
Compare Global Cities:

New York City, Toronto, London, and Paris

August 19, 2019



Introduction

New York City, Toronto, London and Paris are all global cities. They share characteristics of large population, diverse cultures, important roles in economics, and etc. Every year, numerous people visit there, seek job opportunities there, even start up their own business there.

I have never been to these cities except Paris. So, I'm going to compare them to find their similarities and dissimilarities. Expecting I would enjoy more if I had the chance to travel to.

Data Source

Overall look

New York City has 5 boroughs and 57 communities. Each community has numbers of neighborhoods. But their names and borders are not officially defined, and they change from time to time.

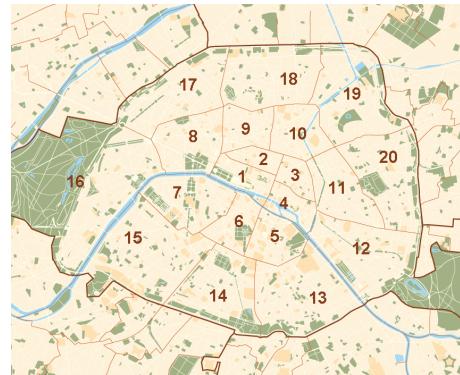


Toronto has 140 officially recognized neighbourhoods in 9 districts. Old Toronto is divided into 4 districts.

London area is the largest, it has 32 boroughs and City of London. So we can count as 33 boroughs.



Paris only have 20 arrondissements, somewhat like boroughs in New York City or London, but they are rather small. And it's hard to define smaller units under arrondissement.



But if we analyze nearby venue categories based on this range, it seems make little sense. The areas of smallest units of each city range broad.

City	Population (million)	Area (km2)	Sub-district numbers	Sub-sub-district numbers
New York City	20.0	784	5	57
Toronto	2.7	630	9	140
London	8.8	1572	33	None
Paris	2.1	105	20	None

What if we compare the four cities' center parts? Considering Paris is too small, so take the whole Paris as the center part. (Divide 12th and 16th arrondissement in Paris to two parts respectively. I'll explain this later in the report.)

Now it seems that venues analysis based on center parts and their sub-districts makes more sense.

City	Borough / District	Area (km2)	Sub-district numbers
New York City	Manhattan	59	40
Toronto	Old Toronto	97	38
London	Inner London	319	183
Paris	Paris	105	22

New York City & Toronto

Manhattan's and Old Toronto's data comes from the courses introduced before.

