Battle of the Neighborhoods

Comparing and Clustering Neighborhoods of New York City and Toronto

Capstone Project by Ajay Bhatnagar

Introduction

New York City and Toronto are both very densely populated and diverse cities which are also the financial capitals of their respective countries. New York City is located in the state of New York of the United States of America while Toronto is located in the Ontario province of Canada. It would be interesting to **compare and contrast the neighborhoods of New York City and Toronto**.

Specifically, we would like to find out answers to questions such as:

- What are the most common venue categories in each city?
- What venue categories are most widespread in each city?
- How many neighborhoods of these two cities are similar to each other?
- How many neighborhoods of these two cities are quite unique to themselves?
- Are these two cities very similar to each other or very different?

Audience:

People who are considering moving from either New York City to Toronto or vice-versa.

P**lanners and officials of these cities and their neighborhoods** to be able to look at data trends and decide which types of businesses to promote based on what appears to be working in the similar neighborhoods of the other city

Data Acquisition and Wrangling

Neighborhoods Data

New York City - 306 Neighborhoods

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	NYC_Wakefield	40.894705	-73.847201
1	Bronx	NYC_Co-op City	40.874294	-73.829939
2	Bronx	NYC_Eastchester	40.887556	-73.827806
3	Bronx	NYC_Fieldston	40.895437	-73.905643
4	Bronx	NYC_Riverdale	40.890834	-73.912585

Toronto - 103 Neighborhoods

	Borough	Neighborhood	Latitude	Longitude
0	North York	YYZ_Parkwoods	43.752935	-79.335641
1	North York	YYZ_Victoria Village	43.728102	-79.311890
2	Downtown Toronto	YYZ_Regent Park, Harbourfront	43.650964	-79.353041
3	North York	YYZ_Lawrence Manor, Lawrence Heights	43.723265	-79.451211
4	Downtown Toronto	YYZ_Queen's Park, Ontario Provincial Government	43.661790	-79.389390

Data Acquisition and Wrangling

Understanding Foursquare Venue Categories

Foursquare Venue Categories

```
Level 1 Categories: 10
Level 2 Categories: 459
Level 3 Categories: 370
```

Level 1 and Level 2 Venue Categories

	Level1	Level2 Count
0	Arts & Entertainment	36
1	College & University	23
2	Event	12
3	Food	92
4	Nightlife Spot	7
5	Outdoors & Recreation	62
6	Professional & Other Places	43
7	Residence	5
8	Shop & Service	145
9	Travel & Transport	34

Data Acquisition and Wrangling

Venues Data

New York City (~ 26,000) and Toronto (~8,500)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	NYC_Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
1	NYC_Wakefield	40.894705	-73.847201	Pitman Deli	40.896744	-73.844398	Food
2	NYC_Wakefield	40.894705	-73.847201	Julio C Barber Shop 2	40.892648	-73.855725	Salon / Barbershop
3	NYC_Wakefield	40.894705	-73.847201	Pittman Ave bodega	40.896744	-73.844398	Convenience Store
4	NYC_Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	YYZ_Parkwoods	43.752935	-79.335641	Church Of Our Saviour	43.751496	-79.337078	Church
1	YYZ_Parkwoods	43.752935	-79.335641	Three Valleys Public School	43.750595	-79.337341	School
2	YYZ_Parkwoods	43.752935	-79.335641	GTA Restoration | Emergency Water Damage Plumb...	43.753567	-79.351308	Construction & Landscaping
3	YYZ_Parkwoods	43.752935	-79.335641	Mo's Ride	43.755123	-79.334583	General Travel
4	YYZ_Parkwoods	43.752935	-79.335641	Bruno's Fine Foods	43.745608	-79.336772	Grocery Store

Methodology

The objective of this project is to compare and contrast the neighborhoods of New York City and Toronto.

In previous step we have collected all the required data:

- **Neighborhood Data** for New York & Toronto using JSON file, web scraping a wiki page & latitude/longitude coordinates using geocoder with ArcGIS provider.
- **Venues Data** for New York City and Toronto using Foursquare API and removing venues of any categories not relevant to our analysis.
- Details about **Venue Categories** used by Foursquare to categorize Venues - these categories will be used as the features for our clustering analysis.

In the next step, we will begin our analysis of the neighborhoods of New York City and Toronto by first **visualizing the neighborhoods of each city**, exploring venues of each city by finding out the **total number of distinct venue categories** in each city, **most common and most widespread venue categories** across the neighborhoods of each city. We will then **prepare the data needed for the clustering algorithm** and we will also generate the **top 10 most common categories in each neighborhood**.

