus of analysis and comparison with Singapore.

Brooklyn is New York City's most populous borough, the state's most populous county, and the second-most densely populated county in the United States. An estimated 2,648,403 residents in 2020, representing 30.8% of New York City's population.⁷ The borough has an area of 250 km², it shares a land border with the borough of Queens at the western end of Long Island and has tunnel connections to Manhattan and Staten Island.⁸ Brooklyn has recently become an avant-garde destination for hipsters, experiencing gentrification and a decrease in housing affordability. Since the 2010s, Brooklyn has "evolved into a thriving hub of entrepreneurship, high technology startup firms, postmodern art and design."

⁴ "Singapore", Wikipedia <https://en.wikipedia.org/wiki/Singapore>

⁵ "Demographics of Singapore", Wikipedia https://en.wikipedia.org/wiki/Demographics_of_Singapore

⁶ "New York City", Wikipedia https://en.wikipedia.org/wiki/New_York_City

⁷ "Demographics of Brooklyn", Wikipedia https://en.wikipedia.org/wiki/Demographics_of_Brooklyn

⁸ "Brooklyn", Wikipedia <https://en.wikipedia.org/wiki/Brooklyn>

Methodology

We shall approach this problem as follows:

- Collect the city data via Wikipedia and available sources
- Use Geopy library and Nominatim API to determine coordinates of each location
- Use FourSquare API we will find all venues for each neighborhood
- Use venue categories and frequency to sort each venue by location
- Visualize the neighborhoods using Folium library
- Cluster the neighborhoods using the K-Means Clustering algorithm
- Study the cluster groupings of each city
- Compare the clusters of both cities and draw conclusions on what makes them good urban places to live in

I have decided to use K-Means Clustering, as it would enable me to gain more information about the types of neighborhoods in the cities by grouping the similar neighborhoods together. Similar neighborhoods would have similar types of venues. The number of neighborhoods inside a cluster would show the level of homogeneity. The different clusters would show the differences between the types of neighborhoods. To sum up, the clustering algorithm would provide a good overall characterization of the cities.

Data Sources

Foursquare

To work on this comparative study, I shall utilize location data⁹ from Foursquare to get information about venues in Singapore and New York.

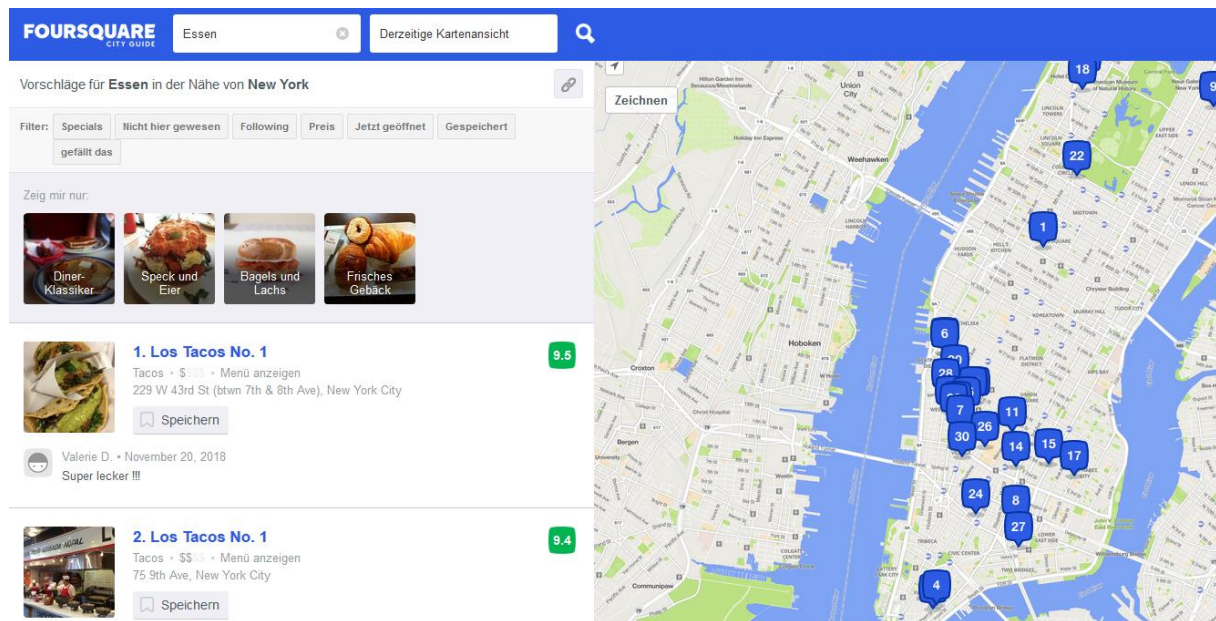
I'd like to determine the presence and availability level of venues that fall under key indicators in top-level cities: recreation and public transport. These are classified under Foursquare's venue categories such as:

- | | |
|------------------------|------------------------|
| - Train station | - Shopping mall |
| - Bus station | - Supermarket |
| - Restaurants and cafe | - Flea market |
| - Movie theater | - Beach |
| - Concert hall | - Park |
| - Museum | - Gym / Fitness center |
| - Botanical garden | - Bike trail |

⁹ "Venue Categories", Foursquare <https://developer.foursquare.com/docs/build-with-foursquare/categories/>

Building sustainable cities

Comparing Singapore and New York City



Extract from a Foursquare search for restaurants in New York City. Foursquare.com.

Wikipedia

The Foursquare location data will be supplemented by statistics taken from Wikipedia. For example, the information on Singapore's neighborhoods were scraped from the Wikipedia page "Postal codes in Singapore".¹⁰

Geopy library

The search API (<https://nominatim.openstreetmap.org/>) returns a location's coordinates from a textual description. For example, the Geopy library can obtain the geographical coordinates of the list of locations in Singapore from Wikipedia.

Data Acquisition and Cleaning

1. Singapore

As already mentioned, the first step is to scrape the neighborhood data from Wikipedia and clean the dataset, leaving only the information needed.

General location	
0	Raffles Place, Cecil, Marina, People's Park
1	Anson, Tanjong Pagar
2	Bukit Merah, Queenstown, Tiong Bahru
3	Telok Blangah, Harbourfront
4	Pasir Panjang, Hong Leong Garden, Clementi New...

Each row in the "General location" column contains several neighborhoods. I separated these so each neighborhood is in its own row. There are 79 neighborhoods in total.

