# THE BATTLE OF NEIGHBOURHOODS

# CITY OF TORONTO

Henry – September 09, 2019

# 1. Introduction

Problem

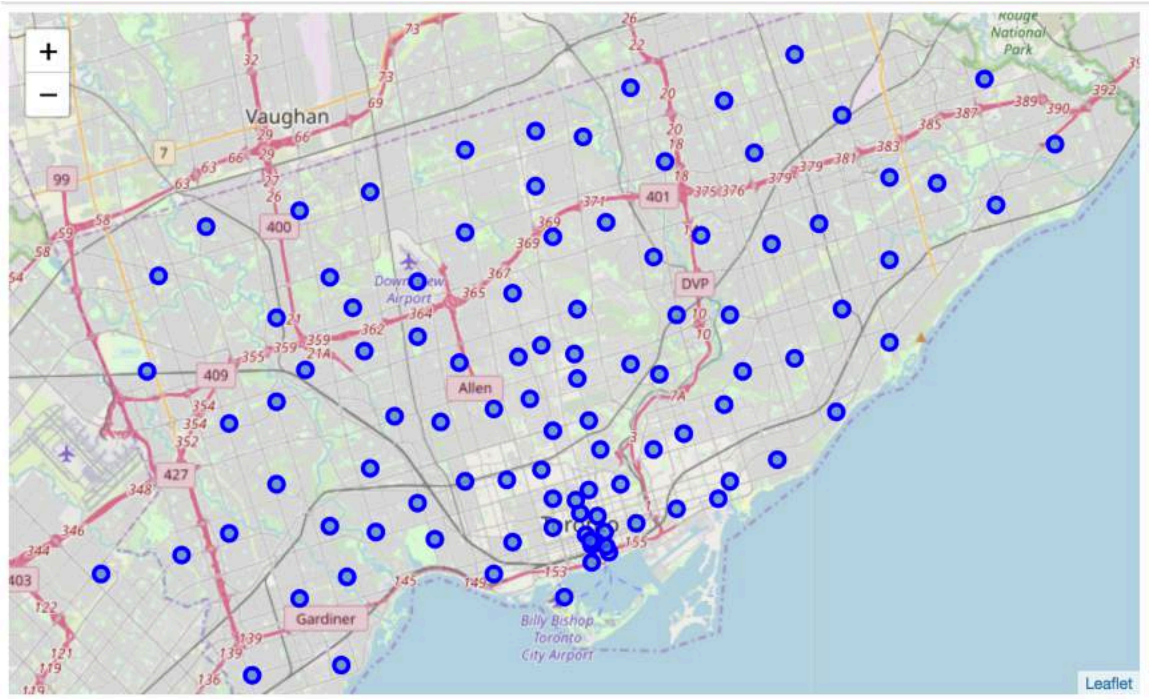
Which neighbourhoods have the similar features from where a family with young children can purchase a detached residential property?

Who are interested?

Families with young children of school age who are looking for a residential property in City of Toronto.

# Data Acquisition and Cleaning

- FSA postcodes of City of Toronto from Wikipedia
- Geospatial coordinates of each postcode



# 2. Data Acquisition and Cleaning (cont.)

- Venues in each postcode from Foursqaure
- Limit top 10 venue categories

| | |
|---|---|
| Coffee Shop | 303 |
| Café | 169 |
| Park | 117 |
| Pizza Place | 113 |
| Restaurant | 102 |
| Italian Restaurant | 97 |
| Bakery | 94 |
| Sandwich Place | 78 |
| Hotel | 68 |
| Bar | 68 |

