

The Battle of Neighborhoods

-- Clustering and Comparing the Neighborhoods of New York City, Toronto and Paris
Tianyan Wang Columbia University

1) Introduction:

In this project, we utilized location data to analyze, cluster and compare neighborhoods in New York, Toronto and Paris to detect the differences and similarities of the distribution of business over these cities.

We identified the most dense and welcome neighborhood business that these cities share in common. Business landers may refer to this study to obtain basic ideas about which kind of business is more likely to thrive and start with and which is not very desirable in these world-famous cities. Stakeholders with stocks relevant to these cities can better understand certain local economic drives there and improve their decision models.

It's well recognized that these cities are most popular over the seas. With an estimated 2020 population of 8,336,817 distributed over about 784 square kilometers, New York City is also the most densely populated major city in the United States. Meanwhile, Toronto, the capital of Canada, is often described as "New York City run by the Swiss" with world-class theater, shopping and restaurants. Paris is the capital and most populous city of France, with an estimated population of 2,148,271 residents as of 2020, in an area of 105 square kilometers, and has been one of Europe's major centers of finance, diplomacy, commerce, fashion, science and arts. A large amount of capitals and opportunities are flowing through these dynamic cities. Therefore, it's worth it to explore and analyze the patterns of the business distribution to deliver insights for commercial kickoff, extension and development.

2) Data:

The data to be used in this project is basically the neighborhood data, and venue information obtained from Foursquare. We're supposed to gather the neighborhoods in each city and then check the venues inside each neighborhood and which category they belong to.

The features in the neighborhood dataframe should contain the name of neighborhoods, the location (longitude, latitude) and hopefully the postal code.

The features in the venue information dataframe should include the name of the venue, the neighborhoods it currently belongs to, and the venue's type of business. The dataframe will be merged and aggregated with one-hot encoding to find the business distribution of these cities, which is the fundamentals of the following clustering operations.

3) Methodology:

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different.

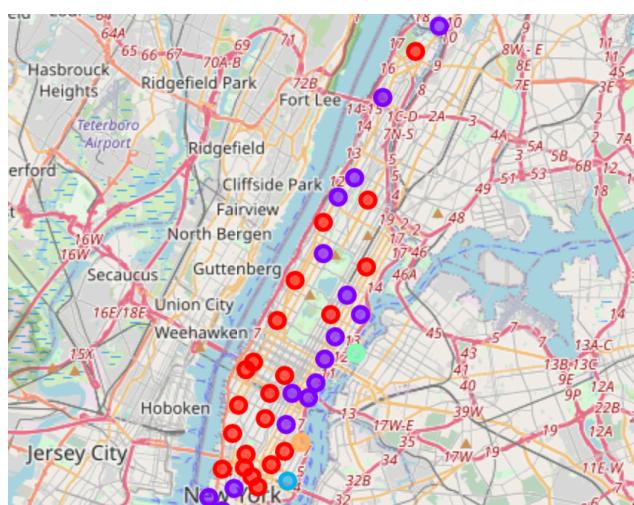
Kmeans algorithm is an iterative algorithm that tries to partition the dataset into K predefined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid.

Another way, which is more advanced in data mining, is the hierarchical clustering. Hierarchical clustering is separating data into groups based on some measure of similarity, finding a way to measure how they're alike and different, and further narrowing down the data.

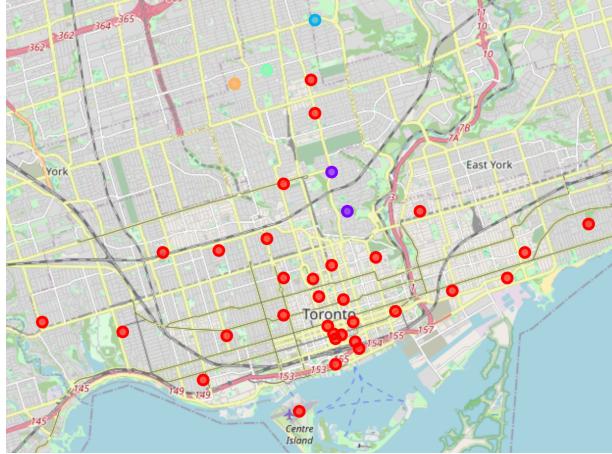
4) Results:

First, I utilized K-means method to cluster the neighborhoods based on the type of business of their venues for each city. The results are as follows.

