Identifying where we should open a high-end fitness centre in Toronto

Michael Mehta

July 2021

# 1. Introduction:

## 1.1 Background:

In this project, we were approached by a land developer interested in building a high-end fitness centre in Toronto. Toronto is the most populous city in Canada and the financial centre of the country. Resultingly, there is a large market for high end fitness centres. There are several existing fitness centres such as the *Adelaide club* and the *Granite club* that cater to this market, but we want to evaluate where they are and what their strategy and if we can carve a niche, and where we should build our facility.

## 1.2 Problem

We want identify by postal code FSA which areas could support a new high-end fitness club. In order to do this, we will take area population, affluence and competitive intensity (distance to existing competitors) for relevant markets to identify which markets have sufficient demographics to support a club like this.

# 2.Data Acquisition and Cleaning

## 2.1 Data Sources

Our data sources here are:

* **Wikipedia** Toronto postal codes: this gives us the latitude and longitude per FSA postal code in Toronto. This allows us to call Foursquare accurately.
* **Foursquare**: This allows us to determine what the existing neighborhood characteristics per FSA postal code looks like. This includes what existing commercial features in the neighbourhood like how many cafes or parks or gyms there are in that neighbourhood. In our case a neighbourhood is defined as a 5km radius around the centre point of the FSA postal code.
* **Stats Canada**: this gives us population data by FSA postal code from the 2016 census.
* **Housepricehub:** this gives us average house price sold in the past year by FSA. This is done as a proxy to neighbourhood affluence.

## 2.2 Data Cleaning

Data taken from Wikipedia, Stats Canada and HousePriceHub were scraped from their respective sites and combined into a single table called *neighbourhoods.* These tables were merged based on the FSA (the first 3 digits of the postal code).

Both Stats Canada and HousePriceHub had data Canada-wide. As this analysis was solely looking at Toronto, we started by reducing to postal codes that begin with the letter ‘M’. These are the postal codes that correspond to the greater Toronto area, from Etobicoke to Scarborough and north to the southern border of Markham.

Timing wise of the Stats Canada is from the 2016 census. The results from the 2020 census have not yet been released, so we are running under the assumption that the proportional populations have remained relatively consistent since then. This could be a potential source of error, particularly for newly developed and developing neighbourhoods, and downtown neighbourhoods that are growing in population density.

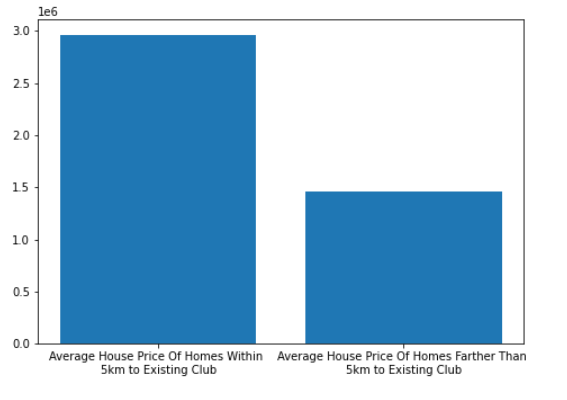
Unfortunately, these datasets were not always comprehensive. While Stats Canada Population data had information per FSA postal code, HousePriceHub did not. In order to work around for this, we estimated the house price for FSAs without information with the Toronto-average house price. This happened in 13 of the FSAs (including non-residential FSAs with 0 population). This allowed us to be able to analyze all of Toronto.

# 3. Exploratory Data Analysis

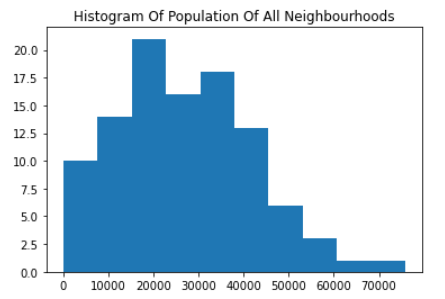
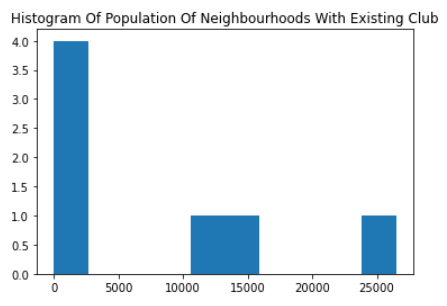
## 3.1 Feature Selection and Data Analysis

We wanted to look at the full range of commercial demographics per neighbourhood. In addition, we wanted to determine how much affluence as well as local population impacts the success of existing luxury fitness centres.

