

# NEW YORK CITY AND TORONTO CITY NEIGHBORHOOD SIMILARITY

A decorative graphic consisting of several parallel white lines of varying lengths, slanted diagonally from the bottom left towards the top right, set against a blue-to-white gradient background.

# PREDICT SIMILARITY OF NEIGHBORHOODS BELONGING TO TORONTO AND NEW YORK CITY BASED ON THE MOST COMMON AND USED SERVICES AND VENUES THAT EACH NEIGHBORHOOD OFFER

- ▶ Predicting neighborhood similarity is useful for giving decision-making support to those who have to move from one city to another
  - ▶ Avoid incurring additional costs due to a poor choice of destination and having to relocate after a short time

# DATA ACQUISITION AND CLEANING

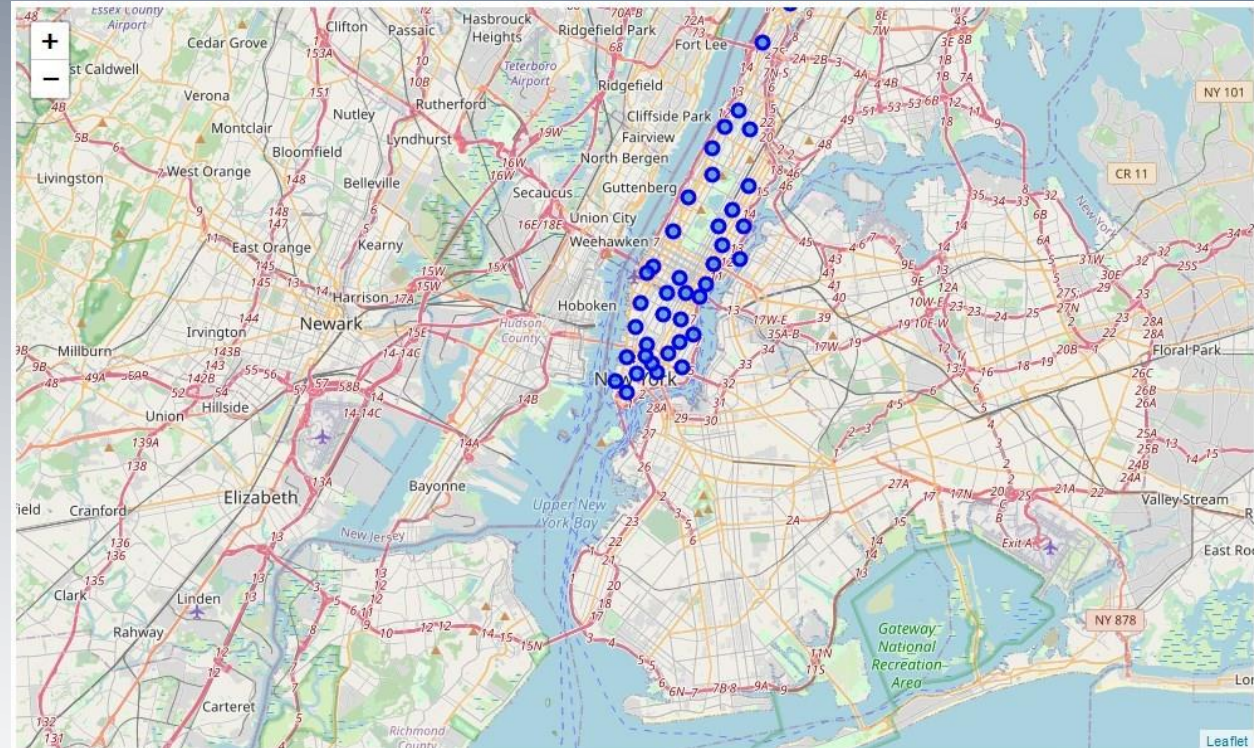
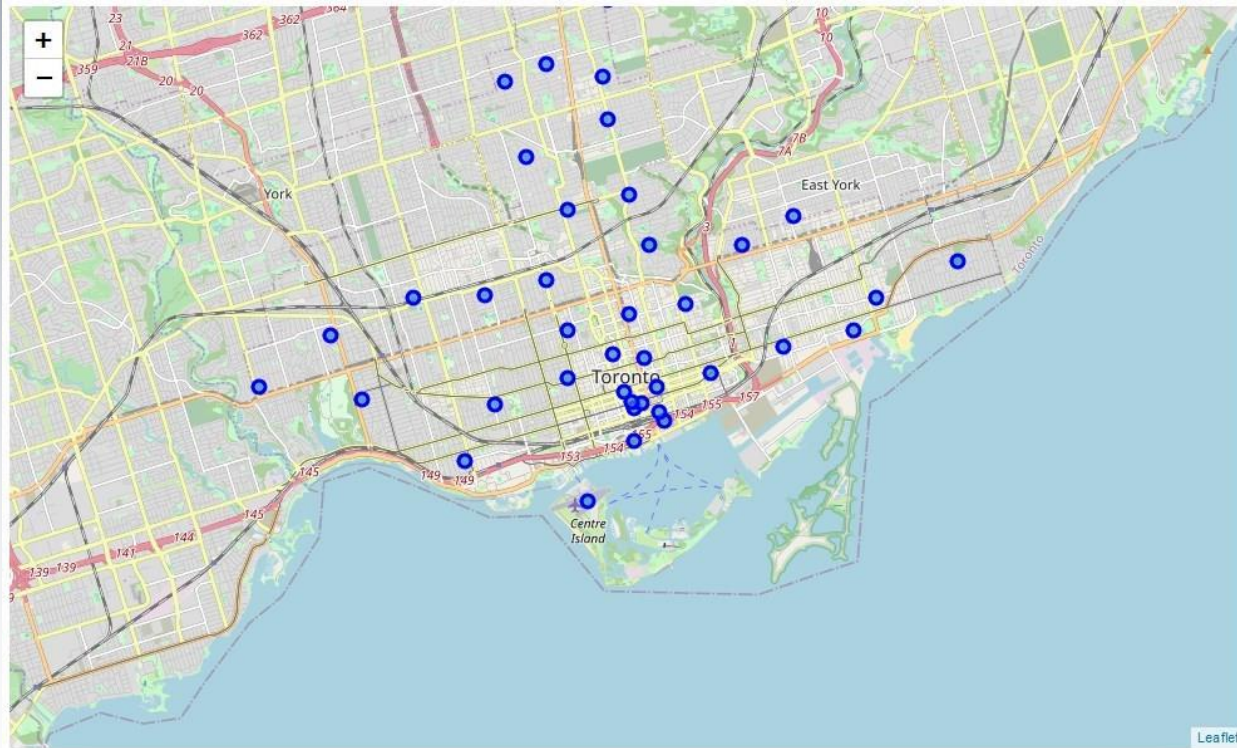
- ▶ Borough name, Neighborhood name, Postal code for Toronto city are scraped from **List of postal codes of Canada: M** and latitude and longitude for each location from **[https://cocl.us/geospatial\\_data](https://cocl.us/geospatial_data)**
- ▶ Borough name, Neighborhood name, latitude and longitude for each location for New York city are scraped from **<https://cf-courses-data>**
- ▶ **Foursquare API** to get the most popular venues for the aforementioned cities.
- ▶ filter NewYork Data taking Borough of Manhattan and for Toronto dataset filtered on Borough containing the string “Tor”