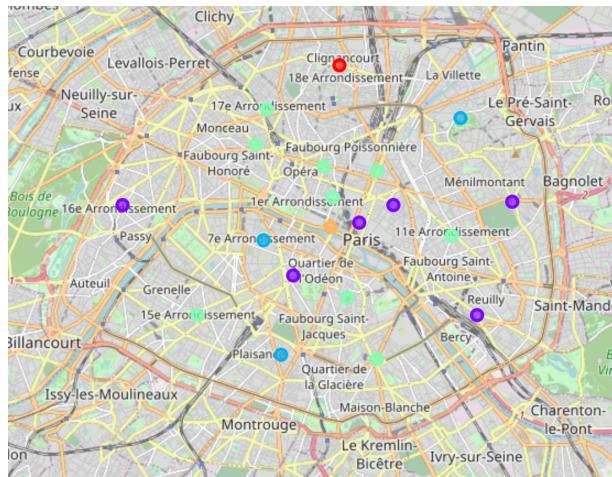
New York City has seen diverse categories of neighborhoods around midtown and southern part. Toronto has similar neighborhoods in the central area while Paris has more neighborhoods of main different functions.



Clustering of NYC neighborhoods



Clustering of Toronto neighborhoods



Clustering of Paris Neighborhoods

Then I aggregated all the cities' neighborhood data and do the overall clustering, in order to find the similarity of the business distribution in these three cities and got 6 clusters.

Cluster 0

	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
78	Paris 02 Bourse	French Restaurant	Bistro	Italian Restaurant	Bakery	Pizza Place	Clothing Store	Hotel	Restaurant	Wine Bar	Wine Shop
81	Paris 05 Panthon	French Restaurant	Bar	Italian Restaurant	Pub	Creperie	Sushi Restaurant	Plaza	Bakery	Lebanese Restaurant	Ice Cream Shop
83	Paris 07 Palais-Bourbon	French Restaurant	Café	Restaurant	Coffee Shop	Japanese Restaurant	Park	Embassy / Consulate	Brasserie	Bistro	Bakery
84	Paris 08 l'Île	French Restaurant	Hotel	Pizza Place	Bistro	Sandwich Place	Wine Shop	Sushi Restaurant	Bakery	Chinese Restaurant	Diner
85	Paris 09 Opéra	French Restaurant	Furniture / Home Store	Hotel	Bistro	Bar	Concert Hall	Candy Store	Mexican Restaurant	Fast Food Restaurant	Salad Place
90	Paris 14 Observatoire	French Restaurant	Wine Shop	Bar	Bakery	Vietnamese Restaurant	Ice Cream Shop	Pizza Place	Restaurant	Tea Room	Brasserie
91	Paris 15 Vaugirard	French Restaurant	Coffee Shop	Bar	Hotel	Indian Restaurant	Restaurant	Lebanese Restaurant	Bakery	Indie Movie Theater	Japanese Restaurant
93	Paris 17 Batignolles-Monceau	Hotel	French Restaurant	Pizza Place	Bar	Thai Restaurant	Gym / Fitness Center	Theater	Japanese Restaurant	Bakery	Asian Restaurant
94	Paris 18 Butte-aux-Cailles	Bar	French Restaurant	Seafood Restaurant	Café	Vietnamese Restaurant	Wine Bar	Sandwich Place	Restaurant	Beer Bar	Gastropub
95	Paris 19 Butte-aux-Cailles	French Restaurant	Historic Site	Park	Moroccan Restaurant	Scenic Lookout	Diner	Falafel Restaurant	Embassy / Consulate	Doner Restaurant	Discount Store

Cluster 1

	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
58	Toronto_Moore Park Summerhill East	Park	Yoga Studio	Distribution Center	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant	Escape Room	Electronics Store	Dumpling Restaurant
64	Toronto_Rosedale	Park	Yoga Studio	Distribution Center	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant	Escape Room	Electronics Store	Dumpling Restaurant

Cluster 2

	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
56	Toronto_Lawrence Park	Photography Studio	Distribution Center	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant	Escape Room	Electronics Store	Dumpling Restaurant	Donut Shop

Cluster 3

	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
59	Toronto_North Toronto West Lawrence Park	Sushi Restaurant	Yoga Studio	Dog Run	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant	Escape Room	Electronics Store	Dumpling Restaurant

Cluster 4 (*partial result*)