¹⁰ "Postal codes in Singapore", Wikipedia https://en.wikipedia.org/wiki/Postal_codes_in_Singapore

Building sustainable cities

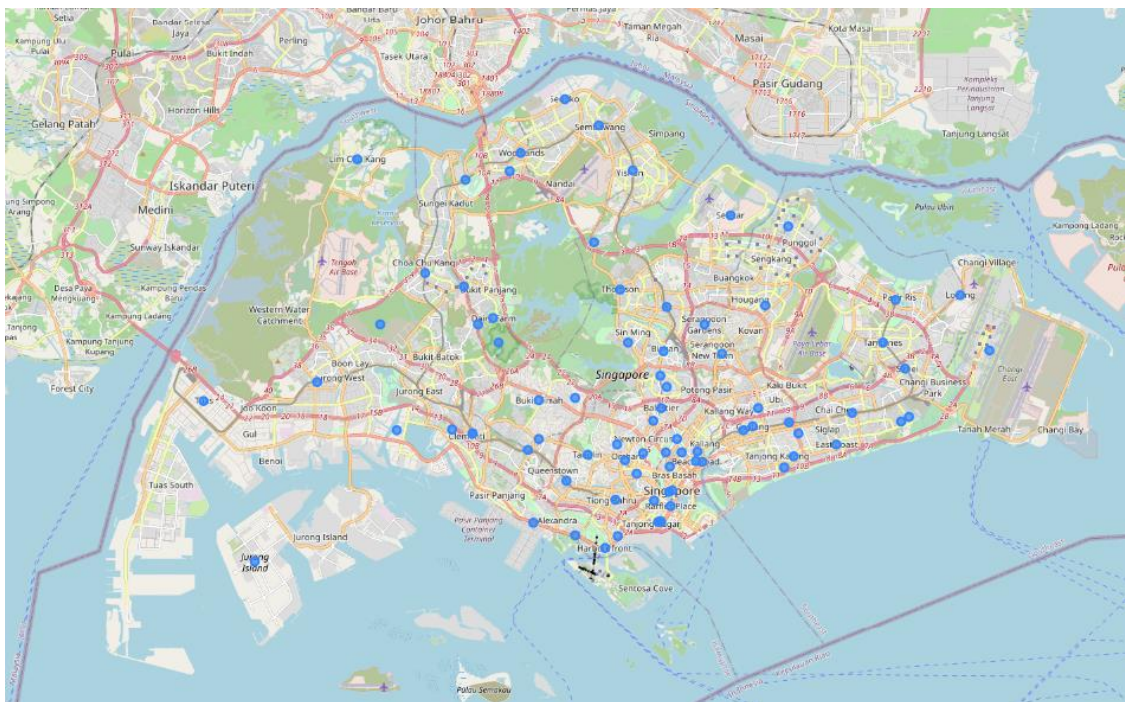
Comparing Singapore and New York City

Because some neighborhood names are similar to those outside of Singapore, I decided to add the country name “Singapore” to the address to obtain the correct geographical coordinates.

Using the Geopy library, the latitude and longitude values of each neighborhood were retrieved.

	Neighborhood	Country	Address	Latitude	Longitude
0	Raffles Place	Singapore	Raffles Place, Singapore	1.283542	103.851460
1	Cecil	Singapore	Cecil, Singapore	1.276955	103.847016
2	Marina	Singapore	Marina, Singapore	1.290475	103.852036
3	People's Park	Singapore	People's Park, Singapore	1.285810	103.844160
4	Anson	Singapore	Anson, Singapore	1.276629	103.847247

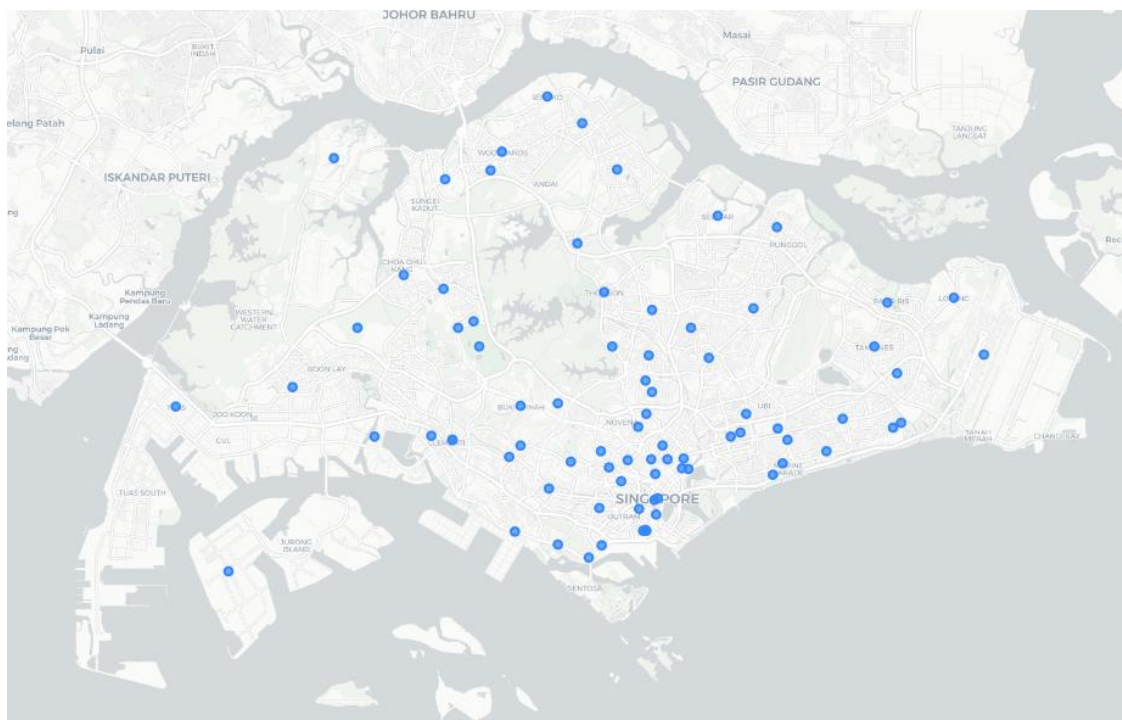
The map of Singapore and its neighborhoods can be visualized using Folium:



Map of Singapore and its neighborhoods using Open Street Maps tiles.

Building sustainable cities

Comparing Singapore and New York City



Map of Singapore and its neighborhoods using CARTODBPOSITRON tiles.

A function to call Foursquare to get the list of venue names, their location coordinates and categories was run and the resulting dataframe grouped these venues by neighborhood. There were 1,442 venues in 201 unique categories returned by Foursquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Raffles Place	1.283542	103.85146	The Fullerton Bay Hotel	1.283878	103.853314	Hotel
1	Raffles Place	1.283542	103.85146	City Hot Pot	1.284173	103.851585	Hotpot Restaurant
2	Raffles Place	1.283542	103.85146	CULINARYON	1.284876	103.850933	Comfort Food Restaurant
3	Raffles Place	1.283542	103.85146	Virgin Active	1.284608	103.850815	Gym / Fitness Center
4	Raffles Place	1.283542	103.85146	1-Altitude	1.284794	103.851151	Cocktail Bar

The number of venues found for each neighborhood ranged from one to fifty.

