

Toronto Neighborhood Visualization Maps

Tool for Potential Movers

Applied Data Science Capstone (Coursera)

Patrick de Guzman

September 2, 2019

Table of Contents

1. Introduction & Problem Statement

a. Problem Statement

b. Target Audience

2. Data Sources

a. Types of Data Required

b. Data Sources, Quality, & Cleaning Requirements

3. Methodology

a. Dimensions of Consideration

b. Choropleth Mapping

c. K-means Clustering

4. Results & Observations

5. Conclusion

6. References

7. Appendix

i. Age Distribution Map

ii. Income Distribution Map

iii. Average Household Size

iv. Average Unit Size

v. Neighborhood Character Cluster Listing (Cluster 0 – 11)

vi. Neighborhood Character Cluster Listing (Cluster 12 – 19)

1. Introduction & Problem Statement

The following report was prepared as a submission for the Applied Data Science Capstone course offered by IBM through Coursera.

Using data from the Foursquare API, neighborhoods in Toronto, Ontario were clustered based on venue-types and plotted against choropleth maps visualizing other dimensions of consideration for potential movers to the city.

a. Problem Statement

The motivation behind this project was to provide a tool for potential new-movers to visualize Toronto's neighborhood and reduce the burden and difficulty of conducting neighborhood research when searching for a new home.

For potential new-movers, the task of conducting research can be daunting and time-consuming. However, this task is critical to long-term satisfaction in a new resident's place of stay, especially in a diverse city like Toronto, Ontario where neighborhoods can differ wildly and appeal to very different demographics.

b. Target Audience

The main target audience for these visualizations are primarily individuals actively planning a move and searching for an apartment in Toronto, Ontario, but also extend to a secondary target audience of individuals contemplating a move, but unsure of whether their personality and interests align with what Toronto's neighborhoods have to offer.

2. Data Sources

a. Types of Data Required

To create maps using Python's 'folium' package, we require the following data:

- A listing of Toronto's main neighborhoods (segregated by postal code)
- A listing of coordinates for each of Toronto's neighborhoods
- GeoJSON data to identify the folium choropleth map boundaries
- Venue type data per each neighborhood
- City of Toronto data corresponding to statistics on: age, income, household sizes, and unit sizes

b. Data Sources, Quality, & Cleaning Requirements

Toronto Postal Codes & Neighborhoods

The listing of Toronto boroughs, neighborhoods, and postal codes was obtained from a Wikipedia listing of Canadian postal codes beginning with ‘M’ [1]. This data was scraped from the Wikipedia site using the ‘BeautifulSoup’ package in Python.

Additional cleaning was required as many postal codes contained neighborhoods labelled as ‘Not Assigned’. In these cases, the missing neighborhood values were simply replaced with the borough name. In addition, duplicate postal codes were found as some postal codes were tagged to several neighborhoods. To develop a set of unique postal codes, the neighborhoods of duplicate codes were combined as a list per postal code.

A sample of the tidy postal code data is shown below:

	PostalCode	Borough	Neighborhoods
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 1: Head of Toronto Postal Code set scraped from Wikipedia

Toronto Neighborhood Coordinates

Neighbourhood coordinates were obtained directly from Coursera through the link noted in [5]. This data corresponds directly to the Toronto postal code set created and will be used in querying the Foursquare API for venues.

A sample of the final neighborhood dataset (with coordinates) is shown below:

	PostalCode	Borough	Neighborhoods	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 2: Head of Toronto Postal Code set with accompanying coordinates

GeoJSON of Neighborhood Boundaries

A GeoJSON file of Toronto's neighborhood boundaries was found through an open-source Github [2]. These boundaries were developed using the Neighborhood Planning Areas data from the City of Toronto website.

These coordinates are used in preparing the choropleth maps to identify neighborhood boundaries.