	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	newyork_Battery Park City	Park	Coffee Shop	Gym	Sandwich Place	Food Court	Food Truck	Gourmet Shop	Boat or Ferry	Gastropub	Boutique
1	newyork_Carnegie Hill	Café	Italian Restaurant	Gym / Fitness Center	Gym	Coffee Shop	Sushi Restaurant	Bar	Wine Bar	Mobile Phone Shop	Shoe Store
2	newyork_Central Harlem	African Restaurant	Fried Chicken Joint	Caribbean Restaurant	Deli / Bodega	Burger Joint	Breakfast Spot	Ethiopian Restaurant	Beer Bar	Library	Lounge
3	newyork_Chelsea	Coffee Shop	French Restaurant	Theater	Hotel	Pizza Place	Cocktail Bar	Noodle House	Nightclub	Fish Market	Sushi Restaurant
4	newyork_Chinatown	Bubble Tea Shop	Chinese Restaurant	Bakery	Hotpot Restaurant	Salon / Barbershop	Vietnamese Restaurant	Noodle House	Sandwich Place	Cocktail Bar	Shanghai Restaurant

Cluster 5

	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
65	Toronto_Roselawn	Health & Beauty Service	Yoga Studio	Distribution Center	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Ethiopian Restaurant	Escape Room	Electronics Store	Dumpling Restaurant

To get a more clear idea of the business distribution, 6 frames are generated to show the percentage of the business venues in each cluster.

For cluster 0 which is generally a cluster inside Paris, French restaurants are very hot spots, followed by bars, hotels and pizza places.

For cluster 1, 2, 3 and 5, only respectively one Toronto neighborhood is extracted at each cluster. This can be attributed to the cultural construction in these areas. For example,

cluster 1 is for leisure and cluster 2 is for Asian style food and cluster 5 is for large hospitals.

Business Distribution in Cluster 0

	Percentage
French Restaurant	20.563380
Bar	7.323944
Hotel	7.323944
Pizza Place	3.380282
Italian Restaurant	3.098592
Bakery	3.098592
Restaurant	2.535211
Café	2.253521
Bistro	2.253521
Japanese Restaurant	1.971831
Wine Bar	1.971831

Business Distribution in Cluster 1

	Percentage
Park	100.0

Business Distribution in Cluster 2

	Percentage
Sushi Restaurant	100.0

Business Distribution in Cluster 3

	Percentage
Photography Studio	100.0

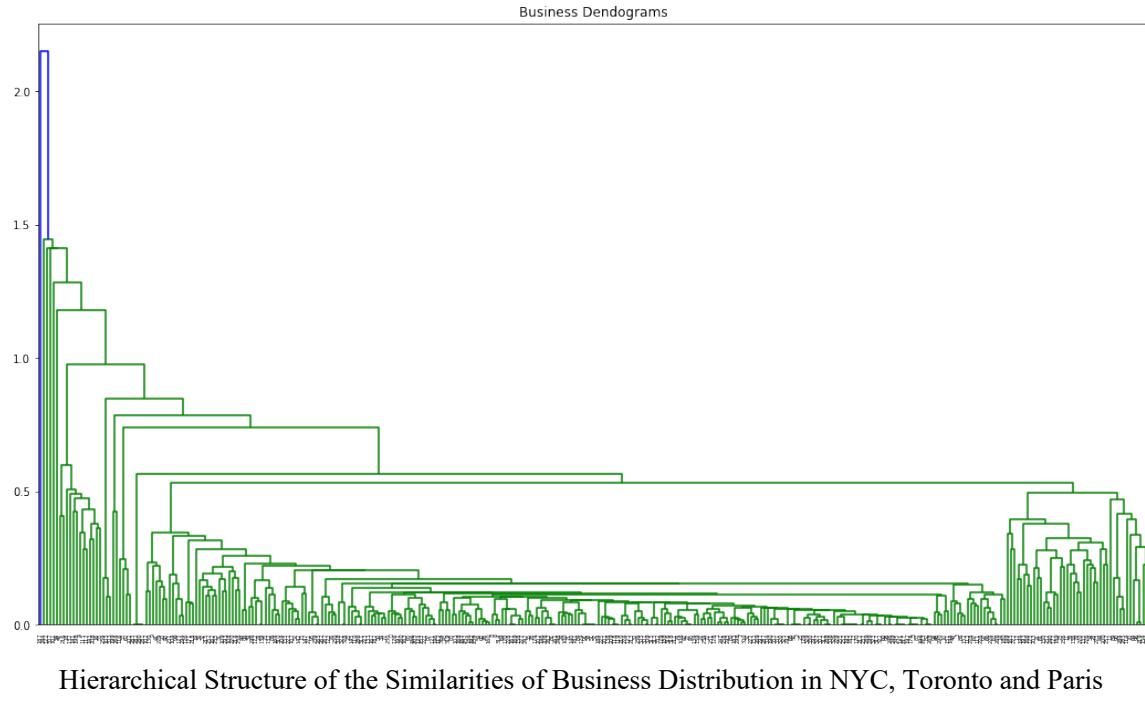
Business Distribution in Cluster 4

	Percentage
Coffee Shop	5.980066
Café	3.554817
Italian Restaurant	3.089701
Hotel	2.192691
Bar	2.026578
Pizza Place	2.026578
Sandwich Place	2.026578
Bakery	1.993355
American Restaurant	1.893688
Japanese Restaurant	1.727575
Park	1.661130

