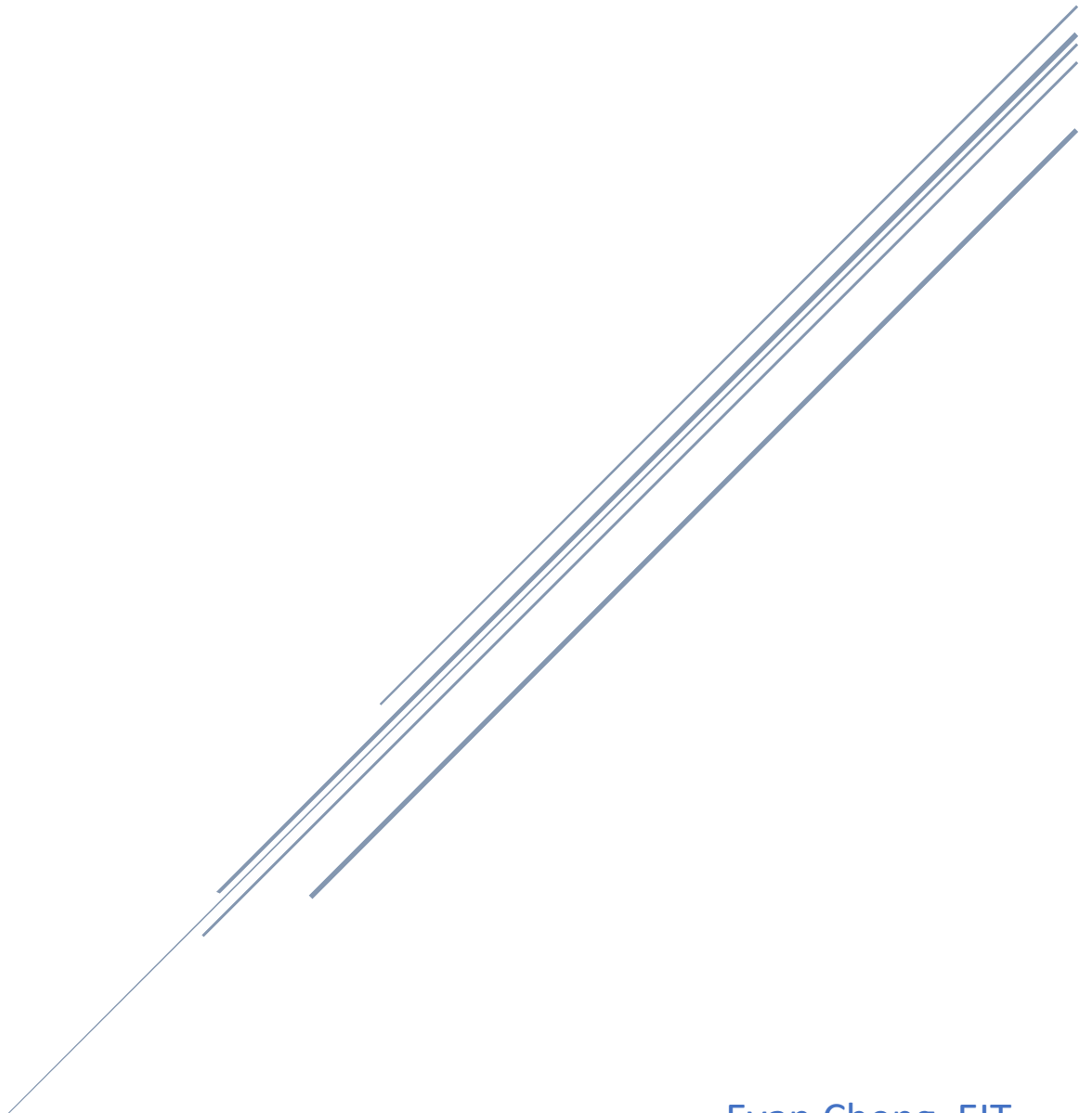


IBM APPLIED DATA SCIENCE CAPSTONE PROJECT:

The Battle of the Neighborhoods – Starting a Brazilian Jiu Jitsu
Academy



Evan Cheng, EIT
May 16, 2020

Introduction

Starting a business is a daunting venture, and one of the difficult decisions entrepreneurs face is determining the ideal location for their business. This is especially true for Brazilian Jiu Jitsu (BJJ) gyms and academies. BJJ is a martial art that has grown tremendously in popularity over the last two decades, with the rise of the Ultimate Fighting Championship (UFC) and the sport of mixed martial arts (MMA). As a result, gyms are popping up across the Canada, making it increasingly difficult to open a new gym. Gyms must be easily accessible and be in large enough spaces to house lots of mats and training space.

The aim of this project is to examine what makes BJJ gyms in the Greater Toronto Area (GTA) successful and use that information to determine the best location to open a new BJJ academy. Entrepreneurial BJJ junkies looking to open their own gyms would benefit greatly from this analysis, and so would the BJJ communities, with new gyms being opened in areas where they have the best opportunity to thrive.

To solve this problem, we will be leveraging Foursquare location data to collect information on existing gyms in the GTA, and the venues and features of their surrounding neighborhoods.

Data

The Foursquare API will be the main source of data for this analysis. To determine what makes existing BJJ gyms successful, we will be looking at all existing gyms in the GTA. We will require geographic and other categorical data from Foursquare including: the category of the venue (to ensure that it is a martial arts gym), the latitude and longitude (to find other venues nearby), and the address of the gym. Below is a sample of the data collected on BJJ gyms in the GTA.

Table 1 - Sample data collected on BJJ gyms in Toronto

| | name | categories | address | crossStreet | lat | lng | city | state | formattedAddress |
|----|-------------------------------------------|-------------------|-------------------------|--------------------------|-----------|------------|-------------|-------|---------------------------------------------------|
| 0 | Bushido Jiu Jitsu | Martial Arts Dojo | 5030 Maingate Dr. | Eglinton Ave. E. | 43.629700 | -79.628599 | Mississauga | ON | [5030 Maingate Dr. (Eglinton Ave. E.), Mississ... |
| 1 | Salvosa Brazilian Jiu-Jitsu Academy | Martial Arts Dojo | 31 Progress Ave Unit #1 | NaN | 43.775312 | -79.172351 | Scarborough | ON | [31 Progress Ave Unit #1, Scarborough ON M1P 4... |
| 2 | Buckley Jiu-Jitsu | Martial Arts Dojo | 1029 Speers Rd | NaN | 43.433083 | -79.703606 | Oakville | ON | [1029 Speers Rd, Oakville ON L6L 2X5, Canada] |
| 3 | Milton Jiu-Jitsu | Martial Arts Dojo | 310 Main St | NaN | 43.515812 | -79.879993 | Milton | ON | [310 Main St, Milton ON, Canada] |
| 4 | Bama - Burlington Academy of Martial Arts | Martial Arts Dojo | 3295 Mainway Dr # 4 | Guelph Line | 43.373619 | -79.804549 | Burlington | ON | [3295 Mainway Dr # 4 (Guelph Line), Burlington... |
| 5 | Toronto BJJ & MMA Academy | Martial Arts Dojo | NaN | Bloor West & Crawford | 43.662768 | -79.421832 | Toronto | ON | [Bloor West & Crawford, Toronto ON, Canada] |
| 6 | toronto bjj | Martial Arts Dojo | 813 Bloor W | NaN | 43.662828 | -79.421429 | Toronto | ON | [813 Bloor W, Toronto ON, Canada] |
| 8 | Body Of Four BJJ | Martial Arts Dojo | 2486 Dufferin street | Dufferin And Castlefield | 43.706720 | -79.468644 | Toronto | ON | [2486 Dufferin street (Dufferin And Castlefiel... |
| 9 | Toronto BJJ | Martial Arts Dojo | NaN | NaN | 43.694518 | -79.293606 | NaN | NaN | [Canada] |
| 10 | Openmat Mixed Martial Arts | Martial Arts Dojo | 593 Yonge St. | Wellesley St. | 43.666172 | -79.384767 | Toronto | ON | [593 Yonge St. (Wellesley St.), Toronto ON M4Y... |

