



SELECTING CANDIDATE NEIGHBORHOODS

IBM APPLIED DATA SCIENCE CAPSTONE

MARCEL MENDES REIS

SCRIPT

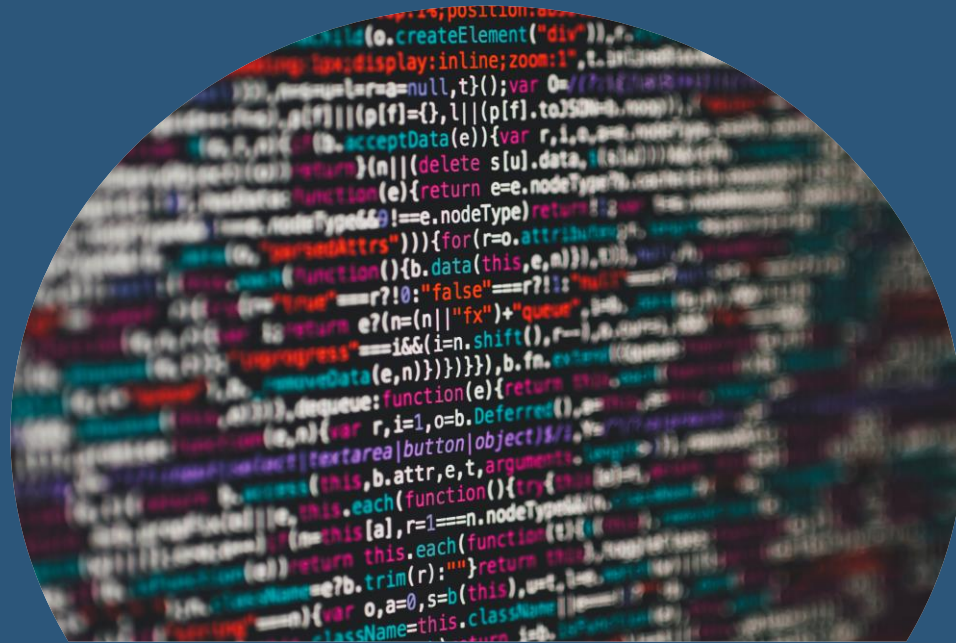
1. Description of the Business Problem & Discussion of the Background
2. Data acquisition
3. Methodology, Analysis and Results



DESCRIPTION OF THE BUSINESS PROBLEM & DISCUSSION OF THE BACKGROUND

DESCRIPTION OF THE BUSINESS PROBLEM & DISCUSSION OF THE BACKGROUND

- A Luxury Fitness Club chain wants to expand its business and wants research about potential neighborhoods in TORONTO - CA for further field surveys, 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. As well as to get an overview profile about the potential neighborhoods.



DATA ACQUISITION AND CLEANING

DATA ACQUISITION AND CLEANING

1) Neighborhoods' Demographics

- The data is sourced from a number of Census tables released by Statistics Canada.
- In these profiles, "neighbourhood" refers to the City of Toronto's 140 social planning neighbourhoods.
- I downloaded all available data, but to calc the target indicator I'm going to use just the following:

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

Neighbourhood Information:	Population and dwellings:	Income of households in 2015:	Age characteristics:	Commuting duration:
1 - Neighbourhood Number	3 - Population, 2016 9 - Land area in square kilometres	1055 - Total - Household after-tax income groups in 2015 for private households - 100% data 1056 - Under \$5,000 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	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)	1973 - Total - Commuting duration for the employed labour 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

DATA ACQUISITION AND CLEANING

1) Neighborhoods' Demographics

Private households by household size	Marital status for the population aged 15 years and over	Time leaving for work for the employed labour force aged 15 years and over	Main mode of commuting for the employed labour force aged 15 years and over	All occupations
69 – 1 person 70 – 2 persons 71 – 3 persons 72 – 4 persons 73 – 5 or more persons	78 – Married 79 – Living common law 81 – Never married 82 – Separated 83 – Divorced 84 – Widowed	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.	1967 - Car, truck, van - as a driver 1968 - Car, truck, van - as a passenger 1969 - Public transit 1970 - Walked 1971 - Bicycle 1972 - Other method	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

DATA ACQUISITION AND CLEANING

2) Neighborhoods' Geographical Coordinates

- 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

DATA ACQUISITION AND CLEANING

3) Information about Gyms

- To get location and other information about the gyms in Toronto I used the Foursquare's explore API. Using the Foursquare's explore API (which gives venue recommendations).
- 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.



METHODOLOGY, ANALYSIS AND RESULTS

HOW WOULD THEY SELECT THE POTENTIAL NEIGHBORHOODS?

GYM TARGET INDICATOR

We created an indicator which represents the number of target people per one unit of gym per one square kilometer. This indicator will be applied to each neighborhood and will be the final criteria to select the neighborhoods.

$$\text{gym_target}+1 = E / ((A \times 1/B) \times (C \times D))$$

where:

E - number of 'Gym' and 'Gym / Fitness Center' inside the neighborhood, 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.

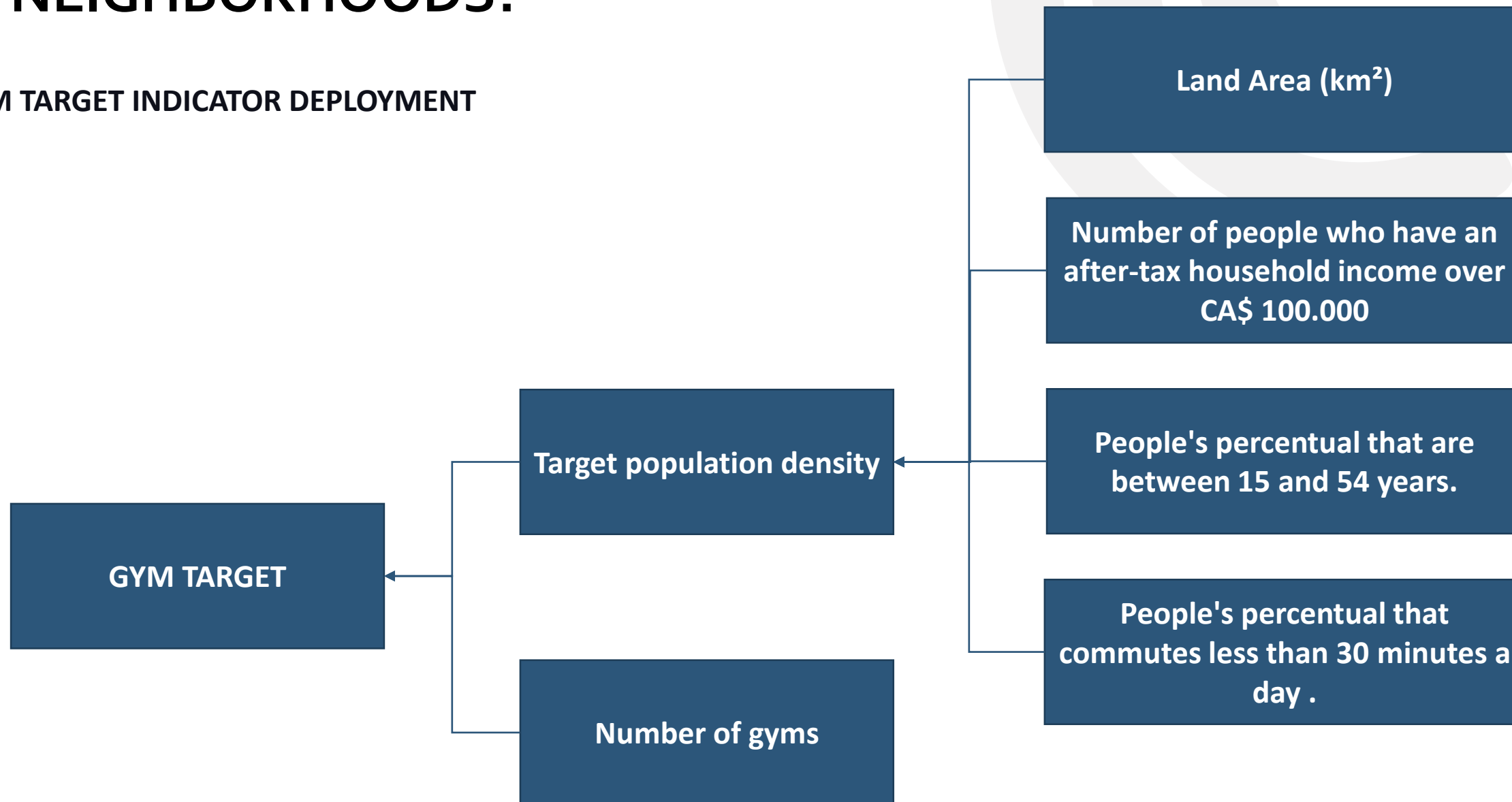
C = People's percentual that are between 15 and 54 years.

