Analysis of Toronto Neighborhoods by Available Sports and Recreational Facilities

Lorenzo Felipe

April 1, 2021

1. Introduction

For most people moving to a new home, choosing the appropriate neighborhood is one of the most important things to think about. One thing a person must consider is the availability of different establishments and amenities within the area. For sporty or physically active people, having facilities such as a Basketball Court or a Yoga Studio is important sometimes even necessary.

What we would like to answer in this project is what neighborhoods are the most ideal place for a person who is physically active to live in? Furthermore, we would like to see which neighborhoods are similar in terms of the sports and recreational facilities that they can provide. For this project we would be considering all the neighborhoods in the City of Toronto.

2. Data

To solve the problem, data of all nearby venues for each neighborhood in Toronto will be extracted. Since the venues of interest are sports and recreational venues, only a portion of the Foursquare data will be used.

For the purpose of this project, we will only keep the column of the following recreational venues: 'Athletics & Sports', 'Baseball Field', 'Baseball Stadium', 'Basketball Court', 'Basketball Stadium', 'Beach', 'Climbing Gym', 'College Gym', 'College Rec Center', 'College Stadium', 'Curling Ice', 'Dance Studio', 'Dog Run', 'Escape Room', 'Field', 'Golf Course', 'Gym', 'Gym / Fitness Center', 'Hockey Arena', 'Indoor Play Area', 'Martial Arts School', 'Other Great Outdoors', 'Park', 'Playground', 'Pool', 'Skate Park', 'Skating Rink', 'Stadium', 'Swim School', 'Tennis Court', 'Trail', 'Yoga Studio'.

Figure 1: Sample Data

Out[6]:

|	Neighborhood	Athletics & Sports	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym	College Rec Center	College Stadium	Curling Ice	Dance Studio	Dog Run	Escape Room	Field	
0	Agincourt	0.000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
1	Alderwood, Long Branch	0.125	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
2	Bathurst Manor, Wilson Heights, Downsview North	0.000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
3	Bayview Village	0.000	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	

3. **Methodology**

For the analysis, we will be using k-means Cluster Analysis. First, we will choose the optimal number of clusters "k" by plotting using the Elbow Method. This method plots the SSE or Sum of squared errors vs the number of clusters and chooses the number of cluster with the greatest reduction in SSE. Second, we will use the optimal number of clusters "k" on the selected data. Finally, once Neighborhoods are already grouped by cluster, we will look at the common characteristics of each cluster and analyze which cluster is most ideal for a person or a family that is physically active to live in.

4. **Results**

Upon inspection of the data, 29 of the 100 neighborhoods listed have no nearby sports or recreational facility. We decided to remove these 29 neighborhoods so that they will not affect the cluster analysis.

To choose the appropriate number K that we will use for our Cluster analysis, we will create an Elbow Plot to determine the most optimal K for the data.
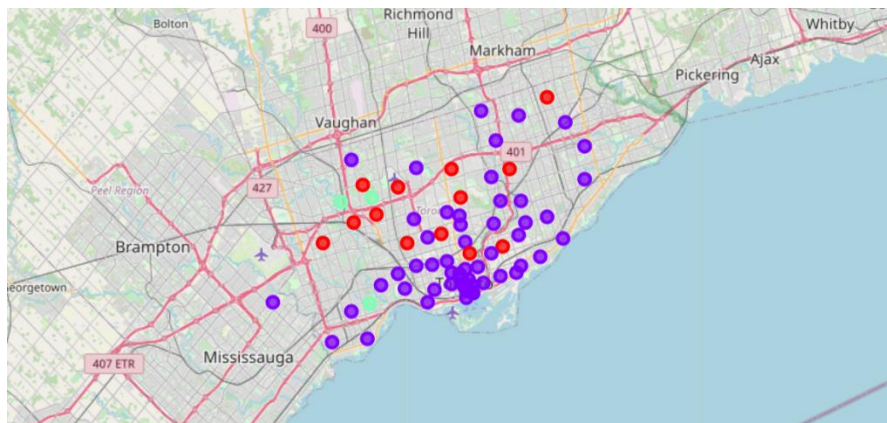
Figure 2: Elbow Plot
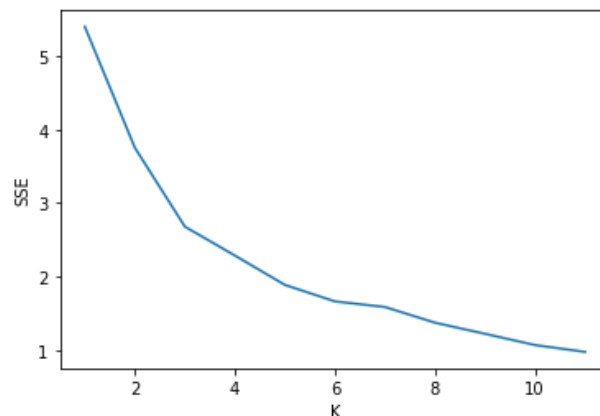
Based on the plot above, the most optimal K to use is K=3. However, some may also consider K=6 as an optimal elbow point. For this project, we will choose K=3 for our Cluster Analysis.