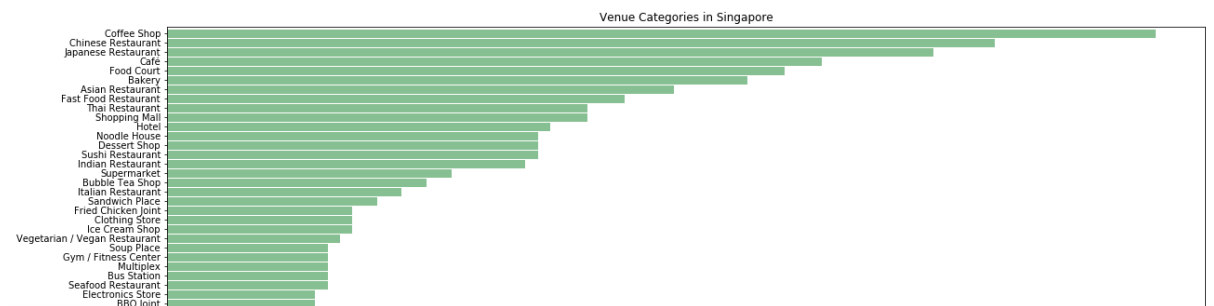
Neighborhood	
Aljunied	11
Amber Road	10
Ang Mo Kio	26
Beach Road	20
Braddell	8
Bukit Panjang	22
Bukit Timah	5

The top 10 largest number of venues found fell under the main “Food” category:

Building sustainable cities

Comparing Singapore and New York City

Venue Category	Venue
Coffee Shop	80
Chinese Restaurant	67
Japanese Restaurant	62
Café	53
Food Court	50
Bakery	47
Asian Restaurant	41
Fast Food Restaurant	37
Thai Restaurant	34
Shopping Mall	34



I ranked the 10 most common venues per neighborhood:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Aljunied	Vegetarian / Vegan Restaurant	Coffee Shop	Steakhouse	Kitchen Supply Store	Dim Sum Restaurant	Train Station	Seafood Restaurant	Japanese Restaurant	Martial Arts School	Women's Store
1 Amber Road	Gym	Café	Pool Hall	Club House	Chinese Restaurant	Tennis Court	Bar	Dance Studio	Halal Restaurant	Electronics Store
2 Ang Mo Kio	Dessert Shop	Snack Place	Bubble Tea Shop	Japanese Restaurant	Burger Joint	Coffee Shop	Food Court	Shopping Mall	Sandwich Place	Fried Chicken Joint
3 Beach Road	Vietnamese Restaurant	Dessert Shop	French Restaurant	Food Court	BBO Joint	Café	Multiplex	Taiwanese Restaurant	Thai Restaurant	Coffee Shop
4 Braddell	Bakery	Women's Store	Fast Food Restaurant	Food Court	Pharmacy	Bus Line	Metro Station	Food Stand	Food & Drink Shop	Food

Out of 79 neighborhoods, 6 had no venues, so they were dropped from the dataset, leaving 73 neighborhoods for clustering.

2. New York City – focusing on Brooklyn

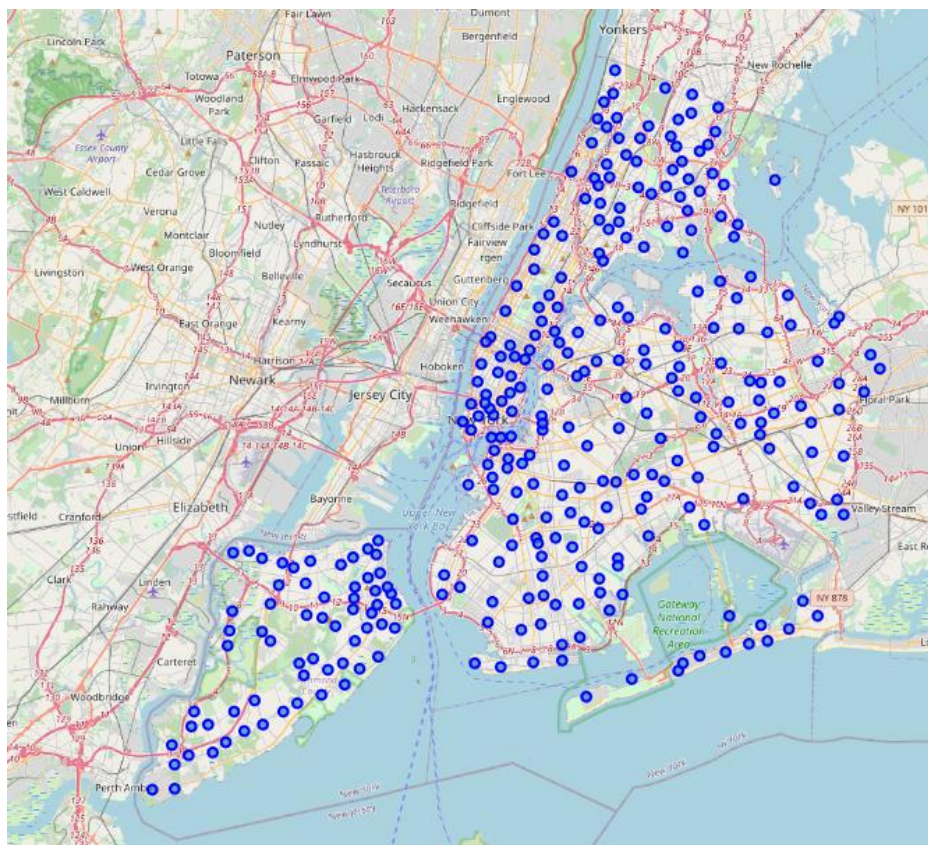
Using data provided by the course, New York City's 5 boroughs, 306 neighborhoods and their respective geographical coordinates can be obtained:

