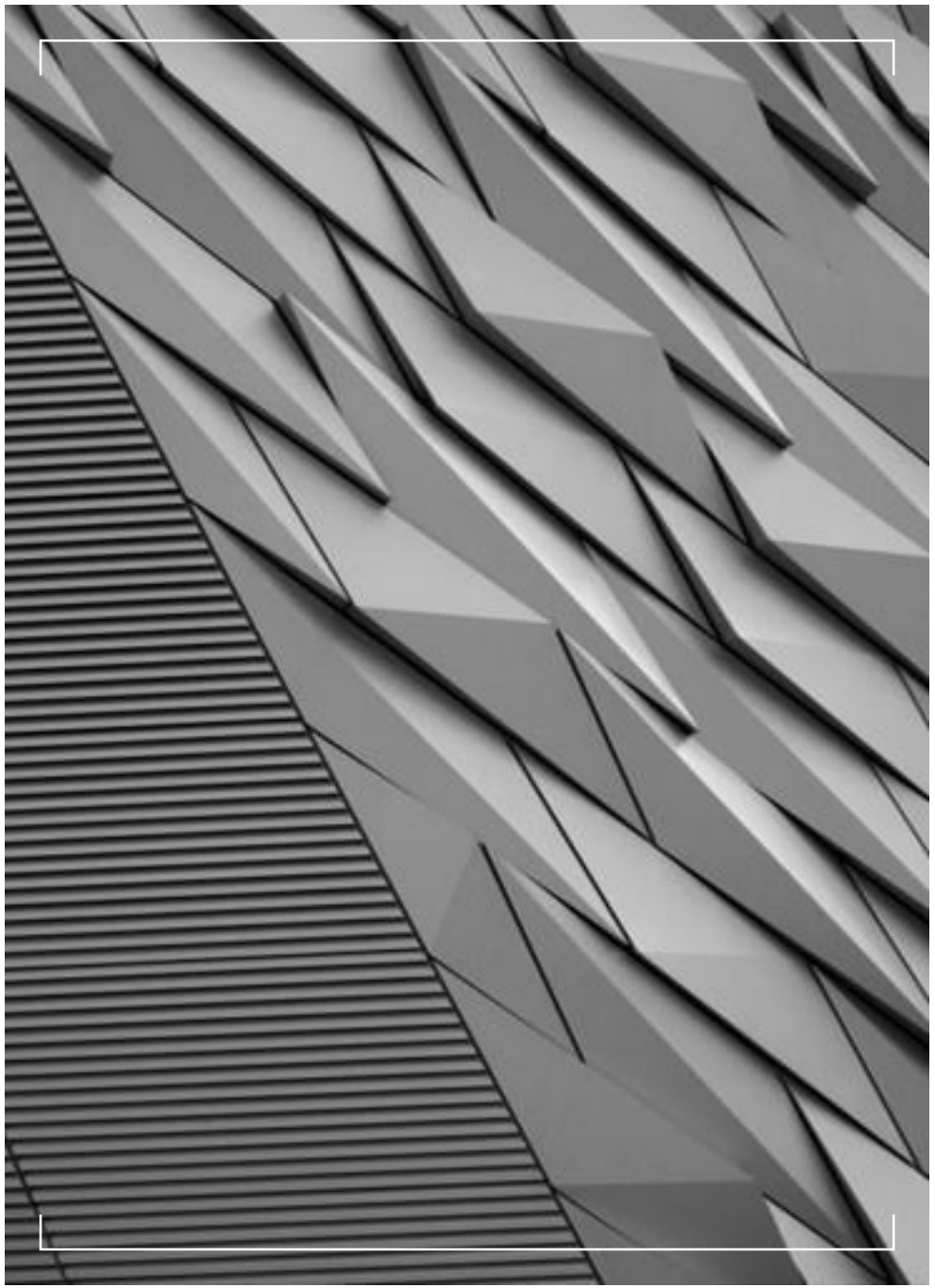
IBM Capstone Project

Building sustainable cities

Comparing the world-leading cities Singapore and New York

Jhoanna Chan-Punzalan

24 August 2020



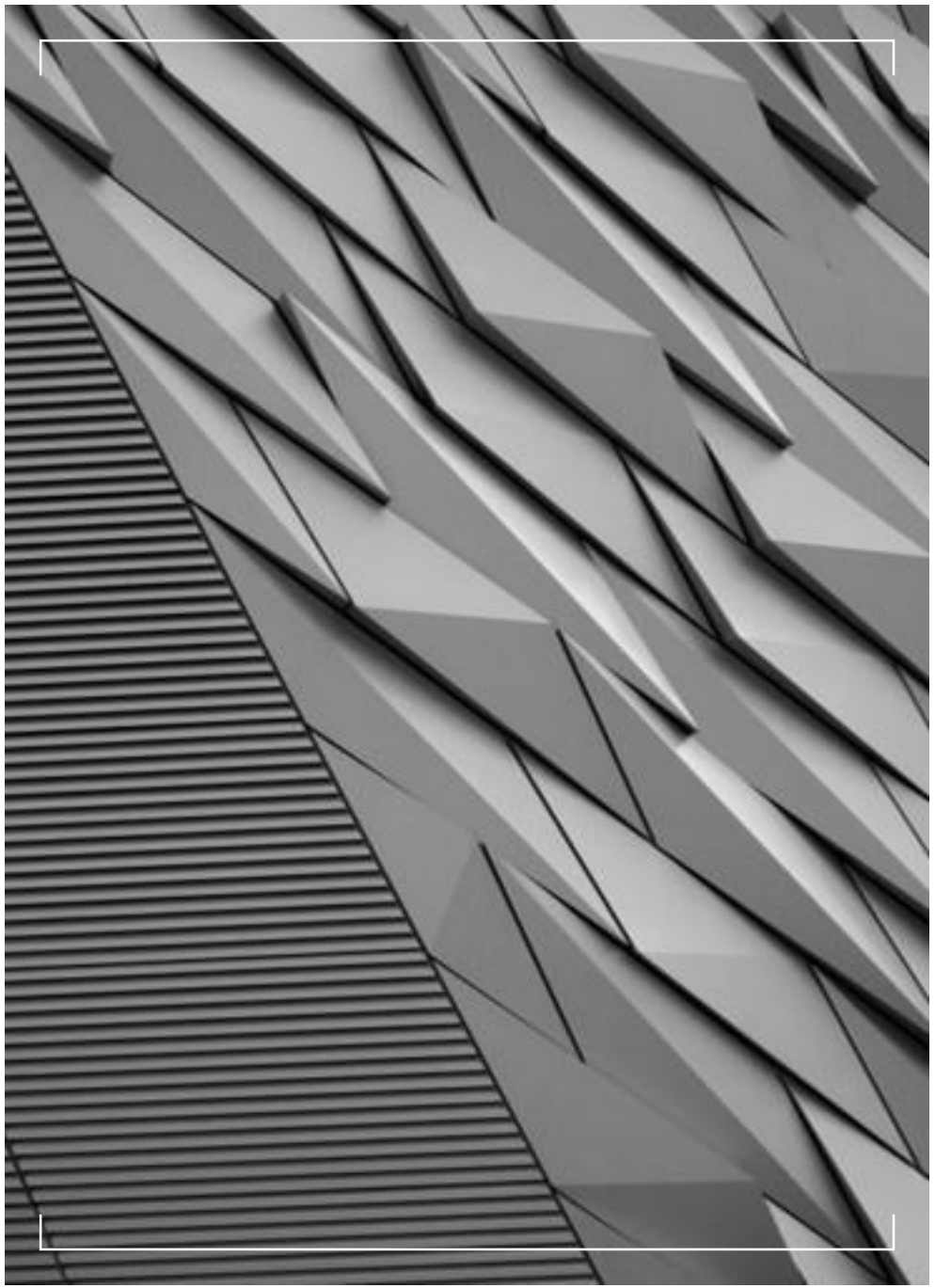
Business Problem

- Globalization leads to greater urbanization
- Cities need to adapt to maintain their attractiveness as a place to live, work and visit – to have a healthy population and sustainable economic growth
- Challenge for city governments and urban planners: How to build a top-tier, future-ready city in anticipation of a big population growth? How should they invest in and develop their cities?

