Selecting Candidate Neighborhoods

1. Introduction

1.1 Problem

A Luxury Fitness Club chain wants to expand its business and demands a research about potential neighborhoods they could open a new unit in TORONTO - CA. To accomplish that, they hired a consultancy company to help them.

1.2 Goals

The central aim of this project is to point out 10 neighborhoods out of 140, where the market competition for the target residents could be lesser, reducing then, the costs of infield surveys, which will not be covered in this project. They also want to know an overview profile about those potential neighborhoods.

Knowing that Toronto has 140 neighborhoods, firstly, the hired company wants to select potential candidates for further field surveys and then reducing then, the field survey costs that will be done in the future, like target properties price's, access easiness and others, which will not be covered in this project.

The central aim of this project is to point out 10 neighborhoods out of 140, where the market competition for the target residents could be lesser reducing then, the costs of infield surveys.

2. Data acquisition and cleaning

2.1 Neighborhoods' Demographics

To get information about Toronto neighborhoods' demographics, I downloaded the provided data on https://open.toronto.ca/dataset/neighbourhood-profiles/.

The Census of Population is held across Canada every 5 years and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labor. City of Toronto Neighborhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighborhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighborhood.

In these profiles, "neighborhood" refers to the City of Toronto's 140 social planning neighborhoods. These social planning neighborhoods were developed by the City of Toronto to help government and community organizations with local planning by providing socio-economic data at a meaningful geographic area. The boundaries of these social planning neighborhoods are consistent over time, allowing for comparison between Census years. Neighborhood level data from a variety of other sources are also available through the City's Wellbeing Toronto mapping application and here on the Open Data portal. Each data point in this file is presented for the City's 140 neighborhoods, as well as for the City of Toronto as a whole. The data is sourced from a number of Census tables released by Statistics Canada. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information. For definitions of terms and concepts referenced in this data set, users should consult the reference materials produced by Statistics Canada for the 2016 Census, available online at: http://www12.statcan.gc.ca/census-recensement/2016/ref/index-eng.cfm.

Where I used the following _id and Characteristic:

Neighborhood Information:

- 1 - Neighborhood Number

Population and dwellings:

- 3 - Population, 2016
- 9 - Land area in square kilometers

Income of households in 2015: ($)

- 1055 - Total - Household after-tax income groups in 2015 for private households - 100% data
- 1056 - Under 5000
- 1057 - 5,000 to 9,999
- 1058 - 10,000 to 14,999
- 1059 - 15,000 to 19,999
- 1060 - 20,000 to 24,999
- 1061 - 25,000 to 29,999
- 1062 - 30,000 to 34,999
- 1063 - 35,000 to 39,999
- 1064 - 40,000 to 44,999
- 1065 - 45,000 to 49,999
- 1066 - 50,000 to 59,999
- 1067 - 60,000 to 69,999
- 1068 - 70,000 to 79,999
- 1069 - 80,000 to 89,999
- 1070 - 90,000 to 99,999
- 1071 - 100,000 and over

Age characteristics:

- 10 - Children (0-14 years)
- 11 - Youth (15-24 years)
- 12 - Working Age (25-54 years)
- 13 - Pre-retirement (55-64 years)
- 14 - Seniors (65+ years)
- 15 - Older Seniors (85+ years)

Commuting duration:

- 1973 - Total - Commuting duration for the employed labor force aged 15 years and over in private households with a usual place of work or no fixed workplace address - 25% sample data
- 1974 - Less than 15 minutes
- 1975 - 15 to 29 minutes
- 1976 - 30 to 44 minutes
- 1977 - 45 to 59 minutes
- 1978 - 60 minutes and over

Private households by household size:

- 69 – 1 person
- 70 – 2 persons
- 71 – 3 persons
- 72 – 4 persons
- 73 – 5 or more persons

Marital status for the population aged 15 years and over:

- 78 – Married

- 79 – Living common law
- 81 – Never married
- 82 – Separated
- 83 – Divorced
- 84 - Widowed

Time leaving for work for the employed labor force aged 15 years and over:

- 1980 - Between 5 a.m. and 5:59 a.m.
- 1981 - Between 6 a.m. and 6:59 a.m.
- 1982 - Between 7 a.m. and 7:59 a.m.
- 1983 - Between 8 a.m. and 8:59 a.m.
- 1984 - Between 9 a.m. and 11:59 a.m.
- 1985 - Between 12 p.m. and 4:59 a.m.

Main mode of commuting for the employed labor force aged 15 years and over:

- 1967 - Car, truck, van - as a driver
- 1968 - Car, truck, van - as a passenger
- 1969 - Public transit
- 1970 - Walked
- 1971 - Bicycle
- 1972 - Other method

All occupations:

- 1923 - 0 Management occupations
- 1924 - 1 Business, finance and administration occupations
- 1925 - 2 Natural and applied sciences and related occupations
- 1926 - 3 Health occupations
- 1927 - 4 Occupations in education, law and social, community and government services
- 1928 - 5 Occupations in art, culture, recreation and sport
- 1929 - 6 Sales and service occupations
- 1930 - 7 Trades, transport and equipment operators and related occupations
- 1931 - 8 Natural resources, agriculture and related production occupations
- 1932 - 9 Occupations in manufacturing and utilities

2.2 Neighborhoods' Demographics

To get all the geographical coordinates of the neighborhoods, I downloaded the provided data on "https://open.toronto.ca/dataset/neighbourhoods/", where I retrieved the following for each neighborhoods:

- AREA_SHORT_CODE - ID of each neighborhood
- LONGITUDE - Longitude of each neighborhood
- LATITUDE - Latitude of each neighborhood
- geometry - polygon data about the neighborhoods' geographical limit

2.3 Information about gyms

To get location and other information about the gyms in Toronto I'm using Foursquare's explore API. Using the Foursquare's explore API (which gives venue recommendations), I'm fetching details about the venues up present in Toronto and collected their names, categories, and locations (latitude and longitude).

From Foursquare API "https://developer.foursquare.com/docs" I retrieved the following for each venue:

- Name: The name of the venue.

- Category: The category type as defined by the API.
- Latitude: The latitude value of the venue.
- Longitude: The longitude value of the venue.

3. Methodology

In this project, we will direct our effort on detecting neighborhoods of Toronto that have low gym density, particularly low density of our target clients.

With that in mind we'll describe our target client based on past researches conducted during the strategic plan in 2019:

- People who have an after-tax household income over CA$ 100.000
- People aged between 15 and 54
- People that commute less than 30 minutes a day
    - In this regard there are some scientific research that supports that thought like those that are cited here: https://engage.gov.bc.ca/app/uploads/sites/391/2018/08/Closer-Commutes.pdf

To select our future candidates, I propose the following methodology that was approved by the stakeholders.