To understand what makes these gyms successful, we need to understand the neighborhoods that they operate in. To do this, we will need to gather information on the venues within 1 km of the BJJ gyms. For

all the venues within this radius of the BJJ gyms, we will need to collect similar data, including their categories, labels, and distance from the BJJ gyms, to better understand how they impact the BJJ gym.

Table 2 Sample data collected on venues nearby BJJ gyms

| | venue_name | nearest_gym | distance_to_gym | venue_lat | venue_lng | venue_address | venue_city | venue_cat |
|---|---------------------|-------------------|-----------------|-----------|------------|-------------------------------|-------------|----------------------|
| 0 | On the bun | Bushido Jiu Jitsu | 83 | 43.629843 | -79.627587 | 14-5030 Maingate Dr | Mississauga | Burger Joint |
| 1 | Shandra's Roti Shop | Bushido Jiu Jitsu | 107 | 43.630173 | -79.627433 | 5030 Maingate Dr. | Mississauga | Caribbean Restaurant |
| 2 | Battle Arts Academy | Bushido Jiu Jitsu | 494 | 43.625405 | -79.630183 | 4880 Tomken Rd | Mississauga | Gym / Fitness Center |
| 3 | Starbucks | Bushido Jiu Jitsu | 244 | 43.627514 | -79.628820 | 1016 Eglinton Ave E | Mississauga | Coffee Shop |
| 4 | Studio 89 | Bushido Jiu Jitsu | 274 | 43.627425 | -79.627279 | 1065 Canadian Place, unit 104 | Mississauga | Café |

Methodology

Data was gathered from the Foursquare API in two phases. In the first data pull, any venues within 40 km of Toronto (effectively the GTA) that were labelled as ‘Martial Arts Dojos’ were collected. This list of venues was then refined to include any venue with a name containing either ‘Brazilian Jiu Jitsu’, ‘BJJ’, or ‘MMA’. This list of BJJ gyms became the ‘gym data set’. The name, geographic coordinates, category label, and address of each gym was collected and tabulated.

The second data pull from Foursquare gathered all venues within a kilometer of each gym. The same information was collected for each venue: its name, category label or classification, geographic coordinates, and distance from the nearest gym. This second data set became the ‘venue data set’.

Once both data sets were pulled using the Foursquare API, each gym in the GTA needed to be reviewed to determine what types (categories) of venues were closest to it by using the ‘one hot encoding’ method. This method allowed the number of occurrences of each category label to be counted for all the venues surrounding a given gym. For example, in the table shown below, we can see that the gym ‘Salvosa Brazilian Jiu-Jitsu Academy’ has 1 bar nearby, while ‘Milton Jiu Jitsu’ is nearby 1 bank and 1 beer store. Note that this is only a small sample of the data collected.

Table 3 Sample of data showing how many different types of venues are nearby a given BJJ gym

| | name | categories | BBQ Joint | Bagel Shop | Bakery | Bank | Bar | Baseball Field | Beach | Beer Bar | Beer Store |
|---|-------------------------------------------|-------------------|-----------|------------|--------|------|-----|----------------|-------|----------|------------|
| 0 | Bushido Jiu Jitsu | Martial Arts Dojo | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | Salvosa Brazillian Jiu-Jitsu Academy | Martial Arts Dojo | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | Buckley Jiu-Jitsu | Martial Arts Dojo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Milton Jiu-Jitsu | Martial Arts Dojo | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 4 | Bama - Burlington Academy of Martial Arts | Martial Arts Dojo | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

Once it was understood which venue types were nearby each gym, the k-means algorithm was implemented through the scikit-learn library to perform clustering on the gyms. The k-means algorithm was selected because of its speed and simplicity. The purpose of using the algorithm was to use machine learning to cluster the different gyms together based on how similar the surrounding shops and venues were. This shed light on the different areas that BJJ gyms commonly operate in. Because of the small number of gyms in the GTA (17 BJJ gyms total), a value of $k = 3$ was used for clustering. Table 4 shows a small portion of the results of the k-means clustering. Notice each gym has been assigned a cluster label from 0 to 2.

