New York vs Toronto: A Comparative Analysis

Angelina R. Joseph

Datasets:

- New York neighbourhoods: Sourced from IBM Course Data
- Toronto neighbourhoods: Sourced from Wiki list of postal codes of Toronto
- Foursquare API used to obtain venues within each neighbourhood

Introduction:

What do the world's most famous North American Cities have in common? Where do they differ? Through this analysis the aim is to understand the unique and common characteristics of both the cities, cluster neighbourhood basis their geographical locations, and understand the distribution of various venues and amenities within those clusters. This will help us answer questions like which of these is a city of choice for opening up a new restaurant, a donut franchisee, a gym etc. and which cluster of neighbourhoods should be chosen for that purpose.

Problem Statement:

Which posh North American neighbourhood should be chosen for opening the next big fine dining restaurant?

Target Audience:

Businessmen, Restaurateurs, City planners

Methodology:

1. Setting up environment and libraries
2. Statistical Data exploration
3. Visualizing neighbourhoods using folium
4. Clustering
   a. Identifying optimal clusters using elbow method
   b. Clustering and profiling
   c. Identifying the clusters with the most neighbourhoods
5. Understanding stores and attractions near dense clusters using Foursquare API

Results:

The city of New York has nearly thrice the number of neighbourhoods in Toronto. The most frequently occurring neighbourhoods in these cities are Queens and North York respectively.
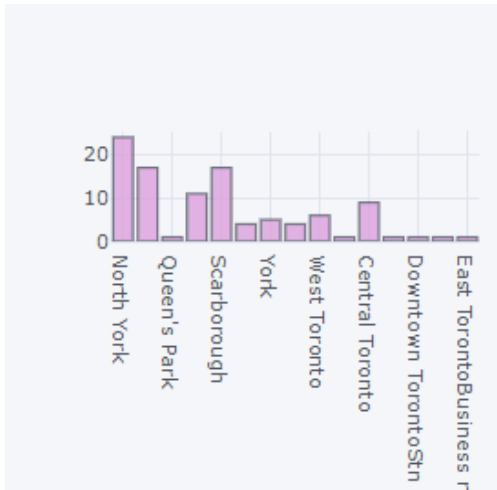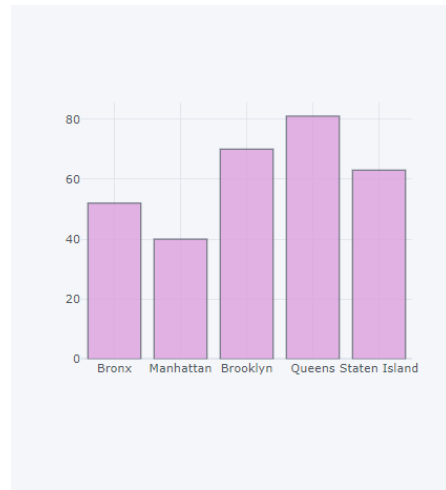
Figure 1: NYC Frequency Distribution

