

# Recommending Areas for Gymnasium in Toronto

Fahida V P

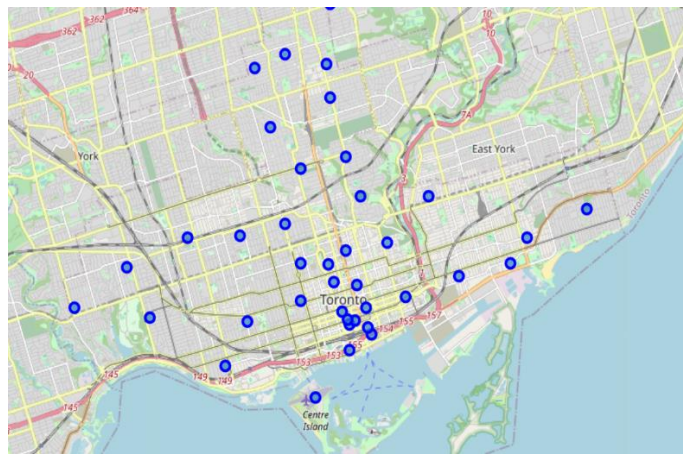
15 December 2020

## 1. Introduction

### 1.1 Background

Toronto is a global city filled with vast opportunity and is home to an array of distinctive and dynamic neighbourhoods that reflect the diversity of its population. Toronto's growing population coincides with increased development and investment in the city and surrounding region. That is why the Toronto area is an excellent place to grow your business.

The project aims at finding best places in Toronto to start up a Gymnasium. Gym, a shortened form of gymnasium, refers to facilities intended for indoor sports and exercise. Joining a gym might be one of the best investments someone can make for your health and nowadays one of the popular fitness centres around the world.



**Map of Toronto and its neighbourhoods**

### 1.2 Problem

Finding a best place to start a new venture is really challenging for the management. Sometimes it may be because of the existence vast number of similar ventures or because of the lack of information about the population or about the nature of people living in that area. In this project, our problem is to build a recommendation system for suggesting optimal locations for a new gymnasium in the neighbourhoods of Toronto, Canada.

### 1.3 Interest

People who are planning to start a new gymnasium would be extremely interested in getting an idea about optimal locations for competitive advantages and business values. If one of their start up venture become successful, then they can extend their business to similar areas in future.

## 2.Data Acquisition and Cleaning

### 2.1 Data Sources

Based on definition of our problem, factors that will influence our decision include, number of existing gymnasiums in the neighbourhoods, population in the neighbourhood, population density, average income etc. Toronto neighbourhood data and population information could be obtained by scraping Wikipedia. That is from this [link](#) and this [link](#).

Nearby venues of each neighbourhood can be obtained from [www.foursquare.com](http://www.foursquare.com). Thus, we can filter out the number of gymnasiums in every neighbourhoods from the received data.

### 2.2 Data Cleaning

Toronto neighbourhood data scraped contains some neighbourhood's data missing and I decided to drop it. Another problem was multiple entries for single neighbourhood field, so I separated it into multiple fields. Collected latitude and longitude information of each neighbourhoods and merged both together.

Population related data was also scraped and merged with the above dataset. After this step dataset looked like this:

	Borough	Neighborhood	Latitude	Longitude	Population	Land area (km2)	Average Income
0	North York	Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0
1	North York	Victoria Village	43.725882	-79.315572	17047.0	4.72	29657.0
2	North York	Lawrence Manor	43.718518	-79.464763	13750.0	2.14	36361.0
3	Scarborough	Malvern	43.806686	-79.194353	44324.0	8.86	25677.0
4	North York	Don Mills	43.745906	-79.352188	21372.0	8.99	47515.0

Next step is to find nearby venues of each neighbourhood from foursquare API. By sending location details (latitude, longitude) we will get venues closer to each neighbourhood and add that data with the above table. So, our final table looks as:

Neighborhood	Latitude	Longitude	Population	Land area (km2)	Average Income	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0	43.753259	-79.329656	Donalda Golf & Country Club	43.752816	-79.342741	Golf Course
Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0	43.753259	-79.329656	Galleria Supermarket	43.753520	-79.349518	Supermarket
Parkwoods	43.753259	-79.329656	26533.0	4.96	34811.0	43.753259	-79.329656	Island Foods	43.745866	-79.346035	Caribbean Restaurant

## 3. Methodology

In this project we will direct our efforts on detecting areas of Toronto that have lesser number of Gymnasiums. In first step we have collected the required data: location and category of every venues from neighbourhoods of Toronto.

