Identifying Ideal Gym and Fitness Studio Locales

June 17, 2020

1. Introduction

1.1   Background

Gyms and active lifestyles are growing in popularity as a form of preventative self-care from the many adverse health effects of obesity. Health risk examples linked to obesity include early mortality, certain cancer types, diabetes, various coronary diseases, and stroke. There are also meaningful economic costs associated with obesity; impacting healthcare and transportation industries most significantly. The latest 2019 data from NHANES suggests that nearly 40% of adults are categorized as "obese;" the highest it has ever been. As with most real estate investments, location is paramount. With the required investment to start a Gym affiliate at a historical low, it would be advantageous to accurately predict and identify ideal locations for new Gyms and fitness studios.

1.2   Problem

This project aims to recommend the optimal location(s) for a new Gym or Fitness Studio based on collected data primarily from the Foursquare API. Data of interest includes locations of currently established Gyms, some measure of visit frequency (e.g. number of check-ins, unique visitors, etc.), viability of venue reputation based on likes and/or reviews, and analysis of similar or comparable venue category statistics such as lists or trending venue statistics over time. The data will then be cleansed and manipulated into layouts suitable for statistical regression analyses and K-means Clustering.

1.3   Interest

Gym owners and aspiring affiliate owners everywhere would welcome the ability to identify optimal Gym locations based on dynamic landscape and social interest variables. Gym franchises, such as Gold's Gym, Planet Fitness, etc. may also be interested in capitalizing on unrealized ideal locales for their franchisees.

2. Data Source & Cleansing

Given our problem described above, factors that will influence our analysis are:
- number of existing Gyms or Fitness Studios located in a given neighborhood
- Popularity and relative "foot traffic" in each respective neighborhood

The following data sources will be needed to extract/generate the required information:
- Wikipedia page consisting of all Toronto-area neighborhoods
- Toronto neighborhood geocoding JSON to link neighborhoods to respective latitude and longitude coordinates.
- number of Gyms/Fitness Studios and their type and location in every neighborhood will be obtained using the Foursquare API.

3. Methodology

In this project we would like to identify areas around Toronto that have Gym/Fitness Center density, particularly those with a low number of Gyms/Fitness Centers. We began by identifying areas in Toronto that have low Gym / Fitness Center density. We are particularly interested in those with a low number of Gyms.

First, we collected the required data: location and category of every Gym / Fitness Center around Toronto via the Foursquare API.

Secondly, we explored the geo-density of Gyms/Fitness Centers across different Toronto neighborhoods. We will leverage heatmaps to identify the neighborhoods with low number of fitness facilities and focus our attention on them.

We then focused on the most promising locations and create clusters of locations that meet some basic requirements established in discussion with stakeholders: we looked into consideration locations with no more than one facility within a radius of 500 meters, and we want locations without fitness facilities in radius of 800 meters. We then created a map of all such locations but also leverage k-means clustering to identify neighborhoods which should be a starting point for final exploration and search for optimal venue location by interested parties.

4. Results

We can see that Gyms are the 13th ranked most frequent venue in the provided Toronto data set with 32 existing locations. Fitness Studios, also included in the analysis but not pictured below, totaled 8 existing locations. Each location was mapped and examined for location density, identifying gyms and fitness centers that either a) do not current exist within a Toronto neighborhood or b) were not within 500m of a competing location i.e. a Gym or Fitness Studio already existed.

```
venue_counts=toronto_gyms['Venue Category'
venue_counts.head(30)
```

```
(2128, 7)
```

Out[13]:
```
Coffee Shop              179
Café                     100
Restaurant                64
Park                      56
Pizza Place               50
Italian Restaurant        44
Japanese Restaurant       43
Hotel                     43
Bakery                    43
Sandwich Place            39
Clothing Store            32
Gym                       32
Grocery Store             29
Sushi Restaurant          28
Bar                       28
```

After the one-hot-encoding methodology was applied, were able to perform a K-means clustering analysis on the data (pictured below). Our end result was 57 "opportunity pockets" around Toronto neighborhoods that fit our acceptance criteria of:

- no more than one facility within a radius of 500 meters, and
- locations without fitness facilities within a radius of 800 meters

5. Discussion

The analysis above shows there are numerous "opportunity pockets" of low fitness facility neighborhood density around Toronto. Surprisingly, there are a number of areas in the downtown Toronto area that fit our acceptance criteria.

There are **57 zones** surrounding Toronto in total that fit our new fitness facility location criteria based on number of and distance to existing fitness facilities. This analysis purely looked at areas close to Toronto with low fitness facility density - it is feasible that low density could be explained or is justified in any of these pockets, reasons which would make them unsuitable for an additional fitness facility regardless of area disposition. The recommended "opportunity pockets" should only be treated as a starting point for more detailed analysis.

6. Conclusion

The objective of this project was to identify Toronto neighborhoods with low number of fitness facilities (Gyms and Fitness Centers) in order to aid stakeholders in narrowing down the search for optimal location. By calculating fitness facility density distribution from Foursquare data, we have identified several neighborhoods that justify potential real estate investment and further investigation. K-means clustering of those neighborhoods was then performed in order to create major areas of interest for stakeholders to investigate further.