

# Clustering of Neighbourhoods in Toronto to Aid Selection of Residence

---

Oladayo K. Ogunnoiki

5th May 2020

COGNITIVE CLASS.AI Coursera

1. Introduction
2. Data acquisition and exploration
3. Methodology
4. Results
5. Conclusion

# Introduction

---

**Problem statement:** How can an individual find their next ideal neighbourhood of residence

**Objective:** Segment and cluster all neighbourhoods in Toronto based on census and non-census data

## Data acquisition and exploration

---

The data for this project is grouped into two classes:

- Census data. For example, population size, population density etc.
- Non-Census data. For example, bus stations, cafes etc.

# Census data

The census data was retrieved from the City of Toronto website.  
Below is sample of the census data:

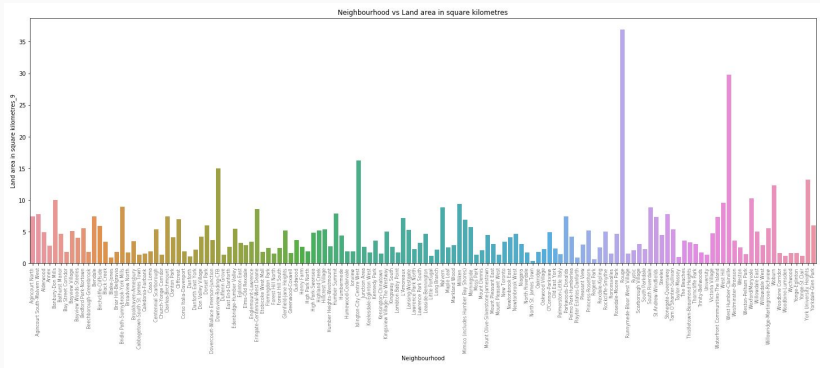
_id	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderswood	Annex --	Willowdale West	Wilmerville-Morningside-Richview	Woburn	Woodbine Corridor	Woodbine-Lumsden	Wychwood	Yonge-Eglinton	Yonge-St. Clair	York University Heights	Yorkdale-Glen Park
0	1	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129	128	20	95 --	37	7	137	64	60	94	100	97	27	31
1	2	Neighbourhood Information	City of Toronto	TSN82020 Designation	NaN	No Designation	No Designation	No Designation	No Designation	No Designation	No Designation	N/A	No Designation	No Designation	No Designation	No Designation	No Designation	N/A	Emerging Neighbourhood
2	3	Population	Census Profile 99-316-X2016001	Population, 2016	2,731,571	29,113	23,797	12,054	30,526 --	16,936	22,156	53,495	12,541	7,965	14,349	11,817	12,528	27,593	14,804
3	4	Population	Census Profile 99-316-X2016001	Population, 2011	2,615,000	30,279	21,988	11,954	29,177 --	15,004	21,343	53,300	11,703	7,826	13,986	10,578	11,652	27,713	14,687
4	5	Population	Census Profile 99-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60% --	12.90%	3.80%	0.30%	7.20%	0.50%	2.80%	11.70%	7.50%	-0.40%	0.80%

5 rows x 145 columns



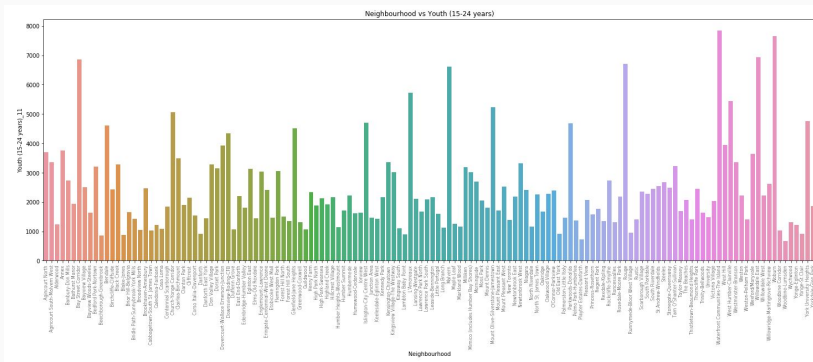
## Census data

Below is a plot of the Neighbourhood vs the Neighbourhood Land Area:



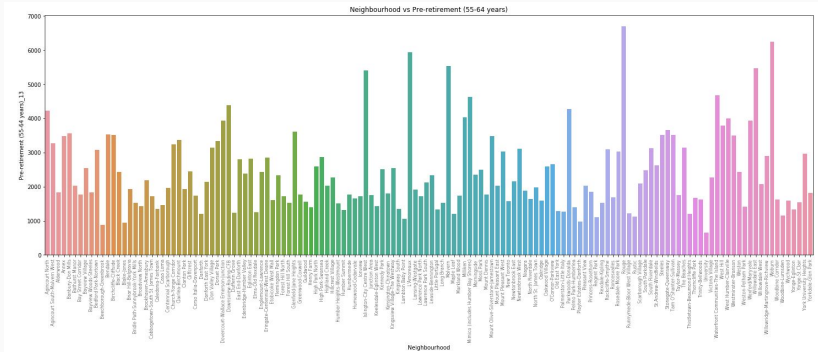
## Census data

Below is a plot of the Neighbourhood vs the population of Youths (15-24years):