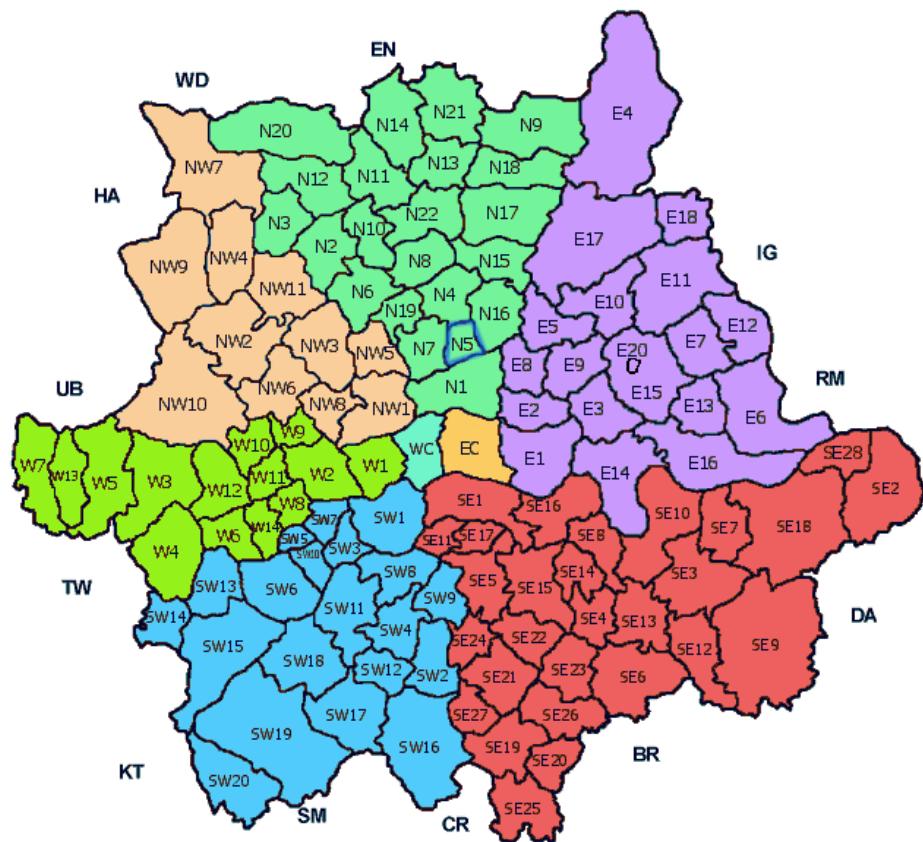
London

Similarly to Toronto, London has post codes representing each area.

London postcodes CSV file download from here:

https://www.doogal.co.uk/london_postcodes.php.

Data from the CSV is rather large. It needs to be cleaned.



Paris

Paris have 20 arrondissements, some kind like boroughs, but much smaller and hard to split into smaller units like neighborhood. So I will use arrondissement directly instead neighborhood. To get coordinates, I need to do web scape work.

This page <http://zip-code.en.mapawi.com/france/1/arrondissement-de-paris/3/150/> concludes each arrondissement's link. For example, click on "Paris 01 Louvre", we get this:

Postcode / zip code of Paris 01 Louvre

75001

Administrative division of Paris 01 Louvre

District :Île-de-France A8

County :Paris 75

District :Arrondissement de Paris

GPS coordinates of Paris 01 Louvre

Latitude : 48.8592

Longitude : 2.3417

So, I need to extract info from this page and the other 19 pages to get Paris arrondissements data.

Since 12th and 16th arrondissements have areas apparently larger than others. From the map, we can tell that it is because of green land part. So, i will add another two coordinates to Paris data, 48.8624°N, 2.2492°E for Bois de Boulogne in 16th arrondissement, and 48.8283°N, 2.4330°E for Bois de Vicennne in 12th arrondissement.

Explore & Cluster

Aggregate all the four cities' sub-district data into one dataframe with column names: City, District, Sub-district, Latitude, Longitude.

For example:

City	District	Sub-district	Latitude	Longitude
New York City	Manhattan	Marble Hill	40.876551	-73.910660
New York City	Manhattan	Chinatown	40.715618	-73.994279
New York City	Manhattan	Washington Heights	40.851903	-73.936900
New York City	Manhattan	Inwood	40.867684	-73.921210
New York City	Manhattan	Hamilton Heights	40.823604	-73.949688

Totally there are 283 rows / sub-districts.

Using Foursquare to get each coordinates's nearby venues.

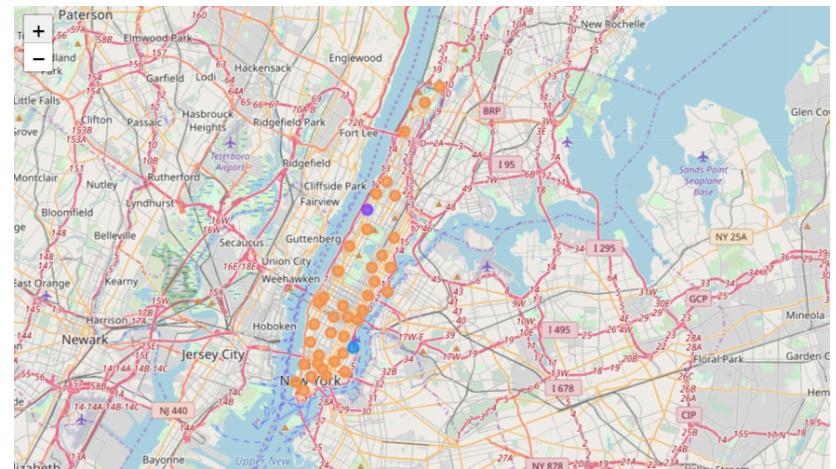
City	District	Sub-district	Sub-district Latitude	Sub-district Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
New York City	Manhattan	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
New York City	Manhattan	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
New York City	Manhattan	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
New York City	Manhattan	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
New York City	Manhattan	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

Then collect the top 10 popular venue categories into a dataframe.

