Capstone Project

The Battle of Neighborhoods

Finding the most suitable neighborhood in Totronto

1. Introduction

- The purpose of this project is to suggest a more suitable neighborhood in a new city for a new immigrant or a visitor for a staying over months, in terms of living costs, social presence, traffic connectivity.
- Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571, it is the most populous city in Canada and the fourth most populous city in North America. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario, while the Greater Toronto Area (GTA) proper had a 2016 population of 6,417,516. Toronto is an international center of business, finance, arts, and culture and is recognized as one of the most multicultural and cosmopolitan cities globally.

2. Data description

- **Postal codes of Toronto**
  - Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - This data source contains the borough, neighbourhoods in terms of postal codes.
- **Location coordinates with respect to postal codes of Toronto**
  - Data Link: http://cocl.us/Geospatial_data
  - This data source contains the locations of boroughs/neighbourhoods (latitude and longitude) in terms of postal codes.
- **Foursquare API**
  - The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes.
- **Housepricehub (web scrapping)**
  - Data Link: https://housepricehub.com/cities/city/Toronto
  - This data source is added for fetching the average house prices in city Toronto.

2. Data description

- **Data specifications:**
  - Neighborhood
  - Neighborhood Latitude
  - Neighborhood Longitude
  - Venue
  - Name of the venue e.g. the name of a store or restaurant
  - Venue Latitude
  - Venue Longitude
  - Venue Category
  - House price
  - Numbers of schools

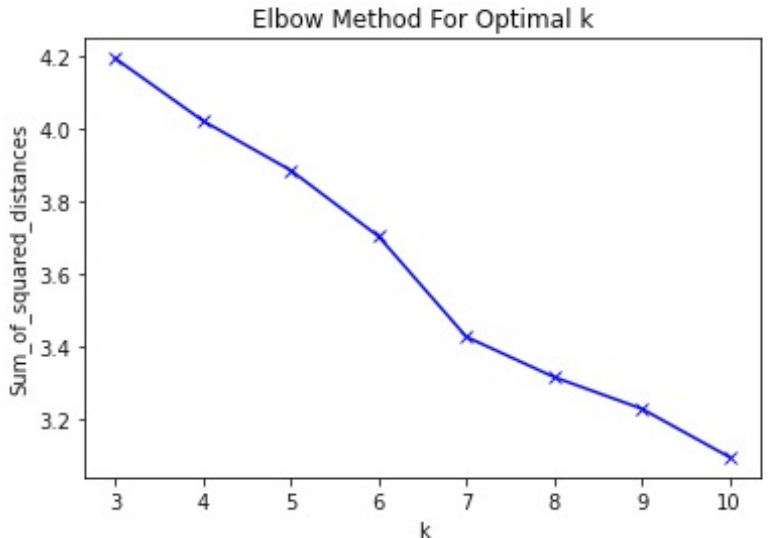
3. Methodology

- Applying K-Means Clustering Approach with k optimization
  - Best K = 7

Appling K-means on Toronto_grouped

```
In [109]: df_clustering = Toronto_grouped.drop('Neighbourhood', 1)

In [110]: df_clustering.shape
Out[110]: (93, 333)
```