Building sustainable cities

Comparing Singapore and New York City

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

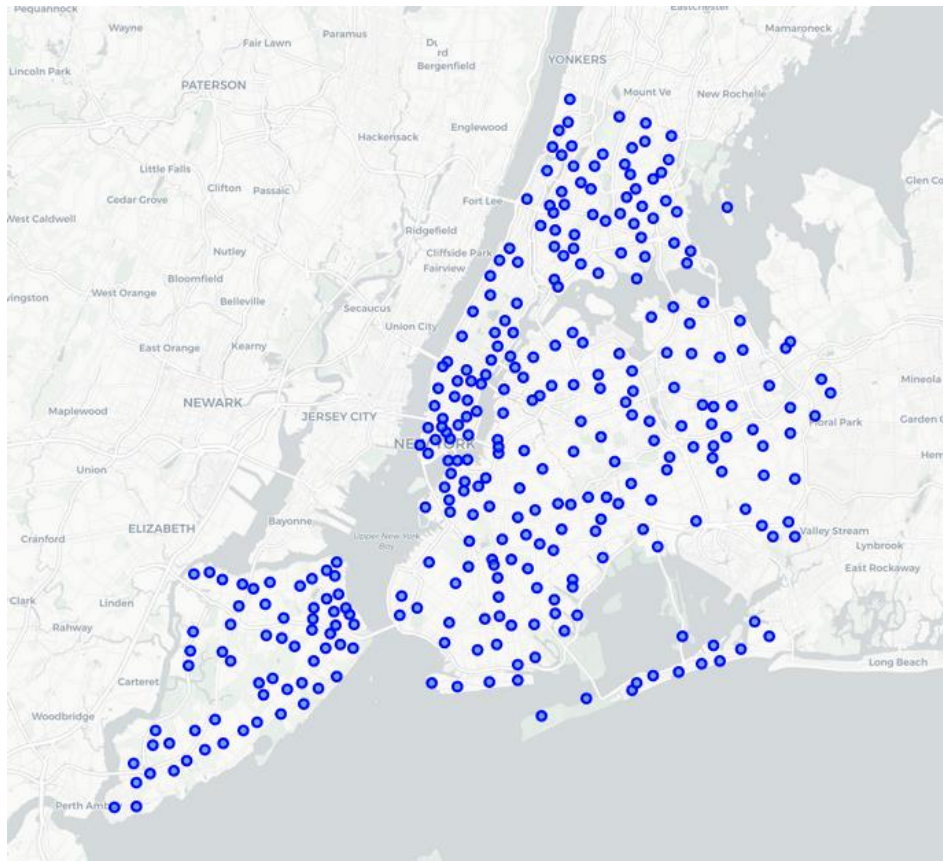
The map of New York City and its 306 neighborhoods can be visualized using Folium:



Map of New York City and its neighborhoods using Open Street View tiles.

Building sustainable cities

Comparing Singapore and New York City



Map of New York City and its neighborhoods using CARTODBPOSITRON tiles.

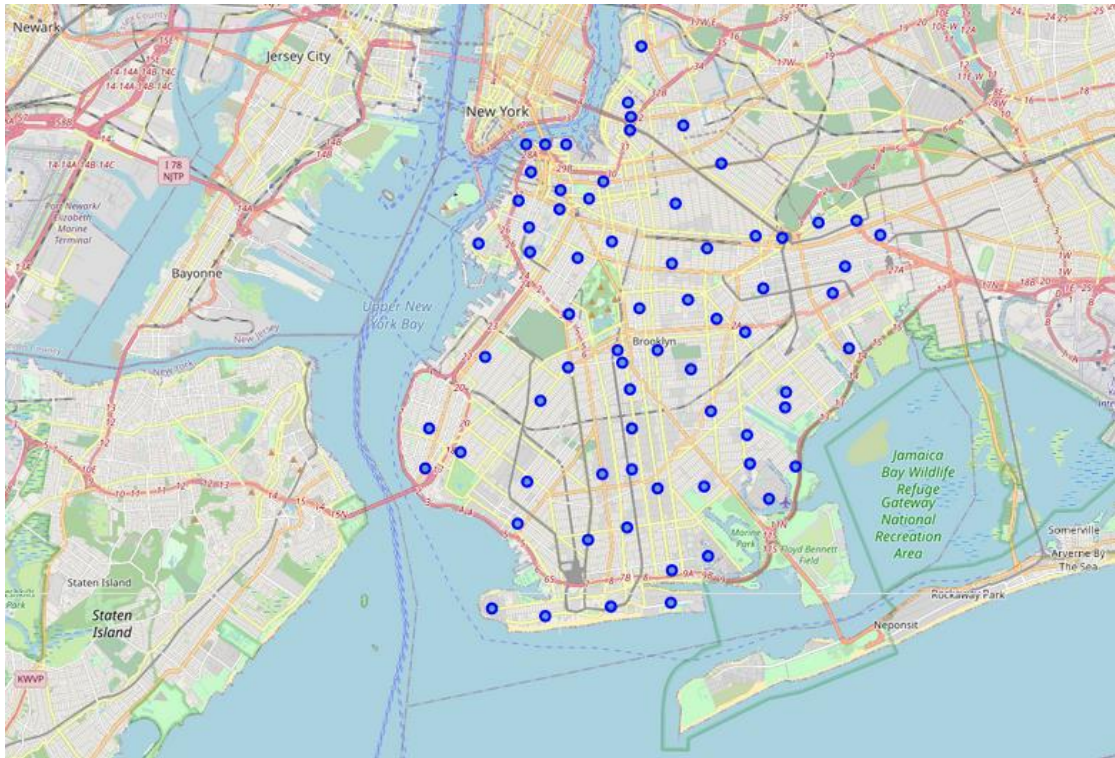
New York City has 306 neighborhoods while Singapore has 79 neighborhoods. To make a fair comparison, we can focus our analysis on one borough of New York with 70 neighborhoods, Brooklyn.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471
5	Brooklyn	Brighton Beach	40.576825	-73.965094
6	Brooklyn	Sheepshead Bay	40.586890	-73.943186
7	Brooklyn	Manhattan Terrace	40.614433	-73.957438
8	Brooklyn	Flatbush	40.636326	-73.958401
9	Brooklyn	Crown Heights	40.670829	-73.943291