Second step in our analysis will be calculation and exploration of 'number of gyms' across different neighbourhoods of Toronto. So, From the data we need to identify neighbourhoods which do not have at least one gym.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders. We will take into consideration, locations with not even one gym, and we want locations with high population and high population density. We will present map of all such locations but also create clusters using k-means clustering of those locations to identify general zones or neighbourhoods which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

### 3.1 Exploratory Data Analysis

#### 3.1.1 Finding locations with Gym

From the dataset above, 'Venue Category' column indicates the type of each venue. Venue Category may be Gym, Restaurant, Supermarket etc. Here we are looking for 'Gym' category and get all the rows with Gym as venue category. So, the resulting dataset would contain all the neighbourhoods with at least one gym.

Neighborhood	Agincourt	Alderwood	Bathurst Manor	Bedford Park	Birch Cliff	Brockton	Church and Wellesley	Cliffside	Davisville	Don Mills	...	St. James Town	The Annex	The Beaches
Venue Category	1	2	2	1	2	1	2	5	4	2	...	2	2	2

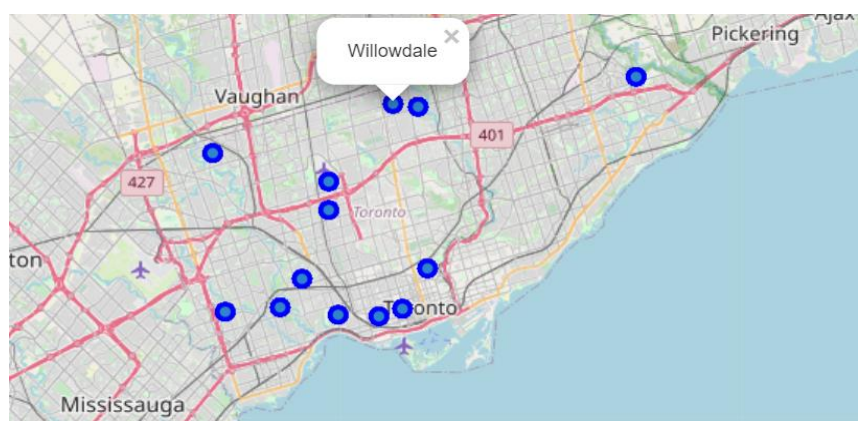
1 rows × 33 columns

The above table shows neighbourhood name and number of gyms. It has 33 columns, which means that 33 neighbourhoods have at least one gym.