Let’s first look at affluence, from HousePriceHub, we can look at the average price of a house in neighborhoods with an existing club within 5km and those that do not. Clubs were deemed to be exclusive if they were rated 4+ on Foursquares expensiveness scale. Majority of these clubs are 5/5 using this scale.



*Figure 1. Shows the average home price within 5km of an existing club vs average home price more than 5km from an existing club.*

From the chart above, existing high-end clubs exist in neighbourhoods with housing prices about double those not within a 5km radius of existing clubs ($3M in club neighborhoods and $1.5 in non-club). It speaks to the importance of surrounding affluence on their business model as well as the types of neighbourhoods we’d want to look at for our client.   


*Figure 2: shows the population count in neighbourhoods with existing exclusive fitness clubs.*

Looking into population surrounding the clubs, we can see that the fitness clubs fall into one of a few categories.

1. **Commercial**: Downtown in a very commercial area such as *Toronto Athletic Club, Equinox Club or Adelaide Club*. Where there are few residences, but a lot of business and daily foot traffic. These have low populations in the surrounding neighborhood.
2. **Residential**: In strictly residential areas amongst primarily detached homes, reducing population density, such as *The Granite Club and Rosedale Club*. These have mid-level populations but are extremely affluent.
3. **Hybrid**: In a downtown residence area that include multi-unit homes such as condos. These are a both close to expensive homes (Yorkville is an expensive neighborhood) and also pretty close to the financial sector similar to group 1. They likely cater to both the white-collar financial district workers coming to and from work as well as the people who live in the surrounding areas.

These different niches of luxury fitness clubs would have different strategies for their market segment and as such will likely end up in different clusters. A neighborhood that could support a *Granite Club* model may not succeed if trying an *Adelaide Club* model. As such, population is another variable that will need to be included in our analysis.

# 4. Predictive Modeling

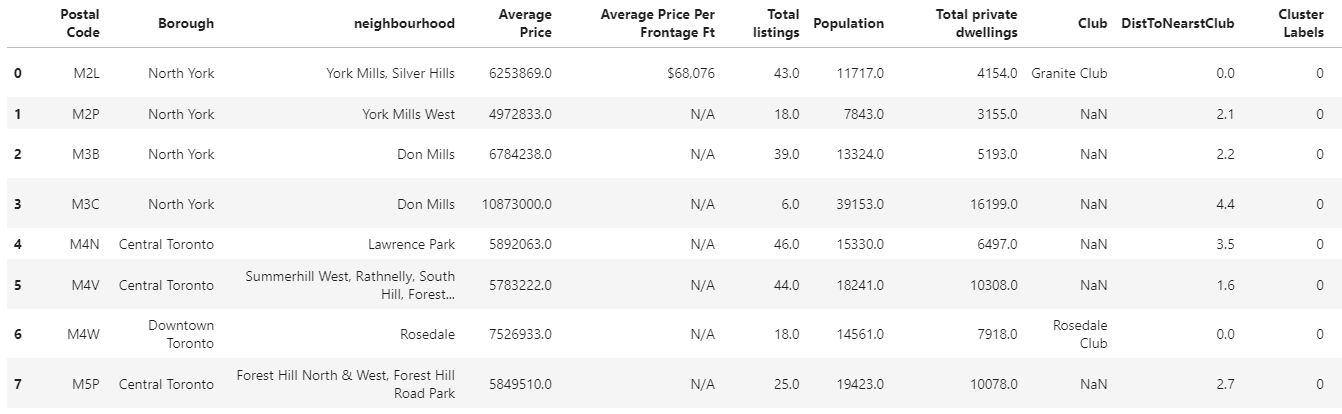
## 4.1 Model Selection

For this project, we are trying to discern which neighborhoods could potentially support a new high-end fitness facility. The basis of this is looking at existing successful facilities and their neighbourhood makeup, their surrounding affluence as well as their surrounding population. Because of this, a clustering method seemed most appropriate. Specifically, a k-means cluster was chosen as the algorithm. We play around with the k in the k-means cluster to determine hierarchal splits and to see stability of clusters as we introduce more.

## 4.2 K-means clustering

We are going to include all Foursquare information as well as population and neighbourhood affluence. These are summarized together in the *toronto\_grouped* dataframe. There are 103 FSAs in Toronto.

### 2 clusters:

Breaking the neighbourhoods into two clusters results in the following group of neighborhoods and everything else. These neighborhoods are some of the most expensive in Toronto including Lawrence Park, Don Mills (including Bridal Path area) and Rosedale, these neighborhoods are geographically close together and are already supported by two long-standing clubs: *The Granite Club* and *Rosedale Club.* All neighborhoods are within 5km of one of these. There doesn’t seem to be much opportunity as clients in these areas who would want to join one of these clubs are likely already at one of these. 