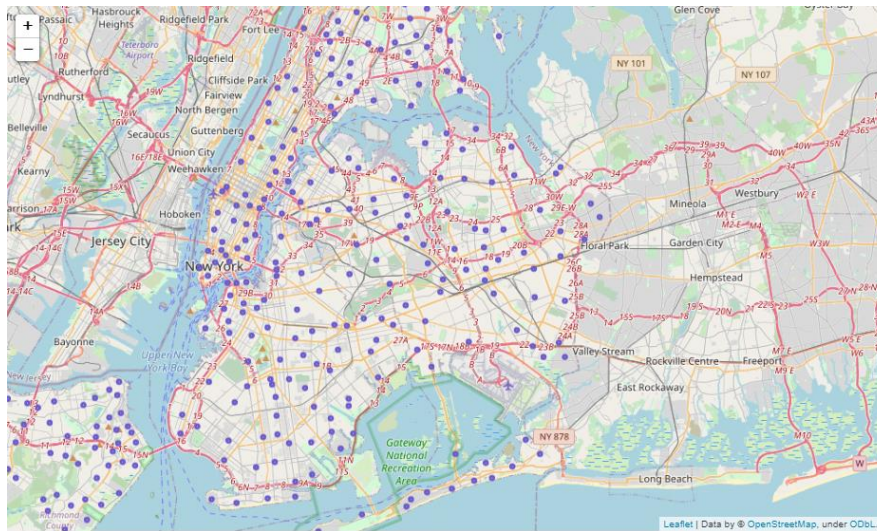*Figure 2: Toronto Frequency Distribution*
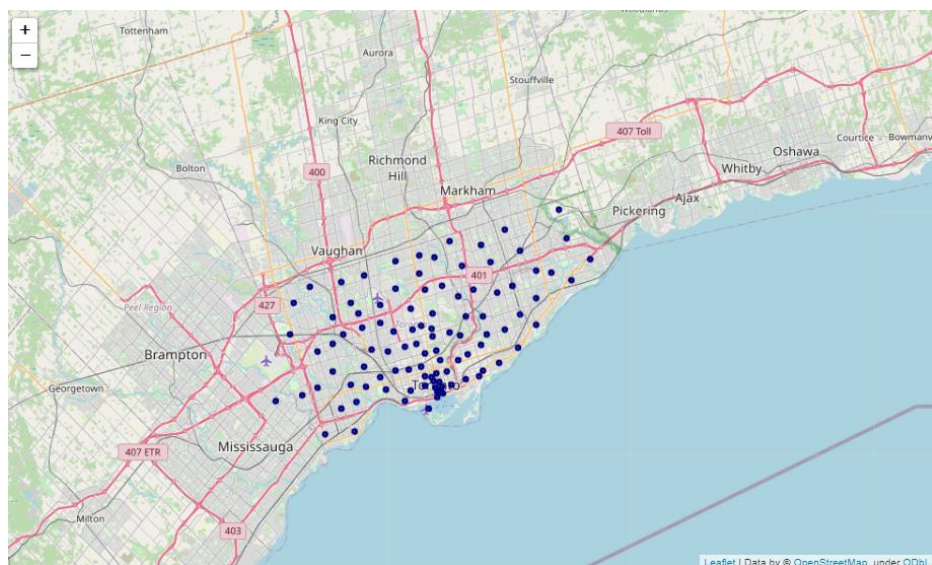

Figure 3: New York Locations

Figure 4: Toronto Locations

K-Means clustering algorithm was used to cluster neighbourhoods based on their geographical location. In order to identify the optimal number of clusters, the elbow method was adopted using an Inertia plot for each of the cities.
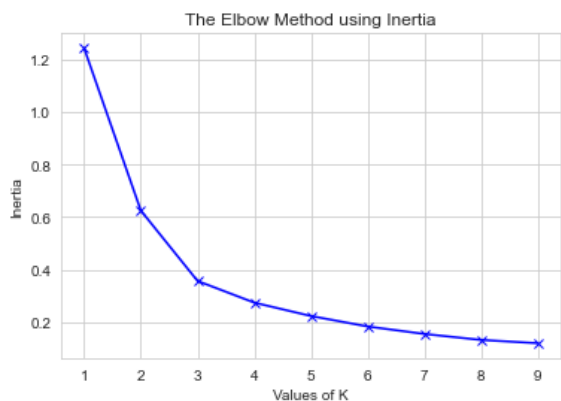

