Comparing neighborhoods between New York City and Toronto

Airton Carlos Nunes Raimundo

March 21, 2020

What is the best way to choose a neighborhood that provides the services that I am used to consuming?

- In the context of globalization, migratory currents are increasingly strong and people are increasingly experiencing the cultural diversity that the world provides.
- However many people have habits of consumption that are part of their lives and they would like to maintain them even during a change of country.
- The results presented here aim to reach people in temporary or permanent change with little taste for changes in their essential services.

Data acquisition and cleaning

- Initially we will need a dataset containing all the neighborhoods in both cities, these datasets can be downloaded here and here. The second part will be collecting the locations of New York and Toronto using the geopy library to help us visualize our data. To top it off, let's get information from nearby vendors using the Foursquare API.

- The data comes from different sources, so it was necessary to adjust column names and eliminate unnecessary data.

Category selection

- For each neighborhood of the two cities we search for all commercial establishments within a radius of 500 meters. After that we group the categories into more generalist groups and select the 22 most representative categories.

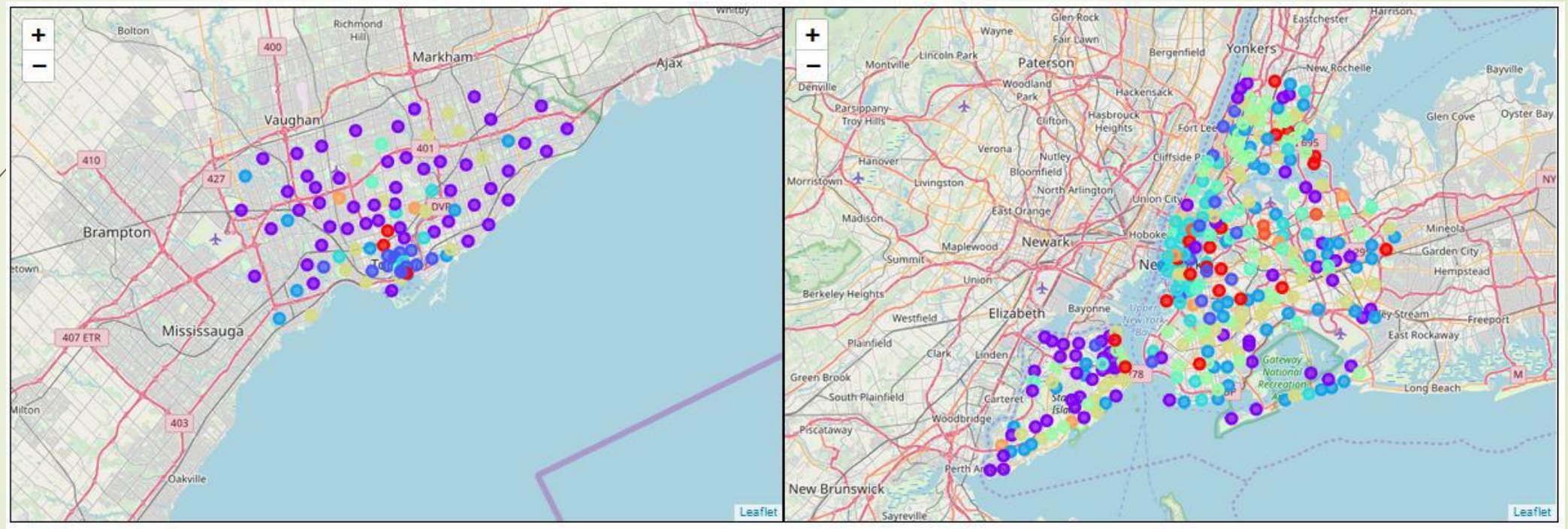
Restaurant, Coffee Shop, Bar, Fast Food, Park, Grocery Store, Bakery, General Store, Gym, Clothing Store, Supermarket, Pharmacy, Bank, Spa, Theater, Shopping Mall, Bus Line, Hotel, Tea Room, Pet Store, Art Gallery, Bookstore.