## Census data

Below is a plot of the Neighbourhood vs the population of Pre-retirement residents (55-64years):



# Non-Census data

The Non-census data was retrieved leveraging the Foursquare and Google-Maps API. Below is a sample of the Non-census data:

	Neighbourhood	Alghan Restaurant	Airport Service	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Video Store	Vietnamese Restaurant	Volleyball Court	Warehouse Store	Whisky Bar	Wine Bar	Wine Shop	Wings Joint	Women's Store	Yoga Studio
0	Agincourt North	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Agincourt North	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Agincourt North	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Agincourt North	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Agincourt North	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 269 columns

# Merged data

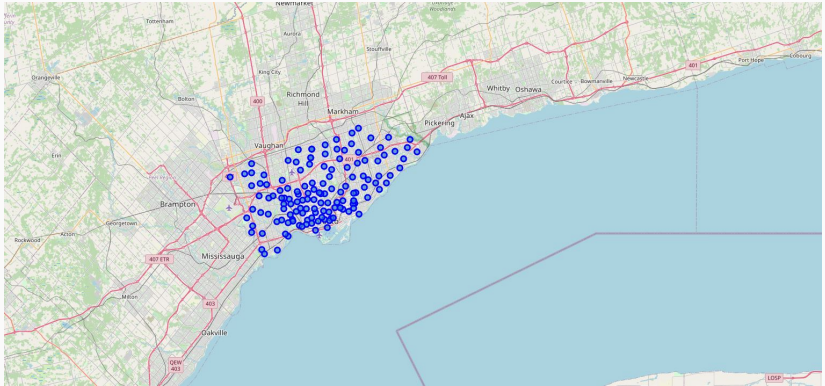
Leveraging Pandas concat function, the census and non-census data are merged. Below is a sample of the merged data:

	Neighborhood	Alghan Restaurant	Airport Service	American Restaurant	Antique Shop	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	External migrants_2374	Total_Mobility status 5 years ago - 25% sample data_2375	Non-movers_2376	Movers_2377	Non-migrants_2378	Migrants_2379	Internal migrants_2380	Intraprovincial migrants_2381	Interprovincial migrants_2382	External migrants_2383
0	Agriculture North	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.191104	0.373337	0.557859	0.149208	0.201777	0.109482	0.040482	0.048436	0.028708	0.213190
1	Agriculture South-Northern West	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.153213	0.263503	0.371889	0.152969	0.210152	0.106431	0.048547	0.048473	0.049043	0.201943
2	Alderswood	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.014827	0.002965	0.185068	0.024279	0.037056	0.016904	0.028293	0.036660	0.013158	0.005112
3	Annex	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.266995	0.377250	0.351385	0.288841	0.348731	0.242218	0.220042	0.200815	0.309809	0.231595
4	Barbury-Don Mills	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.116969	0.346117	0.487753	0.172347	0.254315	0.106235	0.072406	0.083503	0.049043	0.157484

5 rows x 2594 columns

# Merged data

Below is map highlighting all the neighbourhoods in Toronto:



# Methodology

---

The problem statement is a unsupervised machine learning problem. It requires the use of clustering algorithms. Two clustering algorithms are used in this project:

- K-means clustering
- HDBSCAN



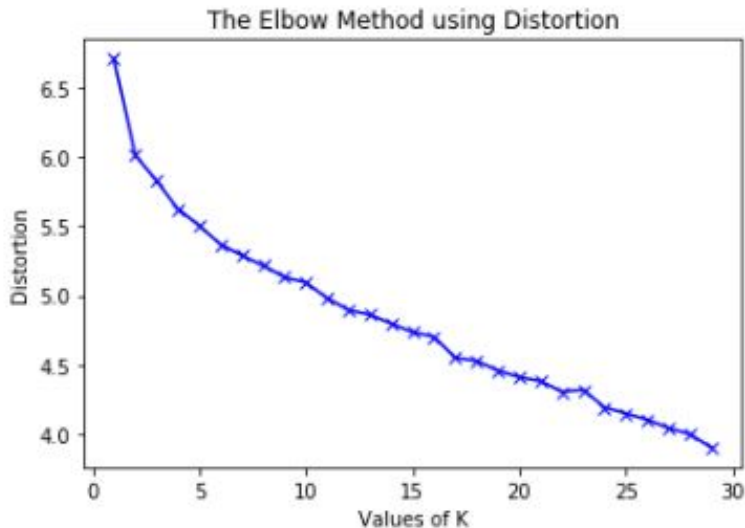
K-means clustering data assumption:

- Round and spherical clusters
- Equally sized and dense clusters
- Clusters with high density at the center
- Absence of noise and outliers in the data

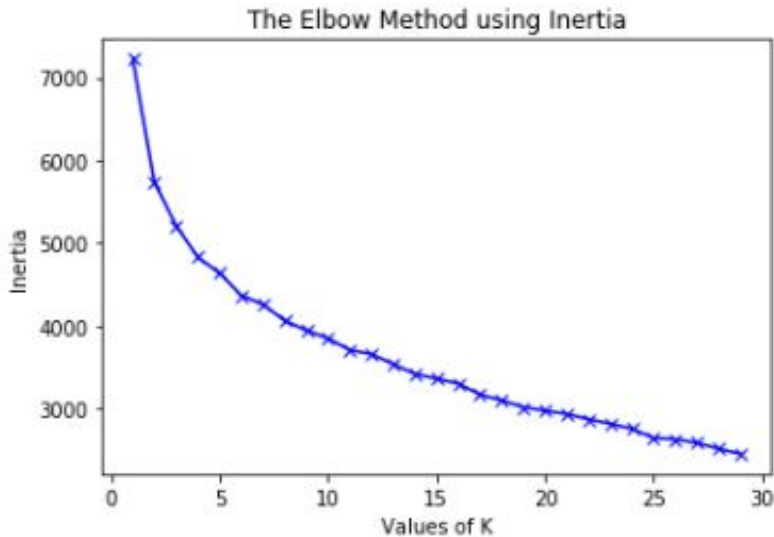
## Results

---

## K-Means: Elbow Method (Distortion)

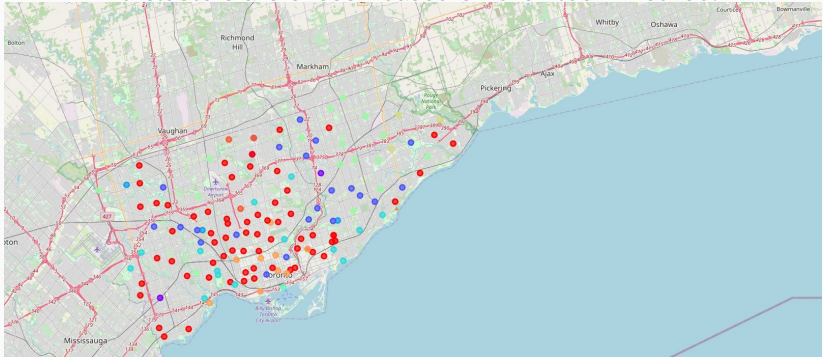


## K-Means: Elbow Method (Inertia)



# K-Means: Map

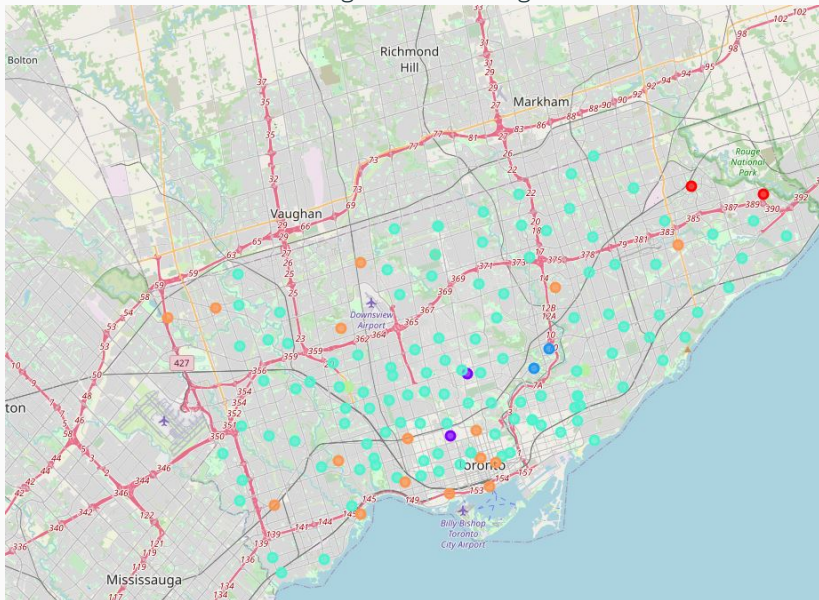
10 Clusters are chosen based on the Elbow method.



The K-Means assumptions are not met.  
It's not clear the optimal number of clusters  
from the elbow method.