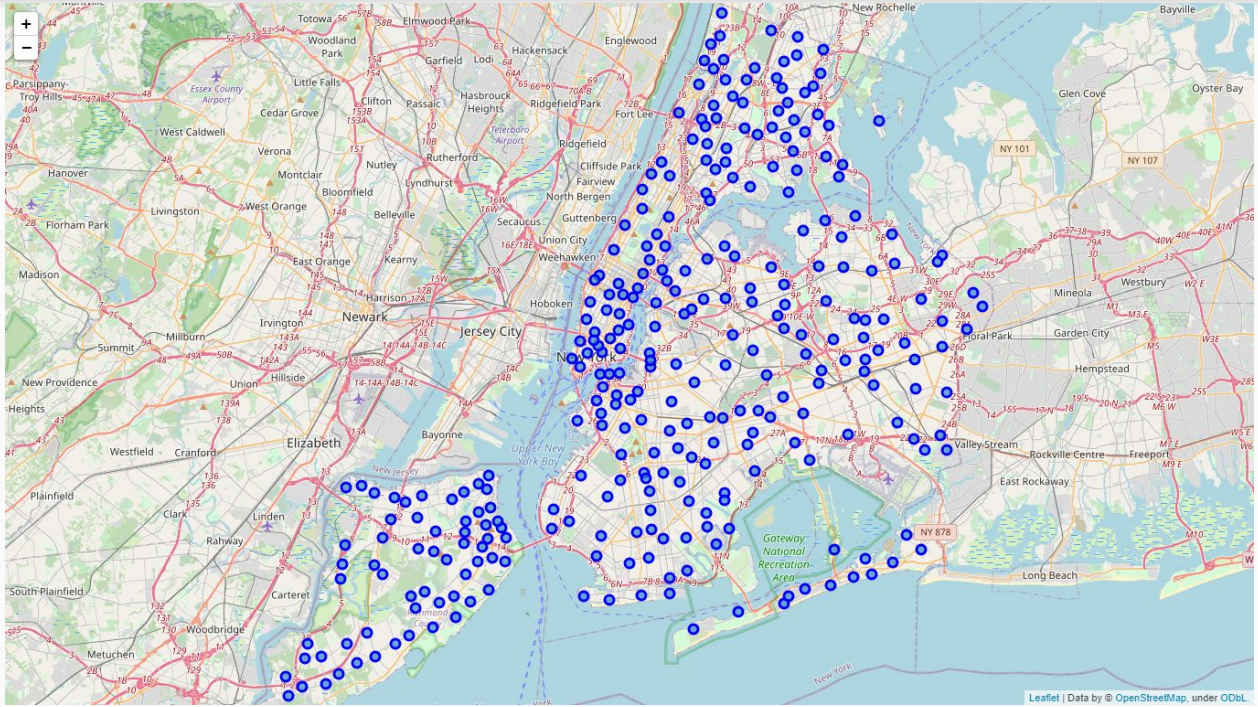
Methodology (cont’d)

In the next step, we will begin our analysis of the neighborhoods of New York City and Toronto by first **visualizing the neighborhoods of each city**, exploring venues of each city by finding out the **total number of distinct venue categories** in each city, **most common and most widespread venue categories** across the neighborhoods of each city. We will then **prepare the data needed for the clustering algorithm** and we will also generate the **top 10 most common categories in each neighborhood**.

In the final step we will carry out the clustering of all neighborhoods of both cities by first combining all relevant datasets for both cities into one. In order to use **K-Means Clustering**, we will generate a dataset of all features (all venue categories) to do the clustering on and we will use the **Elbow Method** as well as the **Silhouette Score Method** to determine the **optimal number of clusters for our dataset**. We will then run the clustering on the optimal number of clusters. Finally, we will then do **analysis of each cluster to name and explain them** so that they can be used appropriately in our Results and Discussion section to meet the project objective.