#### 3.1.2 Finding locations with no Gym

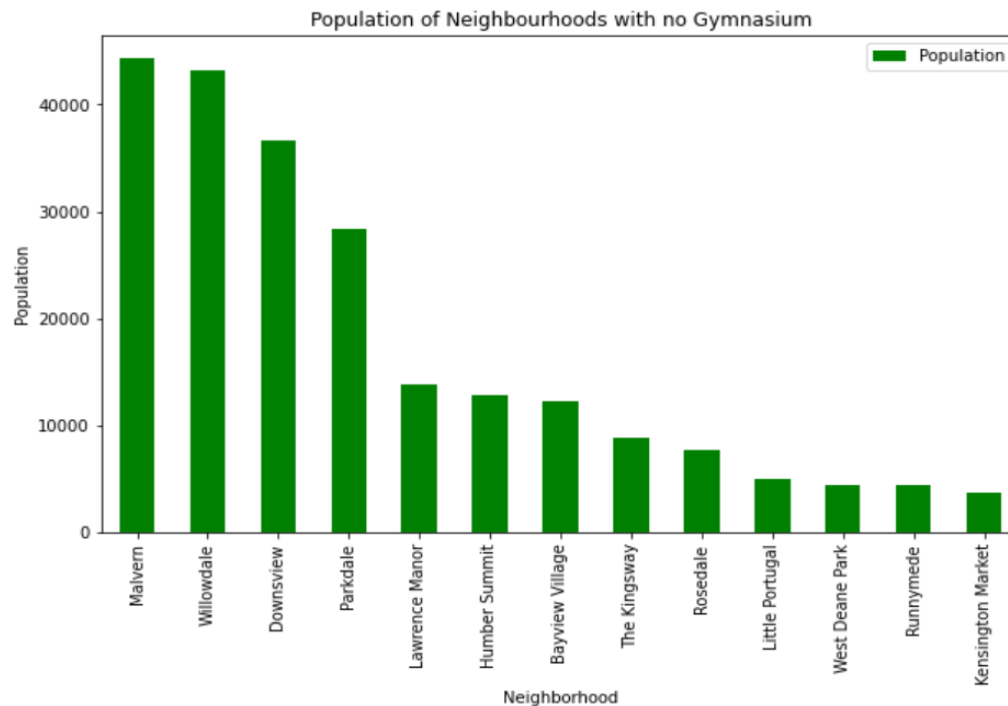
Initially, we selected a dataset with 46 neighbourhoods and 33 has at least one gym, so the rest of the 13 do not have a gym. That 13 neighbourhoods are:

```
{'Bayview Village',  
'Downsview',  
'Humber Summit',  
'Kensington Market',  
'Lawrence Manor',  
'Little Portugal',  
'Malvern',  
'Parkdale',  
'Rosedale',  
'Runnymede',  
'The Kingsway',  
'West Deane Park',  
'Willowdale'}
```

