**Introduction/Business Problem**

New York City is a hub of business, culture, industry, and much more. It’s not everybody’s ideal preference to live in such a big, busy city, but for some it is a place one cannot live without. Within it, Manhattan is at the heart. It contains iconic landmarks such as the Empire State Building, Times Square, and Broadway.

If I were to ever move to Manhattan, I would not be able to survive in the Concrete Jungle without some open areas to roam. There is a plethora of busy city streets filled with vendors, speed walkers who won’t look up to avoid you, and traffic jams galore. With running as my primary method of keeping fit, I would need parks, trails, scenic areas, etc. to maintain sanity.

**Data**

For practice of data gathering and manipulation, as well as utilizing machine learning techniques, I will be looking for the healthiest neighborhoods of Manhattan for which to settle. I will leverage the Foursquare API to get all venues in Manhattan neighborhoods in order to extract from it venues that fit the ‘Outdoors and Recreation’ category as well as a few others I have chosen that were not in that category that to me would fit in ‘recreation.’



(Manhattan neighborhood map)

**Methodology**

After gathering the initial venue data for each neighborhood in Manhattan, I subset the data to only include those in the Health categories I have chosen (‘Outdoors and Recreation’ plus my additional chosen venue types) and used *k*-means clustering to determine what groups of neighborhoods will fit my preferences the best. Below is a sample of the dataframe I used to run *k*-means (with 5 clusters).

****

And here is a snapshot of the resulting dataframe that lists each neighborhood, along with the cluster label and its most common venues.



**Results/Discussion**



(Cluster 1)

The ensuing results of the clustering was very intuitive (if not fairly meaningless). It appears that subsetting our venue dataset to include only venues that can be considered ‘outdoors and recreation’ did not *really* help us in determining which neighborhoods can be the 'healthiest' in this way. Clusters 1 is our only meaningful cluster. That is to say that Clusters 2 through 5 all, at the very least, have the same 5th-10th most common venue type. This indicates to me that within those clusters, they each have very few venues AND they happen to share those venues.

I followed up by taking a closer look at Cluster 1 to see what I can find out.



Besides having similar venues according to the *k*-means clustering algorithm, what set this cluster apart from the others is that they generally have at least 5 venues. Looking at this data, it's easy to say something like, "Why not just look at the neighborhoods that have the most venues in the outdoors and recreation category?" and I'd have to agree with that sentiment. It certainly didn't make for very meaningful analysis in this case, as we could have simply chosen Battery Park with 17 venues. However, it's good to know that there are many other options to look at that have a substantial amount of recreational activies to do. Many if not all of these neighborhoods can lead to living a healthy lifestyle in Manhattan.

**Conclusion**

Just to have a bit more closure, let's take a look at the neighborhood with the most venues, Battery Park, and see what other context we can gather to indicate its health level.

Here's an excerpt from Battery Park City's Health section of its Wikipedia page (<https://en.wikipedia.org/wiki/Battery_Park_City#Health>):

* The concentration of fine particulate matter, the deadliest type of air pollutant, in Battery Park City and Lower Manhattan is 0.0096 milligrams per cubic metre (9.6×10−9 oz/cu ft), more than the city average. Sixteen percent of Battery Park City and Lower Manhattan residents are smokers, which is more than the city average of 14% of residents being smokers. In Battery Park City and Lower Manhattan, 4% of residents are obese, 3% are diabetic, and 15% have high blood pressure, the lowest rates in the city—compared to the citywide averages of 24%, 11%, and 28% respectively. In addition, 5% of children are obese, the lowest rate in the city, compared to the citywide average of 20%.
* Ninety-six percent of residents eat some fruits and vegetables every day, which is more than the city's average of 87%. In 2018, 88% of residents described their health as "good," "very good," or "excellent," more than the city's average of 78%. For every supermarket in Battery Park City and Lower Manhattan, there are 6 bodegas.

So we have some bad with the good in Battery Park. People seem to eat healthier and there is a lower rate of obesity, diabetes, and highblood pressure that elsewhere in the city. This could be in some way aided by the higher number of outdoors and recreaction venues. However, there is a higher rate of smokers as well as a higher concentration of air pollutants than the city average. If it were me looking for a new place to live, as a runner, I would have a hard time with the air pollutants and smoking, so it really comes down to personal preference here. Luckily there are plenty of other options from which to choose.