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

The higher the better!

HOW WOULD THEY SELECT THE POTENTIAL NEIGHBORHOODS?

GYM TARGET INDICATOR DEPLOYMENT

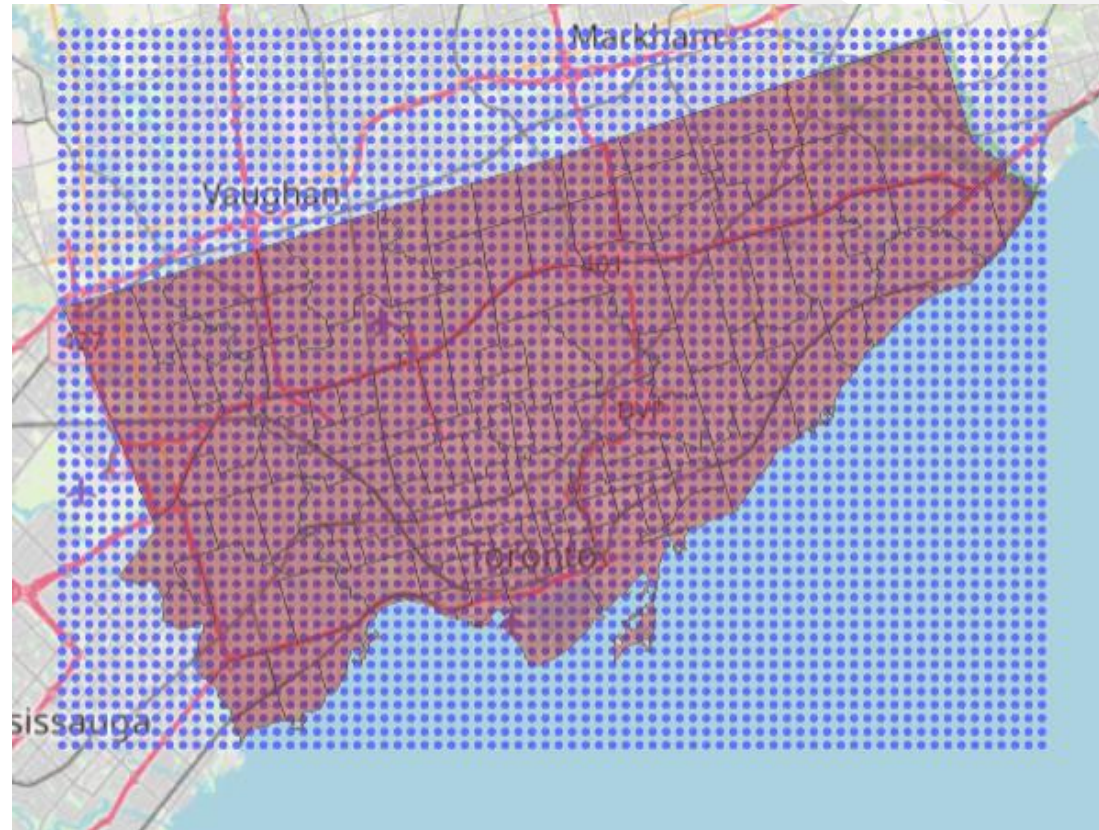


HOW WOULD THEY COLLECT AND ASSIGN EACH VENUE TO THE CORRECT NEIGHBORHOOD?

GRID GENERATOR

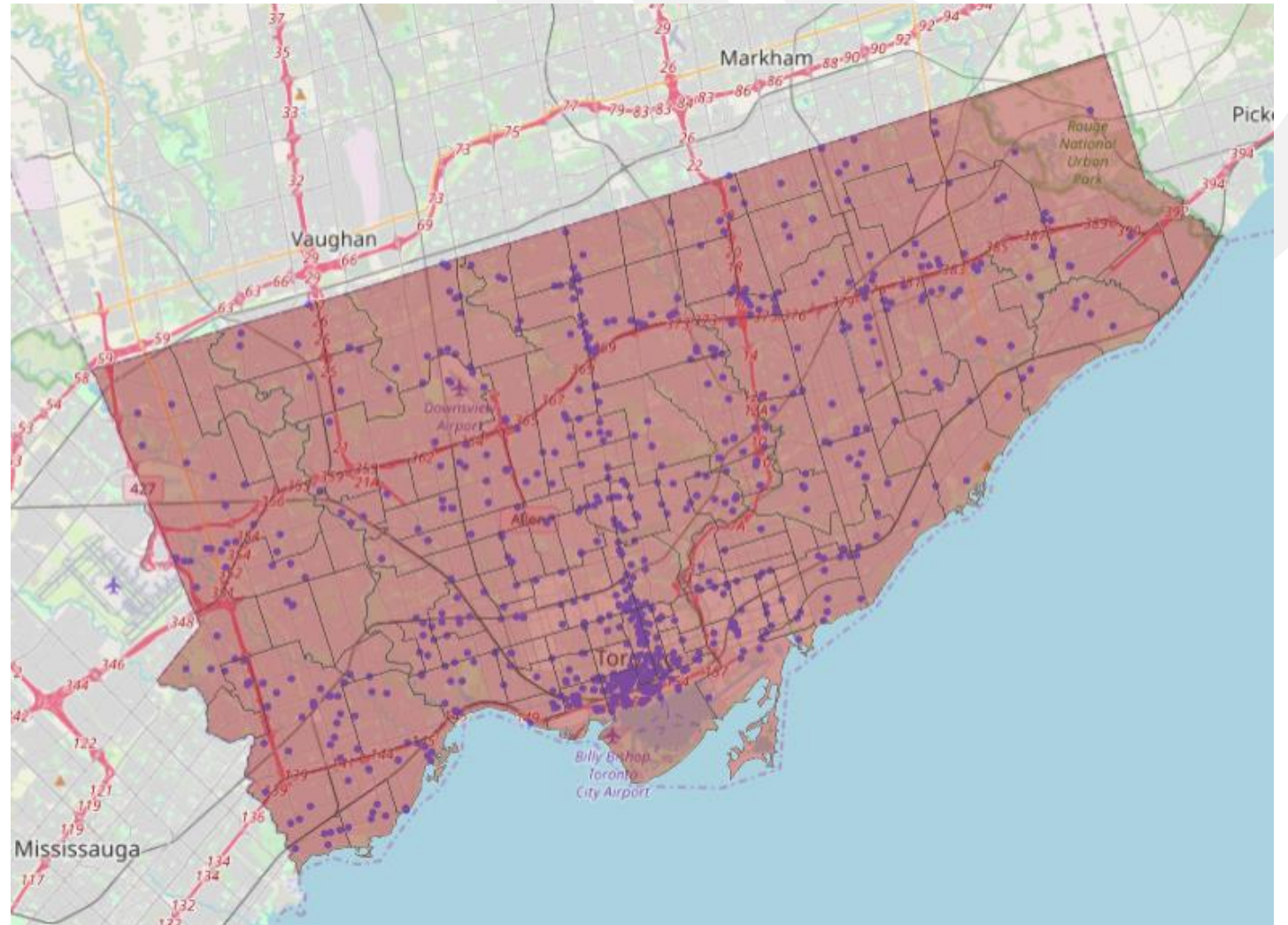
With Toronto's geojson shapefile, I got the shape bounds and generated 3996 equally spaced points of which, we will scan the GYMS in a radius of 1000 meters using the foursquare API.

