# Capstone Project – The Battle of Neighborhoods

By: Tracy Ji

# Introduction:

- Many people constantly seek new job opportunities within the same community they live in or across the city or even a different city itself. Let's say a person got an interesting job offer from a different city , say New York and he/she lives in Downtown Toronto currently. It would be really helpful to seek a place to live which is most similar to the current living location of that person.Some of the popular location categories one might look for in the proximity of a living area are grocery stores, colleges or/and schools, parks, restaurants, coffee-shops, hospitals and other community areas like religious places, community halls, libraries, etc. So I'll find out what are borough-neighborhoods that are very similar to a person's current location.

# Business Intention & Similarity

- This information can help the person decide on which neighborhood he/she would love to live in once he/she moves to New York after accepting a new job offer.
- But how do we compare 2 cities and measure similarity? Fortunately, Foursquare offers venue category to each venue, this information can be used to count the number of venues for each category and compare it with all neighborhoods of New York.

# a. New York Data

- All the venues of New York City are available in a dataset provided by PLUTO which is a part of the Department of City Planning (DCP). It contains Burough, Latitude, Longitude for each venue in New Yor City. Using Foursquare API, we can get Neighborhood information for each venue in the dataset. Another way to get New york city dataset is from Foursquare API with getting latitude and longitude of the center of New York City and getting all venues within a 500-mile radies and filtering out all venues with New York City as city name for the venue.

By passing the latitude and longitude into Foursquare API, we can get venue information like **Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category**

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

There are a total of 431 unique venue categories and 10103 venues in New York City

## b. Toronto Data

- A similar exercise can be done for to get all venues in Toronto by Foursquare API or as we did int he previous exercise, download the Toronto Zip Codes from Wikipedia, parse the wikipedia webpage using BeautifulSoup package. Using the geolocator or the csv file provided by the assignment, we can get latitude and longitiude for the neighborhoods and iterate through the Foursquare API to get all venues in each neighborhood.

By passing the latitude and longitude into Foursquare API, we can get venue information like **Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category**

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

There are a total of 273 unique venue categories and 4342 venues in Toronto

# Methodology

**Data Wrangling**

Below steps will be followed to clean the data from cities datasets

Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned.

More than one neighborhood can exist in one postal code area. Split the neighborhoods as one per line.

If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

# Data Analysis

- **Find common venue categories to New York and Toronto**
- Given a neighbourhood in New York, if a similar neighborhood needs to be found in Toronto, we need to consider all the similar venue categories to selected burough of New York and Toronto before grouping.

Toronto venues common to New York: 243
New York unique venue that will be discarded: 188
Toronto unique venues that will be discarded: 30

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

after cleaning the dataframe:
Total venues in Toronto now: 4201
Total venues in New York City now: 9233

**Manhattan neighborhood**
let's group rows by neighborhood and by taking the maximum of occurrence of each category, this gives the idea whether that venue is present (with value 1) or not (with value 0).

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

**Toronto neighborhood**

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

**Finding similar neighborhood**

Let's take an example of a place where a person currently resides in Downtown Toronto Regent Park and find a similar neighborhood in New York by cosine similarity

# Matrix Multiplication

This appoach will be used to amplify similar venue categories and discard dissimilar venue categories when comparing 2 neighborhoods

For our problem we need to find NY neighborhoods with maximum matching categories with given neighborhood in Toronto and discard dissimilar venue categories

# Methodology: Matrix Multiplication

Steps to get matching locations:

- isolate columns from dataframes to get common venue categories columns
- multiply given index of toronto with transpose of NY data values
- sort the dot product of matrices to get most matched
- save the matching locations into a dataframe

For example, if the person deciding to move NY is currently staying at Harbourfront, Downtown Toronto.

