# Comparing the Cities of New York and Toronto using neighborhood venues from Foursquare

Halimatou B. A.

July 2020

# Introduction

- New York City and Toronto are two similar cities: they are both in the same continent, not very far from each other, multicultural cities and financial hubs in their respective countries
- We will compare the two cities using their neighborhoods venues to cluster them and then analyse the results to see similarities that brought different neighborhoods to be clustered together or not
- This could be of interest to people who are from either of the two cities and looking into relocating to the other city, for instance, businesses present in one of the cities and looking into opening into the other city etc...

# Data (1)

We obtained data for both cities through the following sources:

- Toronto: `https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M`
- New York City: `https://cocl.us/new_york_dataset`
- Foursquare API

After Cleaning the data, the venues were retrieved from Foursquare with a limit of 100 within a radius of 500 miles.

# Data (2)

A total of 12254 venues were retrieved consisting of 459 categories.
We used the technique of one-hot-encoding to generate dummy variables
for each of the category. The following dataframe resulted, with venues
grouped by neighborhood:

```
ny_t_grouped = ny_t_onehot.groupby('Neighborhood').mean().reset_index()
ny_t_grouped.head()
```
]:

|   | Neighborhood | Yoga Studio | Accessories Store | Adult Boutique | Afghan Restaurant | African Restaurant | Airport | Airport Food Court | Airport Gate | Airport Lounge | Airport Service | Airport Terminal | America Restaurar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Agincourt | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00000 |
| 1 | Allerton | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00000 |
| 2 | Annadale | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.07692 |
| 3 | Arden Heights | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.00000 |
| 4 | Arlington | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.12500 |

Confirming size:

```
ny_t_grouped.shape
```
]: (396, 459)

Figure: Dataframe after one-hot- encoding

## Methodology

To cluster the different neighborhoods, we used an unsupervised machinelearning algorithm, Kmeans method from the package Scikit-Learn.
To find the optimal number of clusters k, we used the elbow method. The elbow method did not give a clear elbow shape, so we settled for k= 5.



Elbow Method For Optimal k
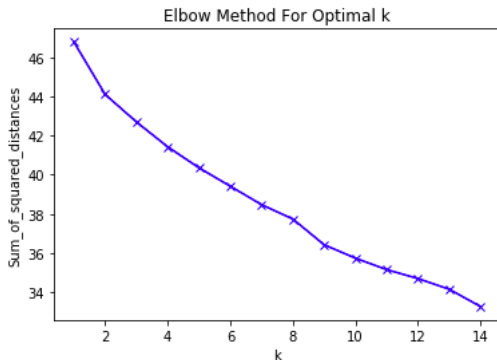
# Results (1)

After executing the code we obtain the following 5 clusters:

| City | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|------|-----------|-----------|-----------|-----------|-----------|
| **New York** | 3 | 84 | 2 | 215 | 1 |
| **Toronto** | 8 | 12 | 0 | 77 | 3 |

Figure: Cluster compostion

We notice that the majority of the neighborhoods in both cities are in cluster 4, followed by cluster 2.
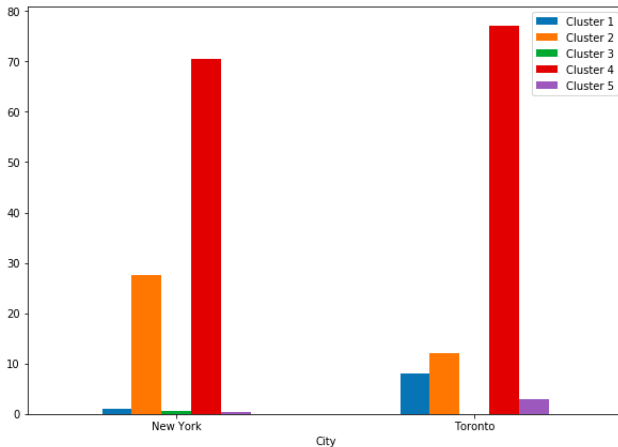
# Results (2)



Figure: Proportion of neighborhood in each cluster per city in percentage

# Results (3)

Nearly 90 and 98 percent of the neighborhoods in Toronto and New York city respectively are in clusters 2 and 4, and cluster 3 contains only 2 cities from New York.

| City | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|---|---|---|---|---|---|
| **New York** | 1.0 | 27.5 | 0.7 | 70.5 | 0.3 |
| **Toronto** | 8.0 | 12.0 | 0.0 | 77.0 | 3.0 |

Figure: Proportion of neighborhoods in each cluster per city in percentage
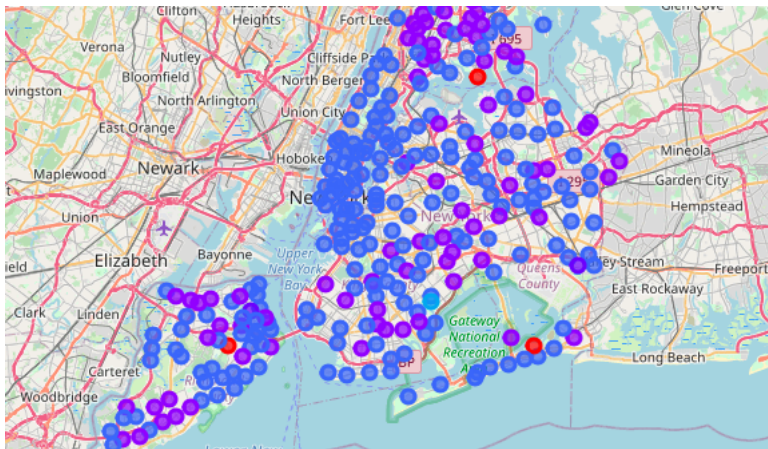
# Results (4)



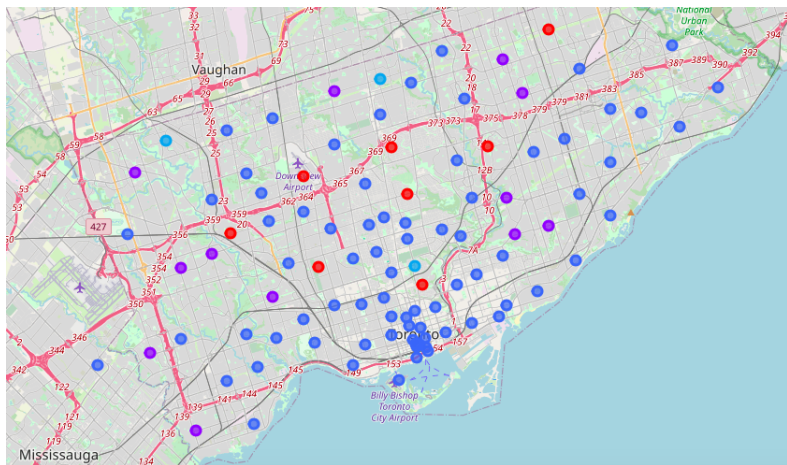Figure: Clusters in New York City

# Results (5)



Figure: Clusters in Toronto

# Discussion

- The results are satisfactory in the sense that we find similar neighborhoods in both cities with the exception of one very small cluster in New York City

- We can say that Toronto and New York city are relatively similar despite the huge size difference between the two cities

- They are both clustered into 2 main groups, but New York city has a 2 neighborhoods that formed a cluster without a neighborhood from Toronto

# Conclusion

Based on our analysis, a person moving between the two cities is almost sure to find a similar neighborhood in either of the two cities. Out of the 5 clusters only one is not found in Toronto, but since it is very small (2 neighborhoods), it can almost be treated as outliers.