Analyzing Cities: A Case Study of Singapore and New York

MICHAEL W.D. KURZ

“You have an impeccable argument if you said that Singapore, Hong Kong, and Tokyo are food capitals. They have a maximum amount of great stuff to eat in the smallest areas.”

- ANTHONY BOURDAIN

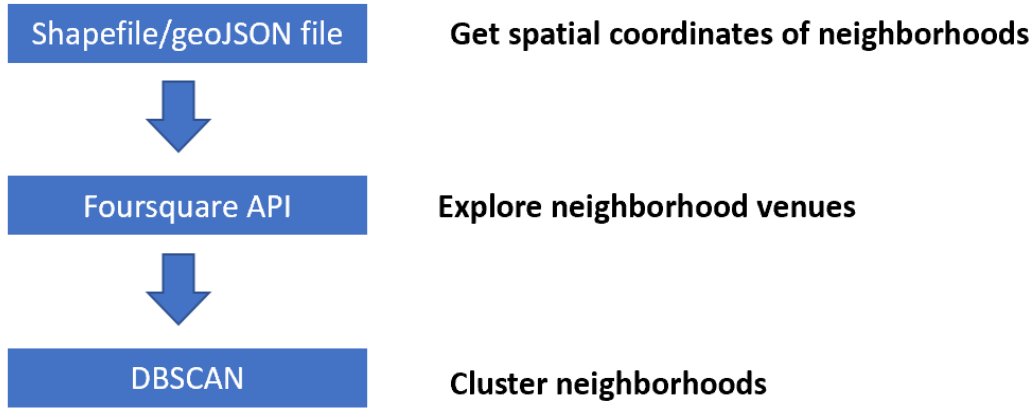
Analyzing Cities

Different neighborhoods of major cities are often described by residents as having their own DNA

This preliminary analysis puts this sentiment to the test in two major cities: Singapore and New York City

Clusters of similar neighborhoods within both cities are identfied using the DBSCAN algorithm based on Foursqure venues data