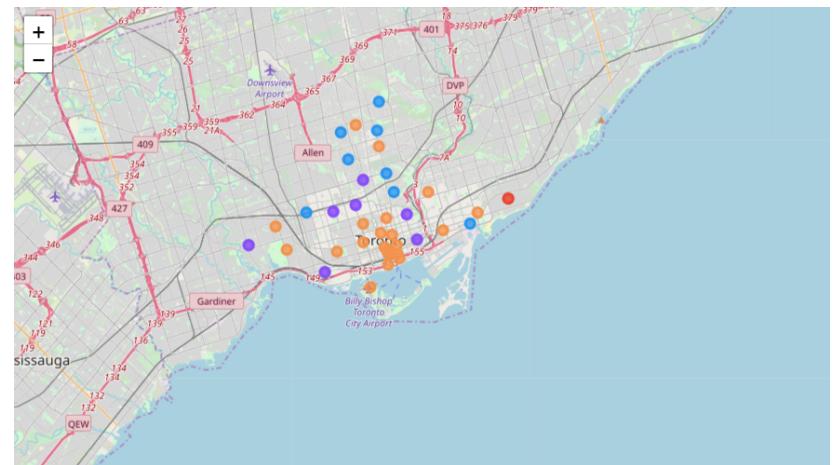
Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
Batignolles-Monceau	French Restaurant	Bar	Wine Bar	Hotel	Restaurant	Italian Restaurant	Thai Restaurant	Vietnamese Restaurant	Bakery	BBQ Joint
Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Wine Shop	Gym	Italian Restaurant	Clothing Store	Ice Cream Shop	Sushi Restaurant
Bois-de-Boulogne	Park	Forest	French Restaurant	Garden	Yoga Studio	Falafel Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant
Bois-de-Vincenne	Music Venue	Forest	Farmers Market	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service
Bourse	French Restaurant	Bistro	Cocktail Bar	Wine Bar	Hotel	Italian Restaurant	Café	Pizza Place	Boutique	Burger Joint

In this project, I would like to find similarities between sub-districts, that is to cluster the most similar points. K-means is popular for cluster analysis. It aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Here I run k-means to cluster sub-districts into 6 clusters.