The format of the first entry within the GeoJSON file is shown below:

```
{'type': 'Feature',
  'properties': {'DAUID': '35200879',
    'PRUID': '35',
    'CSDUID': '3520005',
    'HOODNUM': 81,
    'HOOD': 'Trinity-Bellwoods',
    'FULLHOOD': 'Trinity-Bellwoods (81)'},
  'geometry': {'type': 'Polygon',
    'coordinates': [[[-79.40428280044927, 43.64797961606815],
      [-79.403956753622, 43.64718271074494],
      [-79.42236786578222, 43.643467621011894],
      [-79.42640543946513, 43.65360764326518],
      [-79.41868792113178, 43.65521730993704],
      [-79.41769878521191, 43.65524323486715],
      [-79.41514736685951, 43.65496322517198],
      [-79.40767889826175, 43.65646442447146],
      [-79.40428280044927, 43.64797961606815]]]]}
```

Figure 3: Sample entry from GeoJSON of Toronto neighborhood coordinate boundaries

Venue Type Data per Neighborhood

The Foursquare API was scraped to identify the venue types surrounding the central coordinates of each Toronto neighborhood [4].

The query included the latitude and longitude of each Toronto neighborhood, returning a limit of 100 results per neighborhood based on a 250-meter radius.

However, slight caution should be exercised when formulating conclusions on the character of neighborhood clusters as some neighborhoods returned very few venue type data points from the Foursquare API request.

City of Toronto Statistics: Age, Income, Household Sizes, Unit Sizes

Data from the City of Toronto website was collected pertaining to age, income, household sizes, and unit sizes across neighborhoods [3].

For mapping, each key dimension of data was extracted as a subset and averaged per neighborhood to allow for viewing of distributions across Toronto in choropleth-map format.

It is noted that the data was sourced by the City of Toronto from the 2016 census, indicating potential deviance from current statistics (2019).

Samples of the subsets are shown below:

	hood	neighbourhood_number	children_0-14_years	youth_15-24_years	working_age_25-54_years
	Agincourt North	129	26880	72247.5	446547.5
	Agincourt South-Malvern West	128	21525	65520.0	393617.5
	Alderwood	20	12320	24082.5	206190.0
	Annex	95	16520	73125.0	594080.0
	Banbury-Don Mills	42	25235	53235.0	426995.0

Figure 4: Sample from subset of age data

	hood	neighbourhood_number	1_person	2_persons	3_persons	4_persons
0	Agincourt North	129	1350	2370	1995	1750
1	Agincourt South-Malvern West	128	1610	2325	1680	1335
2	Alderwood	20	1105	1440	885	795
3	Annex	95	7885	5220	1540	885
4	Banbury-Don Mills	42	4360	3820	1755	1515

Figure 5: Sample from subset of household size data

	hood	neighbourhood_number	under_\$10,000_including_loss	10,000_to_19,999
	Agincourt North	129	26122387.5	97121762.5
	Agincourt South-Malvern West	128	22947705.0	69597680.0
	Alderwood	20	6974302.5	23474217.5
	Annex	95	19973002.5	56848105.0
	Banbury-Don Mills	42	16598340.0	48973367.5

Figure 6: Sample from subset of income data

hood	neighbourhood_number	no_bedrooms	1_bedroom	2_bedrooms
Agincourt North	129	2.000000e-12	965	3580
Agincourt South-Malvern West	128	1.500000e-12	915	4470
Alderwood	20	2.000000e-12	325	2290
Annex	95	8.550000e-11	6995	9110
Banbury-Don Mills	42	5.000000e-12	3010	8490

Figure 7: Sample from subset of unit size data

3. Methodology

In order to accomplish the task of aiding the research process of potential new-movers, key dimensions of consideration need to be identified.

a. Dimensions of Consideration

For new-movers, dimensions of consideration are likely to include the following:

Income distributions (to gain an understanding of where residents of similar income classes can afford to live),

Age distributions (to gain an understanding of where residents of similar age groups choose to live),

Household sizes (to gauge expectations of the normal living statuses of Toronto residents, as well as gain an understanding of where families vs. singles/couples live in the city),

Unit sizes (to gauge expectations of the types and sizes of units available, loosely related to affordability).

