

A scenic view of the Toronto skyline across the water, featuring the CN Tower and several high-rise buildings under a cloudy sky. The water in the foreground is dark blue with gentle ripples.

Exploring Neighborhood and Built Environment as Social Determinants of Health in the City of Toronto

In this project we worked together with the Toronto Public Health division to research the effect of neighborhood and built environment as social determinants of health. In order to do so we used data science to determine if there are any significant differences between the neighborhoods, and if so, generate a cluster of neighborhoods that are in need of more help so that the TPH division can include any convenient measures to their 2020-2024 Strategic Plan. Their main objective is to address the unique needs of their community by implementing public policy and practices that enhance the health of individuals, communities and the entire city.

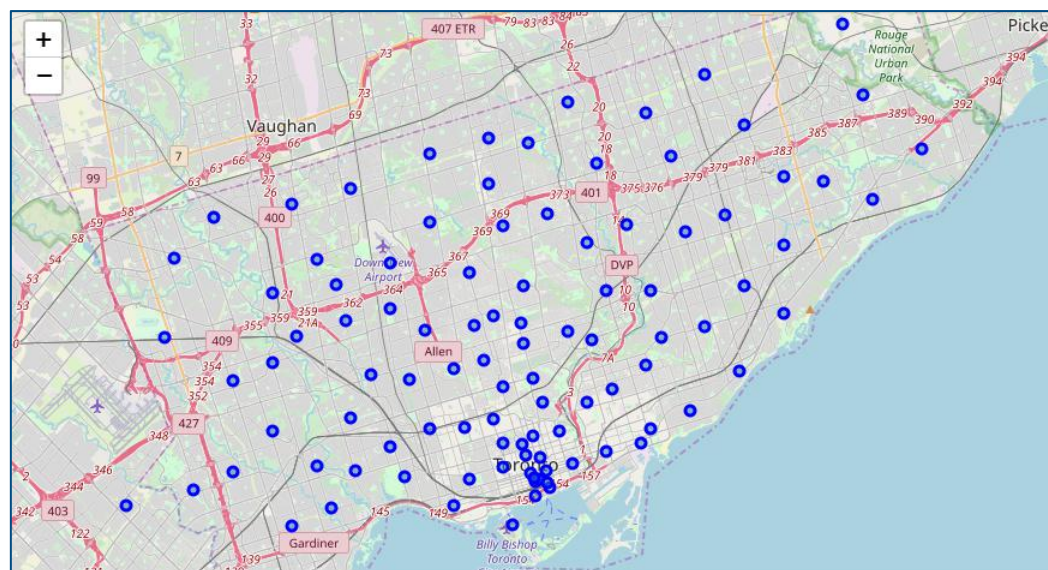
In order to do so, we research some variables that were related to neighborhood and built environment and could potentially have an effect on health. These were:

- Economic Status
- Access to foods that support healthy eating patterns
- Access to parks, fitness or recreation centers
- Crime and Violence
- Environmental Conditions
- Access to Health Care
- Access to Education

The data was extracted/scraped from the following sources:

- Data of neighborhoods and postal codes:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Location data for the neighborhoods: http://cocl.us/Geospatial_data
- Location data for the venues: Foursquare API
- Income data: Canadian Census (<https://www.statcan.gc.ca/eng/start>)
- Crime rates: Tonto Police Service
(<https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates>)
- Air Quality Health Index (AQHI) data:
http://www.airqualityontario.com/science/data_sets.php