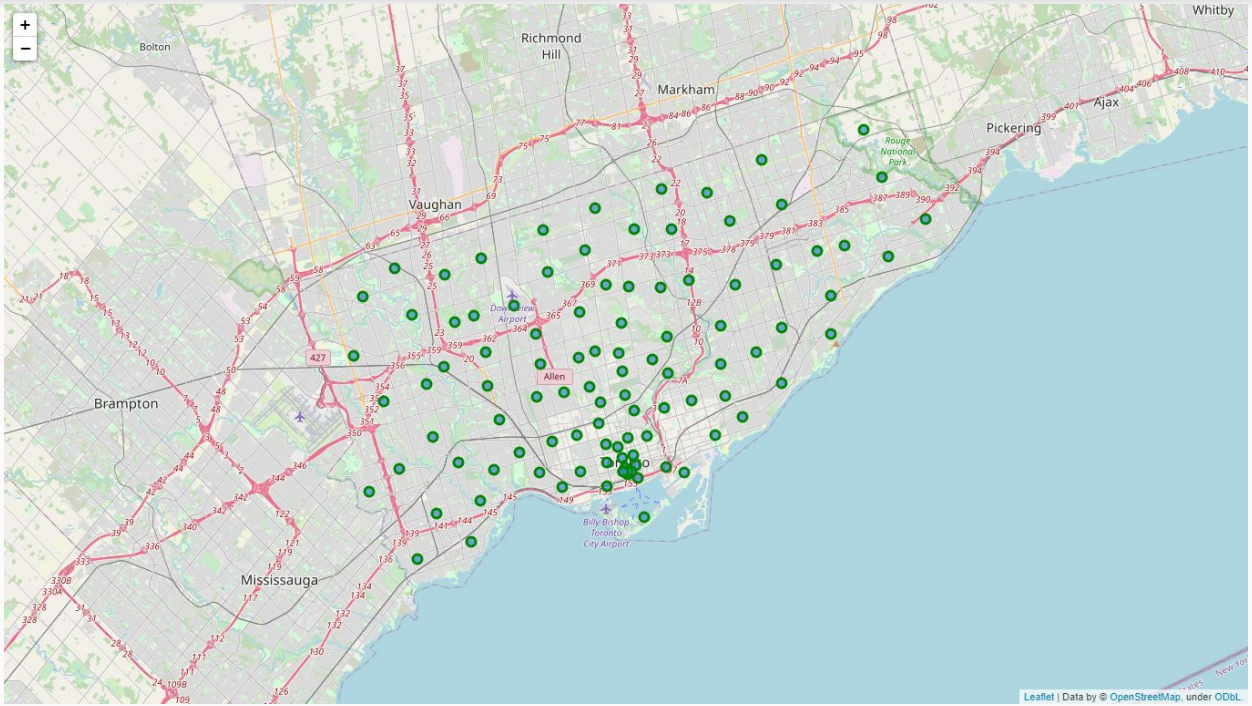
Exploratory Data Analysis

Visualizing New York City Neighborhoods



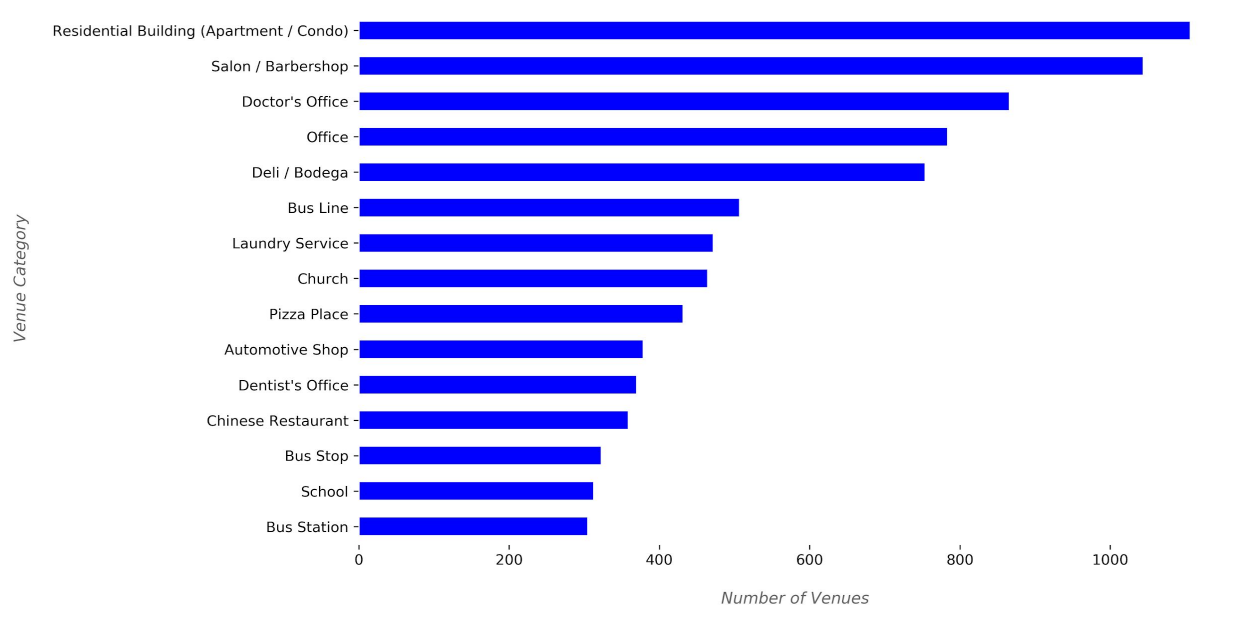
Exploratory Data Analysis (cont'd)

Visualizing Toronto Neighborhoods



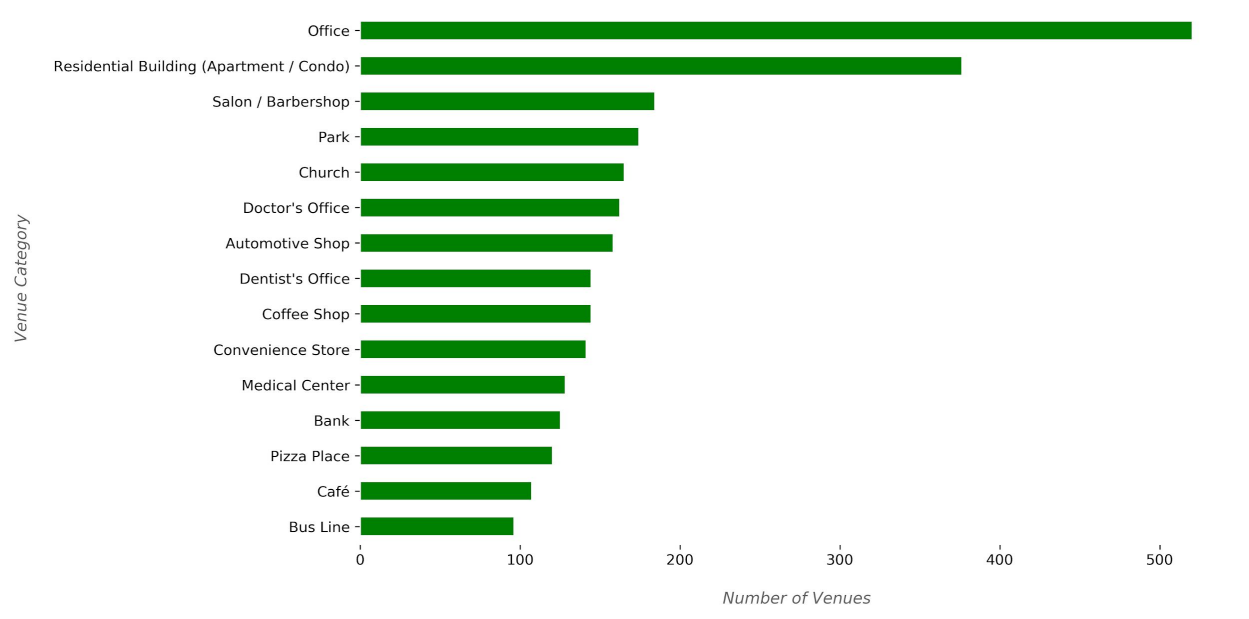
Exploratory Data Analysis (cont'd)