In addition, new movers would likely consider the ‘character’ of neighborhoods, especially in a diverse city like Toronto, Ontario where the variety of neighborhoods appeal to a broad range of demographics and lifestyles. In this case, ‘character’ is defined as the composition of the most common venue types in any given neighborhood.

b. Choropleth Mapping

Under these considerations, choropleth maps were developed for each of the dimensions noted to highlight the distribution across the city of Toronto. In addition, for each choropleth map, markers were placed and colored based on k-means clusters using the most common venue-types within each neighborhood [6][7].

c. K-means Clustering

To perform k-means clustering based on the venue types per each Toronto neighborhood, one-hot encoding was used to transform the occurrence of each venue type into numerical form and allow for the sorting of the most common categories.

After ranking venue types by frequencies, the data was reshaped to show the 1st to 10th most common venue types by neighborhood (see sample below).

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adelaide, King, Richmond	Steakhouse	Coffee Shop	Hotel	Bar	Asian Restaurant
1	Agincourt	Sandwich Place	Concert Hall	Field	Festival	Fast Food Restaurant
2	Alderwood, Long Branch	Pizza Place	Pharmacy	Dance Studio	Coffee Shop	Donut Shop

Figure 8: Sample from Venue Type Frequency Table

4. Results & Observations

From review of the choropleth maps, we note the following observations which may be insightful to the potential new-mover:

Age: Younger residents tend to stay near the city center while the average age of residents appears to increase approaching the midtown area (Midtown neighborhoods: Deer Park, Summerhill, Rosedale), as well as some of the neighborhoods along the outskirts of the map (see Age Distribution map, Appendix i).

Income: Average income increases around the Midtown and Uptown neighborhoods of Toronto (Davisville, Summerhill, Rosedale, Roselawn), as well as around the Harbourfront area, potentially indicating higher housing costs (see Income Distribution map, Appendix ii). However, in combination with the age distribution map, since average age also increases in these areas, it may simply indicate that older residents (with greater experience and higher-paying jobs, and possibly new families) are choosing to live Midtown/Uptown as opposed to within the Downtown core.

The income map must also be cross-examined with the household size distribution map. This is because in the Downtown core (where income appears to be relatively low compared to Midtown and Uptown), low average income and low average age, combined with high household sizes would suggest that low-income individuals are cramming into expensive units to live in the Downtown core.

Household Size: From review of the Household Size map (Appendix iii), the average number of individuals within a unit remains relatively low within the downtown core. Since the data used in this analysis does not cover housing costs, living costs, or density of residential zoning within each neighborhood, more research will have to be conducted to expand on this insight.

Average Bedrooms: From review of the Average Bedroom map (Appendix iv), unit size increases away from the Downtown core, consistent with our household size findings above (Appendix iii).

Key Neighborhood Character Clusters: From the k-means clustering algorithm used to develop the neighborhood markers found in each choropleth map, 3 distinct groups appear (see complete listing of neighborhood character clusters in Appendix v & vi).

First, Cluster #0 (Red) is mostly characterized by coffee shops/cafes, fast food, and Italian restaurants/pizza joints.

Second, Cluster #17 (Orange) is characterized by fast food restaurants, festival events, farmers markets, and fields/parks.

Lastly, Cluster #2 (Dark Purple) is mostly characterized by falafel restaurants, festival events, fields/parks/dog parks, and yoga studios.

5. Conclusion

A noted limitation of the methodology used includes the quantity of data obtained from the Foursquare API as some neighborhoods returned far fewer venue data points compared to others. Therefore, character listings may not perfectly portray the true nature of the neighborhoods with fewer venue-type data points.