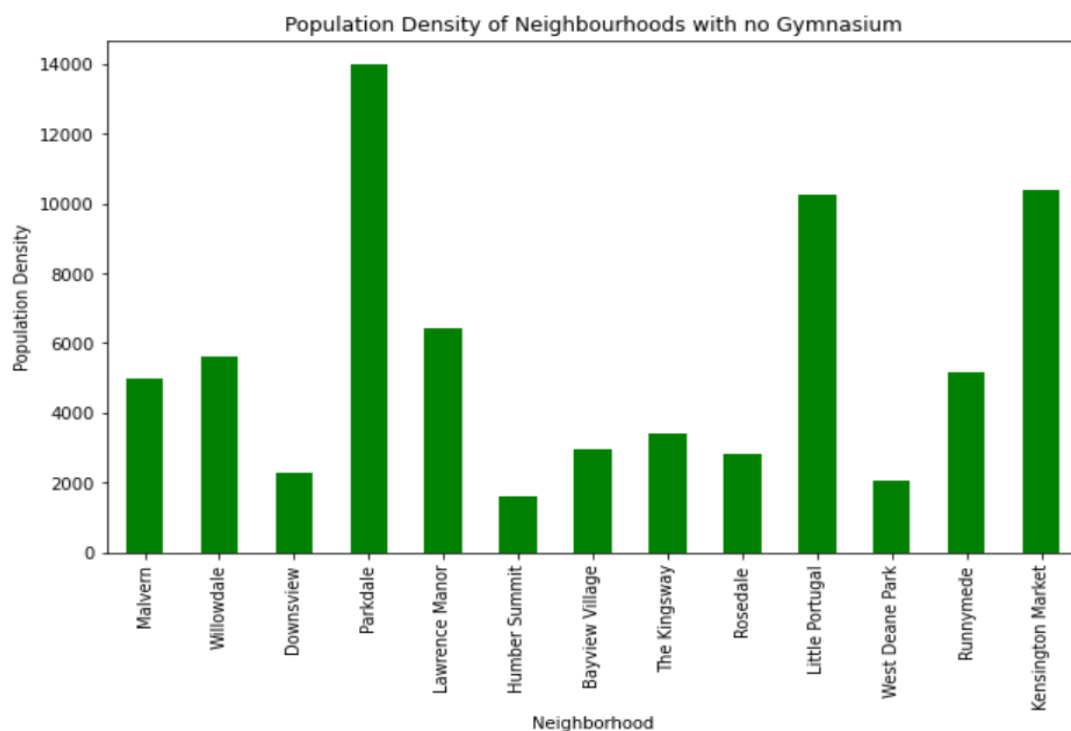


### 3.1.3 Explore locations with no Gym

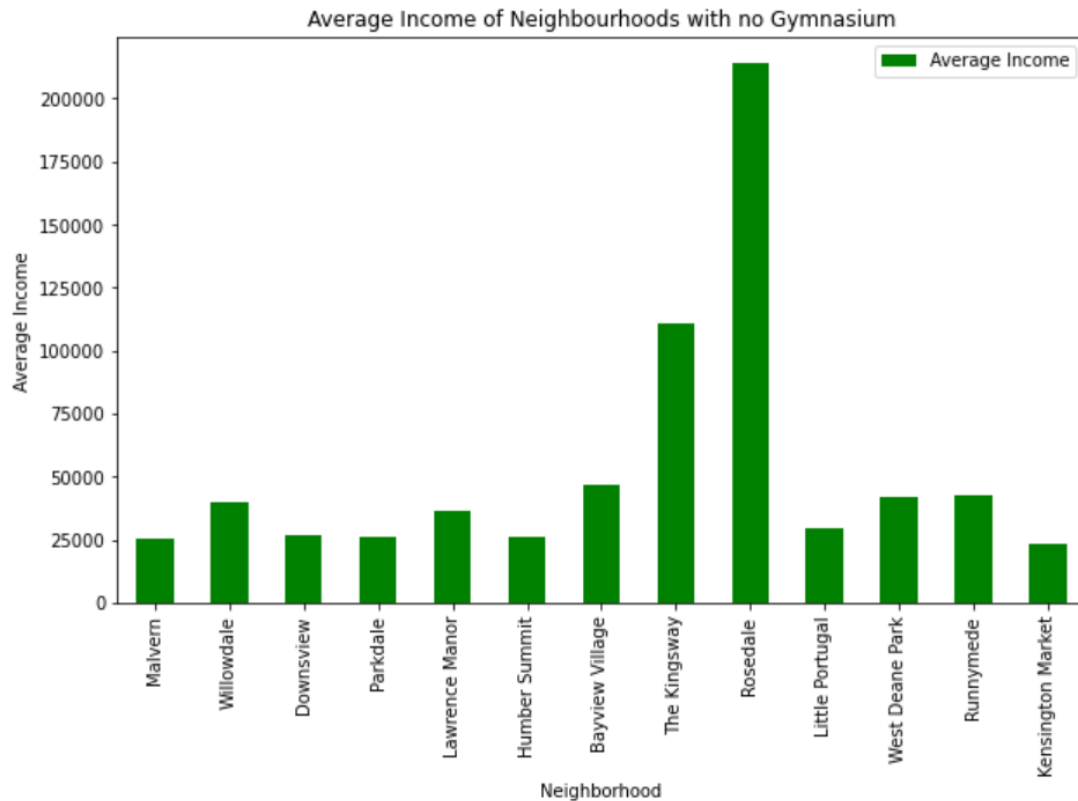
We found location which do not have a gym. Next, we must explore it to get best out of those 13 neighbourhoods to start up a gym. Our exploration is based on population, population density and average income of each neighbourhood. Because those factors may affect the success of a gym in a location.



From the above graph, we could see among neighbourhoods which have no gym, **Malvern, Willowdale, Downsview, Parkdale** have highest population.