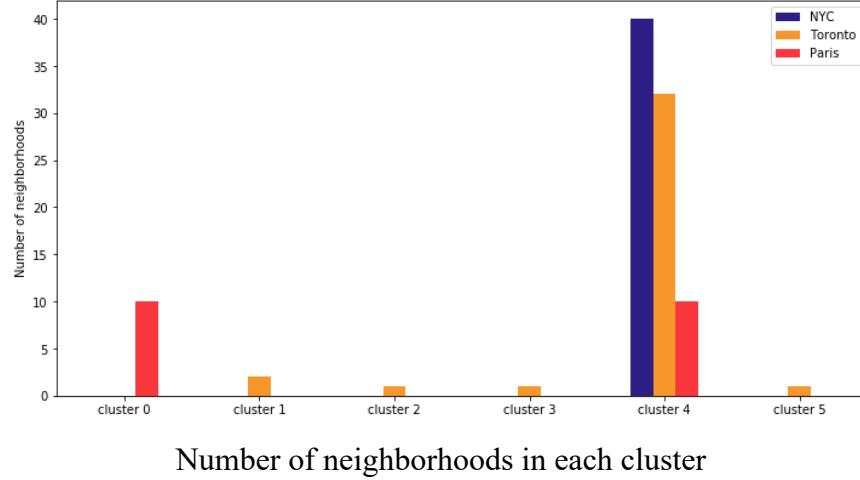
Business Distribution in Cluster 5

	Percentage
Health & Beauty Service	100.0

To show a hierarchical structure of the similarities of business distribution in these three cities, I also introduced the shc method. The result is as below.

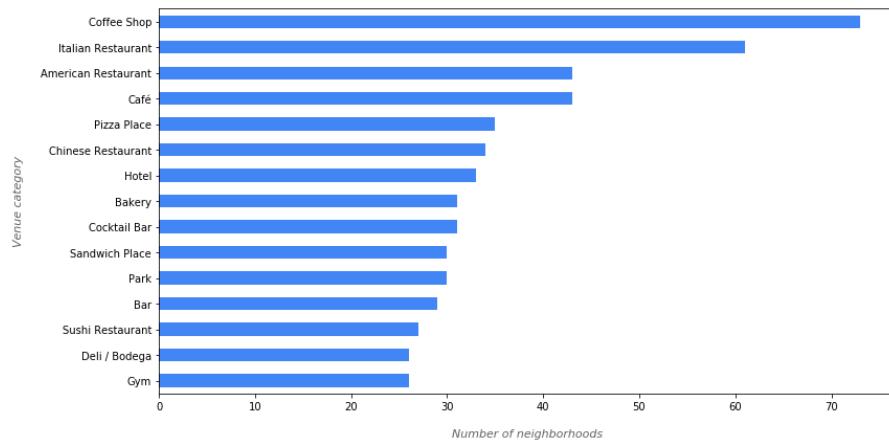


I also analyzed the components of each cluster to find out how each city's neighborhoods resemble one another.

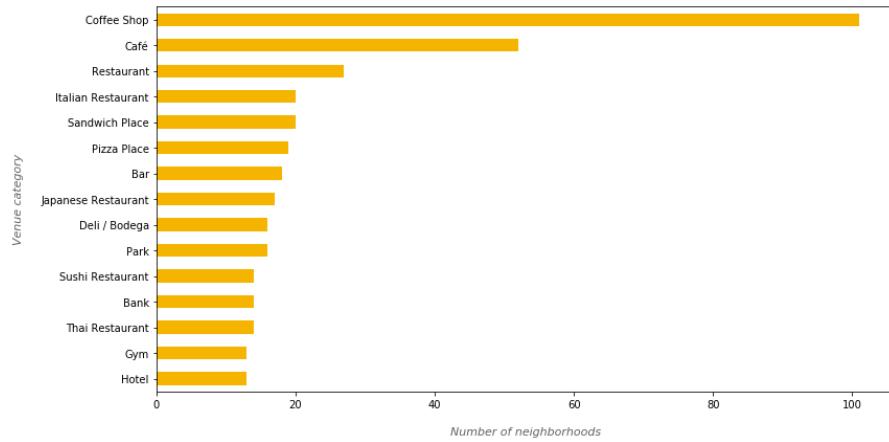


From this graph we can see that despite some unique neighborhoods in Toronto and Paris, all the New York neighborhoods have some kind of reflection with both some of Toronto neighborhoods and certain Paris neighborhoods. The similarities basically lies in the cluster 4, while the differences are distributed in the other clusters.

