

Capstone Project-The Battle of Neighborhoods

(Choosing the best neighborhood for a Japanese customer)

2. Data Acquisition and Cleaning

2.1 Data sources

The following data sources were used to obtain the necessary information for the analysis:

- The information of Toronto neighborhoods and boroughs were obtained from Wikipedia https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
- The coordinates of Toronto neighborhoods were obtained using Geospatial dataset which can be obtained from http://cocl.us/Geospatial_data.
- Foursquare API was used to explore the available venues in each neighborhood in Toronto.

2.2 Data cleaning

Data was downloaded and scrapped from multiple sources to obtain the required information for the analysis and applying the clustering algorithm.

Beautiful Soap library was used for performing web scraping for Wikipedia site to obtain Toronto neighborhood and borough information. Data cleaning was done to convert the scrapped data into tabular DataFrame format. The neighborhoods having missing boroughs name were excluded from the analysis. The neighborhoods having the same postal code were combined together. Some neighborhoods were having missing values (unassigned), these neighborhoods were assigned the same name as the borough.

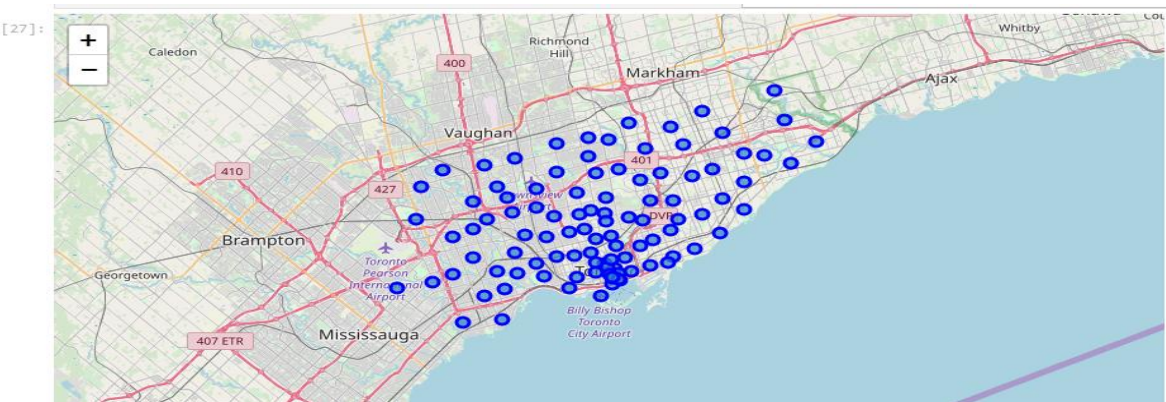
The below shows the Toronto borough and neighborhood information after cleaning

| | Postalcode | Borough | Neighbourhood |
|---|------------|------------------|----------------------------------|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Harbourfront, Regent Park |
| 3 | M6A | North York | Lawrence Heights, Lawrence Manor |
| 4 | M7A | Queen's Park | Queen's Park |

The neighborhood and borough information obtained from Wikipedia was combined with the coordinate information to be to explore the venues available within 2 KM radius in each neighborhood using Foursquare API.

| 0]: | Postalcode | Borough | Neighbourhood | Latitude | Longitude |
|-----|------------|------------------|----------------------------------|-----------|------------|
| 0 | M3A | North York | Parkwoods | 43.753259 | -79.329656 |
| 1 | M4A | North York | Victoria Village | 43.725882 | -79.315572 |
| 2 | M5A | Downtown Toronto | Harbourfront, Regent Park | 43.654260 | -79.360636 |
| 3 | M6A | North York | Lawrence Heights, Lawrence Manor | 43.718518 | -79.464763 |
| 4 | M7A | Queen's Park | Queen's Park | 43.662301 | -79.389494 |

103 neighborhoods are available in Toronto. These neighborhoods will be clustered using k-Mean clustering to obtain similar neighborhoods



The venues within 2 KM radius in each neighborhood are obtained using Foursquare API

[33]:

| | Neighbourhood | Neighbourhood Latitude | Neighbourhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|------------------|------------------------|-------------------------|------------------------|----------------|-----------------|----------------------|
| 0 | Parkwoods | 43.753259 | -79.329656 | Brookbanks Park | 43.751976 | -79.332140 | Park |
| 1 | Parkwoods | 43.753259 | -79.329656 | KFC | 43.754387 | -79.333021 | Fast Food Restaurant |
| 2 | Parkwoods | 43.753259 | -79.329656 | Variety Store | 43.751974 | -79.333114 | Food & Drink Shop |
| 3 | Victoria Village | 43.725882 | -79.315572 | Victoria Village Arena | 43.723481 | -79.315635 | Hockey Arena |
| 4 | Victoria Village | 43.725882 | -79.315572 | Tim Hortons | 43.725517 | -79.313103 | Coffee Shop |

The most 10 venue categories for each neighborhood are obtained to be able to cluster neighborhoods based on the similarities in the venue catogeries

| | Neighbourhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|-----------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 0 | Adelaide, King, Richmond | Coffee Shop | Café | Bar | Steakhouse | Thai Restaurant | Asian Restaurant | Burger Joint | Restaurant | Breakfast Spot | Hotel |
| 1 | Agincourt | Lounge | Furniture / Home Store | Breakfast Spot | Skating Rink | Doner Restaurant | Dim Sum Restaurant | Diner | Discount Store | Dive Bar | Dog Run |
| 2 | Agincourt North, L'Amoreaux East, Milliken, St... | Park | Playground | Dog Run | Department Store | Dessert Shop | Dim Sum Restaurant | Diner | Discount Store | Dive Bar | Doner Restaurant |
| 3 | Albion Gardens, Beaumont Heights, Humbergate, ... | Grocery Store | Fried Chicken Joint | Pharmacy | Coffee Shop | Pizza Place | Fast Food Restaurant | Sandwich Place | Beer Store | Dive Bar | Dessert Shop |
| 4 | Alderwood, Long Branch | Pizza Place | Coffee Shop | Skating Rink | Pharmacy | Pool | Pub | Sandwich Place | Gym | Airport Terminal | Doner Restaurant |

Cluster Neighborhoods