The Battle of the Neighborhoods

Why Comparing Cities using venues info?

- Moving from one city to another is not an easy decision to make. In the world, there is a large number of cities and all of them have something that makes them unique and different than the rest. Several factors need to be consider and one of them could be how similar is the other city and their people, compared to the place we live.

- It can be said that if two cities share same types of most common venues to go then they are similar and therefore their people have similar preferences for certain kind of places and this implies that they have the same habits. Therefore, to have this information in hand will be helpful for a person in this situation because it will be a big contributing factor at the time to make a such important decision.

- For this study case, a family residing in Paris has to decide whether to move to New York or Toronto because the family head has received two very similar job offers from these cities and they will choose where to move based in how similar or dissimilar they are compared to Paris.

Data Acquisition and Cleaning

Data sources for Toronto, New York & Paris neighborhoods and geolocations from:

- URL: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- URL: https://cocl.us/Geospatial_data
- URL: https://cocl.us/new_york_dataset
- URL: https://opendata.paris.fr/explore/dataset/arrondissements/download/?format=json&timezone=Asia/Dubai
- Foursquare API credentials.

Data Cleaning:
Acquired data was in different formats, therefore different workflows were used on each case to obtain dataframes (for each city) to start working with. These, mainly contained the columns 'Neighborhood', 'Latitude' and 'Longitude'

Exploratory Data Analysis (Toronto example)

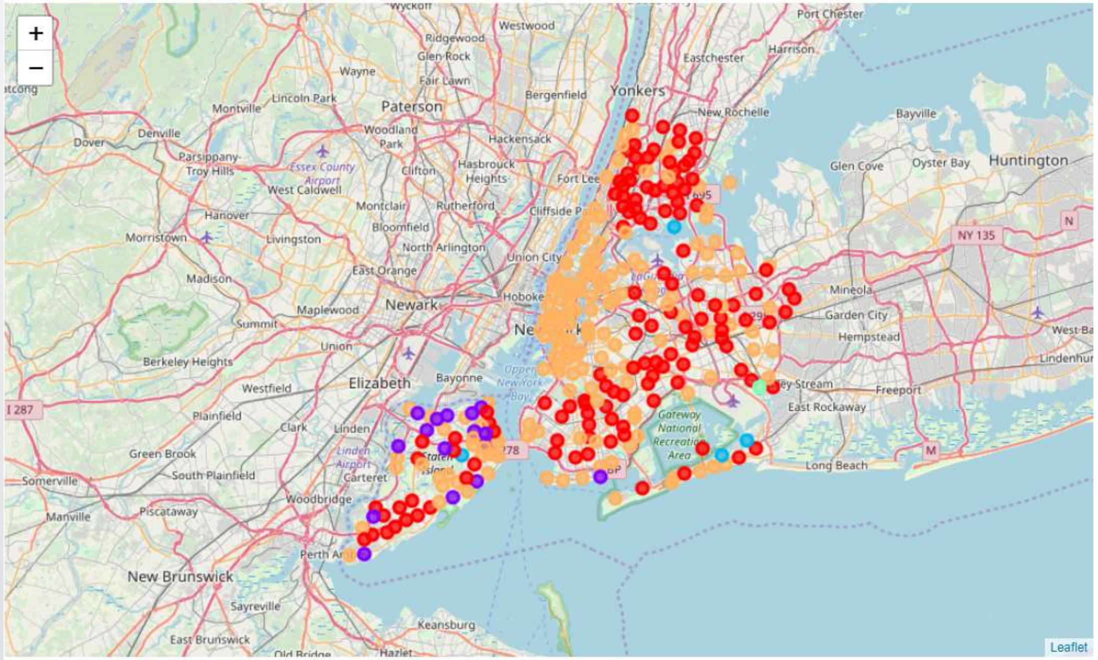
From this:
(After Foursquare)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	KFC	43.754387	-79.333021	Fast Food Restaurant
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