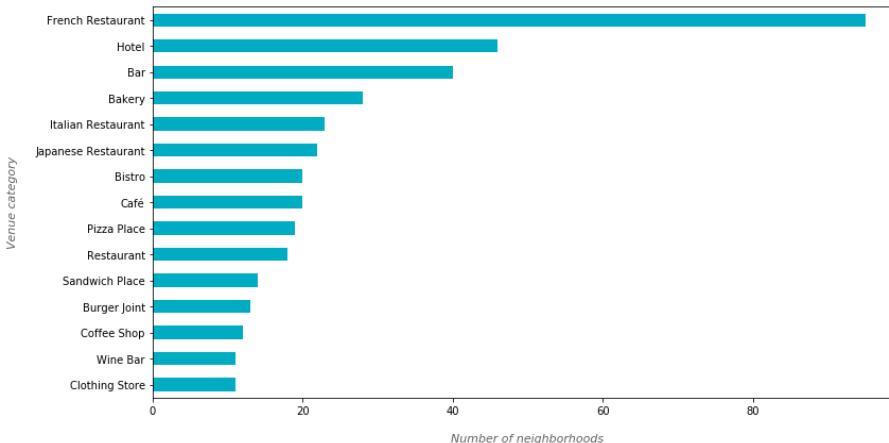
Finally, I organized the most and least popular business in each cities to unveil the trends there for the consideration of the potential peer competition and opportunities.



Top Hot spots in New York City



Top Hot spots in Toronto



Top Hot spots in Paris

The Least common spots in New York City

Venue Category	Count
Storage Facility	1
Sports Club	1
Medical Center	1
Health & Beauty Service	1
Antique Shop	1
Fast Food Restaurant	1
Metro Station	1
Discount Store	1
Fountain	1
Molecular Gastronomy Restaurant	1
Outdoor Sculpture	1
Climbing Gym	1
College Bookstore	1
Social Club	1
Ethiopian Restaurant	1

The Least common spots in Toronto

Venue Category	Count
Fruit & Vegetable Store	1
History Museum	1
Butcher	1
Performing Arts Venue	1
Discount Store	1
Indie Movie Theater	1
Thrift / Vintage Store	1
Opera House	1
Jewelry Store	1
Falafel Restaurant	1
Distribution Center	1
Playground	1
Board Shop	1
Martial Arts School	1
Speakeasy	1

The Least common spots in Paris

Venue Category	Count
Hookah Bar	1
Comic Shop	1
Spanish Restaurant	1
Nightclub	1
Gourmet Shop	1
Arcade	1
Sporting Goods Shop	1
Ramen Restaurant	1
Alsatian Restaurant	1
Vegetarian / Vegan Restaurant	1
South American Restaurant	1
Science Museum	1
Church	1
Beer Store	1
Scenic Lookout	1

5) Recommendations:

Based on the analysis of similarities, the neighborhoods in cluster 4 enjoy a dense distribution of coffee shops, cafés and Italian restaurants. Hotels, bar and sandwich stores are also recommended there. These places are mainly about leisure, meeting, staying and catering. Potential opportunities and competition are crucial for market research there, if business owners want to land expanded shops there. It's also plausible to launch a successful business considering the types that are not common in one neighborhood but are common in most of others in the cluster.

Based on the analysis of the uniqueness, Paris and Toronto have some neighborhoods of their own culture concepts or functions. It's desirable to develop more business relevant to the core of these cultural infrastructures. For example, the neighborhood cluster 2 is sharing the sushi culture, so the raw material business, sushi making and processing business and restaurant promotion business are more likely to be developed around this neighborhood.

6) Conclusions:

In this project, we utilized the location data to extract the patterns (distribution) of business in the neighborhoods of New York City, Toronto and Paris. Based on the result of clustering method and analysis of differences and similarities of the business density, valid recommendations are made to help business owners start, expand and develop business in these cities and leave protocol for solving market entering issues.