In addition, data from the City of Toronto was sourced from the 2016 census and may not perfectly reflect the current landscape of the city. However, changes from the 3-year window of 2016 to current day are not expected to be drastic regarding the dimensions discussed in our analysis (i.e., age, income, household sizes, and unit sizes).

Plans for extending this analysis and project involve the inclusion of household price data, as well as the expansion to other major Canadian cities (e.g., Vancouver, Montreal).

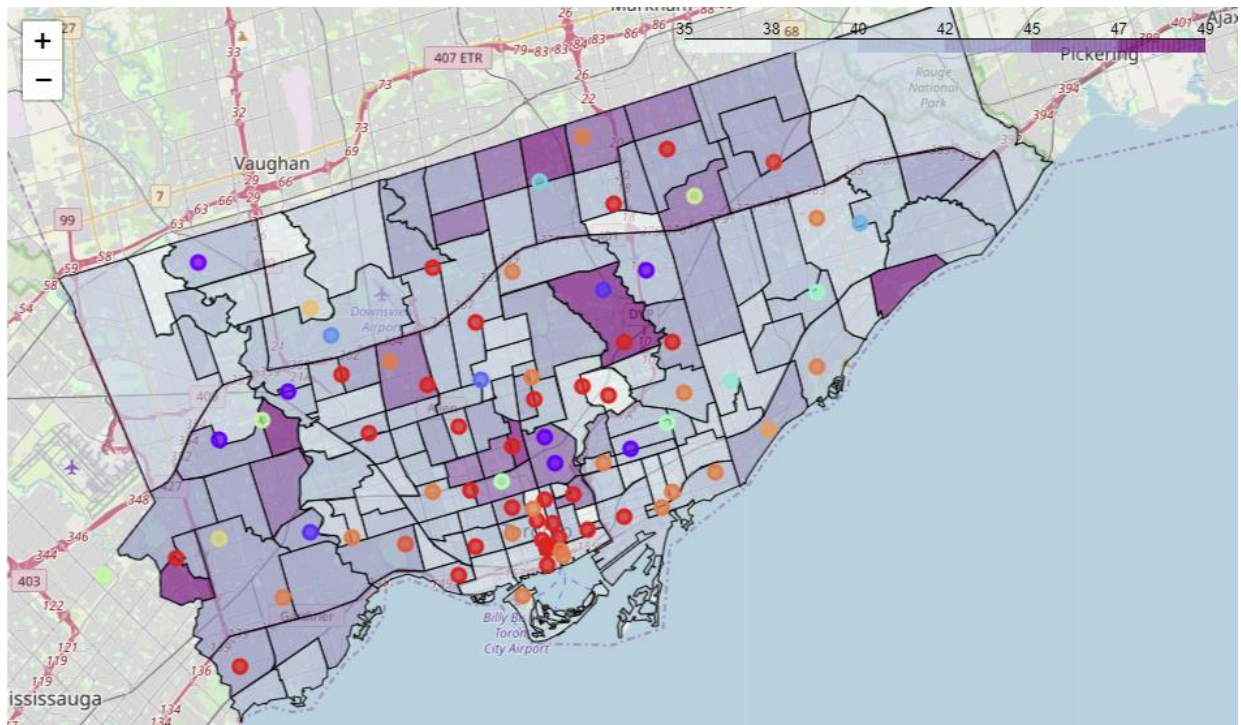
Although the maps do not provide a comprehensive tool to allow for immediate decision-making for the potential new-mover, a high-level overview of Toronto's character and distributions is provided and should aid in the research process by providing a better understanding of the city to those completely unfamiliar.