We created an indicator that will be applied to each neighborhood and will be the final criteria do select the neighborhoods' sample for future studies (project's goal).

$$gym\_target+1 = E / ((A \times 1/B) \times (C \times D))$$

where:

E - number of 'Gym' and 'Gym / Fitness Center' inside the neighboorhood, plus one (representing our possible new unit)

A = number of people who have an after-tax household income over CA$ 100.000

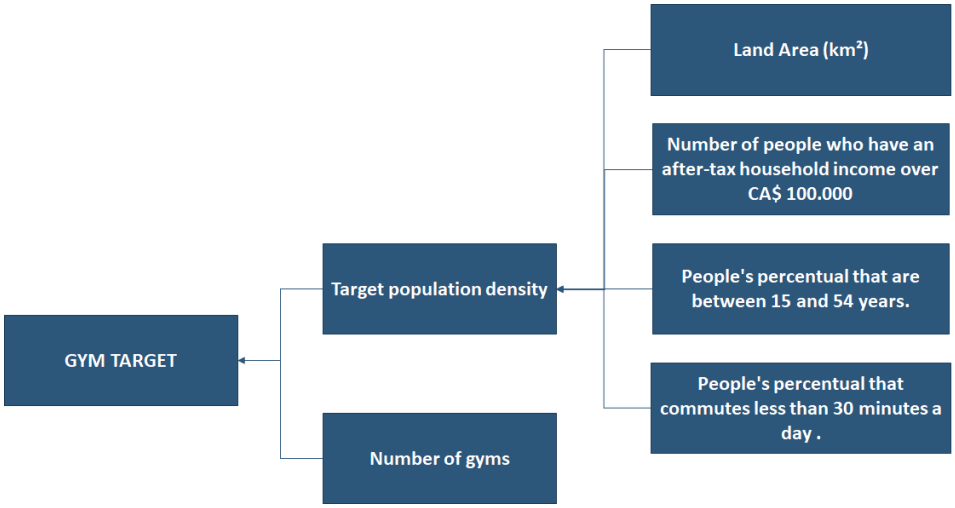
B = Land area in square kilometers.

C = People's percentual that are between 15 and 54 years.

D = People's percentual that commutes less than 30 minutes a day .

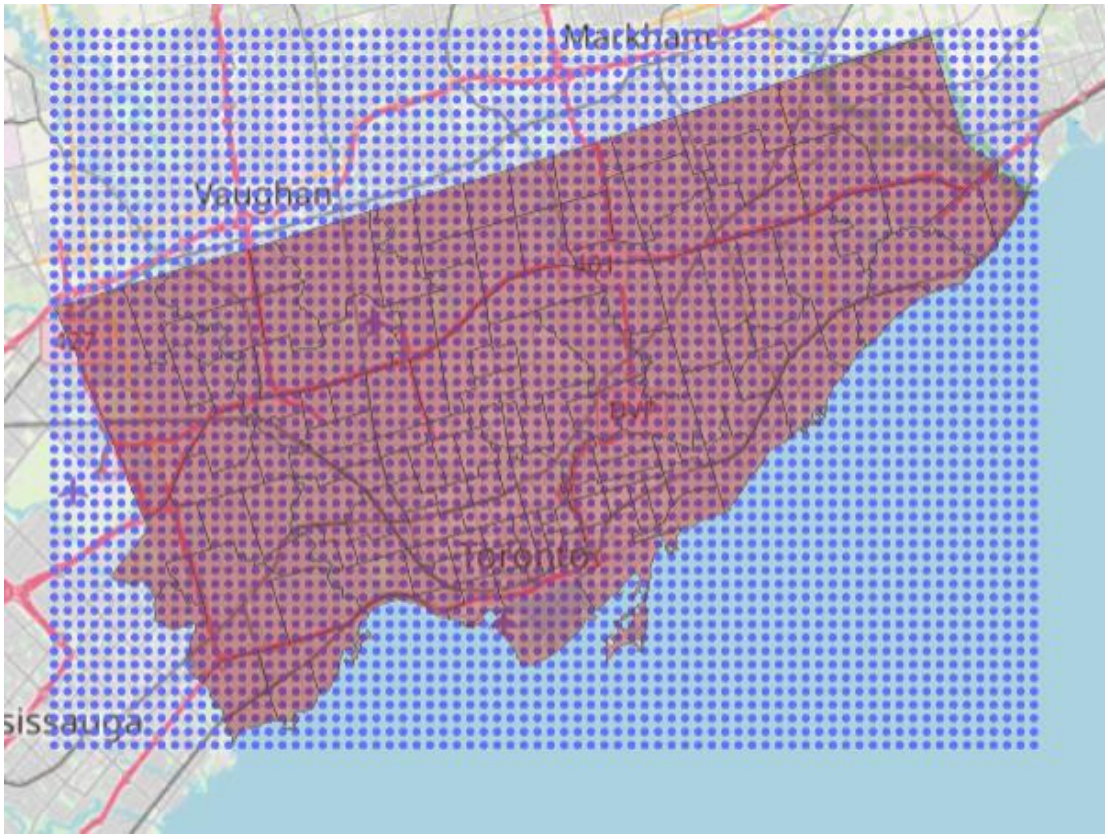
The result will be an indicator that represents the number of target people per one unit of gym per one square kilometer.

It is important to deploy that indicator graphically to understand how the variables are related:



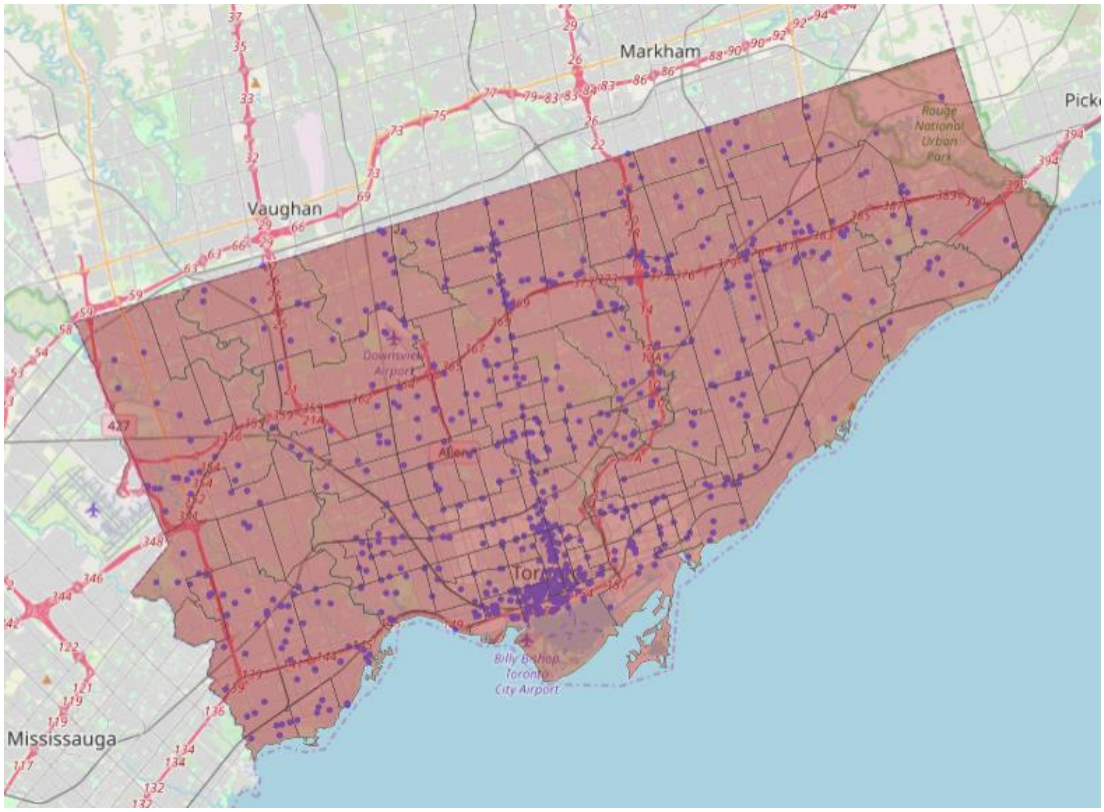
3.2 **Grid of points to search over:**

Once, Foursquare API doesn't have an easy option to search inside a polygon, I took a decision to create a grid of geographical points over the limits of the neighborhood's geometry, as follow:



It was generated 3996 equally spaced points.

3.3. Gyms' location:



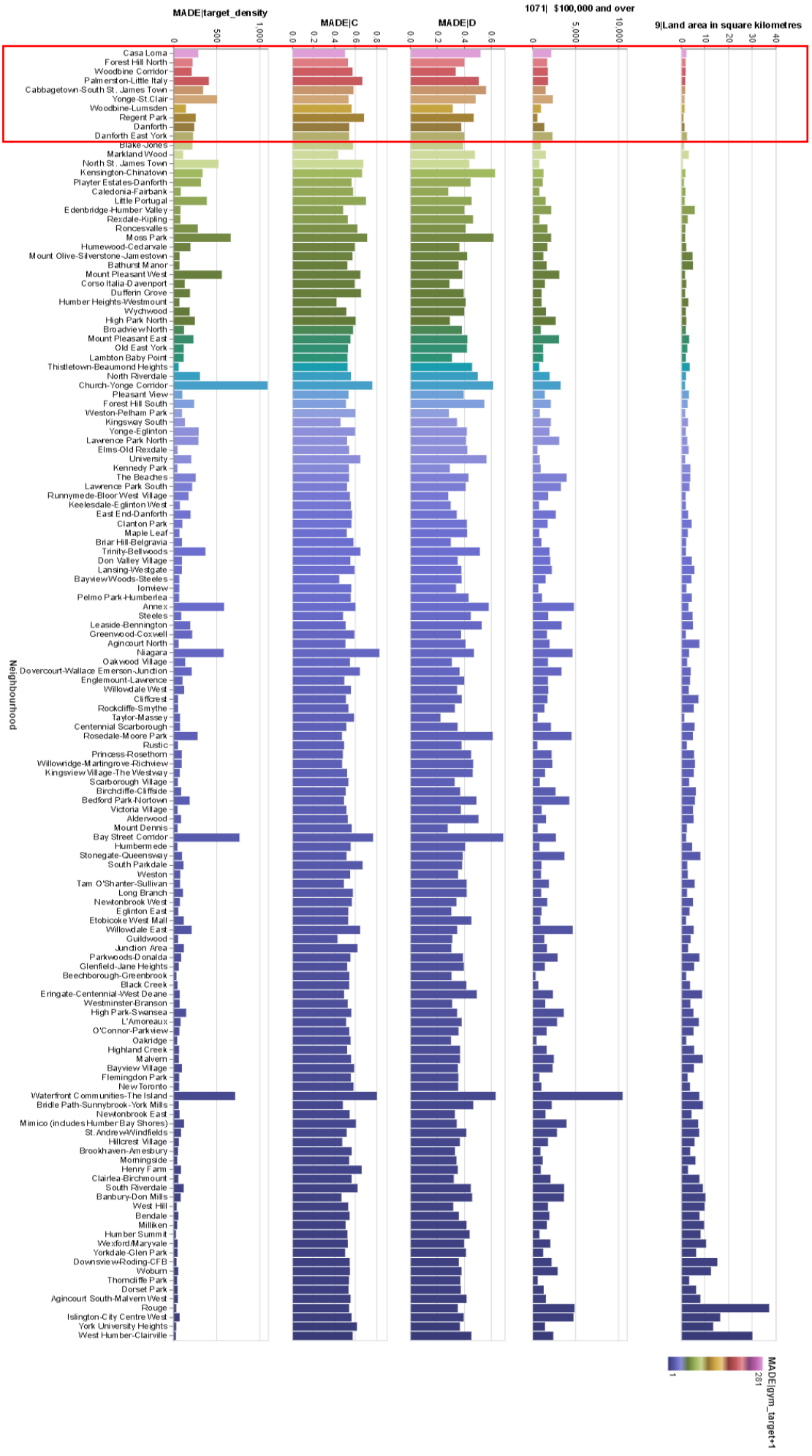
4. Results

4.1 Selected Neighborhoods:

The neighborhoods are sorted by the target indicator as you can see in the second column above:

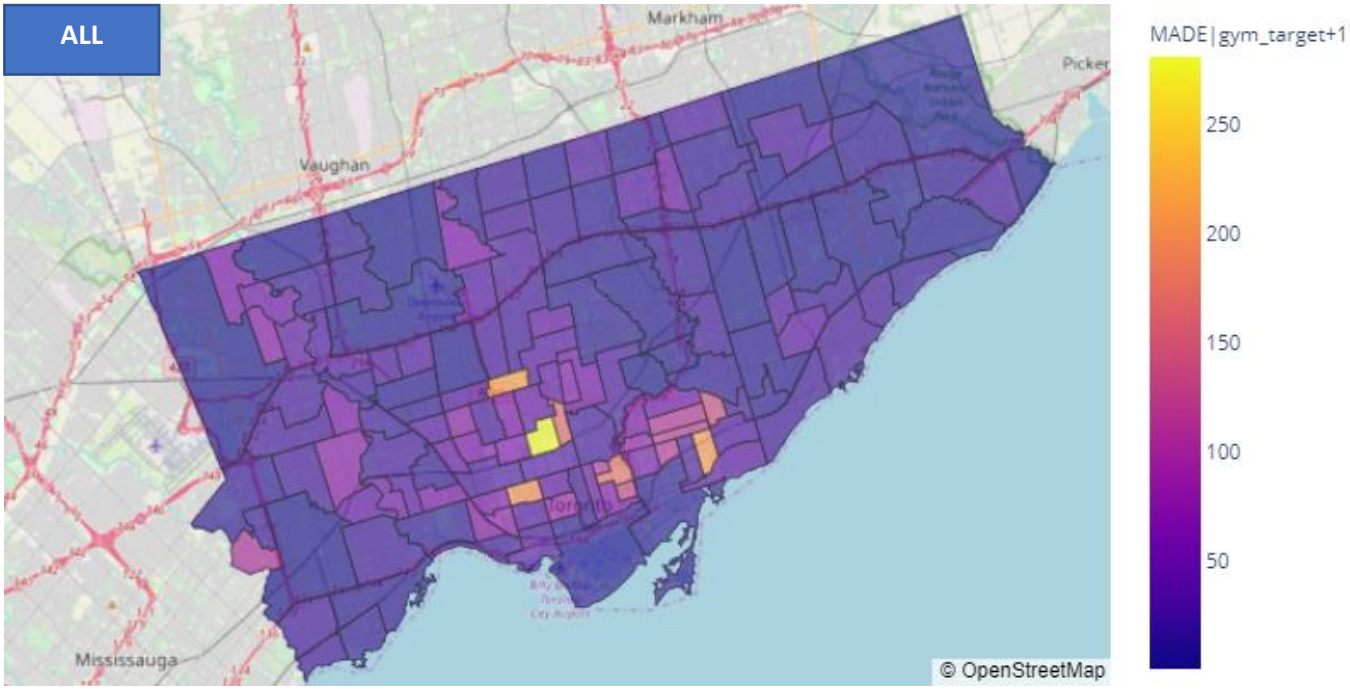
	Selected Neighbourhood	MADE|gym_target+1	MADE|target_density	n_gyms	MADE|C	MADE|D	9|Land area in square kilometres	1071| $100,000 and over
0	Casa Loma	280.75	280.75	0	0.5	0.52	1.93	2120
1	Forest Hill North	214.21	214.21	0	0.52	0.4	1.59	1640
2	Woodbine Corridor	203.47	203.47	0	0.57	0.33	1.6	1730
3	Palmerston-Little Italy	201.65	403.3	1	0.66	0.5	1.44	1750
4	Cabbagetown-South St. James Town	168.33	336.66	1	0.58	0.56	1.4	1470
5	Yonge-St.Clair	166.59	499.78	2	0.53	0.48	1.17	2305
6	Woodbine-Lumsden	136.85	136.85	0	0.56	0.31	1.17	925
7	Regent Park	125.39	250.77	1	0.68	0.47	0.64	510
8	Danforth	116.8	233.59	1	0.54	0.37	1.13	1320
9	Danforth East York	110.03	220.06	1	0.53	0.4	2.18	2265

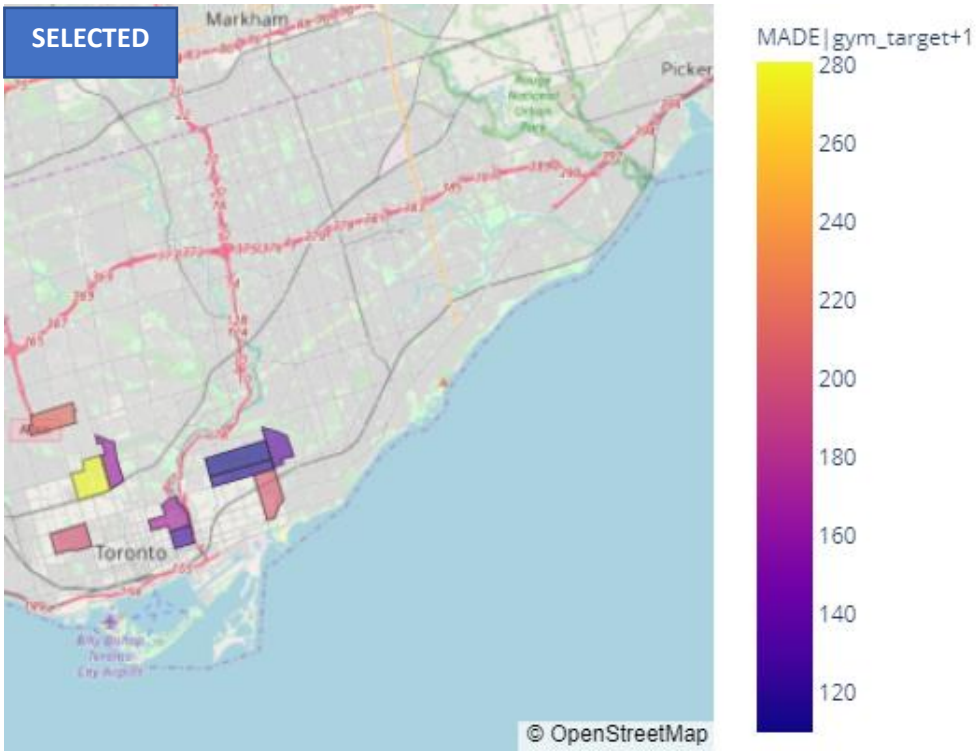
4.2 Graphical results: (Selected Neighborhoods inside the red retangle)