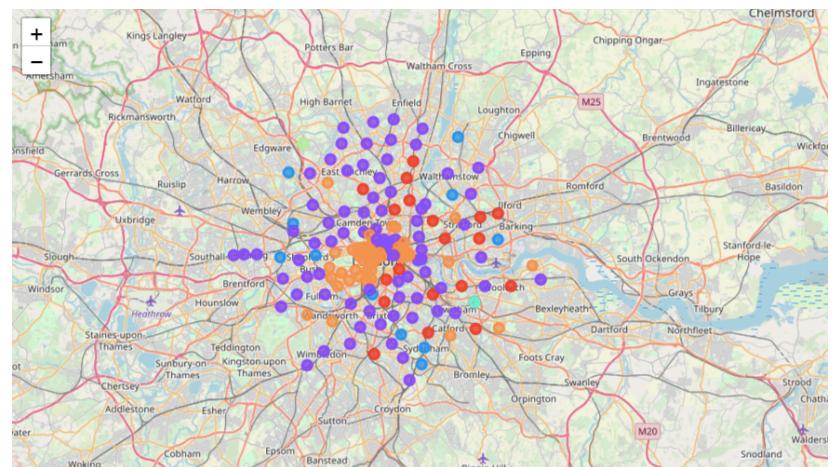
Manhattan



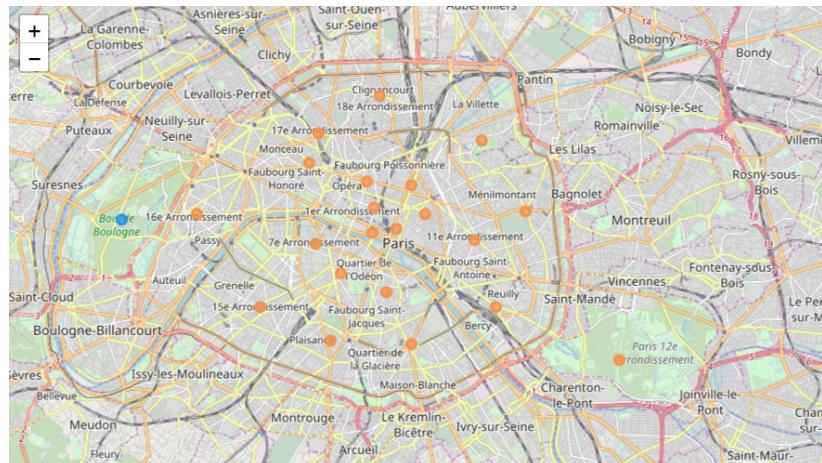
Old Toronto



Inner London



Paris



Cluster 1 --- red points

These points surround with lots of pubs/bars, cafes/coffee shops and restaurants, might be very cozy places to visit. Most in London, one in Toronto. Maybe living in Toronto postcode M4E area would be somewhat like living in those London parts.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
Toronto	East Toronto	M4E	Trail	Pub	Health Food Store	Other Great Outdoors	Neighborhood	Duty-free Shop	Eastern European Restaurant	Electronics Store	Fabric Shop	Empanada Restaurant
London	E	E12	Gym / Fitness Center	Restaurant	Turkish Restaurant	Gas Station	Pub	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Fabric Shop	Empanada Restaurant
London	E	E13	Café	Pub	Bus Station	Gym	Flea Market	Fishing Store	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store
London	E	E3	Pub	Bus Stop	Grocery Store	Hotel	Locksmith	Café	Light Rail Station	Bar	Burger Joint	Convenience Store
London	E	E7	Grocery Store	Pub	Café	Fast Food Restaurant	Bus Stop	Train Station	Comfort Food Restaurant	Chinese Restaurant	Fish & Chips Shop	Moving Target
London	E	E9	Pub	Coffee Shop	Café	Yoga Studio	Train Station	Grocery Store	Food & Drink Shop	Park	Pizza Place	Bus Stop
London	N	N15	Arepas Restaurant	Hardware Store	South American Restaurant	Bus Stop	Fried Chicken Joint	Pub	Café	Electronics Store	Empanada Restaurant	English Restaurant
London	N	N16	Pub	Café	Restaurant	Pizza Place	Italian Restaurant	Gift Shop	Cocktail Bar	Bakery	Coffee Shop	Grocery Store

Cluster 2 --- purple points

One in New York Manhattan, several in Old Toronto, and lots in London. Very close to center downtown. Surrounded with lots of pubs and cafes, having numbers of restaurants, most of which are fast food and foreign cuisines.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
New York City	Manhattan	Morningside Heights	Bookstore	Coffee Shop	Park	American Restaurant	Deli / Bodega	Food Truck	Sandwich Place	Burger Joint	Tennis Court	Frozen Yogurt Shop
Toronto	Central Toronto	M4V	Pub	Coffee Shop	Vietnamese Restaurant	Sports Bar	Fried Chicken Joint	Supermarket	Sushi Restaurant	Liquor Store	Bagel Shop	Restaurant
		Downtown										
City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
New York City	Manhattan	Marble Hill	Sandwich Place	Coffee Shop	Discount Store	Yoga Studio	Supplement Shop	Steakhouse	Spa	Shopping Mall	Seafood Restaurant	Clothing Store
New York City	Manhattan	Chinatown	Chinese Restaurant	Cocktail Bar	Salon / Barbershop	Vietnamese Restaurant	American Restaurant	Bakery	Dumpling Restaurant	Ice Cream Shop	Bubble Tea Shop	Spa
New York City	Manhattan	Washington Heights	Café	Bakery	Mobile Phone Shop	Grocery Store	Chinese Restaurant	New American Restaurant	Coffee Shop	Latin American Restaurant	Deli / Bodega	Italian Restaurant
New York City	Manhattan	Inwood	Café	Mexican Restaurant	Lounge	Pizza Place	Pharmacy	Park	Spanish Restaurant	Wine Bar	Chinese Restaurant	Frozen Yogurt Shop
New York City	Manhattan	Hamilton Heights	Pizza Place	Mexican Restaurant	Café	Coffee Shop	Yoga Studio	Chinese Restaurant	Liquor Store	Sushi Restaurant	Bakery	School
New York City	Manhattan	Manhattanville	Mexican Restaurant	Park	Seafood Restaurant	Italian Restaurant	Coffee Shop	Café	BBQ Joint	Bank	Bar	Grocery Store
New York City	Manhattan	Central Harlem	African Restaurant	Chinese Restaurant	Seafood Restaurant	Cosmetics Shop	Art Gallery	Public Art	French Restaurant	American Restaurant	Bar	Gym