Figure 5: Elbow Method for NYC Segments
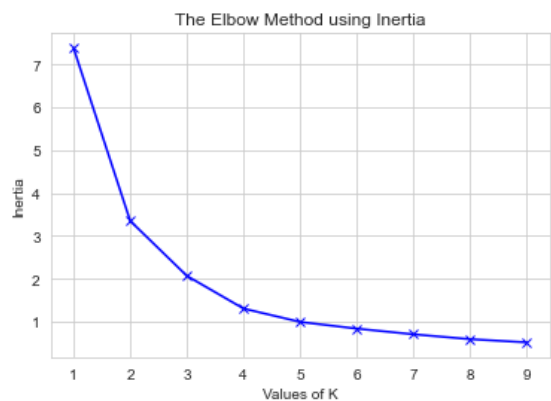
Figure 6: Elbow Method for Toronto Segments

Interestingly, the optimal number of clusters was found to be six for both the cities. The densest clusters for both the cities were identified.

```
In [213]: NYC['cluster_label'].value_counts()

Out[213]: 4    68
          2    64
          0    57
          5    51
          3    44
          1    22
```

```
In [214]: TR['cluster_label'].value_counts()

Out[214]: 3    28
          2    19
          1    18
          5    13
          0    13
          4    12
          Name: cluster_label, dtype: int64
```
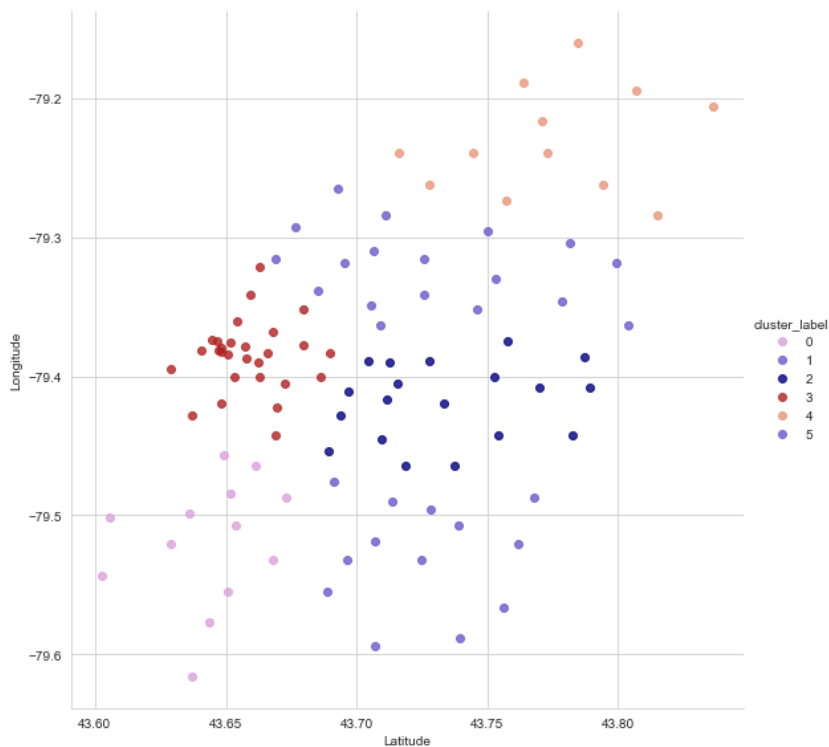