Table 4 Most common venues, and cluster label for each BJJ gym in the GTA

| | name | categories | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Cluster |
|---|-------------------------------------------|-------------------|---------------------------|---------------------------|-----------------------|-----------------------|-----------------------|---------|
| 0 | Bushido Jiu Jitsu | Martial Arts Dojo | Middle Eastern Restaurant | Coffee Shop | Caribbean Restaurant | Bakery | Burger Joint | 0 |
| 1 | Salvosa Brazillian Jiu-Jitsu Academy | Martial Arts Dojo | Hotel | Italian Restaurant | Park | Grocery Store | Burger Joint | 2 |
| 2 | Buckley Jiu-Jitsu | Martial Arts Dojo | Hobby Shop | Latin American Restaurant | Diner | Pub | Sandwich Place | 0 |
| 3 | Milton Jiu-Jitsu | Martial Arts Dojo | Pub | Italian Restaurant | American Restaurant | Sandwich Place | Sporting Goods Shop | 1 |
| 4 | Bama - Burlington Academy of Martial Arts | Martial Arts Dojo | Park | Pizza Place | Pet Store | Coffee Shop | Gas Station | 0 |

Results

To visualize how the k-means algorithm had clustered the gyms of the GTA and their surrounding shops and venues, the Folium library was used to create a color-coded map as shown below. The gyms are plotted with colored markers, with different colors representing different clusters. The venues surrounding each gym are plotted in black.