# 2. Data Acquisition and Cleaning (cont.)

- Elementary and Secondary Schools in each postcode from Foursqaure
- School ratings from Fraser Institute

| | Postcode | Latitude | Longitude | Avg Rating Elementary Schools | Avg Rating High Schools |
|---|---|---|---|---|---|
| 0 | M1B | 43.806686 | -79.194353 | 6.240000 | 5.933333 |
| 1 | M1C | 43.784535 | -79.160497 | 6.950000 | 4.125000 |
| 2 | M1E | 43.763573 | -79.188711 | 5.033333 | 5.033333 |
| 3 | M1G | 43.770992 | -79.216917 | 4.075000 | 5.750000 |
| 4 | M1H | 43.773136 | -79.239476 | 5.200000 | 3.800000 |
| ... | ... | ... | ... | ... | ... |
| 98 | M9N | 43.706876 | -79.518188 | 6.500000 | 0.000000 |
| 99 | M9P | 43.696319 | -79.532242 | 5.900000 | 7.800000 |
| 100 | M9R | 43.688905 | -79.554724 | 5.900000 | 7.266667 |
| 101 | M9V | 43.739416 | -79.588437 | 5.200000 | 7.966667 |
| 102 | M9W | 43.706748 | -79.594054 | 3.850000 | 7.450000 |

103 rows × 5 columns

# 2. Data Acquisition and Cleaning (cont.)

- Transportation services in each postcode from Foursqaure
- High count signifies easy transportation access

| | Postcode | Latitude | Longitude | Transportation Count |
|---|---|---|---|---|
| 0 | M1B | 43.806686 | -79.194353 | 3 |
| 1 | M1C | 43.784535 | -79.160497 | 1 |
| 2 | M1E | 43.763573 | -79.188711 | 5 |
| 3 | M1G | 43.770992 | -79.216917 | 0 |
| 4 | M1H | 43.773136 | -79.239476 | 1 |
| ... | ... | ... | ... | ... |
| 98 | M9N | 43.706876 | -79.518188 | 0 |
| 99 | M9P | 43.696319 | -79.532242 | 1 |
| 100 | M9R | 43.688905 | -79.554724 | 7 |
| 101 | M9V | 43.739416 | -79.588437 | 3 |
| 102 | M9W | 43.706748 | -79.594054 | 0 |

103 rows × 4 columns

# 2. Data Acquisition and Cleaning (cont.)

- 2016 Census from Statistic Canada

| Data | HTML headers attribute value |
|---|---|
| Population size | 'L1000 geo1 total1' |
| Median age | 'L2033 geo1 total1' |
| Average household size | 'L3017 geo1 total1' |
| Without children in a census family | 'L6003 geo1 total1' |
| With children in a census family | 'L6004 geo1 total1' |
| Number of households | 'L6001 geo1 total1' |
| Median total income | 'L13001 geo1 total1' |
| Non-immigrants | 'L18001 geo1 total1' |
| ImmigrantsCensus | 'L18002 geo1 total1' |
| Non-permanent residents | 'L18010 geo1 total1' |

| | Postcode | Census Population Size | Census Median Age | Census Average Household Size | Census Households With Children (%) | Census Number of Households | Census Median Family Income | Census Non Immigrants (%) | Census Immigrants (%) | Census Non Permanent Residents (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M1B | 66108 | 38.2 | 3.3 | 80.10 | 20230 | 69126.0 | 38.21 | 59.92 | 1.87 |
| 1 | M1C | 35626 | 44.0 | 3.1 | 70.14 | 11275 | 109785.0 | 54.12 | 44.89 | 0.99 |
| 2 | M1E | 46943 | 42.2 | 2.7 | 72.57 | 17160 | 62047.0 | 50.74 | 47.79 | 1.47 |
| 3 | M1G | 29690 | 37.2 | 3.0 | 76.57 | 9765 | 54450.0 | 40.54 | 56.03 | 3.43 |
| 4 | M1H | 24383 | 38.1 | 2.7 | 70.77 | 8985 | 58492.0 | 37.27 | 57.95 | 4.77 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 98 | M9N | 25074 | 39.0 | 2.4 | 73.68 | 10170 | 50545.0 | 49.56 | 47.74 | 2.71 |
| 99 | M9P | 20874 | 45.9 | 2.6 | 67.12 | 7905 | 73425.0 | 55.01 | 43.82 | 1.17 |
| 100 | M9R | 33743 | 40.6 | 2.7 | 70.89 | 12335 | 63032.0 | 47.34 | 50.69 | 1.97 |
| 101 | M9V | 55959 | 35.9 | 3.3 | 80.15 | 16805 | 59760.0 | 34.74 | 62.55 | 2.71 |
| 102 | M9W | 40684 | 38.3 | 2.9 | 74.47 | 13705 | 65826.0 | 44.18 | 52.79 | 3.02 |

