
APPLIED DATA SCIENCE CAPSTONE:

Where to open a new gym
in Toronto?

Justin Yu



Business Problem

Where Is The Best Location To Open A New Gym In Toronto?

- The trend of living a healthy lifestyle has transformed the daily lives of many, and the popularity of working out is growing exponentially
- Where can a gym investor open a new gym in order to face the lowest level of competition?
- Target Audience: Gym Investors



Data Acquisition

Wikipedia:

**List of
neighborhoods in
Toronto**

**Geocoder
Package:**

**Latitudes &
Longitudes of
neighborhoods**

Foursquare API:

**Top 100 venues in
each
neighborhood**

Data Acquisition

Combining the three sources of data, a comprehensive data set can be fed into the machine learning algorithm

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant
Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café
Parkwoods	43.753259	-79.329656	A&W	43.760643	-79.326865	Fast Food Restaurant
Parkwoods	43.753259	-79.329656	Bruno's valu-mart	43.746143	-79.324630	Grocery Store

Data Science Methodology

Data Preparation

- Web Scraping & Foursquare API

K - Means Clustering

- Cluster neighborhoods into 5 clusters
- Target: Cluster with lowest gym prevalence

Exploratory Data Analysis

- Visualize all neighborhoods in a map
- Feature engineering

Recommendation

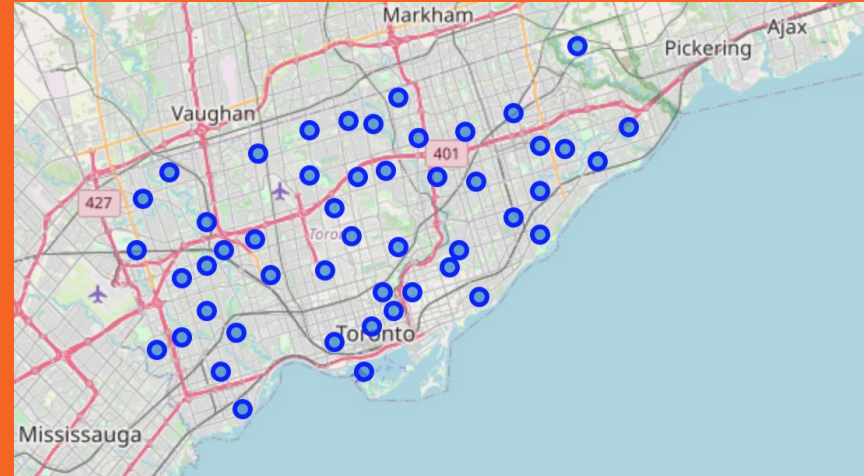
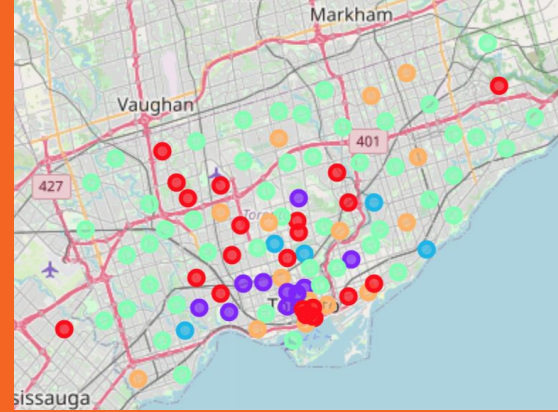
- Visualize neighborhoods in target cluster in a map
- List of neighborhoods with great investment potential

Results: Target Cluster

The target cluster is represented in **mint green** in the map above

Focusing on the neighborhoods in the target cluster in the map below, these neighborhoods scatter in different boroughs in Toronto → autonomy for investors to choose a familiar neighborhoods from the neighborhood candidates in the target cluster

Story for illustration purposes only



Findings

Cluster 3 has a much lower level of competition:

- **70% lower than the second most ideal cluster**
- **10x less competitive than the most competitive cluster**
- **82% lower than the average level of competitiveness of all other clusters**

Gym Mean	
Cluster Labels	
3	0.006849
1	0.022275
4	0.029426
0	0.035258
2	0.067376

Recommendation

- Neighborhoods in Cluster 3 are the ideal locations to open a new gym
- Conduct in - depth analysis on other factors (e.g. demographics, population, average income level, etc) to decide on which neighborhood in Cluster 3 fits the business plan of the gym investor

