

Battle-of-Neighborhoods

1-Introduction

This project will analyze Toronto, Canada and New York City, New York. Suppose I want to open a company in one these cities. This project will help me to decide which city to choose in order to local businesses. This project will explore similarities and differences between Toronto and New York City.

2-Data

For this project we need the following data :

- Toronto data that contains list Boroughs, Neighborhoods along with their latitude and longitude. Data source: https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto
- New York City data that contains neighborhoods. Data source: https://geo.nyu.edu/catalog/nyu_2451_34572
- New York City Latitude and Longitude. Data source: Python Geolibrar
- Restaurants in each neighborhood of Toronto and to get the most common venues . Data source: Foursquare API
- GeoSpace data which is showing the Second-level Administrative Divisions of the Toronto. I used it for create map. Data source: <https://open.toronto.ca/dataset/neighbourhoods/>

3-Methodology

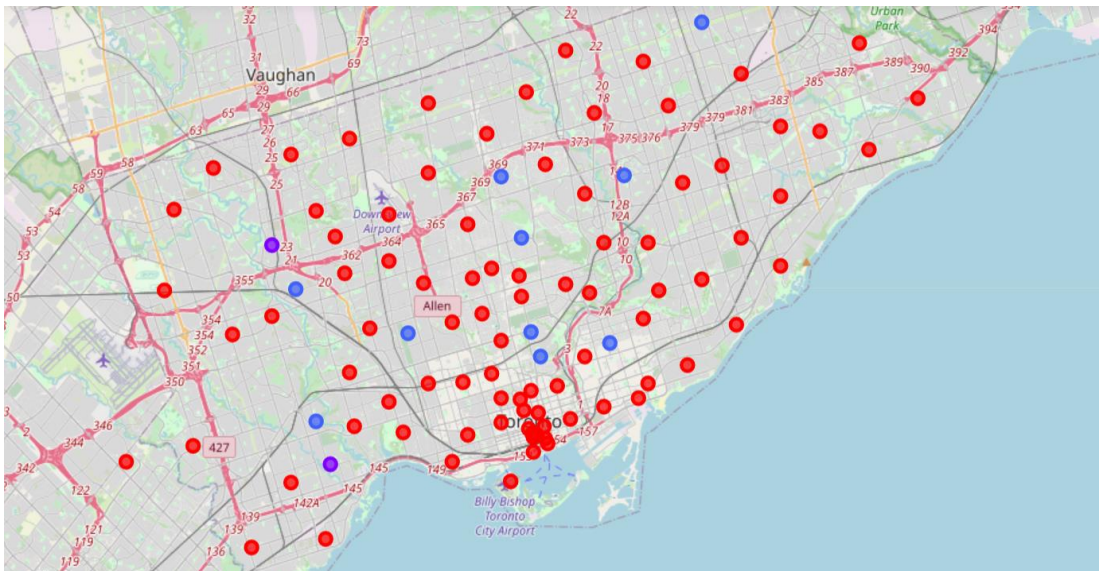
Work Flow:

1. HTTP requests made to this Foursquare API server using zip codes of Toronto and New York city neighborhoods to pull Latitude and Longitude information.

2. Due to http request limitations the number of places per neighborhood parameter be set to 100 and radius parameter set to 500.
3. For the purpose of map visualization, folium library used.
4. Venues which are from Foursquare API grouped by neighborhoods and list first 10 common categories.
5. Since the dimensions are too high Principal Component Analysis implemented in order to speed up the machine learning algorithm.
6. There is not any reason from elbow method silhouette score methos used to determine number of clusters for KMeans algorithm.
7. Unsupervised machine learning algorithm KMean clustering applied to from the clusters of different categories of places. These clusters from each of neighborhoods analyzed individually and comparatively to derive the conclusions.

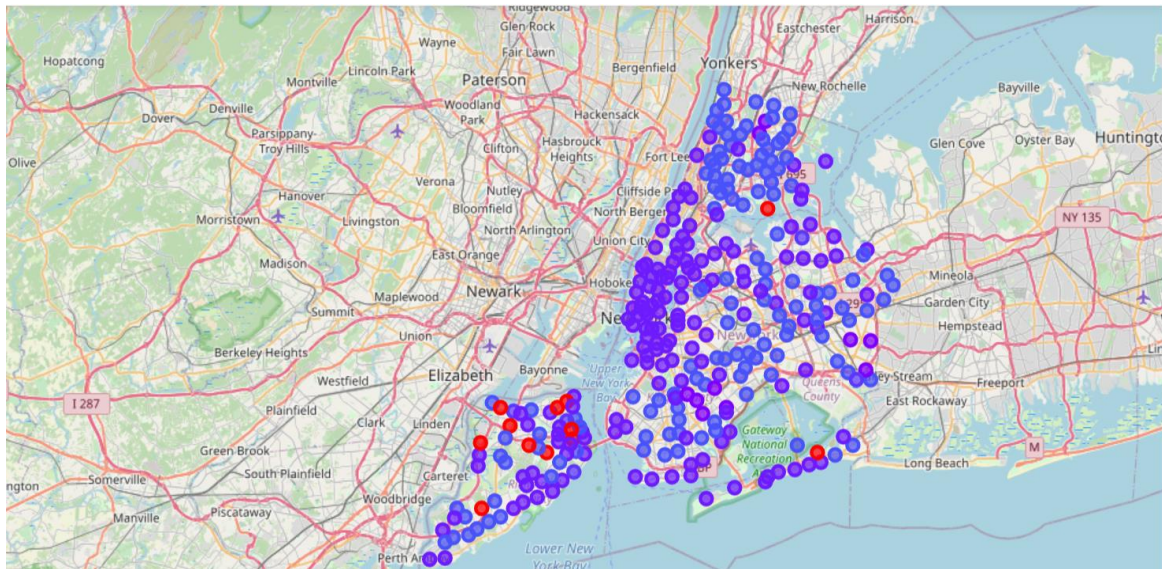
4-Results

Toronto, Canada



KMeans algorithm is used to group neighborhoods in Toronto into 3 clusters. Cluster_0 has 102 neighborhoods and most common venues are international cuisines, parks, bars, coffee shops. Cluster_1 has 2 neighborhood and most common venues are distribution centers, sport areas like baseball field, yoga studio etc. Cluster_2 has 10 neighborhoods and most common venues are parks, gyms, distribution centers.

New York City, New York



KMeans algorithm is used to group neighborhoods in New York into 5 clusters. Cluster_0 12 neighborhoods and most common venues are parks, yoga studios, bus stops. Cluster_1 has 10 neighborhoods and most common venues are caribbean and Chinese restaurants. Cluster_2 has 150 neighborhoods and most common venues are international restaurants and bars. Cluster_3 has 5 neighborhoods and most common venue is Deli/Bodega shops. Cluster_4 has 128 neighborhoods and most common venues are pizza places, Italian and chinese restaurants.

5-Discussion

Toronto has 10 boroughs and 103 neighborhoods. The geographical coordinates are 43.7170226, -79.4197830350134. 2132 venues found in Toronto's 103 neighborhoods.

New York City has 5 boroughs and 306 neighborhoods. The geographical coordinates are 40.7127281, -74.0060152. 9955 venues found in New York's 306 neighborhoods.

Many of the neighborhoods are homogenous and are very similar to each other. Both cities clustering has result one of the cluster's has majority of neighborhoods. Also, New York has significantly higher venues than Toronto.

6- Conclusion

In conclusion, based on the results that I got from analyze I will choose New York to set up my company. New York has more active social life for employer's interests.