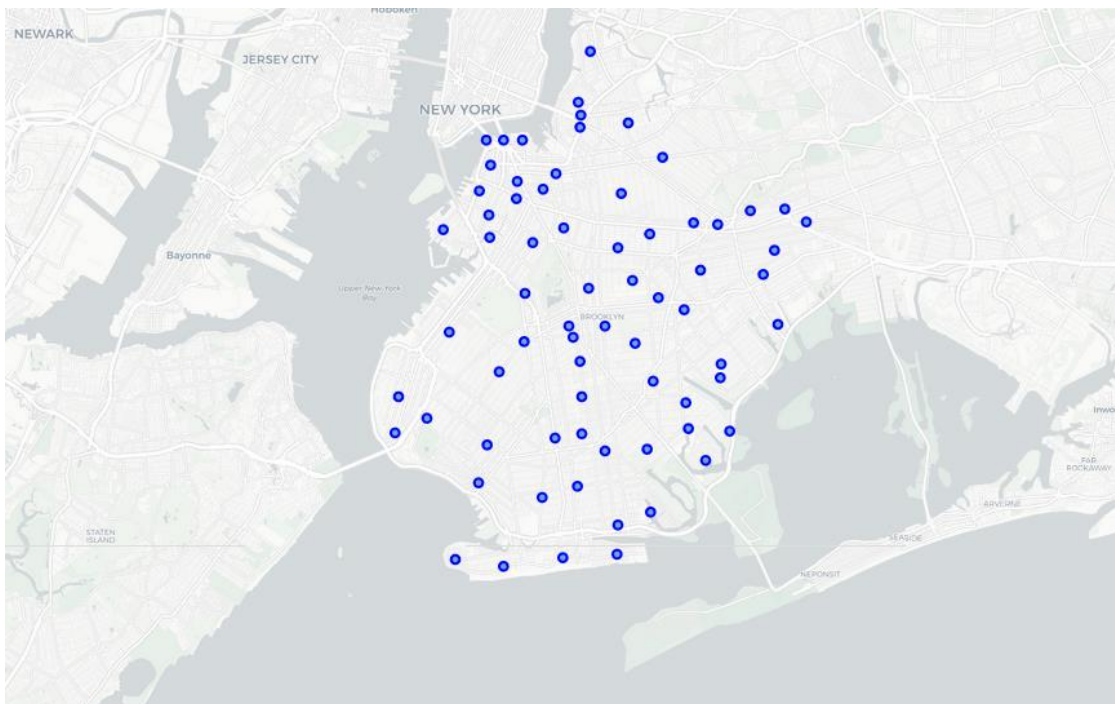
Brooklyn's 70 neighborhoods can be visualized on the map using Folium:

Building sustainable cities

Comparing Singapore and New York City



Map of Brooklyn and its neighborhoods using Open Street Maps tiles.



Map of Brooklyn and its neighborhoods using CARTODBPOSITRON tiles.

A function to call Foursquare to get the list of venue names, their location coordinates and categories was run (like the function used for Singapore above) and the resulting dataframe grouped these venues by neighborhood. There were 924 venues in 210 unique categories returned by Foursquare.

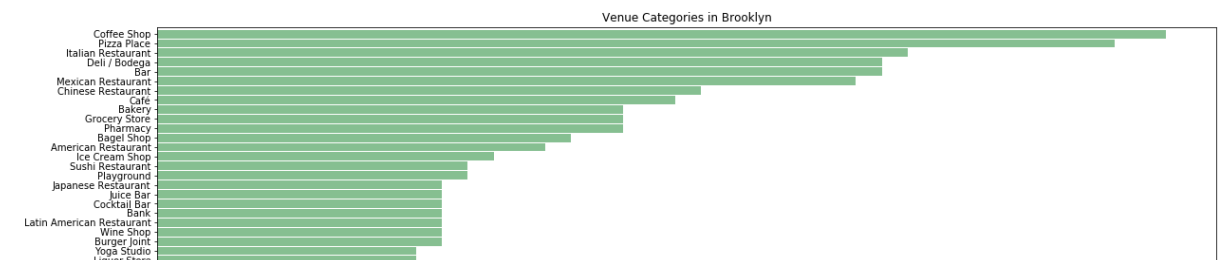
Building sustainable cities

Comparing Singapore and New York City

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore
4	Bay Ridge	40.625801	-74.030621	Brooklyn Market	40.626939	-74.029948	Grocery Store

The most venues found by Foursquare fall under the main category “Food”. The resulting dataframe and bar chart present the following top venue categories:

Venue Category	Venue
Coffee Shop	39
Pizza Place	37
Italian Restaurant	29
Deli / Bodega	28
Bar	28
Mexican Restaurant	27
Chinese Restaurant	21
Café	20
Bakery	18
Grocery Store	18



I ranked the 10 most common venues per neighborhood:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bath Beach	Ice Cream Shop	Video Store	Gas Station	Liquor Store	Donut Shop	Italian Restaurant	Kebab Restaurant	Middle Eastern Restaurant	Diner	Hookah Bar
1	Bay Ridge	Italian Restaurant	Pizza Place	Spa	Chinese Restaurant	Pool Hall	Seafood Restaurant	Sushi Restaurant	Caucasian Restaurant	Hookah Bar	Café
2	Bedford Stuyvesant	Coffee Shop	Boutique	Italian Restaurant	Café	Juice Bar	Yoga Studio	Fast Food Restaurant	French Restaurant	Food Truck	Food Stand
3	Bensonhurst	Italian Restaurant	Chinese Restaurant	Moving Target	Yoga Studio	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop	Flower Shop
4	Boerum Hill	Furniture / Home Store	Thrift / Vintage Store	French Restaurant	Men's Store	Middle Eastern Restaurant	Spa	Kids Store	Cocktail Bar	Sushi Restaurant	Coffee Shop

Out of 70 neighborhoods, 1 had no venues, so they were dropped from the dataset, leaving 69 neighborhoods for clustering.

The data is now ready for K-Means Clustering.

