# Clustering Data Analysis of Toronto's Neighborhoods (Postal Code)

# Table of Content

# Introduction

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2.8 million as of 2016. The city is home to the Toronto Stock Exchange, the headquarters of Canada's five largest banks, and the headquarters of many large Canadian and multinational corporations. Its economy is highly diversified with strengths in technology, design, financial services, life sciences, education, arts, fashion, aerospace, environmental innovation, food services, and tourism.*

Toronto is a very popular hub for Startup companies and our goal in this analysis is to investigate potential postal code (neighborhoods) to open a new start up company in Toronto. This analysis will help us to select the best priced location to buy a property in Toronto considering the available amenities.

* https://en.wikipedia.org/wiki/Toronto

# Data acquisition and cleaning

- List of postal codes in Toronto and neighborhoods associated with them
  https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- List of Multiple Listing Service (MLS) districts in Toronto and neighborhoods associated with them
  https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Toronto
- Most recent MLS district housing pricing in Toronto (April 2020)
  http://trreb.ca/files/market-stats/home-price-index/TREB_MLS_HPI_Public_Tables_0420.pdf
- All data sources are scraped, cleaned and processed yielding a consolidated list of 101 postal codes, their associated neighborhoods, and their average housing price.

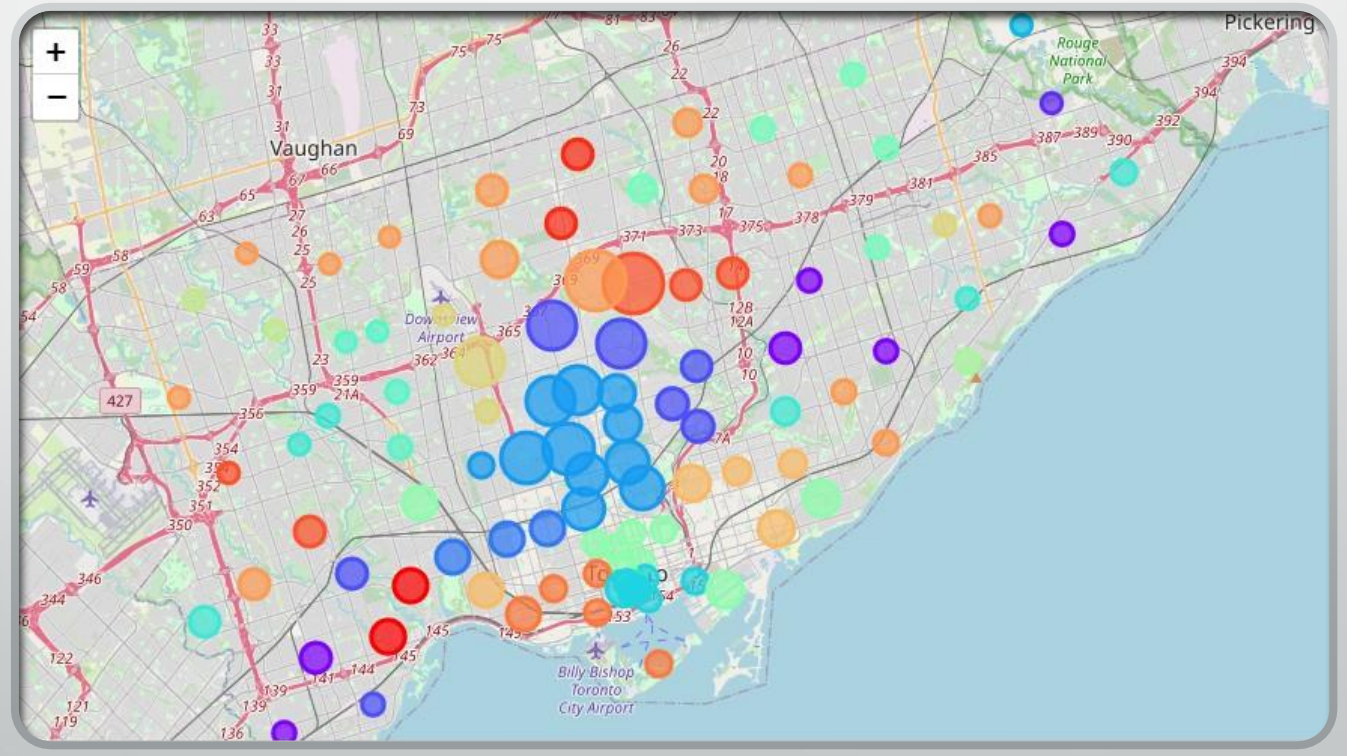
# Creation of Data Clusters

- The top 100 venues for each postal code within a radius of 2000 m of the postal code center is found using foursquare API calls.
- Mean frequency of existence of various venue categories is calculated per postal code. The top ten venue categories per each postal code are identified.

| PostalCode | 1st Most Common Venue Category | 2nd Most Common Venue Category | 3rd Most Common Venue Category | 4th Most Common Venue Category | 5th Most Common Venue Category | 6th Most Common Venue Category | 7th Most Common Venue Category | 8th Most Common Venue Category | 9th Most Common Venue Category | 10th Most Common Venue Category |
|---|---|---|---|---|---|---|---|---|---|---|
| M1B | Zoo Exhibit | Fast Food Restaurant | Restaurant | Pizza Place | Zoo | Gas Station | Bus Station | Caribbean Restaurant | Liquor Store | Supermarket |
| M1C | Coffee Shop | Grocery Store | Pharmacy | Pet Store | Trail | Hotel | Fast Food Restaurant | Sandwich Place | Bank | Breakfast Spot |
| M1E | Pizza Place | Coffee Shop | Restaurant | Hotel | Grocery Store | Bank | Fast Food Restaurant | Park | Gymnastics Gym | Breakfast Spot |
| M1G | Coffee Shop | Fast Food Restaurant | Bank | Sandwich Place | Pizza Place | Discount Store | Supermarket | Beer Store | Indian Restaurant | Chinese Restaurant |
| M1H | Coffee Shop | Clothing Store | Bank | Sandwich Place | Fast Food Restaurant | Gas Station | Pharmacy | Indian Restaurant | Restaurant | Bookstore |

# Creation of Data Clusters (Cont'd)

- K-means clustering technique is used to split 101 Toronto postal codes into 20 different clusters based on their access level to various venue categories.
- The result was plotted using  Folium API calls and the average price calculated for each postal code was used to specify the size of markers on the map.



# Conclusion

The results in indicate that:

- There many areas in the city of Toronto with access to similar amenities as downtown area.
- Midtown area housing price is higher than both downtown and uptown areas.
- For start up company depending on their nature of work and venue categories of interest, the may consider exploring either downtown or uptown neighborhood.