# HDBSCAN: Map

6 Clusters are generated using HDBSCAN.



# Sample Neighbourhoods in their clusters

Neighbourhood	Latitude	Longitude	Population, 2016_3	Population, 2011_4	Population Change 2011-2016_5	Total private dwellings_6	Private dwellings occupied by usual residents_7	Population density per square kilometre_8	Land area in square kilometres_9	...	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	HDB Cluster Labels
102 Roncesvalles	43.646317	-79.449058	0.141516	0.182707	0.110128	0.101524	0.111190	0.203577	0.030162	...	Grocery Store	Bakery	Food & Drink Shop	Eastern European Restaurant	Café	Sushi Restaurant	Bookstore	The Restaurant	American Restaurant	3
21 Casa Loma	43.676843	-79.410363	0.074002	0.085336	0.204898	0.066381	0.072325	0.107276	0.041404	...	History Museum	Café	Coffee Shop	Castle	Burger Joint	Donut Shop	Steakhouse	Jewish Restaurant	Middle Eastern Restaurant	3
35 East End-Danforth	43.678162	-79.309332	0.249494	0.306240	0.171141	0.151735	0.171364	0.161088	0.061420	...	Asian Restaurant	Café	Gas Station	Flower Shop	Hungarian Restaurant	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	3
54 Humber Summit	43.760100	-79.571785	0.098406	0.128825	0.112416	0.033594	0.032725	0.012246	0.205374	...	Park	Gym	Dive Bar	Doctor's Office	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Yoga Studio	3
23 Church-Yonge Corridor	43.672650	-79.387839	0.417335	0.486407	0.305369	0.432013	0.446911	0.508399	0.025775	...	Spa	Café	Sushi Restaurant	Japanese Restaurant	Boutique	French Restaurant	Hotel	Women's Store	Coffee Shop	-1
6 Bay Street Corridor	43.657290	-79.384364	0.323918	0.274423	0.696242	0.352149	0.326038	0.301680	0.038652	...	Middle Eastern Restaurant	Italian Restaurant	Sandwich Place	Bubble Tea Shop	Hotel	Burger Joint	Diner	Restaurant	Clothing Store	-1
125 Weston	43.700167	-79.516264	0.192379	0.249285	0.110738	0.118546	0.129901	0.142256	0.057033	...	Coffee Shop	Pharmacy	Discount Store	Diner	Bank	Sandwich Place	Social Field	Middle Eastern Restaurant	Fried Chicken Joint	3
20 Danforth	43.686552	-79.307341	0.052059	0.063079	0.167785	0.029541	0.033459	0.173610	0.019468	...	The Restaurant	Breakfast Spot	Coffee Shop	Middle Eastern Restaurant	Gaming Cafe	Thrift / Vintage Store	Gas Station	Sushi Restaurant	Music Store	3
100 Rexdale-Kipling	43.719857	-79.070600	0.066004	0.085307	0.134228	0.026802	0.031386	0.073681	0.005759	...	Pizza Place	Sandwich Place	Auto Workshop	Department Store	Bakery	Donut Shop	Doctor's Office	Dog Run	Doner Restaurant	3
16 Bixie Path-Sunnybrook-York Mills	43.735914	-79.371899	0.045318	0.047480	0.233221	0.015356	0.015500	0.000000	0.232794	...	Yoga Studio	Distribution Center	Faerie Restaurant	Event Space	Ethiopian Restaurant	Elementary School	Electronics Store	Egyptian Restaurant	Eastern European Restaurant	3
17 Broadview North	43.688529	-79.353278	0.082951	0.106297	0.117450	0.067908	0.079069	0.132252	0.035087	...	Pizza Place	Bakery	Flower Shop	Farm Store	Sandwich Place	Bank	Discount Store	Pharmacy	Cheeserie	3
118 Trinity-Bellwoods	43.650096	-79.417073	0.168178	0.220093	0.102349	0.107311	0.111793	0.197084	0.035920	...	Cocktail Bar	Bar	Park	Wine Bar	Ice Cream Shop	Clothing Store	Spa	Boutique	Breakfast Spot	3
74 Maple Leaf	43.714802	-79.479429	0.059559	0.079147	0.114064	0.020264	0.023776	0.088668	0.057562	...	Hobby Shop	Construction & Landscaping	Convenience Store	Electronics Store	Doner Restaurant	Donut Shop	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	3
45 Glenfield-Jane Heights	43.706822	-79.304340	0.403027	0.531300	0.078859	0.171729	0.190705	0.111458	0.131067	...	Coffee Shop	Yoga Studio	Eastern European Restaurant	Doctor's Office	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant	Egyptian Restaurant	4



## Conclusion

---

- HDBSCAN is the suitable clustering algorithm for this project.
- The neighbourhoods are grouped into 6 clusters.

The End  
Thank you!