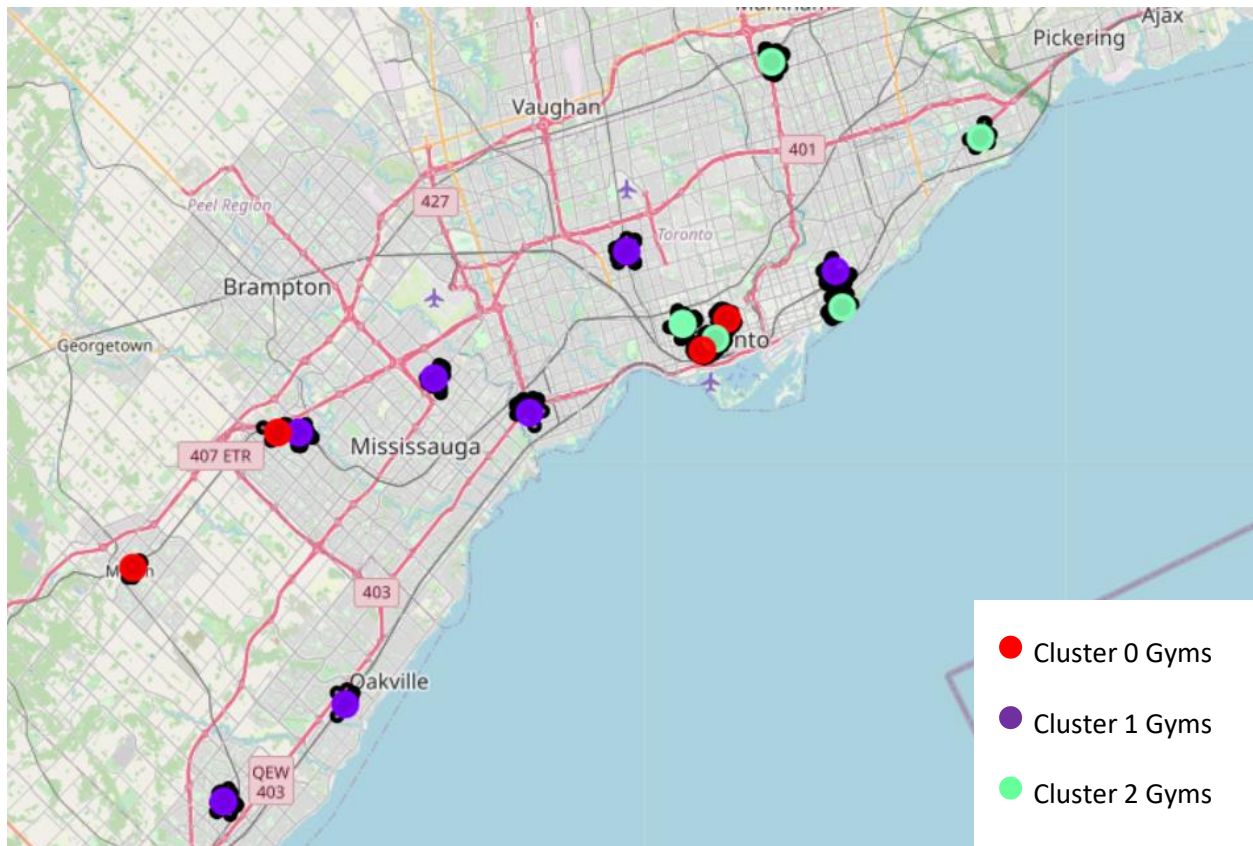


Figure 1 Map showing clustered BJJ gyms in the GTA

The Folium map allows closer inspection of any gym on the map, as demonstrated below.



Figure 2 Snippet of Folium map for Salvosa Brazilian Jiu-Jitsu Academy, showing cluster number, and nearby venues

After using the k-means algorithm to separate the gyms into 3 distinct clusters, word clouds were used to visualize the most common types of venues surrounding gyms in each cluster. For each gym, there were over 100 unique category labels. This meant that traditional data visualization techniques like a histogram would have been ineffective at conveying information, with histograms having a horizontal axis with over 100 labels! Word clouds were used to visualize the data instead, since they more effectively convey textual data. A word cloud was created for each cluster, to understand which venues were most popular around the BJJ gyms in that cluster. The word clouds for the venues surrounding each cluster are shown below.



Figure 3 Word cloud representing venues most commonly nearby gyms in Cluster 0



Figure 4 Word cloud representing venues most commonly nearby gyms in Cluster 1



Figure 5 Word cloud representing venues most commonly nearby gyms in Cluster 2

Additionally, the mean proximity of venues from BJJ gyms was calculated for each cluster, as shown in the bar chart below.