GENERATED GRID PLOT

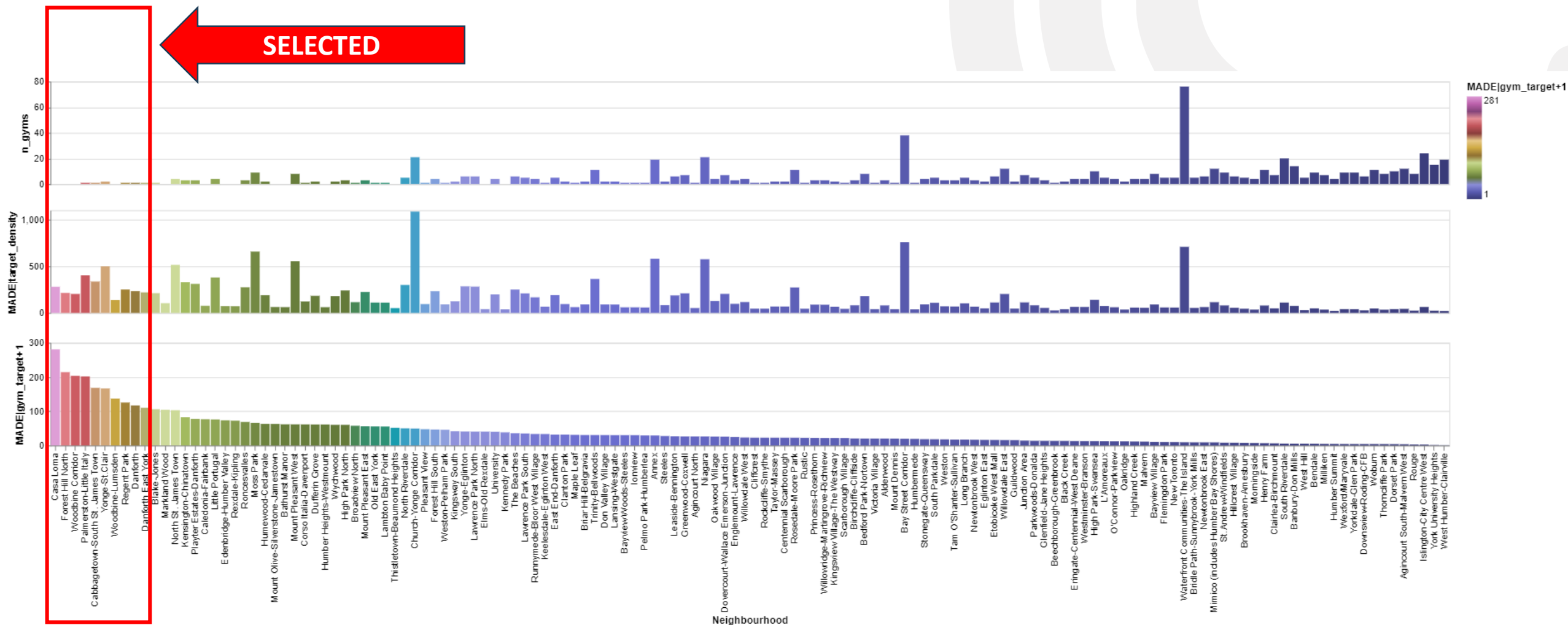


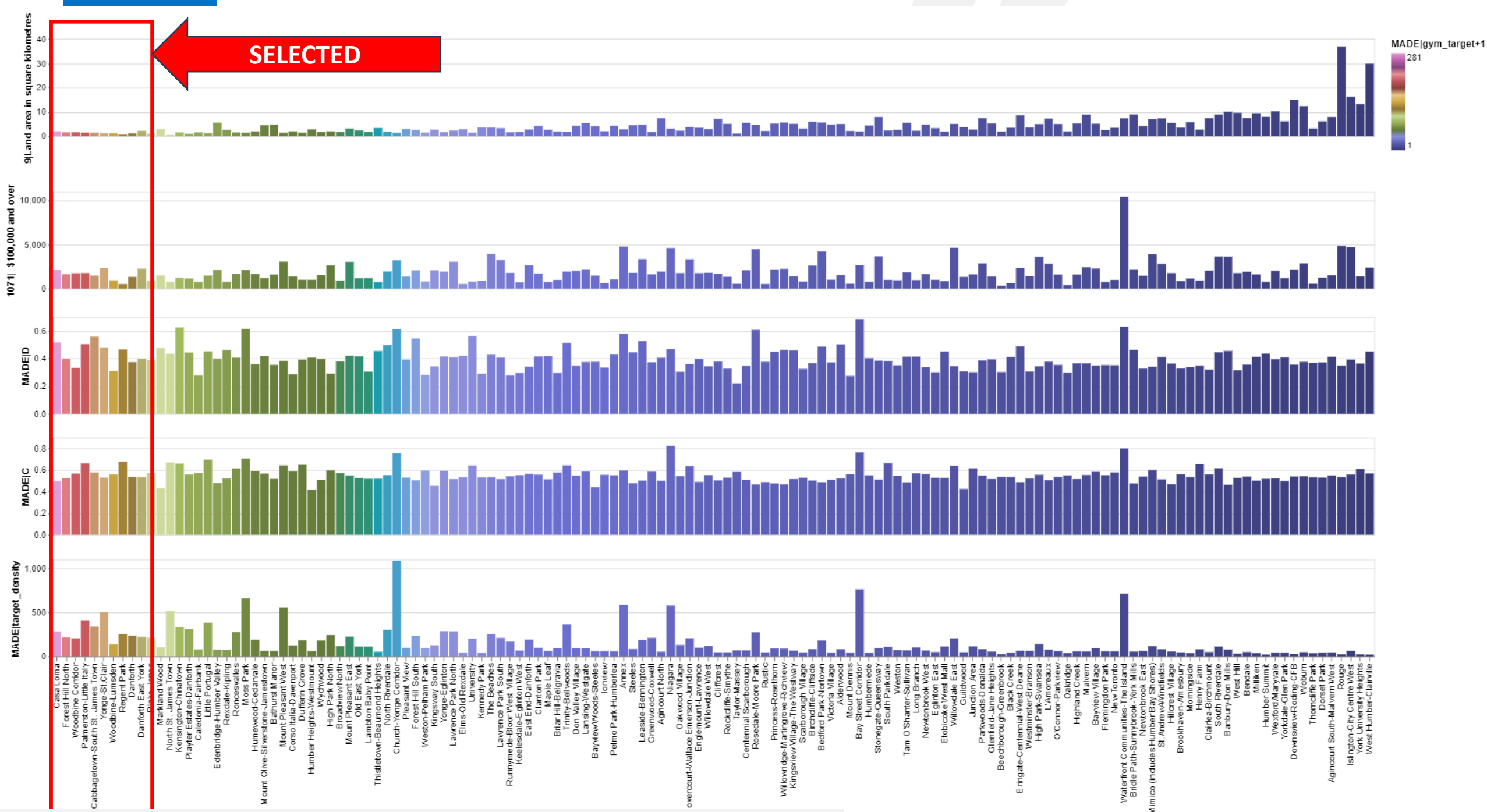
HOW WOULD THEY COLLECT AND ASSIGN EACH VENUE TO THE CORRECT NEIGHBORHOOD?