	Borough	Neighborhood	Latitude	Longitude	City	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Downtown Toronto	Regent Park, Harbourfront	43.65426	-79.360636	Toronto	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
2	Downtown Toronto	Regent Park, Harbourfront	43.65426	-79.360636	Toronto	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Downtown Toronto	Regent Park, Harbourfront	43.65426	-79.360636	Toronto	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
2	Downtown Toronto	Regent Park, Harbourfront	43.65426	-79.360636	Toronto	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
2	Downtown Toronto	Regent Park, Harbourfront	43.65426	-79.360636	Toronto	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa



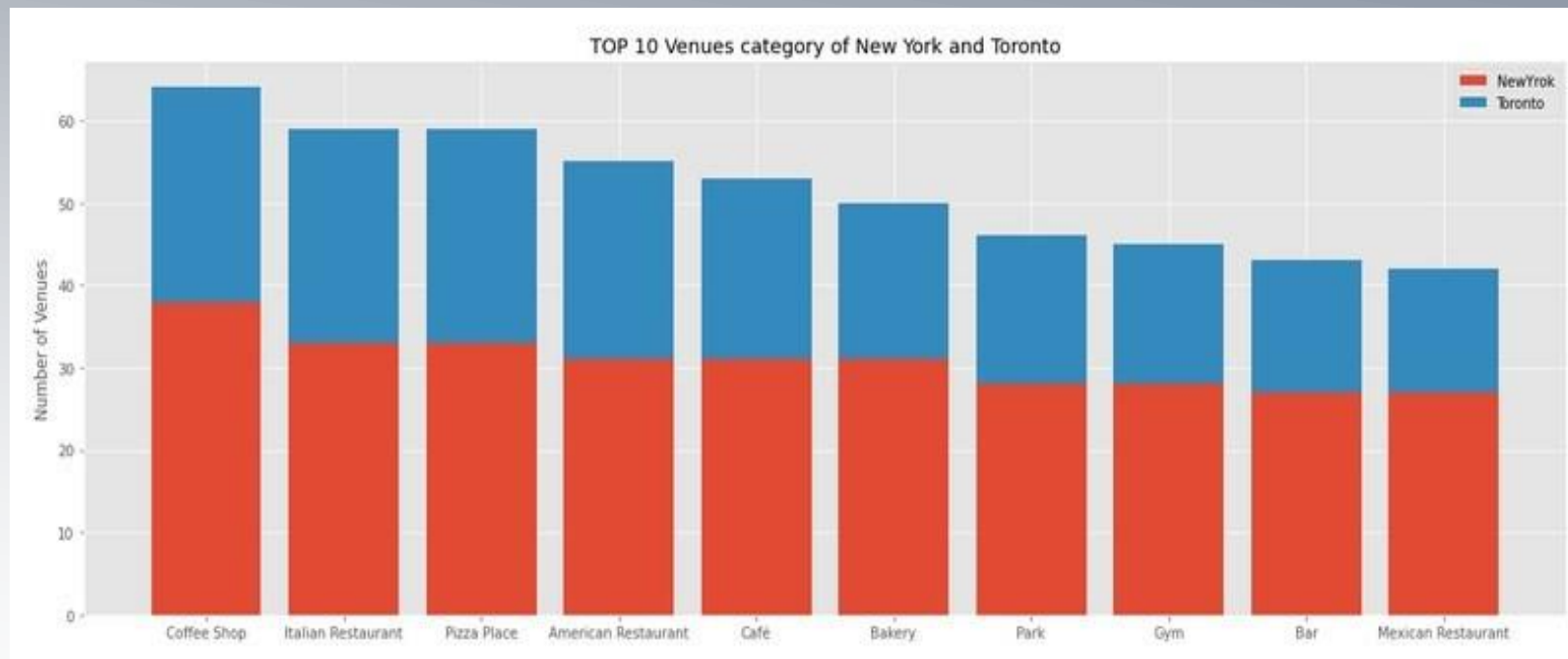
# NEIGHBORHOOD LOCATION USING PYTHON FOLIUM LIBRARY

- Python Folium Library used to display Neighborhood location



# TOP MOST COMMON VENUES IN THE CITY OF TORONTO AND NEW YORK CITY

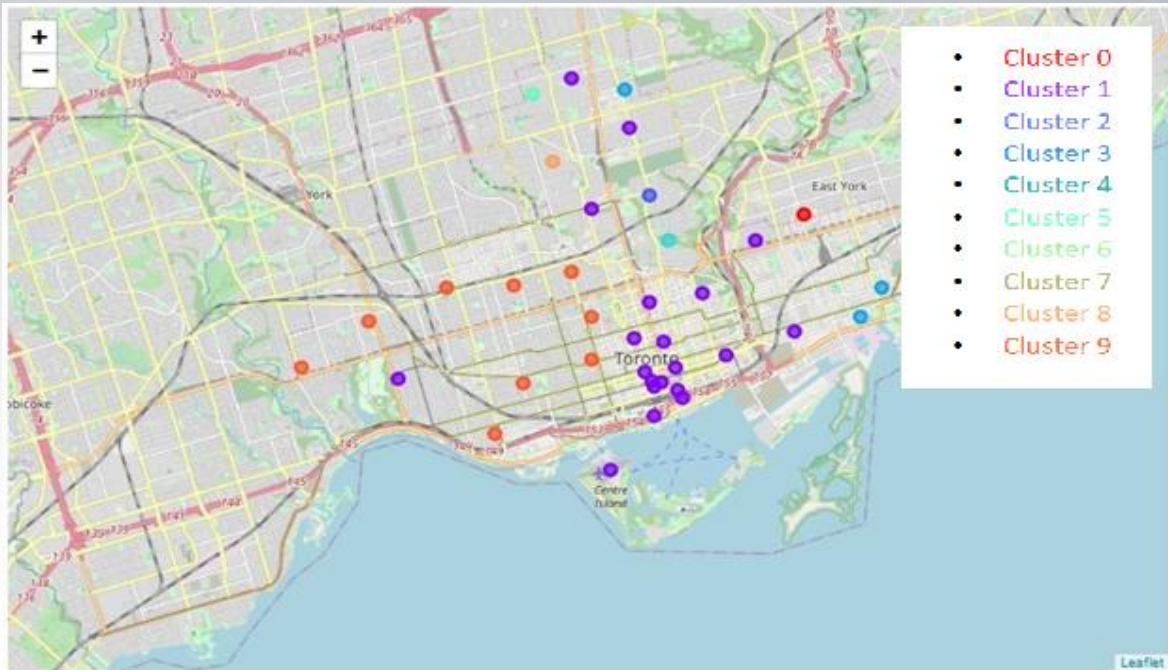
- ▶ Matplotlib Library used to display most common venues in the dataset





# K-MEANS ALGORITHM

- ▶ K-means is one of the simplest unsupervised learning algorithms that solve the clustering problem. Is a way to classify a given data set through a certain number of clusters fixed apriori.
- ▶  $k=10$  number of cluster is considered
  - Most populated cluster are: cluster 1, cluster 4 and cluster 9
  - Cluster 9 include more evenly dataset belonging to both cities



# CONCLUSION

- ▶ While dealing with unlabeled data and unsupervised technique it is was not possible to divide the dataset into train and test set in order to evaluate most accurate algorithm and therefore apply different technique.
- ▶ Next step would be to include an additional dataset to include people satisfaction of the suggested similarities among the different neighborhood to evaluate and afterword tune the model to increase its accuracy