The Index of this location in the toronto dataframe is 3
Harbourfront location Latitude : 43.6542599
Harbourfront location Longitude: -79.3606359

# Results

- The top 5 locations in New York that match the Harbourfront, Downtown Toronto.
- The bar graph shows comparision of the venue categories and their number of occurrences to the matched locations in New York. This gives a fair idea how each matched place is New York is close to the current living place in Toronto with number of categories matched as closely as possible.

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |


| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

# Observations

The following observations are being made after analyzing the data above.

- New York has double in venue categories than Toronto.
- There are more Buroughs in Toronto than New York but there are more neighborhoods in New York than Toronto.

# The matched locations observations are as follows:

- This location matches 9/10 most frequent venues in comparision to Toronto.
- Park is missing in this location but this neighborhood is the best match of all the matched locations.
- 11 venues that are missing in this location - Park,Performing Arts Venue,Mexican Restaurant,Historic Site,French Restaurant,Farmers Market,Antique Shop,Electronics Store,Chocolate Shop,Beer Store,Art Gallery
- This location matches 9/10 most frequent venues in comparision to Toronto.
- Park is also missing in this location.
- 11 venues that are missing in this location - Park,Performing Arts Venue,Mexican Restaurant,Historic Site,French Restaurant,Farmers Market,Antique Shop,Electronics Store,Chocolate Shop,Beer Store,Art Gallery
- This location matches 8/10 most frequent venues in comparision to Toronto.
- It has Park but no Pub and no Breakfast Spot. Although Breakfast spot is absent, there are Cafés and Restaurants and Coffee Shops in the area.
- 11 venues that are missing in this location - Pub,Breakfast Spot,Shoe Store,Restaurant,Hotel,Event Space,Farmers Market,Dessert Shop,Chocolate Shop,Beer Store,Art Gallery
- This location matches 7/10 most frequent venues in comparision to Toronto.
- It has Park but no Theater no Pub and no Breakfast Spot. Although Breakfast spot is absent, there are Cafés and Restaurants and Coffee Shops in the area. Pub and Theater is in the adjecent neighborhoods and comparitively New York neighborhoods are much closer compared to the Toronto.
- 13 venues that are missing in this location - Pub,Theater,Breakfast Spot,Shoe Store,Restaurant,Performing Arts Venue,Historic Site,Event Space,French Restaurant,Antique Shop,Beer Store,Bank,Asian Restaurant
- This location matches 7/10 most frequent venues in comparision to Toronto.
- It doesn't have Café,Theater or Breakfast Spot. Although Breakfast spot is absent, there are Cafés and Restaurants and Coffee Shops in the area. Theater is in adjecent neighborhoods.
- 13 venues that are missing in this location - Theater,Breakfast Spot,Café,Shoe Store,Performing Arts Venue,Mexican Restaurant,Historic Site,Event Space,Farmers Market,Antique Shop,Electronics Store,Chocolate Shop,Art Gallery

# Mapping of the matched locations using Folium

| | Borough | Neighborhood |
|---|---|---|
| 0 | Bronx | 52 |
| 1 | Brooklyn | 70 |
| 2 | Manhattan | 40 |
| 3 | Queens | 81 |
| 4 | Staten Island | 63 |

# Conclusion

In this resport, I analyzed the relationship between neighborhoods and their venue categories in differnt cities New York and Toronto. I identified the latitude and longitude from different data sources and used Foursquare to retrieve all venues in the locations, using this information, I could list the most frequent locations for each neighborhood. This gave some idea on neighborhood characteristics that will be helpful to choose a location in New York, given the person is currently residing in Toronto. The Matrix multiplication methodology used eliminiates dissimilar venues and multiplies similar categories giving it a comparitive advantage over other methodologies. For example, according to the example used above, if a person is currently living in HarbourFront, Downtown Toronto, the best place to choose is Murray Hill, Manhattan in New York since it matches most venue categories with HarbourFront, Downtown Toronto according to the graph above.

Thanks for watching and pls leave a good mark to let me pass. Thank you!