To this:
(Top ten most common venues for each 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	Adelaide, King, Richmond	Coffee Shop	Café	Thai Restaurant	Steakhouse	Bar	Gym	Breakfast Spot	Asian Restaurant	American Restaurant	Restaurant
1	Agincourt	Lounge	Clothing Store	Breakfast Spot	Skating Rink	Drugstore	Discount Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant	Drugstore
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Pizza Place	Fast Food Restaurant	Beer Store	Sandwich Place	Fried Chicken Joint	Coffee Shop	Pharmacy	Comfort Food Restaurant	Dim Sum Restaurant
4	Alderwood, Long Branch	Pizza Place	Coffee Shop	Skating Rink	Dance Studio	Pharmacy	Pub	Sandwich Place	Gym	Airport Service	Dessert Shop

Neighborhoods Clusterization (New York example)



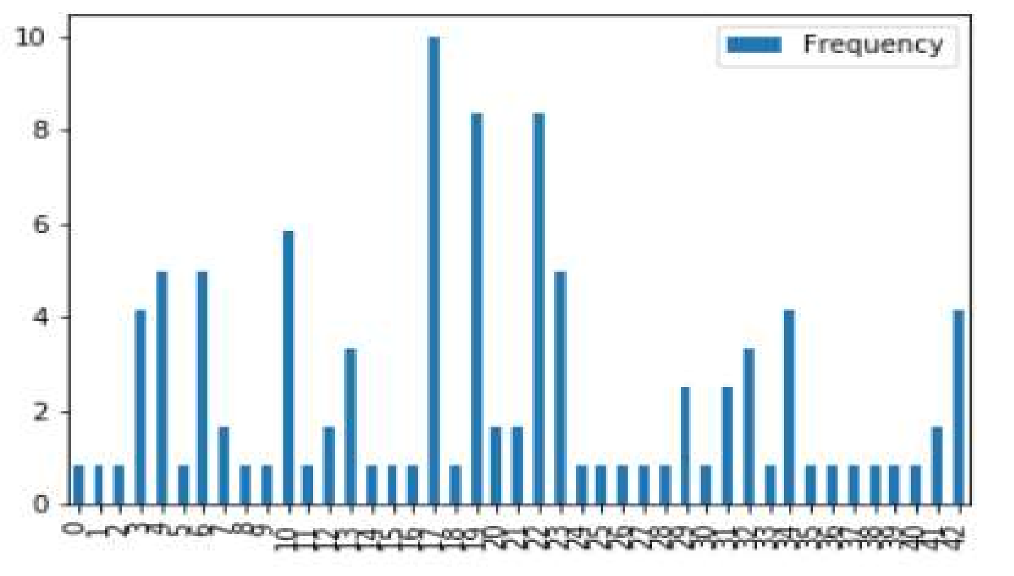
- ‘Most common venue’ feature was used to sgment Neighborhoods into 5 clusters using ‘k-means’ algorithm.
- Geopy library was used to get reference map latitude and longitude. Afterwards, the emerging clusters were mapped on it by using ‘Follium’ library to visualize them.

Representative Clusters Examination, Paris example

- The clusters with more neighborhoods were identified as the most representatives.
- Paris Cluster #3, with 12 neighborhoods and 120 top-ten ‘Most common venues’ (‘MCV’) observations.
- The neighborhood’s most common venues were grouped and their occurrence were counted and the relative frequency was computed, resulting in a new dataframe (43 rows).
- Bar charts was produced.