103 rows × 10 columns

# 3. Methodology and Results

- Normalized data
- School ratings dived by 10
- Percentage divided by 100

| | Postcode | Coffee Shop | Café | Park | Pizza Place | Restaurant | Italian Restaurant | Bakery | Sandwich Place | Bar | Hotel | Avg Rating Elementary Schools | Avg Rating High Schools | Transportation Count | Census Population Size |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M1B | 0.083333 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.624000 | 0.593333 | 0.061224 | 0.871023 |
| 1 | M1C | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.2 | 0.000000 | 0.000000 | 0.2 | 0.0 | 0.695000 | 0.412500 | 0.020408 | 0.469399 |
| 2 | M1E | 0.133333 | 0.0 | 0.000000 | 0.200000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.503333 | 0.503333 | 0.102041 | 0.618509 |
| 3 | M1G | 0.333333 | 0.0 | 0.333333 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.407500 | 0.575000 | 0.000000 | 0.391188 |
| 4 | M1H | 0.090909 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.090909 | 0.000000 | 0.0 | 0.0 | 0.520000 | 0.380000 | 0.020408 | 0.321264 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 98 | M9N | 0.000000 | 0.0 | 0.000000 | 0.166667 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.650000 | 0.000000 | 0.000000 | 0.330369 |
| 99 | M9P | 0.100000 | 0.0 | 0.000000 | 0.100000 | 0.0 | 0.0 | 0.000000 | 0.100000 | 0.0 | 0.0 | 0.590000 | 0.780000 | 0.020408 | 0.275031 |
| 100 | M9R | 0.071429 | 0.0 | 0.000000 | 0.071429 | 0.0 | 0.0 | 0.000000 | 0.071429 | 0.0 | 0.0 | 0.590000 | 0.726667 | 0.142857 | 0.444589 |
| 101 | M9V | 0.076923 | 0.0 | 0.000000 | 0.230769 | 0.0 | 0.0 | 0.000000 | 0.076923 | 0.0 | 0.0 | 0.520000 | 0.796667 | 0.061224 | 0.737302 |
| 102 | M9W | 0.000000 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.000000 | 0.000000 | 0.0 | 0.0 | 0.385000 | 0.745000 | 0.000000 | 0.536042 |

103 rows × 23 columns

# 3. Methodology and Results (cont.)

- K-Means clustering with K=7

Cluster label 0: 7
Cluster label 1: 18
Cluster label 2: 7
Cluster label 3: 8
Cluster label 4: 34
Cluster label 5: 4
Cluster label 6: 25