Benchmark Cities

For the purpose of this project, two high-ranked cities shall be studied:

1. Singapore
   (ranked 25 in Mercer Quality of Living Ranking 2019)
2. New York City
   (ranked 44)



Study Approach

- 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

Data acquisition and cleaning

Tools: Foursquare API, Wikipedia, Geopy library

Data scraping from Wikipedia

Singapore neighborhood data scraped

Geopy library

Get geographical coordinates for each neighborhood

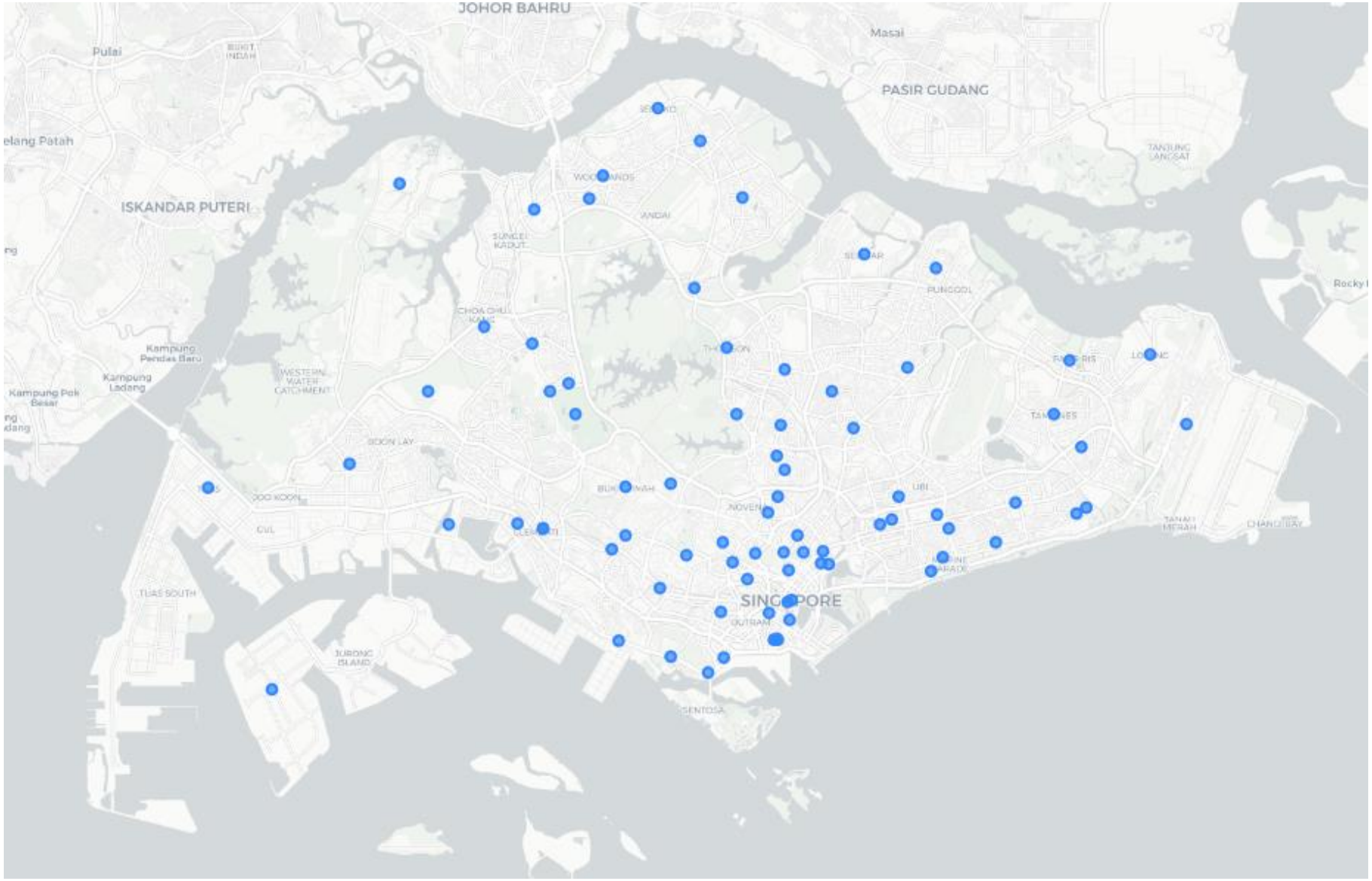
Get Foursquare venues data

Use Python function to retrieve venues, their category names

Transform to useable dataframe

Run function to list all relevant venues – to be used for clustering

Mapping Singapore‘s neighborhoods



Finding Venues in Singapore using Foursquare

Getting venues for each neighborhood and ranking the most common venues

	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

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

	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	Martial Arts School	Dim Sum Restaurant	Seafood Restaurant	Japanese Restaurant	Kitchen Supply Store	Train Station	Fish & Chips Shop