GYM'S LOCATION



WHAT ARE THE RESULTS?





WHAT ARE THE RESULTS?

SELECTED NEIGHBORHOODS

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

MIN

MAX

source:

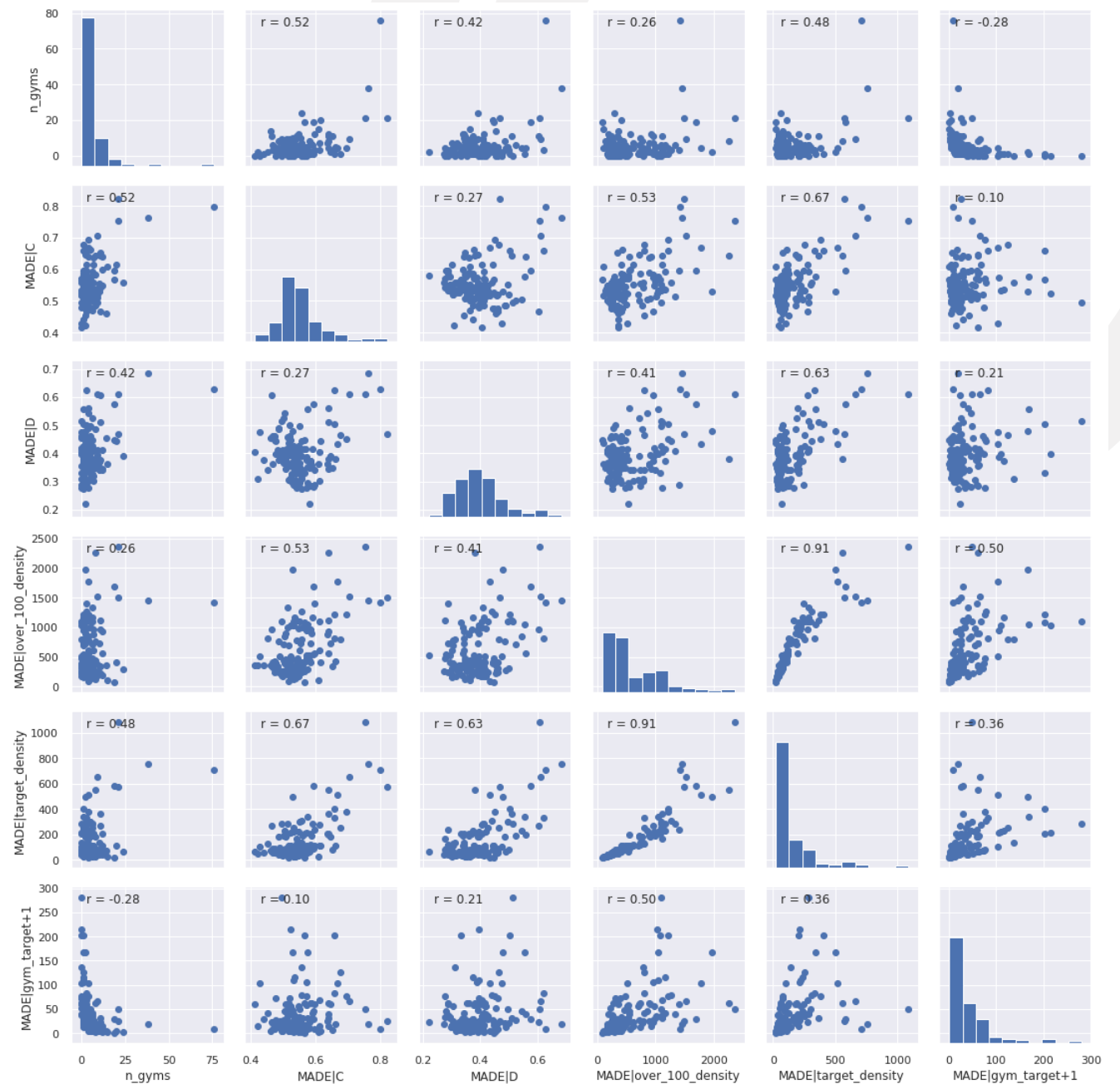
<https://open.toronto.ca/dataset/neighbourhood-profiles/>

<https://open.toronto.ca/dataset/neighbourhoods/%22>

FOURSQUARE API

WHAT ARE THE RESULTS?

VARIABLES PAIR PLOT



source:

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

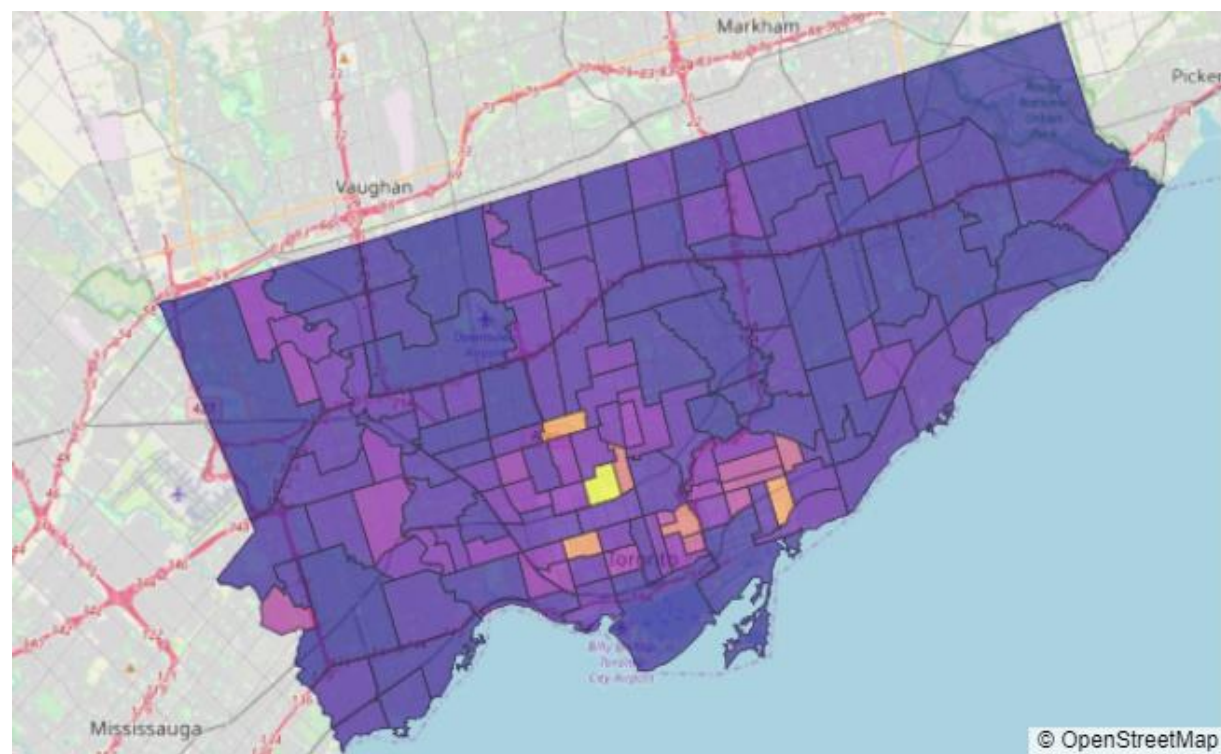
<https://open.toronto.ca/dataset/neighbourhoods/%22>