Building sustainable cities

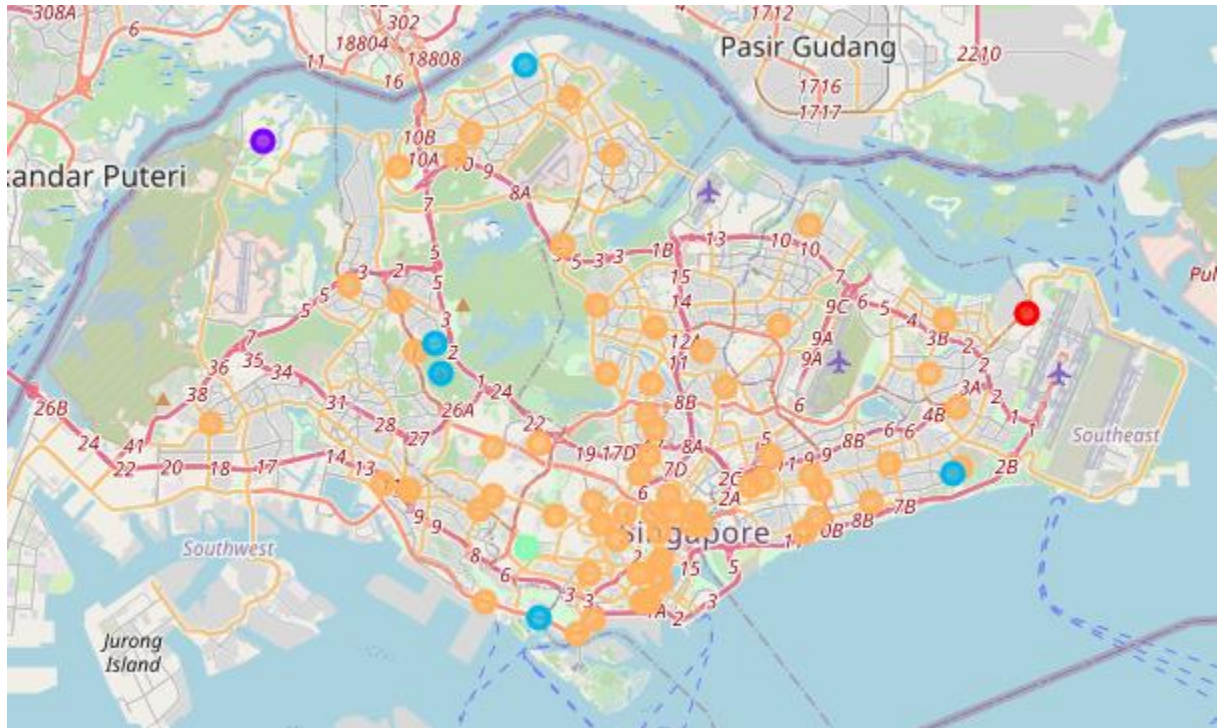
Comparing Singapore and New York City

Analysis: K-Means Clustering

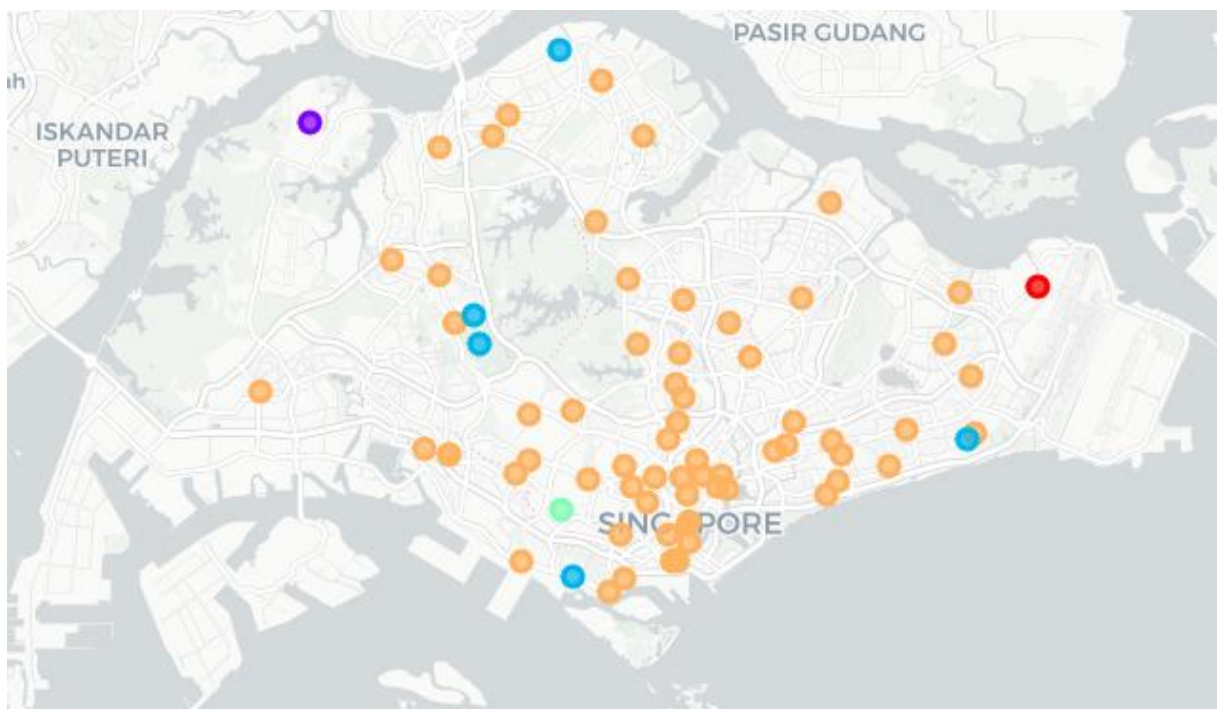
1. Singapore

The cleaned-up data is now ready for K-Means Clustering. For the purpose of this analysis, I used 5 clusters to group the various neighborhoods.

The resulting clusters can be visualized on the map using Folium.



Map of Singapore and its clustered neighborhoods using Open Street Maps tiles.



Map of Singapore and its clustered neighborhoods using CARTODBPOSITRON tiles.

Analyzing the clusters

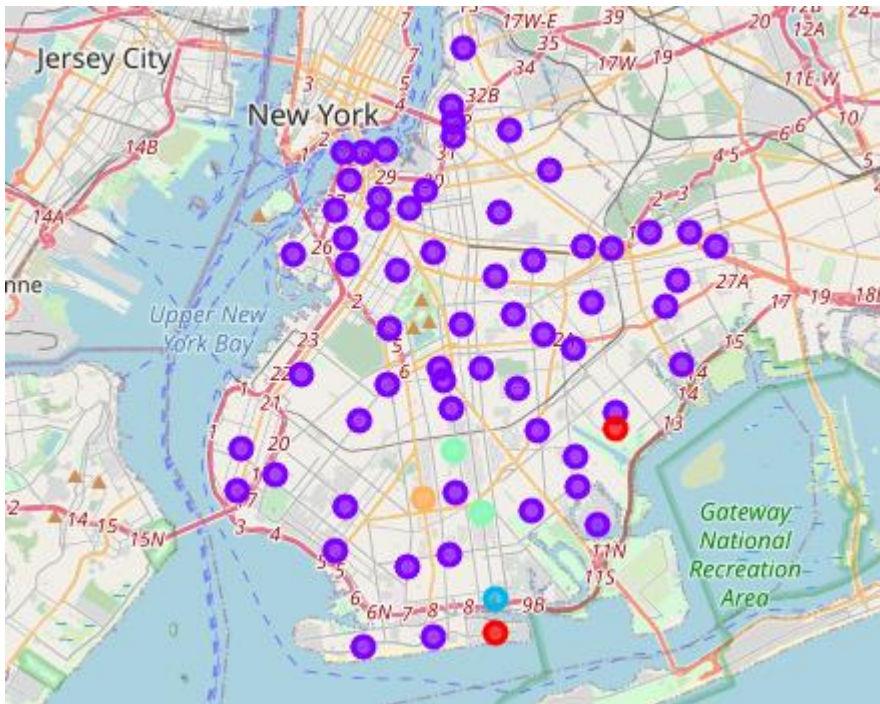
- Cluster 1 (red) – Loyang
Foursquare returned only 1 venue: 'Bus stop 98069', with the category 'Bus Station'.
- Cluster 2 (purple) – Lim Chu Kang
Foursquare returned only 1 venue: 'Neo Cheng Soon Farm', with the category 'Farm'.
- Cluster 3 (blue) – 5 neighborhoods: Telok Blangah, Bukit Timah, Kew Drive, Dairy Farm, Senoko.
Foursquare returned 4 venues for Telok Blangah: Metro Station, Bus Stop, Bus Station, Basketball Court.
For Bukit Timah, 5 venues: 2 x Trail, Scenic Lookout, Hill, Rest Area.
Foursquare returned 3 venues for Kew Drive: Food and Drink Shop, Trail, Bus Stop.
For Dairy Farm, 3 venues were found: Trail, Garden, Public Art.
For Senoko, 2 venues were: Shopping Mall and Candy Store.
The similarities of the neighborhoods seem to be from the outdoors and transport categories: Trail, Bus and Train Stops.
- Cluster 4 (green) – Queenstown
Foursquare returned only 1 venue in Queenstown: 'Queenstown MRT Station', with the category 'Train station'.
- Cluster 5 (orange) – 65 neighborhoods
The most common venues for this cluster are very similar; they fall under the main category 'Food'. Coffee shops, food courts and restaurants of various cuisines are usually the top ranked venues. In a few cases, stores and hotels are in the number one spot.