1	Amber Road	Bus Stop	Café	Chinese Restaurant	Club House	Bar	Women's Store	Fish & Chips Shop	French Restaurant	Food Stand	Food Court
2	Ang Mo Kio	Dessert Shop	Japanese Restaurant	Snack Place	Bubble Tea Shop	Sushi Restaurant	Coffee Shop	Fast Food Restaurant	Multiplex	Sandwich Place	Seafood Restaurant
3	Beach Road	Vietnamese Restaurant	Food Court	Dessert Shop	Hotel	Coffee Shop	Burger Joint	Café	Multiplex	Taiwanese Restaurant	Thai Restaurant
4	Braddell	Fast Food Restaurant	Bus Station	Food Court	Metro Station	Bakery	Filipino Restaurant	French Restaurant	Food Stand	Food & Drink Shop	Food

Mapping New York City's neighborhoods

Focusing on Brooklyn

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.

Finding Venues in Brooklyn using Foursquare

Getting venues for each neighborhood and ranking the most common venues

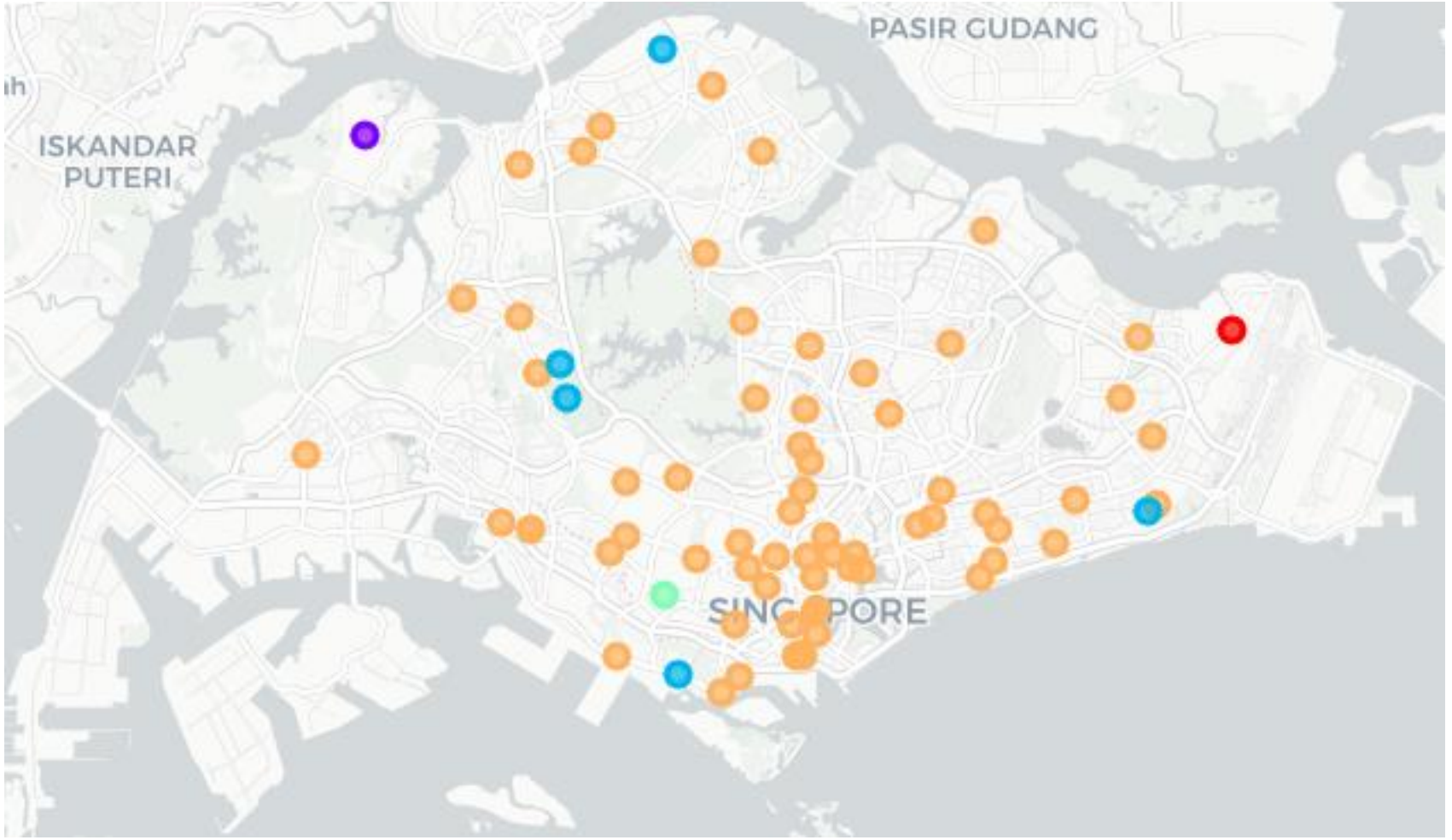
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
2	Bay Ridge	40.625801	-74.030621	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