FOURSQUARE API

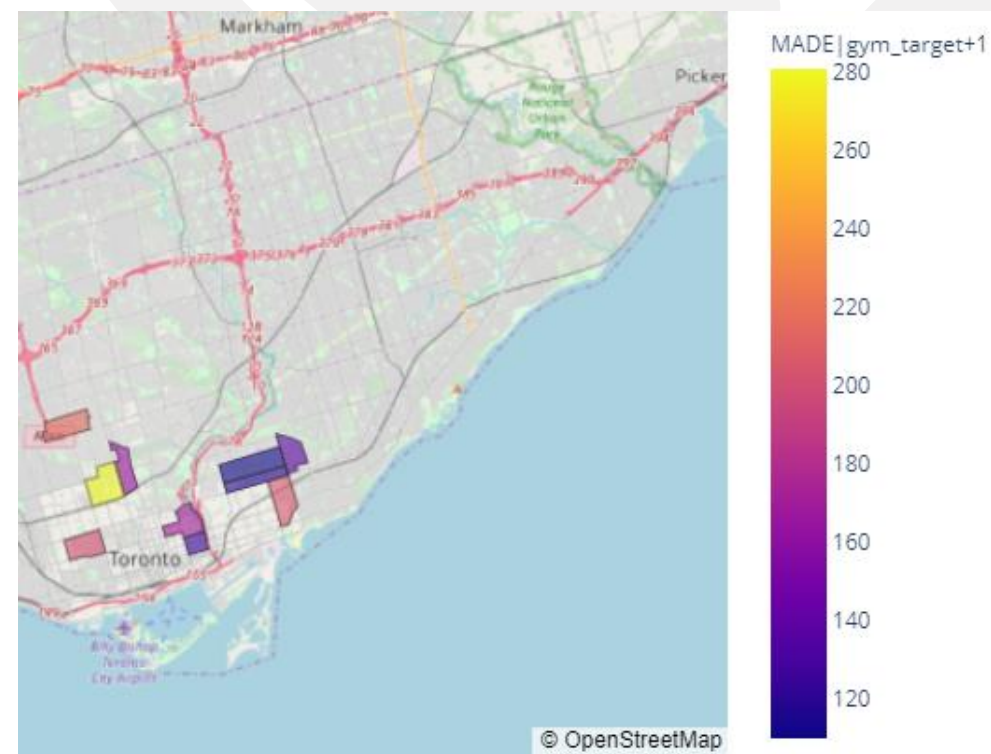
WHAT ARE THE RESULTS?

GYM TARGET INDICATOR - choropleth

ALL NEIGHBORHOODS



SELECTED NEIGHBORHOODS



source:

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

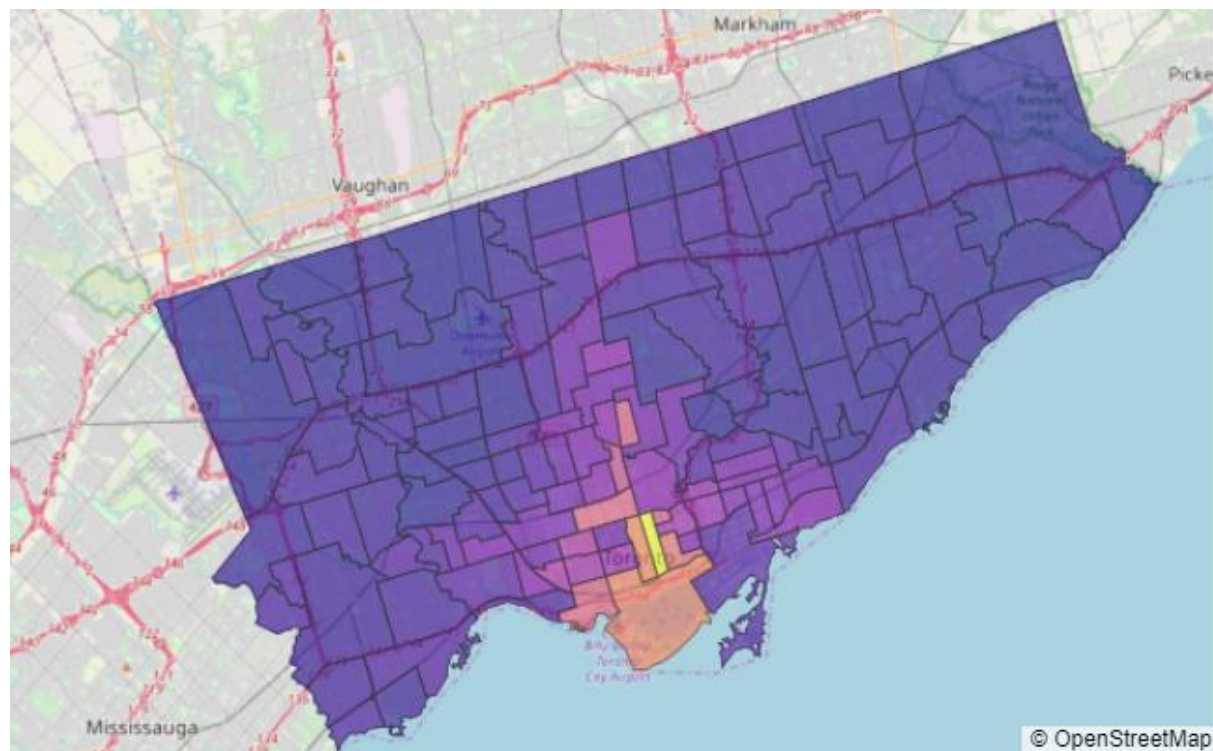
<https://open.toronto.ca/dataset/neighbourhoods/%22>

FOURSQUARE API

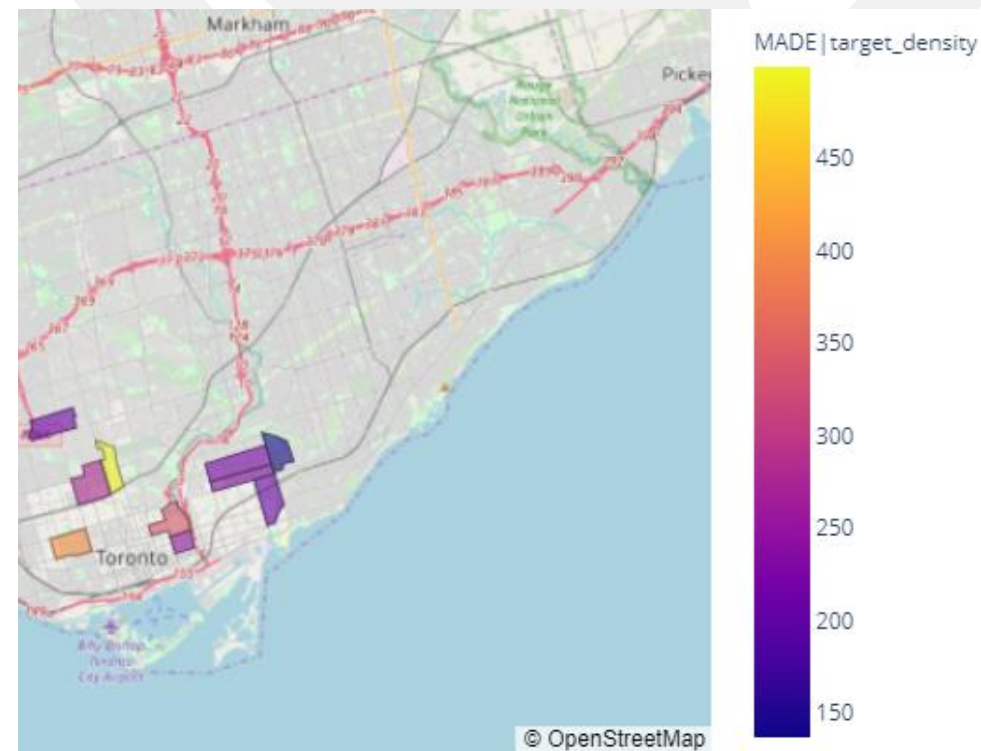
WHAT ARE THE RESULTS?