6. References

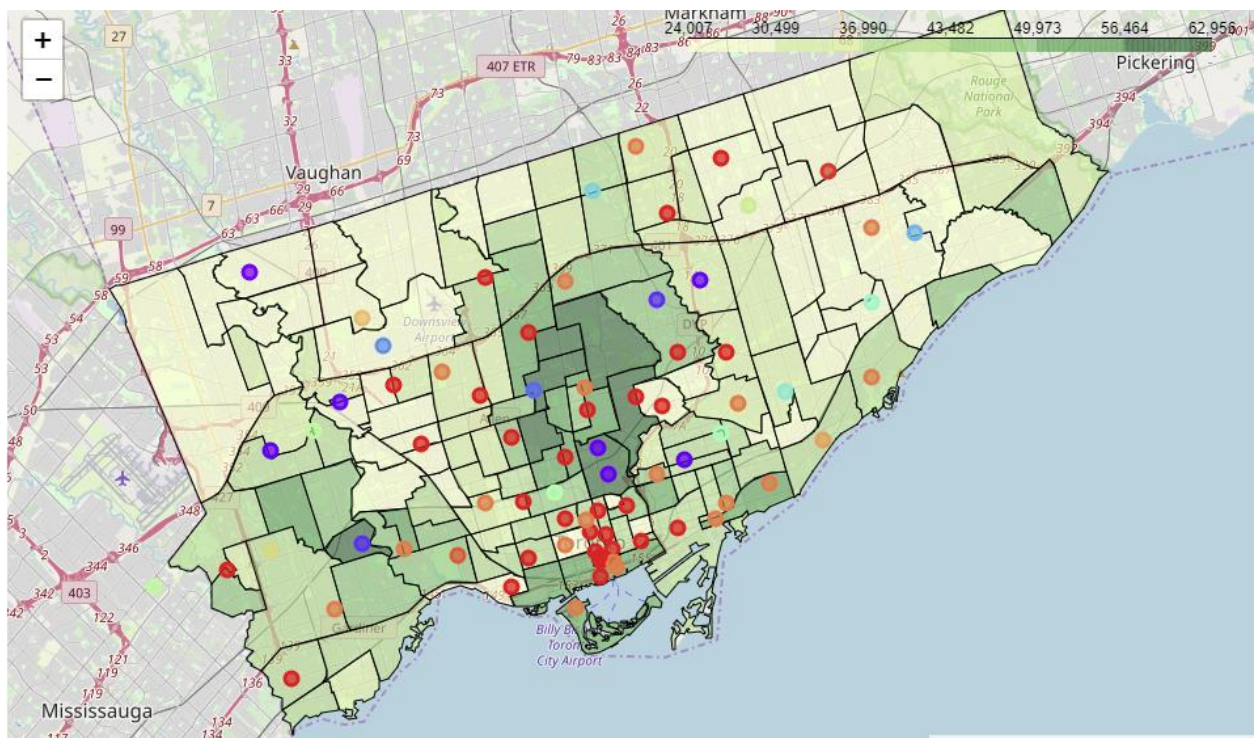
- [1] Toronto Postal Code & Neighborhoods (Wikipedia), url: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- [2] Toronto GeoJSON Data (GitHub: Adamw523), url: <http://adamw523.com/toronto-geojson/>
- [3] Toronto Neighborhood Data (City of Toronto): Age, Income, Household Sizes, Unit Sizes, url: <https://open.toronto.ca/dataset/neighbourhood-profiles/>
- [4] Toronto Neighborhood Venue Data (Foursquare), url: <https://foursquare.com/>
- [5] Geospatial Data (Coursera), url: https://cocl.us/Geospatial_data
- [6] Final Visualization Maps (nbviewer), url: <https://nbviewer.jupyter.org/github/Patrickdg/Toronto-Neighborhood-Visualization-for-Potential-Movers/blob/master/Neighborhood%20Visualization%20Maps.ipynb>
- [7] Source code (GitHub), url: <https://github.com/Patrickdg/Toronto-Neighborhood-Visualization-for-Potential-Movers>