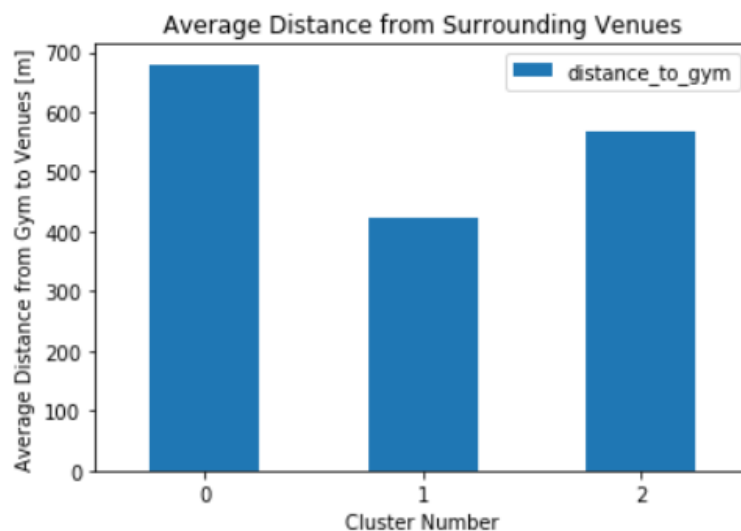


Figure 6 Average distance between gyms and surrounding venues for each cluster

Discussion

Visual Inspection of Word Clouds

From the word clouds shown in Figures 3-5 above, we can see that regardless of cluster, coffee shops, bars and cafés are frequently located near BJJ gyms. A likely reason for this is that coffee shops and bars usually have high volumes of foot traffic, which provides natural advertisement for the gym as people walk to grab their morning coffee, or walk home from an evening at the local bar. This leads to an important realization for prospective gym owners: it is critical to place your gym in areas with high foot traffic, ideally near coffee shops and cafés, regardless of neighborhood.

The word cloud for Cluster 0 indicates that Cluster 0 gyms are found near fast food restaurants (including pizza parlors and sandwich shops), and train/subway stations. Cluster 1 gyms are near more recreational neighborhoods, surrounded by art and yoga studios, other fitness gyms, and parks. Cluster 2 gyms are also commonly located near a variety of food establishments, with a stronger emphasis on nightlife and drinking venues like pubs and cocktail bars.

Analysis of Clusters using Geospatial Data

Taking a closer look at the Folium map from Figure 1 above, we can more closely analyze a few zones of interest to draw more conclusions from the data.