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

	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	Check Cashing Service	Italian Restaurant	Liquor Store	Gas Station	Flower Shop	Donut Shop	Diner	Pizza Place	Hookah Bar
1	Bay Ridge	Italian Restaurant	Mediterranean Restaurant	Pizza Place	Spa	Greek Restaurant	Optical Shop	Sushi Restaurant	Bookstore	Lounge	Pool Hall
2	Bedford Stuyvesant	Coffee Shop	Boutique	Italian Restaurant	Café	Juice Bar	Yoga Studio	Factory	Food	Flower Shop	Fish Market
3	Bensonhurst	Italian Restaurant	Park	Yoga Studio	Event Service	Food	Flower Shop	Fish Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
4	Boerum Hill	Furniture / Home Store	Pharmacy	Thrift / Vintage Store	Men's Store	Middle Eastern Restaurant	French Restaurant	Spa	Kids Store	Sushi Restaurant	Cocktail Bar

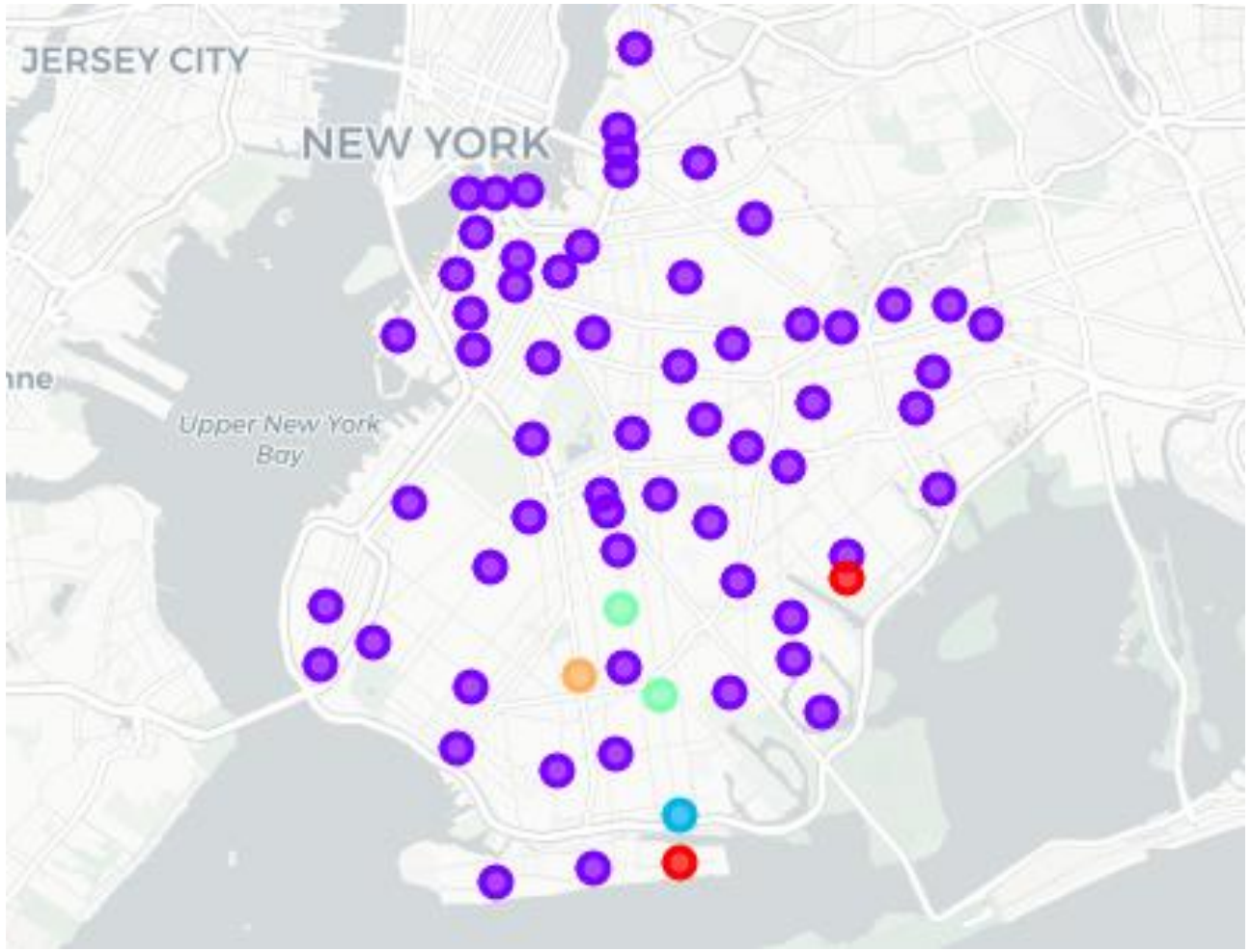
Clustered neighborhoods in Singapore

Result: 1 cluster contains the majority of neighborhoods; the other 4 clusters have 1-5 neighborhoods only



Clustered neighborhoods in Brooklyn

Result: 1 cluster contains the majority of neighborhoods; the other 4 clusters have 1-5 neighborhoods only



Analysis: K-Means Clustering

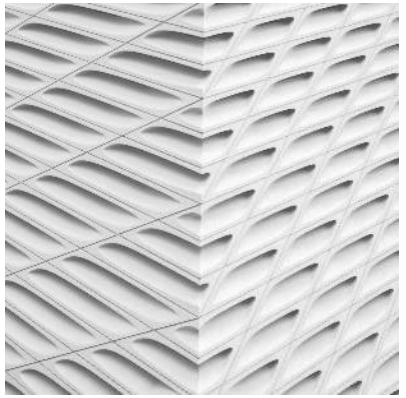
Cluster descriptions - using k = 5

Singapore

- Cluster 1 (red) – Loyang >> 1 venue 'Bus Station'.
- Cluster 2 (purple) – Lim Chu Kang >> 1 venue 'Farm'.
- Cluster 3 (blue) – 5 neighborhoods: Telok Blangah, Bukit Timah, Kew Drive, Dairy Farm, Senoko.
  >> Similarities of the 5 neighborhoods seem to be from the outdoors and transport categories: ‘Trail’, ‘Bus’ and ‘Train Stops’.
- Cluster 4 (green) – Queenstown >> 1 venue 'Train station'.
- Cluster 5 (orange) – 65 neighborhoods
  >> 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 top-ranked.

Brooklyn

- Cluster 1 (red) - Manhattan Beach and Paerdegat Basin
  >> Similarities of these 2 neighborhoods: 'Beach' category.
- Cluster 2 (purple) – 61 neighborhoods in Brooklyn