Workflow for clustering of neighborhoods in each city

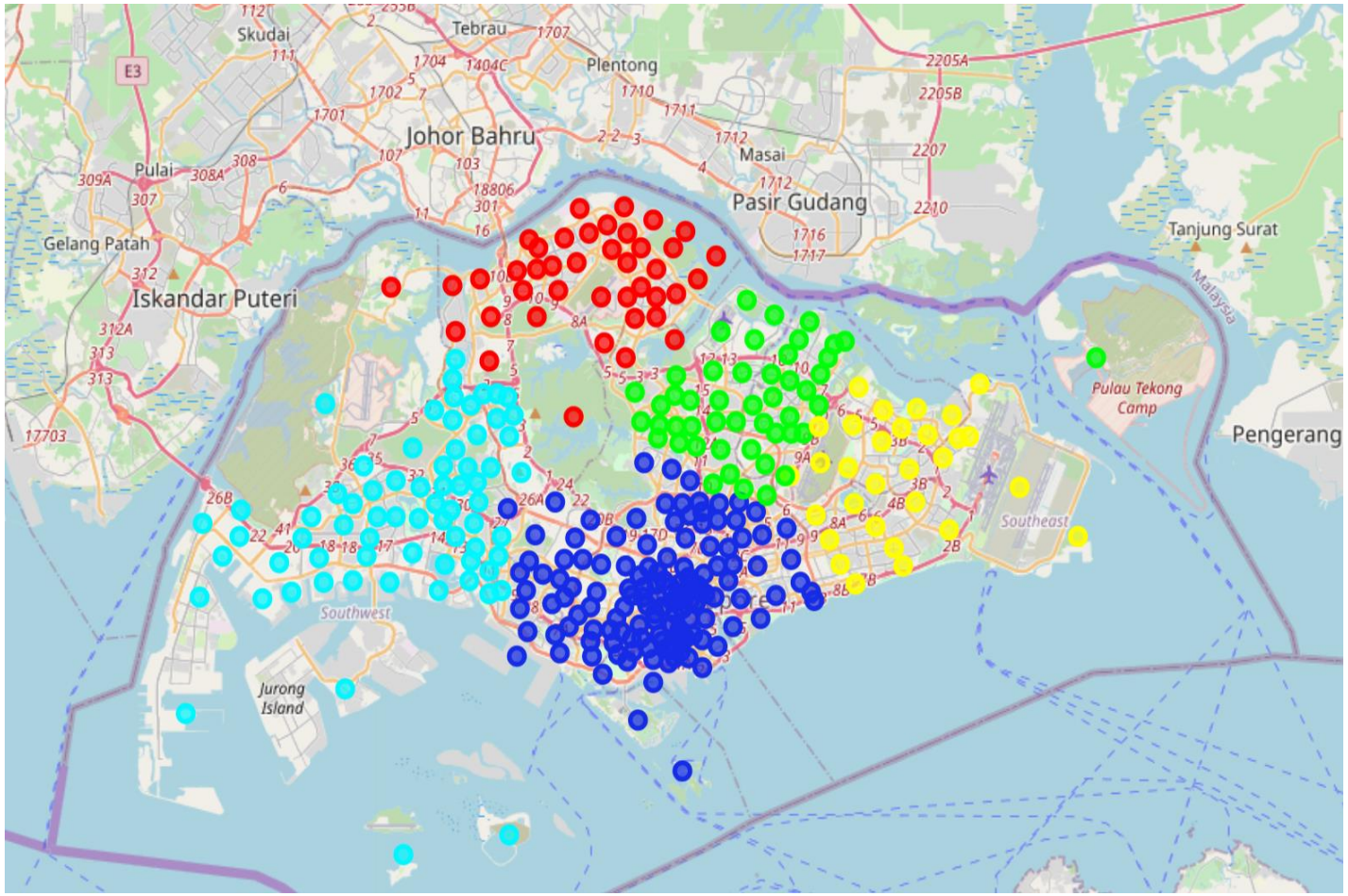


Singapore

323 neighborhoods of **Singapore's regions:**

Central region (blue), East region (yellow), North region (red), North-East region (green), and West region (aqua)