	Most common venue	Frequency
0	Art Gallery	1
1	Art Museum	1
2	Asian Restaurant	1
3	Bakery	5
4	Bar	6

	Most common venue	Frequency
0	Art Gallery	0.833333
1	Art Museum	0.833333
2	Asian Restaurant	0.833333
3	Bakery	4.166667
4	Bar	5.000000

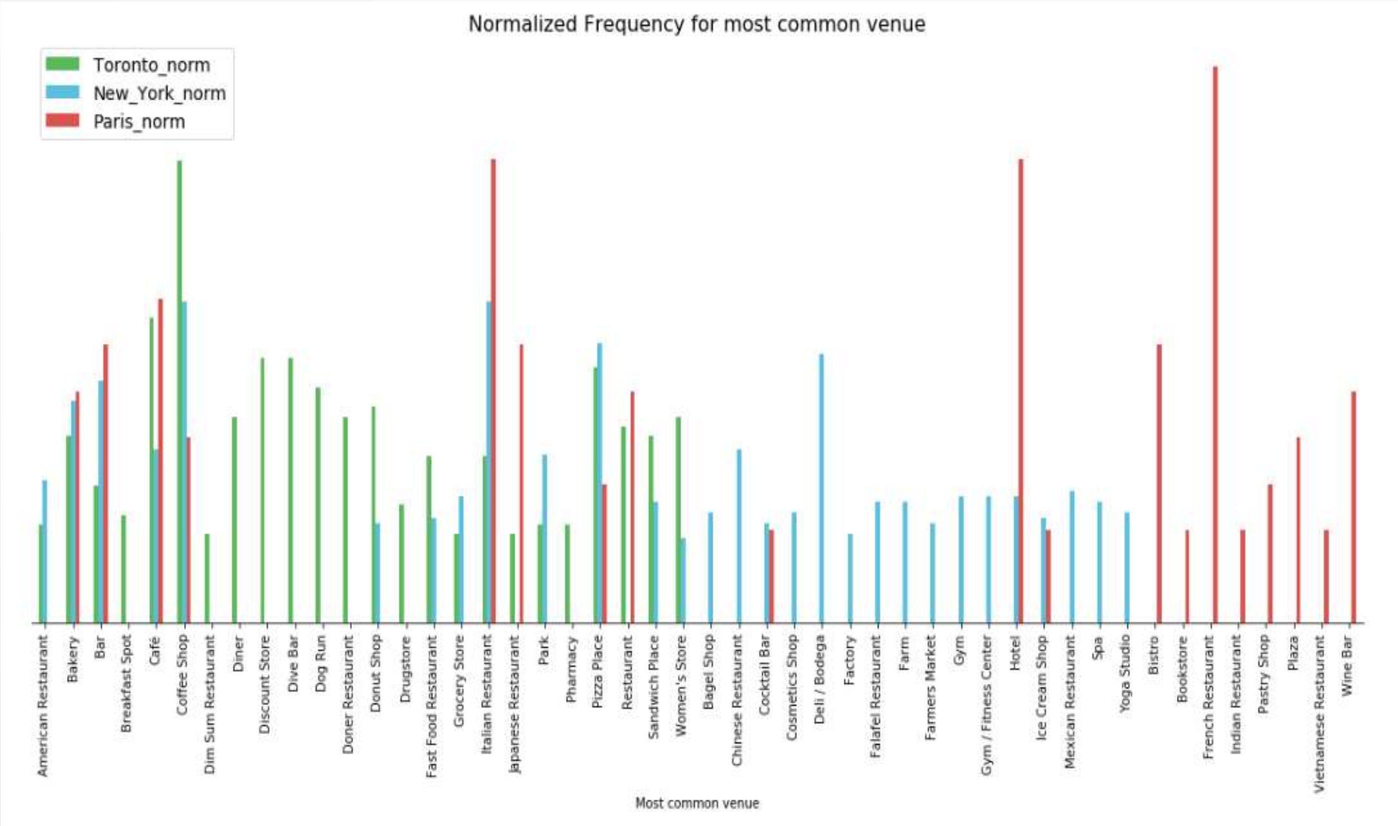


Comparative Model

- Minimum frequency cut-off value of 1% was applied resulting in a reduction of the rows data.
- A 'city' column was added for identification and then dataframes were merged
- Further transformations included addition of columns, frequency and normalized frequency allocation and grouping by 'Most common venues' to develop a final dataframe with 48 categories to compare the 03 cities subjected to study.

