

Toronto Area Clustering using K-Mean

1. Introduction

- a. Toronto is a big city and Canada's biggest city. Toronto is considered one of the most culturally diverse cities in the world, with more than half the population born outside of the city I want to explore and extract the similarity and dissimilarity of each area (Postal Code) in Toronto that can help tourists who want to visit Toronto by providing insight.

2. Data

- a. Toronto Postal Code
 - i. this data provides information about Borough, Neighborhood and Postal Code
 - ii. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- b. Toronto Area Latitude & Longitude
 - i. this data provides information about Postal Code and latitudes and longitudes of each area
 - ii. https://cocl.us/Geospatial_data
- c. Foursquare Venue
 - i. this data provides information about venues locate in each area of Toronto, including Venue, Venue Latitude, Venue Longitude and Venue Category
 - ii. <https://api.foursquare.com/v2/venues>

3. Methodology

- a. Collect Toronto Data

	Postal Code	Borough	Neighborhood	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	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

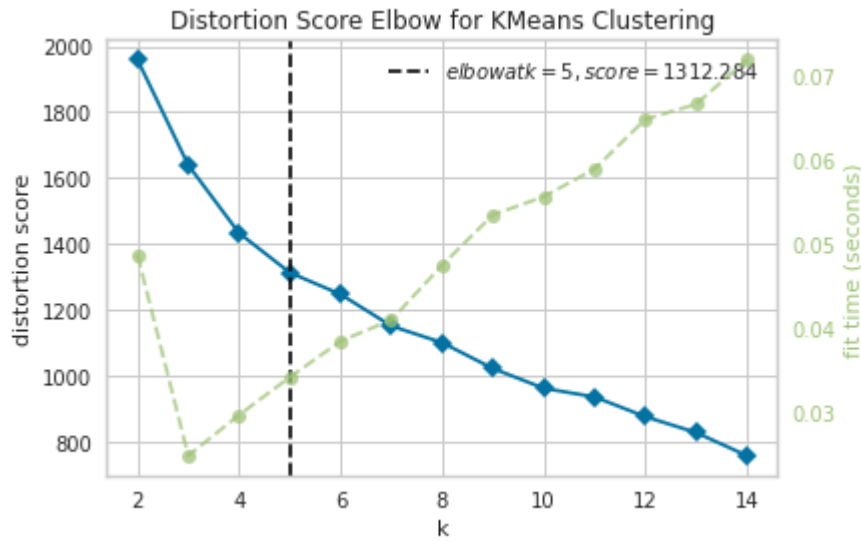
b. Using Foursquare API find all nearby venue within 500 meters

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M3A	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	M3A	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	M3A	43.753259	-79.329656	Corrosion Service Company Limited	43.752432	-79.334661	Construction & Landscaping
3	M4A	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	M4A	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

c. Filter noise venue

	Venue Category	Count
0	Coffee Shop	174
1	Café	98
2	Restaurant	71
3	Park	51
4	Pizza Place	49
5	Italian Restaurant	45
6	Bakery	45
7	Sandwich Place	42
8	Hotel	41
9	Japanese Restaurant	39
10	Clothing Store	34
11	Gym	33
12	Sushi Restaurant	30
13	Grocery Store	29
14	Bar	29
15	American Restaurant	26

d. Find the optimum value of K, we use Elbow technique



4. Analysis

a. Cluster 0

Venue Category	Label	American Restaurant	Art Gallery	Asian Restaurant	Bakery	Bank	Bar	Beer Bar	Beer Store	Bookstore	Breakfast Spot	Brewery	Bubble Tea Shop	Burger Joint	Burrito Place	Café
District																
M2J	0	1.0	0.0	1.0	2.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0
M5B	0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	2.0	0.0	0.0	3.0	1.0	1.0	3.0

b. Cluster 1

[illegible]

5. Result and Discussion

- a. The analysis results are from analyzing the clustering result from K-means ($k=5$ which is picked from using the elbow technique). Then, the result will provide outstanding attributes of each cluster

6. Conclusion

This project point to solve the problem of tourists who want to plan their visit to Toronto, by using results of cluster analysis to help as assistance in making decisions on planning their vacation. Then, the result will provide outstanding attributes of each cluster