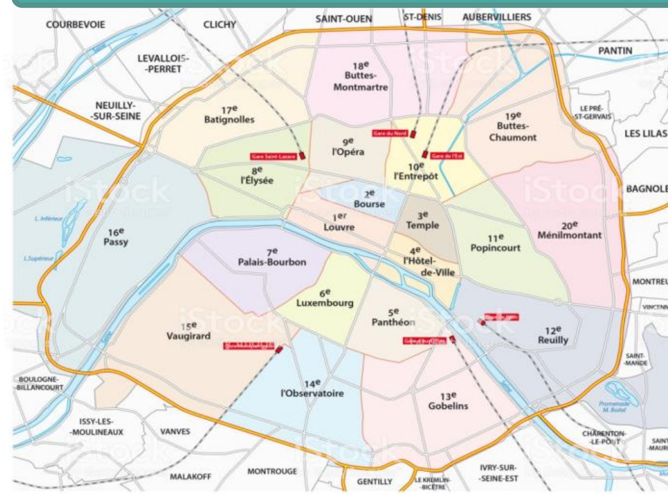
Can we define which neighborhoods are more likely to move on without much changes in the closest venues? If so, which are the neighborhoods more likely to have the same venues

Toronto

Manhattan

Paris



- The mail goal will be exploring the neighborhoods of New York City, Toronto city and Paris in order to define with neighborhoods are related bearing in mind the most common venues close to each neighborhood.
- The idea comes from a family whose want to move having similar venue close to their home. This will work for a family that wants to move from New York, Toronto or Paris, to any of the neighborhoods of these cities. The target audience for this report are:
- Target Public
    - Potential families in New York, Toronto, or Paris that want to move to a similar neighborhood in those cities.
    - Entrepreneurs who wants to open a new venue consider the lack or excess of similar venues in each neighborhood.
    - Learners who would be interested in Clustering and a location foursquare application

Data Description - To consider the problem we can list the data as below:

- **The Toronto Neighborhood Coordinates**

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
https://cocl.us/Geospatial_data

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

- **The New York Neighborhood Coordinates**

https://cocl.us/new_york_dataset

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

- **The Paris Neighborhood Coordinates**

https://opendata.paris.fr/explore/dataset/arrondissements/download

	Neighborhood	Latitude	Longitude
0	Bourse	48.8682792225	2.34280254689
1	Temple	48.86287238	2.3600009859
2	Reuilly	48.8349743815	2.42132490078
3	Louvre	48.8625627018	2.33644336205
4	Hôtel-de-Ville	48.8543414263	2.35762962032

- **The venues of each neighborhood**

Foursquare API, limit of 100 venues and a radius of 500 m of the center of each neighborhood

The Methodology is divided in five phases

Get the list of the location of each neighborhood within Toronto, Manhattan and Paris, the coordinates of the center of each city and a list of venues close to each neighborhood with the location and the Venue Category

Create a DataFrame in which defines the recurrence of each venue category on each neighborhood with help of get_dummies for pandas libraries, and then calculate per neighborhood how many venues there are.

Download Data → **Clean data** → **Model Data** → **K Means** → **Map and Summarize**

Understand the data and check for any inconsistency or improve opportunities. In this case, at evaluate the venues category, we figured it out that Coffee Shops can be obtained as “Coffee Shops” and “Café”, and Gym can be obtained also as “Gym / Fitness Center”.

Define which are the main segments of neighborhood due to the recurrence of venues. With trial and error, I defined the number of clusters of 6. After this, the segments are mapped for each city to take a view of

Result (1/2)