7. Appendix

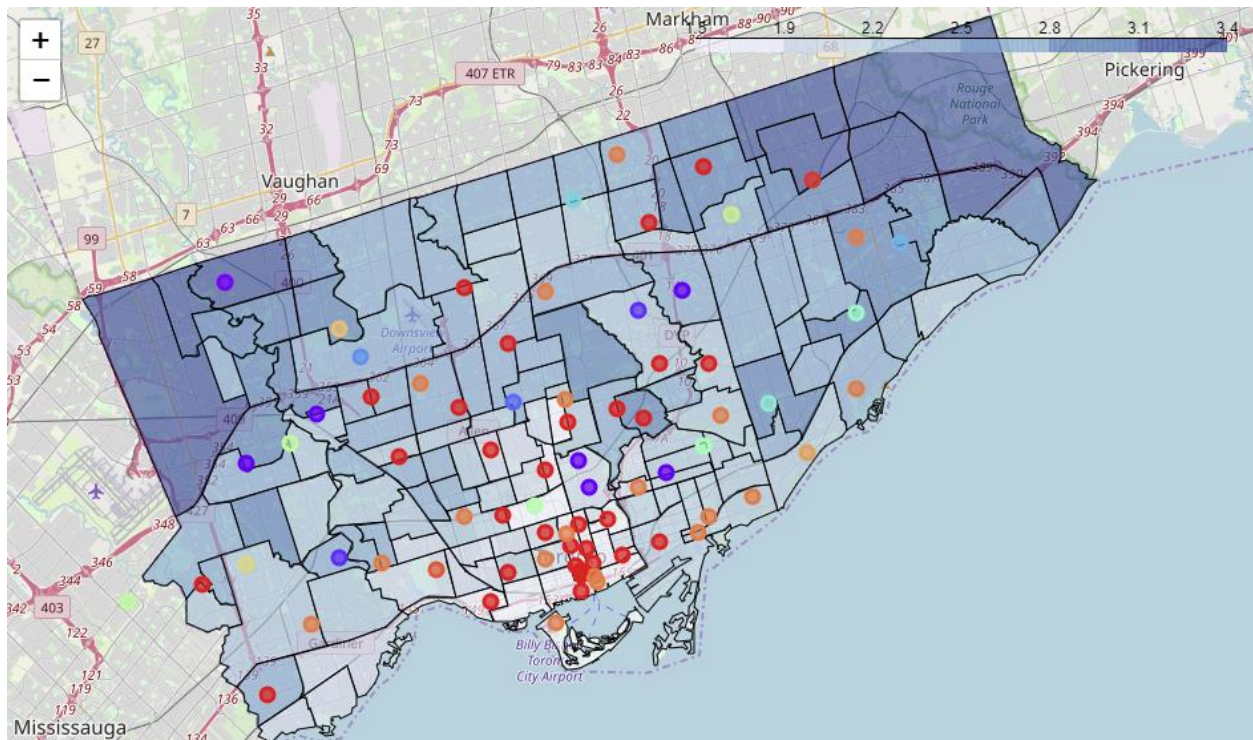
i. Age Distribution Map



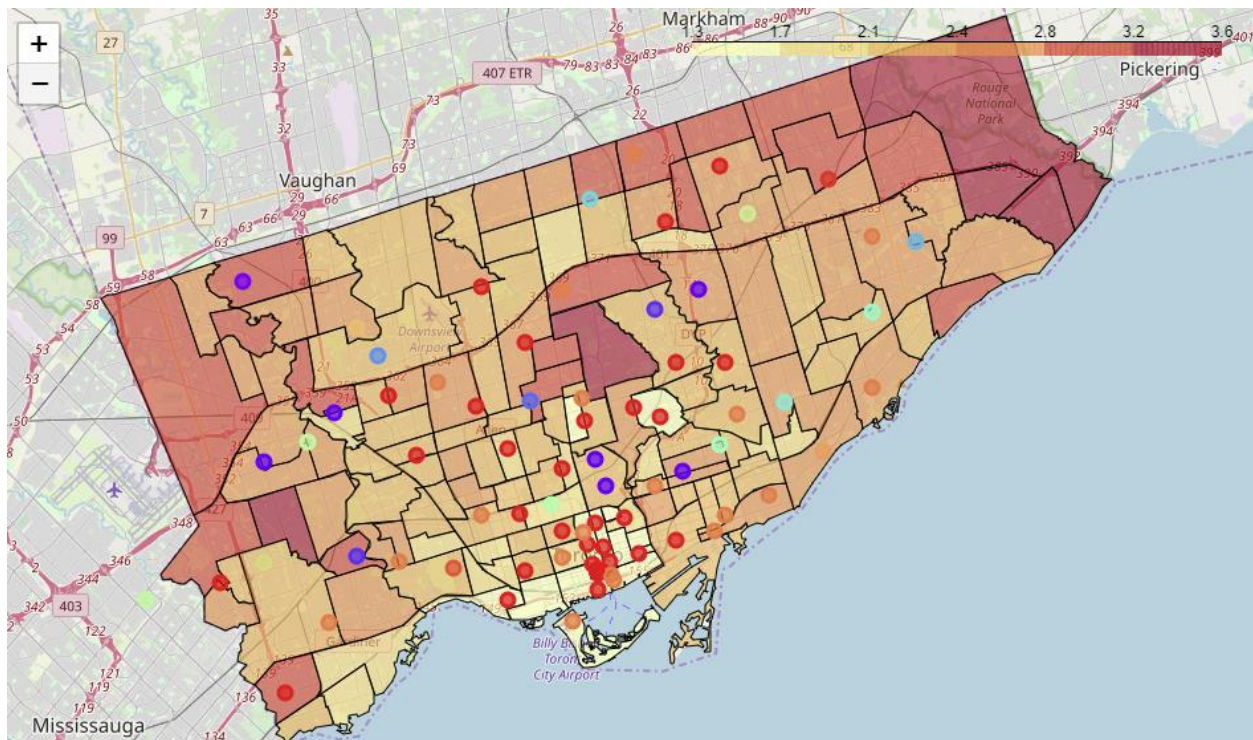
ii. Income Distribution Map



iii. Average Household Size



iv. Average Unit Size



v. Neighborhood Character Cluster Listing (Cluster 0 – 11)