POPULATION TARGET DENSITY - choropleth

ALL NEIGHBORHOODS



SELECTED NEIGHBORHOODS



source:

<https://open.toronto.ca/dataset/neighbourhood-profiles/>

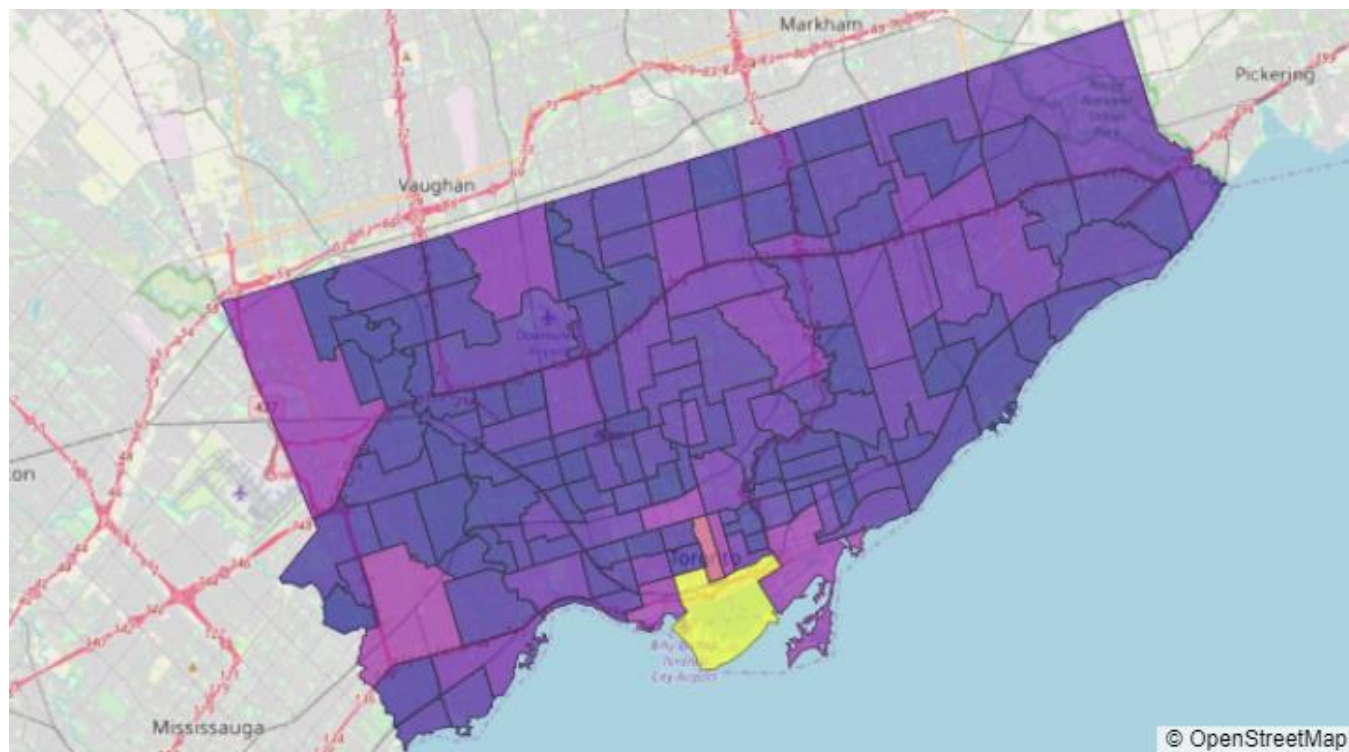
<https://open.toronto.ca/dataset/neighbourhoods/%22>

FOURSQUARE API

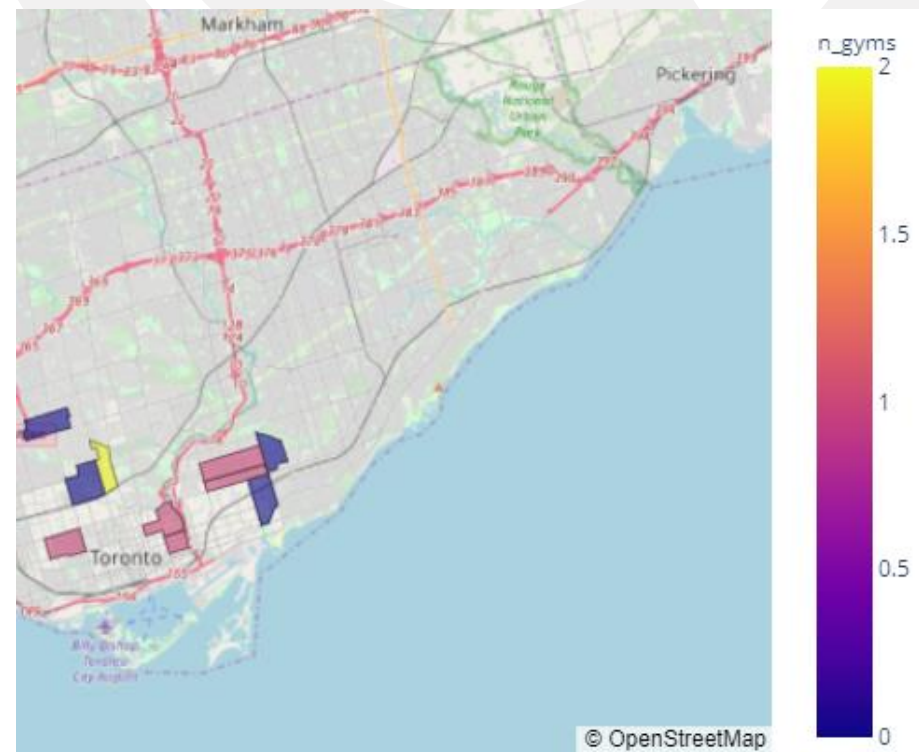
WHAT ARE THE RESULTS?