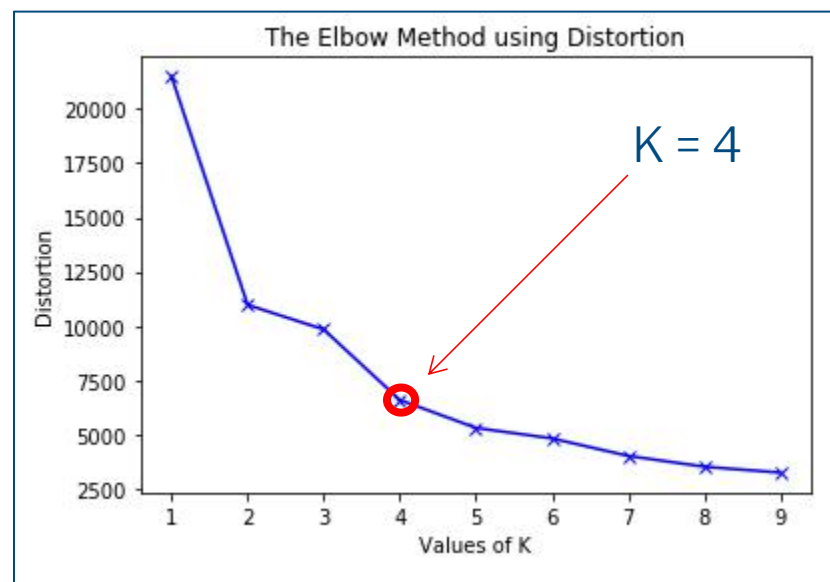
1. Uploaded and cleaned the data including Toronto's neighborhoods, their postcode and what borough they belonged to.
2. Added location data (Lat/Lon) and visualized it in the map



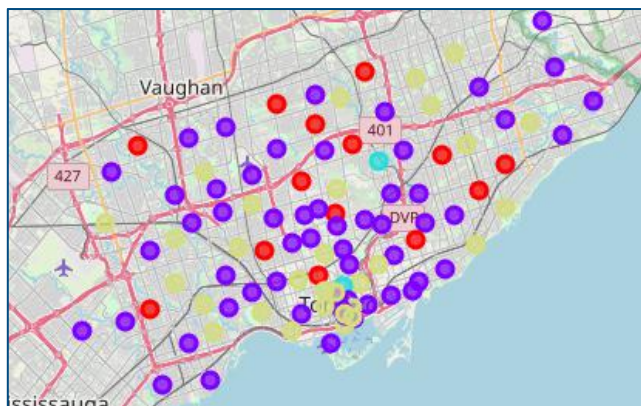
3. Uploaded and added income, crime and air quality data
4. After close examination we dropped the air quality column and grouped the rates of the different crimes under a total
5. Finally, we added the location data of the different venues for each neighborhood by using the Foursquare API

	Borough	Neighbourhood	Postcode	Latitude	Longitude	Avg Income	Total Crime	Healthy Food Places	Un-healthy Food Places	Excercise & Outdoor Places	Health Services	Education Buildings
0	Scarborough	Rouge, Malvern	M1B	43.806686	-79.194353	25750	5487.6	0	0	6	0	1
1	Scarborough	Highland Creek, Rouge Hill, Port Union	M1C	43.784535	-79.160497	35239	628.3	4	0	4	1	0
2	Scarborough	Guildwood, Morningside, West Hill	M1E	43.763573	-79.188711	19687	3104.9	5	1	2	3	0

We performed **k-means clustering** to group the neighborhoods by their similarity for the variables previously mentioned. But before we performed the elbow technique to compute the optimal value of K.



Results



- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4

	Avg Income	Total Crime	Healthy Food Places	Unhealthy Food Places	Exercise & Outdoor Places	Health Services	Education Buildings
Cluster 1	28,264	3,870	2.8	2.5	2.2	1.43	1.2
Cluster 2	98,553	2,066	3.7	1.3	3.9	2.16	2.1
Cluster 3	184,383	1,454	5.3	1.0	4.6	3.28	2.7
Cluster 4	46,953	3,240	3.5	1.7	3.1	1.86	1.9

- Checked that certain factors like income, crime rate and access to various services have a direct impact on health and vary from one neighborhood to another
- Built useful model to determine which neighborhoods in Toronto are in greater need of help by the government (**Cluster 1**)
- Proposed possible explanations to the inequalities between the areas and recommended solutions that could be included in the strategic plan of the Toronto Public Health for the period 2020-2024