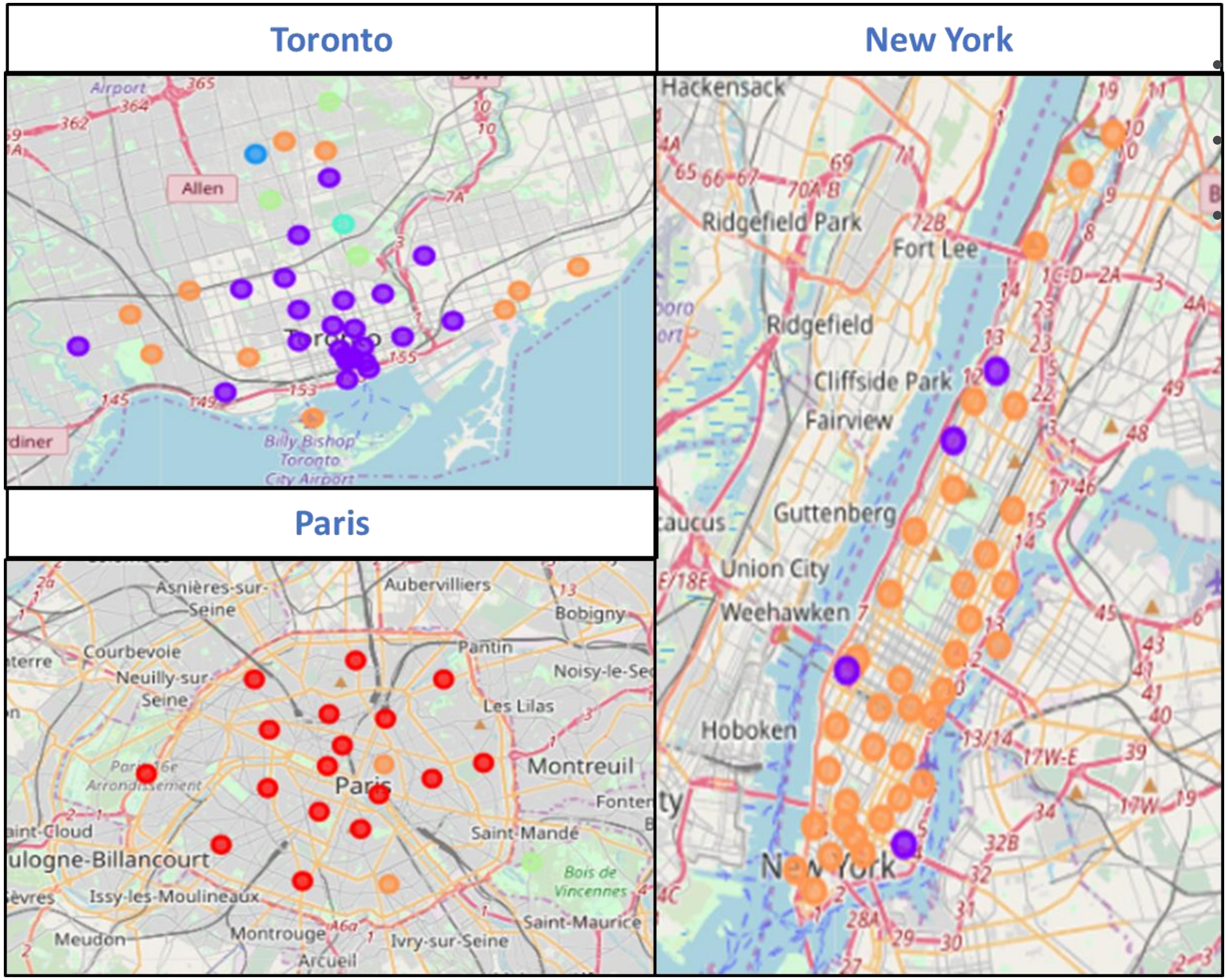
- The results from the k-means clustering shows that we can categorized neighborhoods into 6 cluster based on the frequency of occurrence for each venue category:

Cluster	N. of neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	17	French Restaurant	Hotel	Coffee Shop	Italian Restaurant	Plaza
1	31	Coffee Shop	Italian Restaurant	Park	Pizza Place	Restaurant
2	1	Garden	Home Service	Zoo	Farmers Market	Duty-free Shop
3	4	Park	Trail	Bus Line	Swim School	Jewelry Store
4	44	Coffee Shop	Gym	Italian Restaurant	Park	Bakery
5	1	Summer Camp	Playground	Zoo	Farmers Market	Duty-free Shop

- As coffee shop and Italian Restaurant are the most common place the conclusion would not consider much of these categories. Also, we can see that the cluster 0, 1 and 4 are the main cluster, the 2, 3 and 5 are cluster with neighborhoods that are almost unique.
- The first Cluster are neighborhoods, above Coffee Shops and Italian Restaurant, with French Restaurants and Hotels, the second have Parks, Pizza places, and the fourth with Gyms, parks and Bakeries.

Results (2/2)

- The following maps shows the clustering for each city. The red points are from cluster 0, purple for cluster 1 orange for cluster 4



As graph shows, there is a strong relationship between the city and the cluster, Toronto for cluster 1, Manhattan for cluster 4 and Paris for cluster 0. Even thou, there are some neighborhoods that can be compared.

Discussion and Conclution

Discussion

- As observations shows from the maps in the Result section, there are a strong pattern for each country, which drive the clustering algorithm. Also, the frequency of Coffee Shops, Italian Restaurant and French Restaurant are the most commons venues which is also a driven for the clustering.
- There are three cluster with a low number of neighborhoods within. This can be explained because in those places none of the most frequent category are this neighborhood.

Conclusion

- If anyone from any country would like to move for another neighborhood with a similitud in the most common places I would recommend to look forward in the same country. However, in some cases there are similar neighborhoods in another city to move.