We have assigned a Cluster for each neighborhood. We will then plot the location of each neighborhood and color label them according to cluster.

Figure 3: Cluster Map

We will then check the contents of each cluster. We will then label the cluster depending on the characteristics of each.

Cluster 1: Park Cluster

The first cluster has 13 Neighborhoods. The common similarity for each neighborhood is that all of them have a high number of nearby Parks. We will be calling this Cluster the "Park Cluster" for simplicity

Figure 4: Park Cluster Sample Data

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North York	0.0	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
21	York	0.0	Park	Pool	Yoga Studio	Dog Run	Dance Studio	Curling Ice	College Stadium	College Rec Center	College Gym	Climbing Gym
35	East York/East Toronto	0.0	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
40	North York	0.0	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
46	North York	0.0	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
49	North York	0.0	Basketball Court	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Stadium	Beach	Climbing Gym	College Gym
61	Central Toronto	0.0	Park	Swim School	Yoga Studio	Dog Run	Dance Studio	Curling Ice	College Stadium	College Rec Center	College Gym	Climbing Gym
64	York	0.0	Park	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym

Cluster 2: Variety Cluster

The first cluster has 55 Neighborhoods. This is the cluster with the most number of neighborhoods. There are not much similarity within this cluster other than they have a variety of available sports and recreational venues nearby. We will then call this cluster the "Variety Cluster".

Figure 5: Variety Cluster Sample Data

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	North York	1.0	Hockey Arena	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
2	Downtown Toronto	1.0	Park	Yoga Studio	Gym / Fitness Center	Dog Run	Dance Studio	Curling Ice	College Stadium	College Rec Center	College Gym	Field
4	Queen's Park	1.0	Yoga Studio	Park	Gym	Dog Run	Dance Studio	Curling Ice	College Stadium	College Rec Center	College Gym	Field
7	North York	1.0	Gym	Yoga Studio	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym
8	East York	1.0	Athletics & Sports	Gym / Fitness Center	Field	Baseball Field	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym

Cluster 3: Baseball Cluster

The third cluster has 3 Neighborhoods. The common similarity in this cluster is that Baseball Field is the Number 1 Most Common Venue in all three neighborhoods. We will then label this Cluster as the "Baseball Cluster".

Figure 6: Baseball Cluster Sample Data

Out[20]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
53	North York	2.0	Baseball Field	Yoga Studio	Trail	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym	College Rec Center
57	North York	2.0	Baseball Field	Yoga Studio	Trail	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym	College Rec Center
101	Etobicoke	2.0	Baseball Field	Yoga Studio	Trail	Baseball Stadium	Basketball Court	Basketball Stadium	Beach	Climbing Gym	College Gym	College Rec Center

5. Discussion

In the very beginning of our analysis, we have already seen that there are 29 out of the 100 neighborhoods we checked that have no nearby sports and/or recreational venues. We can safely say that this 29 neighborhoods are not ideal for Physically Active people and/or families.

Among the 3 clusters that we were able to derive, the "Variety Cluster" has the most variety when it comes to available amenities. This cluster of neighborhoods would be good enough for people who do not have a particular sport or activity in mind and is ok with any physical activity.

Cluster 1 or the "Park Cluster" would be ideal for people who enjoy parks or any outdoor activity. There are plenty of activities that can be done in parks such as jogging, walking, biking or outdoor exercise. Furthermore, parks can be enjoy by people of all ages. Thus, this cluster would be suitable for families who are physically active since all members of the family can easily participate in Park activities.

The last cluster is the "Baseball Cluster" which has Baseball Fields as the most common sports venue. This cluster would be ideal for people and families who enjoy playing or watching baseball.

6. Conclusion

In conclusion, there are many neighborhoods in Toronto that are ideal for physically active or sporty families. Among the 100 Neighborhoods observed, 29 are not ideal for physically active or sporty families as they do not have any nearby sports or recreational venues. The result of our analysis is that 3 clusters is optimal. The three clusters that we have created are labelled as the Park Cluster, Variety Cluster and Baseball Clusters. Park Cluster is ideal for Physically Active Families while the Baseball Cluster is ideal for any Baseball loving individuals or families. The

Variety Cluster is ideal for anyone who do not have a particular sport or physical activity in mind as they have a variety of available sports or recreational venues.