Most Common Venue Categories - New York City



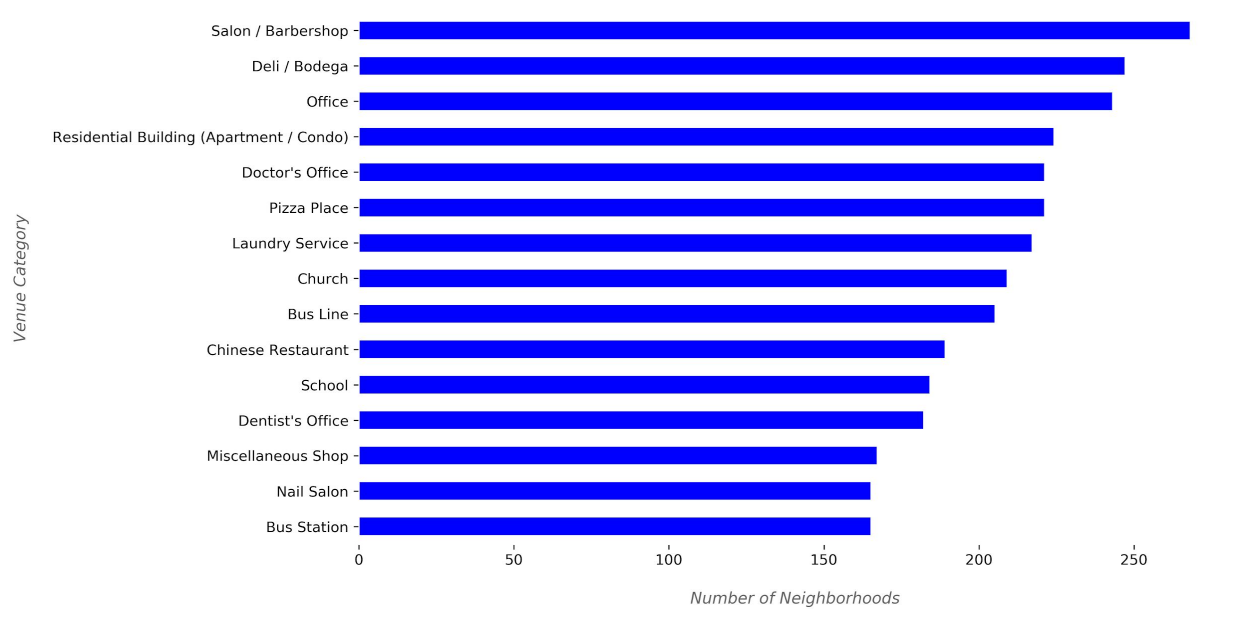
Exploratory Data Analysis (cont'd)

Most Common Venue Categories - Toronto



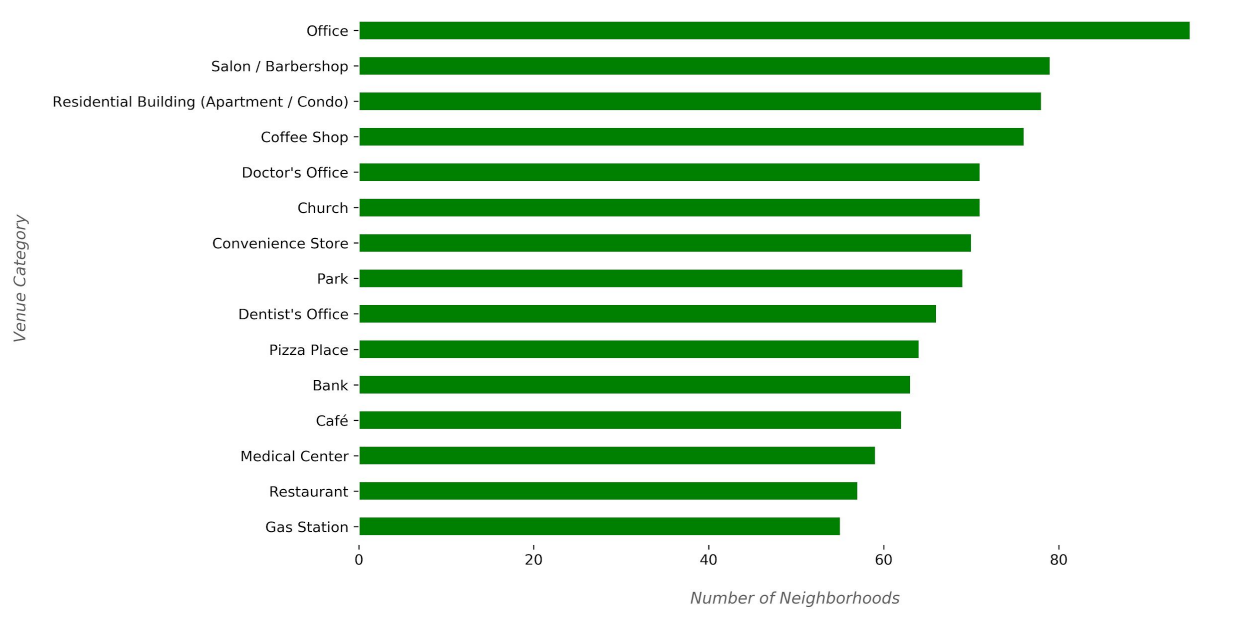
Exploratory Data Analysis (cont'd)