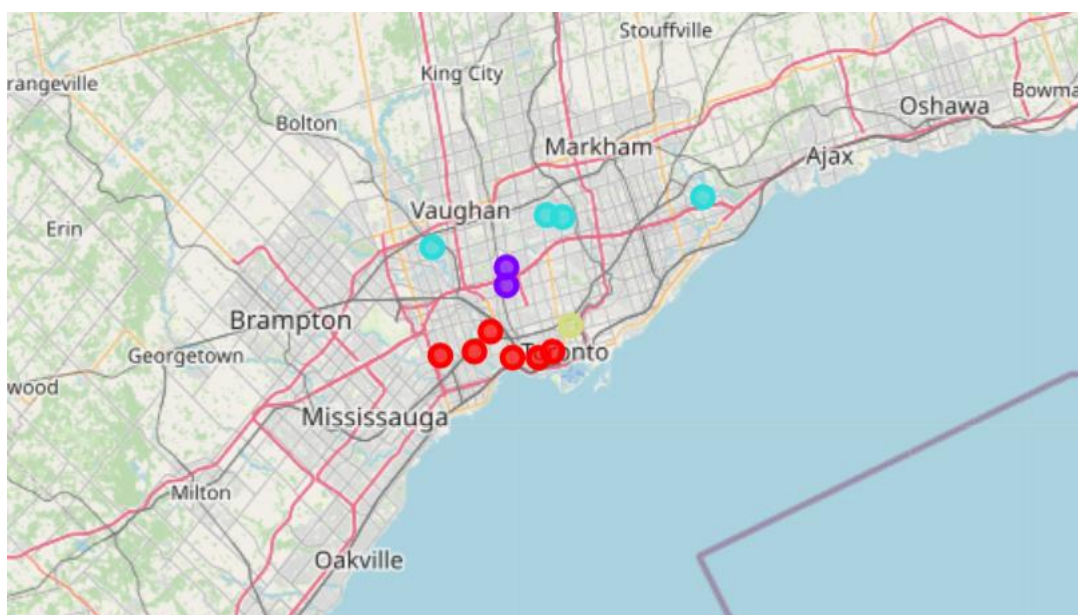
Among high populated neighbourhoods such as **Malvern, Willowdale, Downsview, Parkdale**, first three has low population density. Only **Parkdale** has high population density among four. **Parkdale, Little Portugal, Kensington Market** have high population density.



From above two plots we have seen that **Rosedale** has low population and population density, but it has high average income as compared to others.

### 3.2 Machine Learning: Clustering

Let us choose all 13 neighbourhoods with no gym for clustering. Clustering algorithm groups similar neighbourhoods together and dissimilar neighbourhoods into different clusters. Here I have used k-means clustering algorithm. Four clusters of neighbourhoods have been created. In this project, Clustering algorithm grouped similar neighbourhoods based on the categories of venues exist, population, average income and land area of each neighbourhood.



This could be useful for stakeholders in future when they expand their venture to other neighbourhoods.



## 4. Results

From the above analysis, let us say that Parkdale would be a best location to start a new gymnasium, because which have high population and high population density. Also let us choose other high-density neighbourhoods like Little Portugal, Kensington Market as well. By considering population we can consider Malvern, Willowdale, Downsview in addition to Parkdale. If average income is also considered, then let us choose Rosedale too. So out of 13 neighbourhoods with no gymnasiums 7 neighbourhoods are optimal locations to start a gym. They are:

- Parkdale
- Malvern
- Willowdale
- Downsview
- Little Portugal
- Kensington Market
- Rosedale

The result of clustering is as shown below. From this similar neighbourhoods can be identified.

	Neighborhood	Cluster
0	West Deane Park	0
1	Little Portugal	0
2	Runnymede	0
3	Parkdale	0
4	Kensington Market	0
5	The Kingsway	0
6	Lawrence Manor	1
7	Downsview	1
8	Malvern	2
9	Bayview Village	2
10	Humber Summit	2
11	Willowdale	2
12	Rosedale	3

## 5. Discussion

All 13 neighbourhoods selected do not have at least one gymnasium or 7 optimal neighbourhoods selected after processing might be best locations. But, of course, this does not imply that those zones are actually optimal locations for a new gymnasium. Purpose of this analysis was to only provide info on areas not crowded with gymnasiums and have great population density and income. it is entirely possible that there is particularly good reasons for small number of gymnasiums in any of those areas, reasons which would make them unsuitable

for a new gymnasium regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition, but also other factors and all other relevant conditions met.

## **6. Conclusion**

Purpose of this project was to identify Toronto neighbourhoods with low number of or no gymnasium at all. So, it would be helpful for stakeholders in search for optimal location for a new Gymnasium. From the data of nearby venues of each neighbourhood received from Foursquare API, I identified locations with no gymnasiums at all. So, I took those areas and performed comparisons based on population, population density and average income of the selected areas.

Clustering of the areas with no gymnasiums have also done to identify similar locations, so that stakeholders can start more than one gym in similar areas, or it would be useful in future.