NUMBER OF GYMS - choropleth

ALL NEIGHBORHOODS



SELECTED NEIGHBORHOODS



source:

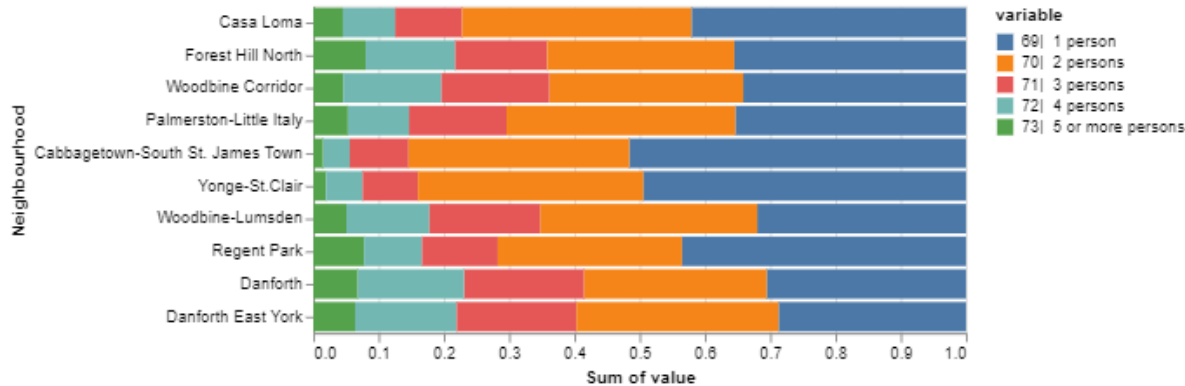
<https://open.toronto.ca/dataset/neighbourhood-profiles/>

<https://open.toronto.ca/dataset/neighbourhoods/%22>

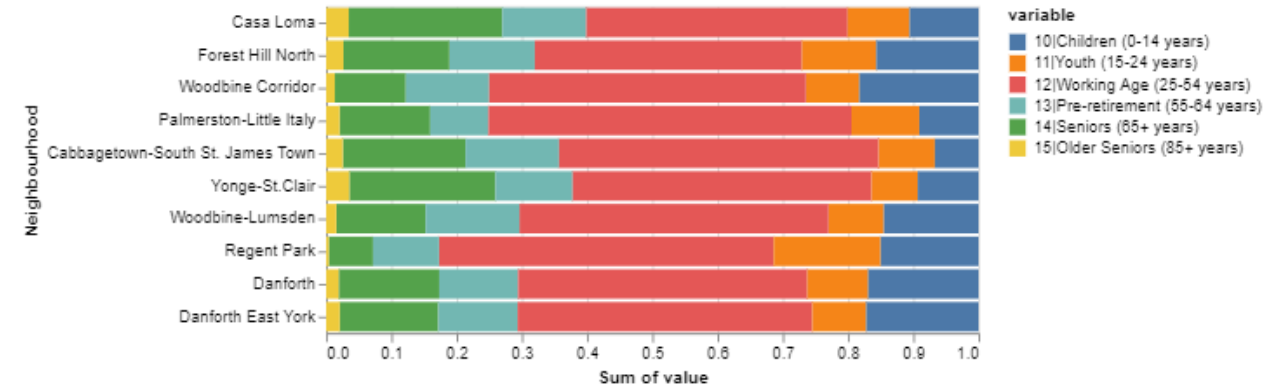
FOURSQUARE API

NOW THAT WE KNOW THE TARGET NEIGHBORHOODS, LET'S TAKE A LOOK ON SOME INTERESTING DEMOGRAPHIC DATA!

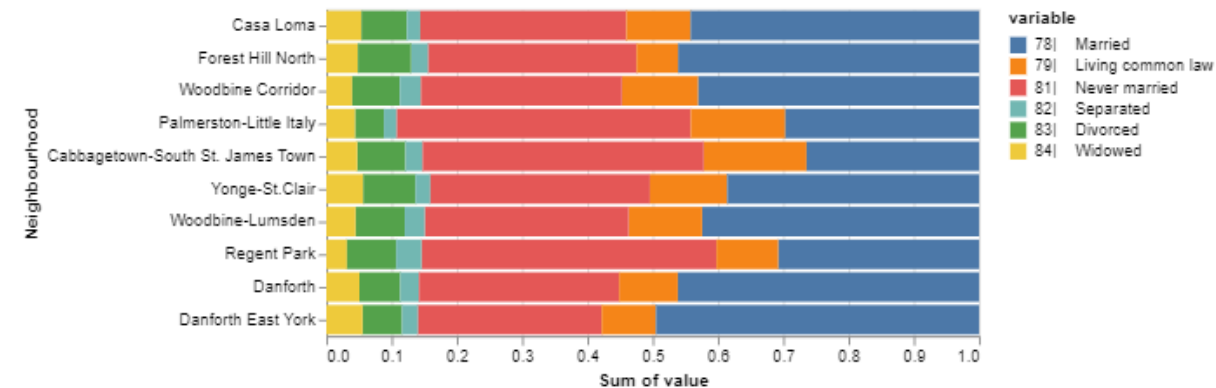
Private households by household size



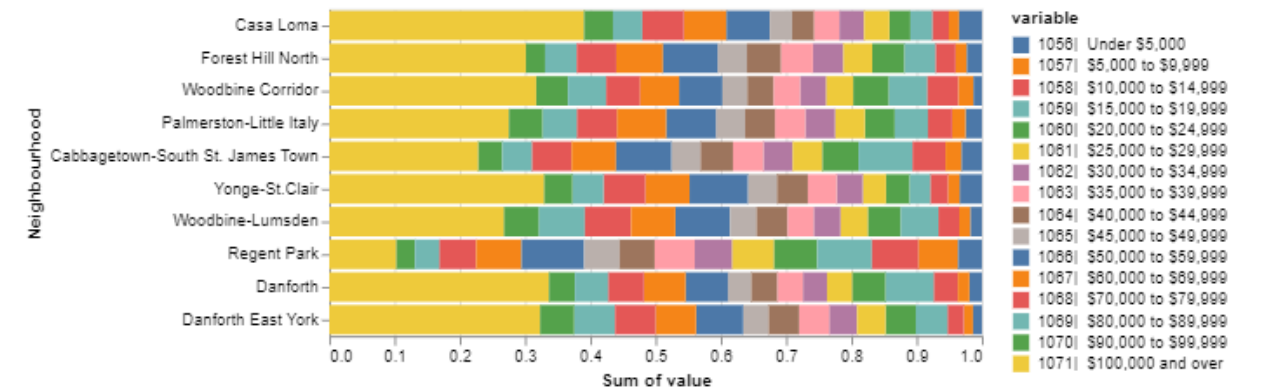
Population age composition



Marital status for the population aged 15 years and over

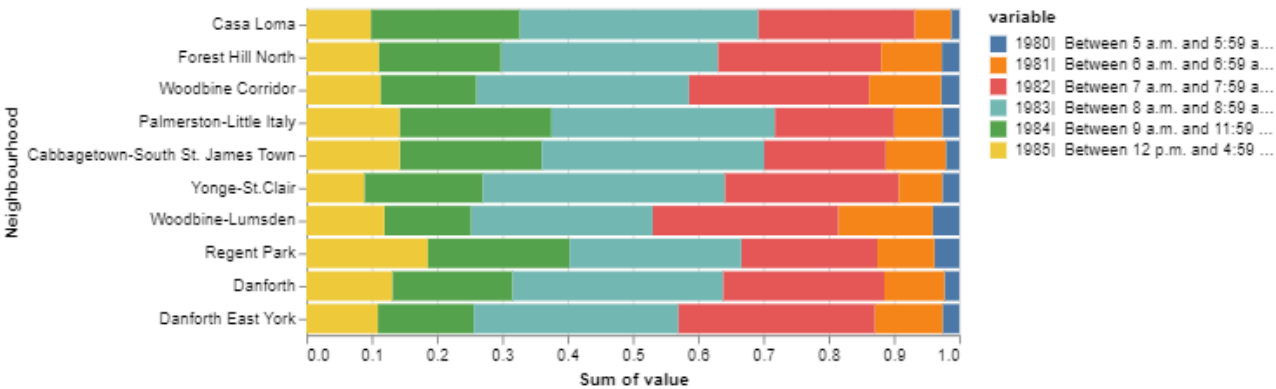


Income of households in 2015 (after tax)