Building sustainable cities

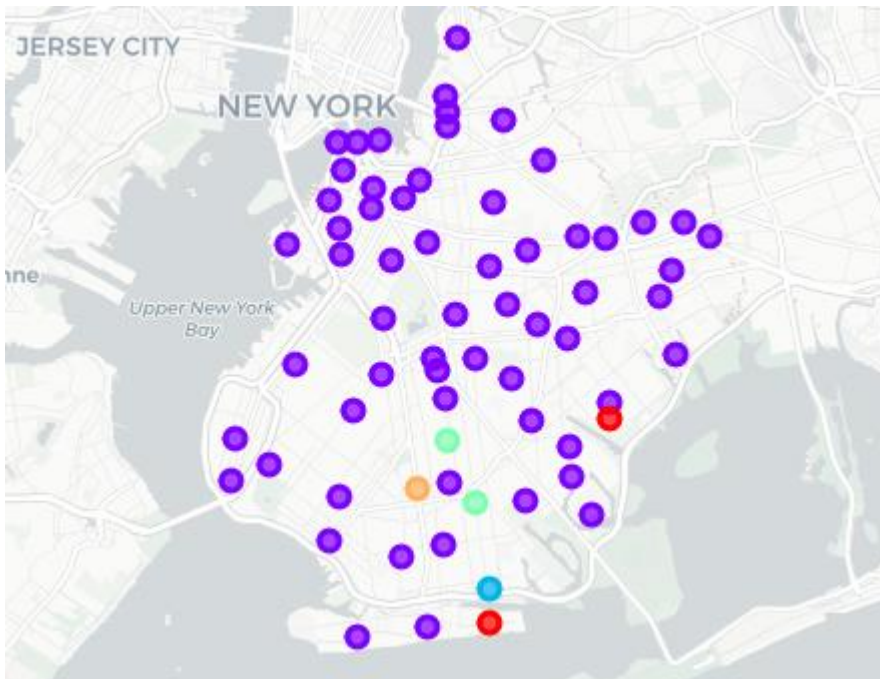
Comparing Singapore and New York City

2. Brooklyn

As above, I used 5 clusters to group the various neighborhoods. The resulting clusters can be visualized on the map using Folium.



Map of Brooklyn and its clustered neighborhoods using Open Street Maps tiles.



Map of Brooklyn and its clustered neighborhoods using CARTODBPOSITRON tiles.

Analyzing the clusters

- Cluster 1 (red) - Manhattan Beach and Paerdegat Basin
Foursquare returned 4 venues for Manhattan Beach: Beach, Ice Cream Shop, Bus Stop, Harbor / Marina. Only 1 venue for Paerdegat Basin was returned: 'Manhattan Beach' with the category 'Beach'.

Building sustainable cities

Comparing Singapore and New York City

The similarities of these 2 neighborhoods seem to come from the 'Beach' category.

- Cluster 2 (purple) – 61 neighborhoods in Brooklyn
This cluster comprises of 61 neighborhoods, the great majority of all neighborhoods in Brooklyn. The most common venues are in the 'Food' and 'Shop & Service' categories: Restaurants of various cuisines, stores and shops offering different products.
- Cluster 3 (blue) – Sheepshead Bay
Foursquare returned only 1 venue for this neighborhood: 'Emmons Avenue' with the category 'Lake'.
- Cluster 4 (light green) – Midwood and Madison
The top 10 most common venues are exactly the same, for instance: Bus station, yoga studio, frozen yogurt shop, etc.
Checking the venues returned by Foursquare, both Midwood and Madison only had 1 venue each: 'Bus station'. This accounts for them being clustered together.
- Cluster 5 (orange) – Ocean Parkway.
Foursquare returned 1 venue: 'Sour hour' with the category 'Outdoors & Recreation'.

Results

Comparing Singapore and New York City (Brooklyn)

The above methodology and analysis provide a good initial approach to finding answers to our questions. Singapore is a highly urbanized city-state and Brooklyn is also a highly urbanized borough of New York City. From their reputations, both offer a great variety of food of various cuisines, all sorts of shopping areas, hotels and high-rises yet with plenty of natural spots.

After reviewing the resulting clusters generated by the K-Means Clustering algorithm, one can conclude that Singapore is a highly urbanized city-state where the residents live, work and play throughout the island. There is a large variety of eating spots and shopping areas and these are decentralized, meaning each neighborhood has roughly the same distribution and availability of these venues. There are a few exceptions, as seen in the Clusters 1 to 4. It could also be that Foursquare doesn't have much information on venues in these four clusters' neighborhoods, so the algorithm separated them from the first cluster.

The clustering algorithm for Brooklyn provided almost the same results with Singapore. The first cluster comprises a large majority of the neighborhoods, while clusters 1, 3-5 are the exceptions. This implies that many of Brooklyn's neighborhoods are highly similar to one another since it's a highly urbanized borough like Singapore.

I ran the K-Means Clustering algorithm several times for both cities, and the clustering results differed for the 4 remaining clusters. The big cluster with the large majority of neighborhoods existed each time the algorithm was run, but the other small clusters differed in the neighborhoods chosen. As mentioned above, one reason for this is that Foursquare has very few information on several neighborhoods, so the algorithm must have assigned new neighborhoods to a different cluster each time it was run.

Discussions

This is a project that has made use of various data science skills: from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to statistical models (K-Means Clustering) and map visualization (Folium).

Based on the results of this study, an important insight which can be used by urban planners and city governments is the topic of *decentralization*.