MADE|target_density
MADE|C
MADE|D
1071| $100,000 and over
9|Land area in square kilometres
Neighbourhood
Casa Loma
Forest Hill North
Woodbine Corridor
Palmerston-Little Italy
Cabbagetown-South St. James Town
Yonge-St.Clair
Woodbine-Lumsden
Regent Park
Danforth
Danforth East York
Blake-Jones
Markland Wood
North St. James Town
Kensington-Chinatown
Playter Estates-Danforth
Caledonia-Fairbank
Little Portugal
Edenbridge-Humber Valley
Rexdale-Kipling
Roncesvalles
Moss Park
Humewood-Cedarvale
Mount Olive-Silverstone-Jamestown
Bathurst Manor
Mount Pleasant West
Corso Italia-Davenport
Dufferin Grove
Humber Heights-Westmount
Wychwood
High Park North
Broadview North
Mount Pleasant East
Old East York
Lambton Baby Point
Thistletown-Beaumond Heights
North Riverdale
Church-Yonge Corridor
Pleasant View
Forest Hill South
Weston-Pelham Park
Kingsway South
Yonge-Eglinton
Lawrence Park North
Elms-Old Rexdale
University
Kennedy Park
The Beaches
Lawrence Park South
Runnymede-Bloor West Village
Keelesdale-Eglinton West
East End-Danforth
Clanton Park
Maple Leaf
Briar Hill-Belgravia
Trinity-Bellwoods
Don Valley Village
Lansing-Westgate
Bayview Woods-Steeles
Ionview
Pelmo Park-Humberlea
Annex
Steeles
Leaside-Bennington
Greenwood-Coxwell
Agincourt North
Niagara
Oakwood Village
Dovercourt-Wallace Emerson-Junction
Englemount-Lawrence
Willowdale West
Cliffcrest
Rockcliffe-Smythe
Taylor-Massey
Centennial Scarborough
Rosedale-Moore Park
Rustic
Princess-Rosethorn
Willowridge-Martingrove-Richview
Kingsview Village-The Westway
Scarborough Village
Birchcliffe-Cliffside
Bedford Park-Nortown
Victoria Village
Alderwood
Mount Dennis
Bay Street Corridor
Humbermede
Stonegate-Queensway
South Parkdale
Weston
Tam O'Shanter-Sullivan
Long Branch
Newtonbrook West
Eglinton East
Etobicoke West Mall
Willowdale East
Guildwood
Junction Area
Parkwoods-Donalda
Glenfield-Jane Heights
Beechborough-Greenbrook
Black Creek
Eringate-Centennial-West Deane
Westminster-Branson
High Park-Swansea
L'Amoreaux
O'Connor-Parkview
Oakridge
Highland Creek
Malvern
Bayview Village
Flemingdon Park
New Toronto
Waterfront Communities-The Island
Bridle Path-Sunnybrook-York Mills
Newtonbrook East
Mimico (includes Humber Bay Shores)
St.Andrew-Windfields
Hillcrest Village
Brookhaven-Amesbury
Morningside
Henry Farm
Clairlea-Birchmount
South Riverdale
Banbury-Don Mills
West Hill
Bendale
Milliken
Humber Summit
Wexford/Maryvale
Yorkdale-Glen Park
Downsview-Roding-CFB
Woburn
Thorncliffe Park
Dorset Park
Agincourt South-Malvern West
Rouge
Islington-City Centre West
York University Heights
West Humber-Clairville
MADE|gym_target+1
1
281

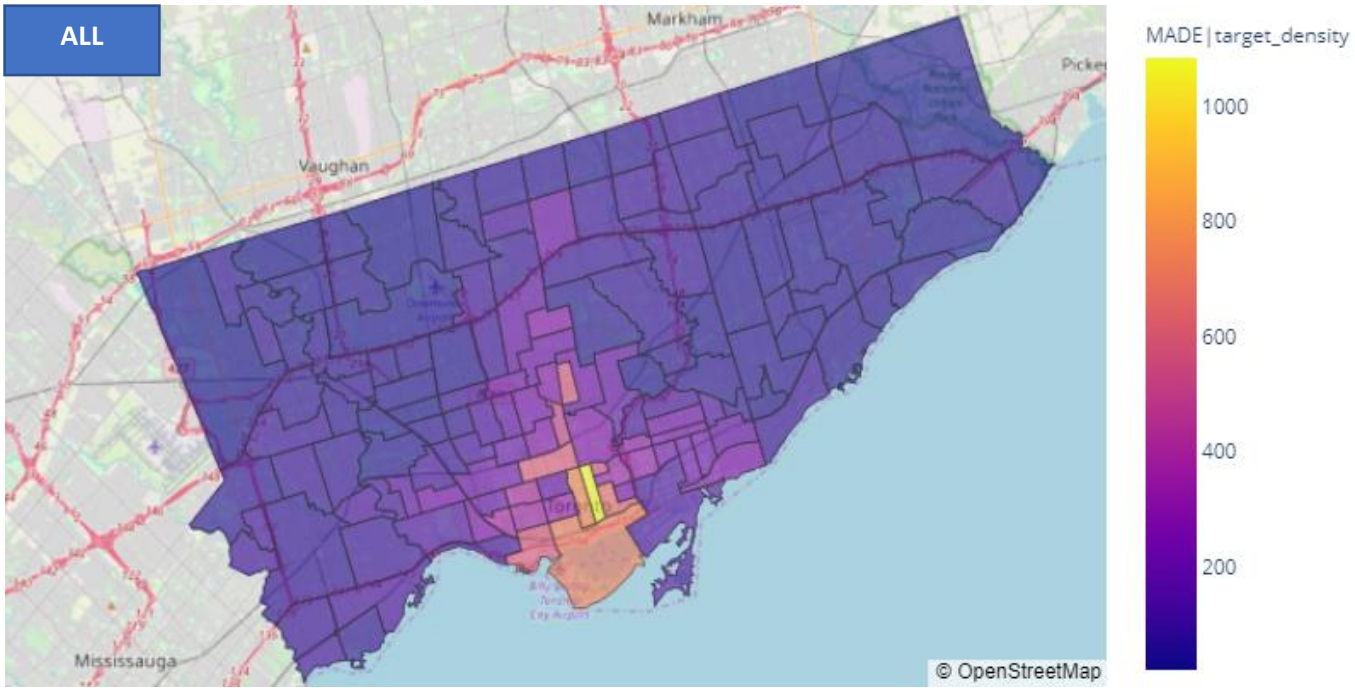
4.3 Choropleth Map results:

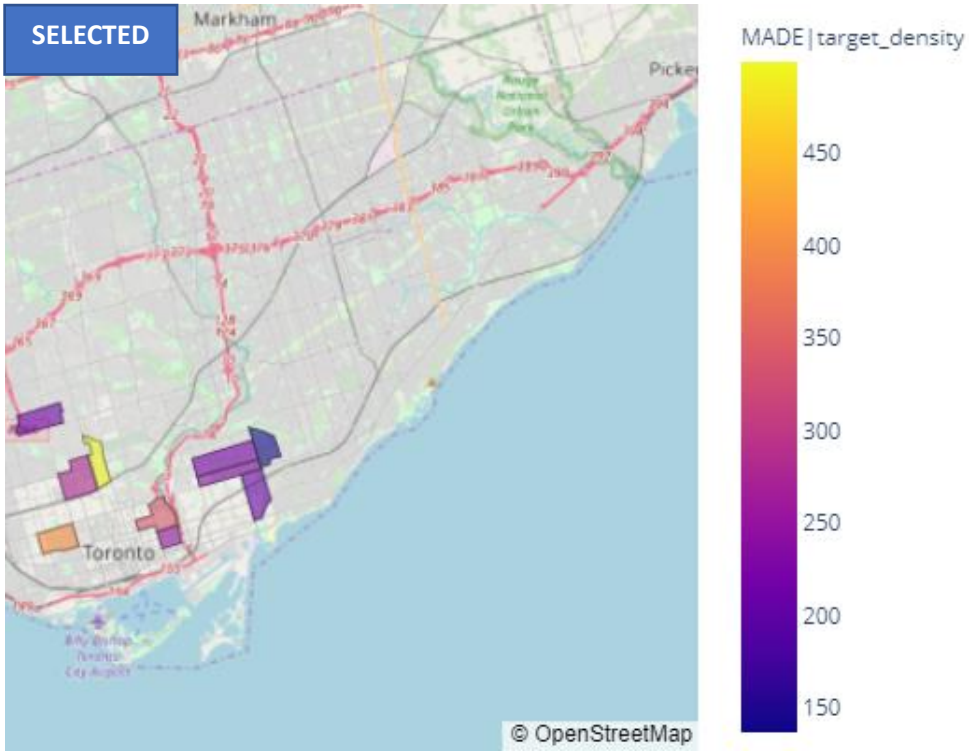
GYM TARGET INDICATOR:



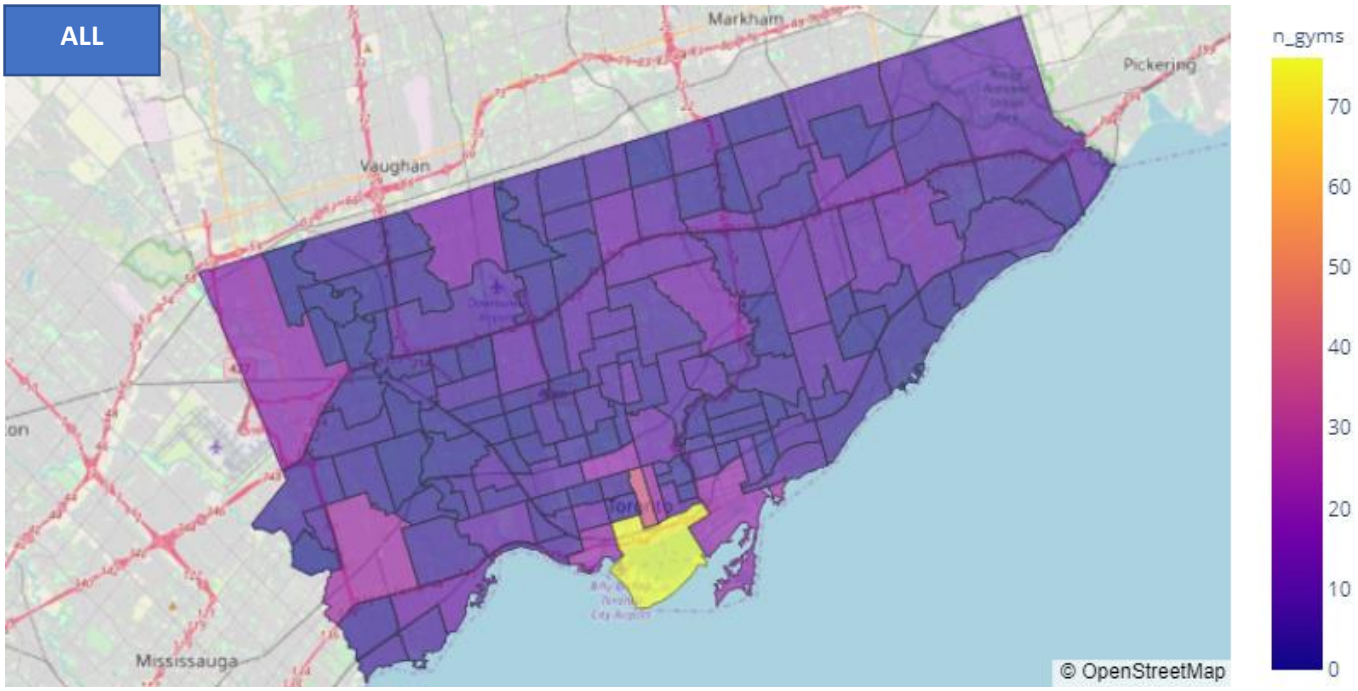


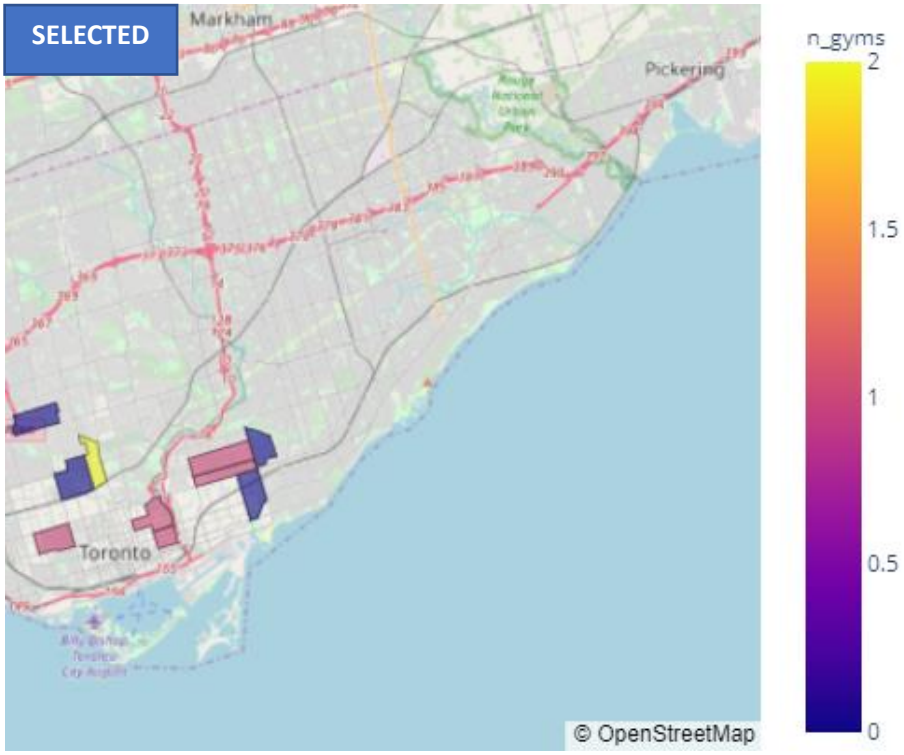
POPULATION TARGET DENSITY





NUMBER OF GYMS

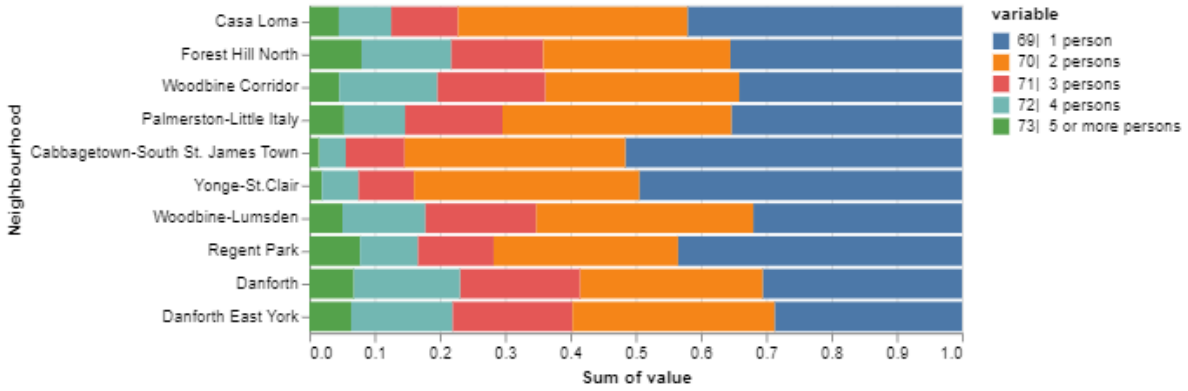




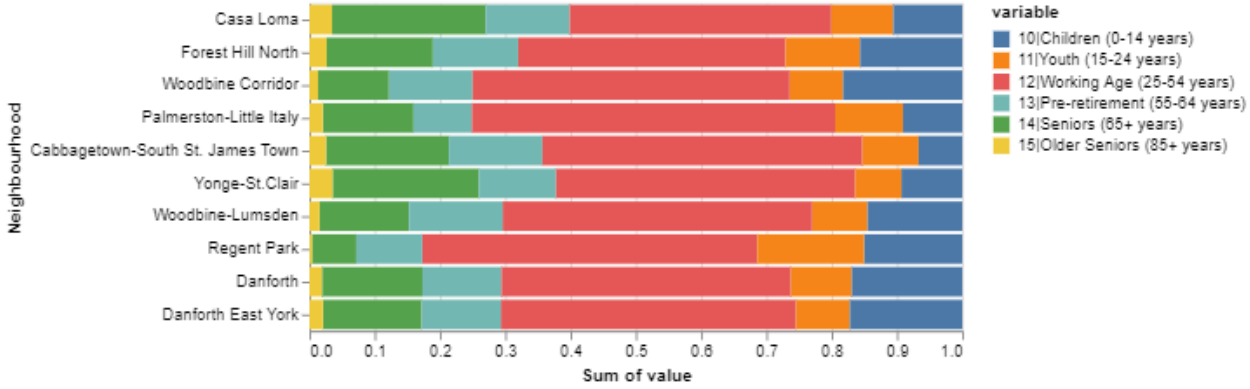
4.4 General Data about the selected Neighborhoods:

The bars represent the percentual of residents in each category.