---- Cluster #0 ----	---- Cluster #4 ----	---- Cluster #8 ----
0 Coffee Shop	0 Electronics Store	0 Falafel Restaurant
1 Restaurant	1 Donut Shop	1 Festival
2 Café	2 Festival	2 Business Service
3 Fast Food Restaurant	3 Ethiopian Restaurant	3 Yoga Studio
4 Sandwich Place	4 Falafel Restaurant	4 Farmers Market
5 Pizza Place	5 Health & Beauty Service	5 Ethiopian Restaurant
6 Farmers Market	6 Farmers Market	6 Field
7 Festival	7 Field	7 Fast Food Restaurant
8 Italian Restaurant	8 Fast Food Restaurant	8 Convenience Store
9 Falafel Restaurant	9 Fish & Chips Shop	9 Fish & Chips Shop
---- Cluster #1 ----	---- Cluster #5 ----	---- Cluster #9 ----
0 Electronics Store	0 Falafel Restaurant	0 Dumpling Restaurant
1 Pizza Place	1 Flower Shop	1 Electronics Store
2 Festival	2 Festival	2 Falafel Restaurant
3 Dog Run	3 Yoga Studio	3 Festival
4 Yoga Studio	4 Farmers Market	4 Dog Run
5 Farmers Market	5 Ethiopian Restaurant	5 Farmers Market
6 Ethiopian Restaurant	6 Field	6 Ethiopian Restaurant
7 Field	7 Baseball Field	7 Field
8 Falafel Restaurant	8 Fast Food Restaurant	8 Playground
9 Fast Food Restaurant	9 Fish & Chips Shop	9 Fast Food Restaurant
---- Cluster #2 ----	---- Cluster #6 ----	---- Cluster #10 ----
0 Falafel Restaurant	0 Flower Shop	0 Electronics Store
1 Festival	1 Festival	1 Beer Store
2 Farmers Market	2 Ethiopian Restaurant	2 Festival
3 Field	3 Yoga Studio	3 Yoga Studio
4 Park	4 Farmers Market	4 Bus Stop
5 Yoga Studio	5 Korean Restaurant	5 Farmers Market
6 Ethiopian Restaurant	6 Field	6 Field
7 Fast Food Restaurant	7 Falafel Restaurant	7 Dance Studio
8 Electronics Store	8 Fast Food Restaurant	8 Fast Food Restaurant
9 Dog Run	9 Fish & Chips Shop	9 Fish & Chips Shop
---- Cluster #3 ----	---- Cluster #7 ----	---- Cluster #11 ----
0 Falafel Restaurant	0 Electronics Store	0 Fish & Chips Shop
1 Festival	1 Shopping Plaza	1 Donut Shop
2 Dog Run	2 Falafel Restaurant	2 Festival
3 Ethiopian Restaurant	3 Festival	3 Yoga Studio
4 Yoga Studio	4 Yoga Studio	4 Farmers Market
5 Farmers Market	5 Farmers Market	5 Ethiopian Restaurant
6 Pool	6 Ethiopian Restaurant	6 Field
7 Field	7 Field	7 Falafel Restaurant
8 Fast Food Restaurant	8 Fast Food Restaurant	8 Fast Food Restaurant
9 Electronics Store	9 Concert Hall	9 Hotel

vi. Neighborhood Character Cluster Listing (Cluster 12 – 19)

---- Cluster #12 ----

0	Festival
1	Yoga Studio
2	Farmers Market
3	Ethiopian Restaurant
4	Field
5	Coffee Shop
6	Falafel Restaurant
7	Fast Food Restaurant
8	Donut Shop
9	Dog Run

---- Cluster #16 ----

0	Pizza Place
1	Festival
2	Donut Shop
3	Yoga Studio
4	Farmers Market
5	Ethiopian Restaurant
6	Field
7	Falafel Restaurant
8	Fast Food Restaurant
9	Café

---- Cluster #13 ----

0	Electronics Store
1	Falafel Restaurant
2	Festival
3	Shopping Mall
4	Yoga Studio
5	Farmers Market
6	Ethiopian Restaurant
7	Field
8	Concert Hall
9	Fast Food Restaurant

---- Cluster #17 ----

0	Fast Food Restaurant
1	Festival
2	Farmers Market
3	Field
4	Fish & Chips Shop
5	Falafel Restaurant
6	Yoga Studio
7	Ethiopian Restaurant
8	Dumpling Restaurant
9	Bank

---- Cluster #14 ----

0	Electronics Store
1	Print Shop
2	Falafel Restaurant
3	Festival
4	Dog Run
5	Yoga Studio
6	Farmers Market
7	Ethiopian Restaurant
8	Field
9	Fast Food Restaurant

---- Cluster #18 ----

0	Fish Market
1	Yoga Studio
2	Festival
3	Light Rail Station
4	Garden
5	Farmers Market
6	Ethiopian Restaurant
7	Field
8	Falafel Restaurant
9	Fast Food Restaurant

---- Cluster #15 ----

0	Donut Shop
1	Home Service
2	Festival
3	Yoga Studio
4	Farmers Market
5	Field
6	Falafel Restaurant
7	Fast Food Restaurant
8	Fish & Chips Shop
9	Wings Joint

---- Cluster #19 ----

0	Dumpling Restaurant
1	Electronics Store
2	Falafel Restaurant
3	Festival
4	Farmers Market
5	Ethiopian Restaurant
6	Field
7	Sandwich Place
8	Fast Food Restaurant
9	Concert Hall