Most Widespread Venue Categories - New York City



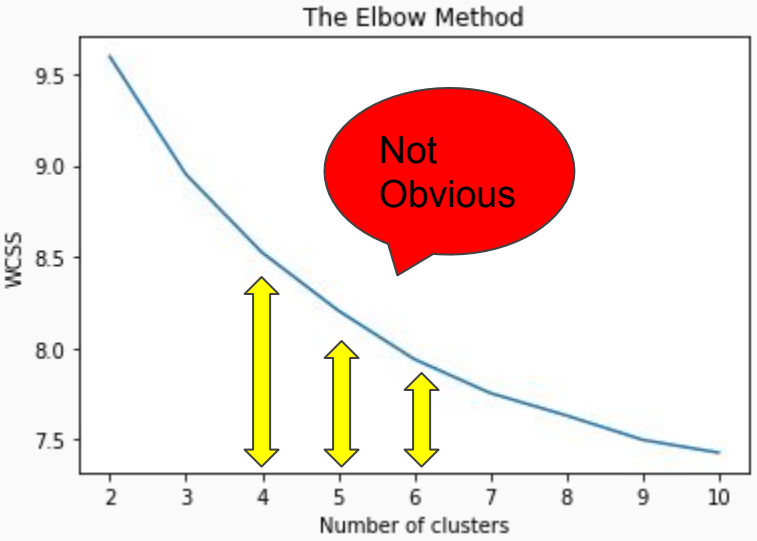
Exploratory Data Analysis (cont'd)

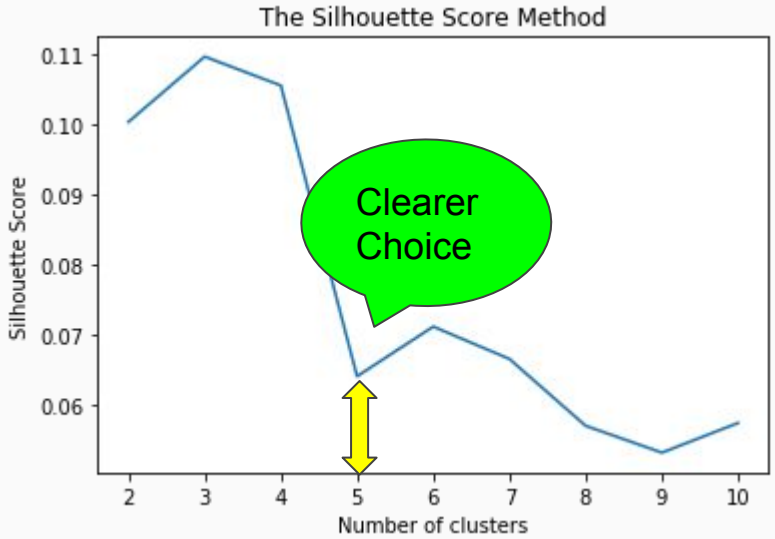
Most Widespread Venue Categories - Toronto



Clustering the Neighborhoods

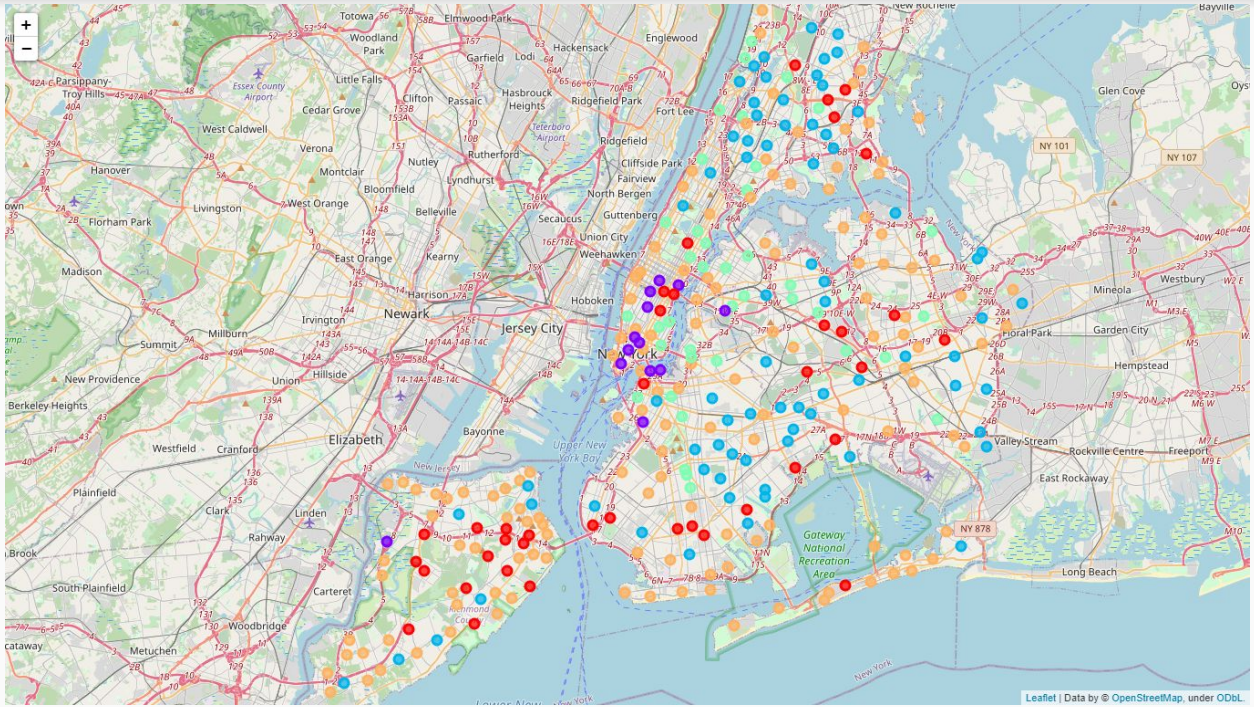
Determining Optimal Number of Clusters for K-Means