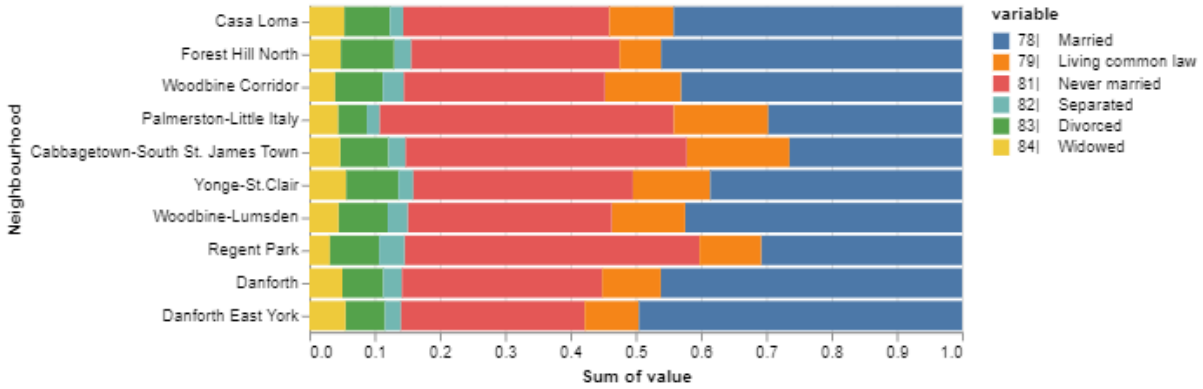
Private households by household size



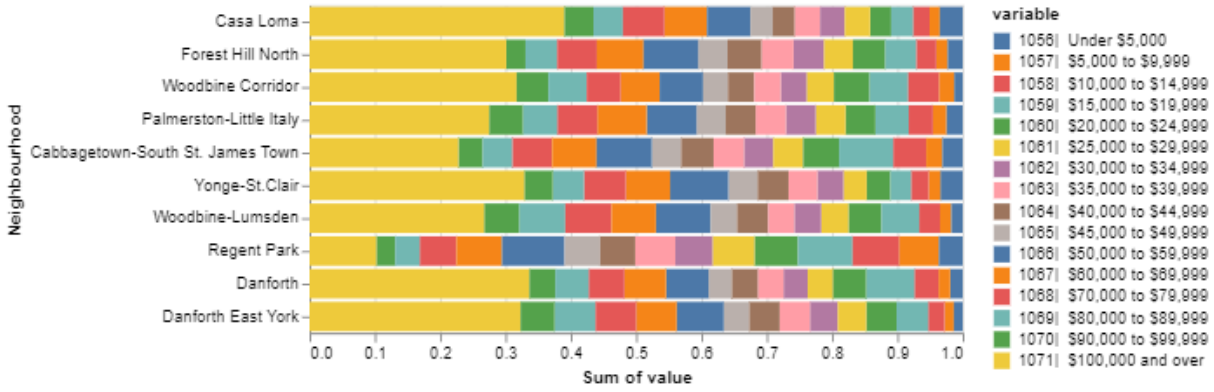
Population age composition



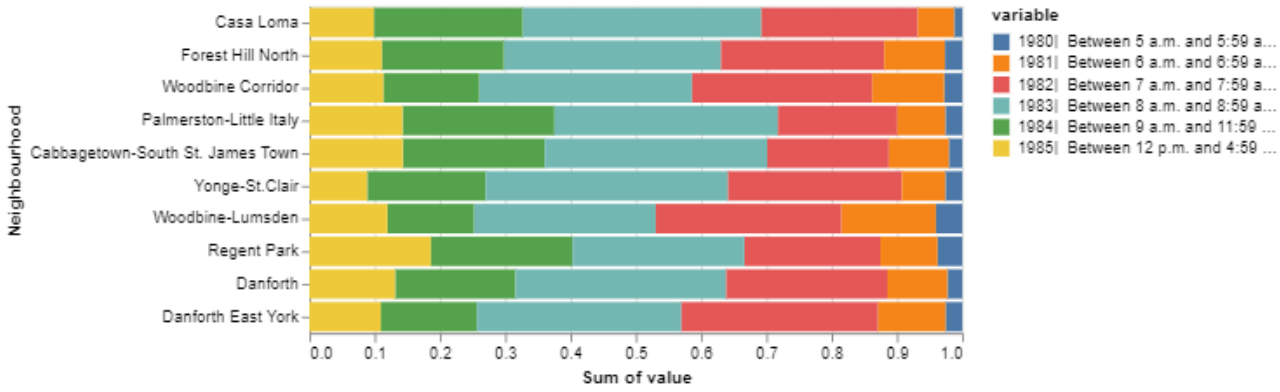
Marital status for the population aged 15 years and over



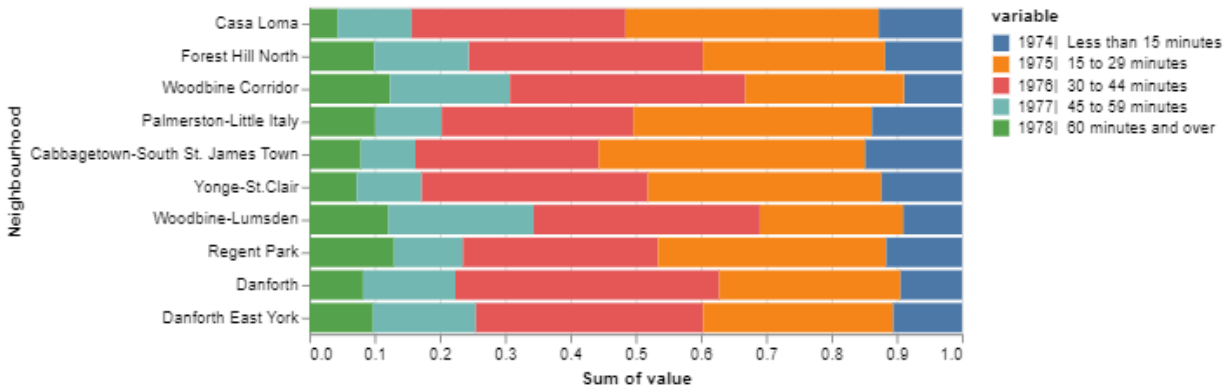
Income of households in 2015 (after tax)



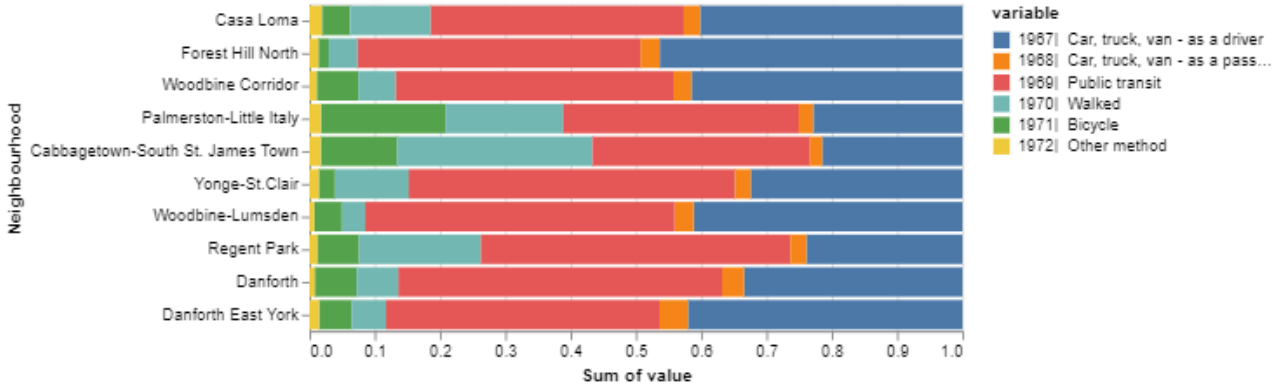
Time leaving for work for the employed labour force



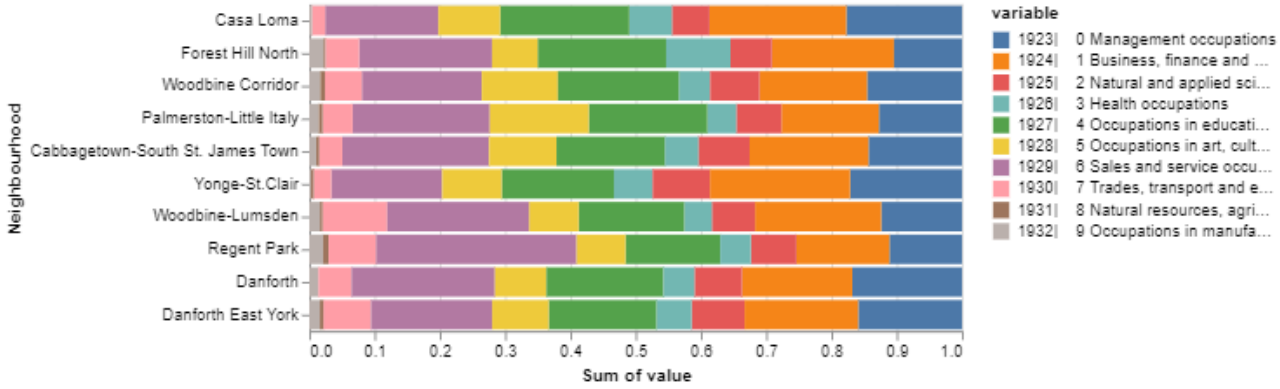
Commuting duration for the employed labour force



Main mode of commuting for the employed labour force



All occupations



5 Conclusion and Future Directions:

The required ten neighborhoods were selected for further field surveys.

It can be useful to run some supervised machine learning algorithms to understand the Profile's variable importance in the number of gyms for helping to feed back the analysis.