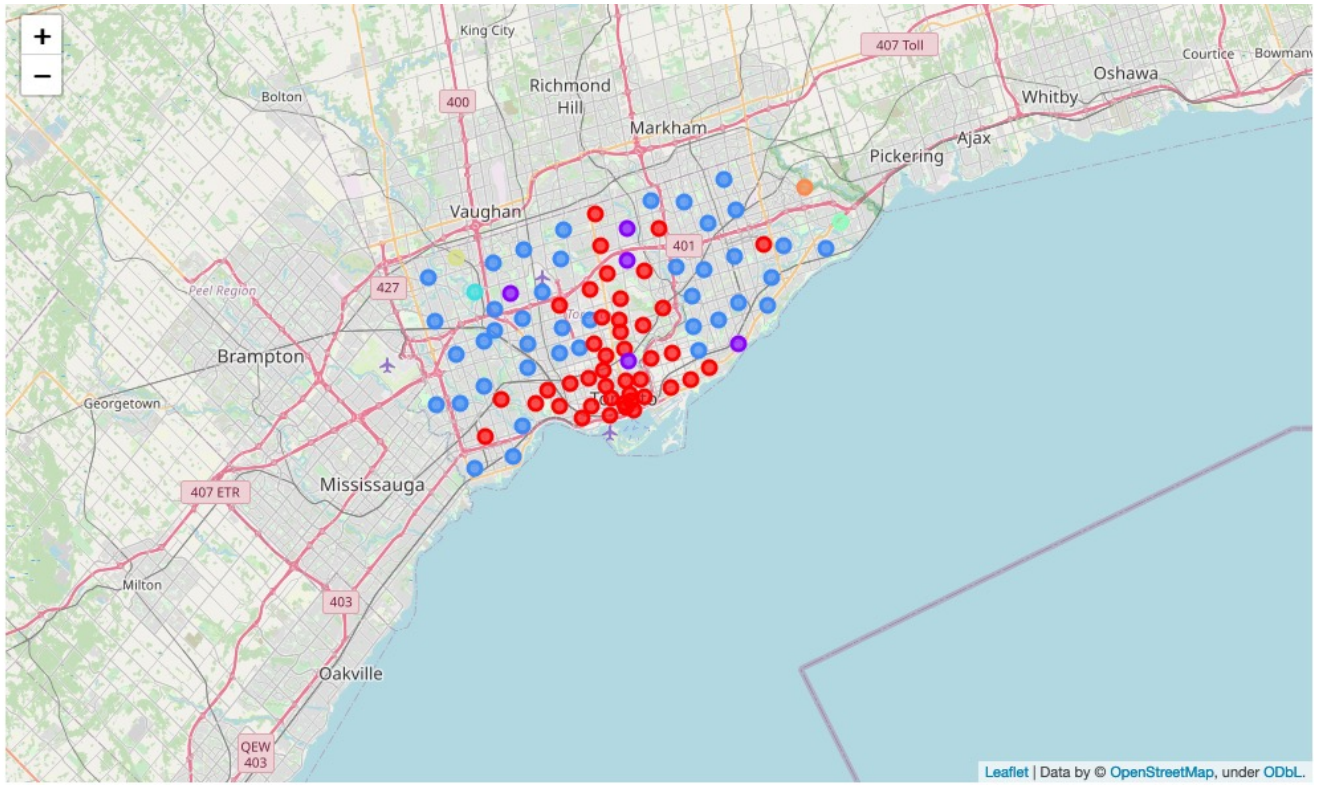
3. Methodology

- Most Common venues in the neighborhoods:

	Cluster Labels	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	2	Agincourt	Chinese Restaurant	Shopping Mall	Sandwich Place	Pool	Seafood Restaurant	Sushi Restaurant	Supermarket	Latin American Restaurant	Lounge	Breakfast Spot
1	2	Alderwood, Long Branch	Pizza Place	Coffee Shop	Bank	Pharmacy	Park	Sandwich Place	Beer Store	Bar	Seafood Restaurant	Liquor Store
2	2	Bathurst Manor, Wilson Heights, Downsview North	Pizza Place	Gas Station	Bank	Coffee Shop	Park	Bridal Shop	Trail	Supermarket	Sandwich Place	Sushi Restaurant
3	1	Bayview Village	Park	Café	Japanese Restaurant	Bank	Trail	Chinese Restaurant	Fish & Chips Shop	Financial or Legal Service	Elementary School	Food
4	0	Bedford Park, Lawrence Manor East	Italian Restaurant	Coffee Shop	Pet Store	Juice Bar	Sandwich Place	Bridal Shop	Pharmacy	Liquor Store	Sports Club	Intersection

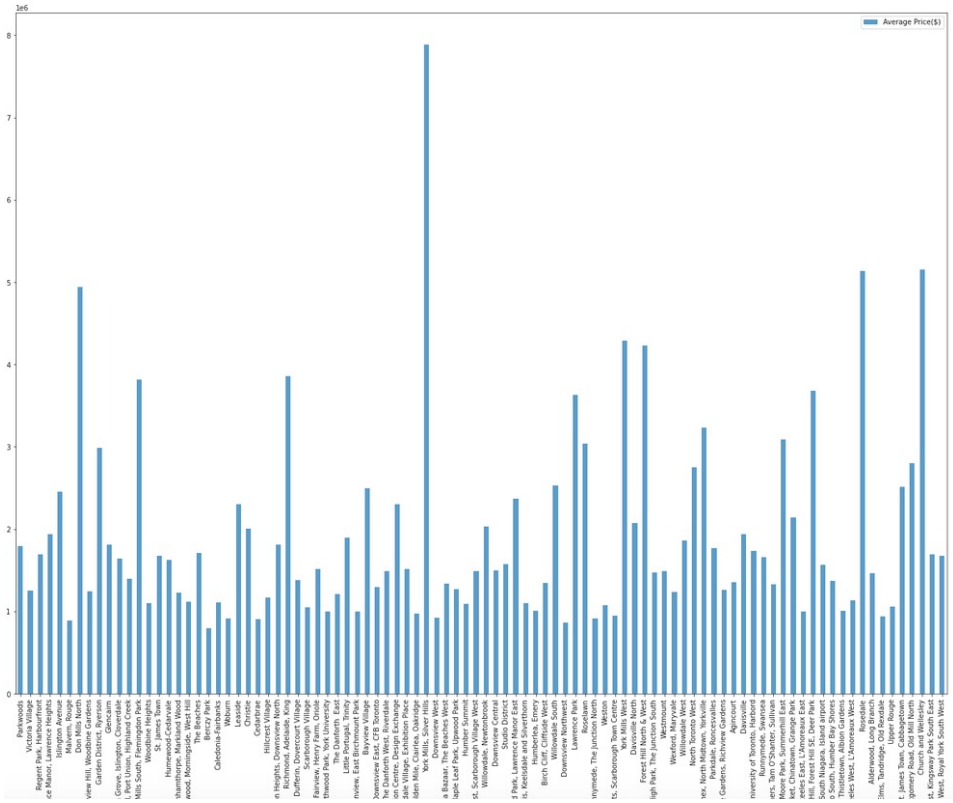
4. Results

- Neighborhoods clustering in map visualization



4. Results

- Average Housing Price for each neighborhood in Toronto



5. Conclusion

- The overview of the categories of neighborhoods in city Toronto underneath,
  - Cluster 0: downtown with high living cost or crowded space. A reversed 'T' formed region can be obviously discovered. The price tier of this region is about 1M-3M. This cluster should be the best choices of people working in the downtown and can bear the small space with affordable price.
  - Cluster 1: Luxury areas. The price tier can approach to '>5M'. This cluster is suitable for rich people.
  - Cluster 2: Mid-high living cost. it is teared apart by Cluster 0 from the middle. This cluster is the best for the mid-class.
  - Cluster 3-7: neighbours far away from the center with different features.