Data Science Capstone Report

**Introduction/Business Problem**

Looking at the clustered neighbourhoods in Toronto, I focused on one of my clusters in which every neighbourhood had a park as its most or second most frequent venue. These neighbourhoods also had many playgrounds and dog parks, and often had schools. This cluster seems to be mainly suburban areas. I also noticed that every neighbourhood in this Toronto had a yoga studio in the top four most common venues, making yoga the most popular business.

My idea is to apply this clustering to other cities and find the suburban areas that don’t have many yoga studios. Any neighbourhood in the suburban cluster that doesn’t have a yoga studio in its top 7 most common venues will be recommended as a good area to open a yoga studio.

**Data**

This problem will be solved using foursquare location data to find the common venues in each of a city’s neighbourhoods. The neighbourhoods will be clustered based on common venues. I hope to automate the process of detecting which cluster contains the suburban areas by looking for parks and yoga studios in the most common venues, then pick out neighbourhoods in the cluster that have an unfilled niche for yoga studios, where a new one might have little competition and high demand.