- >5.000 venues
- 152 venue categories

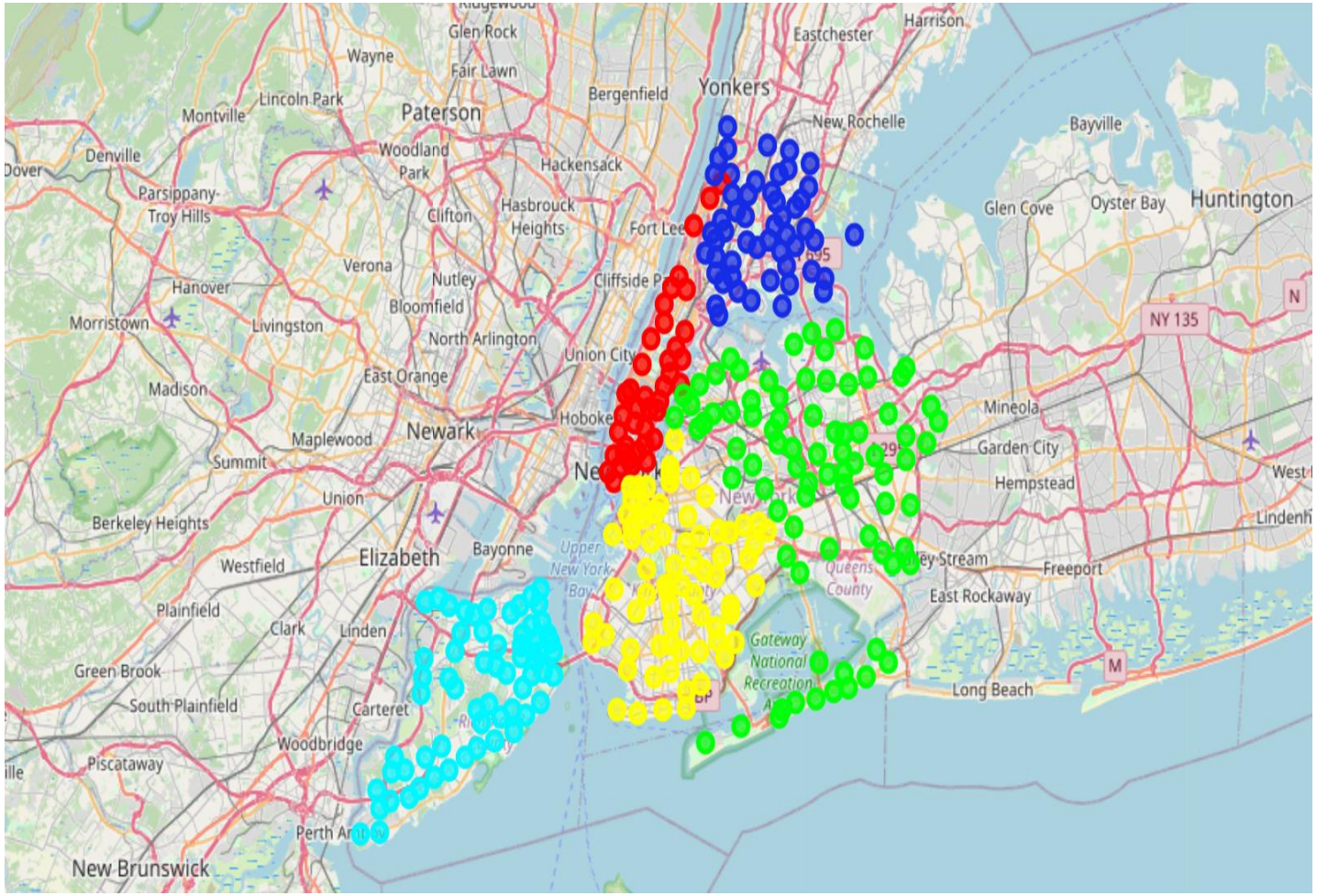


New York City

306 neighborhoods of **New York's boroughs:**

Bronx (blue), Brooklyn (yellow), Manhattan (red), Queens (green), and Staten Island (aqua)

- >5.000 venues
- 193 venue categories



Methodology

Within cities:

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Idea: Find areas of high density within the dataset that are separated by areas of low density

Define distance/dissimilarity of neighborhoods via Manhattan distance based on frequency of venue categories

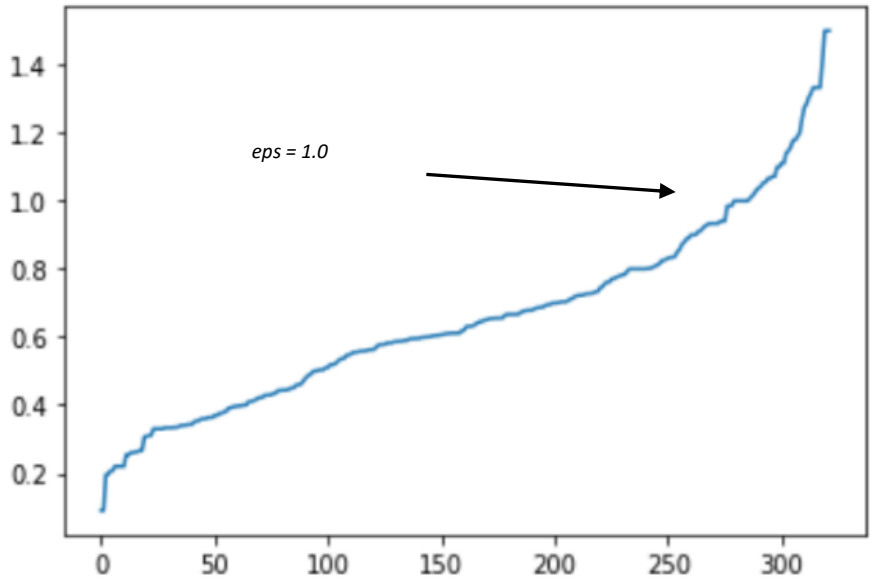
Across cities:

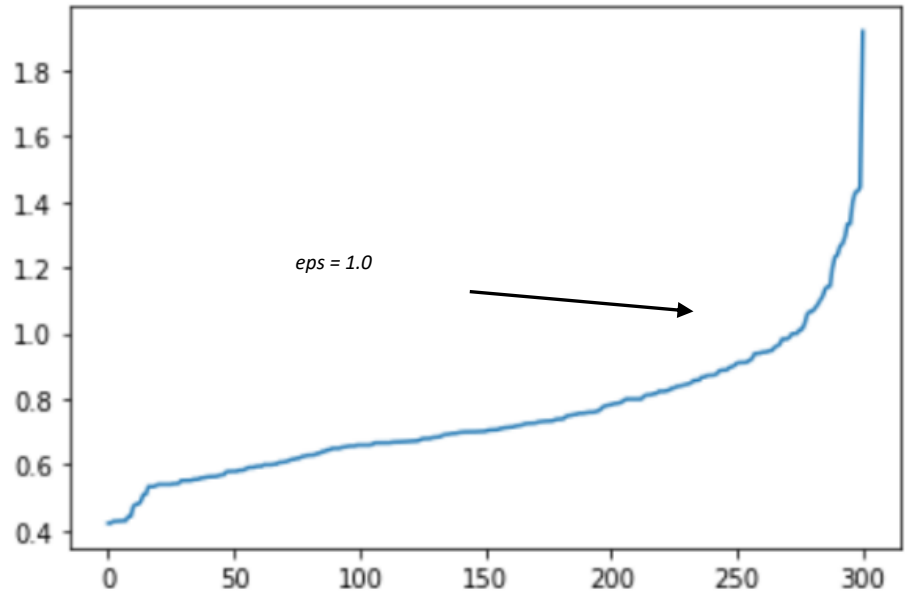
Agglomerative hierarchical clustering

Present complete hierarchy nested groups of clusters found within the cities

Results can be easily visualized in a dendrogram

DBSCAN tuning





Number of neighborhoods in clusters vs. regions/boroughs

Cluster	Sing_1	Sing_2	Sing_3	Sing_4	Sing_5	Total
# of neighborhoods	31	283	5	2	2	323
Regions	Central	North	North-East	East	West	Total
# of neighborhoods	134	41	48	30	70	323

Cluster	Nyc_1	Nyc_2	Nyc_3	Nyc_4	-	Total
# of neighborhoods	27	268	3	2	-	306
Borough	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
# of neighborhoods	52	70	40	81	63	306

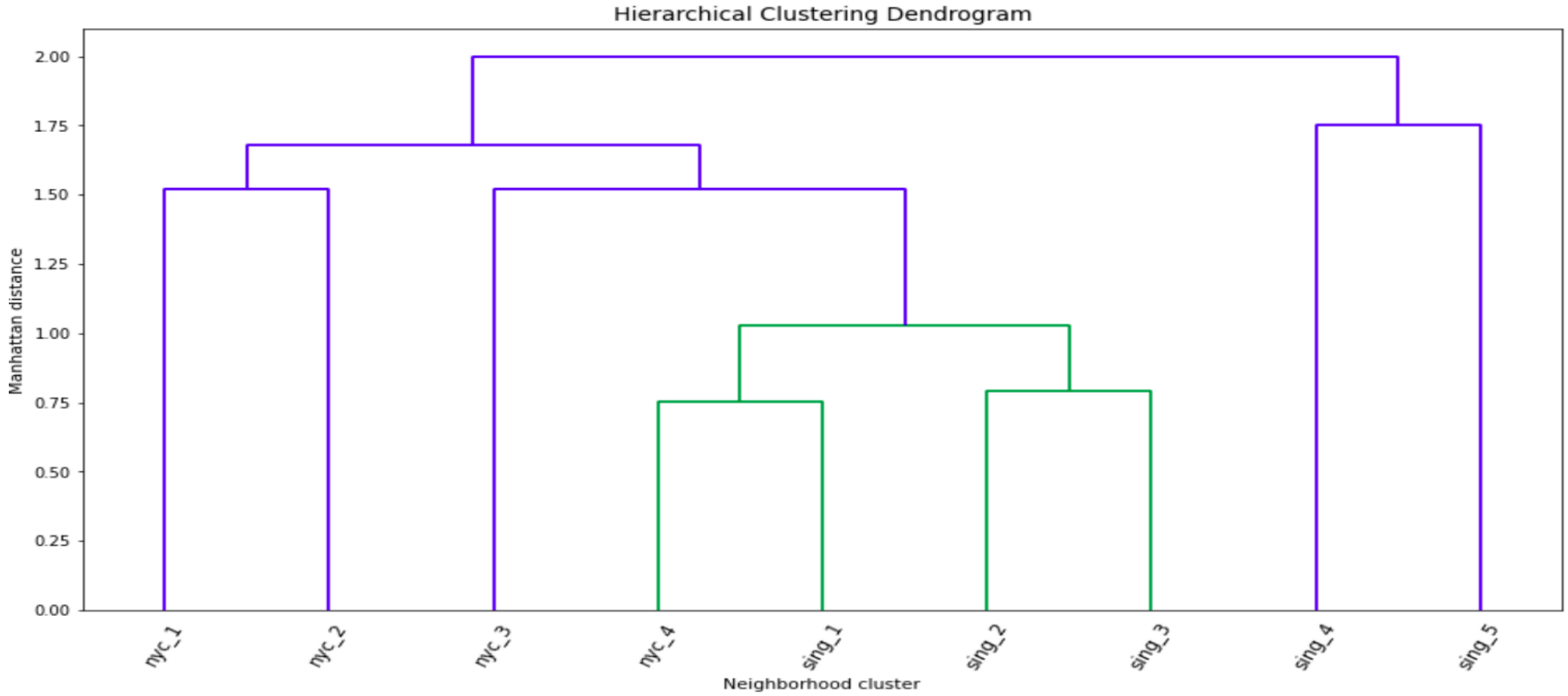
Most common venue categories in Singapore's clusters