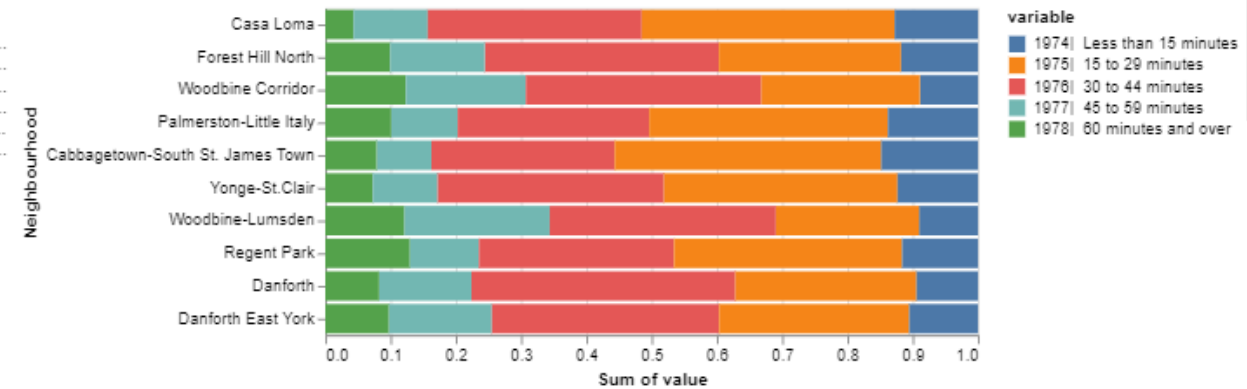


NOW THAT WE KNOW THE TARGET NEIGHBORHOODS, LET'S TAKE A LOOK ON SOME INTERESTING DEMOGRAPHIC DATA!

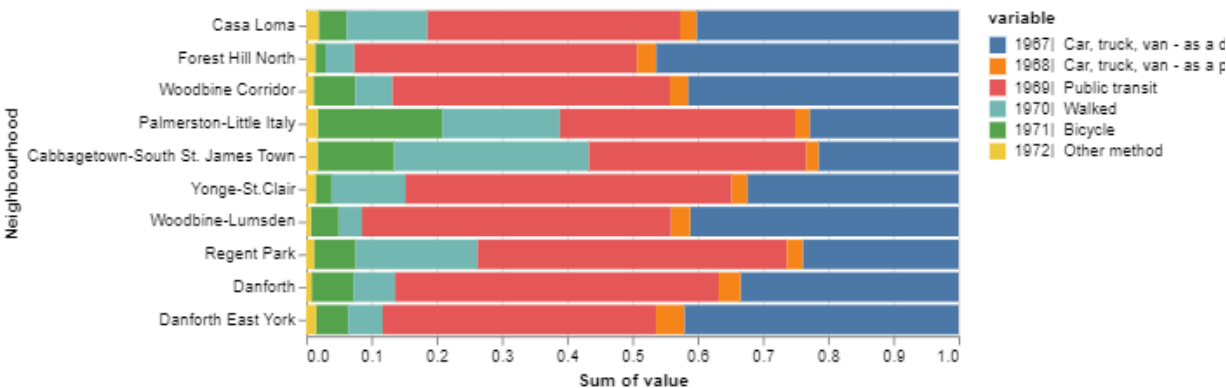
Time leaving for work for the employed labour force



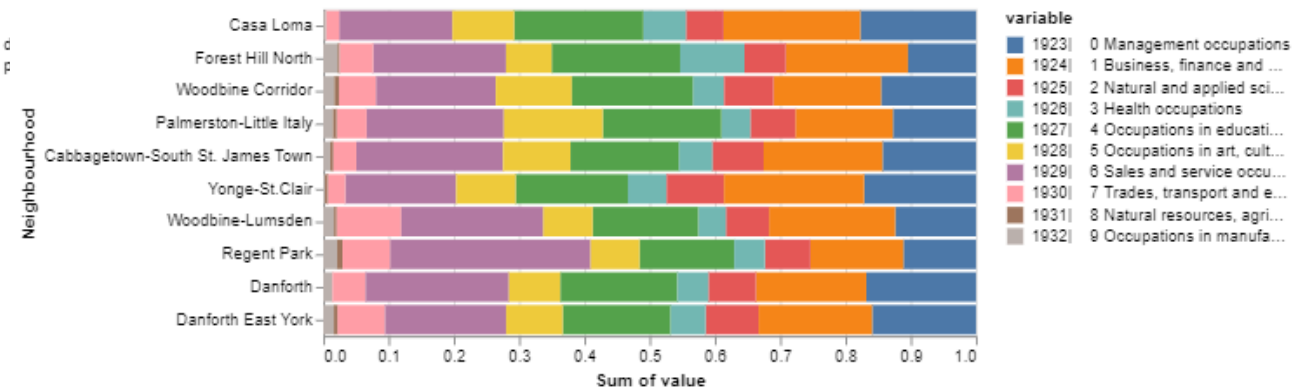
Commuting duration for the employed labour force

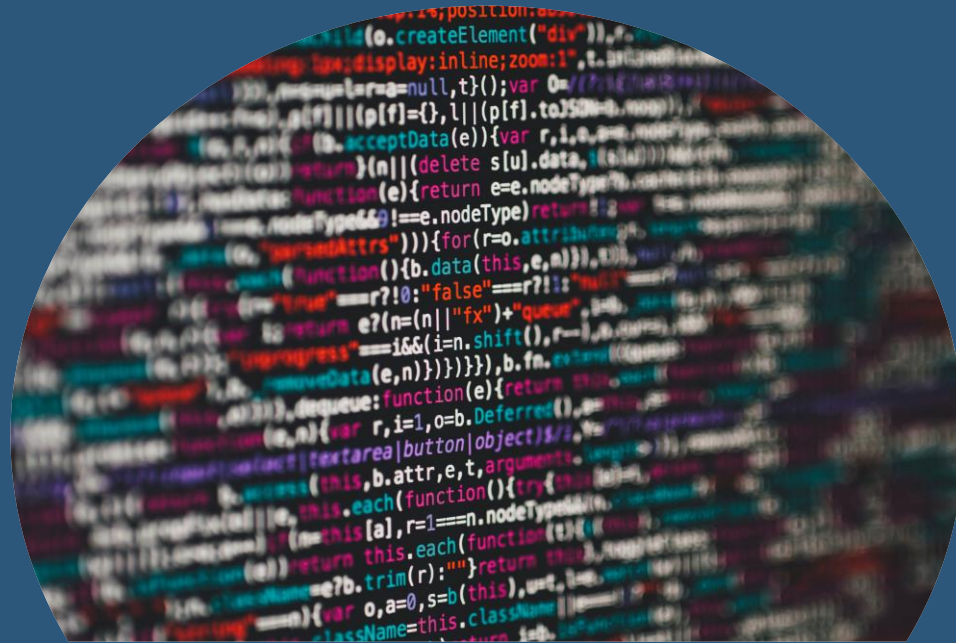


Main mode of commuting for the employed labour force



All occupations





CONCLUSION AND FUTURE DIRECTIONS

CONCLUSION AND FUTURE DIRECTIONS

- 10 neighborhoods were selected
- 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.



THANK YOU



MMREIS@YAHOO.COM.BR



[HTTPS://WWW.LINKEDIN.COM/IN/MARCEL-MENDES-REIS-1780668/](https://www.linkedin.com/in/marcel-mendes-reis-1780668/)