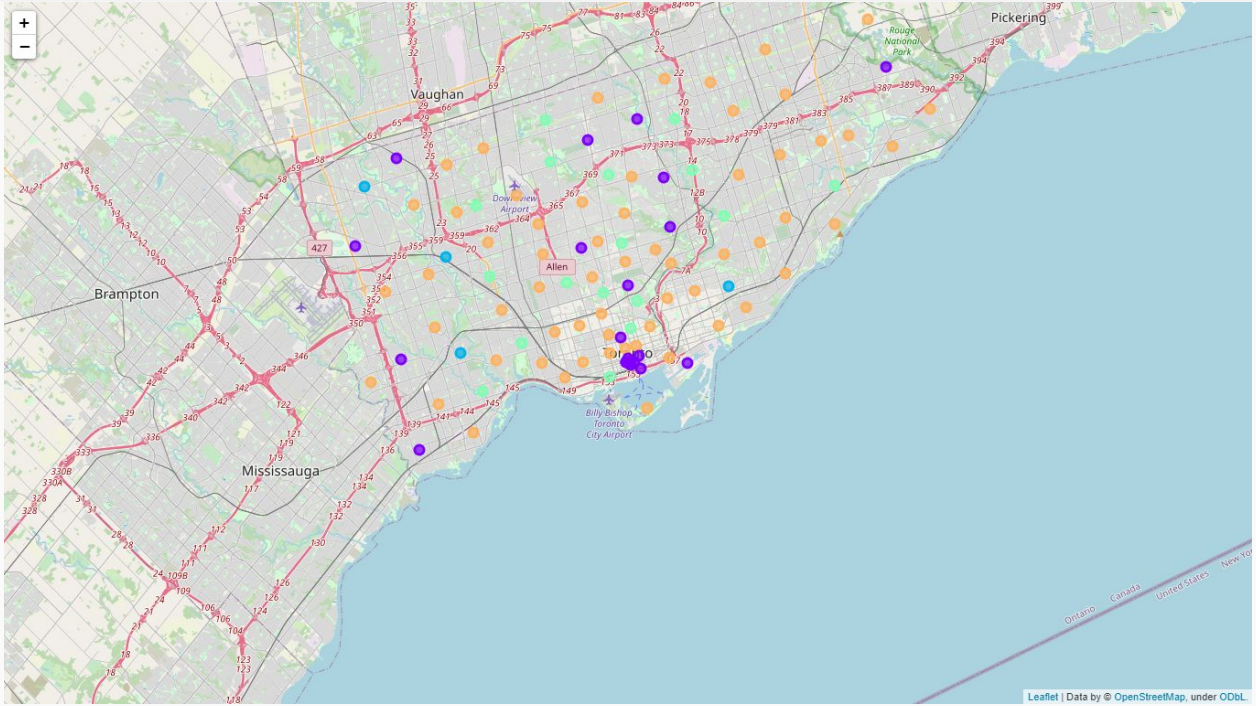
Clustering the Neighborhoods

Visualizing New York City Neighborhoods - Clustered



Clustering the Neighborhoods

Visualizing Toronto Neighborhoods - Clustered



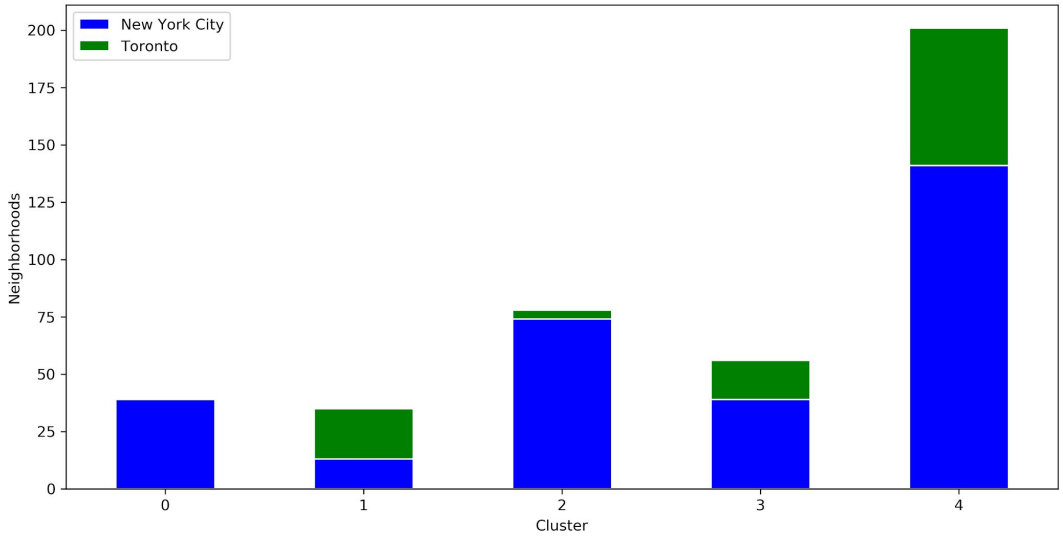
Clustering the Neighborhoods

Analyzing the Neighborhoods of each city across the clusters

Cluster Labels	City	
0	New York City	39
1	New York City	13
	Toronto	22
2	New York City	74
	Toronto	4
3	New York City	39
	Toronto	17
4	New York City	141
	Toronto	60

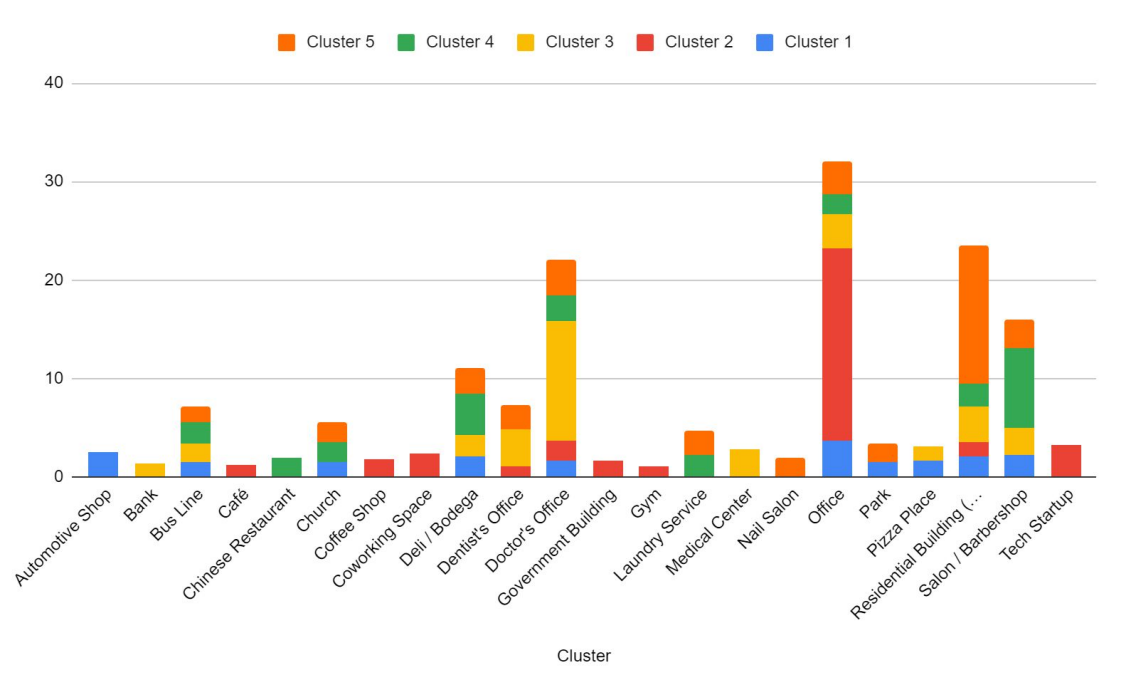
Clustering the Neighborhoods

Analyzing the Neighborhoods of each city across the clusters



Clustering the Neighborhoods

Analyzing the Clusters and Venue Category Frequencies



Clustering the Neighborhoods

Analyzing & Naming Cluster 1 (Cluster Label = 0)

“Neighborhoods around Medical Facilities”

Doctor's Office	13.113695
Residential Building (Apartment / Condo)	4.748062
Dentist's Office	3.779070
Office	3.068475
Medical Center	2.680879
Salon / Barbershop	2.616279
Deli / Bodega	2.357881
Bus Line	2.164083
Laundry Service	1.647287
Bus Stop	1.518088

Clustering the Neighborhoods

Analyzing & Naming Cluster 2 (Cluster Label = 1)

“Neighborhoods that are Business and Commercial Centers”

Office	15.815486
Doctor's Office	2.009885
Tech Startup	2.009885
Coworking Space	1.878089
Residential Building (Apartment / Condo)	1.878089
Coffee Shop	1.647446
Automotive Shop	1.614498
Government Building	1.285008
Café	1.252059
Gym	1.252059

Clustering the Neighborhoods

Analyzing & Naming Cluster 3 (Cluster Label = 2)

“Neighborhoods that are sparsely Residential with above average access to services”