  >> 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 >> 1 venue for this neighborhood with the category 'Lake'.
- Cluster 4 (light green) – Midwood and Madison
  >> Both only had 1 venue each: 'Bus station'
- Cluster 5 (orange) – Ocean Parkway >> 1 venue 'Outdoors & Recreation'.



Results

Slogan oder Unterüberschrift
für Übersicht

- Based on the clustering results, it can be said that both Singapore and Brooklyn are highly urbanized cities where residents can live, work and play throughout the city.
- One cluster contains a big majority of the neighborhoods – meaning that these neighborhoods are homogenous.
- The remaining 4 clusters are the exceptions. Some are have mostly outdoor spots or only have transportation venues.

Finding: Decentralizing cities is key

Decentralizing critical venues such as retail, food, nature, workplaces reduces congestion, improves service offerings and allows for better accessibility for many residents.

In the end, spreading out such important services enables a higher quality of living. This is what future-centric cities should work towards, especially as more and more people move into them in the next years and decades.

Limitations and Suggestions for Future Research

Observations during the project

Foursquare data limited

Foursquare results are not an accurate representation of the variety of venues Singapore and Brooklyn have. Results change every time the query is run.

More focused data sources needed

Type of information provided by Foursquare is lacking and more data sources more focused on urban planning factors are required.

Larger study with more cities needed

Study is limited to two cities. A larger-scale study covering more cities would glean more accurate insights to our business problem

Conclusions

- 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.
- 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.