**The Battle of the Neighborhoods**

**Introduction:**

1. **The problem:** a client wants to open a new gym in Toronto and wants to know the best location to do so. The location can mean the difference between having a successful business and a failing one. The location will determine how many potential clients there are in the area and how much competition the business will have to deal with.
2. **The data:** the Toronto neighborhood data will be collected once again from the Wikipedia page on Toronto postal codes. The coordinates will be collected from the Geocoder package. Using the FourSquare API, the number of gyms in each neighborhood can be collected. Clustering the neighborhoods will allow us to determine the best cluster to place our gym in.

**Methodology:**

1. Data collection (from Wikipedia, Geocoder, FourSquare):
   1. Neighborhood name, postal code from Wikipedia
   2. Latitude and longitude from Geocoder
   3. Venue data from FourSquare
2. Data cleaning:
   1. Reorganize/merge dataframes to get the best data possible, allowing us to plot the neighborhoods onto the map of Toronto:

Map

Description automatically generated

1. Clustering:
   1. Utilize elbow method to determine best K value (3 in this case):

Chart, line chart

Description automatically generated

* 1. Using KMeans clustering to get 3 clusters:

A picture containing map

Description automatically generated

**Results:**

The main results can be summarized by 2 figures:

Chart, box and whisker chart

Description automatically generated

**Discussion:**

We can see that in cluster 1, we have the largest number of neighborhoods with the lowest number of average gyms. We therefore select a neighborhood from cluster 1 to build a new gym in. One possible location to focus on is to the south of the map as it has a higher density of red dots. This could allow the gym to be in a more convenient location for these neighborhoods.

**Conclusion:**

This project has extending what was done previously to answer the question of optimal location for gyms in Toronto. This method can be applied to various other questions involving the opening of new businesses. It can also be applied to any location in the world with datasets like the ones utilized here.