Performing K-means algorithm

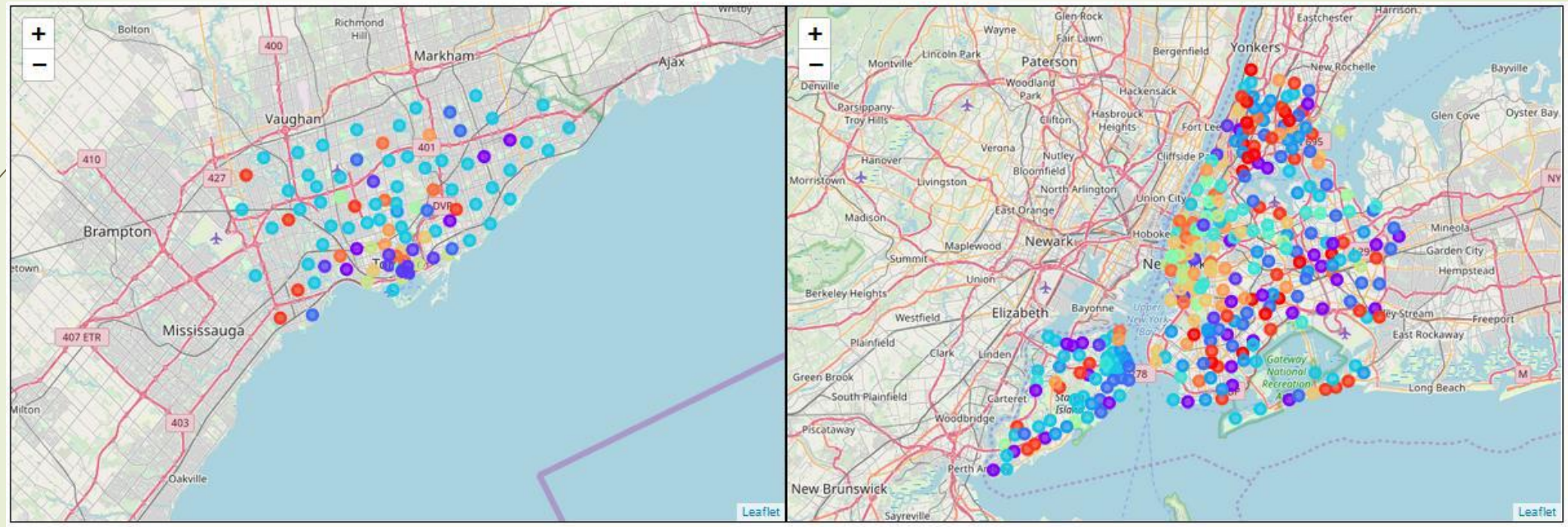
- To perform k-means clustering we first count the number of occurrences of each category in each neighborhood and then apply the algorithm using K = [10, 15, 20].

- The choice of these K's was based on the number of neighborhoods in Toronto that is just over 100.

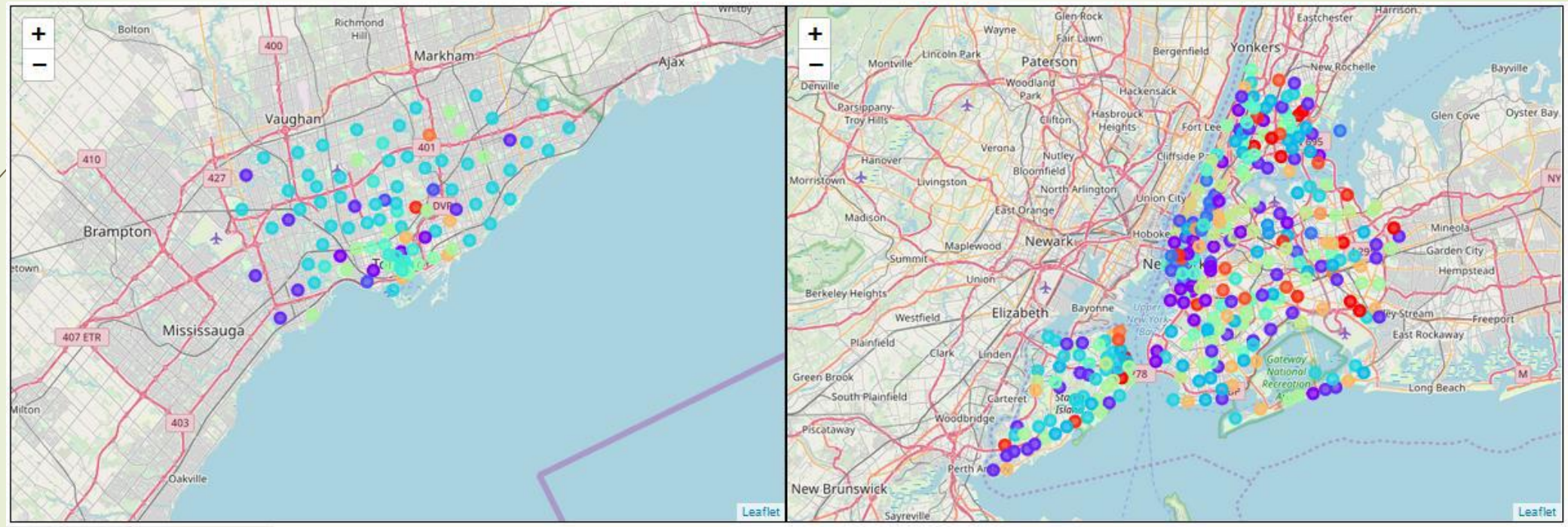
Toronto and New York City
K = 10



Toronto and New York City
K = 15



Toronto and New York City
K = 20



Results and discussion

- The results show us that the city of Toronto is very homogeneous in terms of the main services offered by the city and depends very little on the chosen K, on the other hand, New York City has much more variety.

- Our graphical presentation of the results allows us to directly conclude which are the most similar neighborhoods between the two cities.

Conclusion

- This work satisfactorily demonstrated a procedure to compare the neighborhoods of two cities using:
  - Public databases
  - Machine learning
  - Data visualization tools
- With the practical objective of choosing the most similar neighborhood.