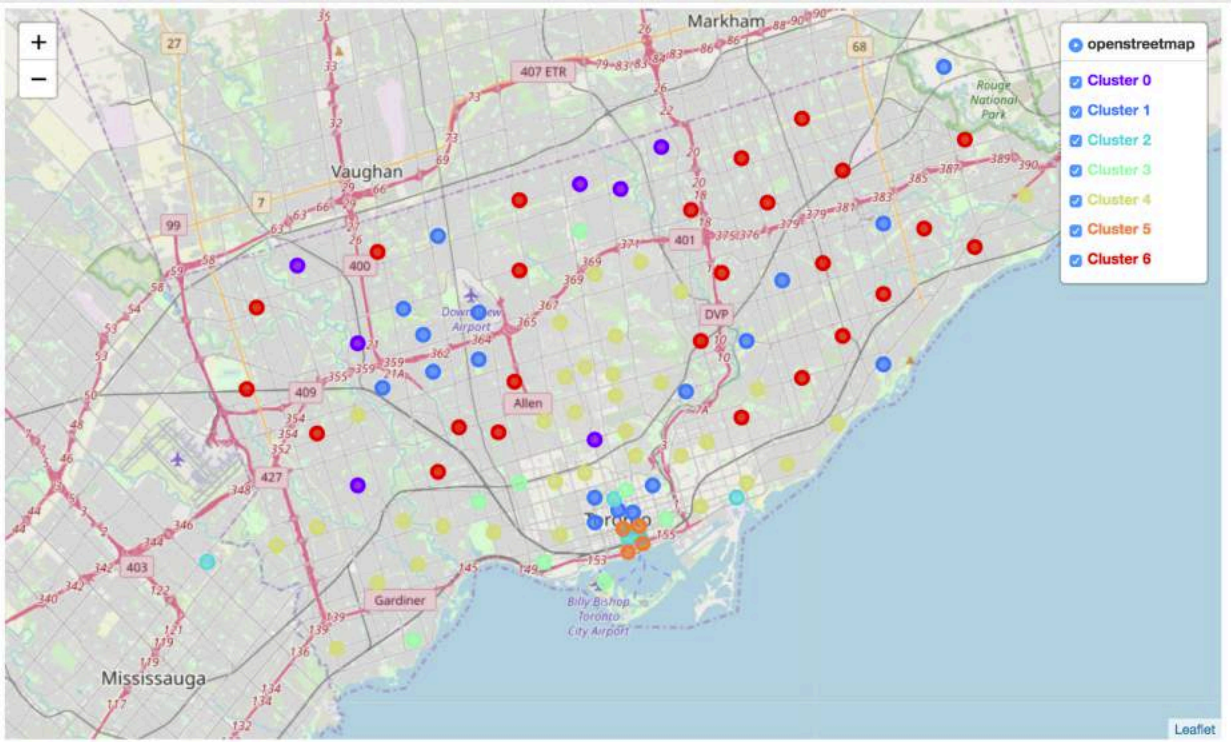
| Cluster Labels | Avg Rating Elementary Schools | Avg Rating High Schools | Transportation Count | Census Population Size | Census Median Age | Census Average Household Size | Census Households With Children (%) | Census Number of Households | Census Median Family Income | Census Non Immigrants (%) | Census Immigrants (%) | Census Non Permanent Residents (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.000000 | 6.909841 | 2.428571 | 24102.428571 | 42.142857 | 2.571429 | 63.782857 | 9640.000000 | 69550.285714 | 42.551429 | 53.604286 | 3.844286 |
| 1 | 6.426481 | 4.830992 | 3.888889 | 18951.388889 | 37.816667 | 2.511111 | 67.159444 | 7534.722222 | 56140.000000 | 43.737778 | 51.050000 | 5.212222 |
| 2 | 5.076531 | 5.841111 | 30.428571 | 6.428571 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 3 | 7.224896 | 6.532054 | 3.875000 | 45069.875000 | 36.225000 | 1.887500 | 47.798750 | 23288.125000 | 61639.250000 | 54.820000 | 39.465000 | 5.716250 |
| 4 | 7.443186 | 6.889511 | 2.500000 | 22639.000000 | 41.911765 | 2.338235 | 61.495588 | 9636.323529 | 88599.441176 | 65.060294 | 32.721765 | 2.217941 |
| 5 | 5.737500 | 5.771250 | 38.500000 | 7154.750000 | 33.975000 | 1.625000 | 31.440000 | 4322.500000 | 86254.750000 | 55.617500 | 38.465000 | 5.920000 |
| 6 | 6.014133 | 6.015467 | 3.040000 | 42533.920000 | 40.116000 | 2.776000 | 72.959600 | 15151.600000 | 60056.360000 | 40.668800 | 56.274000 | 3.056400 |