Salon / Barbershop	8.503961
Deli / Bodega	3.900762
Residential Building (Apartment / Condo)	2.660290
Church	2.510835
Bus Line	2.465999
Laundry Service	2.436108
Doctor's Office	2.196981
Pizza Place	2.032581
Chinese Restaurant	2.002690
Nail Salon	1.808399

Clustering the Neighborhoods

Analyzing & Naming Cluster 4 (Cluster Label = 3)

“Neighborhoods that are densely Residential with good access to services”

Residential Building (Apartment / Condo)	14.517189
Salon / Barbershop	3.109262
Doctor's Office	2.912196
Office	2.758923
Laundry Service	2.255310
Deli / Bodega	2.123933
Church	1.839282
Park	1.642216
Dentist's Office	1.642216
Bus Line	1.335669

Clustering the Neighborhoods

Analyzing & Naming Cluster 5 (Cluster Label = 4)

“Neighborhoods with balanced Residential and Business presence”

Office	2.927847
Residential Building (Apartment / Condo)	2.605844
Salon / Barbershop	2.391175
Automotive Shop	2.134764
Deli / Bodega	1.967800
Church	1.920095
Pizza Place	1.747168
Bus Line	1.717352
Doctor's Office	1.669648
Bus Stop	1.443053

Results and Discussion

Most Common and Most Widespread Venue Categories

Our analysis of **Most Common Venue Categories** of the two cities shows that **New York City has Residential as the topmost common venue category overall while Toronto has Office as the topmost common venue category**. In New York City, Salon / Barbershop, Doctor's Office, Office and Deli / Bodega are the other categories in top 5 common while Toronto has Residential, Salon / Barbershop, Park and Church in top 5 common.

Our analysis of **Most Widespread Venue Categories** (that exist in most neighborhoods) shows that **New York City has Salon / Barbershop as the most widespread venue category while Toronto has Office as the most widespread venue category**. In New York City, Deli / Bodega, Office, Pizza Place & Doctor's Office are the other widespread categories in top 5 widespread while Toronto has Salon / Barbershop, Residential, Coffee Shop and Doctor's Office in top 5 widespread.

- **In New York City, Residential is the Most Common Venue Category but it is only the 4th Most Widespread Venue Category which means that this venue category is highly concentrated in some of the 306 neighborhoods.**
- **In Toronto, Office is the Most Common Venue Category and it is also the Most Widespread Venue Category which means that this venue category is highly concentrated in most of the 103 neighborhoods.**

Results and Discussion

Cluster Analysis and Naming

There are **3 clusters which have similar neighborhoods from both New York City and Toronto**.

- **“Neighborhoods with balanced Residential and Business presence”**
  - New York City = 141 & Toronto = 60
- **“Neighborhoods that are Business and Commercial Centers”**
  - New York City = 13 & Toronto = 22
- **“Neighborhoods that are densely Residential with good access to services”**
  - New York City = 39 & Toronto = 17

There are **2 clusters which have neighborhoods that are quite distinct in New York City**.

- **“Neighborhoods around Medical Facilities”**
  - New York City = 39 & Toronto = 0
- **“Neighborhoods that are sparsely Residential with above average access to services”**
  - New York City = 74 & Toronto = 4

Conclusion

- What are the most common venue categories in each city?

  New York City has Residential as the topmost common venue category overall while Toronto has Office as the topmost common venue category. In New York City, Salon / Barbershop, Doctor's Office, Office and Deli / Bodega are the other categories in top 5 while Toronto has Residential, Salon / Barbershop, Park and Church in top 5

- What venue categories are most widespread in each city?

  New York City has Salon / Barbershop as the most widespread venue category while Toronto has Office as the most widespread venue category. In New York City, Deli / Bodega, Office, Pizza Place & Doctor's Office are the other widespread categories in top 5 widespread while Toronto has Salon / Barbershop, Residential, Coffee Shop and Doctor's Office in top 5 widespread

Conclusion (cont'd)

- How many neighborhoods of these two cities are similar to each other?

  There are 3 sets of neighborhoods of these two cities that are very similar to each other

  - *"Neighborhoods with balanced Residential and Business presence"*
    - *New York City = 141 & Toronto = 60*
  - *"Neighborhoods that are Business and Commercial Centers"*
    - *New York City = 13 & Toronto = 22*
  - *"Neighborhoods that are densely Residential with good access to services"*
    - *New York City = 39 & Toronto = 17*

Conclusion (cont'd)

- How many neighborhoods of these two cities are quite unique to themselves?

There are 2 sets of neighborhoods of these two cities that are very unique to themselves and all but one of them are in New York City

- *"Neighborhoods around Medical Facilities"*
  - *New York City = 39 & Toronto = 0*
- *"Neighborhoods that are sparsely Residential with above average access to services"*
  - *New York City = 74 & Toronto = 4*

Conclusion (cont'd)

- Are these two cities very similar to each other or very different?

The fact that 3 out of the 5 clusters have a significant number of each city's neighborhoods (~ 63% of New York City & ~ 96% of Toronto) in the similar cluster buckets indicates that these two cities are very similar to each other.

There are 2 sets of neighborhoods almost all in New York City (they represent ~ 37% of the neighborhoods of that city) that are quite unique to themselves and have insignificant (~ 4%) similar neighborhoods in Toronto.

Thank You