

Capstone project: New York- Toronto- Paris clustering

In this project, we suppose our client is working in New York and get two offers of jobs in Toronto and Paris.

He wants to choose the city that has more similarity to New York to work. We will measure the similarity among New York, Toronto and Paris based on the clusters.

- First, for each city, we divide the set of neighborhoods into clusters, then assign each cluster with the name by using the top of venues.
- Second, we consider the clusters of each city to see which city Paris or Toronto is more similar to New York.

Data collected

We need data include:

list of neighborhoods New York city from

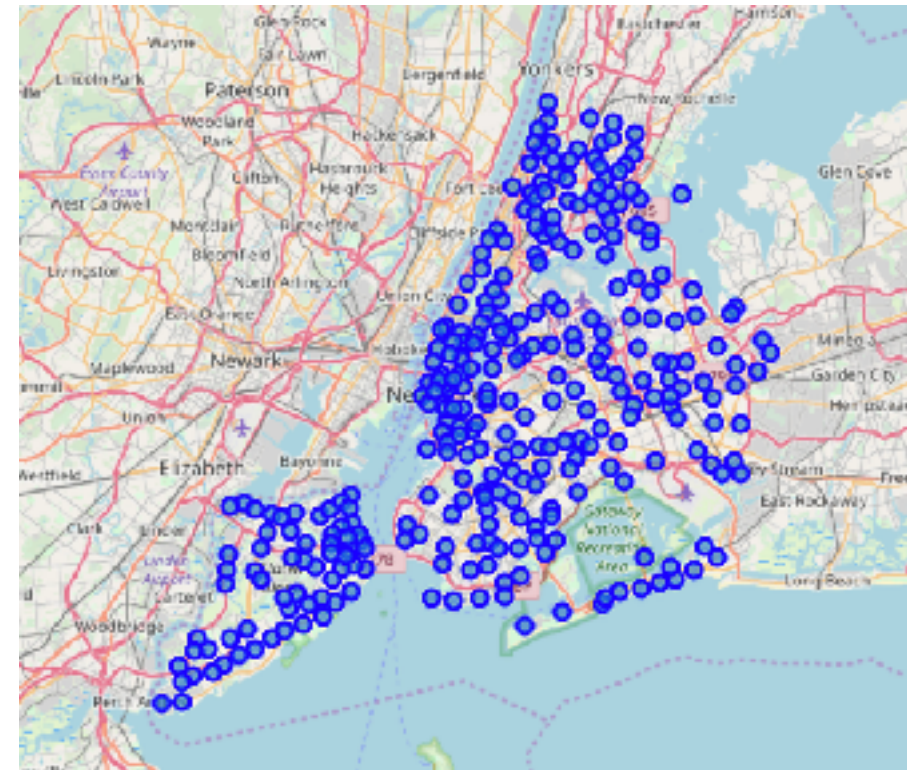
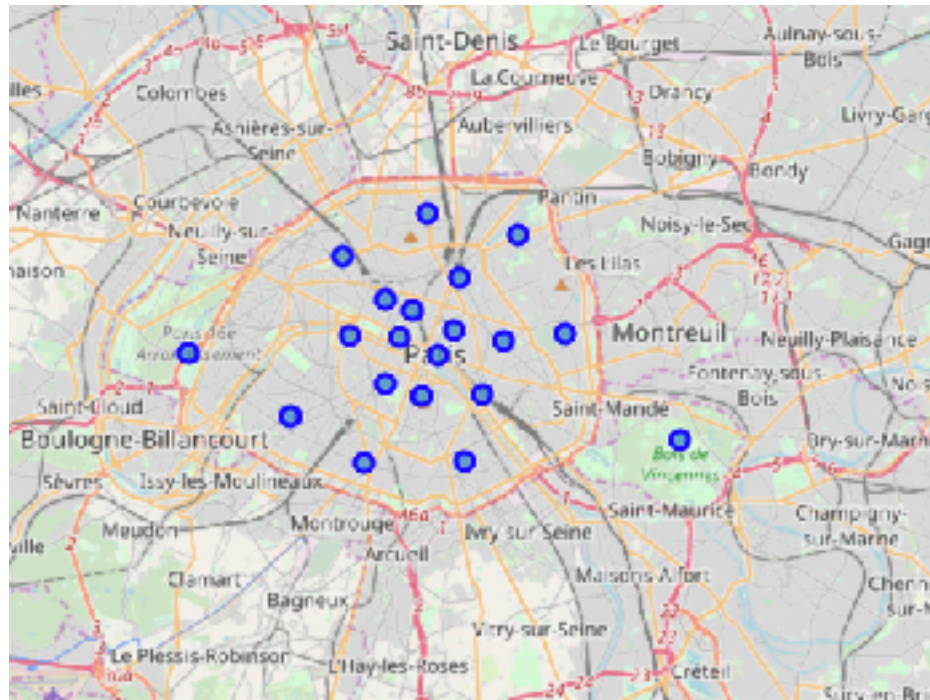
- list of neighborhoods Toronto city from <https://en.wikipedia.org/wiki>
- list_of_postal_codes_of_Canada:_M
- list of neighborhoods Paris city that is given

For each city, we use the Google maps API to search the latitude and longitude of each neighborhoods.