*Figure 3: Lawrence Park Cluster shows the first cluster of neighborhoods in Toronto based on demographics, population and affluence.*

The remaining 95 neighborhoods are still clustered together, so we definitely want to dig deeper.

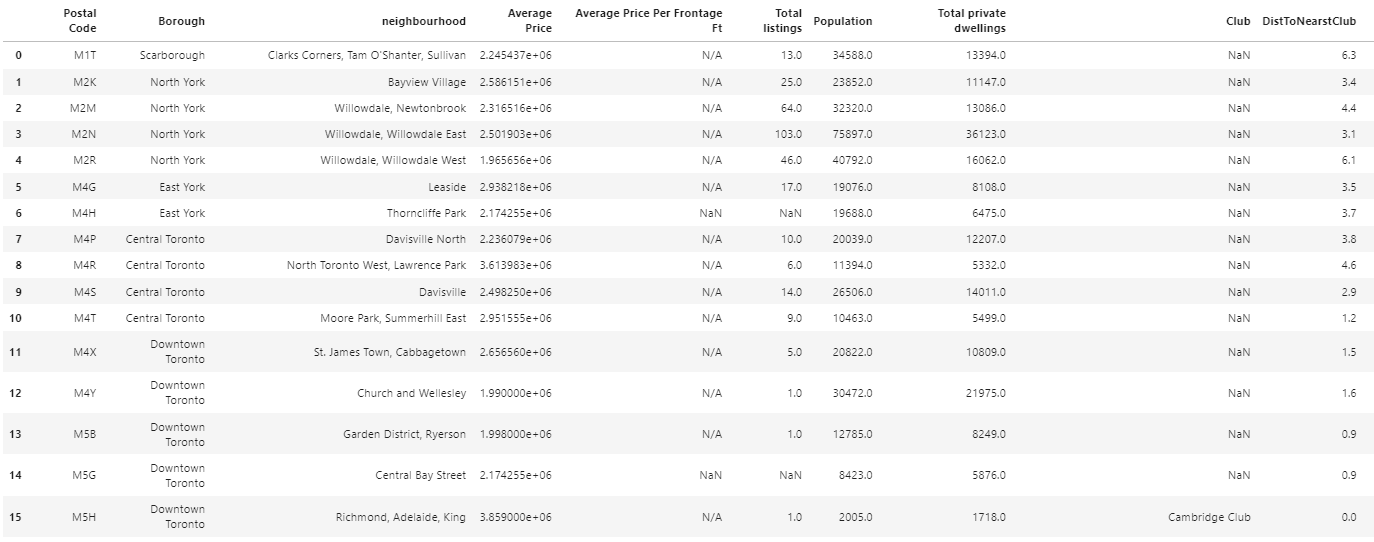
### 3 clusters:

Breaking down a little more, we still get the exact same cluster as shown in Figure 3. The remaining cluster has therefore been split into two. The first has no existing fitness facilities. These are all relatively inexpensive (all under the Toronto average) and are mostly in suburban areas, with a lot of Etobicoke, York and Scarborough FSAs.



*Figure 4: Suburbs Cluster is a cluster of neighborhoods that have housing prices under the Toronto average and have not historically supported a high-end fitness facility*

The third cluster has the remaining existing clubs. These neighborhoods tend to be more expensive (majority are more expensive than the Toronto average). These tend to cluster near downtown and midtown. Most of these neighborhoods are geographically close together and are within 5km of an existing club. However, Postal codes *M8X, M8Y, M9A* which are along Bloor in Etobicoke seem to have the requisite affluence as well as being 8km or more away from existing facilities.

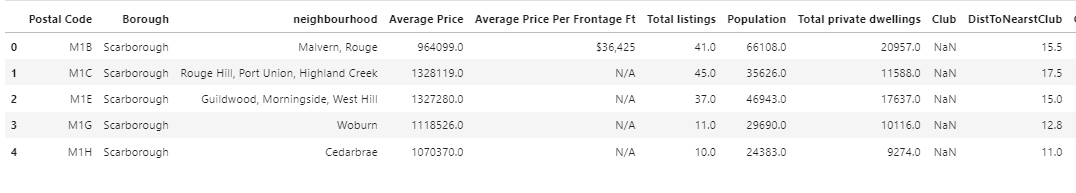


*Figure 5 Potential Club Cluster shows the first cluster that reasonably shows neighborhood traits suitable for another club these include M8X, M8Y, M9A.*

### 4 Clusters

As before, the Lawrence Park cluster from Figure 3 is largely unchanged. This indicates that these neighborhoods are pretty unique compared to the rest of Toronto. Similarly, the cluster from Figure 5 remains unchanged. The cluster from Figure 4 was divided with 5 neighborhoods on the edge of Scarborough becoming their own cluster.