Figure 7: Toronto Clusters
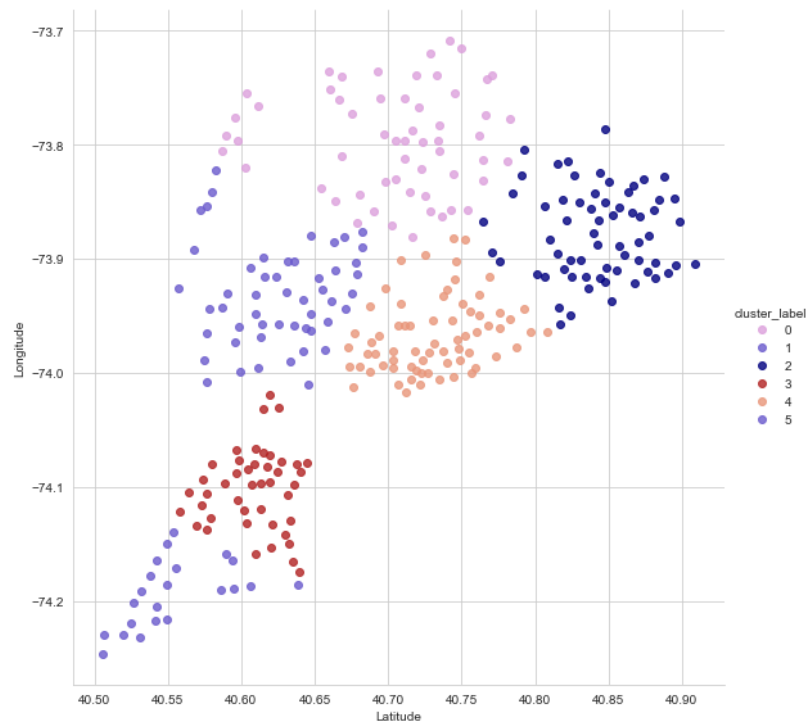
Figure 8: New York Clusters

The clusters formed in both the cities using K-Means are homogenous within the groups and heterogeneous across groups. Clusters in Toronto are sparse due to lesser data points while the clusters in New York are denser and clearly defined.

Discussion:

Upon analysing the locations near the cluster centres for the densest clusters in both of the cities, it was found that there were 48 venues returned by Foursquare for New York while 83 venues were returned for Toronto, despite significantly lesser neighbourhoods.

The most common type of venue near Toronto was found to be Coffee Shops, which were around 14 venues whereas New York's clusters saw a wide variety of different venues with no single category of venue dominating within a cluster.

```
In [284]:  nearby_venues.categories.value_counts()

Out[284]:  Mobile Phone Shop            3
           Sandwich Place              3
           Clothing Store              3
           Pizza Place                 3
           Fried Chicken Joint         2
           Storage Facility            2
           Coffee Shop                 2
           Department Store            2
           Donut Shop                  2
           Mexican Restaurant          2
           Caribbean Restaurant        2
           Kids Store                  1
           Laundromat                  1
           Ice Cream Shop              1
           Multiplex                   1
           Performing Arts Venue       1
           Shoe Store                  1
           Gym                         1
           Supplement Shop             1
           Bank                        1
           American Restaurant         1
           Chinese Restaurant          1
           Discount Store              1
           Music Store                 1
           Gym / Fitness Center        1
           Hardware Store              1
           Concert Hall                1
           Park                        1
           Intersection                1
```

Figure 9: New York Cluster Venues

```
In [265]:  nearby_venues.categories.value_counts()

Out[265]:  Coffee Shop                 14
           Burger Joint                 3
           Bubble Tea Shop              3
           Sandwich Place               3
           Hotel                        3
           Sushi Restaurant             3
           Italian Restaurant           3
           Movie Theater                2
           Department Store             2
           Salad Place                  2
           Thai Restaurant              2
           Restaurant                   2
           Falafel Restaurant           2
           Tea Room                     1
           Chinese Restaurant           1
           Gastropub                    1
           Deli / Bodega                1
           Breakfast Spot               1
           Mexican Restaurant           1
           Juice Bar                    1
           Clothing Store               1
           Bookstore                    1
           Fast Food Restaurant         1
           Poke Place                   1
           Indian Restaurant            1
           Portuguese Restaurant        1
           Art Museum                   1
           Art Gallery                  1
           Wine Bar                     1
           Pharmacy                     1
           Discount Store               1
```

Figure 10: Toronto Cluster Venues

Conclusion:

Both the cities already have varying kinds of venues with denser clusters of Toronto having higher number of venues than New York. If a location were to be chosen for opening a new branch of a Fine Dining restaurant, Cluster 3 of Toronto seems to be an ideal choice for such a venue since the area is a hub of restaurants and has the potential to attract new customers.