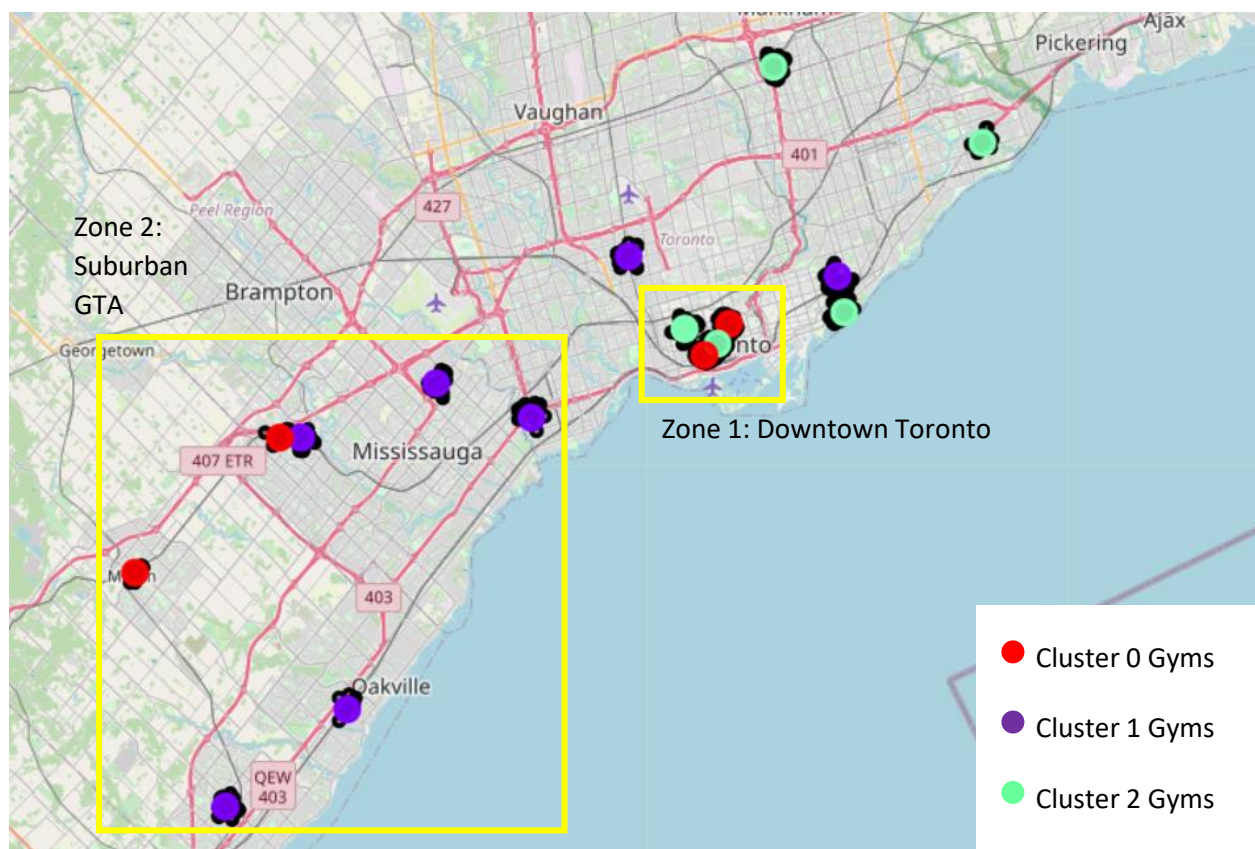


Figure 7 Folium map showing zones of interest

Zone 1: Downtown Toronto

From Zone 1, it is evident that only Cluster 0 and Cluster 2 gyms are found in Toronto's downtown core. The downtown area of most major cities tends to have high concentrations of restaurants and nightlife. This is consistent with the analysis of the word clouds for Clusters 0 & 2, where both clusters tended to have gyms nearby food venues and drinking establishments. The map below takes a closer look at Zone 1, showing the distribution of nearby venues to BJJ gyms in the downtown core of Toronto.

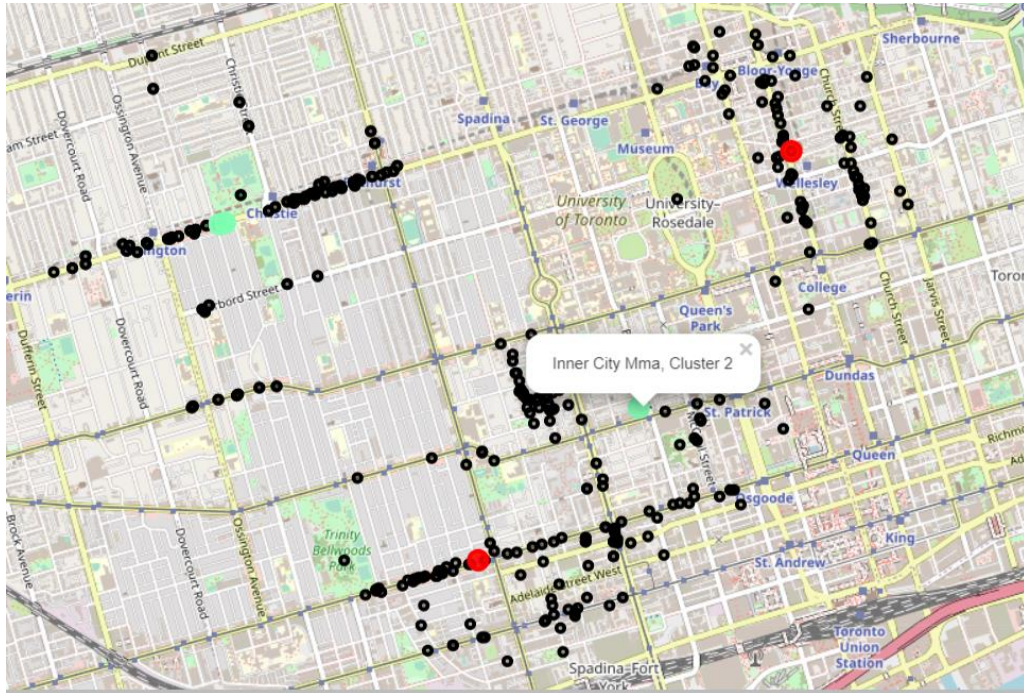


Figure 8 Folium map showing BJJ gyms in Zone 1 (Downtown Toronto)