Top of venues for each neighborhood is collected by using Foursquare API.

The neighborhoods

Paris



New York



Toronto

Analysis

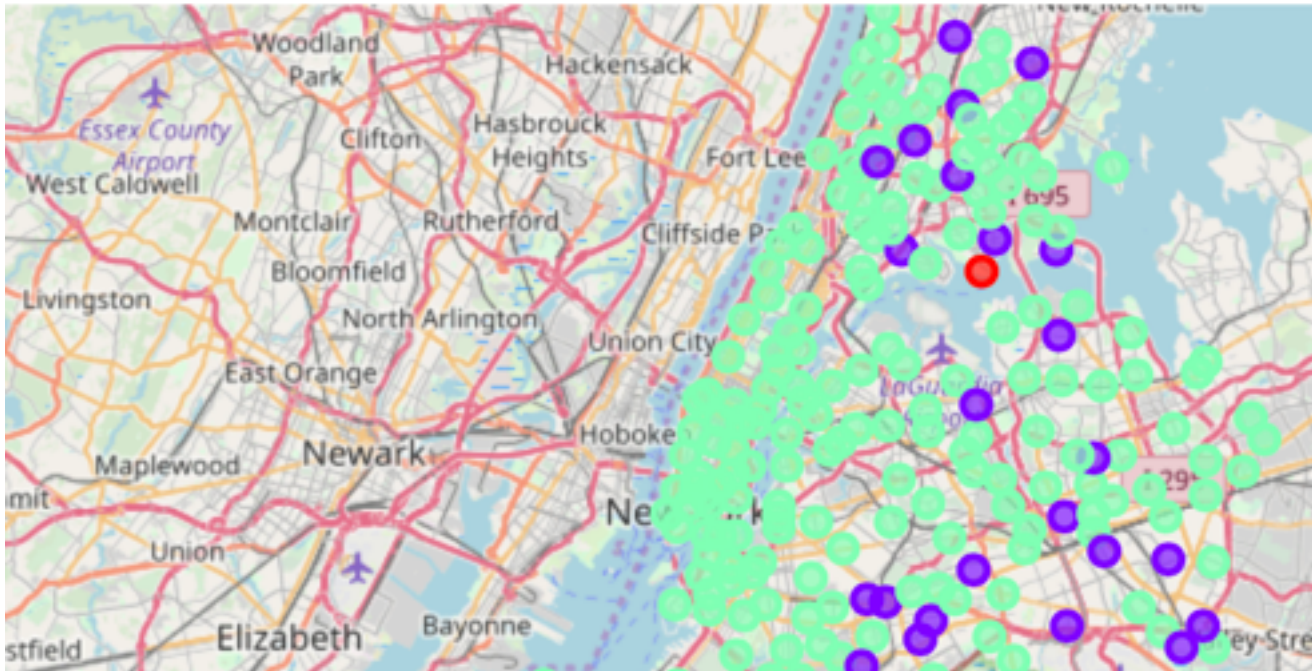
Use the k-means cluster algorithm in scikit learning library to cluster the set of neighborhoods for each city:

- collect the popular venues of every neighborhoods with the radius 500m and limit 30.
- give the top of venues of every neighborhoods and we use hot-coding to code the venues.
- cluster neighborhoods based on the information of the top venues. Then we will assign the name of each cluster for the 1st common venues.

Clustering the neighborhoods into 5 cluster



Toronto



New York



Paris

Names of clusters

label	name
0	Bus Stop
1	Deli / Bodega
2	Supermarket
3	Pizza Place
4	Park

Toronto

label	name
0	French Restaurant
1	Italian Restaurant
2	French Restaurant
3	Basketball Court
4	Vietnamese Restaurant

Paris

label	name
0	coffee shop
1	cafe
2	Airport Service
3	Grocery Store
4	Park

New York

Conclusion

- there are four clusters of restaurants in Paris while only two clusters for restaurants in New York and Toronto
- both of NewYork and Toronto have the 'Park' cluster and one cluster for buying food: supermarket and grocery store.
- both of NewYork and Toronto have cluster for transport: airport service and bus stop.
- both of NewYork and Toronto have cluster for favourite food: Pizza place, Deli/Bodega and coffee,cafe.

New York city is similar to Toronto. Our client should choose Toronto to work.