Most common venue	Frequency	Toronto	New_York	Paris	Toronto_norm	New_York_norm	Paris_norm
American Restaurant	2.917625	1.219512	1.698113	0.000000	2.207506	3.195266	0.000000
Bakery	9.125249	2.317073	2.641509	4.166667	4.194260	4.970414	5.208333
Bar	9.600399	1.707317	2.893082	5.000000	3.090508	5.443787	6.250000
Breakfast Spot	1.341463	1.341463	0.000000	0.000000	2.428256	0.000000	0.000000
Café	11.689293	3.780488	2.075472	5.833333	6.843267	3.905325	7.291667
Coffee Shop	12.901519	5.731707	3.836478	3.333333	10.375276	7.218935	4.166667
Dim Sum Restaurant	1.097561	1.097561	0.000000	0.000000	1.986755	0.000000	0.000000
Diner	2.560976	2.560976	0.000000	0.000000	4.635762	0.000000	0.000000
Discount Store	3.292683	3.292683	0.000000	0.000000	5.960265	0.000000	0.000000
Dive Bar	3.292683	3.292683	0.000000	0.000000	5.960265	0.000000	0.000000
Dog Run	2.926829	2.926829	0.000000	0.000000	5.298013	0.000000	0.000000
Doner Restaurant	2.560976	2.560976	0.000000	0.000000	4.635762	0.000000	0.000000
Donut Shop	3.877895	2.682927	1.194969	0.000000	4.856512	2.248521	0.000000
Drugstore	1.463415	1.463415	0.000000	0.000000	2.649007	0.000000	0.000000
Fast Food Restaurant	3.331032	2.073171	1.257862	0.000000	3.752759	2.366864	0.000000
Grocery Store	2.606995	1.097561	1.509434	0.000000	1.986755	2.840237	0.000000
Italian Restaurant	14.242982	2.073171	3.836478	8.333333	3.752759	7.218935	10.416667
Japanese Restaurant	6.097561	1.097561	0.000000	5.000000	1.986755	0.000000	6.250000
Park	3.232091	1.219512	2.012579	0.000000	2.207506	3.786982	0.000000

Comparison: Toronto & New York vs. Paris



- Paris has 19 'MCV' categories.
- 06/19 'MCV' are shared simultaneously among Toronto, New York and Paris.
- Paris shares 03 'MCV' exclusively with New York and 02 'MCV' exclusively with Toronto.

Conclusions

- It was possible to compare 03 cities based on the ‘Most common venues’ (MCV) existing on their neighborhoods.

- The results are very interesting and it is clear that even though there are some similarities among these three multi-cultural cities, there are also differences among them.

- There are 08 categories of ‘MCV’ (Bistro, Bookstore, French Restaurant, Indian Restaurant, Pastry Shop, Vietnamize Restaurant and Wine Bar) which represents 42%, and makes Paris distinctive from New York and Toronto. However, there are also similarities because Paris shares 09 (47%) common venues with New York and 08 (42%) with Toronto.

Way Forward

- Moving from one city to another always represents a challenge. In this case, the analysis shows that people from New York shares more similarities with Paris than Toronto in terms of 'MCV' but the degree similarity is close to Toronto one (47% vs. 42%).

- More study including other variables need to be taken account. For instance, French is a language that it is spoken widely in Toronto but not in New York. Similarly, parameters like weather, living cost, etc. need to be considered.

- With regards to the workflow used to compare the cities, this can be improved, specially in the last stage where only visual analysis of the comparative bar chart was used to determine the degree of similarity between the cities. This will be very helpful when analyzing simultaneously more number of cities and larger amount of 'Most common venue' categories.