Zone 2: Suburban GTA

Zone 2 demonstrates that most gyms located in the suburban areas of the GTA belong to Cluster 1. From the word cloud for Cluster 1, it was concluded that Cluster 1 gyms are located near other recreational venues like fitness gyms, yoga studios, and parks. While an area like this may exist in downtown Toronto, it is quite common for suburban areas to have large strip malls and complexes that house a variety of recreational venues. The bar chart shown in Figure 6 shows Cluster 1 gyms have a lower mean distance to their surrounding venues, which supports the theory that Cluster 1 gyms tend to be located in strip malls or similar complexes. Having a BJJ gym inside a strip mall or similar complex would reduce the mean distance to nearby venues.

Putting All the Pieces Together:

After analyzing the data from the word clouds, the clustered maps, and the mean distance between gyms and nearby venues, the clusters, as determined by the k-means algorithm, can be described as shown in the table below.

Table 5 Description of BJJ gym clusters based on data analysis

| | Cluster Description |
|-----------|---------------------------------------------------------------------------------|
| Cluster 0 | Downtown Toronto: located near fast food establishments and train stations |
| Cluster 1 | Suburban GTA Gym: located in strip malls near other recreational fitness venues |
| Cluster 2 | Downtown Toronto: located near restaurants and nightlife (bars and pubs) |

The description of each cluster of gyms makes logical sense. In the downtown area of an urban metropolis like Toronto, it would be sensible to place your gym in areas of high foot traffic, where passersby can see the gym, and poke their head in and ask questions if they desire. This means the ideal location for gyms downtown would be near train stations, fast food establishments, restaurants, and nightlife. The k-means Clusters 0 & 2 support this hypothesis. In suburban areas, it makes sense to locate a gym in a strip mall, or similar complex. This would give a gym owner access to sufficient space and allow the gym to be located near other venues that may attract similar customers and students. For example, Cluster 1 gyms tend to be near yoga studios and other fitness gyms. It seems likely that a consumer who is interested in yoga may also be interested in learning BJJ. Having the gym next to a yoga studio could help draw consumers who are inclined to enjoy physical activity and exercise. Being in a strip mall usually also means having access to advertising signage visible from main roads and highways.

The k-means clustering has provided interesting insights on where the ideal locations are for opening BJJ gyms. Potential gym owners that wish to open a school downtown should ensure they choose a location near coffee shops, bars, restaurants, and public transportation. If a gym owner wishes to open an academy in a suburban region, they should be sure to open their academy in a strip mall or similar complex that is near other fitness venues like weightlifting gyms or yoga studios.

Discussion

The k-means algorithm has provided valuable initial insights into the types of venues that are common around BJJ gyms throughout the GTA. However, due to the textual nature of the data, quantifying accuracy and validity of the results is challenging. In future phases of this project, steps should be taken to produce more quantitative models of the data to evaluate the assertions made in this report. For example, the large number of unique category labels could be clustered into more general categories (ie. food, transportation, retail). This would provide more numerical insight into how much more frequently fast food restaurants appear compared to coffee shops around BJJ gyms.

Conclusion

The objective of this study was to determine the ideal location to open a BJJ gym in the GTA. To do this, the Foursquare API was used to pull data about existing gyms in the GTA, and the venues that were within 1 km of these gyms. The gyms in the GTA were clustered using the k-means algorithm. Using word clouds and Folium interactive maps, the categorical data for each gym and its surrounding venues was analyzed. From the analysis, it can be concluded that the ideal location for a BJJ gym depends on whether it is in a downtown core, or a suburban environment. Gyms located downtown are ideally placed near restaurants, bars, pubs, coffee shops, and public transportation. This provides natural advertisement via high levels of pedestrian traffic. Gyms located in suburban regions are ideally placed in strip malls that house other recreational fitness centers like yoga studios and weightlifting gyms. This allows the BJJ gym to attract fitness-minded consumers that would be most eager to try BJJ.