Cluster 3 --- blue points

These are park areas. Out of parks, you can visit a lot of nice restaurants in styles of Japanese, Eastern European, English, or Ethiopian.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
New York City	Manhattan	Stuyvesant Town	Park	Bar	Playground	Cocktail Bar	Heliport	Basketball Court	Gas Station	Harbor / Marina	Pet Service	Boat or Ferry
Toronto	Central Toronto	M4N	Park	Bus Line	Swim School	Photography Studio	Yoga Studio	Falafel Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant
Toronto	Central Toronto	M4P	Gym	Gym / Fitness Center	Asian Restaurant	Clothing Store	Sandwich Place	Hotel	Park	Grocery Store	Breakfast Spot	Food & Drink Shop
Toronto	Central Toronto	M4T	Trail	Summer Camp	Playground	Yoga Studio	Falafel Restaurant	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant
Toronto	Downtown Toronto	M4W	Park	Trail	Building	Playground	Yoga Studio	Falafel Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant
Toronto	Central Toronto	M5N	Music Venue	Garden	Home Service	Farm	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant	English Restaurant
Toronto	Central Toronto	M5P	Park	Trail	Jewelry Store	Sushi Restaurant	Yoga Studio	Fabric Shop	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store

Cluster 4 --- cyan blue points

This particular point only exist in London. Top 3 categories are furniture/home store, yoga studio and farm. And also we can find multi cultural flavor restaurants.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
London	SE	SE3	Furniture / Home Store	Yoga Studio	Farm	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Ethiopian Restaurant

Cluster 5 --- green points

This point also only exist in London. Similarly to cluster 4, we can find multi cultural flavor restaurants. The difference might because of the top 2 categories, park and pharmacy.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
London	NW	NW7	Park	Pharmacy	Yoga Studio	Fabric Shop	Dumpling Restaurant	Duty-free Shop	Eastern European Restaurant	Electronics Store	Empanada Restaurant	English Restaurant

Cluster 6 --- orange points

Apparently these are city center parts. Almost all parts of Manhattan and Paris are belong to this genre.

City	District	Sub-district	1st popular category	2nd popular category	3rd popular category	4th popular category	5th popular category	6th popular category	7th popular category	8th popular category	9th popular category	10th popular category
New York City	Manhattan	Marble Hill	Sandwich Place	Coffee Shop	Discount Store	Yoga Studio	Supplement Shop	Steakhouse	Spa	Shopping Mall	Seafood Restaurant	Clothing Store
New York City	Manhattan	Chinatown	Chinese Restaurant	Cocktail Bar	Salon / Barbershop	Vietnamese Restaurant	American Restaurant	Bakery	Dumpling Restaurant	Ice Cream Shop	Bubble Tea Shop	Spa
New York City	Manhattan	Washington Heights	Café	Bakery	Mobile Phone Shop	Grocery Store	Chinese Restaurant	New American Restaurant	Coffee Shop	Latin American Restaurant	Deli / Bodega	Italian Restaurant
New York City	Manhattan	Inwood	Café	Mexican Restaurant	Lounge	Pizza Place	Pharmacy	Park	Spanish Restaurant	Wine Bar	Chinese Restaurant	Frozen Yogurt Shop
New York City	Manhattan	Hamilton Heights	Pizza Place	Mexican Restaurant	Café	Coffee Shop	Yoga Studio	Chinese Restaurant	Liquor Store	Sushi Restaurant	Bakery	School
New York City	Manhattan	Manhattanville	Mexican Restaurant	Park	Seafood Restaurant	Italian Restaurant	Coffee Shop	Café	BBQ Joint	Bank	Bar	Grocery Store
New York City	Manhattan	Central Harlem	African Restaurant	Chinese Restaurant	Seafood Restaurant	Cosmetics Shop	Art Gallery	Public Art	French Restaurant	American Restaurant	Bar	Gym

Conclusion

Manhattan and Paris are highly centered, with obviously city characteristics. They are very "city". But we all know that this two place look quite different. Furthermore research has to be done to distinguish them. I have never been to London, but now I know that London is much bigger than Paris, much centered than New York, and may be a little bit like Toronto since this two cities share more clusters and used to divide by post code or postal code. London also has some very special places that the other cities don't have, where you can enjoy more east-European and north-African foods. London is very diversity. Toronto might be more cozy to live, where I can find cafe centered areas or park area.

To improve the research, I can do below:

- Paris can be divided into more small districts to compare with other city districts on a same level. This might take some work even manual work. Smaller districts help to outstanding characteristics.
- Toronto should do the same as Paris. Dividing by postal code is not enough.
- Expand to the whole city and their outer areas. Big cities are always influence deeply to their conjunction parts, like great London and great Paris.