# 3. Methodology and Results (cont.)

Averages of each cluster

| Cluster Labels | Avg Rating Elementary Schools | Avg Rating High Schools | Transportation Count | Census Population Size | Census Median Age | Census Average Household Size | Census Households With Children (%) | Census Number of Households | Census Median Family Income | Census Non Immigrants (%) | Census Immigrants (%) | Census Non Permanent Residents (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.000000 | 6.909841 | 2.428571 | 24102.428571 | 42.142857 | 2.571429 | 63.782857 | 9640.000000 | 69550.285714 | 42.551429 | 53.604286 | 3.844286 |
| 1 | 6.426481 | 4.830992 | 3.888889 | 18951.388889 | 37.816667 | 2.511111 | 67.159444 | 7534.722222 | 56140.000000 | 43.737778 | 51.050000 | 5.212222 |
| 2 | 5.076531 | 5.841111 | 30.428571 | 6.428571 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 3 | 7.224896 | 6.532054 | 3.875000 | 45069.875000 | 36.225000 | 1.887500 | 47.798750 | 23288.125000 | 61639.250000 | 54.820000 | 39.465000 | 5.716250 |
| 4 | 7.443186 | 6.889511 | 2.500000 | 22639.000000 | 41.911765 | 2.338235 | 61.495588 | 9636.323529 | 88599.441176 | 65.060294 | 32.721765 | 2.217941 |
| 5 | 5.737500 | 5.771250 | 38.500000 | 7154.750000 | 33.975000 | 1.625000 | 31.440000 | 4322.500000 | 86254.750000 | 55.617500 | 38.465000 | 5.920000 |
| 6 | 6.014133 | 6.015467 | 3.040000 | 42533.920000 | 40.116000 | 2.776000 | 72.959600 | 15151.600000 | 60056.360000 | 40.668800 | 56.274000 | 3.056400 |

# 3. Methodology and Results (cont.)



# 4. Data Analysis

**Cluster 0: Good Schools and Families with Children**

- Top venues: Eateries and parks
- Good secondary schools but no elementary school rating found
- Many families with children (63.7%)
- Medium population density

**Cluster 1: Good Elementary Schools and Young Families with Children**

- Top venues: Eateries and parks
- Good elementary schools but poor secondary school ratings
- Young families in the neighbourhoods (Census Median Age < 40)
- Many families with children (67%)
- Low population density
- Look like these were new neighbourhoods in development

# 4. Data Analysis (cont.)

**Cluster 2: Postcodes with insufficient census data**

- There were no census data (zero most of the cases) in these postcode areas because the census population size was very small (less than 15)
- No characteristics could be determined

**Cluster 3: Downtown Toronto, Average Schools, and High Population Density**

- Top venues: Eateries (coffee, pizza, and restaurants) and hotels
- Household Size was small, at 1.89 persons per household
- Both Elementary and Secondary School Ratings were average to good
- High Population Density due to the highest Census Population Size with small Household Size, most likely high rise apartments

# 4. Data Analysis (cont.)

**Cluster 4: Good Schools, High Family Income, and Non-Immigrants**

- Top venues: Eateries and parks
- Good elementary and secondary school ratings
- High family income
- Predominantly non-immigrants (Census non-Immigrants = 65%)
- Older neighbourhoods located South of Highway 401

**Cluster 5: Downtown Toronto, High Family Income, and Young Families**

- Top venues: Eateries and hotels
- Schools were poor with both ratings below 6
- Household Size was small, at 1.62 persons per houshold, which was consistent with the low Family With Children (31%)
- High Census Median Family Income (over $82k)
- Look like mostly young (Census Median Age = 34) professional couples without children lived in these neighbourhood

# 4. Data Analysis (cont.)

**Cluster 6: Newer Neighbourhoods with High Population Density and Families with Children**

- To venues: Eateries and parks
- Schools were average
- High population density with large Census Population Size (over 42k persons) and the highest Census Average Household Size (2.78 persons per household)
- 73% families with children, significantly higher than other clusters

**Additional Observations in General about Toronto**

- It seems elementary and secondary school ratings were poor in general, mostly 6 or below
- Non-immigrants and immigrants ratios was close to 50:50 in average
- Household size was small with average below 3

# 5. Discussion

**Issues on Data Quality**

- Foursquare venues were more on leisure places
- Did not have school ratings for all schools
- 2016 Census data were 3 years old

**Difficulties on Data Acquisition**

- Could not scrap and match communities from HouseSigma to get median pricing for residential properties

# 6. Conclusion

A family can determine which features are the most important and then search the neighbourhoods in the cluster, which has the features the family is looking for.

However, there are "human factors" that cannot include in "mathematical analysis" which sway the house purchasing decision. Examples are: nearness to other family members or friends, closeness to employment sites, ambient of the neighbourhood (such as country style) , ravines or old large trees, etc.