The relevant clusters remained the same as the 3 cluster example.



*Figure 6 Easter Scarborough Cluster shows that the eastern edge of Scarborough, close to the Toronto Zoo, is relatively unique in Toronto, and likely wouldn’t support a club with Average Home prices much lower than average and low population density.*

### 5 clusters

Similar to above, most clusters stay the same. The one cluster to separate was the Potential Club Cluster shown in Figure 5. It now differentiates between the clubs that are right downtown which contains current clubs like *The Adelaide Club* and *Toronto Athletic Club.* These however are again geographically clustered and have existing club infrastructure.

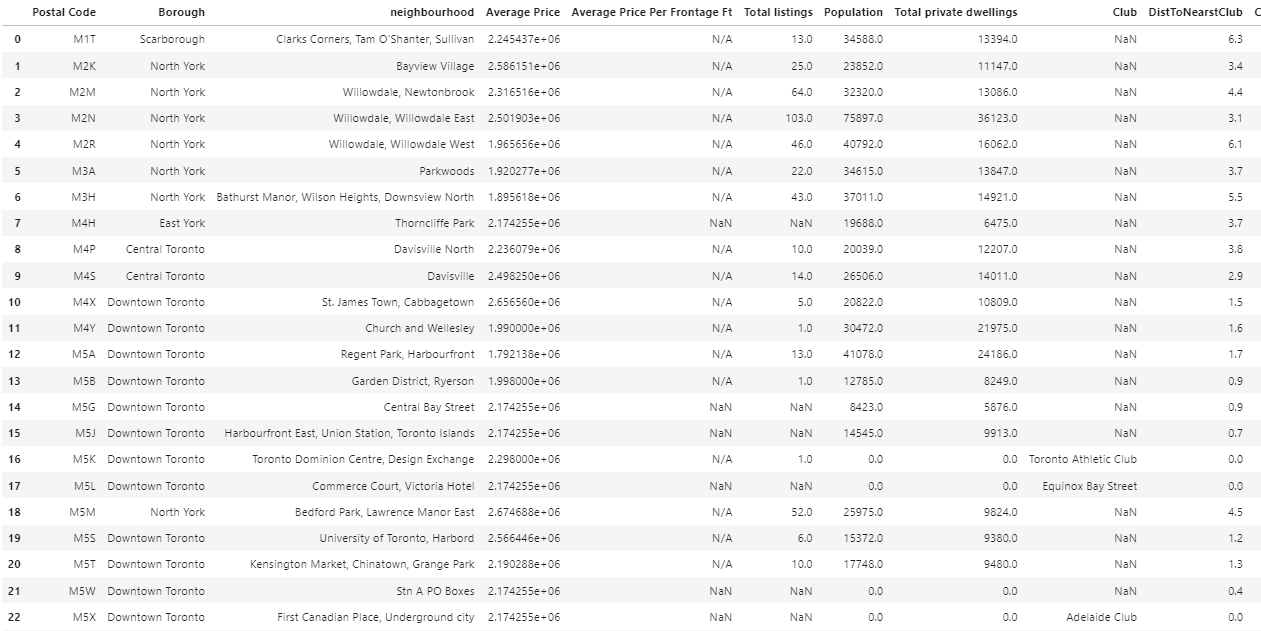


Figure 7 *Downtown clubs show neighborhoods that are similar to those that can support a downtown facility. Unfortunately almost all of them are right downtown and are close to existing clubs.*

The remaining clubs were sorted into a new cluster which seem to follow the more hybrid club model. These are mixed residential and business space and are more expensive than the average Toronto neighborhood. All except *M8X* and *M9A* are pretty close to existing clubs. But those two neighborhoods show promise. There is surrounding affluence and decent local population. In addition, they are on Bloor with an existing subway line, similar to the clubs in the in this cluster and there is a fair bit of commercial activity in that area. I think if I were to open a new club, this would be the area I would be interested in.

### 

*Figure 8 Hybrid Model Neighborhoods. These show neighborhoods that support the hybrid model of club, such as the Yacht club.*

# Conclusion

We analyzed every FSA in Toronto and came to the conclusion that we will advise our client that postal codes *M8X* and *M9A* are the most promising to open a new high-end fitness facility. These areas have a large affluent local base, but are still near commercial activity and are easily accessible being on the subway line. Furthermore, based on the clubs in the same cluster as these postal codes, the successful niche would be one of a hybrid model that tries to do work for both families as well as white-collar professionals on their way to or from work.