Cluster Labels DB	1th Most Common Venue	2th Most Common Venue	3th 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
sing_1	Coffee Shop	Public Transport	Shop & Service	Trail	Bar	Chinese Restaurant	Tourist Accommodation	Park	Gym	Seafood Restaurant
sing_2	Shop & Service	Coffee Shop	Chinese Restaurant	Food Court	Bar	Japanese Restaurant	Public Transport	Asian Restaurant	Tourist Accommodation	Fast Food Restaurant
sing_3	Coffee Shop	Bar	Gym	Food General	Professional & Other Places	Playground	Pizza Place	Food Court	Campground	Brewery
sing_4	Other Great Outdoors	Island	Public Transport	Beach	Pier	Tourist Information Center	Exhibit	Factory	Fast Food Restaurant	Filipino Restaurant
sing_5	River	Seafood Restaurant	Public Transport	Burger Joint	Coffee Shop	Harbor / Marina	Food Court	Fried Chicken Joint	French Restaurant	Food Truck

Most common venue categories in New York City's clusters

Cluster Labels DB	1th Most Common Venue	2th Most Common Venue	3th 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
nyc_1	Beach	Shop & Service	Bar	Tourist Accommodation	Park	Dessert Shop	Thai Restaurant	Pizza Place	Gym	Public Transport
nyc_2	Shop & Service	Bar	Gym	Coffee Shop	Pizza Place	Dessert Shop	Italian Restaurant	Deli / Bodega	Chinese Restaurant	Bakery
nyc_3	Athletics & Sports	Dessert Shop	Playground	Park	Gym	Harbor / Marina	Mexican Restaurant	Deli / Bodega	Pizza Place	Chinese Restaurant
nyc_4	Caribbean Restaurant	Shop & Service	Gym	Music Venue	Pizza Place	Bakery	Thai Restaurant	Dessert Shop	Playground	Chinese Restaurant

How similar are clusters across cities?



Conclusions

Remarkable similarities in clustering patterns across Singapore and New York City

For both cities, most neighborhoods are simialar and fall into one lare cluster

The second largest clusters in each city are “low-density” clusters comprised of relatively dissimilar clusters (from each other and from neighborhoods in other clusters)

However, the two “super” clusters comprising most neighborhoods are relatively different from each other

→ In many aspects Singapore and New York seem rather similar and especially similarities of neighborhoods within each city are high.