As seen in the cities of Singapore and New York City, most of the neighborhoods were clustered into one, since their characteristics are very similar to each other. Decentralizing critical venues such as retail, food, nature, workplaces reduces congestion, improves service offerings and allows for better accessibility for many residents. This can also apply to government offices. In the end, spreading out such important services enables a higher quality of living. This is what future-centric cities should aspire to and work towards, especially as more and more people move into them in the next years and decades.

Limitations and Suggestions for Future Research

I observed, however, limitations to my study's approach:

1. Foursquare's data does not capture the actual variety and number of venues

When I was conducting a general venue search in Singapore, I received the following warning text: "There aren't a lot of results near you. Try something more general, reset your filters, or expand the search area." It was a surprising message given that Singapore is a small, transparent, mobile-phone-centric society where almost all information is online. Most of its citizens are on social media and use mobile phones in their daily lives.

In reviewing the K-Means Clustering results, it must be pointed out that the Foursquare results are not an accurate representation of the variety of venues Singapore and Brooklyn have.

Based on my experience living in Singapore for several years, the public transport system – metro rail, light rail, bus network - is comprehensive. According to the Land Transport Authority, Singapore has "more than 130 stations across six MRT lines span the island. This 200km system has over three million daily ridership. In addition, there are more than 40 stations across two LRT lines and this 28km system has over 200 thousand daily ridership."¹¹ On top of this, there are 5,008 bus stops in the country.¹² The Foursquare results under the "Travel and Transport" venue category, on the other hand, are only very few in number.

Regarding the "Outdoors and Recreation" category, I also believe that the Foursquare results do not provide an accurate representation. For example, Singapore has more than 350 parks and 4 nature reserves.¹³ The Foursquare results in the "Outdoors & Recreation" category are again very few.

¹¹ "Rail network", Land Transport Authority

https://www.lta.gov.sg/content/ltagov/en/getting_around/public_transport/rail_network.html

¹² "Bus stops", Land Transport Guru Blog, 1 January 2018 <https://landtransportguru.net/bus-stops/>

¹³ "13 Nature Reserves & Parks In Singapore For A Dose Of Greenery And Scenic Walking Trails", The Smart Local, 24 August 2019 <https://thesmartlocal.com/read/nature-reserves-parks/>

Building sustainable cities

Comparing Singapore and New York City

Venue	
Venue Category	
Bus Station	15
Bus Line	5
Metro Station	7
Bus Stop	8
Train Station	3
Light Rail Station	1

Venue	
Venue Category	
Trail	5.0
Park	8.0
Scenic Lookout	4.0
Playground	4.0
Garden	1.0

Also, I noticed that the same in the venue results of Brooklyn. Regarding public transportation and outdoor areas, the borough has many more venues than what Foursquare returned during my query.

According to Wikipedia, there are 170 New York City Subway stations in Brooklyn.¹⁴ Moreover, there are 55 local Brooklyn bus routes operated by the New York City Transit Authority,¹⁵ which connect many neighborhoods to the subway stations as well. Checking the Foursquare results under the “Travel and Transport” venue category, only very few were returned.

In the “Outdoors and Recreation” category, I also believe that the Foursquare results do not accurately capture the many outdoor spaces in the borough. A quick Google search revealed that Brooklyn has more than 40 parks.¹⁶

Venue	
Venue Category	
Bus Station	5.0
Bus Line	2.0
Metro Station	2.0
Bus Stop	2.0

Venue	
Venue Category	
Trail	2.0
Park	9.0
Scenic Lookout	4.0
Playground	12.0
Other Great Outdoors	3.0
Outdoors & Recreation	2.0
Beach	3.0

Interestingly, Foursquare found 400 less venues in Brooklyn than in Singapore, yet 7 more venue categories.

For future studies, I recommend that the use of Foursquare as a LBSN be further supplemented with other LBSN channels like Google Maps API and open data sources from city governments and independent journals.

¹⁴ “List of New York City Subway stations in Brooklyn”, Wikipedia

https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Brooklyn

¹⁵ “List of bus routes in Brooklyn”, Wikipedia https://en.wikipedia.org/wiki/List_of_bus_routes_in_Brooklyn

¹⁶ “Here Are Some of the Most Beautiful Parks in Brooklyn, NY”, At Home in Brooklyn Blog, October 2016
<https://athomeinbrooklyn.com/blog/2016/10/parks-in-brooklyn-ny/>

2. Type of information provided by Foursquare is insufficient

In answering the business problem which I posed at the start of this project, the information provided by Foursquare is lacking and more data sources more focused on urban planning considerations are required.

Foursquare's information is user-generated, meaning that its information is more consumer-oriented. Providing information on stores and restaurants, such as opening hours, ratings and reviews, is its forte; however, Foursquare does not capture information such as general traffic level to the venues or their accessibility to public transport.

The theme of urban planning is undoubtedly complex, and data on issues like ease of mobility via public transport, housing availability, affordability level of the city, work opportunities are required in order to answer the question: what makes a world-class city? Data and insights from academic studies such as "Green infrastructure planning: Unveiling meaningful spaces through Foursquare users' preferences"¹⁷ and "Scalable Urban Data Collection From The Web"¹⁸ focusing on mobility in a city, could serve as useful examples to further explore this topic.

3. A larger comparative study of more cities would be beneficial

This study is limited in the sense that it only looks at 2 cities: Singapore and New York City (particularly Brooklyn). These cities provide a quick view of what conditions are needed for a sustainable urban area, but it would certainly be more effective to have a larger dataset with more cities in the high, middle and low rankings in the Mercer Index. Comparing and contrasting a larger group of cities would provide better insights.

Conclusions

In this study, I analyzed two cities with the objective to find out more about what makes them top-tier urban places despite having big populations. The Foursquare query and K-Means Clustering approach I used is simple but provides a general understanding of the cities' neighborhoods. This can be easily replicated to other cities to gain more understanding about how to further improve urban planning and land use policies.

Decentralizing critical venues is a key finding from this project. Spreading out retail, restaurants, outdoor sites make for better living conditions and improve accessibility.

As mentioned in the Discussion section, there are limitations as to the information gathered from Foursquare, so it's highly recommended to use other data sources like other LBSN platforms and open data sets. More data would build a more accurate picture and provide actionable insights for urban planners.

¹⁷ Pablo Martí, Clara García-Mayor, Almudena Nolasco-Cirugeda, Leticia Serrano-Estrada, "Green infrastructure planning: Unveiling meaningful spaces through Foursquare users' preferences", Land Use Policy, 6 June 2020 <https://www.journals.elsevier.com/land-use-policy>

¹⁸ Rijurekha Sen, Daniele Quercia, Carmen Vaca, Krishna Gummadi, "Scalable Urban Data Collection From The Web", 2016, Association for the Advancement of Artificial Intelligence, www.aaai.org